ISSN 2359-0386

5ª EDIÇÃO

ERI·GO

ESCOLA REGIONAL DE INFORMÁTICA

ANAIS

ANAIS

2017

Goiânia - GO

Dados Internacionais de Catalogação na Publicação (CIP)
GPT/BC/UFG
Bibliotecária responsável: Adriana P. Aguiar / CRB1: 3172

E74 Escola Regional de Informática de Goiás (05. : 2017 : Goiânia, GO).
Anais da V Escola Regional de Informática de Goiás [recurso eletrônico] / Organizadores, Antonio Carlos de Oliveira Júnior, e Vinicius Cunha Martins Borges. – Dados eletrônicos. – Goiânia : Instituto de Informática, UFG, 2017.
399 p. : il.

ISSN: 2359-0386
Disponível também em World Wide Web:
< http://erigo.sbc.org.br/p/152-anais-eri-go>

1. Redes sociais – Relações humanas. 2. Internet - Twitter. 3. Bullying – Assédio virtual. I. Oliveira Júnior, Antonio Carlos de. II. Borges, Vinicius Cunha Martins. III. Título.

CDU: 004

COMITÊ

ORGANIZAÇÃO GERAL LOCAL

Antonio Carlos de Oliveira Júnior (UFG/Goiânia) - Coordenador
Marcelo Akira Inuzuka (UFG/Goiânia) - Vice-Coordenador

COORDENAÇÃO DO COMITÊ CIENTÍFICO

Vinicius da Cunha M. Borges (UFG/Goiânia) - Coordenador
Luciana de Oliveira Berretta (UFG/Goiânia) - Vice-Coordenadora
Renato de Freitas Bulcão Neto (UFG/Goiânia)

COMITÊ DE AVALIAÇÃO TÉCNICA E CIENTÍFICA

Adailton Araújo (UFG/Goiânia)
Adriano Braga (IFGoiano/Ceres)
Alessandro Silva (IFG/Anápolis)
Alex da Rosa (UnB)
Alisson Cardoso (EMC/UFG)
Ana Claudia Loureiro Monção (UFG/Goiânia)
Antônio Oliveira Jr (UFG/Goiânia)
Cleon Pereira Júnior (IFGoiano/Posse)
Cristiane Bastos Rocha Ferreira UFG/Goiânia)
Daniel Stefany Duarte Caetano (UFU)
Edison Morais (UFG/Goiânia)
Edmar Yokome (UEG/Santa Helena)
Eliomar Lima (UFG/Goiânia)
Fabian Cardoso (Fesurv/Rio Verde)
Fernando Federson (UFG/Goiânia)
Flávio Vieira (EMC/UFG)
Flávio Geraldo Coelho Rocha (EMC/UFG)
Francisco Xavier (Fatesg)
Gilmar Arantes (UFG/Goiânia)
Iwens Sene Jr (UFG/Goiânia)
Juliano Lopes de Oliveira (UFG/Goiânia)
Leandro Freitas (IFG/Inhumas)
Leonardo Andrade Ribeiro (UFG/Goiânia)
Luciana Berretta (UFG/Goiânia)
Luiz Henrique Rauber Rodrigues (Senac/Santa Cruz do Sul/RS)
Marcelo Akira Inuzuka (UFG/Goiânia)
Marcelo Ricardo Quinta (UFG/Goiânia)
Marcos Aurélio Batista (UFG/Catalão)
Marcos Caetano (UnB)
Marcos Wagner de Souza Ribeiro (UFG/Jataí)
Marcus Vieira (EMC/UFG)
Marcus Vinícius Gonzaga Ferreira (UFG/Goiânia)
Maria José P. Dantas (PUC-Goiás)
Márcio Dias (UFG/Catalão)
Nádia Silva (UFG/Goiânia)
Otávio Calaça Xavier (IFG/Goiânia)
Renato de Freitas Bulcão Neto (UFG/Goiânia)
Ricardo Rocha (UFG/Catalão)
Rodrigo Cândido Borges (IFG/Inhumas)
Rogério Salvini (UFG/Goiânia)
Ronaldo Del-Fiaco (UEG/Anapólis)
Rogério Silva (IFG/Inhumas)
Roney Lima (IFG/Jataí)
Rosana Santos (UFG/Goiânia/Tribunal)

Rubens Pereira (UFG/Goiânia)
Sérgio Teixeira de Carvalho (UFG/Goiânia)
Taciana Novo Kudo (UFG/Goiânia)
Telma Woerle Lima Soares (UFG/Goiânia)
Tércio Alberto Santos Filho (UFG/Catalão)
Thales Nicolai Tavares (UFSM, RS)
Thierson Rosa (UFG/Goiânia)
Vinicius da Cunha M. Borges (UFG/Goiânia)
Vinicius Garcia (UFSM, RS)
Vinicius Sebba Patto (UFG/Goiânia)
Vinicius Petrucci (UFBA/Salvador)
Wanderley Alencar (UFG/Goiânia)
Wendell Bento Geraldes (IFG/Luziânia)
William Ferreira (UFG/Goiânia)

REVISORES EXTERNOS
Ernesto Veiga (INF/UFG)
Leonardo Morais (INF/UFG)
Mayke Arruda (INF/UFG)
Rafael Tomaz Parreira (IFG/Jataí)

E-mail de Contato: erigo@inf.ufg.br

DIRETOR DO INSTITUTO DE INFORMÁTICA - UFG

Sérgio Teixeira de Carvalho

Apresentação

A 5ª Edição da Escola Regional de Informática de Goiás (ERI-GO 2017) é um evento promovido pela UFG (Instituto de Informática), IFGoiano e IFG. A ERI-GO 2017 foi realizada em conjunto com a Escola Regional de Sistemas de Informação de Goiás (ERSI-GO) e com o Fórum Goiano de Software Livre (FGSL), contribuindo assim para uma maior integração entre a comunidade goiana de pesquisadores e discussão de temas relevantes nas áreas de Informática em geral, aproximando estudantes, pesquisadores, profissionais e empresários. Esta edição do evento tem como objetivos: estimular ações de pesquisa, ensino e extensão sobre informática em geral nas instituições de ensino superior e técnico como também centros de pesquisa no Estado de Goiás, baseado em trilhas dos principais cursos de Informática em Goiás, no Brasil e no mundo de acordo com ACM (Association for Computing Machinery), por exemplo, Sistema de Informação, Engenharia de Software, Ciência da Computação e Engenharia a fins; tornar público os projetos de SL desenvolvidos regionalmente; e promover a cooperação entre produtores e usuários de software livre visando ampliar o desenvolvimento tecnológico e a inovação em soluções de SL no Estado. A ERI-GO 2017 ocorreu nos dias 16,17 e 18 de novembro de 2017, em Goiânia (Goiás). 21 (Vinte e Um artigos) completos (apresentação oral), 5 (cinco) resumos expandidos (apresentação de poster) e 7 (sete) artigos curtos (apresentação de poster) foram incluidos nos anais selecionados pelo comitê de avaliação técnica e científica para a ERI 2017, os quais tratam de vários temas de pesquisa e desenvolvimento científico e tecnológico em Software Livre e Informática de forma geral. Além disso, os 3 (três) melhores artigos do evento foram selecionados e premiados, 1º Lugar com artigo intitulado "Localização Usando Beacons em Ambientes Internos para Monitoramento Remoto de Pacientes", 2º Lugar com artigo intitulado "Aplicação de Algoritmos Genéticos para Estimação de Parâmetros de Painéis Fotovoltaicos" e 3º Lugar com artigo intitulado "Análise, Implementação e Simulação de Propriedades da Cicloide em Software Livre de Geometria Dinâmica para o Ensino de Matemática", a classificação levou em conta a avaliação da apresentação e a avaliação dos revisores. Os anais também estão disponibilizados eletronicamente nos sites: http://www.inf.ufg.br/~erigo/4/o/anais-v-erigo-2017.pdf e http://erigo.sbc.org.br/p/152-anais-eri-go.

Sumário

Artigos Completos 11

11 Aplicando técnicas de aprendizado de máquina para detecção automática de bullying no Twitter
Gaspar Silva
Nádia Silva
Márcio Dias

25 Métodos para análise de sentimentos em tweets: um estudo comparativo
Alison Ribeiro
Nádia Silva

39 Avaliação do Distúrbio Disfórico Pré-Menstrual e Transtorno Afetivo Bipolar uma abordagem usando Regras de Associação
Giovanna Castro
Rogerio Salvini
Rodrigo Dias

53 Otimização de Interligação de Religadores por Fibra Óptica para Automação de Redes Auto-recuperáveis de Energia
Marcos de Carvalho
Maykon Renan Pereira da Silva
Flávio Henrique Teles Vieira

67 Aplicação de Algoritmos Genéticos para Estimação de Parâmetros de Painéis Fotovoltaicos
Gilberto Filho
Ricardo Franco
Flávio Henrique Teles Vieira

79 Modelo de Simulação de Sistema Fotovoltaico Ligado à Rede com Redução da Injeção de Harmônicos via Algoritmo Fuzzy
Henrique Correa
Ricardo Franco
Flávio Henrique Teles Vieira
Marcelo Stehling de Castro

93 Otimização da Alocação de Recursos em Redes Sem Fio com Agregação de Portadora e F-OFDM Utilizando Programação Evolucionária
Marcus Vinícius Gonzaga Ferreira
Flávio Henrique Teles Vieira

105 Algoritmo de Alocação de Recursos de Rede baseado na Função de Autocorrelação de um Modelo de Tráfego Multifractal
Maykon Renan Pereira da Silva
Flávio Geraldo Coelho Rocha
Flávio Henrique Teles Vieira

119 Avaliação Comparativa de Tecnologias de Modulação para redes 5G baseadas em Sistemas LTE
Dalton Foltran de Souza
Vinicius da Cunha Martins Borges
Flávio Henrique Teles Vieira
Diego Abrahão

133 Algoritmo Adaptativo Baseado em Modelo Autorregressivo Transformado para Geração de Séries Sintéticas de Tráfego de Rede
Alisson Cardoso
Flávio Henrique Teles Vieira

147 Estimação de Frequências em Sinais Eletroencefalográficos Utilizando Autovalores
Lucas Fiorini Cruz
Marcela Magalhães
Rodrigo Pinto Lemos

159 Localização Usando Beacons em Ambientes Internos para Monitoramento Remoto de Pacientes
Sergio T. Carvalho
Ana Paula de Sousa Oliveira

173 Análise, Implementação e Simulação de Propriedades da Cicloide em Software Livre de Geometria Dinâmica para o Ensino de Matemática
Rosário Ribeiro
Antônio Oliveira Jr

185 Uso de Softwares Educacionais como Recurso Didático-Pedagógico para o Ensino da Geografia
Raimundo N. de A. S. Neto
Tiago Nogueira
Sônia Regina Rezende

199 Ambientes Virtuais Tridimensionais no Processo de Ensino-Aprendizagem de Libras: Uma Revisão Sistemática da Literatura
Tiago Nogueira
Emily Mendonça
Júlia Araújo
Júlia Moura
Deller Ferreira

211 Esteganografia, A Escrita Oculta: Um Mapeamento Sistemático
Hayanna Oliveira
Leonardo Silva
Rafael Alves
Gilton Silva

223 Gestão de Continuidade do Negócio (GCN): Normas, Recomendações e Modelos
Mariane Batista
Edison Morais
Ana Claudia Loureiro Monção

237 Modelagem ontológica de informações meteorológicas para aplicações de IoT
Murilo Barcelos Silva
Renato Bulcão Neto

251 Uma Abordagem para Garantia da Consistência de Middleware Específico de Domínio Dirigido a Modelos
Marcelo Fortes
Fabio Costa

265 IFLivros – Sistema de Gerenciamento de Livros para o IFG Câmpus Luziânia
Ernane Martins
Jorge Tawnan Leles Vilela
Reinoldo Oliveira Sommer

279 Uma abordagem de entrega contínua para empresas de desenvolvimento de software
Rodolfo Bueno
Adailton Araújo

Resumos Expandidos 293

293 Balanceamento de Carga de Rede: Alto Desempenho e Escalabilidade para Serviços Computacionais
Raimundo N. de A. S. Neto
Tiago Nogueira
Cláudio Silva

303 Extração De Edificações Em Imagens de Baixa Resolução Para Atualização De Mapa Urbano Básico
Paulo Victor Dos Santos
Dyego Almeida
Cristhiane Gonçalves
Ronaldo Costa
Leandro Oliveira

313 Aprimorando a Plataforma UbiCare por meio da Integração do Plano de Cuidados Ubíquo e do Coletor de dados Vitais via Serviços Web
Natália Junqueira
Sergio T. Carvalho

323 Uma Linguagem de Modelagem para o Desenvolvimento de Aplicações Ubíquas em um Contexto de Assistência Domiciliar à Saúde
Helberth Borelli
Sergio T. Carvalho

333 Aplicação de mineração de dados para identificação de evasão de clientes
Ygor Escatolim Castor Santana
Nádia Silva
Walison Cavalcanti

Artigos Curtos 343

343 Computação Evolucionária Aplicada à Otimização de Trajetos e Alocação de Torres de Linhas de Transmissão de Energia Elétrica
Caio Jose Povoa
Pedro Henrique Palhares
Leonardo Brito

351 Identificação de surtos de altas frequências em sinais de Eletroencefalografia utilizando Transformada de StockWell e Decomposição em Valores Singulares
Marcela Magalhães
Lucas Fiorini Cruz
Rodrigo Pinto Lemos

359 Using Sentiment Analysis to Real-time Forecast Box Office Revenue
Rodrigo Sasse David
Deborah Fernandes
Márcio Giovane Cunha Fernandes

367 Requisitos Essenciais de Usuários de um Sistema de Informação em Saúde Bucal para a Atenção Básica
Beatriz Silveira
Fabio Lucena
Renata Braga
Rejane Ribeiro-Rotta

375 Derivando Requisitos de Exigências para Certificação de Sistemas de Informação em Saúde
Leonardo Neves
Fabio Lucena
Renata Braga
Rejane Ribeiro-Rotta

383 BPM das Coisas Uma abordagem prática
Walmir Melo
Gilmar Arantes

391 EasYii Gii, Um Gerador de código para aplicações MVC, em PHP, baseado em banco de dados relacionais
Matheus Morais
Alessandro Silva

Index of Authors 399

Aplicando técnicas de aprendizado de máquina para detecção automática de *bullying* no Twitter

Gaspar Monteiro e Silva[1], **Nádia Félix F. da Silva**[1], **Márcio de Souza Dias**[2]

[1]Instituto de Informática – Universidade Federal de Goiás – Goiânia, GO – Brasil

[2]Departamento de Ciência da Computação
Universidade Federal de Goiás – Catalão, GO – Brasil

gaspar.monteiro@gmail.com,nadia@inf.ufg.br,marciosouzadias@ufg.br

Abstract. *Machine Learning techniques can be used to automatically infer non-trivial information from large datasets. Social networks have become increasingly popular. As a consequence, they are becoming an important source of data to the application of Machine Learning techniques. This paper intends to show how when applying these techniques to a dataset extracted from Twitter and manually annotated, it is possible to automatically detect when a Tweet contains what we describe as a bullying trace. Several configurations of classifiers and feature representations were studied and compared in order to assess which one is the best to this task.*

Resumo. *Técnicas de aprendizado de máquina podem ser utilizadas para inferir automaticamente informações não triviais a partir de grandes quantidades de dados. As redes sociais vêm ganhando popularidade e tornando-se importantes fontes de dados para a aplicação de tais técnicas computacionais. O objetivo deste trabalho é mostrar que aplicando técnicas de aprendizado de máquina em uma base de dados extraída do Twitter é possível detectar automaticamente se um texto possui ou não um traço de bullying. Vários classificadores e conjuntos de atributos diferentes foram estudados e comparados, com o objetivo de identificar qual é o mais apropriado para esta tarefa.*

1. Introdução

A prática de *bullying* consiste em ações negativas persistentes por parte de colegas (daí seu outro nome, *peer victimization*), sejam eles de escola, de trabalho ou de qualquer outro ambiente social [Olweus 1994]. Este assédio pode ser físico, verbal ou relacional [Nylund et al. 2007, Archer and Coyne 2005]. Os episódios geralmente são face a face, porém instâncias de *bullying* em meios virtuais vêm gradativamente se proliferando, estimuladas principalmente pelas interações entre usuários de redes sociais, tais episódios são conhecidos como *cyberbullying* [Slonje and Smith 2008].

As consequências do *bullying* ou *cyberbullying* para as vítimas vão de um maior risco de desenvolver doenças relacionadas ao estresse, males psicológicos, como transtornos emocionais e de personalidade, a até mesmo suicídio [Rigby 2001]. Para a sociedade como um todo, o *bullying* pode implicar em episódios de violência, como tiroteios em escolas [Leary et al. 2003] que são particularmente comuns nos Estados Unidos. No Brasil também foram notificados casos de tiroteios em escolas motivados de alguma forma

por *bullying*, como o Massacre de Realengo [UOL Notícias 2011] e o caso do atirador de Taiúva [Folha de SP 2003]. Em 2009, uma pesquisa realizada pelo Instituto Brasileiro de Geografia e Estatística (IBGE), denominada de Pesquisa Nacional de Saúde Escolar [Malta et al. 2010], concluiu que 5.4% dos estudantes relataram ter sofrido *bullying* com alta frequência nos últimos 30 dias, enquanto que 25.4% mencionaram episódios com uma periodicidade menor.

A lei $N^o13.185$ de 6 de novembro de 2015 [Presidência da República 2015] foi instituída no Brasil para tentar combater o *bullying* em escolas e agremiações, e um dos seus objetivos é integrar os meios de comunicação de massa com as escolas e a sociedade, como forma de identificação e conscientização do problema. As redes sociais, um desses meios de comunicação de massa, são veículos para persuasão e formação de opiniões, sendo fortemente usadas para difusão de conteúdo.

Utilizar da grande quantidade de dados disponíveis nas redes sociais para estudar casos de *bullying* é um campo de estudo promissor, apesar de ainda pouco explorado. O crescimento da *Internet* e do conteúdo que é gerado por seus usuários, que publicam suas opiniões em uma linguagem coloquial e em muitos casos utilizando de artifícios gráficos, para tornar ainda mais sucintos seus diálogos é um importante desafio. Esse cenário é observado no Twitter[1], uma ferramenta de comunicação que pode facilmente ser usada como fonte de informação para ferramentas automáticas de detecção de *bullying*.

Os trabalhos encontrados na literatura ([Xu et al. 2012a] e [Xu et al. 2012b]) abordam o problema para língua inglesa sob o ponto de vista de algoritmos de Aprendizado de Máquina Supervisionados (classificadores), e não apresentam comparações ou extensões para outros idiomas, como por exemplo o Português do Brasil. Desafios importantes de pesquisa tanto do ponto de vista de Processamento de Língua Natural(PLN), quanto da aplicação de técnicas de Aprendizagem de Máquina(AM) e Análise Léxica(AL) são inerentes à língua. Por este motivo, este trabalho visa nortear futuros estudos para a língua portuguesa que utilizem o Twitter, ou mesmo outras redes sociais para o estudo de *bullying*. Especificamente, é possível enumerar como contribuições deste estudo: (i) a criação de um córpus, construído a partir do Twitter, contendo textos que possuam traços de *bullying* para o português, bem como a identificação do papel do escritor neste texto (podendo o autor ter o papel de vítima, relator ou praticante do ato de *bullying*); (ii) a aplicação de técnicas de aprendizado de máquina para modelagem computacional do problema de inferência e detecção de textos contendo traços de *bullying* em redes sociais; e (iii) uma vez identificado que tal texto se trata de um *traço de bullying*, prover a identificação automática do papel do escritor (classificar o autor do texto em: vítima, praticante ou relator/narrador de um episódio de *bullying*).

Este trabalho está organizado conforme segue: Na Seção 2 são apresentados os principais trabalhos relacionados. Na Seção 3 são definidos os papéis das pessoas envolvidas em episódios de *bullying*, bem como a fundamentação teórica. Na Seção 4, é apresentado brevemente como o córpus foi criado, os experimentos realizados, os métodos e ferramentas utilizadas, além dos resultados dos experimentos realizados. A Seção 5 apresenta algumas considerações finais e indicações de trabalhos futuros.

[1]https://twitter.com/

2. Trabalhos relacionados

Como mencionado anteriormente, [Xu et al. 2012a] e [Xu et al. 2012b] abordam o problema de detecção de traços de *bullying* para a língua inglesa. Os autores anotaram um córpus contendo *tweets* que potencialmente possuíam traços de *bullying* e usaram ferramentas de Processamento de Linguagem Natural e de Aprendizado de Máquina para obter um classificador para detectar automaticamente instâncias de *bullying*. Usando SVM linear [Tan et al. 2005] juntamente com *unigramas e bigramas*[2], foi obtido uma acurácia de 81.3% na detecção de traços de *bullying* para a língua inglesa. É importante citar que não foi feito nenhuma radicalização de palavras(utilização das partes básicas de uma palavra) e as *stopwords*(palavras sem conteúdo, por exemplo, preposição, conjunção etc) não foram removidas. O córpus utilizado continha 1.762 *tweets*, sendo destes, 39% anotados com a classe "contém traços" ou "não contém traços" de *bullying*. Outro experimento realizado pelos autores destes artigos foi classificar os *tweets* conforme o papel do escritor do *tweet*. Usando SVM linear juntamente com unigramas e bigramas, o classificador obtido alcançou uma acurácia de 61%.

Em [José et al. 2012], o autor realizou a coleta de *tweets* relacionados a professores, usando os termos de pesquisa "meu professor" e "minha professora". Após pré-processar os *tweets*, 300 *tweets* foram anotados como Positivo, Negativo ou Neutro. Estes *tweets* anotados foram usados para treinar um classificador usando Naive Bayes [Tan et al. 2005]. Tal classificador obteve uma taxa de acurácia de 87.1%. Apesar do artigo não explicitar como foi feita a anotação e não prover uma matriz de confusão, além de ter um conjunto de dados relativamente pequeno (300 *tweets*), os resultados do classificador se mostraram promissores o bastante para demonstrar que usar classificadores ao lidar com *tweets* em português pode ser uma alternativa viável para a análise de *bullying*.

Em [Bellmore et al. 2015], os autores utilizam do grande volume de *tweets* sobre *bullying* para tentar responder algumas questões: Quem são os autores de relatos? O que eles relatam? Porque o fazem? De onde partem os relatos e quando eles são feitos? Foram analisados 9.764.583 *tweets* usando técnicas de aprendizado de máquina para prover respostas a estas perguntas.

As redes sociais criaram um ambiente propício ao *cyberbullying*. Logo vários estudos foram feitos para entender melhor como o *cyberbullying* funciona. Em [Nalini and Sheela 2015], os autores utilizam técnicas de aprendizado de máquina para detectar automaticamente instâncias de *cyberbullying* no Twitter utilizando um esquema de peso baseado em seleção de atributos (*feature selection*).

Em [Chatzakou et al. 2017], os autores utilizam além dos textos dos *tweets*, metadados, como número de seguidores, número médio de postagens diárias, data da última vez postada, data da criação de conta, dentre outros, para extrair características de *cyberbullies* no Twitter. Entre as conclusões apresentadas, estão o fato de que usuários agressivos mostram um comportamento similar ao de *Spammers*(usuários que enviam grandes quantidades de mensagens repetidas, geralmente de conteúdo mercadológico), por exemplo, em termos de número de seguidores, amigos e tamanho da rede. Tais usuários não postam muitos *tweets* e usam poucas *hashtags*(#) e URLs em seus *tweets*. Além disso,

[2]Unigramas são termos comportos por apenas uma palavra extraídos dos textos utilizados nas análises, enquanto que bigramas são termos que possuem duas palavras.

estes usuários costumam estar no Twitter há muito tempo. Por outro lado, os usuários comuns são bastante populares, considerando-se número de seguidores, amigos e tamanho da rede. Eles participam de muitas comunidades e usam um grande número de *hashtags* e URLs.

Os trabalhos relacionados apresentados nesta seção evidenciam o grande relevância do objeto de estudo neste artigo. Assim, para a língua portuguesa é necessário prover um córpus que funcionará como um *benchmark* para detecção de traços de *bullying* em português, bem como análises fundamentadas e comparativas dos principais algoritmos de classificação aplicados à estes dados, e cuja a acurácia já foi comprovada para a língua inglesa, mas não comprovada para o Português do Brasil.

3. Identificação de traços de *bullying*

Neste trabalho, se entende como traço de *bullying* um texto que mencione um episódio de *bullying* sofrido por alguém, ou um indivíduo relatando que alguém sofreu *bullying* em algum momento, mesmo que não detalhe o episódio, ou ainda instâncias de *cyberbulling*. Esta distinção é necessária pois é comum usuários de redes sociais se referirem a *bullying* de forma jocosa ou irônica. Há também casos onde o termo *bullying* é usado de forma diferente das definições por este trabalho adotadas. Por exemplo, o *tweet "Agora q o Mexico tá sofrendo bullying econômico do Trump, ele vai procurar um psicólogo e depois desabafar fazendo textao no facebook"*, menciona que o atual presidente americano estaria fazendo *bullying* com o México. Definimos acima *bullying* como uma ação de indivíduos sobre indivíduos, portanto este *tweet* não se encaixa como contendo traço de *bullying*.

3.1. Papéis presentes em um episódio de *bullying*

Os indivíduos que participam de um episódio de *bullying* possuem papéis específicos [Xu et al. 2012a]. A classificação de participantes segundo estes papéis é importante, pois cada participante do grupo tem uma função distinta no episódio, e a junção de todos os envolvidos é o que torna a prática de *bullying* possível. Assim, com o conhecimento destes papéis é possível trabalhar e contribuir para o combate ao *bullying* efetivamente [Salmivalli 1999].

Salmivalli, em [Salmivalli 1999], listou quatro papéis dos participantes de um episódio de *bullying* em adição ao ***Bully*** (plural ***Bullies***) e a ***Vítima***: (i) ***Assistente***, (ii) ***Reforçador***, (iii) ***Espectador*** e (iv) ***Defensor***. Para o autor, ***bullies*** são os indivíduos que ativamente praticam a ação. Enquanto que as ***vítimas*** são os indivíduos aos quais a ação é direcionada. Os ***assistentes*** são indivíduos que não iniciam um episódio, mas se juntam ao *bully* no ato. ***Reforçadores*** não se juntam diretamente ao episódio, mas reforçam os *bullies* através de *feedback* positivo (Rindo, por exemplo). Os ***espectadores*** presenciam o episódio mas não participam do mesmo. Enquanto que os ***defensores*** ativamente ajudam as vítimas, confrontando *bullies* e confortando a *vítima*.

Em [Xu et al. 2012a], os autores adicionaram dois papéis quanto ao estudo de *bullying* em redes sociais: ***Relator*** e ***Acusador***. Segundo os autores estes papéis são necessários pois em muitos casos não fica claro se o escritor do *tweet* é uma vítima, espectador ou defensor. Os **Relatores** relatam um episódio ocorrido. Podem ou não estarem presentes durante o episódio, enquanto que os **Acusadores** acusam alguém de cometer *bullying*.

Dada a complexidade de identificação dos papéis listados em [Xu et al. 2012a],[Salmivalli 1999], principalmente pela natureza subjetiva da tarefa de anotação neste cenário (além da dificuldade de se identificar *tweets* com papéis muito específicos), neste trabalho são considerados apenas os papéis mais comuns: **Vítima**, **Bully** e **Relator** (Veja Figura 1).

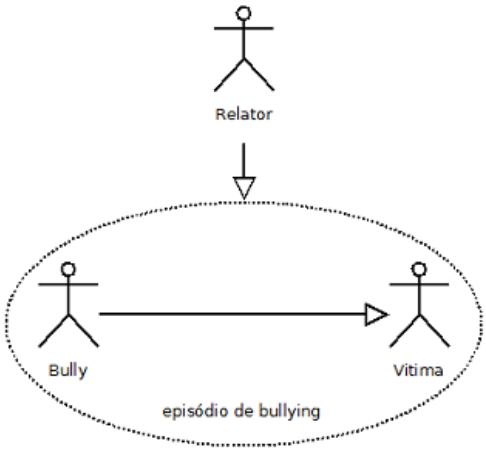


Figura 1. Um episódio de *bullying*

4. Classificação automática de traços de *bullying*

Pessoas que presenciam ou vivenciam um episódio de *bullying* comumente reportam ou comentam o mesmo em redes sociais, o que faz da extração de traços de *bullying* em redes sociais possível.

O Twitter foi a rede social escolhida para este estudo, pois apesar de suas limitações em número de caracteres, a mesma possui várias vantagens. Sua API[3] permite facilmente coletar uma grande quantidade de dados usando palavras chave e além disso é possível monitorar trocas de mensagem e viralizações através de *retweets*[4] e *replies*[5] – Caso um *tweet* seja resposta a outro, e este contenha traço de *bullying*, é possível analisar o *tweet* original. Pode-se também usar os dados da quantidade de *retweets* para fazer um estudo sobre viralização de *tweets* com traço de *bullying*, ou até mesmo o uso de *retweets* como reforçamento do *bullying*.

4.1. Construção do córpus

Para a construção do córpus, inicialmente, foram coletados 48.262 *tweets* usando a API de *streaming* do Twitter. Os *tweets* foram coletados da seguinte forma: Todos os *tweets* que continham as palavras chave ***bullied, buly, bullying, bulin, buli, bullin, bulen, bullen e bulli*** foram recebidos por um *script* em *Python*, utilizando a biblioteca *tweepy*[6]. Foram adicionados as grafias incorretas dos termos em inglês pois os mesmos são frequentemente utilizados por usuários brasileiros. A coleta foi feita em tempo real. Os

[3]https://dev.twitter.com/rest/public

[4]Os *retweets* ou RTs são mensagens repassadas de algum usuário no Twitter para os seus seguidores.

[5]*Replies* são mensagens direcionadas a algum usuário referenciando o nome do usuário.

[6]`http://www.tweepy.org`

tweets que não eram do Português do Brasil foram descartados. Nenhum *tweet* foi duplicado. Após a coleta foram excluídos *tweets* que representavam *spams*[7]. Também foram excluídos *tweets* que possuíam menos de 120 caracteres. Após todos estes filtros restaram 6.851 *tweets* de interesse. Deste montante, 2.000 *tweets* foram anotados por 6 anotadores quanto à presença de traço de *bullying* e quanto ao papel do autor em relação ao episódio. Os 2.000 *tweets* foram divididos em dois grupos de 1.000, cada grupo sendo anotado por 3 anotadores.

De forma a facilitar o processo de anotação e de forma a contornar o problema da subjetividade da anotação (ode haver discordância quanto ao que é ou não um episódio de *bullying*), foi confeccionado um manual para os anotadores, contendo as motivações para o trabalho, diretrizes para anotação, e principalmente uma definição objetiva do que deveria ser considerado como um traço de *bullying*. Os anotadores deveriam então, seguir como base na seguinte definição: "Um *tweet* possui traço de *bullying* quando há menção a um ou mais episódios de *bullying*, que acuse alguém de ser um *bully* sem mencionar episódios específicos, que mencione que o autor ou outra pessoa sofreu *bullying*, mesmo que sem mencionar episódios específicos ou instâncias de *cyberbullying*."

Os anotadores receberam um arquivo contendo o número de identificação ID e o texto do *tweet*, um campo chamado *Bullying* (onde o anotador deveria anotar "sim" caso fosse identificado um traço de *bullying* e "não" caso contrário). Além disso, também foi disponibilizado no arquivo o campo Papel para que o anotador (quando anotar o campo *Bullying* com o valor "sim") identifique se o traço de *bullying* foi escrito por uma vítima, por um *bully* (praticante do episódio de *bullying*) ou por um relator (narrador do episódio de *bullying*). O arquivo final após a anotação é ilustrado na Tabela 1.

4.2. Experimentos

Uma vez de posse do córpus anotado, o próximo passo deste estudo foi levantar as estatísticas do córpus e em seguida, avaliar algoritmos de classificação clássicos já avaliados para outras línguas para o problema de detecção de traços de *bullying*. Neste trabalho foram usados os resultados da anotação do córpus integralmente, ou seja, os textos anotados relativos a ambos os grupos. O total de textos anotados foi de 2.000 *tweets*. Destes 2000 textos, 1294 foram rotulados como contendo traços de *bullying*, e 706 como não contendo traços de *bullying*. Quanto ao papel do autor, 102 foram rotulados como *Bully*, 442 como Vítima e 629 como Relator. Houve discordância total entre os 3 anotadores quanto ao papel em 121 exemplos, estes então não foram alocados em nenhuma classe.

Os métodos de detecção de traços de *bullying* seguem uma sequência de passos comuns à área de mineração de textos [Xu et al. 2012a, Xu et al. 2012b] — ver Figura 2.

Apesar de o *dataset* incluir apenas *tweets* que continham os termos relacionados a *bullying*, várias instâncias não contém traço de *bullying* efetivamente, como por exemplo: "*@USERNAME COMASSIM? SÓ PODE QUEM TIVER 18 EM 2020? Q BULLYING! eu vou ter 16 ;-; LETÍCIA SE TU FOR, ME LEVA MOZONA parei*".

O resultado esperado de um algorítimo de classificação supervisionado é um modelo que será usado para classificar novas instâncias (diferentes das instâncias usadas no

[7]Foi percebido que as palavras chaves utilizadas traziam um grande numero de *tweets* com o mesmo texto: "Morgan Dollar Uncirculated US Mint Gem PQ Silver Coin BU Unc MS++++++ HTTPLINK #silver #junksilver #coins #us #bulli...". Foram removidas todas as ocorrências deste texto na base de dados.

Tabela 1. Arquivo de anotação

ID	**Texto**	**Bullying?** (Sim ou Não)	**Papel** (vítima, *bully* ou relator)
30	Já fui alvo de "bullying", na escola diziam que eu parecia um menino pq eu usava roupas da sessão de meninose. Eu odiava isso!	S	vítima
36	#SouDoTempoQue meninos eram simplesmente meninos, meninas eram meninas, todo mundo se zoava e ngm chorava "bullying mimimi"	N	N/A
45	Tava reclamando da festa, mas na real foi muito boa, encontrei amigos antigos, descobri q os garotos q eu fazia bullying n me odeiam	S	bully
6679	Aluno sofre bullying por ser careca, e diretor o deixa raspar sua cabeça em sala	S	relator
6717	esse povo que vitimiza o snape pq ele sofreu bullying na escola mas apaga tudo o que ele fez com os alunos da gryff...	S	relator
6742	vi em uma entrevista da pabllo falando o bullying que ela sofre por causa da por dela eu fiquei ????? a voz dela é maravilhosa gente	S	relator
6818	uma pagina de youtubers no fb postou que o Christian sofreu bullying por ser magro demais. O que é de total verdade, quem acompanha ele (+)	S	relator

treinamento) conforme classes predeterminadas. Este modelo, também conhecido como classificador, pode ser construído usando varias técnicas diferentes. Neste trabalho, foram usados os classificadores Naive Bayes, Regressão Logística (*Logistic Regression*) e SVM (*Support Vector Machine*). Ao usar SVM, é necessário escolher um *kernel*. Neste trabalho foram utilizados o *kernel* linear e o *kernel* RBF [Cortes and Vapnik 1995]. Ao se utilizar o *kernel* RBF em SVM , é necessário calibrar alguns parâmetros internos do mesmo, nomeadamente *C* e *gamma*. Para este trabalho foi utilizada a abordagem de calibragem de parametros demonstrada em [Hsu et al. 2003]. Após realizar a abordagem citada, os valores obtidos foram C = 8, gamma = 0.01

Seguindo [Xu et al. 2012a], executamos 30 rodadas de cada experimento, sendo os resultados a média aritmética das execuções. Do conjunto de 2.000 tuplas, 300 (15%) foram separados como conjunto de teste. O montante restante de 1.700 tuplas foram utilizados para treinamento. Inicialmente utilizando apenas 100, e aumentando em incrementos de 100 até atingir a totalidade do conjunto de treinamento. Os conjuntos de teste e treinamento foram gerados de forma aleatória em cada execução, garantindo-se que a proporção entre as classes fosse constante em todas as execuções (estratificação).

Os textos foram transformados em dois conjuntos distintos de *features*. No primeiro conjunto, os textos foram transformados em *bag-of-words* ou conjunto de *features* (ver Fase 1 da Figura 2) contendo somente unigramas. Já no segundo conjunto, os textos

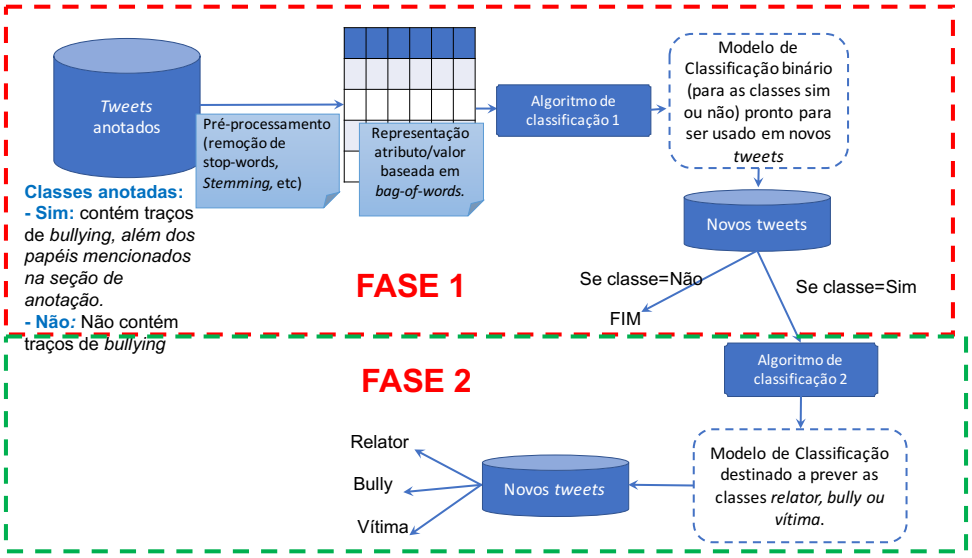


Figura 2. Etapas realizadas para a obtenção de um modelo de classificação de detecção de traços de *bullying*

foram transformados em uma *bag-of-words* contendo unigramas e bigramas. Nenhum pré-processamento adicional foi empregado no conjunto de dados. A *bag-of-words* contendo unigramas, após a extração, apresentou 7.387 termos (atributos), enquanto que a *bag-of-words* contendo unigramas e bigramas apresentou 35.542 termos.

Sumarizando, foram executadas 30 rodadas para cada classificador e cada conjunto de *features*. Sendo cada par classificador+conjunto de *features* sendo testado com 17 tamanhos diferentes, ou seja, treinando o modelo inicialmente em 100 *tweets* e re-treinando o modelo de classificação após incrementos de 100 (até atingir a totalidade do conjunto de treinamento). Ao final foram realizados 4.080 experimentos. A hipótese inicial com essa configuração experimental é que ao adicionar novos dados ao conjunto de treinamento ocorreria um ganho de acurácia gradativo, o que de fato pode ser visualizado nos gráficos das Figuras 3 e 4.

Uma forma de comparar classificadores é comparar suas medidas de acurácia. A acurácia é definida pelo número de instâncias corretamente classificadas dividido pelo total de instâncias. Além da acurácia, outras medidas são utilizadas neste trabalho. Nomeadamente Precisão, *Recall* e *F-score*[Tan et al. 2005]. Estas medidas são especialmente úteis quando não se tem uma base de dados perfeitamente balanceada, ou seja, quando a quantidade de instâncias de cada classe não é igual.

Uma forma de visualizar o desempenho de um classificador é através da sua Matriz de Confusão. Uma Matriz de Confusão é uma matriz que exibe em um eixo as classes preditas pelo classificador, e em outro eixo as classes reais dos dados. Através da Matriz de Confusão é possível observar a eficácia do classificador em relação a cada classe dos dados.

Para executar os experimentos, criar os conjuntos de *features* e avaliar os classificadores obtidos, foi utilizada a linguagem Python[8]. Python possui um módulo específico

[8]https://www.python.org/

para ciência de dados, chamado *scikit-learn* [Pedregosa et al. 2011], que contém toda a implementação dos classificadores utilizados, funções de transformação de texto em *features*, separação de dados para teste e treinamento, e funções de avaliação.

4.3. Resultados

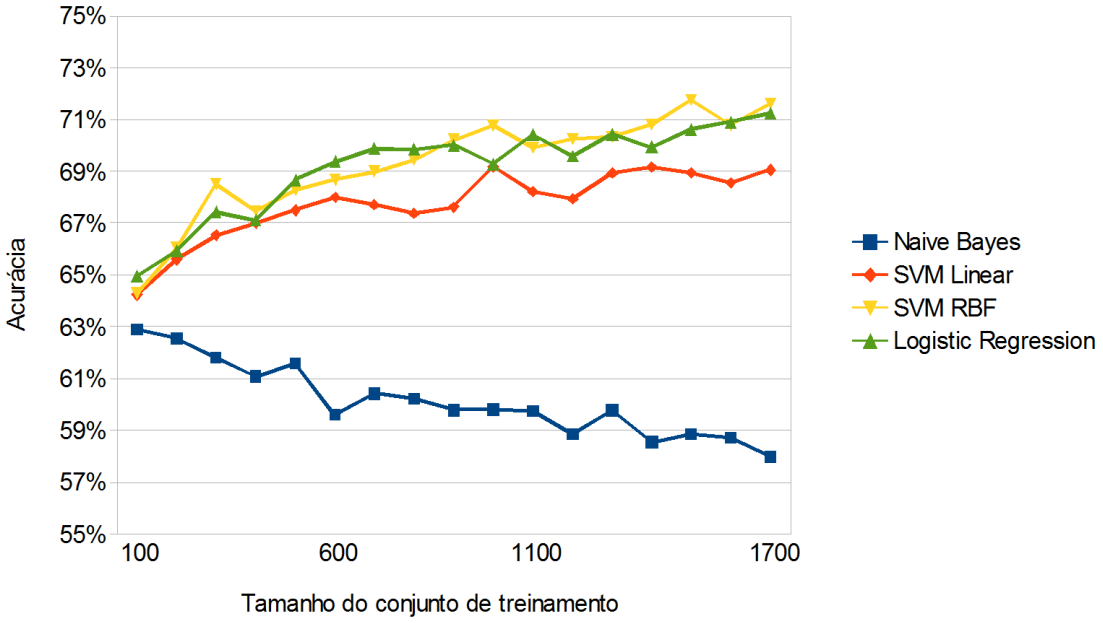


Figura 3. Resultados dos experimentos utilizando somente unigramas

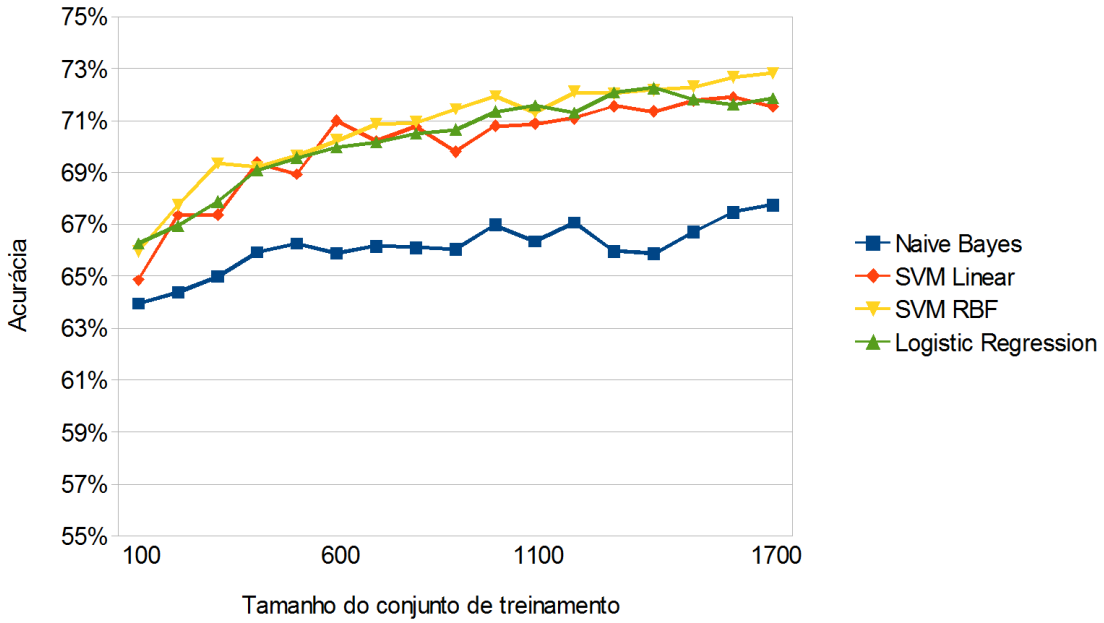


Figura 4. Resultados dos experimentos utilizando unigramas e bigramas

Observando as Figuras 3 e 4, todos os classificadores testados obtiveram melhores resultados utilizando o conjunto de *features* contendo unigramas e bigramas. Quanto à acurácia, SVM com *kernel* RBF obteve os melhores resultados, chegando a 72.8% no maior conjunto de dados. SVM com *kernel* Linear e Regressão Logística foram levemente piores, com 71.5% e 71.8% respectivamente. Naive Bayes foi significativamente pior, com 67.7%.

O classificador mais bem sucedido, SVM com *kernel* RBF, obteve a Precisão de 75%, *Recall* de 87% e F-score de 80%. De forma a melhor ilustrar os resultados, a Figura 5 exibe a matriz de confusão extraída do classificador mais bem sucedido.

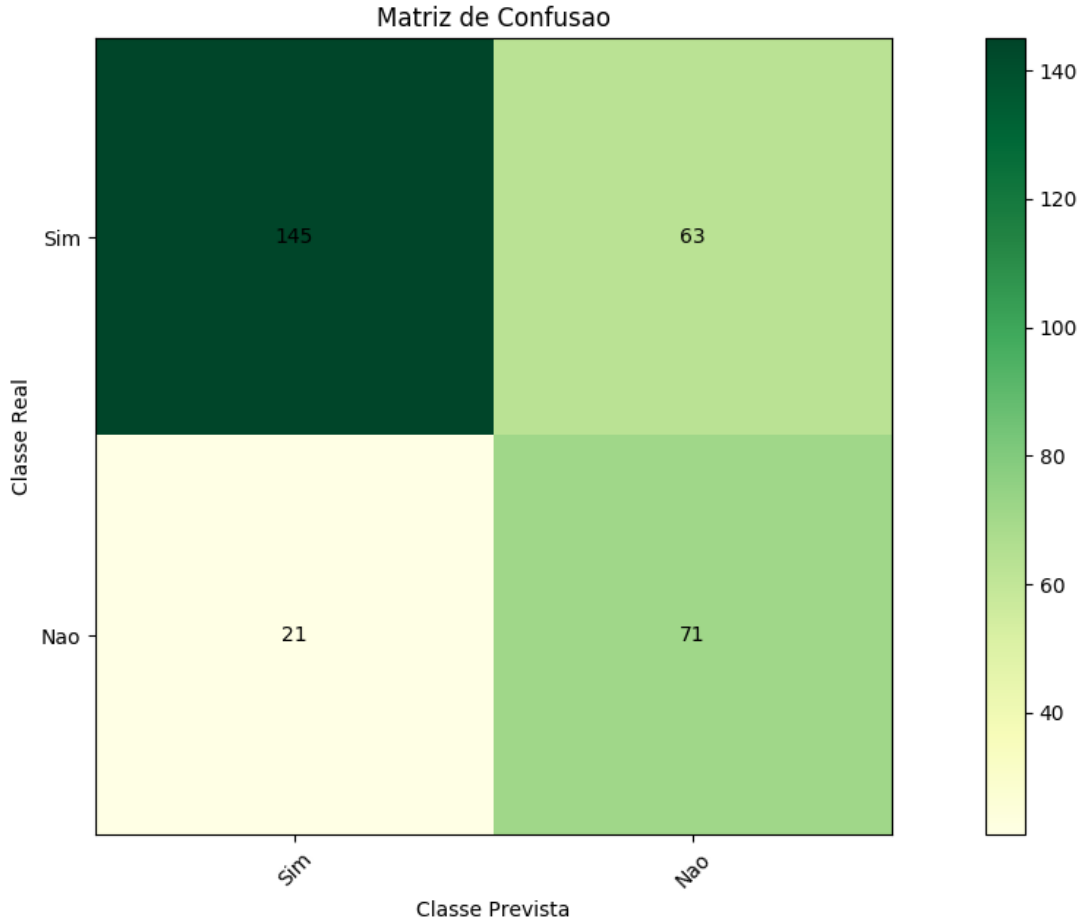


Figura 5. Matriz de Confusão utilizando SVM RBF com atributos compostos de unigramas e bigramas para os dados de teste - 300 *tweets*

Um ponto interessante a se notar é que utilizando apenas unigramas como *features*, o Naive Bayes teve seu desempenho deteriorado com o aumento do conjunto de treinamento.

A Figura 6 mostra a diferença na acurácia entre os conjuntos de *features*, usando o classificador mais bem sucedido como exemplo.

Apesar dos resultados serem positivos, podemos enumerar possíveis motivos para a acurácia ainda estar longe dos 100%:

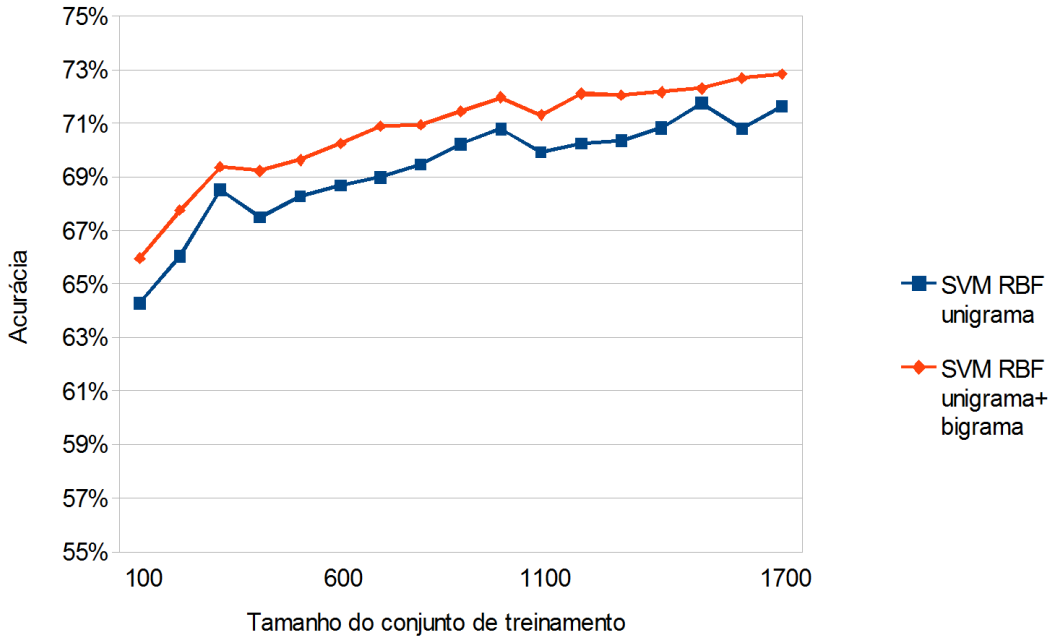


Figura 6. Comparação entre a configuração unigramas VS unigramas+bigramas usando SVM com *Kernel* RBF como algoritmo de classificação

- Identificar textos que contém ou não *bullying* se mostrou uma tarefa difícil até mesmo para os anotadores. A subjetividade e ambiguidade se mostraram barreiras, e o *dataset* anotado não apresentou grande concordância entre os anotadores, mesmo provendo um manual de anotação com diretrizes e uma definição do que deveria ser considerado traço de *bullying*.
- Apesar de termos selecionado apenas *tweets* cuja língua informada era o Português do Brasil, *tweets* erroneamente setados como Português existem no corpus. Estes e outros ruídos podem ter influenciado negativamente na construção dos modelos.
- Informação importante para a tarefa pode estar contida em *tweets* pequenos, portanto a decisão de usar apenas *tweets* com mais de 120 caracteres pode ter influenciado negativamente os modelos construídos.
- Apenas ferramentas preexistentes foram utilizadas. O emprego de ferramentas novas ou customizadas para a tarefa pode gerar melhores resultados.
- Como mencionado anteriormente, a curva de aprendizado continua crescendo, o que significa que um maior conjunto de dados anotados pode aumentar a acurácia.

5. Considerações Finais

Bullying é um problema mundial, sendo que a agressão física ou moral repetitiva deixa sequelas psicológicas na pessoa atingida. A grande quantidade de dados disponíveis no Twitter é um dos aspectos motivadores para o presente trabalho, além do que a compreensão das ferramentas virtuais de comunicação pode levar a um entendimento maior do *bullying* fora da Internet. O estudo de detecção de traços de *bullying* em redes sociais, como o Twitter, para a língua inglesa tem tido grande ajuda do aprendizado de máquina e mineração de dados, uma vez que tais técnicas ajudam a obter informações sobre o

comportamento dos indivíduos nestes ambientes, fornecendo subsídios para o combate e prevenção de tais atos.

Este trabalho demonstrou que é possível utilizar ferramentas para a detecção automática de *Bullying* no Twitter na língua portuguesa. Utilizando uma base de dados de 2000 tweets anotada manualmente, foram testados 4 classificadores e duas representações de features, sendo SVM com *kernel* RBF o que obteve os melhores resultados, com acurácia de 72.8%. É possível ver pelo gráfico que a curva de aprendizado continua aumentando, mostrando que uma maior quantidade de dados para treinamento pode aumentar ainda mais a acurácia dos classificadores, corroborando as conclusões em [Xu et al. 2012a].

Para trabalhos futuros, sugerimos uma análise da eficácia de classificadores na detecção automática não somente da existência de um traço de *bullying*, mas também na inferência do papel do autor do texto no episódio (os papéis de vítima, *bully* e relator foram anotados no *corpus* e são definidos na seção 3.1, mas não foram utilizados nos experimentos aqui apresentados). Além disso, técnicas mais sofisticadas de aprendizado de maquina podem ser testadas nesta tarefa, como *Ensemble Learning* [Dietterich 2000] e *Deep Learning* [Cambria and White 2014].

Referências

Archer, J. and Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Pers Soc Psychol Rev*, 9(3):212–230.

Bellmore, A., Calvin, A. J., Xu, J.-M., and Zhu, X. (2015). The five w's of "bullying" on twitter: who, what, why, where, and when. *Computers in human behavior*, 44:305–314.

Cambria, E. and White, B. (2014). Jumping nlp curves: A review of natural language processing research [review article]. *IEEE Computational Intelligence Magazine*, 9(2):48–57.

Chatzakou, D., Kourtellis, N., Blackburn, J., De Cristofaro, E., Stringhini, G., and Vakali, A. (2017). Mean birds: Detecting aggression and bullying on twitter. *arXiv preprint arXiv:1702.06877*.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3):273–297.

Dietterich, T. G. (2000). *Ensemble Methods in Machine Learning*, pages 1–15. Springer Berlin Heidelberg, Berlin, Heidelberg.

Folha de SP (2003). Segundo a polícia, atirador de taiúva escolheu alvos. `http://www1.folha.uol.com.br/folha/cotidiano/ult95u67698.shtml`. Acessado em 16-09-2017.

Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.

José, R. et al. (2012). Estudo da ocorrência de cyberbullying contra professores na rede social twitter por meio de um algoritmo de classificação bayesiano.

Leary, M. R., Kowalski, R. M., Smith, L., and Phillips, S. (2003). Teasing, rejection, and violence: Case studies of the school shootings. *Aggressive behavior*, 29(3):202–214.

Malta, D. C., Silva, M. A. I., Mello, F. C. M. d., Monteiro, R. A., Sardinha, L. M. V., Crespo, C., Carvalho, M. G. O. d., Silva, M. M. A. d., and Porto, D. L. (2010). Bullying nas escolas brasileiras: resultados da pesquisa nacional de saúde do escolar (pense), 2009. *Ciência & Saúde Coletiva*, 15(suppl 2):3065–3076.

Nalini, K. and Sheela, L. J. (2015). Classification of tweets using text classifier to detect cyber bullying. In *Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India CSI Volume 2*, pages 637–645. Springer.

Nylund, K., Bellmore, A., Nishina, A., and Graham, S. (2007). Subtypes, severity, and structural stability of peer victimization: what does latent class analysis say? *Child development*, 78(6):1706–1722.

Olweus, D. (1994). Bullying at school. In *Aggressive behavior*, pages 97–130. Springer.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Presidência da República (2015). Lei nº 13.185, de 6 de novembro de 2015. `http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2015/Lei/L13185.htm`. Acessado em 16-09-2017.

Rigby, K. (2001). Health consequences of bullying and its prevention. *Peer harassment in school: The plight of the vulnerable and victimized*, page 310.

Salmivalli, C. (1999). Participant role approach to school bullying: Implications for interventions. *Journal of adolescence*, 22(4):453–459.

Slonje, R. and Smith, P. K. (2008). Cyberbullying: Another main type of bullying? *Scandinavian journal of psychology*, 49(2):147–154.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

UOL Notícias (2011). Autor do massacre no rio sofreu bullying, dizem ex-colegas de escola. http://noticias.uol.com.br/cotidiano/ultimas-noticias/2011/04/08/autor-do-massacre-no-rio-sofreu-bullying-dizem-ex-colegas-de-escola.htm Acessado em 16-09-2017.

Xu, J.-M., Jun, K.-S., Zhu, X., and Bellmore, A. (2012a). Learning from bullying traces in social media. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, pages 656–666. Association for Computational Linguistics.

Xu, J.-M., Zhu, X., and Bellmore, A. (2012b). Fast learning for sentiment analysis on bullying. In *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*, WISDOM '12, pages 10:1–10:6, New York, NY, USA. ACM.

Métodos para análise de sentimentos em *tweets*: um estudo comparativo

Alison P. Ribeiro[1], **Nádia F.F. da Silva**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Campus Samambaia
CEP 74690-900 – Goiânia – GO – Brasil

alisonrib17@gmail.com, nadia@inf.ufg.br

Abstract. *Sentiment Analysis or Opinion Mining has become an important field of study due to the enormous amount of texts available in the Social Web, which allows several applications such as monitoring of brands and products, forecasting political campaigns and even applications in the financial market. Several independent methods are explored in the literature. This article aims at comparing techniques already known as machine learning, lexical dictionaries, and part-of-speech, with the aim of indicating to the reader, among such approaches, the one that best suits the particularities of the texts.*

Resumo. *Análise de sentimentos ou Mineração de Opinião tem se tornado um importante campo de estudo devido à enorme quantidade de textos disponíveis na Web Social, o que possibilita diversas aplicações como monitoramento de marcas e produtos, previsão de campanhas políticas e até aplicações no mercado financeiro. Vários métodos independentes são explorados na literatura. Este artigo visa comparar técnicas já conhecidas como aprendizado de máquina, uso de dicionários léxicos, emoticons e part-of-speech, com o objetivo de indicar ao leitor, dentre tais abordagens, a que melhor se adéqua frente às particularidades dos textos.*

1. Introdução

Segundo [Pang and Lee 2008], descobrir o que as pessoas pensam sempre foi motivo de interesse. Devido ao crescimento de ambientes que dispõem de uma grande quantidade de dados subjetivos, a tarefa de classificar sentimentos passou a ser objeto de pesquisas para diversos estudiosos. Dentre estes ambientes, está o Twitter[1], um *microblog* no qual usuários podem divulgar suas ideias e expressar suas opiniões sobre algum evento, podendo ser esportivo, político ou até opiniões direcionadas à empresas e serviços.

O Twitter é uma fonte importante para realização de pesquisas por ser uma plataforma que propicia a difusão de conteúdo. Por meio dos chamados *tweets*, os usuários expõem o que pensam dentro de 140 caracteres. Por causa desse limite de escrita classificar *tweets* é um desafio, pois os usuários utilizam muita linguagem informal, por exemplo, abreviações e gírias [SILVA 2016]. Frente tais desafios em classificar *tweets*, pesquisadores utilizam diversos métodos presentes na literatura afim de minimizar os problemas e buscar resultados satisfatórios. Como não há um consenso estabelecido sobre qual a

[1]https://twitter.com/

melhor abordagem para a classificação de sentimentos, é livre para o pesquisador usufruir de métodos como aprendizado de máquina, dicionários léxicos[2], orientação sintática (aspectos gramaticais) e *emoticons* [Pang et al. 2002]. Os *emoticons* são usados porque permitem reduzir o esforço necessário para identificar a emoção expressa na publicação. A ideia base em usar *emoticons* consiste na ocorrência do *emoticon*, ou seja, verificar se um determinado *tweet* possui *emoticons* que expressam sentimento positivo, negativo ou neutro. A desvantagem dessa técnica é que, nem todos *tweets* possuem *emoticons*.

Para predizer o sentimento de um *tweet*, muitos estudos [SILVA 2016, Go et al. 2009, Hassan Saif, Miriam Fernandez and Alani 2013, Kiritchenko et al. 2014] utilizam a chamada engenharia de atributos, processo que usa o conhecimento de domínio dos dados para determinar caraterísticas que melhor definam um sentimento. Uma vez que tal etapa é realizada, diversos algoritmos são aplicados a fim de decidir qual possui maior acurácia frente o conjunto de dados escolhido. Diversos algoritmos de classificação são aplicados, como SVM (*Support Vector Machine*), MNB (*Multinomial Naive Bayes*), RF (*Random Forest*) e LR (*Logistic Regression*).

O presente trabalho irá analisar a polaridade expressa em *tweets*. Para tanto, as técnicas citadas anteriormente foram aplicadas em duas bases de dados distintas: *Sanders* [Sanders 2011] e *HCR* [Speriosu et al. 2011]. A finalidade deste estudo é contribuir para a análise de sentimentos explorando diversos modelos preditivos e realizando comparações. As comparações foram focadas especialmente na acurácia, embora outras três métricas foram consideradas como precisão, revocação e medida F.

O artigo está organizado da seguinte forma: a Seção 2 descorre sobre os trabalhos relacionados, a Seção 3 explica alguns conceitos relacionados à Análise de Sentimentos, a Seção 4 apresenta uma descrição das bases de dados, a Seção 5 aborda a metodologia do trabalho e como cada método foi aplicado, a Seção 6 mostra os resultados obtidos. Por fim, a Seção 7 apresenta as considerações finais e uma descrição dos trabalhos futuros.

2. Revisão Bibliográfica

Em análise de sentimentos não existe uma convenção de quais métodos e atributos apresentam um melhor resultado na classificação de *tweets* [SILVA 2016]. Desta forma, é uma prática comum realizar diversos experimentos com algoritmos diferentes baseando-se em métodos independentes que se diferem em sua predição de sentimentos de acordo com o seu *viés* [Tan et al. 2005].

Em [Araújo et al. 2013], os autores apresentaram um trabalho no intuito de comparar 8 métodos baseados em léxicos de sentimentos conhecidos na literatura. A partir dos 8 métodos, 7 foram escolhidos para um novo experimento chamado pelos autores de Método Combinado. Na análise feita em cada método, a comparação foi realizada através da acurácia. Os métodos que obtiveram melhores resultados foram o SenticNet [Cambria et al. 2016] e o SentiWordNet [Baccianella et al. 2010]. Já no Método Combinado, os autores realizaram a seguinte estratégia: analisaram a média harmônica (*F-measure*), a precisão e *recall* de cada método e distribuíram diferentes pesos para cada um deles. A combinação foi feita de forma incremental e a medida que cada método

[2]O dicionário léxico é um conjunto de palavras armazenadas em um *dataset*, construído automaticamente ou manualmente, que polariza as palavras que o compõe de acordo com os sentimentos daquele contexto.

foi adicionado, menor foi se tornando o *F-measure* e a acurácia. Como dito pelos autores, usar muitos métodos não garante melhores resultados, ao invés disso, é mais viável escolher o conjunto de métodos que irá lidar melhor com um determinado problema.

Já no trabalho de Reis em [Reis et al. 2015], foi realizado um estudo destinado à análise de sentimentos em 9 idiomas. Para isso, os autores selecionaram 13 métodos baseados em aprendizado de máquina e dicionários léxicos. Neste trabalho, as bases de dados foram traduzidas para o idioma inglês. Os autores tiveram como objetivo verificar a qualidade dos métodos abordados. Para isso, cada método foi comparado em nível de abrangência, isto é, a capacidade de predição de um método em cada idioma. As técnicas que se destacaram foram: SentiWordNet [Baccianella et al. 2010], Sentiment140 *Lexicon* [Mohammad et al. 2013] e SenticNet [Cambria et al. 2016].

Nos estudos de Chaovalit & Zhou em [Chaovalit and Zhou 2005], duas abordagens foram usadas: aprendizado de máquina e *part-of-speech*[3]. A abordagem de aprendizado de máquina foi empregada com *n-gramas*[4] em duas formas, a primeira aplicando validação cruzada com 3 *folds* e a segunda com uma base de dados para teste. A acurácia obtida foi de 84,49% e 66,27%, respectivamente. Já a abordagem de *part-of-speech* alcançou uma acurácia de 77%. O desempenho da técnica de aprendizado de máquina foi influenciado pelo uso de *n-gramas* juntamente com a eliminação de ruídos. Enquanto o método de *part-of-speech*, teve a performance diretamente ligada ao conjunto de *tags* aplicadas na análise. Os testes realizados foram feitos em avaliações de filmes porque segundo os autores é mais desafiador devido as palavras irônicas presentes nas críticas.

Seguindo uma linha comparativa, Hardeniya & Borikar em [Hardeniya and Borikar 2016] propuseram uma estrutura para análise de sentimentos usando uma abordagem baseada em dicionário léxico e trazendo um estudo comparativo sobre técnicas de mineração de opinião, incluindo aprendizado de máquina e léxicos. As comparações foram embasadas em recursos como técnica empregada, dicionários e abordagens de *soft-computing* [Jang et al. 1997]. Os autores propuseram também uma abordagem para dicionário léxico que incorpora uma lógica difusa, este método trata, principalmente, de classificar as avaliações como positivas, negativas ou neutras com base em uma pontuação que é calculada usando os dicionários SentiWordNet [Mohammad et al. 2013] e WordNet [Miller 1995].

Este trabalho segue um caminho diferente. A ideia principal é prover para análise de sentimentos um estudo focado em métodos baseados em dicionários léxicos, aprendizado de máquina, orientação sintática e *emoticons*. Além disso, tais métodos foram comparados com resultados da literatura. Uma comparação também foi realizada entre os dicionários léxicos. Além do mais, levantamos algumas questões a serem respondidas: (i) Dentre os atributos utilizados nos trabalhos relacionados quais são os atributos preditivos mais relevantes nos conjuntos de dados utilizados nos experimentos (*benchmarking* da literatura)? (ii) Com a configuração de atributos proposta neste trabalho, qual algoritmo de aprendizado de máquina possui a maior acurácia *tweets*? (iii) Quantas palavras em comum existem nos léxicos utilizados nos experimentos? (iv) E quantas dessas palavras se contradizem nos léxicos escolhidos?

[3] *Part-of-speech* é um processo de rotulação de elementos textuais – tipicamente palavras e pontuação – com o objetivo de evidenciar a estrutura gramatical de um determinado trecho de texto [Gimpel et al. 2011].

[4] Termos compostos por n palavras.

3. Análise de sentimentos em *tweets*

Análise de Sentimentos é o campo da Ciência da Computação responsável por realizar a análise de opiniões em *tweets*, *reviews* de filmes, comentários referentes a procutos[Liu 2012]. Trata-se de uma área em expansão que engloba diversas técnicas, já citadas na introdução. A classificação de sentimentos consiste em 4 fases: seleção de *tweets* relevantes, pré-processamento, classificação e predição.

Em mineração de opinião, nota-se um predomínio do uso de métodos supervisionados [da Silva et al. 2014, Pawar and Deshmukh 2015, Aisopos et al. 2011, Barbosa and Feng 2010, Go et al. 2009], mais especificamente, classificação e regressão. O problema da classificação pode ser dividido em dois passos: (i) aprender um modelo de classificação sobre um *corpus* de treinamento previamente rotulado como positivo, negativo ou neutro; (ii) prever a polaridade de novas porções de *tweets* com base no modelo resultante.

Os dados de treinamento para a classificação correspondem a um conjunto de instâncias caracterizadas por atributos. O rótulo é denominado atributo alvo, enquanto que os demais são designados como atributos discriminantes ou *features* [Tan et al. 2005]. Em termos de pré-processamento, é necessário extrair de cada porção de texto analisado, as *features* relevantes para a tarefa de classificação e representá-las na forma de um vetor de termos, chamado de *bag-of-words*. A Tabela 1 ilustra uma coleção de *tweets* após a etapa de pré-processamento.

	t1	*t2*	...	*tm*
tweet1	*a11*	*a21*	...	*a1m*
tweet2	*a21*	*a22*	...	*a2m*
...	...	...	...	...
tweetn	*an1*	*an2*	...	*anm*

Tabela 1. Representação de uma *bag-of-words.*

Isso significa que o valor *aij* refere-se ao valor associado ao *j*-ésimo termo do *tweet i*, isto é, *aij* é o valor do termo *tj* no *tweet i* e pode ser calculado pela frequência ou pela presença. Alguns autores utilizam valores binários [SILVA 2016, Pang et al. 2002]. Neste caso, *aij* = 1 significa a presença do termo *j* na mensagem *i* e o valor 0 significa a ausência do termo. Após esta etapa, o algoritmo de classificação recebe como entrada o modelo *bag-of-words*, assim como as classes de cada *tweet*, e apresenta como saída a predição das classes.

Para a avaliação do desempenho do classificador e do comportamento do modelo, primeiro é preciso entender o que é Acurácia (*Accuracy*), Precisão (*Precision*), Revocação (*Recall*) e a Medida F (*F1-Score*). Mas antes, vamos apresentar algumas definições importantes [Tan et al. 2005]:

- True positive (TP): significa uma classificação correta da classe positiva. Por exemplo, a classe real é positiva e o modelo classificou como positiva.
- True negative (TN): significa uma classificação correta da classe negativa. Por exemplo, a classe real é negativa e o modelo classificou como negativa.
- False positive (FP): significa uma classificação errada da classe positiva. Por exemplo, a classe real é negativa e o modelo classificou como positiva.

- False negative (FN): significa uma classificação errada da classe negativa. Por exemplo, a classe real é positiva e o modelo classificou como negativa.

A Precisão significa o número de vezes que uma classe foi predita corretamente (TP), divido pela soma da classe predita positivamente com a classe predita erroneamente como positiva (TP + FP). Como pode ser visto na equação 1. Já a Revocação significa o número de vezes que uma classe foi predita corretamente (TP), dividido pela soma da classe predita positivamente com a classe predita erroneamente como negativa (TP + FN). Como ilustra a equação 2. Enquanto a Medida F consiste na média harmônica entre a precisão e a revocação. Com essa informação podemos dizer a performance do classificador com um indicador apenas. Como mostra a equação 3. Por fim, temos a acurácia que nos mostra como o classificador se saiu de uma maneira geral, pois, esta mede a quantidade de acertos sobre o todo. O cálculo da acurácia é apresentado na equação 4.

$$precision = TP / (TP + FP) \tag{1}$$

$$recall = TP / (TP + FN) \tag{2}$$

$$f1\text{-}score = 2 * ((precision * recall) / (precision + recall) \tag{3}$$

$$accuracy = (TP + TN) / (TP + FP + TN + FN) \tag{4}$$

4. Dados utilizados

Esta seção explana sobre os conjuntos de dados utilizados: *Sanders* [Sanders 2011] e *HCR* [Speriosu et al. 2011]. Estas bases de dados foram escolhidas por serem comumente usadas para pesquisas em análise de sentimentos como em [Speriosu et al. 2011, Hassan Saif, Miriam Fernandez and Alani 2013, SILVA 2016] e, também, por estarem disponíveis publicamente.

4.0.1. Sanders

A base de dados *Sanders* consiste em 5.513 *tweets* classificados manualmente por especialistas, este conjunto de *tweets* foi coletado a partir de quatro tópicos: *@apple*, *#google*, *#microsoft* e *#twitter*. Cada *tweet* possui um rótulo de sentimento: positivo, negativo, neutro e irrelevante. Para este trabalho, apenas os *tweets* classificados como positivo, negativo e neutro foram analisados. Portanto, 3.726 *tweets* foram selecionados, sendo: 570 com sentimento positivo, 653 com sentimento negativo e 2.503 com sentimento neutro.

4.0.2. HCR

A base de dados *HCR* (*Health Care Reform*) foi construída a partir da *hashtag* "*#hcr*". Os *tweets* foram coletados em março de 2010 [Speriosu et al. 2011]. Esta base de dados foi catalogada em 4 sentimentos: positivo, negativo, neutro e irrelevante. O conjunto

de *tweets* foi dividido em dados de treinamento, desenvolvimento e teste, no entanto, mesclamos os dados de treino com os dados de desenvolvimento. Portanto, utilizamos 852 *tweets* para treino e 480 *tweets* para teste. Neste artigo desconsideramos *tweets* dados como irrelevantes.

5. Metodologia

A metodologia adotada neste artigo consiste em criar novas configurações a partir do modelo *bag-of-words* e dos métodos selecionados, juntamente com atributos de unigrama, bigrama e uma junção entre estes. Na construção do modelo *bag-of-words*, foi verificado a presença do termo ao invés da frequência, isto é, uma matriz construída de forma binária. Esta estratégia é mais eficaz para classificação do sentimento [Pang et al. 2002]. Para os testes utilizamos 4 algoritmos de classificação: SVM (*Support Vector Machine*), MNB (*Multinomial Naive Bayes*), RF (*Random Forest*) e LR (*Logistic Regression*). Escolhemos estes algoritmos por serem frequentemente utilizados em trabalhos de mineração de opinião, como em [SILVA 2016, Mullen and Collier 2004, Go et al. 2009], e por demonstrarem resultados satisfatórios em relação ao nível de acurácia.

A seguir explicaremos o processo de classificação desde a etapa de pré-processamento até a validação dos dados. Também detalharemos como cada método foi aplicado ao modelo *bag-of-words* afim de alcançar resultados mais expressivos.

Pré-processamento: Esta etapa consiste em eliminar ruídos e termos que não possuem significado semântico na predição de sentimentos. Para isso, realizamos a remoção de links, remoção de números, remoção de caracteres especiais, remoção de *stopwords* (palavras com baixo poder discriminativo, por exemplo, ”a”, ”é”, ”que”etc.) e remoção de *emoticons* (somente quando usamos as técnicas que não envolvem *emoticons*). Foi aplicada também a padronização dos *tweets* em minúsculo e, por fim, *stemming*. O objetivo do *stemming* é reduzir palavras ao seu radical, por exemplo, a palavra ”*belivies*”será transformada em ”*believ*”[Perkins 2014].

Dicionários léxicos: Para esta técnica foi utilizado três dicionários léxicos: *Opinion Lexicon* [Hu and Liu 2004], *SenticNet* [Cambria et al. 2016] e *SemEval2015-Lexicon* [Kiritchenko et al. 2014].

- *Opinion Lexicon*: Este pacote possui 4.783 léxicos positivos e 2.006 léxicos negativos. Para este método utilizamos a estratégia proposta por Mohammad em [Mohammad et al. 2013], foi feita uma contagem de léxicos positivos e negativos presentes em cada *tweet*. Se o número de palavras positivas for maior que o número de palavras negativas, então, o *tweet* é positivo, caso contrário o *tweet* é negativo. No caso de haver um empate entre palavras positivas e negativas, o *tweets* é neutro.

- *SenticNet*: Este corpus conta com 50.000 palavras classificadas como positivas e negativas. Assim como no *Opinion Lexicon* [Hu and Liu 2004], para o *SenticNet* aproveitamos a mesma estratégia.

- *SemEval2015-Lexicon*: Já este conjunto de léxicos possui 1515 léxicos, cada palavra desse *corpus* possui uma pontuação de número real. Nesta etapa soma-se os pontos de palavras positivas e negativas, e então verifica-se o resultado final se é

um número positivo ou negativo, caso seja um número maior que zero o *tweet* é dado como positivo, caso contrário, o *tweet* é negativo. Um *tweet* é considerado neutro se o resultado da soma for igual a zero.

Emoticons: Neste método utilizamos o *Emoticon Sentiment Lexicon* [Hogenboom et al. 2013], que detém 476 *emoticons*, distribuídos em 179 com sentimento positivo, 278 com sentimento negativo e 20 com sentimento neutro. Neste caso, a estratégia é verificar se o *tweet* possui um *emoticon*. No caso de haver *emoticon* positivo, então o *tweet* é positivo. Já se o *emoticon* encontrado for negativo, o *tweet* é dado como negativo. Por fim, se detectado um *emoticon* neutro, o *tweet* será neutro. A Tabela 2 mostra alguns dos *emoticons* que expressam sentimentos.

Emoticons positivos	Emoticons negativos	Emoticons neutros
:), :'), :'-), :-), :-))	:*(, :-c, :/, :-/, :{	(o;, :l, :-O, ;)
:-*, :-}, :}, :D, :-D, :o)	#-(, :#, #(, %(	;o), ¡:}, }:-), }:)

Tabela 2. *Emoticons* de sentimento pertencentes ao Emoticon Sentiment Lexicon.

Part-of-Speech: Trata-se de uma técnica que permite categorizar cada palavra na respectiva classe sintática, como: verbo, pronome, advérbio, entre outros. A Tabela 3 mostra alguns exemplos de *tags* [Jurafsky and Martin 2014]. Muitos pesquisadores aplicaram *part-of-speech* em seus trabalhos [Go et al. 2009, Barbosa and Feng 2010, Aisopos et al. 2011]. Neste método foi aplicado o pacote de *Stanford* [Manning et al. 2014]. Nesta fase, utilizamos uma técnica chamada tokenização, que divide um *tweet* em palavras e pontuações, então atribuímos a cada *token* uma *tag*. Após a rotulação de cada *tweet*, é feita uma contagem de cada *tag* em cada *tweet*, desse modo, criamos uma matriz que foi concatenada ao modelo de *bag-of-words*.

Tag	Descrição	Exemplo
CC	conjunção	*e, ou, mas*
JJ	adjetivo	*bom, brilhante*
NN	substantivo	*felicidade, livro*
RB	advérbio	*hoje, já, sim*
VB	verbo	*falou, falaria*
SYM	símbolo	*+, %, #*

Tabela 3. Tags do corpus de *Stanford*.

Combinações de métodos: As mesmas estratégias ditas anteriormente foram mantidas, porém, combinamos técnicas na construção do modelo de classificação. As combinações foram realizadas da seguinte forma: *emoticons* + léxico (neste caso escolhemos o *Opinion Lexicon*), os 3 dicionários léxicos, e por último, *part-of-speech* + *emoticons* + léxico (novamente optamos pelo *Opinion Lexicon*).

Para validar os modelos de cada experimento, aplicamos a técnica de validação cruzada [Witten et al. 2016]. Esta técnica consiste em dividir a base de dados em *k* partes, essas partes se chamam *folds*. Uma dessas partes será usada para a fase de treinamento,

isso é feito repetidamente até que o modelo seja treinado e testado com todas as partes. Nos experimentos realizados na base de dados *Sanders*, utilizamos 10 *folds*, dessa forma, o modelo foi treinado com 9 partes e testado com a parte restante. Essa técnica evita problemas de variância nos dados. Enquanto na base de dados *HCR*, nós não aplicamos validação cruzada, pois esta conta com uma base de *tweets* exclusivamente para teste.

Todo o processo de análise de sentimentos, bem como onde se aplica cada método, pode ser visto na Figura 1.

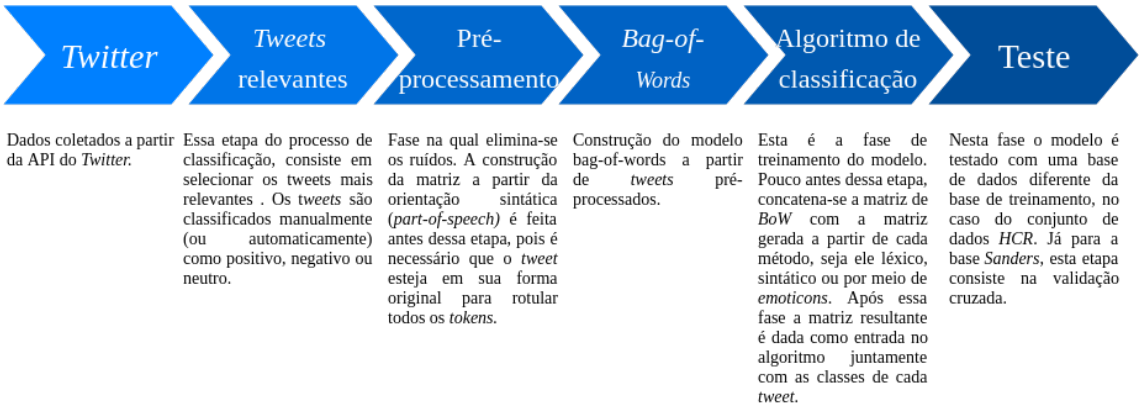


Figura 1. Passos para classificação de *tweets*.

6. Resultados

Os resultados obtidos pelos métodos e pelos algoritmos propostos foram analisados, com o objetivo de sabermos qual configuração se saiu melhor em cada base de dados. Realçamos nas comparações a acurácia, o modelo, o classificador e o atributo (*n-grams*). Tratamos, também, de realizar uma comparação especial entre os dicionários léxicos utilizados no trabalho, afim de responder as questões que levantamos, bem como, entender a diferença de resultados.

Apontamos, em negrito, o melhor resultado de cada configuração. Dentre os resultados, selecionamos o maior para contrastar com a literatura. A Tabela 4 mostra os resultados da base *Sanders*. Enquanto a Tabela 5 apresenta os resultados da base *HCR*.

Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra		
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)
Experimento 1: ***Bag-of-Words***											
Unigram	MNB	70,48	43,22	29,65	35,17	55,14	62,48	58,58	78,96	81,86	80,38
	SVM	76,03	55,45	**50,00**	**52,58**	67,33	57,12	**61,81**	81,83	86,90	84,29
	RF	75,09	61,62	30,70	40,98	70,00	46,09	55,59	77,09	92,77	84,21
	LR	77,08	66,46	37,54	47,98	75,17	49,62	59,78	78,51	93,25	85,24
Bigram	MNB	65,46	35,00	46,67	40,00	49,41	64,17	55,83	82,81	70,08	75,91
	SVM	75,39	60,45	37,54	46,32	72,05	40,28	51,67	77,55	93,17	84,65
	RF	71,28	48,45	21,93	30,19	74,56	26,03	38,59	72,87	94,33	82,22
	LR	75,55	74,11	29,12	41,81	76,01	37,37	50,10	75,61	**96,08**	84,62
Uni+Big	MNB	68,84	43,00	45,79	44,35	51,43	**71,82**	59,94	**83,14**	73,31	77,92
	SVM	76,89	60,57	44,74	51,46	72,73	52,68	61,10	80,01	90,53	84,95
	RF	73,81	59,15	22,11	32,18	76,68	36,75	49,69	74,50	95,25	83,61
	LR	**78,10**	**72,03**	39,30	50,85	**79,50**	48,70	60,40	78,54	94,61	**85,83**
Experimento 2: ***Bag-of-Words + Part-of-Speech Stanford***											
Unigram	MNB	71,93	45,86	12,63	19,81	61,19	55,28	58,09	75,43	89,77	81,98
	SVM	76,22	56,10	**48,42**	51,98	66,95	**59,88**	63,22	**82,00**	86,82	84,34
	RF	73,11	66,89	17,72	28,02	70,34	35,22	46,94	73,68	95,61	83,22
	LR	**77,40**	65,04	39,82	49,40	72,38	54,98	62,49	79,76	91,81	**85,36**
Bigram	MNB	71,93	54,92	11,75	19,36	61,18	46,09	52,28	74,29	92,37	82,35
	SVM	75,60	62,27	41,40	49,74	65,48	47,93	55,35	79,05	90,61	84,44
	RF	71,31	63,56	13,16	21,80	68,75	23,58	35,12	71,75	97,00	82,49
	LR	75,58	69,50	31,58	43,43	66,37	46,25	54,51	77,49	93,25	84,64
Uni+Big	MNB	72,97	56,52	20,53	30,12	60,35	58,50	59,41	76,92	88,69	82,39
	SVM	77,29	61,08	45,44	**52,11**	70,15	57,58	**63,25**	81,16	89,69	85,22
	RF	72,62	**71,20**	15,61	25,61	**76,11**	26,34	39,14	72,44	**97,68**	83,19
	LR	77,27	67,89	38,95	49,50	71,43	52,83	60,74	79,29	92,37	85,33
Experimento 3: ***Bag-of-Words + Opinion Lexicon***											
Unigram	MNB	72,84	50,80	33,51	40,38	59,15	63,86	61,41	79,62	84,14	81,82
	SVM	76,38	56,44	50,00	53,02	66,84	60,80	63,67	82,38	86,46	84,37
	RF	75,60	62,21	33,51	43,56	67,91	52,83	59,43	78,36	91,13	84,26
	LR	77,72	65,83	41,23	50,70	74,68	53,75	62,51	79,68	92,29	85,52
Bigram	MNB	64,30	33,43	**58,60**	42,57	53,32	66,46	59,17	**85,10**	65,04	73,73
	SVM	75,52	67,19	37,72	48,31	64,43	49,92	56,26	78,38	90,81	84,14
	RF	72,30	74,13	18,60	29,73	65,45	33,08	43,95	72,92	94,77	82,42
	LR	75,74	**77,03**	28,25	41,34	74,93	40,28	52,39	75,74	**95,81**	84,60
Uni+Big	MNB	69,57	44,38	51,23	47,56	52,63	**73,51**	61,34	84,42	72,71	78,13
	SVM	78,15	63,35	49,12	**55,34**	73,08	58,19	**64,79**	81,48	89,97	85,51
	RF	74,53	69,74	23,86	35,56	72,70	41,19	52,59	75,04	94,77	83,76
	LR	**78,85**	70,06	42,28	52,74	**76,30**	56,20	64,73	80,32	93,09	**86,23**
Experimento 4: ***Bag-of-Words + SemEval2015-Lexicon***											
Unigram	MNB	71,44	46,17	30,70	36,88	58,05	61,87	59,90	78,57	83,22	80,83
	SVM	75,68	52,65	48,77	50,64	68,52	56,66	**62,03**	81,72	86,78	84,17
	RF	75,85	66,92	31,23	42,58	74,11	44,72	55,78	76,84	94,13	84,61
	LR	76,65	64,11	36,67	46,65	74,00	51,00	60,38	78,44	92,45	84,87
Bigram	MNB	61,76	30,66	**59,12**	40,38	51,19	**59,26**	54,93	**84,29**	63,00	72,11
	SVM	74,40	68,85	29,47	1,28	72,36	35,68	47,79	75,03	94,73	83,74
	RF	71,85	72,59	17,19	27,80	75,40	21,59	33,57	71,62	97,40	82,55
	LR	74,37	**78,95**	23,68	36,44	**84,62**	28,64	42,79	73,46	**97,84**	83,91
Uni+Big	MNB	69,16	42,92	48,95	45,74	52,40	73,51	61,19	84,17	72,63	77,98
	SVM	77,32	63,07	46,14	**53,29**	72,84	52,99	61,35	80,17	90,77	85,14
	RF	74,40	72,68	23,33	35,33	75,40	35,68	48,44	74,40	96,12	83,88
	LR	**77,67**	72,67	37,89	49,83	76,00	49,46	59,93	78,40	94,09	**85,53**
Experimento 5: ***Bag-of-Words + Senticnet***											
Unigram	MNB	71,66	45,17	27,89	34,49	59,05	60,95	59,98	78,26	84,42	81,22
	SVM	75,66	55,36	48,95	51,96	66,06	56,05	**60,65**	81,48	86,86	84,08
	RF	73,81	62,69	29,47	40,10	63,62	43,64	51,77	76,31	91,77	83,33
	LR	76,62	67,11	35,79	46,68	73,41	49,46	59,10	78,07	93,01	84,89
Bigram	MNB	60,74	31,02	**56,49**	40,05	46,30	**58,50**	51,69	83,68	62,29	71,42
	SVM	72,73	63,72	25,26	36,18	71,71	27,57	39,82	73,44	95,33	82,96
	RF	70,80	50,39	22,63	31,23	80,89	19,45	31,36	71,90	95,17	81,91
	LR	73,11	**80,56**	20,35	32,49	**85,06**	22,66	35,79	72,18	**98,28**	83,23
Uni+Big	MNB	68,71	43,55	49,12	46,17	51,08	72,13	59,81	**83,71**	72,27	77,57
	SVM	**77,35**	64,63	44,56	**52,75**	72,90	51,91	60,64	79,81	91,45	85,24
	RF	74,02	69,10	21,58	32,89	72,54	38,44	50,25	74,45	95,25	83,58
	LR	77,32	72,95	35,96	48,18	76,46	48,24	59,15	77,84	94,33	**85,30**
Experimento 6: ***Bag-of-Words + Emoticon Sentiment Lexicon***											
Unigram	MNB	71,28	44,78	26,32	33,15	58,14	60,18	59,14	77,83	84,42	80,99
	SVM	75,95	56,68	49,12	**52,63**	67,17	54,82	**60,37**	81,22	87,57	84,28
	RF	74,37	62,74	28,95	39,62	65,11	44,87	53,13	76,77	92,41	83,87
	LR	76,97	68,11	35,96	47,07	74,72	50,69	60,00	78,20	93,17	85,03
Bigram	MNB	60,82	28,70	**55,44**	37,82	49,49	51,91	50,67	83,04	64,36	72,52
	SVM	73,11	67,97	27,54	39,20	73,53	26,80	39,28	73,44	95,57	83,06
	RF	72,14	70,12	20,18	31,34	72,55	22,66	34,54	72,22	96,88	82,75
	LR	72,79	**79,69**	17,89	29,23	**82,86**	22,21	35,02	72,01	**98,48**	83,19
Uni+Big	MNB	67,77	42,04	46,32	44,07	49,57	**71,36**	58,51	**83,18**	71,71	77,02
	SVM	76,81	64,25	43,51	51,88	71,77	50,23	59,10	79,29	91,33	84,89
	RF	73,94	65,82	22,63	33,68	73,95	37,83	50,05	74,44	95,05	83,49
	LR	**77,91**	73,20	37,37	49,48	78,41	48,39	59,85	78,30	94,85	**85,78**
Experimento 7: ***Bag-of-Words + Opinion Lexicon + Emoticon Sentiment Lexicon***											
Unigram	MNB	73,24	50,56	31,40	38,74	62,17	62,17	62,17	78,85	85,66	82,11
	SVM	76,95	56,29	51,05	53,54	68,33	60,80	**64,34**	**82,91**	87,06	84,93
	RF	75,76	63,73	32,98	43,47	67,82	53,91	60,07	78,40	91,21	84,32
	LR	78,15	65,87	43,68	52,53	75,05	54,82	63,36	80,29	92,09	85,78
Bigram	MNB	66,43	36,13	**52,11**	42,67	54,09	62,79	58,11	82,39	70,64	76,06
	SVM	75,79	67,18	38,07	48,60	65,74	50,84	57,34	78,50	90,89	84,24
	RF	73,22	73,65	19,12	30,36	70,90	35,07	46,93	73,43	95,49	83,01
	LR	75,50	**77,08**	25,96	38,85	74,37	40,89	52,77	75,53	**95,81**	84,47
Uni+Big	MNB	70,59	45,83	45,26	45,54	54,22	**71,82**	61,79	82,81	76,03	79,28
	SVM	77,64	61,56	48,60	**54,31**	71,46	57,89	63,96	81,47	89,41	85,26
	RF	75,31	68,78	24,74	36,39	76,17	45,02	56,59	75,63	94,73	84,11
	LR	**78,56**	69,03	42,63	52,71	**76,55**	54,98	63,99	80,03	92,89	**85,98**
Experimento 8: ***Bag-of-Words + Opinion Lexicon + SemEval2015-Lexicon + Senticnet***											
Unigram	MNB	73,56	51,03	30,35	38,06	63,17	60,95	62,04	78,71	86,70	82,51
	SVM	75,71	54,49	50,00	52,15	66,79	57,58	61,84	81,82	86,30	84,00
	RF	75,50	65,44	31,23	42,28	70,68	47,63	56,91	77,11	92,85	84,25
	LR	77,78	64,44	42,28	51,06	75,48	54,21	63,10	79,88	92,01	85,52
Bigram	MNB	67,77	37,06	46,49	41,25	55,63	61,26	58,31	81,15	74,31	77,58
	SVM	75,90	67,57	39,47	49,83	66,17	47,93	55,60	78,42	91,49	84,46
	RF	72,41	65,31	16,84	26,78	65,87	38,13	48,30	73,51	94,01	82,50
	LR	75,34	**75,83**	28,07	40,97	69,21	44,41	54,10	76,13	94,17	84,19
Uni+Big	MNB	71,98	48,08	46,14	47,09	56,29	**70,60**	62,64	**82,97**	78,23	80,53
	SVM	**78,40**	64,43	**50,53**	**56,64**	72,11	58,19	**64,41**	81,87	80,01	85,75
	RF	75,47	73,60	22,98	35,03	**77,62**	43,03	55,37	75,33	**95,88**	84,37
	LR	78,34	69,01	42,98	52,97	74,63	54,06	62,70	80,09	92,73	**85,95**
Experimento 9: ***Bag-of-Words + Part-of-Speech + Opinion Lexicon + Emoticon Sentiment Lexicon***											
Unigram	MNB	72,97	51,98	16,14	24,63	65,46	55,14	60,03	75,60	90,49	82,38
	SVM	76,54	55,71	**49,65**	52,50	67,92	**60,64**	64,08	**82,47**	86,82	84,59
	RF	74,96	73,44	24,74	37,01	72,15	41,65	52,82	75,39	95,09	84,10
	LR	77,99	65,96	43,86	52,69	71,59	57,12	63,54	80,79	91,11	85,68
Bigram	MNB	72,03	65,66	11,40	19,43	**78,35**	30,47	43,88	71,75	96,68	82,37
	SVM	76,73	62,70	41,58	50,00	68,60	52,53	59,50	80,02	91,05	85,18
	RF	72,79	**78,22**	13,86	23,55	76,19	29,40	42,43	72,37	**97,52**	83,08
	LR	76,44	67,69	34,91	46,06	69,87	50,08	58,34	78,34	92,77	84,95
Uni+Big	MNB	74,15	62,43	20,70	31,09	65,90	57,12	61,20	76,47	90,77	83,01
	SVM	78,34	63,35	49,12	**55,34**	70,66	60,49	**65,18**	82,35	89,65	85,85
	RF	74,02	71,34	20,53	31,88	76,97	35,83	48,90	73,88	96,16	83,56
	LR	**78,48**	68,01	44,39	53,72	72,88	58,04	64,62	80,88	91,57	**85,89**

Tabela 4. Resultados referente a base *Sanders*.

Experimento 1: ***Bag-of-Words***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	60,71	59,71	**56,46**	**58,04**	62,99	76,29	69,01	52,46	31,37	39,26	
	SVM	59,04	51,37	52,08	51,72	66,54	70,42	68,42	49,38	41,24	44,94	
	RF	58,42	52,11	26,81	35,41	60,48	90,73	72,58	50,00	20,00	28,57	
	LR	62,58	57,25	54,35	55,76	**76,72**	76,28	71,75	50,77	36,67	42,58	
Bigram	MNB	53,01	48,91	50,76	49,81	72,87	55,92	63,28	32,69	**49,04**	39,23	
	SVM	62,16	62,14	44,44	51,82	63,56	85,83	73,03	48,57	20,48	28,81	
	RF	55,72	**90,00**	12,33	21,69	55,00	**98,37**	70,55	38,10	08,99	14,55	
	LR	60,29	63,53	38,03	47,58	59,52	89,16	71,38	**60,87**	15,56	24,78	
Uni+Big	MNB	60,91	54,29	54,68	54,48	68,75	69,84	69,29	48,24	45,56	46,86	
	SVM	**64,47**	62,81	51,70	56,72	67,47	80,25	73,31	53,52	41,76	**46,91**	
	RF	54,05	72,00	12,08	20,69	53,99	96,64	69,28	40,00	12,77	19,35	
	LR	62,79	57,04	55,80	56,41	67,34	79,05	72,73	51,02	27,78	35,97	
Experimento 2: ***Bag-of-Words + Part-of-Speech Stanford***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	58,00	59,38	36,77	45,42	57,81	88,66	69,98	55,00	12,50	20,37	
	SVM	58,21	48,97	52,99	50,90	67,19	71,43	69,25	46,99	35,78	40,63	
	RF	57,38	54,67	28,67	37,61	57,80	88,84	70,03	58,82	20,83	30,77	
	LR	61,75	56,91	45,45	50,54	64,58	78,48	70,54	58,57	45,56	**51,25**	
Bigram	MNB	57,17	56,67	26,36	35,98	57,18	94,09	71,13	66,67	02,04	03,96	
	SVM	57,17	48,04	35,25	40,66	60,06	80,93	68,95	57,38	33,02	41,92	
	RF	57,80	**95,24**	14,29	24,84	55,33	**100,00**	71,24	**90,00**	09,78	17,65	
	LR	60,91	52,58	36,43	43,04	63,94	85,77	**73,26**	57,41	32,63	41,61	
Uni+Big	MNB	60,71	57,45	38,75	46,15	60,76	91,39	73,00	75,00	15,46	25,64	
	SVM	59,46	56,56	45,39	50,36	**64,68**	72,50	68,37	47,78	**48,31**	48,04	
	RF	56,96	85,19	15,13	25,70	55,35	97,14	70,52	54,17	15,48	24,52	
	LR	**63,20**	63,03	**55,97**	**59,29**	63,28	84,28	72,28	63,16	30,51	41,14	
Experimento 3: ***Bag-of-Words + Opinion Lexicon***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	60,50	51,27	62,31	56,25	67,50	75,00	71,05	48,84	21,21	29,58	
	SVM	64,24	55,06	**64,93**	**59,59**	**72,36**	72,36	72,36	57,14	**43,56**	49,44	
	RF	59,67	61,19	29,50	39,81	59,40	91,21	71,95	59,57	27,18	37,33	
	LR	64,24	60,43	57,14	58,74	66,90	79,84	72,80	59,62	34,07	43,36	
Bigram	MNB	60,08	57,58	60,90	59,19	69,51	67,98	68,74	41,94	40,21	41,05	
	SVM	62,37	55,71	60,47	57,99	66,12	83,06	73,63	56,76	19,09	28,57	
	RF	56,34	**80,77**	15,67	26,25	55,15	**97,57**	70,47	50,00	09,00	15,25	
	LR	61,75	61,29	49,35	54,68	63,49	83,68	72,20	50,00	23,86	32,31	
Uni+Big	MNB	62,58	58,50	59,72	59,11	70,45	72,20	71,31	47,13	42,71	44,81	
	SVM	**65,70**	58,33	55,80	57,04	68,15	82,57	**74,67**	**70,18**	39,22	**50,31**	
	RF	60,50	66,67	36,00	46,75	58,68	94,09	72,29	70,00	14,89	24,56	
	LR	62,79	59,15	56,00	57,53	65,12	83,05	73,00	57,89	23,16	33,08	
Experimento 4: ***Bag-of-Words + SemEval2015-Lexicon***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	56,21	50,38	49,25	49,81	66,20	73,64	69,72	38,10	26,97	31,58	
	SVM	62,37	57,42	57,05	57,23	69,29	75,54	72,28	48,61	38,04	42,68	
	RF	62,79	62,50	37,16	46,61	63,06	86,78	73,04	71,67	40,66	**49,01**	
	LR	63,62	59,54	54,17	56,73	65,32	83,26	73,21	64,15	32,69	43,31	
Bigram	MNB	55,51	55,03	52,56	53,77	65,77	61,09	63,34	35,45	**45,35**	39,80	
	SVM	59,88	60,19	44,29	51,03	61,03	82,11	70,02	51,06	25,26	33,80	
	RF	51,77	**85,19**	16,08	27,06	49,89	**98,64**	66,26	47,06	06,84	11,94	
	LR	57,38	51,04	36,03	42,24	58,47	89,17	70,63	**68,42**	12,38	20,97	
Uni+Big	MNB	62,37	52,73	**67,97**	**59,39**	**74,46**	67,19	70,64	48,24	42,27	45,05	
	SVM	58,84	46,56	46,56	46,56	65,85	73,33	69,39	53,03	36,84	43,48	
	RF	60,91	71,70	28,15	40,43	59,51	95,63	**73,36**	60,87	14,89	23,93	
	LR	**64,03**	62,02	54,42	57,97	65,75	82,76	73,28	60,00	35,29	44,44	
Experimento 5: ***Bag-of-Words + Senticnet***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	60,91	52,03	49,23	50,59	65,16	81,78	72,53	56,25	25,96	35,53	
	SVM	61,75	52,05	58,91	55,27	68,75	72,43	70,54	56,96	**41,28**	**47,87**	
	RF	58,84	64,71	38,19	48,03	60,47	88,14	71,72	38,46	19,80	26,14	
	LR	**64,86**	62,22	57,53	**59,79**	66,33	81,40	73,10	63,27	33,33	43,66	
Bigram	MNB	54,89	47,80	54,68	51,01	68,12	63,67	65,82	34,41	32,99	33,68	
	SVM	56,76	58,76	38,78	46,72	57,66	82,40	67,72	47,06	23,76	31,58	
	RF	55,72	65,38	12,69	21,25	55,13	96,41	70,14	56,25	09,38	16,07	
	LR	59,88	60,67	39,13	47,58	59,25	89,11	71,18	**68,42**	13,68	22,81	
Uni+Big	MNB	62,16	51,63	**62,70**	56,63	**71,59**	73,54	72,55	48,44	31,63	38,27	
	SVM	61,54	58,45	54,36	55,86	63,73	79,66	70,81	60,00	28,12	38,30	
	RF	59,67	**82,76**	17,91	29,45	59,12	**97,63**	72,86	59,26	17,02	26,45	
	LR	64,66	74,19	44,52	55,65	63,30	90,00	**74,33**	57,38	36,46	44,59	
Experimento 6: ***Bag-of-Words + Emoticon Sentiment Lexicon***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	63,20	52,03	59,23	55,40	69,89	75,88	72,76	59,26	34,04	43,24	
	SVM	59,04	48,30	55,04	51,45	70,20	67,45	68,80	46,07	42,27	**44,09**	
	RF	57,38	68,42	27,08	38,81	55,75	93,56	69,87	57,58	18,27	27,74	
	LR	61,75	54,00	57,45	55,67	69,43	73,31	71,32	48,48	35,96	41,29	
Bigram	MNB	55,93	54,90	57,53	56,19	69,34	58,80	63,64	32,76	**44,71**	37,81	
	SVM	58,21	64,00	32,00	42,67	57,69	88,61	69,88	52,38	23,40	32,35	
	RF	52,39	62,86	14,77	23,91	51,61	**97,40**	67,47	50,00	04,95	09,01	
	LR	58,00	**76,92**	32,47	45,66	54,00	96,43	69,23	**81,25**	12,62	21,85	
Uni+Big	MNB	61,75	55,48	**62,32**	**58,70**	**70,97**	71,26	71,11	44,87	36,46	40,23	
	SVM	60,71	53,44	54,69	54,05	67,24	78,80	72,56	43,86	24,27	31,25	
	RF	55,93	72,92	21,88	33,65	54,34	93,19	68,65	50,00	17,44	25,86	
	LR	**62,99**	62,11	40,97	49,37	61,92	91,03	**73,70**	73,81	30,10	42,70	
Experimento 7: ***Bag-of-Words + Opinion Lexicon + Emoticon Sentiment Lexicon***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	58,21	55,00	52,03	53,47	59,93	80,44	68,69	56,41	20,37	29,93	
	SVM	62,16	59,21	56,25	57,69	66,67	74,11	70,19	53,75	**44,33**	**48,59**	
	RF	59,04	62,89	41,22	49,80	58,52	90,35	71,03	53,12	16,19	24,82	
	LR	62,79	53,44	56,91	55,12	68,38	80,24	73,84	55,93	30,00	39,05	
Bigram	MNB	60,08	55,13	**61,87**	58,31	65,57	75,85	70,33	46,15	22,64	30,38	
	SVM	63,20	56,21	59,72	57,91	68,09	78,05	72,73	56,52	28,57	37,96	
	RF	56,96	56,14	43,84	49,23	57,06	85,59	68,47	61,54	08,08	14,29	
	LR	64,24	60,74	57,34	58,99	64,09	87,34	73,93	**86,96**	19,80	32,26	
Uni+Big	MNB	64,45	56,95	63,24	59,93	**69,34**	78,19	73,50	60,71	33,33	43,04	
	SVM	64,24	62,32	59,31	60,78	67,26	81,47	73,68	54,84	32,69	40,96	
	RF	58,84	59,15	31,11	40,78	59,34	**90,62**	71,72	47,37	10,00	16,51	
	LR	**65,07**	**69,09**	55,07	**61,29**	64,51	88,56	**74,64**	59,57	26,17	36,36	
Experimento 8: ***Bag-of-Words + Opinion Lexicon + SemEval2015-Lexicon + Senticnet***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	62,16	59,18	40,00	47,74	65,76	86,11	74,57	45,28	28,57	35,04	
	SVM	61,33	55,12	49,30	52,04	68,79	75,49	71,99	43,06	37,80	40,26	
	RF	60,08	54,21	43,61	48,33	62,24	84,40	71,65	57,14	20,41	30,08	
	LR	64,86	57,02	55,73	56,37	72,11	79,10	**75,44**	45,76	32,93	38,30	
Bigram	MNB	60,71	57,14	56,34	56,74	64,95	78,10	70,92	46,00	23,71	31,29	
	SVM	62,99	56,67	**59,03**	57,82	67,99	78,10	72,69	54,72	30,53	39,19	
	RF	56,13	53,06	36,36	43,15	57,14	84,55	68,20	52,63	10,87	18,02	
	LR	59,46	59,52	50,68	54,74	60,19	85,53	70,65	51,61	15,24	23,53	
Uni+Big	MNB	62,99	59,35	51,05	54,89	68,40	78,49	73,10	47,14	37,93	42,04	
	SVM	**65,90**	66,17	56,77	**61,11**	68,73	81,12	74,41	54,79	**43,01**	**48,19**	
	RF	60,71	**66,67**	33,99	45,02	59,95	**93,00**	72,90	53,85	16,47	25,23	
	LR	64,66	55,73	57,03	56,37	**69,23**	82,14	75,14	**60,78**	30,69	40,79	
Experimento 9: ***Bag-of-Words + Part-of-Speech + Opinion Lexicon + Emoticon Sentiment Lexicon***												
Atributo	Algoritmo	Acc. (%)	Classe Positiva			Classe Negativa			Classe Neutra			
			Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	Precision (%)	Recall (%)	F1-score (%)	
Unigram	MNB	60,08	54,87	46,27	50,20	61,05	91,70	73,30	**100,00**	05,66	10,71	
	SVM	57,59	48,99	50,69	49,83	65,37	69,14	67,20	48,00	38,30	42,60	
	RF	59,25	58,97	32,62	42,01	58,93	89,84	71,18	64,29	19,15	29,51	
	LR	62,79	58,87	50,34	54,28	65,62	77,46	71,05	57,97	**43,48**	49,69	
Bigram	MNB	55,51	**78,26**	24,49	37,31	53,26	**97,03**	68,77	40,00	02,04	03,88	
	SVM	61,12	59,20	50,68	54,61	65,12	77,54	70,79	49,33	37,37	42,53	
	RF	57,28	60,26	31,54	41,41	57,78	87,95	69,75	41,67	12,05	18,69	
	LR	60,08	51,64	47,01	49,22	65,00	78,95	71,30	52,54	31,00	38,99	
Uni+Big	MNB	60,91	58,33	40,00	47,46	60,33	90,98	72,55	88,24	15,46	26,32	
	SVM	**66,32**	55,19	**65,89**	**60,07**	**73,21**	75,78	**74,47**	64,52	41,67	**50,63**	
	RF	58,84	61,43	30,94	41,15	58,40	93,20	71,80	58,33	07,61	13,46	
	LR	62,58	57,60	50,00	53,53	69,18	79,22	73,86	42,19	32,93	36,99	

Tabela 5. Resultados referente a base HCR.

Na base de *tweets Sanders* nota-se que na maioria dos modelos o classificador LR (*Logistic Regression*) de descatou. Apenas nos experimentos 5 e 8 o melhor classificador foi o SVM (*Support Vector Machine*), com uma acurácia de 77.35% e 78.40%, respectivamente. Se tratando de engenharia de atributos, em 8 dos 9 experimentos apresentaram a melhora combinação como sendo unigrama + bigrama. Somente o experimento 2 mostrou-se melhor com unigrama. Dentre todos os teses, o experimento 3 alcançou a maior acurácia com 78,85%. Em comparação com o resultado de Aston em [Aston et al. 2014], os autores angariaram uma acurácia de 73,30%. Alcançamos uma acurácia melhor devido ao nosso método e, também, ao uso de unigrama + bigrama que aumenta a ocorrência de termos. Enquanto os autores utilizaram uma metodologia chamada MBW (*Modified Balanced Winnow*) com trigrama.

Enquanto no corpus de *tweets HCR*, observa-se que dos 4 classificadores, 2 se destacaram por obterem a melhor acurácia, são eles: SVM (*Support Vector Machine*) e LR (*Logistic Regression*). Além do mais, 8 dos experimentos se saíram melhor com a união de unigrama e bigrama, houve apenas um caso em que o melhor resultado foi com unigrama. O experimento 9 se destacou como a melhor configuração, atingindo 66,32% de acurácia. Para comparação, selecionamos na literatura a melhor acurácia obtida na base *HCR*, 78,67% [Hassan Saif, Miriam Fernandez and Alani 2013]. Este resultado foi melhor devido a configuração adotada, isto é, os autores mesclaram todos os conjuntos de dados e aplicaram validação cruzada, isso possibilitou uma acurácia mais expressiva. Já neste trabalho, na fase de teste foi aplicado um conjunto de dados exclusivo. Portanto, conclui-se que construir uma única base de dados pode alterar o resultado final.

Em um levantamento realizado foi constatado que existem 243 palavras em comum entre os léxicos presentes neste artigo. Dessas palavras, 14 contradizem-se em seu sentimento, como pode ser visto na Tabela 6. No entanto, essas palavras contraditórias não influenciaram os resultados. Comparando os léxicos entre si, nota-se que em ambas as bases de dados, aquele que alcançou a melhor predição foi o *Opinion Lexicon*, com uma acurácia de 78,85% na *Sanders*, e uma acurácia de 65,70% no *HCR*. O *SenticNet* e o *SemEval2015-Lexicon* mostraram dados inferiores. Inferimos, assim, que o desempenho do *Opinion Lexicon* se deve ao fato dos termos subjetivos determinarem um papel fundamental na classificação do sentimento.

Palavras	SemEval2015	SenticNet	Opinion Lexicon
abysmal	negativo	positivo	negativo
enough	negativo	positivo	negativo
flirt	positivo	positivo	negativo
funny	positivo	positivo	negativo
hang	negativo	positivo	negativo
hot	positivo	negativo	positivo
joke	positivo	positivo	negativo
layghable	positivo	negativo	negativo
mediocrity	negativo	positivo	negativo
miss	positivo	negativo	negativo
pleasantly	positivo	negativo	positivo
tired	negativo	positivo	negativo
waste	negativo	positivo	negativo
work	negativo	negativo	positivo

Tabela 6. Palavras contraditórias.

7. Conclusões e trabalhos futuros

Neste trabalho foi apresentado novas configurações na classificação de *tweets* a partir do modelo *bag-of-words*. Métodos já presentes na literatura foram utilizados como: aprendizado de máquina, dicionários léxicos, orientação sintática e *emoticons*. O objetivo central foi explorar novas aplicações dos métodos citados, como também realizar comparações entre os resultados obtidos e resultados de outros trabalhos. Em especial, uma outra comparação direcionada aos dicionários léxicos foi mostrada com o intuito de entender a diferença entre eles. O resultado desse contraste teve como melhor *corpus* de léxicos o *Opinion Lexicon*. Neste artigo foi escolhido 4 classificadores que são constantemente empregados em diversos estudos, como pode ser visto em [Hassan Saif, Miriam Fernandez and Alani 2013, SILVA 2016, Pawar and Deshmukh 2015].

Neste estudo, em cada base de dados foi realizado 9 experimentos e os algoritmos que galgaram a melhor acurácia foram SVM (*Support Vector Machine*) e LR (*Logistic Regression*). Em relação a engenharia de atributos aplicada aos modelos, mostrou-se presente nos resultados destacados a união de unigrama e bigrama.

Como trabalhos futuros pretendemos aplicar algoritmos de regressão, os quais pouco se tem explorado [Nasim 2017, Jiang et al. 2017]. A hipótese é que a análise de sentimentos pode ser melhor abordada quando considerada de forma"não rígida" (a classificação não fica restrita somente às classes positiva, negativa e neutra), isto é, quando a opinião extraída está em um *ranting* de valores (por exemplo $[0, 1]$) discriminaria melhor o seu real sentimento – um texto com um *ranting* de sentimento $= 0.8$ é mais positivo que um texto com *ranting* de sentimento $= 0.6$.

Referências

Aisopos, F., Papadakis, G., and Varvarigou, T. (2011). Sentiment analysis of social media content using n-gram graphs. In *Proceedings of the 3rd ACM SIGMM International Workshop on Social Media*, WSM '11, pages 9–14, New York, NY, USA. ACM.

Araújo, M., Gonçalves, P., and Benevenuto, F. (2013). Measuring sentiments in online social networks. In *Proceedings of the 19th Brazilian Symposium on Multimedia and the Web*, WebMedia '13, pages 97–104, New York, NY, USA. ACM.

Aston, N., Munson, T., Liddle, J., Hartshaw, G., Livingston, D., and Hu, W. (2014). Sentiment analysis on the social networks using stream algorithms. *Journal of Data Analysis and Information Processing*, 2(02):60.

Baccianella, S., Esuli, A., and Sebastiani, F. (2010). Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. In Calzolari, N., Choukri, K., Maegaard, B., Mariani, J., Odijk, J., Piperidis, S., Rosner, M., and Tapias, D., editors, *LREC*. European Language Resources Association.

Barbosa, L. and Feng, J. (2010). Robust sentiment detection on twitter from biased and noisy data. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, COLING '10, pages 36–44, Stroudsburg, PA, USA. Association for Computational Linguistics.

Cambria, E., Poria, S., Bajpai, R., and Schuller, B. W. (2016). Senticnet 4: A semantic resource for sentiment analysis based on conceptual primitives. In *COLING*.

Chaovalit, P. and Zhou, L. (2005). Movie review mining: A comparison between supervised and unsupervised classification approaches. In *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*, pages 112c–112c. IEEE.

da Silva, N. F., Hruschka, E. R., and Hruschka, E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decis. Support Syst.*, 66(C):170–179.

Gimpel, K., Schneider, N., O'Connor, B., Das, D., Mills, D., Eisenstein, J., Heilman, M., Yogatama, D., Flanigan, J., and Smith, N. A. (2011). Part-of-speech tagging for twitter: Annotation, features, and experiments. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers - Volume 2*, HLT '11, pages 42–47, Stroudsburg, PA, USA. Association for Computational Linguistics.

Go, A., Bhayani, R., and Huang, L. (2009). Twitter sentiment classification using distant supervision. *CS224N Project Report, Stanford*, 1(2009):12.

Hardeniya, T. and Borikar, D. A. (2016). An approach to sentiment analysis using lexi cons with comparative analysis of different techniques. *IOSR Journals*, 18(3):53–57.

Hassan Saif, Miriam Fernandez, Y. H. and Alani, H. (2013). Evaluation datasets for twitter sentiment analysis: A survey and a new dataset, the sts-gold. *first ESSEM workshop*.

Hogenboom, A., Bal, D., Frasincar, F., Bal, M., de Jong, F., and Kaymak, U. (2013). Exploiting emoticons in sentiment analysis. In *Proceedings of the 28th Annual ACM Symposium on Applied Computing*, pages 703–710. ACM.

Hu, M. and Liu, B. (2004). Mining and summarizing customer reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '04, pages 168–177, New York, NY, USA. ACM.

Jang, J.-S. R., Sun, C.-T., and Mizutani, E. (1997). Neuro-fuzzy and soft computing: a computational approach to learning and machine intelligence.

Jiang, M., Lan, M., and Wu, Y. (2017). Ecnu at semeval-2017 task 5: An ensemble of regression algorithms with effective features for fine-grained sentiment analysis in financial domain. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 888–893. Association for Computational Linguistics.

Jurafsky, D. and Martin, J. H. (2014). *Speech and language processing*, volume 3. Pearson London.

Kiritchenko, S., Zhu, X., and Mohammad, S. M. (2014). Sentiment analysis of short informal texts. *Journal of Artificial Intelligence Research*, 50:723–762.

Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., and McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations*, pages 55–60.

Miller, G. A. (1995). Wordnet: A lexical database for english. *Commun. ACM*, 38(11):39–41.

Mohammad, S. M., Kiritchenko, S., and Zhu, X. (2013). Nrc-canada: Building the state-of-the-art in sentiment analysis of tweets. *CoRR*, abs/1308.6242.

Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources. In *EMNLP*, volume 4, pages 412–418.

Nasim, Z. (2017). Iba-sys at semeval-2017 task 5: Fine-grained sentiment analysis on financial microblogs and news. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 827–831. Association for Computational Linguistics.

Pang, B. and Lee, L. (2008). Opinion mining and sentiment analysis. *Found. Trends Inf. Retr.*, 2(1-2):1–135.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA. Association for Computational Linguistics.

Pawar, K. K. and Deshmukh, R. R. (2015). Twitter sentiment classification on sanders data using hybrid approach.

Perkins, J. (2014). *Python 3 Text Processing with NLTK 3 Cookbook*. Packt Publishing Ltd.

Reis, J. C., Gonçalves, P., Araújo, M., Pereira, A. C., and Benevenuto, F. (2015). Uma abordagem multilıngue para análise de sentimentos. *IV Brazilian*.

Sanders, N. J. (2011). Sanders-twitter sentiment corpus. *Sanders Analytics LLC*.

SILVA, N. F. F. (2016). Análise de sentimentos em textos curtos provenientes de redes sociais. Tese (Doutorado em Ciências de Computação e Matemática Computacional) - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Paulo, São Carlos. 2016.

Speriosu, M., Sudan, N., Upadhyay, S., and Baldridge, J. (2011). Twitter polarity classification with label propagation over lexical links and the follower graph. In *Proceedings of the First Workshop on Unsupervised Learning in NLP*, EMNLP '11, pages 53–63, Stroudsburg, PA, USA. Association for Computational Linguistics.

Tan, P.-N., Steinbach, M., and Kumar, V. (2005). *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

Witten, I. H., Frank, E., Hall, M. A., and Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Avaliação do Distúrbio Disfórico Pré-Menstrual e Transtorno Afetivo Bipolar uma abordagem usando Regras de Associação

Giovanna Aguiar de Castro[1] , Rogerio Salvini[1] e Rodrigo da Silva Dias[2]

1: Instituto de Informática
Universidade Federal de Goiás (UFG)
{giovannacastro, rogeriosalvini}@inf.ufg.br

2: Instituto de Psiquiatria
Universidade de São Paulo (USP)
rdgdias@gmail.com

Abstract. *Women with Bipolar Disorder (BD) may have worsening symptoms during events of their reproductive life, in particular, those who suffer from Premenstrual Dysphoric Disorder (PMDD), also know as Premenstrual Syndrome (PMS). The presence of PMDD is considered a marker of disease severity. In this study, are explored data from a cohort of 1099 women with BD to find out associations between BD symptoms and PMDD. The rules found shows some associations between PMDD and symptoms of the menstrual period and menstrual cycle of women. These associations, coming from a real database, can guide physicians in more depth studies on this subject.*

Resumo. *Mulheres com Transtorno Afetivo Bipolar (TAB) podem ter os sintomas agravados durante eventos de sua vida reprodutiva, em particular, as que sofrem de Distúrbio Disfórico Pré-Menstrual (PMDD), conhecido também como Tensão Pré-menstrual (TPM). A presença do PMDD nos diagnósticos de TAB é considerado um marcador de gravidade para a doença. Neste trabalho, são explorados dados de uma coorte de 1099 mulheres com TAB para achar associações entre sintomas do TAB e PMDD. Foram encontradas regras que fazem a algumas associações do PMDD com sintomas do período e ciclo menstrual da mulher. As associações encontradas a partir desta base de dados real podem nortear os médicos em estudos mais aprofundados sobre o tema.*

1. Introdução

O Transtorno Afetivo Bipolar (TAB) é uma desordem do humor na qual o paciente manifesta episódios depressivos e de mania, que pode causar impactos negativos na vida dessas pessoas. Possui um caráter crônico, que é capaz de gerar prejuízos na vida pessoal e profissional e levar a uma baixa qualidade de vida e até mesmo ao suicídio [Geddes R and Miklowitz J , Judd L et al. 2005].

A atuação do TAB se diferencia em muito na questão de gênero, se manifestando de forma desigual e acentuada no sexo feminino. O transtorno em mulheres é caracterizado pelo aumento do risco de instabilidade do humor durante os eventos da vida

reprodutiva feminina tal como o pré-menstrual, pré-natal e pré-menopausal. O início do distúrbio em mulheres pode ser identificado entre os 15 e 23 anos de idade, com ocorrências típicas de depressão precedidas por episódios de mania [Bergink et al. 2016, Cirillo C et al. 2012, Hu et al. 2016].

Alguns poucos estudos têm procurado entender a relação entre o TAB e o distúrbio disfórico pré-menstrual (*premenstrual disforic disorder* - PMDD), também conhecido como tensão pré-menstrual (TPM). O PMDD afeta de 3 a 9% das mulheres [Anastasiya Slyepchenko et al. 2017, Wittchen et al. 2002]. Ele foi recentemente reconhecido no DSM-5 (Manual Diagnóstico e Estatístico de Transtornos Mentais, 5ª edição) como um transtorno psiquiátrico independente [American Psychiatric Association 2013], enquanto que no ICD-10 (Classificação Internacional de Doenças, 10ª Revisão) ele é diagnosticado como "Outros distúrbios específicos do humor"[World Health Organization].

A partir de um grande estudo epidemiológico foi encontrado que mulheres com diagnóstico de PMDD são 8 vezes mais propensas a receberem o diagnóstico de TAB, comparado com as que não apresentaram PMDD [Anastasiya Slyepchenko et al. 2017, Wittchen et al. 2002]. Estas mesmas mulheres podem vivenciar mais sintomas severos de TAB durante o período pré-natal e enquanto fazem o uso de pílulas contraceptivas [Anastasiya Slyepchenko et al. 2017].

Neste trabalho é utilizada uma amostra de 1.099 mulheres com TAB que forneceram informações detalhadas sobre eventos de vida reprodutiva como parte do estudo STEP-BD (*Systematic Treatment Enhancement Program for Bipolar Disorder*) [Sachs S et al. 2003]. O objetivo principal é achar padrões e associações entre as características da doença, a prevalência de transtornos psiquiátricos comórbidos e sintomas da vida reprodutiva, entre mulheres com diagnóstico de PMDD, através de regras de associação.

De acordo com esse objetivo, foi escolhido a utilização do Algoritmo *Apriori*, uma vez que, ele faz uso de regras de associação para encontrar padrões dentro da base de dados como um todo. Além disso, o paradigma não supervisionado permite uma análise exploratória dos dados, de forma que as regras de associação trazem uma fácil interpretação das suas resoluções, consequentemente facilitando a análise medica dos resultados.

Este artigo está organizado da seguinte forma: a Seção 2 apresenta conceitos importantes relacionados a Regras de Associação e o algoritmo *Apriori*; a Seção 3 aborda a metodologia empregada para a descoberta de Regras de Associação; a Seção 4 mostra os resultados obtidos neste trabalho; a Seção 5 é utilizada para mostrar a interpretação das regras extraídas; finalmente a Seção 6 mostra a conclusão relativa aos resultados atingidos.

2. Fundamentação Teórica

2.1. Descoberta de Regras de Associação

Regras de associação são padrões descritivos que representam a probabilidade de que um conjunto de itens apareça em uma transação visto que outro conjunto está presente [Brusso 2000].
Mais formalmente, temos o seguinte:

Seja $A = \{a_1, ..., a_m\}$ o conjunto de m itens.

Os itens podem ser, por exemplo, sintomas do TAB, ansiedade, deficit de atenção, tensão pré-menstrual, etc.

Dizemos assim que qualquer subconjunto $I \subseteq A$ é denominado como **conjunto de itens** (*itemset*).

Por exemplo:

$$\text{I} = \{\textit{ansiedade, depressão}\}$$

Regras de associação assumem a forma de:

Se antecedente então consequente

onde:

antecedente e consequente são *itemsets*

Uma **transação** é uma relação entre *itemsets*. Cada transação é definida como:

$$\text{t} = \{\text{A} \rightarrow \text{B}\}$$

onde A e B são um ou mais itens que pertencem ao conjunto I e $A \rightarrow B$ é uma regra de associação.

Por exemplo, a transação abaixo define que quem tem ansiedade também possui depressão:

$$\text{t} = \{\textit{ansiedade} \rightarrow \textit{depressão}\ \}$$

Seja $T = \{t_1, ..., t_n\}$ um conjunto de n transações que é denotado por **banco de dados de transações**.

A descoberta de regras de associação pode ser dividida em duas partes:

1. Encontrar os **conjuntos de itens frequentes** que são aqueles conjuntos de itens que possuem **suporte** igual ou superior ao mínimo estabelecido. **Suporte** é a frequência dos padrões de transações que estão ocorrendo no banco de dados de transações.
2. Gerar as regras de associação baseadas no conjunto de itens frequentes. Deve-se selecionar as regras que estabelecem um grau de **confiança** mínimo. **Confiança** é a probabilidade de B ser verdade quando já sabemos que A é verdade.

Desta forma, o problema de mineração de dados em relação às regras de associação consiste em produzir regras que possuem suporte e confiança mínimos.

O suporte de uma regra $A \rightarrow B$, é explicitado pela Equação 1:

$$\textit{Suporte}(A \rightarrow B) = \frac{\textit{Ocorrências de A e B}}{\textit{Total de transações em T}} \tag{1}$$

A confiança é dada pela Equação 2:

$$\textit{Confiança}(A \rightarrow B) = \frac{\textit{Ocorrências de A e B}}{\textit{Ocorrências de A}} \tag{2}$$

Logo abaixo apresenta-se um exemplo concreto para a criação de regras de associação. A Tabela 1, mostra um exemplo de um banco de dados de transações pertencente a um Hospital Psiquiátrico

As colunas são itens e as linhas são transações. A falta do item é representada por *0* e caso contrário por *1*. Supondo uma confiança mínima pré-estabelecida de *90*% e o suporte mínimo de *55*%.

Tabela 1. Banco de dados de transações de um Hospital Psiquiátrico

Transação	depressão	ansiedade	pmdd
1	1	1	1
2	0	1	0
3	1	1	1
4	0	0	1
5	1	1	1

A partir das transações da Tabela 1, é possível obter a seguinte regra:

$$\textit{depressão, ansiedade} \rightarrow \textit{pmdd}$$

Ou seja, se um paciente dispõe de depressão e ansiedade ele também expressa pmdd.

Calculando o suporte e confiança desta regra temos:

$$\textit{Suporte} = \frac{3}{5} = 0,6 = 60\% \tag{3}$$

$$\textit{Confiança} = \frac{3}{3} = 1 = 100\% \tag{4}$$

2.2. Algoritmo Apriori

Apriori [R. Agrawal and R. Srikant 1994] é um algoritmo para mineração de dados frequentes e o aprendizado de regras de associação em banco de dados transacionais.

A ideia principal do algoritmo *Apriori* é baseada na observação de que se conjuntos de itens (*itemset*) são frequentes, as combinações destes pode gerar novos conjuntos que também sejam frequentes [Bolina 2013].
O algoritmo decompõe-se basicamente em duas partes principais:

1. Concepção do conjunto de itens frequentes: conjuntos que apresentam suporte maior ou igual ao mínimo estabelecido.
2. Concepção de regras de associação confiáveis: regras de associação que obedecem a um grau de confiança mínimo pré-estabelecido.

De forma geral, a sua meta é procurar por relações entre os dados enquanto eles são separados [Vasconcelos and Carvalho 2004].

2.2.1. Concepção do conjunto de itens frequentes

O algoritmo efetua diversas etapas sobre o conjunto de itens totais:

1. Calcula o suporte de cada item usando a Equação 1 e gera *1-itemset*. A partir deste ele seleciona aqueles subconjuntos com suporte maior que o mínimo e cria o conjunto de *1-itemsetFrequente*.

2. Realiza a junção dos conjuntos de *1-itemsetFrequente* para a criação dos conjuntos *2-itemset*. O suporte para esses novos grupos é calculado com uma nova busca nos dados. Chegando assim aos *2-itemsetsFrequentes*.
3. Prossegue com os passos anteriores até onde o conjunto de *k-itens* achado seja diferente de vazio.

Abaixo apresentaremos um exemplo do funcionamento do algoritmo baseado no banco de dados de transações da Tabela 2. Neste exemplo há 5 itens ocorrendo em 5 transações, onde 1 representa a presença do item e 0, a sua ausência. Neste exemplo vamos adotar suporte mínimo = 20%.

Tabela 2. Banco de dados de transações qualquer

Transação	Item1	Item2	Item3	Item4	Item5
1	1	0	0	1	0
2	0	1	0	1	1
3	0	0	1	1	1
4	1	0	1	0	0
5	1	1	0	0	0

Na Tabela 3, temos o conjunto de *1-itemset*, ou seja, os conjuntos que possuem apenas 1 elemento. A Tabela 4 representa os *1-itemsetFrequente*, ou seja, os *1-itemset* com suporte maior que 20%.

Tabela 3. 1-itemset

Item	Suporte
1	0.6
2	0.4
3	0.4
4	0.6
5	0.4

Tabela 4. 1-ItemsetFrequente

Item	Suporte
1	0.6
2	0.4
3	0.4
4	0.6
5	0.4

O conjunto *2-itemsets* (Tabela5) é feito a partir da combinação de elementos de *1-ItemsetFrequente* representados pela Tabela 4.

Tabela 5. 2-itemsets

Item	Suporte
$1,2$	0.2
$1,3$	0.2
$1,4$	0.2
$1,5$	0.0
$2,3$	0.0
$2,4$	0.2
$2,5$	0.2
$3,4$	0.2
$3,5$	0.2
$4,5$	0.4

Tabela 6. 2-ItemsetFrequentes

Item	Suporte
$1,2$	0.2
$1,3$	0.2
$1,4$	0.2
$2,4$	0.2
$2,5$	0.2
$3,4$	0.2
$3,5$	0.2
$4,5$	0.4

O conjunto *3-itemsets* (Tabela 7) é gerado a partir de *2-ItemsetFrequentes* (Tabela 6). Esses novos itens serão formados a partir de dois subconjuntos de 2 elementos.

O conjunto *4-itemsets* é vazio. O que finaliza a execução dessa parte do algoritmo.

Tabela 7. 3-itemset

Item	Suporte
$1, 2, 3$	0.0
$1, 2, 4$	0.0
$1, 3, 4$	0.0
$2, 4, 5$	0.2
$3, 4, 5$	0.2

Tabela 8. 3-ItemsetFrequentes

Item	Suporte
$2, 4, 5$	0.2
$3, 4, 5$	0.2

2.2.2. Concepção de regras de associação confiáveis

Depois de descoberto os conjuntos de itens frequentes é possível originar regras de associação que tem um grau de confiança maior ou igual ao mínimo estabelecido. Estas regras são chamadas de **regras confiáveis**.

A partir do exemplo apresentado anteriormente é plausível a obtenção de regras de associação com base no conjunto de itens frequentes expostos. Para referenciar estes conjuntos usaremos a expressão F_k. Na qual *k* representa o tamanho dos itens frequentes (por exemplo: F_1 é o conjunto de itens frequentes de tamanho 1). De modo formal representamos $F_k = \{I_1, I_2, I_3, ..., I_k\}$. No qual I_k é um subconjunto de itens. Através de F_k com $k \geq 2$ é possível produzir todas as regras (com no máximo *k* regras) que possuem itens de I_k.

Considerando uma confiança mínima de 50%, as regras confiáveis do exemplo dado são apresentadas na Tabela 9.

Tabela 9. Regras confiáveis geradas pelo *Apriori*

Item	Regra	Confiança
$2, 4$	$2 \rightarrow 4$	50%
$2, 5$	$2 \rightarrow 5$	50%
$3, 4$	$3 \rightarrow 4$	50%
$3, 5$	$3 \rightarrow 5$	50%
$4, 5$	$4 \rightarrow 5$	66.6%
$2, 4, 5$	$2 \rightarrow 4, 5$	50%
	$5 \rightarrow 2, 4$	50%
	$2, 4 \rightarrow 5$	100%
	$2, 5 \rightarrow 4$	100%
	$4, 5 \rightarrow 2$	50%
$3, 4, 5$	$3 \rightarrow 4, 5$	50%
	$5 \rightarrow 4, 3$	50%
	$3, 4 \rightarrow 5$	100%
	$3, 5 \rightarrow 4$	100%
	$4, 5 \rightarrow 3$	50%

Tabela 10. Regras não confiáveis

Item	Regra	Confiança
$1, 2$	$1 \rightarrow 2$	$33,3\%$
$1, 3$	$1 \rightarrow 3$	$33,3\%$
$1, 4$	$1 \rightarrow 4$	$33,3\%$
$2, 4, 5$	$4 \rightarrow 2, 5$	$33,3\%$
$3, 4, 5$	$4 \rightarrow 3, 5$	$33,3\%$

Na Tabela 10 são apresentadas as regras geradas pelo *Apriori* mas cujo grau de confiança é inferior ao mínimo estabelecido. Estas regras são denominadas como **regras não confiáveis**.

3. Metodologia

Foram utilizados dados de uma coorte de 1.099 mulheres com TAB provenientes do *Systematic Treatment Enhancement Program for Bipolar Disorder* (STEP-BD).

O STEP-BD é o maior estudo de TAB financiado pelo governo federal dos Estados Unidos da América, caracterizado como um estudo ambulatorial de longo prazo. No qual foram matriculados 4.360 participantes de 22 regiões diferentes do país, ao longo de sete anos (1998-2005). Ele foi projetado com o intuito de descobrir quais tratamentos, ou combinações de tratamentos, são mais eficazes para tratar episódios de depressão e mania, e assim, prevenir episódios recorrentes em pessoas com transtorno bipolar. Logo, é um amplo programa de pesquisa que inclui vários estudos diferentes, de forma que cada um visa em um aspecto diferente do tratamento para a doença [NIH , Sachs S et al. 2003].

A base de dados utilizada é composta por 94 variáveis que abrangem dados sociodemográficos, histórico do TAB, comorbidades psiquiátricas e sintomas menstruais das pacientes.

Todas as variáveis da base foram utilizadas para a geração das regras de associação pelo algoritmo *Apriori*. Os parâmetros utilizados foram: confiança mínima 65%, suporte mínimo 10%, tamanho da regra mínimo 2 e máximo 3.

Os parâmetros estabelecidos de confiança e suporte mínimos se justificam, pois com um suporte baixo é possível conseguir estabelecer eventos raros, ou seja, eventos de pouca recorrência. Já pelo aspecto da confiança mínima estabelecida, é factível ter regras que são realmente confiáveis, uma vez que, esse índice de 65%, permite mostrar a confiabilidade desses eventos.

Para fins deste estudo foram selecionadas apenas regras onde ocorrem a variável PMDD com valor positivo, e com as seguintes características:

- PMDD como antecedente e alguma característica positiva no consequente da regra
- alguma característica positiva como antecedente e PMDD no consequente da regra
- PMDD associado a alguma característica como antecedente e alguma comorbidade no consequente da regra

O motivo desta seleção é entender como a ocorrência de PMDD está associada com outros fatores como, sintomas do TAB, outros transtornos comórbidos, eventos reprodutivos e/ou dados sociodemográficos das pacientes.

A linguagem R foi utilizada para executar o algoritmo (*Apriori*) com os pacotes *arules, arulesViz, devtools, Rsenal*, que podem ser baixados em `https://cran.r-project.org/web/packages/available_packages_by_name.html`. O pacote *Rsenal* em especial pode ser obtido diretamente no site `https://github.com/brooksandrew/Rsenal`.

Os experimentos foram executados num computador com processador $Intel^{®}$ $Core^{TM}$ $i7-3537U$ CPU @ $2.00GHz$ x 4 , com 8GB de memória e sistema operacional Windows 10.

4. Resultados

No total, o *Apriori* encontrou 621.062 regras de associação. Foram então selecionadas para análise as regras segundo os critérios apresentados na Seção 3. As Tabelas 12 a 19 apresentam as regras geradas agrupadas, por tipo de característica.

Na Tabela 11 é apresentado o significado das variáveis que aparecem nas regras.

Tabela 11. Significado das varáveis

Variável	Descrição
PMDD_DSM5	descrição de PMDD de acordo com o DSM5
RACE	Etnia
MCF5Y.CYCLE10	Duração do ciclo em torno de 10 dias durante os primeiros 5 anos menstruais
MCF5Y.MOODEPRS	mudança de humor ou depressão nos primeiros 5 anos do ciclo menstrual
MCF5Y.TENS_IRRIT	tensão ou irritabilidade nos primeiros 5 anos do ciclo menstrual
MCF5Y.TENSION_IRRIT.DEP.MSW	irritabilidade, depressão e mudança de humor nos primeiros 5 anos menstruais
MP_IRRIT_DEP_MSWING	irritação, depressão e mudança de humor no período menstrual
MP_PHYSICAL	cansaço físico durante o período menstrual
MP_IRRITANGR	irritação no período menstrual
MP_BLOATING	inchaço durante o período menstrual
MP_FOODCRAV	inquietação em relação à alimentação no período menstrual
MP_LENGTH10	duração do ciclo em torno de 10 dias
MP_CRYEASLY	choro fácil durante o período menstrual
MP_ABDOMPAN	dores abdominais no período menstrual
MP_BRESTEND	mastalgia (incômodos mamários) no período menstrual
MP_OVERWHEL	sentimento de sobrecarga emocional no período menstrual
MP_ANX	ansiedade no período menstrual
MP_DEPRSION	depressão no período menstrual
MP_LACKENGY	perda de energia no período menstrual
MP_FINAL.SXWORK	interferência no trabalho ou escola no período menstrual
MP_MOODSWING	mudança de humor no período menstrual
MP_DEP_MSWING_PME	depressão e mudança de humor com piora pré-menstrual
PHOBIAP	Fobia passada
PTSDP	Transtorno de stress pós-traumático passado
ALCDEPP	Dependência de álcool passado
ALCABUSP	Abuso de álcool passado
SXPREG	Sintomas severos de humor durante o período de gravidez
SXPPART	Sintomas severos de humor durante o pós-parto
SXOCP	Sintomas severos de humor enquanto faz o uso de pílula anticoncepcional
RAPIDCYCL	Ciclagem rápida
ANTIMAN	Uso de antidepressivo no estado de mania
BDI_BDII	Tipo do TAB
POLARITY1ST	Polaridade do primeiro episódio do TAB
ANYANXPA	Alguma desordem de ansiedade passada
ANYANXCU	Alguma desordem de ansiedade corrente

Tabela 12. PMDD relacionado a dados sociodemográficos

Id	Regra	Confiança	Suporte
1	$PMDD_DSM5 = yes \rightarrow RACE = WhiteorCaucasian$	90.0%	41.0%

Tabela 13. PMDD relacionado diretamente a sintomas dos 5 primeiros anos do ciclo menstrual

Id	Regra	Confiança	Suporte
1	$PMDD_DSM5 = yes \rightarrow MCF5Y.CYCLE10 = Yes$	71.0%	32.0%
2	$PMDD_DSM5 = yes \rightarrow MCF5Y.MOODEPRS = Yes$	79.5%	36.0%
3	$PMDD_DSM5 = yes \rightarrow MCF5Y.TENS_IRRIT = Yes$	80.5%	36.5%
4	$PMDD_DSM5 = yes \rightarrow MCF5Y.TENSION_IRRIT.DEP.MSW = Yes$	84.0%	38.5%

5. Discussão

As regras encontradas pelo algoritmo *Apriori* não mostram nenhuma relação direta entre PMDD e transtornos comórbidos específicos.

Tabela 14. PMDD relacionado diretamente a sintomas do período menstrual

Id	Regra	Confiança	Suporte
1	$PMDD_DSM5 = yes \rightarrow MP_IRRIT_DEP_MSWING = Yes$	100%	45.5%
2	$PMDD_DSM5 = yes \rightarrow MP_PHYSICAL = Yes$	97.0%	44.0%
3	$PMDD_DSM5 = yes \rightarrow MP_IRRITANGR = Yes$	95.0%	43.0%
4	$PMDD_DSM5 = yes \rightarrow MP_BLOATING = Yes$	84.5%	38.5%
5	$PMDD_DSM5 = yes \rightarrow MP_FOODCRAV = Yes$	81.0%	37.0%
6	$PMDD_DSM5 = yes \rightarrow MP_LENGTH10 = Yes$	79.0%	36.0%
7	$PMDD_DSM5 = yes \rightarrow MP_CRYEASLY = Yes$	79.0%	35.5%
8	$PMDD_DSM5 = yes \rightarrow MP_ABDOMPAN = Yes$	74.0%	33.5%
9	$PMDD_DSM5 = yes \rightarrow MP_BRESTEND = Yes$	72.0%	33.0%

Tabela 15. PMDD e sintomas do período menstrual associados mutuamente

Id	Regra	Confiança	Suporte
1	$PMDD_DSM5 = yes \rightarrow MP_OVERWHEL = Yes$	75.0%	34.0%
2	$MP_OVERWHEL = Yes \rightarrow PMDD_DSM5 = yes$	82.5%	34.0%
3	$PMDD_DSM5 = yes \rightarrow MP_ANX = Yes$	82.0%	37.5%
4	$MP_ANX = Yes \rightarrow PMDD_DSM5 = yes$	78.0%	37.5%
5	$PMDD_DSM5 = yes \rightarrow MP_DEPRSION = Yes$	85.0%	39.0%
6	$MP_DEPRSION = Yes \rightarrow PMDD_DSM5 = yes$	79.0%	39.0%
7	$PMDD_DSM5 = yes \rightarrow MP_LACKENGY = Yes$	79.5%	36.0%
8	$MP_LACKENGY = Yes \rightarrow PMDD_DSM5 = yes$	71.5%	36.0%
9	$PMDD_DSM5 = yes \rightarrow MP_FINAL.SXWORK = Yes$	100%	45.5%
10	$MP_FINAL.SXWORK = Yes \rightarrow PMDD_DSM5 = yes$	81.0%	45.5%
11	$PMDD_DSM5 = yes \rightarrow MP_MOODSWNG = Yes$	94.0%	42.5%
12	$MP_MOODSWNG = Yes \rightarrow PMDD_DSM5 = yes$	73.5%	42.5%
13	$PMDD_DSM5 = yes \rightarrow MP_DEP_MSWING_PME = Yes$	98.0%	44.5%
14	$MP_DEP_MSWING_PME = Yes \rightarrow PMDD_DSM5 = yes$	69.5%	44.5%

Tabela 16. Comorbidades

Id	Regra	Confiança	Suporte
1	$PHOBIAP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXCU = Yes$	89.0%	10.5%
2	$PHOBIAP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	100%	11.5%
3	$PTSDP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	100%	13.5%

Tabela 17. Dependência ou abuso de álcool

Id	Regra	Confiança	Suporte
1	$ALCDEPP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	76.0%	10.0%
2	$ALCABUSP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	76.0%	12.5%

Tabela 18. Sintomas relacionados à vida reprodutiva

Id	Regra	Confiança	Suporte
1	$SXPREG = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	71.5%	12.0%
2	$SXPPART = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	72.0%	11.5%
3	$SXOCP = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	71.0%	13.5%

Por outro lado, foram encontradas regras que mostram as seguintes relações diretas com PMDD:

- Variável relacionada a dados sociodemográficos (Tabela 12): raça

Tabela 19. Sintomas do TAB relacionados diretamente ao PMDD

Id	Regra	Confiança	Suporte
1	$RAPIDCYCL = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	75.5%	13.5%
2	$ANTIMAN = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	70.5%	13.0%
3	$MCF5Y.MOODEPRS = Yes, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	69.0%	25.0%
4	$BDI_BDII = BPI, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	71.5%	19.5%
5	$POLARITY1ST = DepressiveOnset, PMDD_DSM5 = yes \rightarrow ANYANXPA = Yes$	70.0%	23.1%

- Variáveis relacionadas ao período menstrual (Tabela 14): irritação, mudanças de humor, depressão, cansaço físico, inchaço, inquietação em relação à alimentação, duração do ciclo em torno de 10 dias, choro fácil, dores abdominais, mastalgia (incômodos mamários).

A regra de associação apresentada na Tabela 12 descreve que as pacientes com PMDD são pessoas brancas. Esta regra aparece em aproximadamente 41% da base e possui confiança de 90%.

A Tabela 13 mostra que a presença de PMDD implica na ocorrência de sintomas relacionados ao período menstrual nos primeiros ciclos de vida reprodutiva, com uma confiança média de 78.8%. Estas características em conjunto com as regras das Tabelas 14 e 15, indicam que estes mesmos sintomas tendem a aparecer durante todos o ciclos menstruais da vida de mulheres com TAB.

Algumas variáveis relacionadas ao período menstrual ocorrem mutuamente com PMDD (Tabela 15). Estas relações mútuas são as seguintes:

- Sentimento de sobrecarga emocional (34.0% de ocorrências na base de dados).
- Ansiedade generalizada (37.5% de ocorrências na base de dados).
- Depressão (38.7% de ocorrências na base de dados).
- Perda de energia (36.2% de ocorrências na base de dados).
- Interferência do rendimento no trabalho ou na escola (45.5% de ocorrências na base de dados).
- Mudanças de humor (42.7% de ocorrências na base de dados).
- Depressão e mudança de humor com piora pré-mestrual (44.5% de ocorrências na base de dados).

O PMDD associado a outras variáveis pode estar relacionado a outros transtornos. Por exemplo,

- Variáveis relacionadas a comorbidades (Tabela 16): fobia passada, estresse pós traumático passado.
- Variáveis relacionadas ao uso do álcool (Tabela 17): dependência de álcool passada, abuso de álcool passado.
- Variáveis relacionadas a sintomas severos de mudança de humor na vida reprodutiva (Tabela 18): gravidez, pós-parto, uso de contraceptivo oral.
- Variáveis relacionadas a sintomas do TAB relativos ao PMDD (Tabela 19): ciclagem rápida, uso de antidepressivo no estado de mania, mudança de humor, TAB tipo I, início depressivo.

As tabelas 16 a 19 possuem em comum o fato de que suas regras implicam em alguma desordem de ansiedade. Este fator é apresentado quando há eventos de:

- fobia passada (11.5% de ocorrências na base de dados)
- transtorno de stress pós-traumático (13.5% de ocorrências na base de dados)
- dependência de álcool passada (10.0% de ocorrências na base de dados)
- abuso de álcool passado (12.5% de ocorrências na base de dados)
- sintomas severos de humor na gravidez (12.0% de ocorrências na base de dados)
- sintomas severos de humor no pós-parto (11.5% de ocorrências na base de dados)
- sintomas severos de humor enquanto faz uso de pílula anticoncepcional (13.5% de ocorrências na base de dados)
- ciclagem rápida (13.5% de ocorrências na base de dados)
- uso de antidepressivo no estado de mania (13.0% de ocorrências na base de dados)
- mudança de humor ou depressão nos primeiros 5 anos menstruais (25.0% de ocorrências na base de dados)
- transtorno afetivo bipolar do tipo 1 (19.5% de ocorrências na base de dados)
- polaridade do primeiro episódio do TAB (23.1% de ocorrências na base de dados)

Estas associações anteriores podem não trazer informação relevante uma vez que as variáveis *ANYANXPA* e *ANYANXCU* não apontam qual a desordem de ansiedade específica.

6. Conclusão

O Transtorno Afetivo Bipolar (TAB) é uma desordem do humor na qual o paciente apresenta episódios depressivos e de mania, e apresenta-se como uma doença de caráter evolutivo crônico e incapacitante. Mulheres com distúrbio disfórico pré-menstrual (*premenstrual disforic disorder* - PMDD) podem ter os sintomas de TAB agravados durante eventos de sua vida reprodutiva.

Neste estudo, foram utilizados dados do *Systematic Treatment Enhancement Program for Bipolar Disorder* (STEP-BD) e o algoritmo *Apriori* para a descoberta de associações entre o PMDD, sintomas reprodutivos, transtornos comórbidos e informações sócio-demográficas de mulheres com TAB.

Das milhares de regras geradas pelo *Apriori*, foram encontradas associações diretas do PMDD com sintomas do período menstrual e sintomas do ciclo menstrual da mulher. No entanto, o algoritmo não encontrou associação entre o PMDD e outras comorbidades específicas (presença ou associação de duas ou mais doenças na mesma paciente).

A presença do PMDD nos diagnósticos de TAB é considerado um marcador de gravidade para a doença [American Psychiatric Association 2013]. Devido a este motivo é de grande importância encontrar as associações de PMDD com outras características em pacientes com TAB.

As regras encontradas a partir de dados reais de pacientes podem mostrar associações importantes que auxiliem os médicos para um estudo mais aprofundando destas relações. Além disso, por outro lado, o possível não aparecimento de associações esperadas pelos médicos a partir destes dados pode trazer também reflexões e apontar novos estudos no tema.

Como trabalho futuro, é possível continuar a exploração da riqueza de informações que existe na base de dados utilizada e o potencial de outros algoritmos de aprendizado de máquina, em especial do paradigma supervisionado, usando a variável PMDD como classe, obtendo assim uma nova abordagem para a análise dos dados.

Referências

National Insititute of Health.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. Amenrican Psychiatric Publishing.

Anastasiya Slyepchenko, Benicio N. Frey, Beny Lafer, Andrew A. Nierenberg, Gary S. Sachs, and Rodrigo S. Dias (2017). Increased illness burden in women with bipolar and premenstrual dysphoric disorder: Data from 1,099 women. *Acta Psychiatrica Scandinavica (under revision)*.

Bergink, V., Rasgon, N., and Wisner, K. L. (2016). Postpartum Psychosis:Madness, Mania, and Melancholia in Motherhood. *The American Journal of Psychiatry*, 173(12):1179–1188.

Bolina, A. C. (2013). Estudo de Algoritmos Paralelos de Regras de Associação usando o framework Mapreduce. Trabalho de Conclusão de Curso do curso de Ciência da Computação da Universidade Federal de Lavras.

Brusso, M. J. (2000). Access Miner: uma proposta para a extração de regras de associação aplicada à mineração do uso da Web. Master's thesis, Universidade Federal do Rio Grande do Sul.

Cirillo C, P., Passos F Benitez, R., Bevilaqua N, M. C., López A Rodriguez, J. R., and Nardi E, A. (2012). Bipolar disorder and Premenstrual Syndrome or Premenstrual Dysphoric Disorder comorbidity: a systematic review. *Revista Brasileira de Psiquiatria*, 34(4):467–479. Enter text here.

Geddes R, J. and Miklowitz J, D. Treatment of bipolar disorder. *The lancet*, 381(9878):1672–1682.

Hu, L.-Y., Shen, C.-C., Hung, J.-H., Chen, P.-M., Wen, C.-H., Chiang, Y.-Y., and Ti-Lu (2016). Risk of Psychiatric Disorders Following Symptomatic Menopausal Transition: A Nationwide Population-Based Retrospective Cohort Study. *Medicine*, 95(6).

Judd L, L., Akiskal S, H., Schettler J, P., Endicott, J., Leon C, A., Solomon A, D., Coryell, W., Maser D, J., and Keller B, M. (2005). Psychosocial Disability in the Course of Bipolar i and ii Disorders: a prospective, Comparative, Longitidinal Study. *Archives of general psychiatry*, 62(12):1322–1330.

R. Agrawal and R. Srikant (1994). Fast Algorithms for mining association rules in large databases. *International Conference on Very Large Data Bases*, 1:487–499.

Sachs S, G., Thase E, M., Otto W, M., Bauer, M., Miklowitz, D., Wisniewski R, S., Lavori, P., Lebowitz, B., Rudorfer, M., Frank, E., Nierenber A, A., Fava, M., Bowden, C., Ketter, T., Marangell, L., Calabrese, J., Kupfer, D., Rosenbaum F, J., and Michael W Otto (2003). Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). *Biological Psychiatry*, 53(11):1028–1042.

Vasconcelos, L. and Carvalho, C. (2004). Aplicação de Regras de Associação para Mineração de Dados na Web. *Instituto de Informática- UFG*.

Wittchen, H.-U., Becker, E., Lieb, R., and Krause, P. (2002). Prevalence, incidence and stability of premenstrual dysphoric disorder in the community. *Psychological Medicine*, 32(01):119–32.

World Health Organization. *International Statistical Classification of Diseases and Related Health Problems*.

Otimização de Interligação de Religadores por Fibra Óptica para Automação de Redes Auto-recuperáveis de Energia

Marcos N. L. de Carvalho[1], **Maykon R. P. da Silva**[1], **Flávio H. T. Vieira**[1]

[1]Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.605-010 – Goiânia – GO – Brasil

{marcosnlc4,maykoon.renan,flaviohtv}@gmail.com

Resumo. *Neste artigo, propõe-se a utilização de um algoritmo genético multiobjetivo NSGA-II (Algoritmo Genético de Classificação Não-Dominada) para dimensionar a rede óptica em malha (Mesh) que interliga os religadores de uma rede de energia inteligente (Smart Grid) que possui a capacidade de se auto-recuperar. Considerando somente distâncias (custo) de enlace, este dimensionamento pode ser efetuado considerando a solução para o Problema do Caixeiro Viajante. Assim, primeiramente apresenta-se neste trabalho uma comparação de desempenho dos seguintes algoritmos utilizados para solucionar o Problema do Caixeiro Viajante (Travelling Salesman Problem - TSP): Força Bruta, Entropia Cruzada, Algoritmo Genético e Algoritmo Genético de Classificação Não-Dominada (NSGA-II). Em seguida, propõe-se a utilização do NSGA-II para dimensionar a rede óptica em malha considerando fatores como custo e vazão. Ou seja, emprega-se o algoritmo NSGA-II para resolver o problema multiobjetivo de maximizar a vazão da rede e minimizar o custo de implantação da rede de fibra.*

Abstract. *In this article, it is proposed the use of a multiobjective genetic algorithm NSGA-II (Non-dominated Sorting Genetic Algorithm) for dimensioning the mesh optical network that connects the reclosers of a smart grid which has the self-healing capacity. Considering only link distances, this dimensioning can be done considering the solution to the Traveling Salesman Problem. Thus, we first present in this work a performance comparison of the following algorithms used to solve the Problem of the Traveling Salesman (TSP): Brute Force, Cross Entropy, Genetic Algorithm, and Nondominated Sorting Genetic Algorithm (NSGA-II) to solve the multiobjective problem of maximizing network throughput and minimizing the cost of deploying the fiber network.*

1. Introdução

O principal objetivo deste trabalho é de propor a aplicação de algoritmos genéticos para o dimensionamento de rede óptica em malha para interligação de religadores em uma rede auto-recuperável, comparando os desempenhos em relação a outros algoritmos da literatura. Para tal, inicialmente procurou-se solucionar o Problema do Caixeiro Viajante (*Travelling Salesman Problem* (*TSP*)), com base no estudo de caso do projeto P&D 368 CELG-D-ANEEL-UFG [de Carvalho vaz et al. 2016]. Esse projeto consiste na implantação de uma Rede Óptica em Malha (*Mesh*) na cidade de Goiânia - GO para automatizar os

religadores de uma rede elétrica inteligente (*Smart Grid*). Essa rede possui a característica de se auto-recuperar (*self-healing*) de eventuais adversidades que possam interromper a distribuição de energia elétrica.

Os seguintes algoritmos foram considerados para a solução do *TSP*: Força Bruta, Entropia Cruzada, Algoritmo Genético e Algoritmo Genético de Classificação Não-Dominada (*Nondominated Sorting Genetic Algorithm* (*NSGA-II*)). Esse último algoritmo possui a vantagem de ser multiobjetivo, característica esta que será discutida mais à frente.

A Figura 1 apresenta o diagrama físico da Rede Óptica em Malha do projeto de P&D 368 a partir de uma imagem via satélite (obtida pelo *Google Earth*) da região de Goiânia - GO, onde os religadores deverão ser instalados [de Carvalho vaz et al. 2016]. Em busca de uma confiabilidade do canal de comunicação próxima a 100% e alta taxa de vazão que possibilite comando em tempo real, empregou-se nesse projeto fibra óptica do tipo monomodo. Além disso, o protocolo utilizado nos nós da rede foi o *AODV* (*Adhoc Ondemand Distance Vector Routing*) [da Cruz 2010]. Esse protocolo é baseado no algoritmo vetor distância, onde o mesmo calcula a menor rota possível baseado no custo de cada enlace.

Figura 1. Diagrama físico da Rede Óptica em Malha do estudo de caso.

As próximas seções apresentam os resultados das simulações dos algoritmos levando em consideração o tempo de execução em segundos. Em um primeiro momento realizou-se as simulações com 4 nós e depois com 12 nós (os mesmos nós apresentados na Figura 1), com o objetivo de se analisar o comportamento dos algoritmos ao se modificar o número de nós.

2. Dimensionamento de Rede Óptica em Malha Através da Solução do *TSP*

A história do Problema Caixeiro Viajante (*TSP*) remete a um vendedor, situado numa determinada cidade, que procura obter o menor caminho possível que o permita visitar somente uma vez as demais cidades de um mapa, retornando à cidade de partida. Numa análise matemática do problema, o *TSP* é conhecido como um clássico problema de otimização combinatória que consiste em partir de um vértice qualquer de um grafo, percorrer todos os outros vértices sem repeti-los, utilizando o menor caminho possível, e retornar para o vértice inicial [Potvin 1996] [Maredia 2010]. Segundo [Potvin 1996], é possível formular um algoritmo de programação linear para este problema, de forma que a solução

ótima será dada pelo caminho de tamanho mínimo, onde as variáveis de decisão, a função objetivo e as restrições são dadas, respectivamente, por:

Variáveis de decisão:

$$x_{ij} = \begin{cases} 1, & \text{se a aresta } i \to j \text{ existe} \\ 0, & \text{caso contrário} \end{cases} \tag{1}$$

Minimizar:

$$\sum_i \sum_j d_{ij} x_{ij} \tag{2}$$

Sujeito a:

$$\sum_j x_{ij} = 1, \quad i = 1, ..., N, \tag{3}$$

$$\sum_i x_{ij} = 1, \quad j = 1, ..., N, \tag{4}$$

$$(x_{ij}) \in X,$$

onde N é o número de vértices, d_{ij} é a distância entre os vértices i e j e x_{ij} são as variáveis de decisão. A expressão $(x_{ij}) \in X$ denota o conjunto de restrições de sub-rotas de parada que restringem as soluções viáveis para aquelas que consistem num único caminho.

O projeto de uma rede de fibra óptica em malha (*Mesh*) capaz de fornecer dados para acionar religadores de energia pode ser relacionado com um típico problema do *TSP*. O primeiro passo para se construir tal rede consiste em interconectar todos os nós (religadores) da rede. Na maioria dos casos, a aplicação do *TSP* tende a formar uma configuração em anel. Essa topologia é realizada conectando todos os nós adjacentes da rede até que todos estejam conectados, garantindo que o menor caminho para se percorrer todos os nós (sem passar mais de uma vez por um mesmo nó) da rede seja obtido. Uma vez que os nós da rede estejam interconectados, basta adicionar as redundâncias (rotas adicionais) para se obter uma rede óptica em malha que satisfaça os critérios do projeto. Esses critérios exigem que cada nó deve estar conectado a no mínimo um nó (uma redundância) e no máximo a quatro nós (quatro redundâncias). Apesar de demandar mais fibra óptica (aumento do custo), as redundâncias permitem obter uma configuração em malha para a rede óptica, possibilitando a formação de subanéis que proporcionam maior confiabilidade ao sistema, garantindo uma maior qualidade de serviço para a rede [de Carvalho vaz et al. 2016].

As informações sobre os algoritmos empregados no estudo, bem como os resultados das simulações aplicadas no estudo de caso são apresentados nas seções a seguir.

3. Solução do *TSP* com Força Bruta

A solução do *TSP* por Força Bruta consiste basicamente em realizar todos os caminhos possíveis na malha, armazenando todas as distâncias obtidas. Com todas as distâncias de todos os caminhos obtidas, basta escolher o caminho onde se obteve a menor distância. Apesar de garantir que sempre a melhor solução será obtida, esse algoritmo é extremamente ineficiente para malhas com muitos nós. Sua complexidade computacional é de

$O((n-1)!)$, onde n é o número de nós da malha. Isso comprova o grande aumento de rotas ao se incrementar o número de nós [Sahalot and Shrimali 2014].

No presente estudo, utilizou-se um programa implementado em MATLAB capaz de solucionar o *TSP* com Força Bruta [Burkardt 2013]. Este código é executado ao se inserir um arquivo de texto contendo a tabela de distâncias dos nós da malha. Assim que a tabela é inserida, o algoritmo de Força Bruta é executado, retornando as menores distâncias, médias e máximas, bem como o número de caminhos verificados e o tempo de execução.

3.1. Resultados da Simulação do *TSP* com Força Bruta

Por meio dos resultados obtidos nas simulações com o algoritmo de Força Bruta apresentados na Figura 2, fica evidente a ineficiência do algoritmo por Força Bruta com a elevação do número de nós da malha, onde o tempo de execução para 12 nós ficou em torno de 8 horas. Apesar dessa ineficiência, a solução do *TSP* via Força Bruta garante sempre o melhor resultado possível ao fim da execução (garante a confiabilidade), uma vez que o mesmo é um algoritmo determinístico.

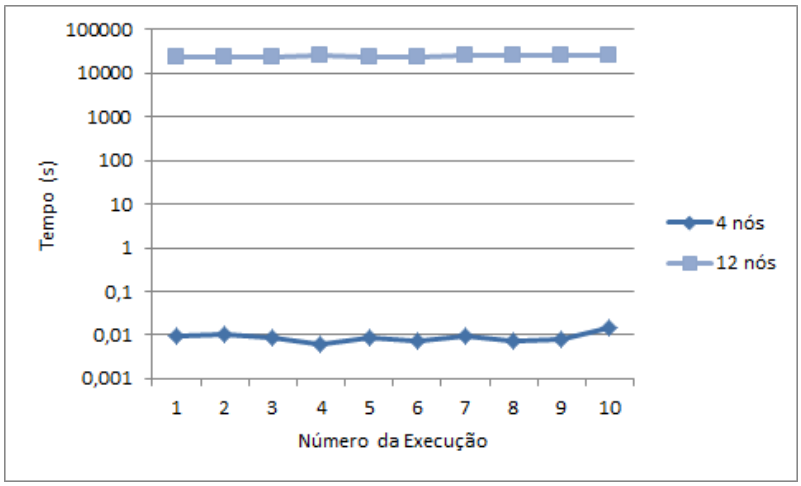


Figura 2. Resultados do tempo de execução das simulações utilizando o algoritmo com Força Bruta.

4. Solução do *TSP* com Entropia Cruzada

De acordo com [de Boer 2005], a Entropia Cruzada (*Cross Entropy* (*CE*)) é um tipo de algoritmo adaptativo que trabalha com a estimativa de probabilidades de eventos raros em redes complexas estocásticas que estejam relacionadas com a minimização da variância. Para se aplicar Entropia Cruzada em problemas de otimização combinatória difíceis, como o *TSP*, deve-se modificar o método original. Isso é obtido por meio da conversão do problema de otimização determinística para um problema relacionado à otimização estocástica. Para obter as amostras aleatórias de forma adaptativa, é necessário a utilização de técnicas de simulação de eventos raros [de Boer 2005]. A complexidade computacional de um algoritmo que utiliza *CE* é da ordem de $O(n \ln n)$ [Rubinstein and Kroese 2004].

Uma implementação de um programa em MATLAB capaz de solucionar o *TSP* via *CE* foi desenvolvida por [MATHS 2004]. As variáveis de entrada para o algoritmo de Entropia Cruzada são:

- N - Número de amostras para gerar cada rodada;
- rho - fração das melhores amostras a se tomar;
- $alpha$ - parâmetro de sensibilidade;
- C - matriz/tabela de distâncias; e
- $traj$ - se 0, posicionamento dos nós. Se 1, transição dos nós.

O algoritmo de Entropia Cruzada fornece o menor caminho (ou mais aceitável) para solucionar o problema do caixeiro viajante.

4.1. Resultados da Simulação *TSP* com Entropia Cruzada

Primeiramente foram realizadas simulações com diferentes valores para o parâmetro de tolerância. Esse parâmetro representa a tolerância a uma resposta longe da ideal [de Carvalho vaz et al. 2016]. Ou seja, trata-se de um parâmetro de tolerância a erros. Os resultados das simulações do *TSP* com Entropia Cruzada com diferentes valores para a tolerância indicaram que quanto maior o valor do parâmetro de tolerância, menor é o tempo de execução. Assim, ao fixar o parâmetro de tolerância em 0,005, foram obtidos os resultados descritos na Figura 3:

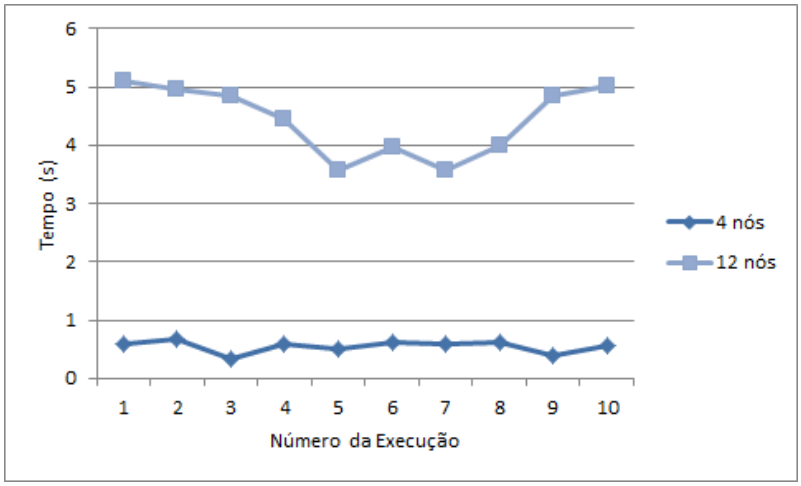


Figura 3. Resultados de tempo de execução das simulações com Entropia Cruzada, com o parâmetro de tolerância fixo.

Os demais parâmetros utilizados, descritos na Seção 4, para realizar essas simulações foram os seguintes:

- $N = 1000$;
- $rho = 0,05$;
- $alpha = 0,8$;
- $traj = 0$.

A Figura 4 ilustra o resultado da malha óptica após a otimização da rede óptica em malha através do algoritmo *TSP* com Entropia Cruzada [de Carvalho vaz et al. 2016].

5. Solução do *TSP* com Algoritmo Genético

Com o objetivo principal de buscar soluções exatas ou aproximadas em problemas de busca ou otimização, os Algoritmos Genéticos (AGs) trabalham com uma busca probabilística baseada nos princípios de uma população natural, tais como seleção, cruzamento e mutação. Assim, os AGs são "cegos", quando aplicados em sua forma mais

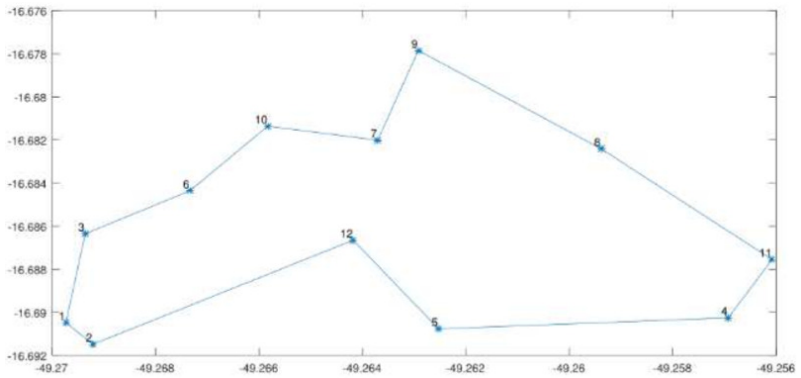


Figura 4. Solução do *TSP* com Entropia Cruzada.

pura, posto que não levam em conta o conhecimento específico do problema que será resolvido [Alves and Pinto 2006].

Para abstrair o problema, cada indivíduo de uma população é chamado de cromossomo ou genótipo, onde estes podem ser representados por uma estrutura de dados como vetor, lista, tabela *hash*, etc. Após ser inicializada, a população passa por um processo de avaliação, onde ocorre a verificação se o critério de término foi atingido. As apurações se a solução ótima global foi alcançada ou o número máximo de gerações foi atingida são utilizadas como critérios de término. Se o critério de término for atingido, o algoritmo retorna os melhores indivíduos alcançados. Caso contrário, deve-se gerar uma nova população por meio de uma seleção baseada na função de avaliação e em seguida, deve-se aplicar recombinação, mutação e realizar a avaliação dos indivíduos descendentes. Depois deve-se verificar se a nova geração da população obteve um critério de parada aceitável. Se não, repete-se todo o processo de geração de uma nova população. A Figura 5 ilustra o fluxograma de um típico AG.

As buscas heurísticas possuem tempo de execução de ordem polinomial. Dessa forma, um típico Algoritmo Genético possui uma complexidade computacional de $O(n^2)$ [Wei 2008].

Uma implementação para a solução do *TSP* via Algoritmo Genético tradicional é proposta por [Vedenyov 2011]. Realizada com o MATLAB, essa implementação plota um conjunto de pontos no plano bidimensional, formando a malha do problema de acordo com a tabela de distâncias informada no código. Cada ponto possui vários caminhos com diferentes tamanhos para os demais pontos. Como o cromossomo é uma possível solução do problema, nessa implementação, vetores de cidades (nós) são utilizados para representar os cromossomos, onde cada cidade (nó) representa um gene. As posições das cidades são informadas por meio de suas coordenadas cartesianas. Os parâmetros exigidos para executar o programa são os seguintes:

- nn - Número de nós;
- ps - Tamanho da população;
- ng - Número de gerações;
- Taxas de mutação;
  - pm - Probabilidade de mutação de troca de 2 cidades aleatórias no caminho (por gene, por geração);

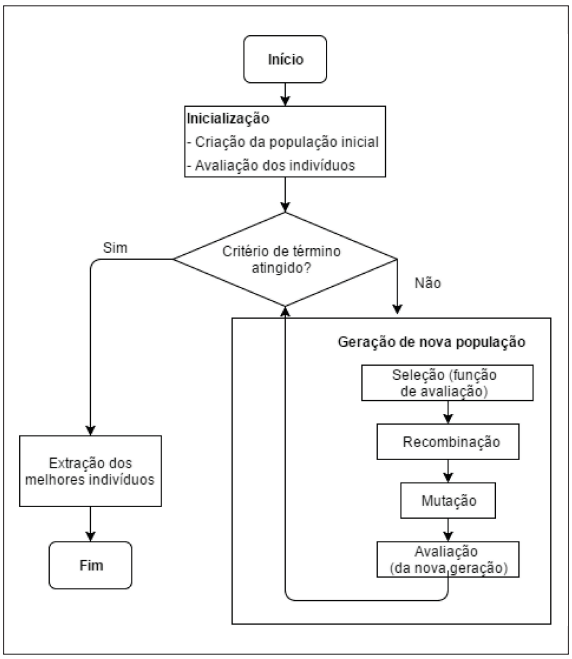


Figura 5. Fluxograma de um típico Algoritmo Genético.

  - $pm2$ - Probabilidade de mutação de troca 2 pedaços de caminho (por gene, por geração); e
  - pmf - Probabilidade de mutação de troca aleatória de caminho.
- Tabela de distâncias entre os nós.

Para resolver o *TSP*, o Algoritmo Genético segue as seguintes etapas que são repetidas a cada geração:

1. Encontrar o tamanho dos caminhos;
2. Encontrar as probabilidades (aptidão);
3. Preparar para a recombinação de acordo com as probabilidades;
4. Recombinação e substituição dos pais pelos filhos; e
5. Aplicação de mutação.

5.1. Resultados da Simulação *TSP* com Algoritmo Genético

Os resultados das simulações com Algoritmo Genético são exibidos na Figura 6. Para essas simulações foram utilizados os seguintes parâmetros:

- $ps = 3000$ (Tamanho da População);
- $ng = 200$ (Número de Gerações);
- Taxas de mutação;
  - $pm = 0,01$ (Probabilidade de mutação na troca de duas cidades aleatórias pelo caminho: por gene, por geração);
  - $pm2 = 0,02$ (Probabilidade de mutação na troca de dois pedaços de caminho: por gene, por geração);
  - $pmf = 0,08$ (Probabilidade de mutação em um pedaço de caminho aleatório).

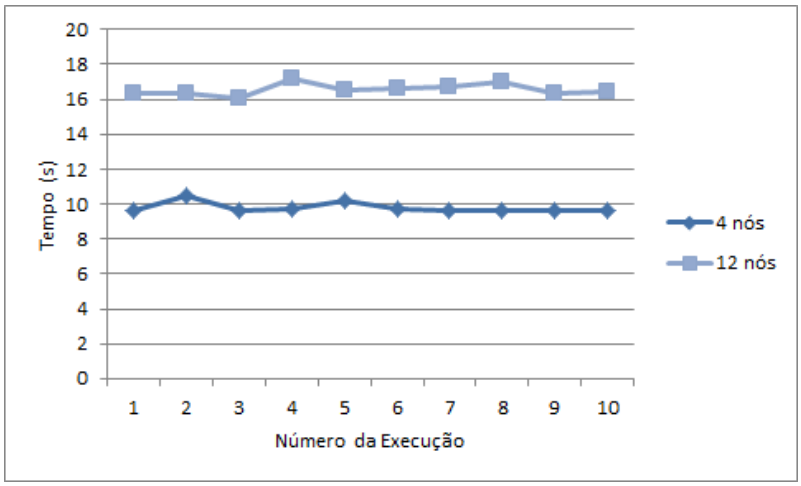


Figura 6. Resultados do tempo de execução das simulações com Algoritmo Genético.

É importante destacar que o tempo de execução apresentado na Figura 6 representa o tempo de execução até a última geração de indivíduos. Assim, foi possível notar que o tempo de execução foi menor em certas simulações, onde o algoritmo convergiu mais rapidamente, ou seja, o algoritmo encontrou o menor caminho na rede antes de chegar na última geração. Como o critério de parada utilizado foi o de obter todas as 200 gerações, os valores efetivos dos tempos de execução registrados na Figura 6 foram relativamente altos.

A Figura 7 apresenta a comparação do tempo médio de execução dos três algoritmos de único objetivo estudados.

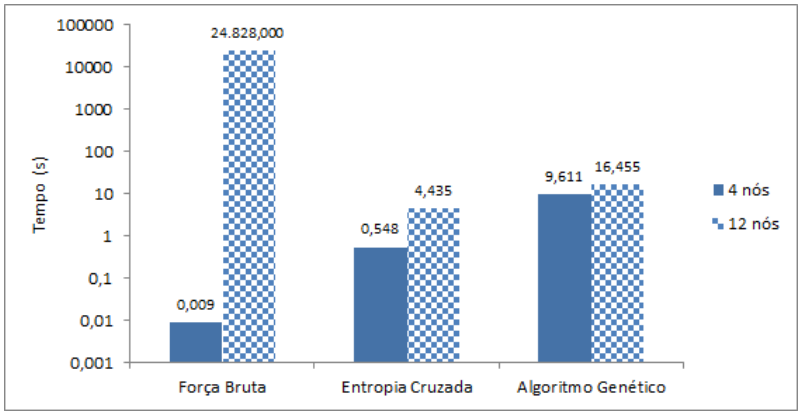


Figura 7. Comparação do tempo médio dos algoritmos de único objetivo.

6. Dimensionamento de Rede Óptica Utilizando *NSGA-II*

Quando um problema de otimização possui múltiplos objetivos, obter apenas uma solução ótima não é suficiente. Nesse caso, é necessário obter um conjunto de soluções ótimas. Esse conjunto é chamado de soluções pareto-ótima.

As soluções clássicas para a solução desse tipo de problema consistem em dividir os problemas de otimização multiobjetivos em problemas simples de único objetivo, de

forma a tratar cada solução pareto-ótima de uma vez. Nesse sentido, junta-se todas as soluções pareto-ótima obtidas individualmente e aplica-se esse método várias vezes até se obter o conjunto ótimo.

O *NSGA-II* é capaz de solucionar problemas de otimização multiobjetivo com restrições com uma complexidade computacional da ordem de $O(MN^2)$. Para isso, o *NSGA-II* utiliza um rápido procedimento de seleção não-dominada com uma abordagem elitista de preservação da diversidade das soluções [Deb et al. 2002].

O primeiro passo desse algoritmo é, para cada solução, calcular duas entidades na seguinte ordem:

1. Contagem dominante, n_p: trata-se do número de soluções que dominam a solução p; e
2. s_p: é o conjunto de soluções que a solução p domina.

No algoritmo de geração de populações do *NSGA-II*, uma população R_t, de tamanho $2N$ é formada pela união da população de pais (P_t) e de descendentes (Q_t). Em seguida, a população R_t é classificada de acordo com a não-dominação. Nesse caso, o elitismo é garantido, visto que todos os membros das populações anteriores e atuais dos pais e dos descendentes estão contidos em R_t. As soluções pertencentes ao melhor conjunto não-dominado F_1 são as melhores soluções na população combinada e devem, portanto, ser enfatizadas mais do que qualquer outra solução na população combinada. Se o tamanho da população F_1 for menor que N, deve-se escolher todos os membros do conjunto F_1 para a nova população P_{t+1}. Os membros restantes de P_{t+1} são escolhidos a partir das frentes não-dominadas subsequentes na ordem de sua classificação. O mesmo procedimento é repetido para os demais conjuntos F_i até que não seja possível acomodar mais nenhum conjunto. Digamos que o conjunto F_l seja o último conjunto não-dominado além do qual nenhum outro conjunto pode ser acomodado. Em geral, o número de soluções em todos os conjuntos de F_1 até F_l seria maior do que o tamanho da população. Para escolher o número exato de membros (N), ordena-se as soluções da última frente F_l utilizando o operador de comparação por aglomeração ($\prec_n$) em ordem descendente e escolhe-se as melhores soluções necessárias para para preencher todos os espaços da população. A Figura 8 apresenta um esquema do processo realizado no *NSGA-II*. A nova população P_{t+1} de tamanho N é utilizada agora para seleção, recombinação e mutação para criar a nova população Q_{t+1} de tamanho N [Deb et al. 2002].

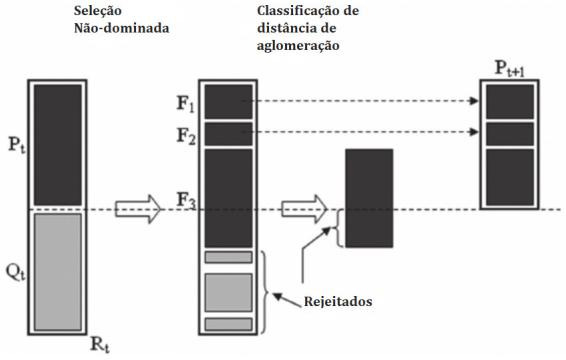


Figura 8. Esquema do algoritmo de geração de populações do *NSGA-II* [Deb et al. 2002].

6.1. Resultados do Dimensionamento de Rede Óptica com NSGA-II

Para o estudo de caso, utilizou-se como variável de decisão uma relação de todos os religadores (nós) da rede, indicando a conexão de cada nó com os demais nós da rede. Para estabelecer essa relação, utilizou-se uma matriz de custos para representar as posições dos religadores. Nessa estrutura, foram indicados, a partir de valores binários, quais nós estavam conectados, onde 1 representa conexão e 0 indica que os nós estão soltos. Em cada geração do algoritmo, uma nova matriz de custos otimizada é gerada. A última geração representa a melhor configuração (menor custo e maior vazão - otimização das funções objetivo) da rede obtida pelo algoritmo.

As funções objetivo otimizadas no problema foram o custo e a vazão. Análogo ao objetivo otimizado nos algoritmos apresentados anteriormente, buscou-se minimizar o custo, no sentido de obter sempre o menor caminho na rede. Já o segundo objetivo consistiu em maximizar a vazão de cada nó da rede, ou seja, a taxa efetiva de pacotes transmitidos por unidade de tempo. Esse objetivo foi adotado para garantir um bom desempenho na transmissão dos dados na rede. Mais informações a respeito de cálculos da vazão podem ser encontradas em [de Castro Souza 2014] [Liu and Liao 2008].

Foram consideradas as seguintes restrições, quando se aplicou o *NSGA-II* para solucionar o problema:

- O número total de conexões em cada nó deve ser maior ou igual a 1, ou seja cada nó deve estar conectado a pelo menos um outro nó da rede; e
- O número máximo de conexão para cada nó é 4, pelo fato de os religadores utilizados na rede possuírem apenas 4 conectores.

Para realizar as simulações, considerou-se 100 gerações e uma população de 50 indivíduos. Além disso, o projeto do estudo de caso utilizou os nós 3, 4, 6 e 10 para conectar diretamente ao *gateway*. Por motivos práticos, considerou-se na implementação da solução com o *NSGA-II* que apenas um nó estaria conectado ao *gateway* de cada vez. Essa implementação foi feita de tal forma a ser flexível quanto a escolha de um nó da rede para se conectar ao *gateway*.

As Figuras 9 e 10 apresentam o conjunto de soluções pareto-ótimas da primeira e da última geração, respectivamente, considerando as restrições, um *buffer* de tamanho 50 (pacotes) e uma taxa de chegada de 1000 pacotes por unidade de tempo para o *gateway* conectado ao nó 3. Esse nó foi escolhido por ser um dos nós que estão conectados ao *gateway* no projeto do estudo de caso. O eixo das abcissas representa o custo a ser minimizado, enquanto que o eixo das ordenadas representa a vazão a ser maximizada.

A partir da comparação dos resultados apresentados nas Figuras 9 e 10, é possível verificar o avanço do algoritmo em obter um conjunto de soluções ótimas. Nesse caso, todas as soluções do algoritmo convergiram para o mesmo resultado, ou seja, na última geração é possível verificar que todas as 50 populações de indivíduos encontram-se no mesmo ponto do gráfico, enquanto que na primeira geração as populações de indivíduos estavam dispersas.

A Figura 11 ilustra um grafo contendo a configuração da melhor rede obtida nessa simulação, apresentando as conexões entre cada um dos nós. Analisando-se o grafo obtido na Figura 11 nota-se que cada nó realiza no mínimo 1 conexão e no máximo 4 conexões, respeitando as restrições estabelecidas pelo problema. Essa solução gerou uma

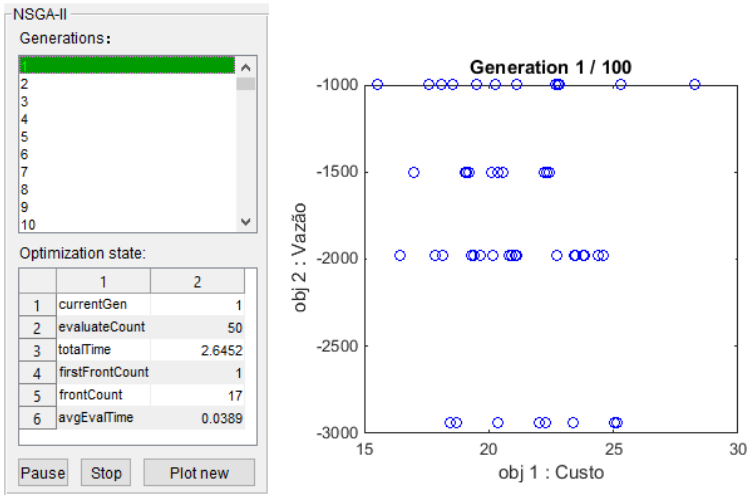


Figura 9. Conjunto de soluções pareto-ótimas obtido na primeira geração da simulação considerando as restrições com o *gateway* conectado ao nó 3.

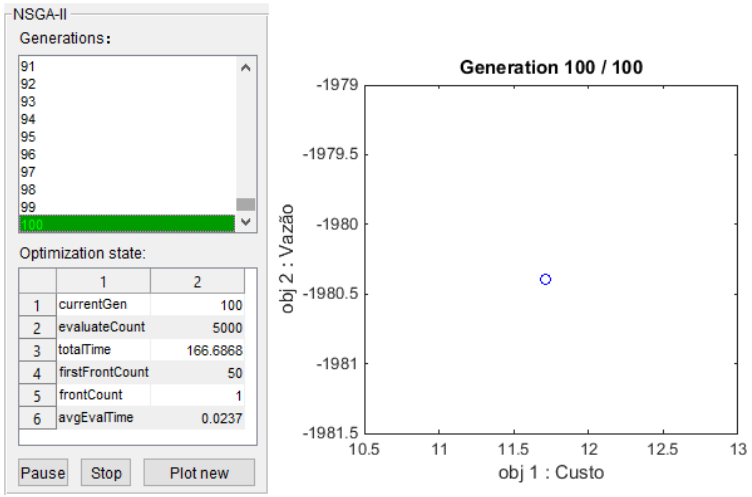


Figura 10. Conjunto de soluções pareto-ótimas obtido na última geração na simulação considerando as restrições com o *gateway* conectado ao nó 3.

vazão de, aproximadamente, 1980 pacotes por unidade de tempo e custo na ordem de 12 quilômetros.

Apesar de a simulação ter encontrado somente um conjunto de soluções ótimas, nem sempre isso ocorre. Em alguns casos, obteve-se mais de um conjunto de soluções ótimas, conforme pode ser observado na Figura 12. Nesse exemplo de simulação, utilizou-se o nó 10 conectado ao *gateway*, considerando as restrições do problema.

Pela Figura 12 pode-se observar que as soluções possuem um certo padrão, onde o custo e a vazão são inversamente proporcionais. Ou seja, a função objetivo da vazão aumenta para valores menores da função objetivo do custo e vice-versa.

Para confirmar os resultados obtidos, foram realizadas simulações também no simulador de rede *Network Simulator 2* (*NS2*) para se comparar a vazão da rede obtida na simulação com o *NSGA-II* (considerando as restrições para o *gateway* conectado ao nó 3

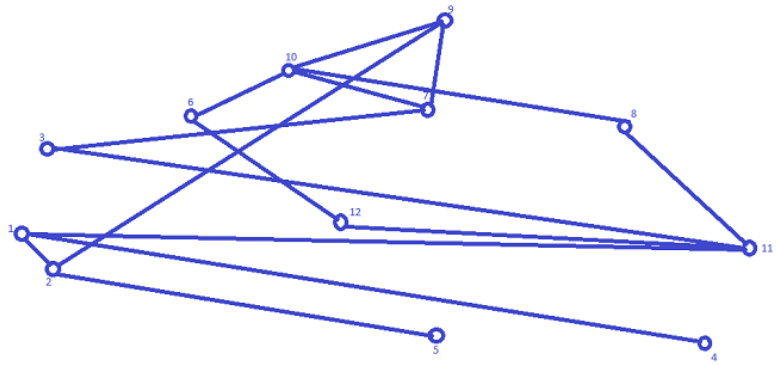


Figura 11. Melhor configuração da rede obtida na simulação considerando as restrições com o *gateway* conectado ao nó 3.

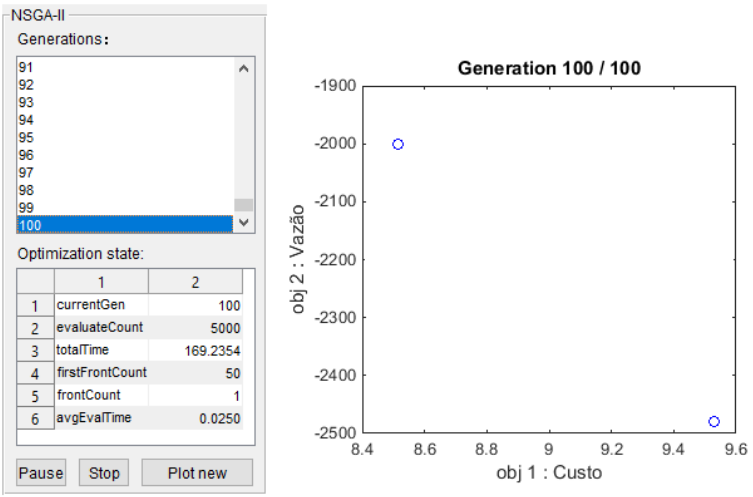


Figura 12. Múltiplas soluções obtidas na última geração da simulação considerando as restrições com o nó 10 conectado ao *gateway*.

- Figura 11) e a vazão obtida pelo simulador *NS2*.

Para realizar essa comparação, utilizou-se os seguintes parâmetros no simulador de rede:

- Taxa de transmissão dos enlaces: Gigabits por unidade de tempo (Esse valor foi escolhido devido às fibras monomodos possuírem taxas de transmissão na ordem de 1×10^9 bits por unidade de tempo);
- Tamanho do pacote: 2KB;
- Tamanho do *buffer*: 50 pacotes;
- Taxa de envio de pacotes: 1000 pacotes por unidade de tempo;
- Protocolo de Roteamento: Vetor Distância; e
- Tipo de Fila: *Drop Tail* que representa o algoritmo (*First In, First Out*);

O resultado obtido pelo simulador foi uma vazão média de aproximadamente 1835 pacotes por unidade de tempo. De acordo com a Figura 10, o resultado obtido pelo *NSGA-II* foi de 1980 pacotes por unidade de tempo. Analisando-se os dois resultados, pode-se considerar uma certa proximidade entre a simulação feita no *NSGA-II* e o *NS2*, já que não

espera-se que o resultado final da vazão média da simulação teórica (com o *NSGA-II*) seja exatamente igual à vazão da simulação prática (com o *NS2*), mas que os valores mostrem uma certa proximidade que torne os resultados teóricos mais confiáveis.

Também utilizou-se o *NS2* para simular qual seria a vazão do anel apresentado na Figura 4, obtido com a aplicação direta do *TSP*. Com isso, verificou-se que a vazão total da rede para a configuração em anel foi de 22000 pacotes por segundo, ou seja uma vazão média de aproximadamente 1830 pacotes por segundo. Assim, a configuração em anel apresentada na simulação da Figura 4 possui uma vazão pouco menor do que a vazão obtida na simulação da rede com o *NSGA-II* apresentada na Figura 11. Esse resultado é mais uma justificativa para o uso da abordagem multiobjetiva.

7. Conclusão

O presente trabalho teve o objetivo de estudar, analisar e otimizar a rede *Mesh* de um sistema *self-healing* de energia elétrica, além de tentar maximizar a transmissão de dados (vazão) da mesma. Para a pesquisa e a análise de resultados, foram utilizados os algoritmos *TSP* para otimizar a distância de nós e enlaces e o *NSGA II* para pesquisa na comunicação dos religadores.

Este estudo mostrou que o algoritmo que utiliza Entropia Cruzada é o que obteve melhor média de tempo de execução. Embora o Algoritmo Genético tenha proporcionado um tempo médio de execução superior ao do Entropia Cruzada, valores de tempo de execução mais baixos podem ser obtidos pelo AG de acordo com o tamanho do população e parâmetros iniciais. Por fim, o algoritmo de Força Bruta foi o que obteve o pior desempenho, o que já era esperado, devido à sua alta complexidade computacional, já previamente conhecida.

Dessa forma, para o caso de um problema com apenas um objetivo (minimizar o custo de fibra óptica, ou seja, obter o caminho que exija o menor comprimento total de fibra) o algoritmo mais adequado para se utilizar no estudo de caso é o que utiliza o método de Entropia Cruzada, uma vez que o mesmo encontrou os caminhos mais curtos possíveis (ou mais aceitáveis) de forma mais rápida.

A inclusão da otimização da vazão como mais um objetivo faz com que o problema torne-se multiobjetivo (maximizar a vazão e minimizar o custo). Para isso, se faz necessário o uso do *NSGA-II*, um algoritmo genético multiobjetivo. O motivo de se maximizar a vazão se deve à busca de uma comunicação veloz entre os religadores, para que quando ocorra falhas na rede, a distribuição de energia não seja interrompida ou, em caso de interrupção, o tempo de recuperação da rede seja o menor possível. Isso é benéfico tanto para o fornecedor quanto para o consumidor, uma vez que a operadora de energia elétrica irá ter menos despesas, evitando ter de pagar multas para a Agência Nacional de Energia Elétrica (ANEEL), além de que os consumidores evitarão os transtornos ocasionados pelas quedas de energia elétrica. Também foi possível perceber que a inclusão da vazão sacrifica em partes o custo (e vice-versa), uma vez que os objetivos demostraram um comportamento inversamente proporcional na simulação com mais de um conjunto de soluções ótimas.

De acordo com as simulações em conjunto com as equações de vazão de Teoria de Filas (vide o trabalho do [de Castro Souza 2014]), percebe-se que a vazão está diretamente ligada a dois parâmetros: taxa de chegada de pacotes e o *buffer* de armazenamento.

Com base nas simulações feitas, essas duas variáveis impactam no comportamento da rede quando se fala em transmissão de dados, além de mudar em vários casos a configuração dos nós e enlaces da rede *self-healing*, tentando sempre achar a melhor configuração possível para que a vazão seja máxima e o custo (configuração de nós e enlaces) seja o menor possível.

Referências

Alves, F. T. and Pinto, P. E. D. (2006). *Aplicacação de Algoritmos Genéticos ao Problema do Percurso do Cavalo*. Universidade Estadual do Rio de Janeiro.

Burkardt, J. (2013). Brute force solution of travelling salesman problem. Florida State University.

da Cruz, J. A. G. (2010). *Análise de Protocolos de Roteamento em Redes de Sensores Sem Fio*. Universidade Federal de Pelotas.

de Boer, P.-T. (2005). *A Tutorial on the Cross-Entropy Method*. University of Twente.

de Carvalho vaz, R. R., Vieira, F. H. T., Queiroz, P., de Deus Junior, G. A., and de Araújo, S. G. (2016). *Aplicação do Método de Entropia Cruzada para Projeto de Redes Ópticas em Malha via Resolução do TSP*.

de Castro Souza, G. B. (2014). Otimização do posicionamento de concentradores gprs em redes elétricas inteligentes utilizando programação linear e teoria de filas. Master's thesis, Universidade Federal de Goiás, Goiânia-GO.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). *A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II*. Indian Institute of Technology.

Liu, T. and Liao, W. (2008). *Location-dependent throughput and delay in wireless mesh networks*.

Maredia, A. (2010). History, analysis, and implementation of traveling salesman problem (tsp) and related problems. Master's thesis, University of Houston-Downtown, Department of Computer and Mathematical Sciences, Houston, Estados Unidos da América.

MATHS (2004). Travelling salesmanl. MATHS.

Potvin, J.-Y. (1996). *Genetic algorithms for the traveling salesman problem*. Centre de Recherche sur les Transports, Université de Montréal.

Rubinstein, R. Y. and Kroese, D. P. (2004). *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer; Softcover reprint of the original, Estados Unidos da América, 1 edition.

Sahalot, A. and Shrimali, S. (2014). A comparative study of brute force method, nearest neighbour and greedy algorithms to solve the travelling salesman problem. *IMPACT: International Journal of Research in Engineering & Technology*, 2(1):59–72.

Vedenyov, M. (2011). Travelling salesman problem with genetic algorithm. MathWorks.

Wei, J.-D. (2008). Approaches to the travelling salesman problem using evolutionary computing algorithms. *InTech - Open Science Open Minds*, pages 63–74.

Aplicação de Algoritmos Genéticos para Estimação de Parâmetros de Painéis Fotovoltaicos

Flávio Henrique Teles Vieira[1], **Gilberto Lopes Filho**[1], **Ricardo Augusto Pereira Franco**[1]

[1] Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG)
Av. Universitária, nº. 1488 – Goiânia – GO – Brasil

flavio@eee.ufg.br, gilbertofilho93@gmail.com, ricardofranco3@gmail.com

Resumo. *A geração de energia fotovoltaica tem se tornado uma ótima opção para a geração de energia elétrica limpa, visto que ela possui alto potencial de produção de energia elétrica em todo o mundo. O equipamento responsável pela transformação dos raios solares em energia elétrica é o painel fotovoltaico. Sua simulação computacional é essencial para a realização de projetos de sistemas fotovoltaicos. Contudo, os datasheets dos painéis não fornecem todas as informações necessárias para a simulação computacional. Portanto, torna-se necessário desenvolver métodos para estimar alguns parâmetros dos painéis. Este trabalho propõe um método de estimação de parâmetros de painéis fotovoltaicos utilizando o Algoritmo Genético, desenvolvido no ambiente MATLAB. A função objetivo do Algoritmo Genético corresponde à minimização do erro quadrático médio entre a curva característica IxV experimental do painel em relação à curva IxV gerada pelos parâmetros estimados. Os resultados obtidos com o Algoritmo Genético são comparados com outros métodos presentes na literatura.*

Abstract. *The photovoltaic energy is an interesting option for clean energy generation in the entire world, since it has a high potential for electricity generation worldwide. The equipment responsible for transforming the solar rays into electrical energy is the photovoltaic panel. The computational simulation of photovoltaic panels is essential to design photovoltaic systems. However, the datasheets of the panels does not provide all information necessary for the computational simulation. Therefore, it becomes necessary to develop methods to estimate some parameters of the panels. This work proposes the application of the Genetic Algorithm to estimate parameters of photovoltaic systems, in the MATLAB environment. The objective function of the Genetic Algorithm corresponds to the minimization of the root mean square error between the IxV experimental curve and the IxV curve generated by the estimated parameters. The results obtained with the Genetic algorithm are compared to those of some methods present in the literature.*

1. Introdução

A crescente demanda por energia elétrica demonstra a dependência da sociedade por este tipo de energia. O setor elétrico vem desenvolvendo novos sistemas para aumentar a

confiabilidade na rede elétrica, aprimorando suas funcionalidades e sua infraestrutura. Há, também, a preocupação com o impacto ambiental na geração de energia, impactos estes que podemos citar a elevação da temperatura como um dos tópicos mais abordados [Aramizu 2010], [da Costa 2010]. Portanto, a geração de energia através da luz solar é uma medida que beneficia tanto a concessionária de energia, o consumidor e o meio ambiente.

A energia solar é uma fonte de energia renovável. Essa geração é realizada por meio da captação da radiação solar, na qual é feita sua conversão em energia elétrica por meio do efeito fotovoltaico na célula fotovoltaica [Aramizu 2010]. Nesse efeito há a excitação de elétrons, devido às características de alguns materiais semicondutores na presença de luz solar, ocorrendo o efeito fotovoltaico e gerando a corrente elétrica desse sistema.

Um conjunto de células fotovoltaicas formam um painel fotovoltaico. O painel fotovoltaico pode ser representado por meio de um circuito elétrico que possui 5 parâmetros desconhecidos [Villalva et al. 2009]. Estes parâmetros não são fornecidos pelo *datasheet*, que é um material contendo informações sobre a placa fotovoltaica e é cedido pelo seu fabricante, sendo necessário, portanto, a sua estimação. Essa estimação trata-se de um problema envolvendo modelagem matemática e algoritmos de otimização. Alguns trabalhos, como o de [Cubas et al. 2014], utilizam formas analíticas aproximadas para o cálculo dos parâmetros. Outros trabalhos, como o de [El-Naggar et al. 2012] e [Askarzadeh and Rezazadeh 2013], utilizam algoritmos de busca e otimização a fim de se estimar esses parâmetros.

Uma grande parte dos algoritmos de otimização convencionais são do tipo determinístico, e dentre eles, existem aqueles que são baseados na informação do gradiente da função. Um exemplo é o método de Newton-Raphson, que possui uma boa performance para funções bem comportadas. Porém, quando se trata de problemas altamente não lineares, não convexos, não diferenciáveis e não suaves, estes métodos determinísticos baseados em gradiente apresentam problemas na convergência, e por muitas vezes ficam presos em ótimos locais. Para contornar esse problema, é necessário utilizar algoritmos que não sejam baseados no gradiente da função, como por exemplo, algoritmos estocásticos. Estes algoritmos, também chamados de algoritmos heurísticos, fazem uma busca aleatória pela melhor solução. Essa busca muitas vezes é, de certa forma, direcionada e garante que o método convirja para uma boa solução [Yang 2008].

Um tipo de algoritmo heurístico é o algoritmo genético. Este algoritmo é baseado em métodos heurísticos de busca aleatória inspirados pela biologia evolutiva como hereditariedade, mutação, seleção natural e recombinação. Os algoritmos genéticos são implementados de tal forma que uma população de representações abstratas de solução é selecionada em busca de soluções melhores. Os algoritmos genéticos utilizam transições probabilísticas e não regras determinísticas [Holland 1992].

Neste trabalho serão utilizadas equações analíticas, propostas pela literatura, para o cálculo de três parâmetros dos módulos e será proposto um algoritmo de busca e otimização dos outros dois parâmetros restantes. A seção 2 mostra a modelagem do painel fotovoltaico utilizada neste trabalho. A seção 3 descreve o algoritmo genético utilizado. A seção 4 apresenta os resultados deste trabalho, comparando-os com outros presentes na

literatura.

2. Modelo do Painel Fotovoltaico

O modelo de um painel fotovoltaico é representado por meio de um circuito elétrico que o define. Essa modelagem se justifica porque é necessário realizar simulações precisas e confiáveis acerca do sistema a ser instalado, devido ao seu alto custo de implantação. A modelagem matemática é definida em função do modelo do painel fotovoltaico escolhido.

O modelo do painel fotovoltaico com um diodo é o modelo escolhido para este trabalho. Este modelo é um modelo complexo, preciso e confiável, satisfazendo os requisitos a confiabilidade da simulação do sistema fotovoltaico ([da Costa 2010]). O circuito elétrico do modelo é apresentado na Figura 1.

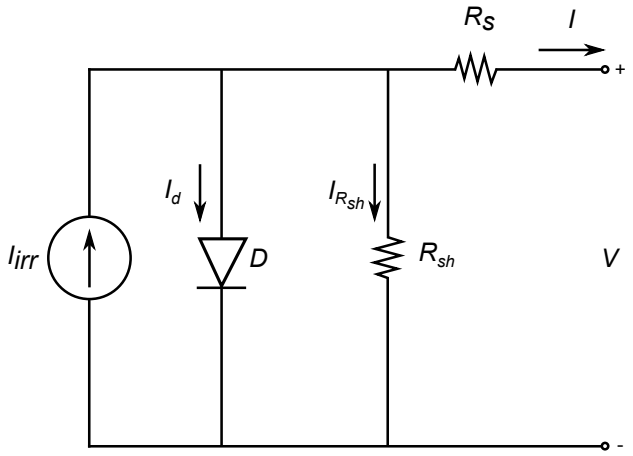


Figura 1. Circuito elétrico que representa o painel fotovoltaico

onde os parâmetros a serem estimados do modelo são: a corrente fotogerada (I_{irr}), a corrente de saturação do diodo (I_0), a resistência em série (R_s), a resistência em paralelo (R_{sh}) e o fator de idealidade do diodo (n). O parâmetro I é a corrente gerada pelo painel, a ser calculada, e V é a tensão entre os terminais do painel.

O modelo do painel fotovoltaico escolhido influencia diretamente o comportamento do sistema fotovoltaico a ser simulado, porque a expressão matemática que define o modelo é determinada pelo circuito elétrico escolhido ([de Blas et al. 2002]). A equação da corrente resultante é chamada de equação característica dada pela Eq. 1.

$$I = I_{irr} - I_0 \left[e^{\left(\frac{V + IR_s}{N_s n V_t} \right)} - 1 \right] - \frac{V + IR_s}{R_{sh}} \tag{1}$$

onde $Id = I_0 \left[e^{\left(\frac{V + IR_s}{N_s n V_t} \right)} - 1 \right]$.

Portanto, como alguns dados não são fornecidos pelo *datasheet*, eles devem ser estimados (I_{irr}, I_0, R_s, R_{sh} e n).

O trabalho [Laudani et al. 2014] propõe obter valores analíticos para alguns parâmetros por meio da manipulação da Eq. 1 no ponto de circuito aberto ($I = 0$, $V = V_{ca}$),

no ponto de curto circuito ($I = I_{cc}, V = 0$), no ponto de máxima potência ($I = I_{mp}$, $V = V_{mp}$) e por meio da derivada da potência em função da tensão no ponto de máxima potência $\left(\frac{\partial P}{\partial V}\Big|_{V=V_{mp}} = 0\right)$. Gerando, assim, equações para três dos cinco parâmetros a serem estimados. Os valores da corrente fotogerada (I_{irr}), da corrente de saturação do diodo (I_0) e da condutância em paralelo (G_{sh}) são dados, respectivamente, por:

$$I_{irr} = \frac{I_{cc}V_{ca}\left(e^{\left(\frac{V_{mp}+R_sI_{mp}}{N_snV_t}\right)} - 1\right) + I_{cc}V_{mp}\left(1 - e^{\left(\frac{V_{ca}}{a}\right)}\right) + I_{mp}V_{ca}\left(1 - e^{\left(\frac{R_sI_{cc}}{a}\right)}\right)}{A_1e^{\left(\frac{R_sI_{cc}}{a}\right)} + A_2e^{\left(\frac{V_{mp}+R_sI_{mp}}{a}\right)} + A_3e^{\left(\frac{V_{ca}}{a}\right)}} \tag{2}$$

$$I_0 = \frac{V_{ca}\left(I_{cc} - I_{mp}\right) - V_{mp}I_{cc}}{A_1e^{\left(\frac{R_sI_{cc}}{a}\right)} + A_2e^{\left(\frac{V_{mp}+R_sI_{mp}}{a}\right)} + A_3e^{\left(\frac{V_{ca}}{a}\right)}} \tag{3}$$

$$G_{sh} = \frac{\left(I_{mp} - I_{cc}\right)e^{\left(\frac{V_{ca}}{a}\right)} + e^{\left(\frac{V_{mp}+R_sI_{mp}}{a}\right)} - (I_{mp})e^{\left(\frac{R_sI_{cc}}{a}\right)}}{A_1e^{\left(\frac{R_sI_{cc}}{a}\right)} + A_2e^{\left(\frac{V_{mp}+R_sI_{mp}}{a}\right)} + A_3e^{\left(\frac{V_{ca}}{a}\right)}} \tag{4}$$

onde $G_{sh} = 1/R_{sh}$; $A_1 = V_{mp}+R_sI_{mp}-V_{ca}$; $A_2 = V_{ca}-R_sI_{cc}$; $A_3 = R_sI_{cc}-R_sI_{mp}-V_{mp}$ e $a = N_snV_t$.

É necessário desenvolver novos métodos para calcular os demais parâmetros (R_s e n) que não possuem equações analíticas. Assim, este trabalho utiliza o conceito de Algoritmos Genéticos para desenvolver um algoritmo que otimize a estimação dos parâmetros R_s e n, minimizando o erro ($RMSE$) entre os pontos experimentais e calculados da curva característica.

3. Algoritmo Genético

O conjunto das possíveis soluções de um algoritmo genético é denominado população. A população é composta por indivíduos, que por sua vez são compostos de genes. Cada indivíduo é uma abstração de uma possível solução, de forma que o número de genes dos indivíduos depende de cada problema [Holland 1992].

O algoritmo genético possui as etapas de seleção, cruzamento e mutação. Na etapa de seleção, são escolhidos os melhores indivíduos para serem os progenitores. [Linden 2006] mostra que há várias formas de se fazer essa seleção, como por exemplo a roleta viciada ou torneio. Na etapa de cruzamento ocorre uma combinação dos genes dos indivíduos escolhidos para serem os progenitores, de forma a gerar um novo indivíduo. A mutação é uma perturbação adicionada aos genes do novo indivíduo, o que faz com que novos tipos de genes sejam introduzidos na população, aumentando o espaço de busca e evitando mínimos locais [Linden 2006].

Foi implementado neste trabalho um algoritmo genético a fim de se estimar os parâmetros n e Rs do circuito mostrado na Figura 1. Para tal fim, utilizou-se as Equações 4, 3 e 2, que estão em função de n e R_s. Para cada iteração, onde se calcula estes dois parâmetros, os outros três parâmetros também são obtidos.

A Figura 2 apresenta um fluxograma do algoritmo. No início é gerado uma população inicial com 30 indivíduos, cada indivíduo com 2 genes. Cada gene representa um dos parâmetros que estão sendo estimados. Em seguida os indivíduos passam pelo simulador, onde cada um é avaliado por uma função de avaliação, que ordena os indivíduos do melhor para o pior.

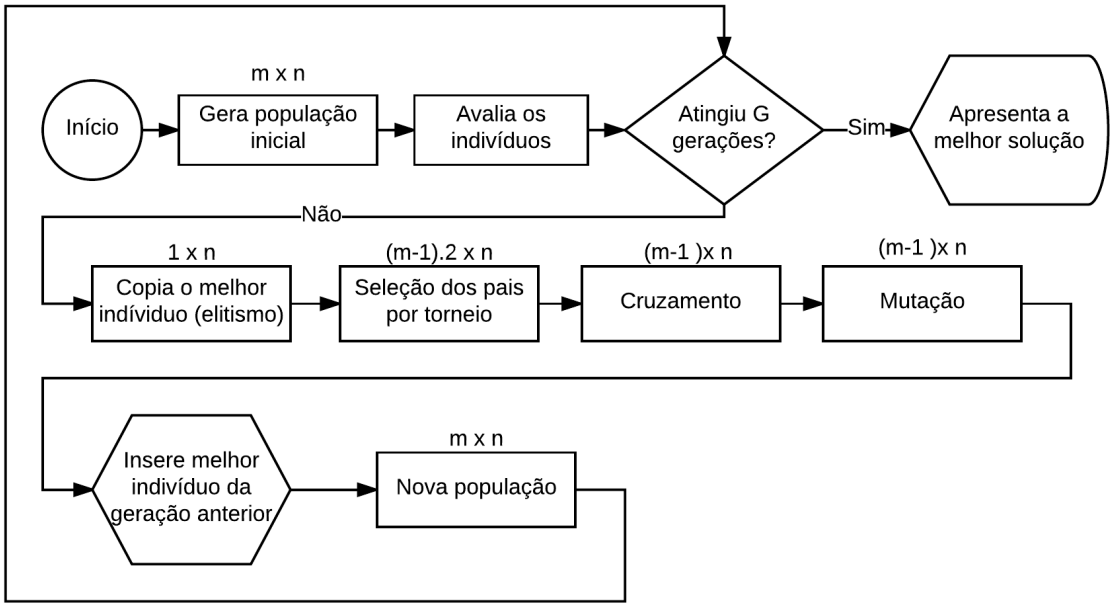


Figura 2. Fluxograma do Algoritmo Genético

A função de avaliação utilizada foi a raiz quadrada do erro quadrático médio, doravante chamada de $RMSE$ (do inglês *root mean square error*). O $RMSE$ foi calculado em relação aos pontos de corrente e tensão disponibilizados no *datasheet* do módulo fotovoltaico Kyocera KC200GT. Isto é, para cada valor do parâmetro n e do parâmetro Rs, plota-se uma curva, e o $RMSE$ é calculado entre esta curva e os pontos do *datasheet*. Quanto menor o $RMSE$, mais adequados estão os parâmetros n e Rs, e mais a curva encontrada se aproxima da curva real.

O critério de parada desse algoritmo foi especificado como um limite de gerações. Neste trabalho adotou-se o número máximo de gerações como 100. Foi adotado também a técnica de elitismo, isto é, a manutenção do melhor indivíduo da geração anterior na próxima geração. Isso garante que no pior caso, o pior indivíduo de uma determinada geração ainda será melhor ou igual que o melhor indivíduo da geração anterior. Isso faz com que o algoritmo convirja para uma boa solução.

A seleção para os progenitores da próxima geração é realizada por torneio, onde se escolhe k indivíduos aleatoriamente, e os dois melhores serão os pais de um indivíduo da próxima geração. O processo é repetido até que haja a quantidade de pais suficientes para manter o número de indivíduos da próxima geração igual aos da geração anterior. A quantidade k de indivíduos é determinado pelo parâmetro τ. Este parâmetro define a porcentagem da população total que irá participar do torneio. Neste trabalho foi adotado $\tau = 70\%$, isto é, dos 30 indivíduos da população, são escolhidos 21 aleatoriamente para

cada torneio.

Na etapa de cruzamento ocorre a troca de genes entre os progenitores para gerar os filhos. Escolhe-se aleatoriamente um número η, que é a variável probabilística de decisão. O valor de η diminui de acordo com o aumento das gerações, porém é um valor randômico, conforme mostra a Figura 3. Caso η seja maior que um β, onde β é um número aleatório entre 0 e 1, o filho receberá o gene do primeiro pai, e caso contrário, receberá o gene do segundo pai. Esse processo ocorre até que o filho tenha a mesma quantidade de cromossomos dos pais.

Na etapa de mutação ocorre uma pertubação que se provoca em um determinado gene para a promoção da exploração no espaço de busca. Nesta etapa foi definido que há 50% de chance dos indivíduos sofrerem mutação. Caso haja mutação, é sorteado um Δ de mutação que é aplicado ao gene. Neste trabalho foi adotado um $\Delta = 5\%$, ou seja, quando ocorre mutação, cada gene pode ter seu valor alterado em até 5%. Este processo é ilustrado na Figura 4.

A nova população é constituída dos filhos gerados pelos pais, adicionado também do melhor indivíduo da geração anterior. Novamente se avalia os indivíduos conforme o $RMSE$, e o algoritmo é recomeçado.

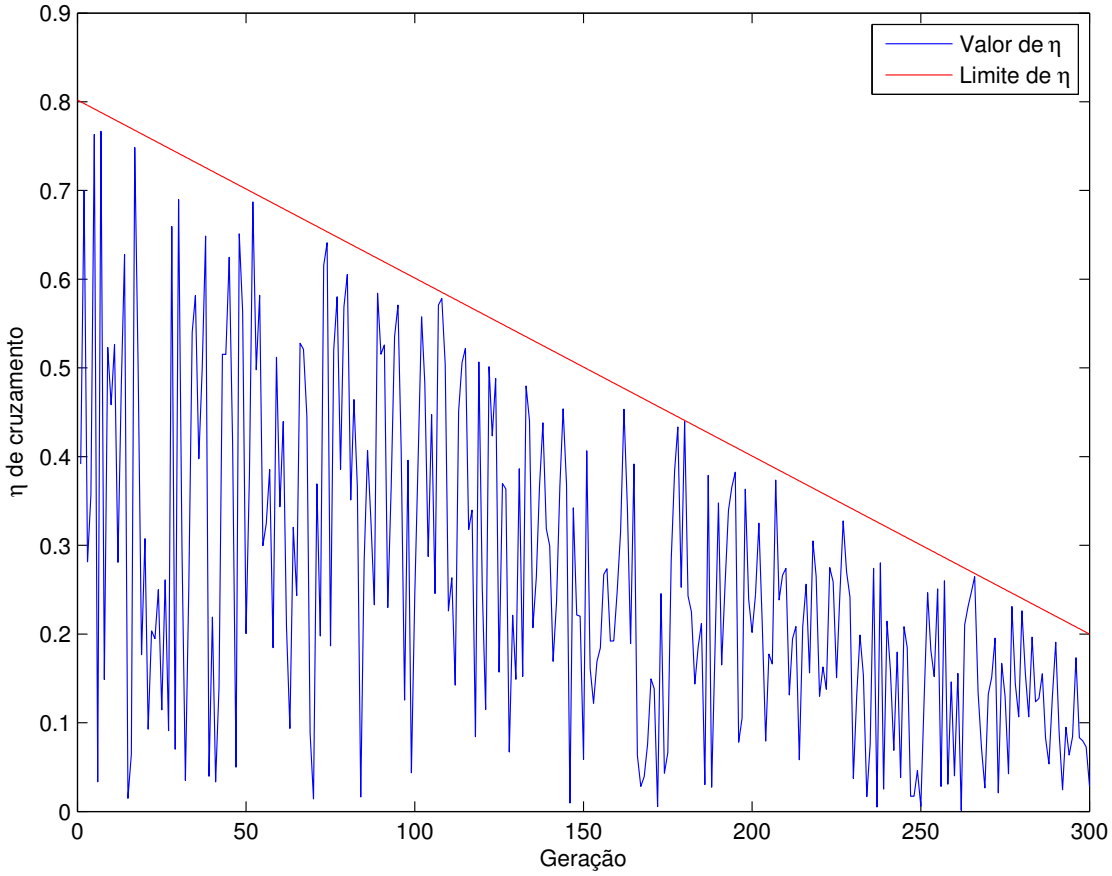


Figura 3. Decréscimo da taxa de cruzamento com o aumento das gerações

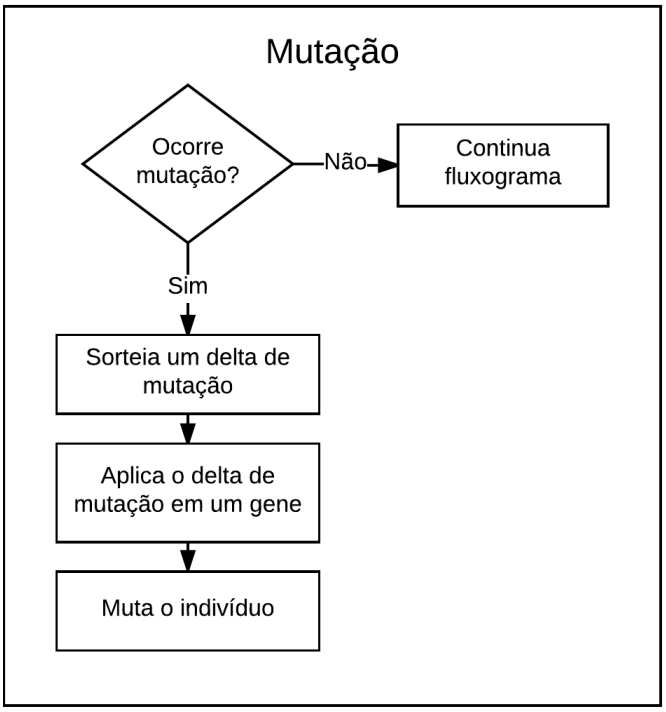


Figura 4. Detalhe da mutação do Algoritmo Genético

4. Resultados

Os resultados foram obtidos por meio do *software* MATLAB e comparados com quatro trabalhos da literatura [Cubas et al. 2014], [Laudani et al. 2014], [Majdoul et al. 2015] e [Villalva et al. 2009] .

Como o algoritmo genético provê um resultado cada vez que ele é executado, para garantir a confiabilidade na solução obtida com o algoritmo, executou-se o algoritmo 10 vezes. Fazendo a média das respostas fornecidas pelo algoritmo, evita-se que os parâmetros obtidos sejam oriundos de um mínimo local da superfície de erro $RMSE$, por exemplo. Tal como n e R_s, analisou-se as estatísticas do $RMSE$ com o intuito de se mostrar que este converge para um certo valor. Mesmo o maior valor obtido, ou seja, no pior caso, o desempenho do algoritmo genético ainda foi superior ao de outros trabalhos da literatura, como será mostrado adiante. A Tabela 1 apresenta uma análise estatística dos dados.

Tabela 1. Estatísticas dos parâmetros estimados

Parâmetro	Média	Desvio Padrão	Variância	Máximo	Mínimo
n	1,2275	$3{,}4398.10^{-3}$	$11{,}8327.10^{-6}$	1,2333	1,2230
Rs	0,1477	$1{,}0730.10^{-3}$	$1{,}1514.10^{-6}$	0,1493	0,1459
RMSE	$9{,}7369.10^{-2}$	$2{,}9054.10^{-6}$	$8{,}4415.10^{-12}$	$9{,}7374.10^{-2}$	$9{,}7366.10^{-2}$

A Tabela 2 apresenta uma comparação dos valores do erro ($RSME$) e parâmetros do painel fotovoltaico obtidos neste trabalho com os de outros presentes na literatura.

É possível perceber que o $RMSE$ deste trabalho foi o menor registrado até então na literatura. Nota-se, também, que o valor do fator de idealidade do diodo (n) está mais próximo dos valores obtidos pelos trabalhos [Cubas et al. 2014], [Villalva et al. 2009] e [Majdoul et al. 2015], mostrando similaridade de valores, ao contrário do valor de (n) obtido por [Laudani et al. 2014]. Os valores deste trabalho dispostos na Tabela 2 são a médias das 10 vezes que o algoritmo foi executado.

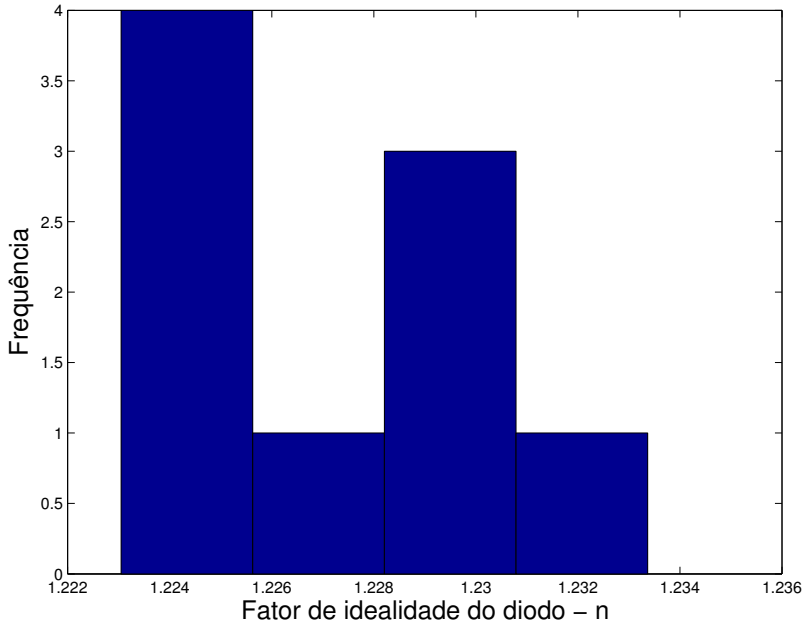
Tabela 2. Comparação dos parâmetros obtidos neste trabalho com outros obtidos na literatura

Trabalhos	RMSE	n	$R_s(\Omega)$	$R_{sh}(\Omega)$	$I_0(\mu A)$	$I_{irr}(A)$
Algoritmo Genético Proposto	9,74 $\cdot 10^{-2}$	1,2275	0,1477	92,2464	32.1221	8,2231
[Laudani et al. 2014]	$9,77 \cdot 10^{-2}$	1,1145	$0,1578$	72,05	4,4279	8,2280
[Cubas et al. 2014]	$25,01 \cdot 10^{-2}$	1,3000	0,2309	594,63	96,930	8,2132
[Villalva et al. 2009]	$21,77 \cdot 10^{-2}$	1,3000	0,2300	566,90	97,561	8,2135
[Majdoul et al. 2015]	$21,17 \cdot 10^{-2}$	1,3000	0,2310	598,00	96,896	8,2100

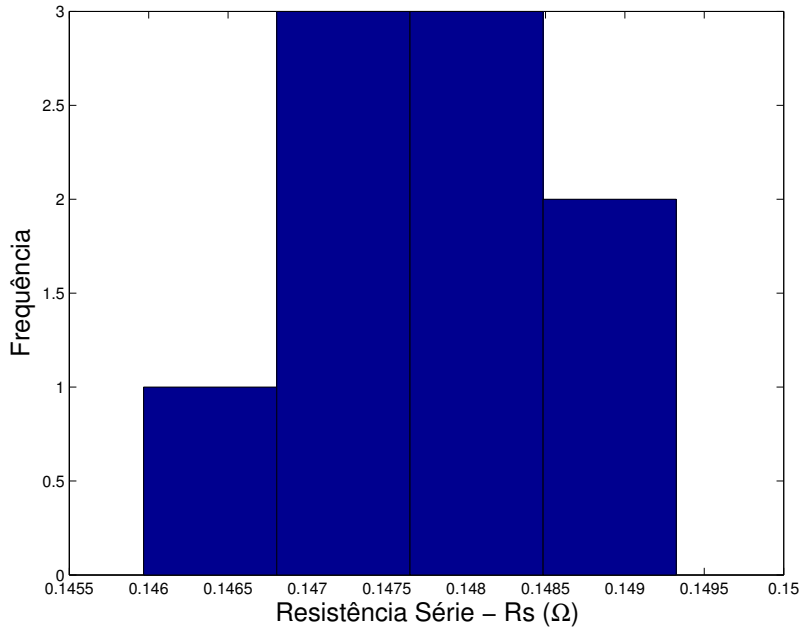
As Figuras 5a e 5b apresentam os histogramas dos valores do fator de idealidade do diodo (n) e da resistência série (R_s) para as 10 execuções do algoritmo. O histograma mostra que os valores obtidos para estes dois parâmetros estão concentrados próximos da média (apresentada na Tabela 2), confirmando que o desvio padrão é realmente pequeno, na ordem de 10^{-3} como foi mostrado na Tabela 1.

Para mostrar a eficiência do algoritmo genético proposto, plotou-se os valores do $RMSE$ a cada iteração, e esta foi comparada com o menor erro obtido da literatura, isto é, o erro foi comparado com o trabalho [Laudani et al. 2014]. Analisando a Figura 6, é possível perceber que o $RMSE$ obtido com a aplicação proposta de algoritmo genético tende a ser menor em todas as gerações. Adotou-se a escala logarítmica nos dois eixos a fim de fornecer melhor visualização do gráfico.

Utilizando os parâmetros obtidos neste trabalho, apresentados na Tabela 2, plotou-se a curva de corrente e tensão utilizando a Equação 1. Essa curva característica é mostrada na Figura 7. Os pontos em círculo foram extraídos do *datasheet*, a curva tracejada é a curva estimada por [Laudani et al. 2014] e a contínua é a curva proposta por este trabalho, obtida por meio da Equação 1.



(a) Histograma do fator de idealidade do diodo (n)



(b) Histograma da resistência série (R_s)

Figura 5. Histogramas do fator de idealidade do diodo (n) e da resistência série (R_s)

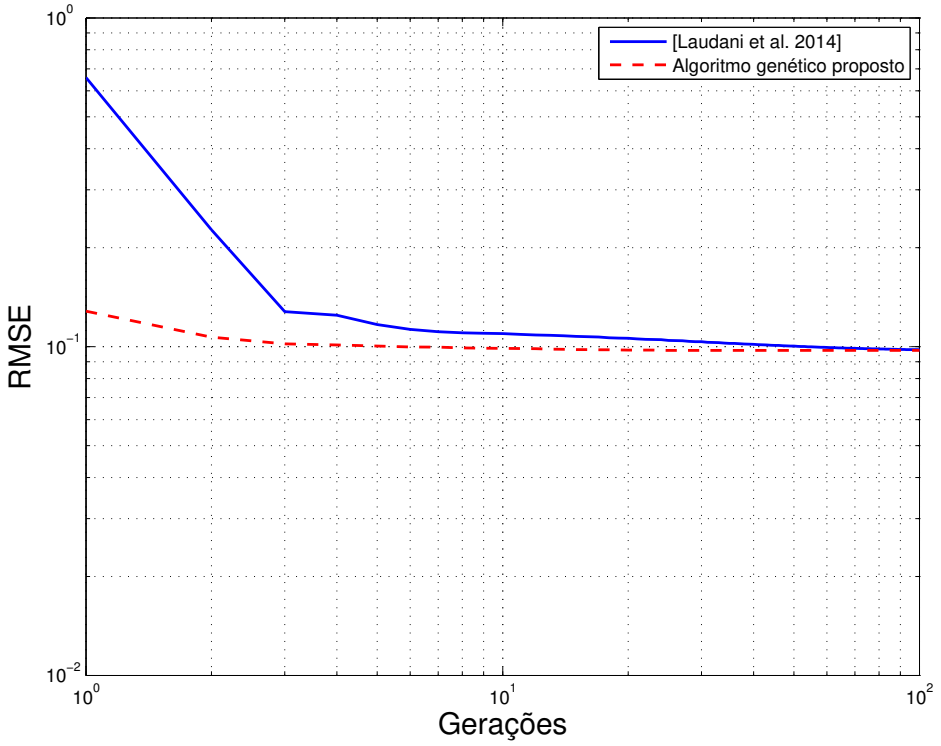


Figura 6. Comparação da convergência do $RMSE$

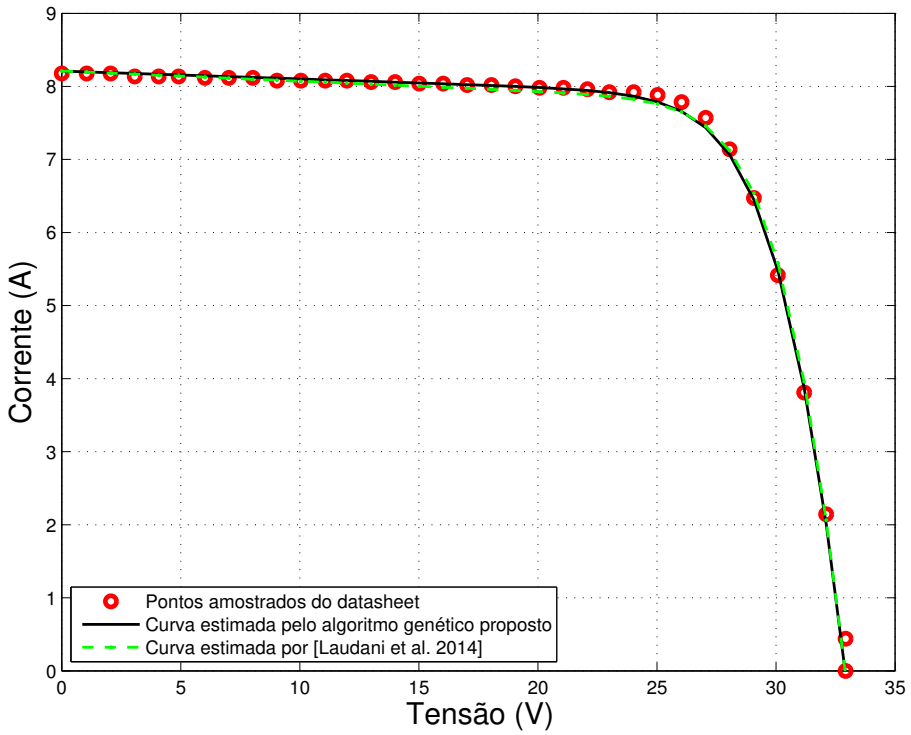


Figura 7. Comparação da Curva calculada de Tensão por Corrente com os pontos amostrados

5. Conclusão

Neste artigo foi proposta a aplicação de um algoritmo genético para realizar a busca dos parâmetros fator de idealidade do diodo (n) e resistência série (R_s) de um painel fotovoltaico. Para avaliar o algoritmo proposto, utilizou-se como métrica a raiz quadrada do erro quadrático médio. Foi mostrado que o desempenho do algoritmo deste trabalho se mostrou superior a outros comparados na literatura, tanto no quesito convergência quanto no valor final.

O algoritmo proposto se mostrou estável e tendendo, na média, para um mesmo valor de erro. Os valores de n e R_S obtidos foram comparados com os de outros trabalhos, e se mostraram coerentes.

Como trabalhos futuros, pretende-se testar outro método de seleção de pais no algoritmo genético, como por exemplo o da roleta viciada. Pretende-se ainda: comparar os resultados do algoritmo genético proposto com os de outros algoritmos evolutivos, como o algoritmo *firefly*; e utilizar outras equações analíticas para o cálculo dos três parâmetros, Gsh, I_0 e I_{irr}.

Referências

Aramizu, J. (2010). *Modelagem e Análise de Desempenho de um Sistema Fotovoltaico em Operação Isolada e em Paralelo com uma Rede de Distribuição de Energia Elétrica*. USP, São Paulo.

Askarzadeh, A. and Rezazadeh, A. (2013). Artificial bee swarm optimization algorithm for parameters identification of solar cell models. *Applied Energy*, 102(Supplement C):943 – 949. Special Issue on Advances in sustainable biofuel production and use - XIX International Symposium on Alcohol Fuels - ISAF.

Cubas, J., Pindado, S., and Victoria, M. (2014). On the analytical approach for modeling photovoltaic systems behavior. *Journal of Power Sources*, 247(Supplement C):467 – 474.

da Costa, W. T. (2010). *Modelagem, estimação de parâmetros e método MPPT para módulos fotovoltaicos*. UFES, Vitória.

de Blas, M., Torres, J., Prieto, E., and Garcı, A. (2002). Selecting a suitable model for characterizing photovoltaic devices. *Renewable Energy*, 25(3):371 – 380.

El-Naggar, K., AlRashidi, M., AlHajri, M., and Al-Othman, A. (2012). Simulated annealing algorithm for photovoltaic parameters identification. *Solar Energy*, 86(1):266 – 274.

Holland, J. H. (1992). *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. The MIT Press.

Laudani, A., Fulginei, F. R., and Salvini, A. (2014). High performing extraction procedure for the one-diode model of a photovoltaic panel from experimental i–v curves by using reduced forms. *Solar Energy*, 103:316 – 326.

Linden, R. (2006). *Algoritmos Genéticos (2a edição)*. BRASPORT.

Majdoul, R., Abdelmounim, E., Aboulfatah, M., Touati, A. W., Moutabir, A., and Abouloifa, A. (2015). Combined analytical and numerical approach to determine the four parameters of the photovoltaic cells models. In *2015 International Conference on Electrical and Information Technologies (ICEIT)*, pages 263–268.

Villalva, M. G., Gazoli, J. R., and Filho, E. R. (2009). Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Transactions on Power Electronics*, 24(5):1198–1208.

Yang, X.-S. (2008). *Nature-Inspired Metaheuristic Algorithms*. Luniver Press.

Modelo de Simulação de Sistema Fotovoltaico Ligado à Rede com Redução da Injeção de Harmônicos via Algoritmo *Fuzzy*

Henrique P. Corrêa[1], Ricardo A.P. Franco[1], Flávio H.T. Vieira[1], Marcelo S. de Castro[1]

[1]Universidade Federal de Goiás
Goiânia – GO – Brazil

Abstract. *In the present work, we propose a computational implementation of a photovoltaic panel simulation model based in a modified version of classical photovoltaic cell equations. The proposed modification includes the effects of temperature variation over the photogenerated current, increasing the precision of the original model. As means of exemplifying the model's application, a comparative analysis is realized between three distinct grid-connected photovoltaic system operational instances: harmonic suppression via LC filtering and through the use of two theoretical fuzzy filters, operating under the logic of Recursive Least Squares and Nearest Neighborhood Clustering algorithms.*

Resumo. *No presente trabalho, propõe-se a implementação computacional de um modelo de simulação de painéis fotovoltaicos baseado em uma versão modificada de um equacionamento clássico da célula fotovoltaica. A modificação proposta inclui o efeito da variação de temperatura sobre a corrente fotogerada, aumentando a precisão do modelo original. Exemplificando-se a aplicação do modelo, é realizada uma análise comparativa da operação de um sistema fotovoltaico em três situações distintas, a saber: com supressão de componentes harmônicas via filtro LC e com filtragem destas por meio de dois filtros ativos teóricos, que operam sobre os sinais de potência em sua entrada via algoritmos Recursive Least Squares e Nearest Neighborhood Clustering.*

1. Introdução

A crescente utilização de sistemas fotovoltaicos no contexto de geração distribuída confere grande relevância ao estudo dos impactos que tais sistemas podem surtir sobre as redes de distribuição de energia elétrica às quais encontram-se conectados. Mais especificamente, tendo-se em vista o caráter não-linear e sazonal da geração fotovoltaica, especial atenção deve ser dispendida à análise da injeção de componentes harmônicas de corrente e modificação nos perfis de tensão na rede, fatores estes que representam avarias à qualidade da energia elétrica suprida, caso não sejam adequadamente controlados.

A simulação computacional de sistemas fotovoltaicos e dos efeitos que sua operação surte sobre a rede de distribuição à qual encontram-se conectados é desejável, tendo-se em vista o fato de permitir a avaliação dos principais parâmetros de qualidade de energia nos barramentos de geração distribuída e, portanto, ser um instrumento útil na realização de estudos preliminares à instalação de sistemas fotovoltaicos.

Neste trabalho, objetiva-se avaliar o impacto sobre a qualidade de energia em uma rede devido à inserção de um painel fotovoltaico na mesma, propondo-se algumas ferramentas para redução da distorção harmônica total (*DHT*). Para tal, é inicialmente proposto

um modelo de simulação de painel fotovoltaico, o qual leva em consideração os principais parâmetros que influenciam as características I-V do painel, sendo também capaz de simular transitórios e aferir parâmetros de qualidade de energia na barra de geração fotovoltaica. O modelo proposto, baseado em [1], introduz uma modificação de equacionamento visando a eliminação de aproximações computacionais e um consequente aumento de precisão.

Apresenta-se também, neste artigo, uma análise comparativa entre diferentes métodos de supressão de harmônicos na tensão entregue pelo sistema fotovoltaico. O mais básico consiste de um filtro de segunda ordem projetado tendo-se em vista a atenuação de harmônicos concomitantemente à manutenção da fundamental. Em seguida, avaliou-se o desempenho de redução de harmônicos com um filtro adaptativo utilizando algoritmo *RLS (Recursive Least Squares)*. O último método de filtragem considerado consiste no uso de um controlador *fuzzy*, capaz de processar sinais de potência, previamente treinado para a eliminação de ruído superposto ao sinal original.

A distorção harmônica provocada por sistemas fotovoltaicos é um tópico analisado em vários trabalhos da literatura. Dentre eles, [2] propõe a inserção de filtros passivos de sintonização simples na rede (*single tuned passive filters*) para mitigar os harmônicos injetados na rede. Em [3], são expostas as limitações do método controle de corrente ressonante (*resonant current control*), sendo proposto um projeto de controle com o objetivo de aprimorá-lo. É proposto, em [4], uma configuração de sistemas fotovoltaicos que utiliza filtro harmônico de dupla-sintonização (*double-tuned harmonic filter*) para a manutenção da qualidade de energia. Em [5], é proposto um filtro de potência *shunt* ativo (*shunt active power filter*) para reduzir-se as distorções harmônicas geradas por circuito inversor e cargas não-lineares. O trabalho [6] analisa um filtro passa-baixas entre inversor e rede elétrica para a redução da distorção harmônica. Um dos intuitos do presente trabalho é contribuir, neste contexto, via análise comparativa do uso de filtragem passiva com filtro LC e por meio de filtros ativos que processam os sinais de potência, como filtro *RLS* e via lógica *fuzzy*.

2. O Modelo de Painel Fotovoltaico

O elemento fundamental que compõe os painéis utilizados em geração distribuída é a célula fotovoltaica, a qual é modelada pelo circuito equivalente ilustrado na Fig. 1. Neste trabalho, adotamos a aproximação $R_{sh} \to \infty$, válida para células de alta qualidade [7]. O equacionamento deste circuito com essa aproximação leva à característica I-V dada por:

$$V = mV_t \cdot ln\left(\frac{I_{ph} + I_o - I}{I_o}\right) - R_s I \tag{1}$$

Os parâmetros utilizados em (1) são apresentados, assim como seus valores, na Tabela I. Parâmetros dependentes da irradiância (S) e temperatura da célula (T) encontram-se expressos para a condição de referência *Standard Temperature Condition (STC)*, $T_o = 25°C$ e $S_o = 1000W/m^2$. Os valores utilizados referem-se à uma célula fotovoltaica do módulo *MSX-60*, tendo sido estimados segundo informações de *datasheet* [8] e o método dado em [9].

Os parâmetros dependentes de T e S são I_{ph}, I_o e V_t. Logo, a variação de temperatura da célula e irradiância promovem alterações na característica I-V. Em [1], é proposto

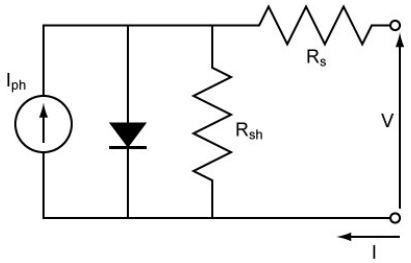


Figura 1. Circuito equivalente da célula fotovoltaica.

Tabela 1. PARÂMETROS DA CÉLULA FOTOVOLTAICA

Parâmetro	Descrição	Valor Simulado
m	Fator de idealidade	$1,5$
V_t	Tensão térmica	$26mV$
I_o	Corrente reversa de saturação	$1,02 \cdot 10^{-6}A$
I_{ph}	Corrente fotogerada	$3,8A$
R_s	Resistência série	$0,003\Omega$

um método de modificação da característica, em função de T e S, por meio de fatores corretivos. Em outras palavras, os parâmetros da Tabela 1, com exceção de I_{ph}, são tomados como fixos e o efeito da variação de T e S sobre a característica é calculado via (2) e (3), em que $V^*(T_o, S_o) = V(T_o, S_o, I_{ph}(T, S))$:

$$\frac{I_{ph}(T,S)}{I_{ph}(T_o,S_o)} = \left[1 + \gamma_T(T - T_o)\right] \frac{S}{S_o} \tag{2}$$

$$\frac{V(T,S)}{V^*(T_o,S_o)} = \left[1 + \beta_T(T - T_o)\right] \left[1 + \beta_T \alpha_S (S - S_o)\right] \tag{3}$$

Os fatores corretivos de (2) e (3) tem seus valores, extraídos do *datasheet* do módulo *MSX-60*, relacionados na Tabela II. Em particular, o parâmetro α_S não é dado explicitamente, mas pode ser calculado por meio da *Nominal Operating Cell Temperature (NOTC)* fornecida no *datasheet* por meio de (4):

$$\alpha_S = \frac{NOTC(^\circ C) - 20}{S_o(W/m^2) - 800} \tag{4}$$

Entretanto, em [1], T (e sua respectiva referência T_o) é tomado como a temperatura ambiente. O coeficiente α_S é utilizado com o intuito de considerar o aquecimento da célula fotovoltaica e sua consequente operação em temperatura superior à ambiente. Vê-se que o efeito deste aquecimento é desconsiderado em (2). Sendo o aquecimento $\Delta T = \alpha_S(S - S_o)$, surge, em (3), o termo espúrio ${\beta_T}^2(T - T_o)\Delta T$.

Este trabalho propõe, para que tais imprecisões sejam evitadas, que T seja considerado como a temperatura da célula. Esta temperatura pode ser facilmente deduzida da temperatura ambiente T_a por meio do coeficiente α_S, pois $T = T_a + \Delta T = T_a + \alpha_S(S - S_o)$. Logo, tendo o aquecimento sido contabilizado previamente, faz-se

$\alpha_S = 0$ em (3). A Fig. 2 ilustra a implementação computacional, via *Simulink*, de (1) a (4), considerando-se a correção proposta ao modelo.

Tabela 2. COEFICIENTES DE TEMPERATURA

Parâmetro	Descrição	Valor Simulado
α_S	Coeficiente de aquecimento	$0,13°C \cdot m^2/W$
β_T	Coeficiente de temperatura de V_{oc}	$-0,38\%/°C$
γ_T	Coeficiente de temperatura de I_{sc}	$0,065\%/°C$

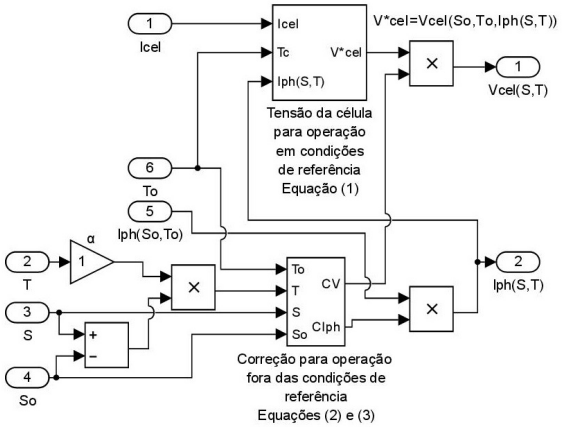


Figura 2. Implementação da célula fotovoltaica.

A obtenção da característica I-V do painel fotovoltaico decorre do fato deste ser uma associação paralela entre N_p *strings* de N_s células fotovoltaicas cada, o que implica em $V = N_s \cdot V_{cel}$ e $I = N_p \cdot I_{cel}$, onde V_{cel} e I_{cel} são os valores de tensão e corrente terminais calculados via (1) a (3). A implementação do painel fotovoltaico, realizada via *Simulink*, encontra-se na Fig. 3.

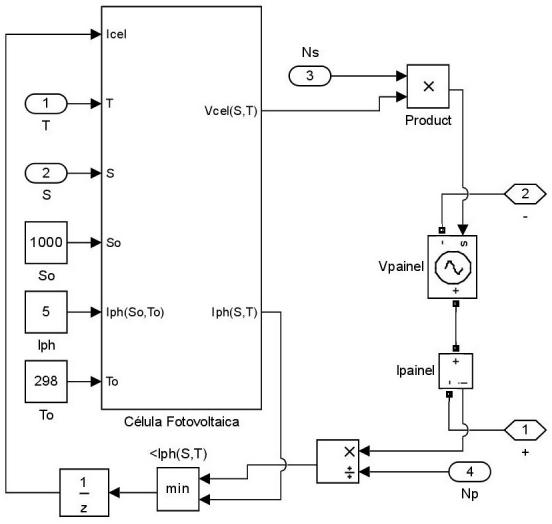


Figura 3. Implementação do painel fotovoltaico.

3. Projeto Elementar e Simulação de um Sistema Fotovoltaico Conectado à Rede

De posse do modelo proposto para o painel fotovoltaico, este foi utilizado conjuntamente à biblioteca *SimPowerSystems* com o intuito de simular-se um sistema fotovoltaico. Este é projetado via considerações teóricas, de modo a ser possível validarem-se os resultados obtidos via simulação. Delineamos, nesta seção, o projeto básico do sistema. Na subseção 3.4, os valores dos parâmetros adotados/projetados são sumarizados.

O sistema hipotético considerado consiste em um painel composto por módulos fotovoltaicos *MSX-60*, distribuídos em 72 *strings* de 12 módulos associadas em paralelo ($N_s = 432$ e $N_p = 12$), conectado em cascata com um conversor *boost* e um inversor trifásico. O conversor *boost* tem seu *duty cycle* controlado por um algoritmo de *Maximum Power Point Tracking (MPPT)*. Um filtro LC é utilizado na saída do sistema para a atenuação de harmônicos gerados pelo inversor.

Toma-se como condição representativa de operação do sistema a *Nominal Operation Cell Temperature (NOCT)* fornecida no *datasheet* do *MSX-60*, $T = 47°C$ e $S = 800W/m^2$. Para essas condições, calcula-se, com auxílio de (1) a (3), $V_{mp,painel} \approx 182,8V$ e $I_{mp,painel} \approx 204,5A$. Logo, em condições usuais, a potência máxima é $P_{mp,painel} \approx 37,4kW$.

Supõe-se que o sistema seja instalado no barramento de uma carga residencial de valor nominal $25kVA$, com fator de potência $cos\theta = 0,92$ e tensão nominal de $220V$. A rede de distribuição considerada é representada por seu equivalente de Thévenin referido à base da carga. A configuração geral do sistema é ilustrada na Fig. 4.

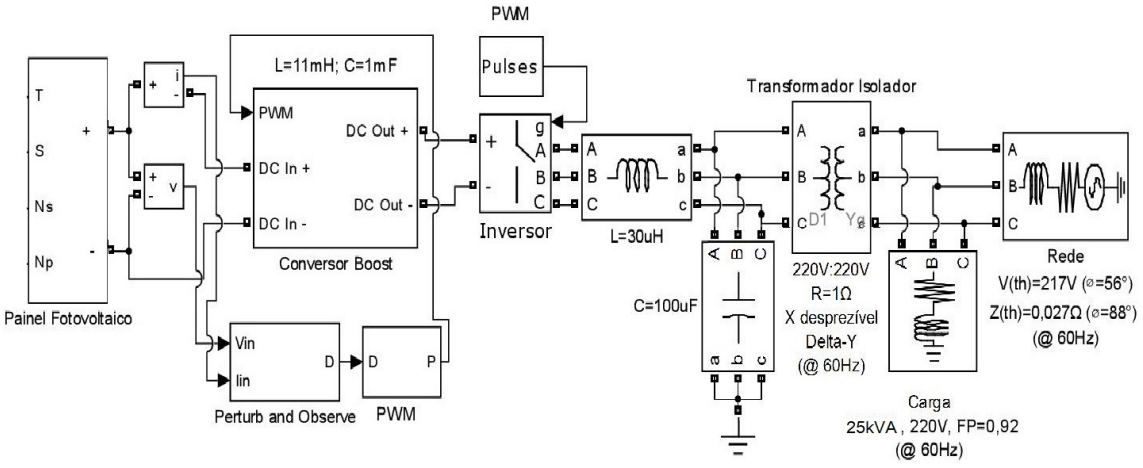


Figura 4. Esquema do sistema simulado.

3.1. Conversor *Boost* e o Algoritmo *Perturb & Observe (P&O)*

O conversor *boost* é um circuito que transforma tensão CC de entrada em um nível CC superior em sua saída. Há uma dependência definida entre o *duty cycle* D do sinal de controle e a razão de transformação $M(D)$ [10].

Para cada $M(D)$ e um mesmo circuito de saída, o painel fotovoltaico perceberá um carregamento diferente. Se D for ajustado de modo que o carregamento sempre corresponda ao máximo de potência da característica I-V do painel, é garantida a extração

ótima de potência dos terminais do painel fotovoltaico. O algoritmo de controle simulado é o algoritmo $P\&O$, com o incremento $\Delta D = \pm 0,1\%$ [11].

Os elementos reativos do conversor foram projetados de acordo com (5) e (6), visando-se a obtenção de ondulações de tensão e corrente iguais a, respectivamente, $5,0V$ e $1,1A$ para $D = 0,65$. As frequências de chaveamento do conversor e de amostragem do algoritmo $P\&O$ são iguais a $f = 10kHz$.

$$\Delta I_L = \frac{V_{in} \cdot D}{L \cdot f} \tag{5}$$

$$\Delta V_{out} = \frac{I_{out} \cdot D}{C \cdot f} \tag{6}$$

3.2. Inversor Trifásico

Considera-se que o inversor trifásico é do tipo seis pulsos, sendo controlado via modulação de largura de pulso (PWM). Opta-se pela utilização da frequência de portadora $f_s = 6,06kHz$, o que implica uma modulação em frequência $m_f = 101$. A escolha de m_f ímpar garante redução de picos de amplitude no conteúdo harmônico gerado e previne a injeção de subharmônicos no espectro da tensão de saída [10].

Como a saída do inversor é conectada em paralelo à rede, é necessário que seu nível de tensão CA de saída seja adequadamente projetado para que seja da mesma ordem de grandeza da rede. Sabe-se que [10], para um inversor trifásico controlado via *PWM* e operando com modulação em amplitude $0 < m_a < 1$, a tensão suprida é dada por (7):

$$V_{LL,1}^{inv} \approx 0,612 \cdot m_a V_d = 0,612 \cdot m_a \cdot \frac{V_{cc}}{1 - D} \tag{7}$$

em que V_d é a tensão de saída do conversor *boost* e V_{cc} é a tensão suprida pelo painel fotovoltaico. Para operação na *NOTC*, $V_{cc} \approx 180V$. Considerando que a simulação foi realizada para uma mesma condição ambiental e de rede, adotou-se o procedimento de estimar-se o *duty cycle* do conversor *boost* com $m_a = 1$ e escolher V_{cc} adequadamente. Para cenários variáveis, estratégias de controle de tensão devem ser consideradas, o que será implementado em trabalhos futuros.

Para prever-se D em regime, é necessário o conhecimento da impedância de entrada $|Z_{in}|$ do inversor. O procedimento adotado para o aferimento desta foi uma simulação acessória, na qual conectou-se uma fonte de tensão DC ideal aos terminais do inversor e registrou-se o valor médio da corrente fluindo nos terminais da fonte. Desse modo, $|Z_{in}|$ pode ser calculado pela razão entre tensão e corrente média de fonte. Por meio desse procedimento, obteve-se $|Z_{in}| \approx 7,10\Omega$.

Como $|Z_{in}| \approx 7,10\Omega$ e o algoritmo *MPPT* atua no sentido de fazer $|Z_{in}| = M^2(D) \cdot \frac{V_{mp,painel}}{I_{mp,painel}}$, obtém-se $D \approx 0,65$. Considerando-se que $V_{cc} \approx 180V$, obtém-se, via (8), a tensão de saída $V_{LL,1}^{inv} \approx 313,9V$.

3.3. Filtro LC

Em inversores trifásicos controlados via *PWM*, a distorção harmônica revela-se predominante em torno dos harmônicos de ordem $k \cdot m_f$, com k inteiro. Como $m_f = 101$, harmônicos de tensão fornecidos pelo inversor terão amplitudes significativas a partir da vizinhança de $6,06kHz$. Logo, estipula-se uma frequência de corte $f_o = 3,0kHz$. Ademais, é necessário verificar se a resistência R do transformador isolador promove um pico de ressonância demasiadamente elevado na vizinhança de f_o, o que poderia aumentar a *DHT*. A teoria clássica de filtros fornece as seguintes equações para o projeto de um filtro RLC passa-baixas, cuja topologia corresponde ao circuito em questão [12]:

$$f_o = \frac{1}{2\pi\sqrt{LC}} \tag{8}$$

$$\zeta = \frac{1}{4\pi RC f_o} \tag{9}$$

Utilizou-se (8) para o projeto de um filtro LC com a especificação de f_o dada e, via (9), verifica-se $\zeta = 0,273$, o que indica um pico de ressonância com magnitude $\approx 1,9dB$ na frequência de corte. Os valores do filtro LC obtidos encontram-se Fig. 4.

3.4. Síntese dos Parâmetros de Simulação

Na Tabela 3, apresentam-se os principais parâmetros para simulação do sistema fotovoltaico, calculados de acordo com os procedimentos previamente indicados.

Tabela 3. Parâmetros do Sistema Fotovoltaico

Parâmetro	Valor
N_s	432
N_p	12
S	$800W/m^2$
T	$47°C$
L	$30\mu H$
C	$100\mu F$
L_{boost}	$11mH$
C_{boost}	$1mF$
m_a	1
$P_{mp,painel}$	$37,4kW$
ΔD	$0,1\%$

4. Mitigação de Harmônicos Gerados Pelo Sistema Fotovoltaico via Controlador Fuzzy

Com objetivo de se avaliar a mitigação de harmônicos gerados pelo sistema fotovoltaico utilizando filtros ativos, dois algoritmos de filtragem adaptativa foram implementados, *NNC (Nearest Neighborhood Clustering)* e *RLS (Recursive Least Squares)*. A implementação *RLS* utilizada é aquela descrita em [15]. Na subseção IV-A, descreve-se o modelo *NNC* utilizado.

4.1. Algoritmo *Nearest Neighborhood Clustering*

A lógica *fuzzy* é uma ferramenta capaz de processar informações vagas, usualmente descritas em linguagem natural, convertendo-as em formas quantitativas, passíveis de processamento por sistemas computacionais [13]. O objetivo da lógica fuzzy é gerar saídas lógicas a partir de um conjunto de entradas imprecisas e ruidosas [9; 10].

O sistema *fuzzy* utilizado consiste na obtenção de um modelo que combine N pares entrada-saída $(x_0^l; y_0^l)$, com $l = 1, 2, ..., N$, com uma devida tolerância $\epsilon > 0$, isto é, $|f(x_0^l) - y_0^l| < \epsilon$ [15]. Logo, o sistema implementado neste trabalho é representado por (10):

$$f(x) = \frac{\sum_{l=1}^{N} y_0^l \cdot exp\left(-\frac{|x-x_0^l|^2}{\sigma^2}\right)}{\sum_{l=1}^{N} exp\left(-\frac{|x-x_0^l|^2}{\sigma^2}\right)}, \tag{10}$$

onde y_0^l corresponde às saídas de uma regra l, x representa os dados de entrada, x_0^l são os centros do modelo, σ é o desvio padrão e N determina a quantidade de regras *fuzzy*.

Os parâmetros do sistema representado por (10) podem ser encontrados utilizando técnicas de *clustering*. Tais técnicas classificam um conjunto de dados em subconjuntos (*clusters*) de acordo com suas propriedades [15].

A técnica de *clustering* utilizada neste trabalho foi a *Nearest Neighborhood Clustering*. Nesta técnica, toma-se como centro do primeiro *cluster* o primeiro dado de entrada. Em seguida, analisa-se se a distância do próximo dado ao centro do *cluster* é menor do que um valor predefinido (r); caso a distância seja inferior a r, insere-se o dado neste *cluster*; senão, cria-se um novo *cluster* com este dado como seu centro. O algoritmo é apresentado a seguir [15]:

- Passo 1: Com o primeiro par de entrada e saída $(x_0^l; y_0^l)$, define-se o centro do *cluster* $x_c^l = x_0^l$ e atribui-se $A^l(l) = y_0^l$ e $B^l(l) = 1$. Finalmente, define-se o valor do raio r.
- Passo 2: Considera-se o *k-ésimo* par entrada-saída $(x_0^k; y_0^k)$ e M *clusters* com centros x_c^k, $k = 1, 2, ..., M$. Calculam-se as distâncias de x_0^k aos M centros dos *clusters*, $|x_0^k - x_c^l|$, $l = 1, 2, ..., M$. Seja a menor distância correspondente a $x_c^{l_k}$. Logo:
  a) Se $|x_0^k - x_c^{l_k}| > r$, define-se novo *cluster* com centro $x_c^{M+1} = x_0^k$. Realizam-se as atribuições $A^{M+1}(k) = y_0^k$, $B^{M+1}(k) = l$ e, para $l = 1, 2, ..., M$, $A^l(k) = A^l(k-1)$ e $B^l(k) = B^l(k-1)$.
  b) Se $|x_0^k - x_c^{l_k}| < r$, faz-se, para $l = 1, 2, ..., M$ com $l \neq l_k$, $A^l(k) = A^l(k-1)$ e $B^l(k) = B^l(k-1)$. Efetuam-se as seguintes atribuições:

$$A^{l_k}(k) = A^{l_k}(k-1) + y_0^k \tag{11}$$

$$B^{l_k}(k) = B^{l_k}(k-1) + 1 \tag{12}$$

- Passo 3: Se x_0^k não for considerado um novo *cluster*, então o sistema *fuzzy* é definido pelos k pares de entrada e saída $(x_0^j; y_0^j)$, $j = 1, 2, ..., k$, sendo descrito

por (13):

$$f_k(x) = \frac{\sum_{l=1}^{M} A^l(k) \cdot exp\left(-\frac{|x-x_c^l|^2}{\sigma}\right)}{\sum_{l=1}^{M} B^l(k) \cdot exp\left(-\frac{|x-x_c^l|^2}{\sigma}\right)} \tag{13}$$

Por outro lado, se x_0^k é considerado um novo *cluster*, então o sistema é definido por (13), com os somatórios variando entre 1 e $M+1$, incluindo-se o *cluster* x_0^k.

- Passo 4: Retorna-se ao Passo 2 para $k = k + 1$.

Nota-se que $B^l(k)$ é igual ao número de pares entrada-saída e $A^l(k)$ corresponde ao somatório de todas as saídas do par entrada-saída associados aos *clusters* atuais. Logo, quando todos os pares entrada-saída forem associadas a centros de *cluster*, o sistema representado por (13) será idêntico ao sistema *fuzzy* ótimo (10). Com a possibilidade de criação de novos *clusters*, é possível que o número de regras *fuzzy* também seja alterado durante o processo de *clustering*, visto que o número de *clusters* depende da distribuição dos dados e de r. A complexidade do sistema é de ordem $O(r)$ [15].

4.2. Redução de THD Utilizando Algoritmo Fuzzy

Assume-se que o sistema *fuzzy* age como um filtro de impedâncias de entrada e saída ideais. Desse modo, a simulação computacional do filtro é realizada por meio do seguinte procedimento: o sistema é inicialmente simulado isento de qualquer tipo de filtro, sendo a tensão da fase $v_a(t)$ registrada e, subsequentemente, tratada pelo algoritmo *fuzzy*. A tensão de saída $v_{a,fz}(t)$ obtida é utilizada como base para uma nova simulação, na qual uma fonte de tensão trifásica balanceada, cuja tensão da fase a é igual a $v_{a,fz}(t)$, alimenta diretamente o transformador isolador.

A entrada do sistema *fuzzy* é a tensão de fase $v_a(t)$ para operação sem filtro. O sistema *fuzzy* utiliza-se de um sinal de referência senoidal. Este sinal é usado pelo sistema para modelar-se um controlador $\mu(k)$ que faz a saída do sistema seguir a saída definida pelo sinal de referência. O controlador é dado por (14) [15]:

$$\mu(k) = -f_k[y(k), y(k-1)] + r(k) \tag{14}$$

Em (14), $f_k[y(k), y(k-1)]$ é o sistema *fuzzy*, representado por (10) e (13), e $r(k)$ é o sinal de referência. O comportamento do sistema *fuzzy* é definido por (15):

$$y_{fuzzy}(k) = y(k-1) + \mu(k) \tag{15}$$

Em (15), $y_{fuzzy}(k)$ é a saída do sistema *fuzzy*, que é definido pela amostra anterior da tensão de entrada $y(k-1)$ e pelo controlador *fuzzy*. Vê-se, portanto, que o sistema *fuzzy* tende a, iterativamente, corrigir o sinal de entrada por meio de seu controlador, eliminando o ruído relativamente a $r(t)$. O sinal de referência é dado por (16):

$$y_{ref}(t) = \sqrt{2} V_{ph,1}^{inv} \cdot cos(2\pi f_s t) \tag{16}$$

Em (16), $f_s = 60Hz$ e $V_{ph,1}^{inv}$ é a fundamental da tensão de fase suprida pelo inversor, aferida para operação do sistema na ausência de filtros. Por análise análoga àquela realizada em III-B, obtém-se $D = 0,59$ e $V_{LL,1}^{inv} \approx 268,7V$. Para o mesmo conversor *boost*, obtém-se, via (5) e (6), $\Delta I_L \approx 0,9A$ e $\Delta V_{out} \approx 5,0V$.

5. Resultados de Simulação

5.1. Modelo de Painel Fotovoltaico

Parâmetros pertinentes na característica I-V de um painel fotovoltaico são tensão e corrente terminais nas situações de operação em curto-circuito, circuito aberto e máxima potência. Não obstante, a potência suprida na condição de máxima transferência de potência é também relevante.

A Tabela 4 mostra a comparação entre os valores calculados por meio do *datasheet* do fabricante, obtidos via simulação com o modelo proposto e via simulação sem a correção de modelo proposta. Esses dados revelam que a correção proposta aumenta a acurácia do modelo original. Os valores são dados para uma célula individual do painel. A corrente de curto-circuito não é dada por ser um dado de entrada do modelo.

Tabela 4. Parâmetros da Característica I-V (*NOTC*)

Parâmetro	Teórico	Simulado	Simulado (original)
V_{oc}	$0,5265V$	$0,5265V$	$0,5173V$
V_{mp}	$0,4352V$	$0,4274V$	$0,4202V$
I_{mp}	$2,8400A$	$2,8441A$	$2,8025A$
P_{mp}	$1,2286W$	$1,2168W$	$1,1788W$

5.2. Filtro LC

O sistema fotovoltaico projetado na Seção III foi simulado com auxílio do modelo proposto. Na Tabela 5, apresenta-se a comparação entre valores de parâmetros que foram previamente projetados e, posteriormente, obtidos na simulação. O parâmetro THD_V é a distorção harmônica total verificada na tensão de linha no ponto de acoplamento comum (*PAC*).

Tabela 5. PARÂMETROS AFERIDOS - FILTRO LC

Parâmetro	Teórico	Simulado
ΔI_L	$1,1A$	$0,8A$
ΔV_o	$5,0V$	$7,0V$
D	$0,65$	$0,67$
$V_{LL,1}^{inv}$	$313,9V$	$297,0V$
THD_V	$-$	2%

Na Fig. 5, encontra-se o espectro da tensão de linha no *PAC*, para operação do sistema fotovoltaico na ausência de filtros. Evidentemente, a *DHT* seria demasiadamente elevada.

Deve-se notar que as equações utilizadas para o projeto do conversor *boost* são derivadas de suposições de operação ideal do conversor (e.g. ausência de impedância interna na fonte DC) [10]. Isto justifica o desvio existente entre valores projetados e simulados de *ripple* e tensão de saída do inversor.

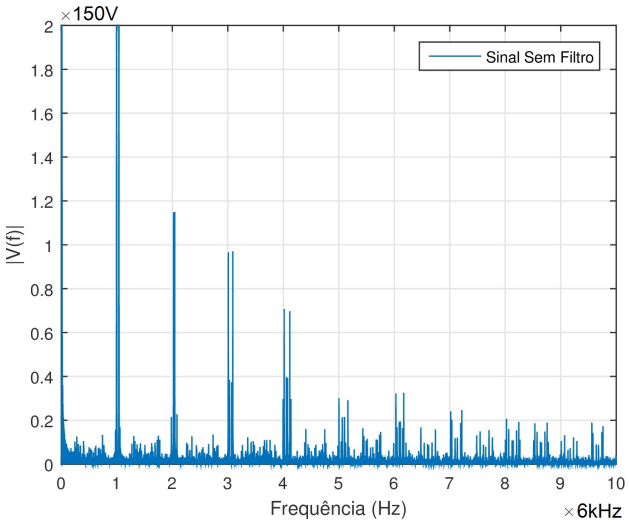


Figura 5. Espectro da tensão de linha - Sem filtros.

A literatura sugere [10] que a *DHT* da tensão de saída de um inversor *PWM* com elevada modulação em frequência, após filtragem apropriada dos harmônicos dominantes, não ultrapassa 5%. O resultado obtido com o filtro LC mostra-se coerente com esta asserção.

5.3. Filtragem com Sistema Fuzzy

Aplicou-se o algoritmo *fuzzy* descrito na seção IV para controlar-se o sinal produzido pelo sistema fotovoltaico conectado, de forma a mitigar os harmônicos na rede. Considerou-se como valores de entrada para o treinamento do algoritmo *fuzzy*, dados referentes a dois ciclos da frequência fundamental do sinal produzido pelo sistema fotovoltaico em regime permanente. Após o treinamento do sistema *fuzzy*, este foi posto a atuar como filtro sobre a forma de onda de entrada.

Na Tabela 6, encontra-se a comparação entre valores teóricos e simulados, dos mesmos parâmetros analisados na seção anterior, para a operação do sistema com o controlador *fuzzy*.

Novamente, a discrepância entre os valores calculados de *ripple* e tensão de linha do inversor apresentam certa divergência com relação àqueles obtidos na simulação devido à condição de idealidade da fonte DC (neste caso, o painel fotovoltaico) não ser satisfeita.

Tabela 6. PARÂMETROS AFERIDOS - FILTRO ATIVO

Parâmetro	Teórico	Simulado
ΔI_L	$0,9A$	$0,7A$
ΔV_o	$5,0V$	$6,0V$
D	$0,59$	$0,58$
$V_{LL,1}^{inv}$	$267,9V$	$234,7V$
THD_V	$-$	$0,035\%(NNC)/0,047\%(RLS)$

Os resultados revelam que a implementação física de um filtro ativo da natureza proposta neste trabalho seria capaz, a depender da natureza do algoritmo *fuzzy* utilizado, de propiciar grandes reduções na injeção de componentes harmônicas na rede de distribuição; de fato, a forma de onda da tensão de linha revelou-se um sinal muito próximo de uma senóide pura, para operação com os filtros *fuzzy*.

O melhor resultado obtido com *fuzzy* pode ser justificado considerando-se que, enquanto a atuação do filtro LC promove a mitigação de frequências $f > f_o$, o controlador *fuzzy*, quando suficientemente robusto, força o sinal de entrada a acompanhar um sinal senoidal de referência, também atenuando fortemente harmônicos de ordem menor ou próxima a f_o. O espectro da tensão no PAC encontra-se nas Figs. 6 e 7, respectivamente, para os filtros *RLS* e *NNC*.

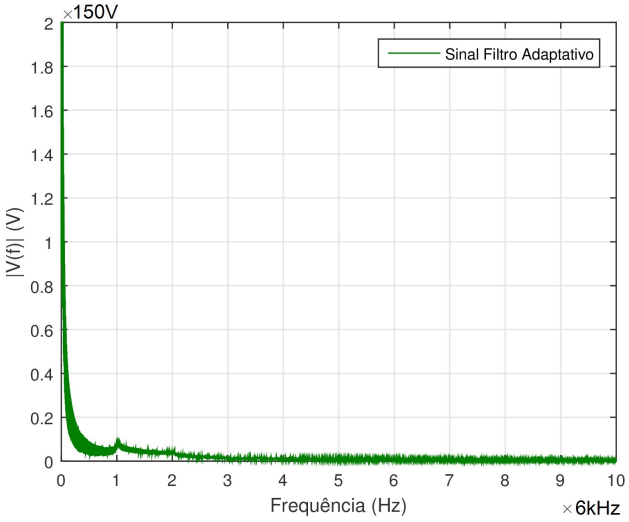


Figura 6. Espectro da tensão de linha - Filtro *NNC*.

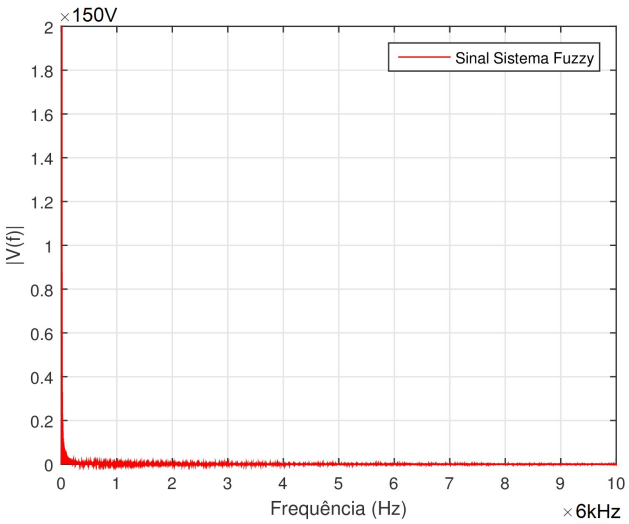


Figura 7. Espectro da tensão de linha - Filtro *RLS*.

6. Conclusão

Neste trabalho, realizou-se uma contribuição à simulação de painéis fotovoltaicos por meio da proposta de um modelo de simulação baseado em uma modificação do equacionamento clássico desenvolvido em [1]. A modificação realizada consistiu na consideração do efeito, sobre a corrente fotogerada, surtido pelo aquecimento da célula com relação à temperatura ambiente. No modelo original, este efeito era desprezado. Verificou-se, por meio de simulações, que a proposta realizada efetivamente aumentou a precisão do modelo original.

Foram realizadas simulações de um sistema fotovoltaico conectado à rede operando em condições distintas de mitigação de harmônicos, isto é, via filtro LC e filtragem ativa com controlador *fuzzy*. Constatou-se que, na ocasião da implementação de um filtro *fuzzy*, é possível reduzir o conteúdo harmônico injetado na rede substancialmente. Para filtragem *fuzzy*, a *DHT* da tensão de linha no barramento local revelou-se duas ordens de grandeza menor do que aquela verificada com a utilização do filtro LC passivo.

7. Referências

1. BURESCH, M. Photovoltaic Energy Systems: Design and Installation. McGraw-Hill, 1983.

2. EL-SAYED, M. M., EL-ELA, A. A. A., EL-SEHIEMY, R. A., Effect of Photovoltaic System on Power Quality in Electrical Distribution Networks, Eighteenth International Middle East Power Systems Conference (MEPCON), 2016.

3. CASTILLA, M., MIRET, J., CAMACHO, A., MATAS, J., VICUÑA, L. G., Reduction of Current Harmonic Distortion in Three-Phase Grid-Connected Photovoltaic Inverters via Resonant Current Control. IEEE Transactions on Industrial Electronics, vol. 60, no 4, april 2013.

4. DAS, S., SADHU, P. K., SHRIVASTAV, A. K., Synchronization and Harmonic Reduction of a Grid-Connected Photovoltaic Generation System. International Conference on Energy, Power and Environment: Towards Sustainable Growth (ICEPE), 2015.

5. MOFTAH, M. A., EL-SAADY, G. IBRAHIN, E. N. A., Active Power Filter for Power Quality Enhancement of Photovoltaic Renewable Energy Systems. Saudi Arabia Smart Grid (SASG), 2016.

6. ADEKOL, O. I., ALMAKTOOF, A. M., RAJI, A. K., Design of a Smart Inverter System for Photovoltaic System Application. International Conference on the Industrial and Commercial Use of Energy (ICUE), pp 310-317, 2016.

7. ISABELLA, O., JAGER, K., SMETS, A., SWAAIJ, R., ZEMAN, M., Solar Energy: The Physics and Engineering of Photovoltaic Conversion, Technologies and Systems. UIT Cambridge Ltd, 2016.

8. *MSX-60* and *MSX-64* Photovoltaic Modules. Disponível em ¡https://www.solarelectricsupply.com/media/¿. Último acesso em 03/2017.

9. FRANCO, R. A. P., VIEIRA, F. H. T., CASTRO, M. S., Estimação de Parâmetros para Modelagem de Painéis Fotovoltaicos Utilizando Algoritmo de Levemberg-Marquardt Modificado. VI Congresso Brasileiro de Energia Solar, 2016.

10. MOHAN, N., UNDELAND, T.M., ROBBINS, W.P., Power Electronics: Converters, Applications and Design. John Wiley and Sons, 2003.

11. MORALEZ, D.S., Maximum Power Point Tracking Algorithms for Photovoltaic Applications. Tese de mestrado, Aalto University.

12. KUO, F.F., Network Analysis and Synthesis. John Wiley and Sons, 1962.

13. KLIR, G. J., YUAN B.. Fuzzy sets and fuzzy logic, theory and applications. Prentice-Hall Press, USA, 2008.

14. ZADEH, L. A., Fuzzy Sets. Information and Control, v. 8, n. 3 (Jun), pp. 338-353, 1965.

15. WANG, L. X., A course in fuzzy systems. Prentice-Hall Press, USA, 1997.

Otimização da Alocação de Recursos em Redes Sem Fio com Agregação de Portadora e F-OFDM Utilizando Programação Evolucionária

Marcus Vinícius Gonzaga Ferreira, Flávio Henrique Teles Vieira

Instituto de Informática – Universidade Federal de Goiás
30332-0250, Goiânia – GO – Brasil

marcusviniciusbr@gmail.com, flaviohtv@gmail.com

Resumo. *Neste artigo, propomos um esquema de alocação de recursos em redes sem fio com o objetivo de maximizar a vazão total do sistema e atender certos parâmetros de QoS (Quality of Service, Qualidade de Serviço) usando EP (Evolutionary Programming, Programação Evolucionária). O desempenho do algoritmo de alocação de recursos é verificado e comparado com outros da literatura utilizando simulações computacionais. Nestas simulações, consideramos também técnicas atuais e candidatas às redes da próxima geração como f-OFDM (filtered-Orthogonal Frequency Division Multiplexing, Multiplexação por Divisão de Frequência Ortogonal filtrado) e agregação de portadora com intuito de mostrar que é possível prover maior taxa de transmissão, melhor utilização eficiente do espectro, melhor cobertura e menor latência em relação às redes 4G.*

Abstract. *In this paper, we propose a resource allocation scheme in wireless networks that aims to maximize the total data rate and attain certain QoS (Quality of Service) parameters using EP (Evolutionary Programming) heuristic. The performance of the resource allocation algorithm is verified and compared to others in the literature using computational simulations. In these simulations, we also consider current and candidate techniques for next generation networks such as f-OFDM (filtered-Orthogonal Frequency Division Multiplexing) and carrier aggregation in order to show that is possible to provide higher data rate, efficient spectrum utilization, better coverage and lower latency in relation to 4G networks.*

1. Introdução

Em um futuro próximo, diferentes tecnologias irão conectar bilhões de máquinas, equipamentos hospitalares, monitores de saúde, carros autônomos e inúmeros outros dispositivos cujas aplicações ainda nem foram concebidas [Rysavy 2015]. Para atender essas demandas, devem ser realizadas melhorias significativas na arquitetura das redes celulares, e as redes sem fio de próxima geração 5G pretendem expandir a capacidade de comunicação [Gupta 2015].

Globalmente, a indústria celular convergiu para a tecnologia LTE (*Long Term Evolution*, Evolução de Longo Prazo), incluindo o LTE-A (LTE-*Advanced*, LTE-Avançado), como interface sem fio. Uma indústria que estava fragmentada em múltiplas tecnologias de rede sem fio agora possui um padrão, resultando em economia e padronização em termos de infraestrutura. Sendo assim, muitos pesquisadores da área

propõem que a tecnologia 5G seja uma extensão dessa interface de comunicação [Olwal 2016] [Rysavy 2015].

Focando em tecnologias com o objetivo de melhorar a taxa de transmissão de dados para suportar redes sem fio da próxima geração, um dos tópicos abordados neste trabalho compreende o conceito de agregação de portadora. Este é um recurso da tecnologia LTE-A que operadores estão implantando para prover maior taxa de dados, melhor cobertura e menor latência [Rostami 2015]. Este trabalho também considera a tecnologia f-OFDM (*filtered-Orthogonal Frequency Division Multiplexing*, Multiplexação por Divisão de Frequência Ortogonal filtrado), que possibilita uma operação fragmentada do espectro e a coexistência eficiente de múltiplas sub-bandas com o objetivo de melhorar a utilização do espectro [Bi 2017] [Wu 2016]. A escolha da tecnologia f-OFDM se justifica por melhorar a taxa de transferência do sistema em 10% ao usar as bandas livres das redes LTE, além de suportar a transmissão assíncrona de diferentes usuários e ser compatível com a tecnologia MIMO (*Multiple Input Multiple Output*, Múltiplas Entradas Múltiplas Saídas) [Abdoli 2015].

Há várias propostas de esquemas de alocação de recursos em redes sem fio [Alasti 2010] [Ferreira 2015] [Guan 2011] [Kausar 2012] [Su 2012]. Em [Su 2012], é proposto um algoritmo para alocação de blocos de recursos em redes LTE baseado na heurística PSO (*Particle Swarm Optimization*, Otimização por Enxame de Partículas), com o objetivo de maximizar a vazão total do sistema. Em [Guan 2011], é proposto um algoritmo com o objetivo de garantir o critério de taxa mínima de transmissão requerida pelo usuário. Em [Ferreira 2015] é proposto um algoritmo com o objetivo de minimizar o parâmetro de retardo e atender critérios de QoS (*Quality of Service*, Qualidade de Serviço). Estes esquemas de alocação, em adição a outros não mencionados, tem o mesmo propósito de maximizar a vazão total através de heurísticas que focam em diferentes parâmetros de QoS como retardo, taxa mínima, etc.

Neste artigo, propomos um esquema de alocação de blocos de recurso para redes sem fio, considerando técnicas de próxima geração, de forma atender o parâmetro de QoS de taxa mínima de transmissão. O problema de alocação de recursos é resolvido aplicando uma heurística EP (*Evolutionary Programming*, Programação Evolucionária). O artigo está dividido da seguinte forma: Primeiramente, apresentamos conceitos que abordam o problema de alocação de recursos, cujo objetivo é maximizar a vazão total do sistema garantindo certos requisitos de largura de banda. Posteriormente, é aplicada a heurística EP para decidir quais blocos de recursos serão alocados para cada usuário, dado o requisito de taxa mínima. O desempenho do esquema de alocação proposto baseado em EP é então validado através de simulações comparativas com outros algoritmos da literatura.

2. Modelo de Sistema e Formulação do Problema

Considere uma rede celular baseada em LTE-A com apenas um eNB (*evolved Node* B, nó evoluído B) consistindo de N UEs (*User Equipments*, Equipamentos de Usuário) como $\alpha = \alpha_1, \alpha_2, \dots, \alpha_N$. Cada usuário $\alpha_n \in \alpha$ pode possuir diferentes capacidades de CA (*Carrier Aggregation*, Agregação de Portadora) que é modelada como $\mu = \{\mu_n | \mu_n \in \{1,2,3,4,5\}\}_{1xN}$, onde μ_n representa o número máximo de CCs (*Component Carriers*, Portadoras Componente) que $\alpha_n \in \alpha$ pode suportar (se $\alpha_n \in \alpha$ é um usuário LTE Release 8, $\mu_n = 1$).

Todos UEs em um dado TTI (*Transmission Time Interval*, Intervalo de Tempo de Transmissão), competem por $\mathcal{M}$ CCs ortogonais sem sobreposição representados por $\beta = \{\beta_1, \beta_2, \dots, \beta_{\mathcal{M}}\}$ onde cada $\beta_m \in \beta$ possui diferente número de RBs (*Resource Blocks*, Blocos de Recurso) e podem ser escritos como $\gamma = \{\gamma_m | \gamma_m \in \{6,15,25,50,75,100\}\}_{1x\mathcal{M}}$, onde γ_m corresponde ao número de RBs em um dado β_m [Rostami 2015].

Assumindo que o índice MCS (*Modulation and Coding Scheme*, Esquema de Codificação e Modulação) máximo de todos UEs em diferentes RBs é modelado como [Rostami 2015]

$$K = \{k_{m,p,n} | k_{m,p,n} \in \{0,1, \dots, h\}\}_{\mathcal{M}x\mathrm{P}xN} \tag{1}$$

onde $k_{m,p,n}$ representa o índice MCS máximo para α_n em β_m no p^{th} RB. O valor de $k_{m,p,n}$ depende do intervalo de qualidade do canal entre 0 e h. Enquanto que o valor do h para *uplink* é 22 e para *downlink* é 28 e $\mathrm{P} = \max[\gamma]$. Também é definida uma matriz de alocação de blocos de recurso sendo uma matriz binária como segue [Rostami 2015]

$$A = \{a_{m,p,n} | a_{m,p,n} \in \{0,1\}\}_{\mathcal{M}x\mathrm{P}xN} \tag{2}$$

Representando o mapa de alocação de RBs onde $a_{m,p,n} = 1$ se e apenas se o p^{th} RB localizado em β_m é alocado unicamente para α_n e $a_{m,p,n} = 0$ caso contrário. A matriz de alocação de recursos A deve satisfazer o requisito de interferência definido em (3), ou seja, dois ou mais UEs não podem utilizar o mesmo RB simultaneamente [Rostami 2015].

$$\sum_{n=1}^{N} a_{m,p,n} \leq 1 \; para \; 1 \leq p \leq \mathrm{P}, 1 \leq \mathrm{m} \leq \mathcal{M} \tag{3}$$

É definida uma matriz de alocação de CC como uma matriz binária $\mathcal{M}xN$ [Rostami 2015] dada por:

$$E = \left\{ e_{m,n} | e_{m,n} = 1 \leftrightarrow \sum_{p=1}^{P} a_{m,p,n} \geq 1 \right\}_{\mathcal{M}xN} \tag{4}$$

onde $e_{m,n}$ representa a condição se β_m é atribuído para α_n ou não. Para representar o MCS alocado para α_n em certo TTI, uma matriz de usuários é definida como segue: [Rostami 2015]

$$B = \{b_{m,n} | b_{m,n} \in \{0,1, \dots, h\}\}_{\mathcal{M}xN} \tag{5}$$

onde $b_{m,n}$ representa o índice MCS alocado para α_n em β_m para cada TTI. O 3GPP TS36.213 [3GPP 2012] especifica a taxa de transmissão correspondente para cada índice MCS. Neste trabalho, para representar a relação entre o índice MCS e a taxa de transmissão, é usada a notação $r = R(b)$ onde R mapeia cada índice MCS para cada taxa de transmissão conforme [3GPP 2012]; em outras palavras, r é a taxa de transmissão obtida para cada UE em um RB com MCS b. É definida então a matriz de taxa de transmissão dos usuários como $R = \{r_{m,n} | r_{m,n} = R(b_{m,n})\}_{\mathcal{M}xN}$ ou seja, uma matriz $\mathcal{M}xN$ onde $r_{m,n}$ apresenta a taxa de transmissão por RB para α_n em β_m [Rostami 2015].

O eNB é responsável por todo procedimento de admissão, escalonamento de recursos e adaptação de *link*. Após receber o CSI (*Channel State Information*, Informação do Estado do Canal) de todos UEs, um mapa de alocação de recursos é construído e o índice MCS é determinado pelo eNB e depois enviado para cada UE através dos canais de controle [Rostami 2015]. De acordo com o índice MCS, o tipo de modulação e a taxa de codificação para cada UE em cada CC atribuído podem ser determinados [Rostami 2015].

Neste artigo, o escalonador é empregado para maximizar a vazão da célula e atender a taxa mínima de transmissão requerida por usuário. Para o n^{th} UE, a função objetivo de otimização é definida como a vazão adquirida sobre o t^{th} TTI, e pode ser calculada como:

$$f^{(t)} = \sum_{n=1}^{N} \sum_{m=1}^{\mathcal{M}} \sum_{p=1}^{P} r_{m,n}^{(t)} \times a_{m,p,n}^{(t)} \tag{6}$$

onde $r_{m,n}^{(t)}$ e $a_{m,p,n}^{(t)}$ são $r_{m,n}$ e $a_{m,p,n}$ no t^{th} TTI respectivamente. Sujeito a [Rostami 2015]

$$\sum_{n=1}^{N} a_{m,p,n} \leq 1, \tag{7}$$

$$\sum_{m=1}^{\mathcal{M}} e_{m,n} \leq \mu_n, \tag{8}$$

$$b_{m,n} \leq k_{m,p,n} \tag{9}$$

$$r_{m,n} \geq r_{m,n}^{min} \tag{10}$$

para $1 \leq p \leq P$, $1 \leq m \leq \mathcal{M}$ e $1 \leq n \leq N$. A equação (7) assegura que cada RB na rede é atribuído para no máximo um UE. A equação (8) garante que o número de CCs atribuídos para cada UE seja menor que sua capacidade de agregação. A equação (9)

garante que o índice MCS para cada UE em cada CC seja menor que o índice MCS máximo suportado para cada RB atribuído em seu CC correspondente. A equação (10) garante que o escalonador atenda a taxa de transmissão mínima requerida para cada usuário.

3. Algoritmo de Alocação de Recursos Utilizando Programação Evolucionária

Um problema de minimização global pode ser formalizado como um par (S, f), onde $S \subseteq R^n$ é um conjunto delimitado no R^n e $f: S \to R$ é uma função n-dimensional de valor real. O problema é encontrar um ponto $x_{min} \in S$ de forma que $f(x_{min})$ seja o mínimo global em S. Mais especificamente, é necessário encontrar $x_{min} \in S$ de forma que [Yao 1999]

$$\forall x \in S: f(x_{min}) \leq f(x) \tag{11}$$

onde f não precisa ser contínuo mas deve ser delimitado. De acordo com o descrito em Bäck and Schwefel [Back 1993], a EP clássica é implementada como descrito no Algoritmo 1.

Algoritmo 1. Algoritmo clássico de Programação Evolucionária [Yao 1999]

1. Gera a população inicial de μ indivíduos e define $k = 1$. Cada indivíduo é tomado como um par de vetores de valores reais $(x_i, \eta_i), \forall i \in \{1, \dots, \mu\}$, onde x_i's são variáveis objetivo e η_i's são valores de desvio padrão para mutações Gaussianas (também conhecidos como parâmetros de estratégia em algoritmos evolucionários auto adaptativos).
2. Avalia o valor de ajuste para cada indivíduo $(x_i, \eta_i), \forall i \in \{1, \dots, \mu\}$ da população baseado na função objetivo $f(x_i)$.
3. Cada pai $(x_i, \eta_i), i = 1, \dots, \mu$ cria um único descendente (x'_i, η'_i): para $j = 1, \dots, n$

$$x'_i(j) = x_i(j) + \eta_i(j) N_i(0,1) \tag{12}$$

$$\eta'_i(j) = \eta_i(j) exp\left(\tau' N(0,1) + \tau N_j(0,1)\right) \tag{13}$$

   onde $x_i(j)$, $x'_i(j)$, $\eta_i(j)$ e $\eta'_i(j)$ denota o j^{th} componente dos vetores x_i, x'_i, η_i e η'_i, respectivamente. $N(0,1)$ denota um número aleatório com distribuição normal de uma dimensão com média zero e desvio padrão 1. $N_j(0,1)$ indica que o número aleatório é gerado outra vez para cada valor de j. Os fatores τ e τ' são comumente definidos como $\left(\sqrt{2\sqrt{n}}\right)^{-1}$ e $\left(\sqrt{2n}\right)^{-1}$.
4. Calcula o valor ajuste para cada descendente $(x'_i, \eta'_i), \forall i \in \{1, \dots, \mu\}$.
5. Realiza a comparação entre o conjunto de pais (x_i, η_i) e descendentes $(x'_i, \eta'_i), \forall i \in \{1, \dots, \mu\}$. Para cada indivíduo, q oponentes são aleatoriamente e

uniformemente escolhidos de todos os pais e descendentes. Para cada comparação, se o valor ajuste de cada indivíduo seja maior que o valor do oponente, ele recebe uma "vitória".

6. Seleciona os μ indivíduos entre (x_i, η_i) e $(x'_i, \eta'_i), \forall i \in \{1, \dots, \mu\}$ com maior número de "vitórias" para serem os pais da próxima geração.
7. Encerra o algoritmo se o critério de parada é satisfeito; caso contrário, $k = k + 1$ e retorna ao passo 3.

O problema de otimização representado pela função objetiva em (6) pode ser resolvido utilizando a EP clássica, porém fazendo algumas considerações. A EP clássica não trabalha com restrições, então a restrição de taxa mínima requerida para cada usuário definida em (10) se torna uma função de penalidade. Quando a taxa mínima requerida para todos os usuários é satisfeita, a função de penalidade é igual a zero, ou seja, a restrição é satisfeita.

$$F = \sum_{n=1}^{N} f^{(t)} - \left[\max\left(r_{m,n}^{min}\right)\right]^2 \sum_{n=1}^{N} \left[min\left(0, \frac{r_{m,n} - r_{m,n}^{min}}{r_{m,n}^{min}}\right)\right]^2 \quad (14)$$

A otimização na heurística EP é implementada avaliando o custo de cada solução sobre a função objetivo, e os menores custos são armazenados e utilizados na resolução do problema de alocação de recursos.

4. Resultados

Nesta seção são apresentados os resultados das simulações do algoritmo proposto de alocação de blocos de recurso em comparação a outros da literatura, implementados através do software MATLAB® versão R2014a.

As condições de canal para cada usuário e RBs em termos de SINR (*Signal-to-Interference-plus-Noise-Ratio*, Relação Sinal Ruído mais Interferência) foram geradas para cada TTI conforme os parâmetros apresentados na Tabela 1.

Tabela 1. Parâmetros de simulação para modelagem de canal [Ni 2013] [3GPP 2008] [3GPP 2011]

Modelo multipercurso	Rayleigh
Perfil de atraso multipercurso	ETU (*Extended Typical Urban,* Urbano Típico Estendido)
Modelo de perda de percurso	$L = 128.1 + 37.6log10(R)$, R em quilômetros
Sombreamento lognormal	Média 0 e desvio padrão 10dB
Distância entre UE e eNB	1km
Densidade de potência do ruído branco	-174dBm/Hz
Potência máxima de transmissão do eNB	46dBm
Ganho da antena do eNB após perda do cabo	15dBi

Ganho da antena do UE	0dBi
Figura de ruído do UE	9dB
Margem de interferência do UE	4dB
Velocidade do UE	3km/h

As simulações foram desenvolvidas considerando os parâmetros para um cenário de transmissão *downlink* apresentado na Tabela 2. A taxa de bit e o índice MCS associado ao SINR são definidos conforme [Huang 2011]. As simulações das heurísticas EP e PSO foram desenvolvidas utilizando 15 e 30 indivíduos respectivamente e 100 iterações máximas como critério de parada.

Tabela 2. Cenário de simulação da transmissão *downlink*

Capacidade CA	[25, 50, 50, 100, 100]
Taxa de transmissão mínima requerida por usuário	0.768Mbits/s
Tamanho do *subframe*	1ms
Número de TTIs simulados	1000
Forma de onda	OFDM / f-OFDM

As simulações foram desenvolvidas comparando os resultados do algoritmo proposto baseado em EP com os seguintes algoritmos de alocação de recursos: algoritmo baseado em PSO [Su 2012], algoritmo QoS *guaranteed* [Guan 2011] e algoritmo de minimização de retardo, denominado no restante do trabalho como *Min-delay* [Ferreira 2015].

A Figura 1 apresenta o resultado comparativo do algoritmo EP utilizando OFDM e f-OFDM. Verifica-se que ao utilizar f-OFDM no algoritmo de alocação, o desempenho em termos de vazão melhora consideravelmente em relação ao OFDM.

O algoritmo proposto EP apresenta os maiores valores de vazão total para todos os cenários simulados com diferentes números de usuários considerados na alocação, como pode ser visto na Figura 2. O resultado se explica pelo fato do algoritmo proposto dar maior prioridade na alocação dos recursos aos usuários com melhores condições de canal, garantindo consequentemente maiores valores de vazão. Os algoritmos QoS *guaranteed* e *Min-delay* apresentam valores similares, maiores que os valores apresentados pelo algoritmo PSO quando se considera mais de 13 usuários.

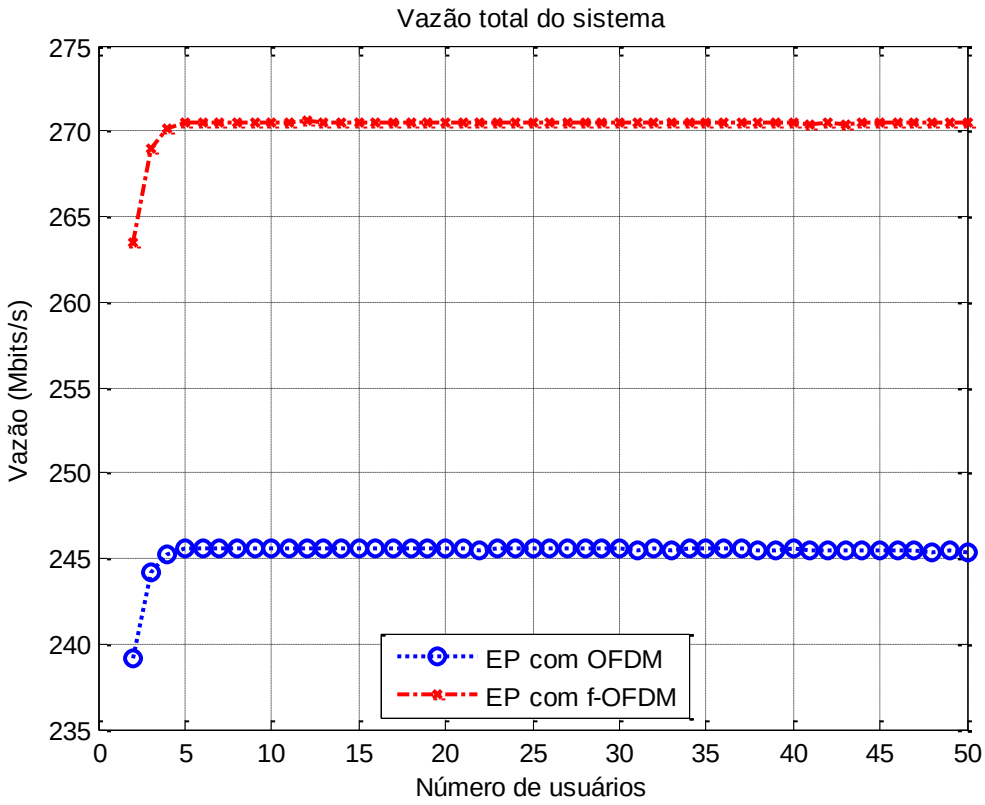


Figura 1. Vazão total utilizando OFDM e f-OFDM

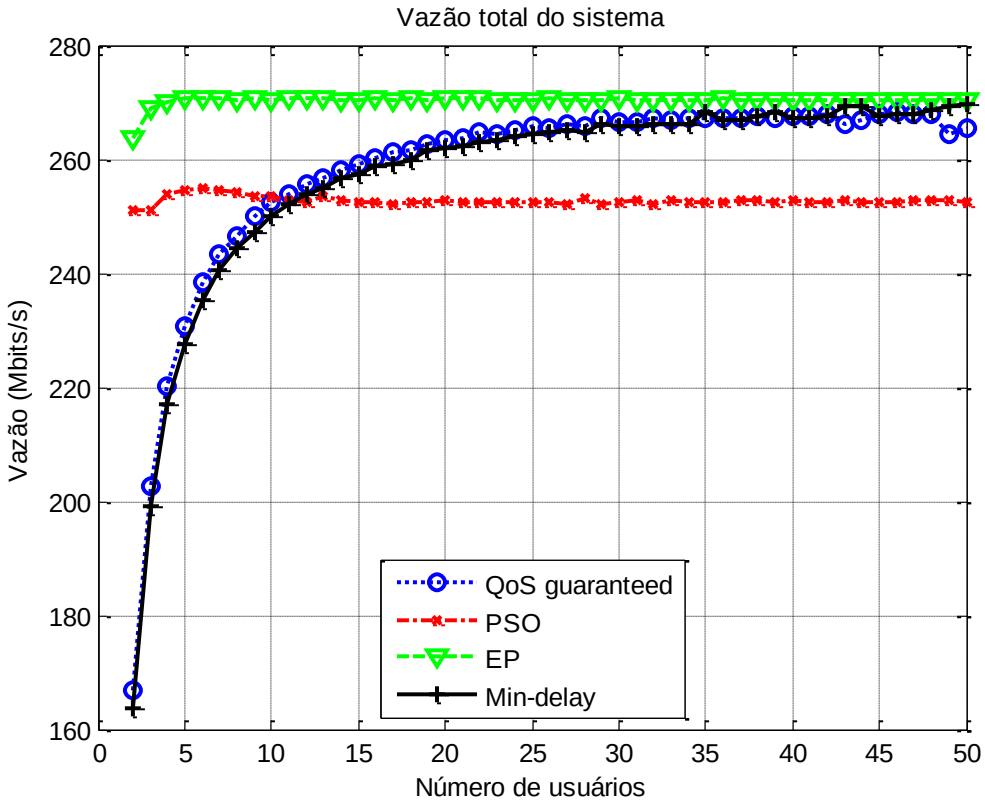


Figura 2. Vazão total utilizando f-OFDM

A Figura 3 mostra que os algoritmos EP e PSO apresentam os maiores valores de vazão média se considerado menos de 10 usuários. Ao se considerar mais de 10 usuários, os algoritmos estudados apresentam valores similares. A vazão média é calculada através da média simples da vazão total em função do número de usuários considerados em simulação.

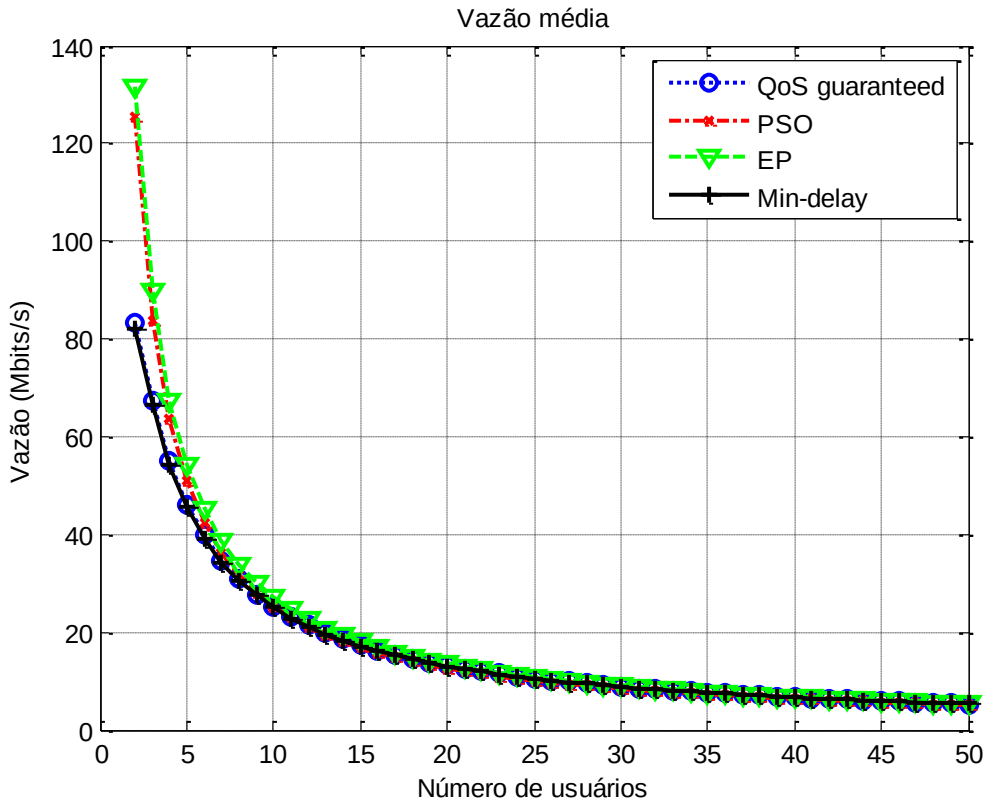


Figura 3. Vazão média utilizando f-OFDM

Em termos de índice *fairness*, calculado conforme [Jain 1984] e apresentado na Figura 4, os algoritmos QoS *guaranteed* e *Min-delay* apresentam valores similares, sendo na maioria das vezes os maiores valores entre os algoritmos comparados. O fraco desempenho do algoritmo proposto se deve ao fato deste dar maior prioridade de alocação aos usuários com melhores condições de canal.

A Figura 5 apresenta o tempo de processamento (em milissegundos) médio para cada simulação com diferentes cenários. As simulações foram realizadas utilizando um microcomputador com a seguinte configuração: Processador Intel Core IR-3570 3.40 GHz, 8Gb RAM, HD SATA III 7200 RPM, Windows 10 64 bits.

Os algoritmos QoS *guaranteed* e *Min-delay* apresentam os menores valores de tempo de processamento. Este fato era esperado, uma vez que os demais algoritmos são baseados nas heurísticas EP e PSO, que objetivam maximizar uma determinada função objetivo, no caso a vazão total, à custa de maior complexidade computacional. O algoritmo EP alcança valores consideravelmente baixos quando comparado com o algoritmo PSO devido ao número reduzido de indivíduos considerados, 15, enquanto o algoritmo PSO considera 30 indivíduos em simulação.

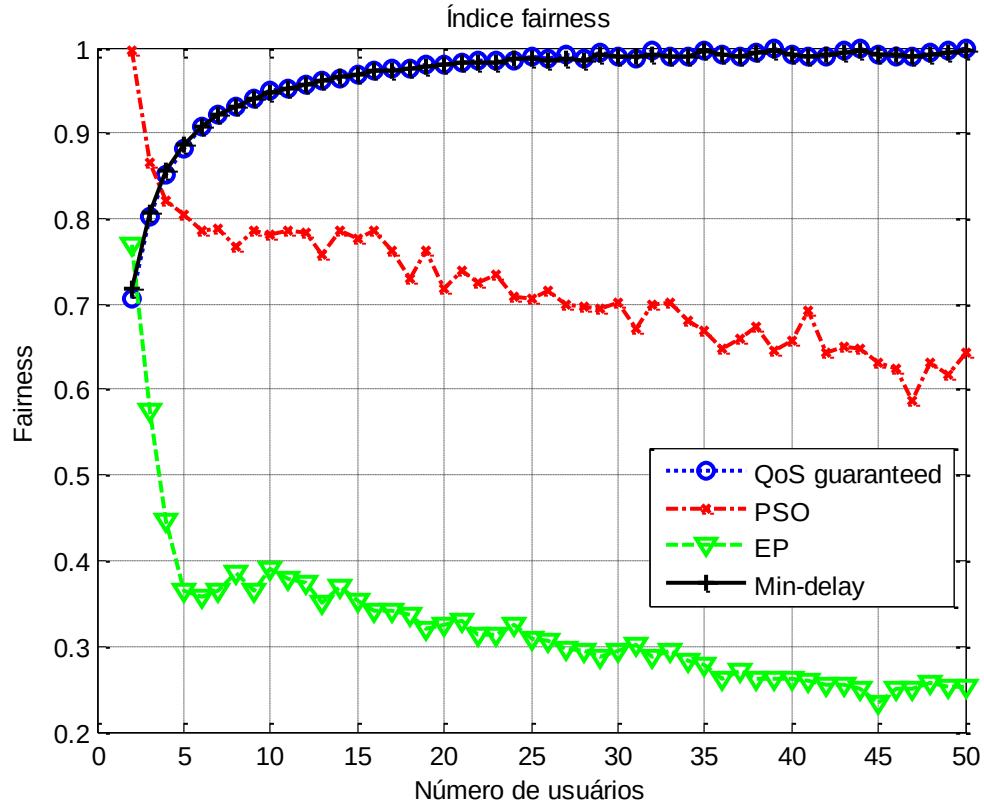


Figura 4. Índice *fairness* utilizando f-OFDM

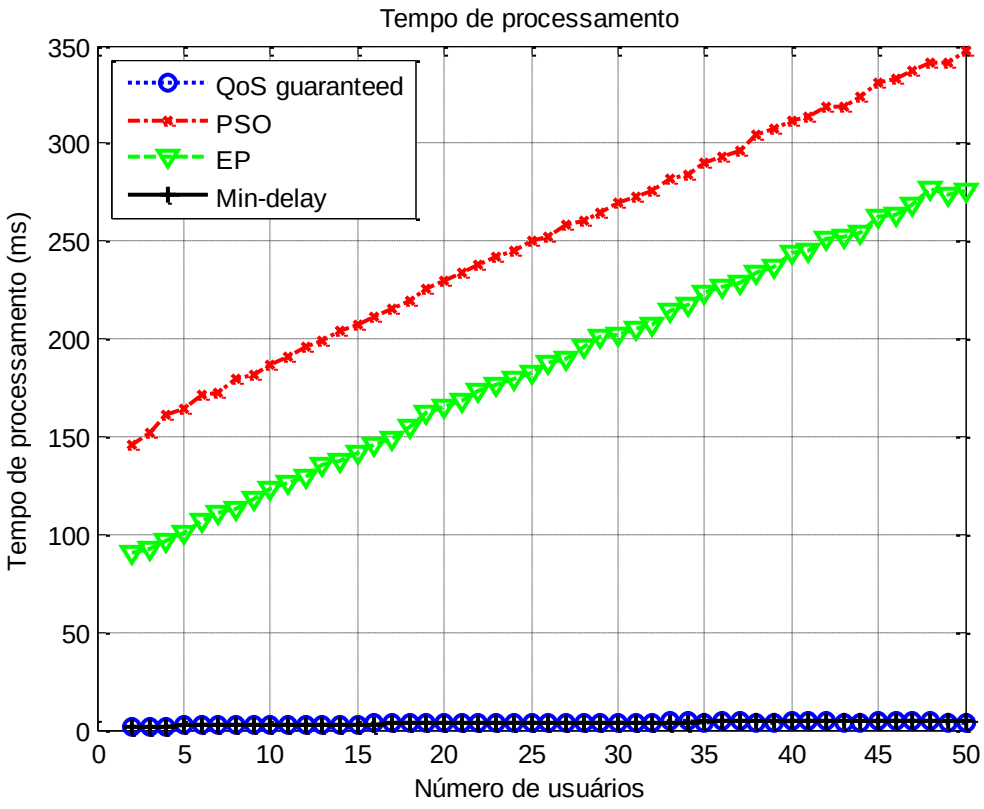


Figura 5. Tempo médio de processamento utilizando f-OFDM

5. Conclusão

Propomos neste artigo um esquema de alocação de blocos de recurso para redes sem fio utilizando tecnologias da próxima geração que objetiva satisfazer parâmetros de QoS aplicando a heurística EP para resolver o problema de alocação de recursos.

Os resultados apresentados nas simulações mostram que ao se utilizar f-OFDM no esquema de alocação, o desempenho do algoritmo proposto em termos de vazão melhora consideravelmente em relação ao OFDM.

Também se verifica que o algoritmo proposto EP possui melhor desempenho que os algoritmos PSO, QoS *guaranteed* e *Min-delay* em termos de vazão total e vazão média, considerando todos os cenários com diferentes números de usuários, às custas de tempo de processamento maior e menor índice *fairness*. O fato se justifica pela característica do algoritmo proposto de dar maior prioridade de alocação dos recursos aos usuários com melhores condições de canal. É importante observar também que o algoritmo EP proposto apresenta melhores resultados no geral que o algoritmo PSO, mesmo considerando metade do número de indivíduos na implementação da heurística.

Referências

3GPP TS 36.104 version 8.3.0 Release 8. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception, 2008.

3GPP TR 36.931 version 9.0.0 Release 9. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) requirements for LTE Pico Node B, 2011.

3GPP TR 36.213. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures, 3rd Generation Partnership Project (3GPP). Jun, 2012.

3GPP. Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description. Valbonne, FR, 2013.

Abdoli, J; Jia, M.; Ma, J.. "Filtered OFDM: A New Waveform for Future Wireless Systems". IEEE 16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), 2015.

Back, T.; Schwefel, H.-P.. "An overview of evolutionary algorithms for parameter optimization", Evol. Comput., vol. 1, no. 1, pp. 1–23,1993.

Bi, Meihua; Jia, Weikang; Li, Longsheng; Miao, Xin; Hu, Weisheng. "Investigation of F-OFDM in 5G fronthaul networks for seamless carrier-aggregation and asynchronous transmission". Optical Fiber Communications Conference and Exhibition (OFC), 2017.

Ferreira, M.V.G.; Vieira, F.H.T.; Abrahão, D.C.. "Minimizing Delay in Resource Block Allocation Algorithm of LTE Downlink". VI International Workshop on Telecommunications: IWT 2015.

Guan, Na; Zhou, Yiqing; Tian, Lin; Sun, Gang; Shi, Jinglin. "QoS guaranteed resource block allocation algorithm for LTE systems". Wireless and Mobile Computing, Networking and Communications (WiMob), IEEE 7th International Conference on, Pag. 307-312, 2011.

Gupta, A.; Jha, R. K.. "A Survey of 5G Network: Architecture and Emerging Technologies". IEEE Access (Volume: 3). Pag. 1206-1232, 2015.

Huang, Jeng-Ji; Lin, Wei-Keng; Ko, Hung-Hsiang. "A Resource Allocation Algorithm for Maximizing Packet Transmission in Donwlink LTE Cellular Systems". TENCON 2011 - 2011 IEEE Region 10 Conference, Pag. 445-449, 2011.

Jain, R.; Hawe, W.; Chiu, D.. "A Quantitative measure of fairness and discrimination for resource allocation in Shared Computer Systems". DEC-TR-301, September 26, 1984.

Kausar, R; Y, Chen; KK, Chai. "Qos aware packet scheduling with adaptive resource allocation for OFDMa based LTE-advanced networks". IET Conference Publications vol. 2011, (586 CP) Pag. 207-212, 2012.

Su, Lin; Wang, Ping; Liu, Fuqian. "Particle swarm optimization based resource block allocation algorithm for downlink LTE systems". Communications (APCC), 18th Asia-Pacific Conference on, Pag. 970-974, 2012.

Ni, M.; Xu, X; Mathar, R. "A Channel Feedback Model with Robust SINR Prediction for LTE Systems. Institute for Theoretical Information Technology", RWTH Aachen University. 7th European Conference on Antennas and Propagation (EuCAP), 2013.

Rostami, S.; Arshad, K.; Rapajic, P.. "A Joint Resource Allocation and Link Adaptation Algorithm with Carrier Aggregation for 5G LTE-Advanced Network". 22nd International Conference on Telecommunications (ICT 2015), 2015.

Rysavy Research/4G Americas. LTE and 5G Innovation: Igniting Mobile Broadband. August, 2015.

Yao, Xin; Liu, Yong; Lin, Guangming. "Evolutionary Programming Made Faster". IEEE Transactions on Evolutionary Computation, Vol. 3, No. 2, July, 1999.

Wang, J.H.; Yin, Z.Y.. "A ranking selection-based particle swarm optimizer for engineering design optimization problems". Structural and Multidisciplinary Optimization, Volume 37, Issue 2, pp 131-147, 2008.

Wu, Dan; Zhang, Xi; Qiu, Jing; Gu, Lian; Saito, Yuya; Benjebbour, Anass; Kishiyama, Yoshihisa. "A Field Trial of f-OFDM toward 5G". Globecom Workshops (GC Wkshps), IEEE, 2016.

Algoritmo de Alocação de Recursos de Rede baseado na Função de Autocorrelação de um Modelo de Tráfego Multifractal

Maykon R. P. da Silva[1], **Flávio G. C. Rocha**[1], **Flávio H. T. Vieira**[1]

[1]Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.605-010 – Goiânia – GO – Brasil

maykoon.renan@gmail.com

{flaviogcr, flavio_vieira}@ufg.br

Abstract. *The autocorrelation function and the Hurst parameter are generally used to estimate characteristics of long-range dependence and self-similarity, respectively, present in computer network traffic traces. These characteristics can reduce performance in terms of quality of service parameters such as throughput, latency and data loss. Existing methods for estimating the autocorrelation function and the Hurst parameter generally use all available samples of a network traffic process. In this paper, we propose an equation for the autocorrelation function based only on the parameters of a multifractal model for the network traffic process. The proposed autocorrelation function presents a power law decay with parameters, among them the Hurst parameter, that can be obtained using linear regression based on ordinary least squares. The value obtained for the Hurst parameter is then used by a dynamic resource allocation algorithm to control network transmission rates. Comparisons with other literature methods are performed, verifying the efficiency of the proposed algorithm in controlling the transmission rates, increasing utilization of the network resources while reducing data loss and buffer occupancy rates.*

Resumo. *A função de autocorrelação amostral e o parâmetro de Hurst são geralmente utilizados para estimar características de dependência de longa duração e autossimilaridade, respectivamente, presentes entre amostras de tráfego de rede de computadores. Essas características podem reduzir o desempenho de redes em atender a parâmetros de qualidade de serviço como vazão, latência e perda de dados. Os métodos existentes para estimação da função de autocorrelação e do parâmetro de Hurst geralmente utilizam todas as amostras disponíveis do tráfego de rede de computadores. Neste artigo, propõe-se uma equação para a função de autocorrelação baseada apenas nos parâmetros de um modelo multifractal para o processo de tráfego de rede. A função de autocorrelação proposta apresenta decaimento segundo uma lei de potência cujos parâmetros, dentre eles o parâmetro de Hurst, podem ser obtidos utilizando regressão linear de mínimos quadrados ordinários. O valor obtido para o parâmetro de Hurst é então utilizado por um algoritmo de controle dinâmico das taxas de transmissão de um servidor de rede. Comparações com outros métodos da literatura são realizadas, verificando a eficiência do algoritmo proposto*

em controlar as taxas de transmissão, aumentando a utilização dos recursos de rede e ao mesmo tempo obter uma redução de buffer que não destoe dos demais métodos.

1. Introdução

Os fluxos de redes de computadores podem englobar diferentes protocolos e conteúdos provenientes de diversas tecnologias de acesso ao meio de comunicação. A heterogeneidade dessas redes proporciona um comportamento de tráfego variável, com presença de rajadas e de difícil predição [Won et al. 2003]. A melhoria no desempenho dessas redes é certamente um dos temas que merecem mais atenção da comunidade científica associada às redes de comunicações. Assim, o desempenho de redes é alvo de pesquisas constantes, e muitos são os trabalhos propondo modelos para descrição do tráfego de redes [Krishna et al. 2003, Lee and Fapojuwo 2005, Rocha and Vieira 2011], mecanismos de predição [Aquino and Barria 2006, Vieira and Lee 2010] e métodos de controle de tráfego [Liu and Hsu 2006].

Dentre os modelos existentes, modelos monofractais e multifractais têm recebido grande atenção devido aos seus desempenhos em termos de análise e modelagem relacionados ao tráfego real de redes [Feldmann et al. 1998, Park and Willinger 2000]. Para esses modelos, o tratamento analítico é geralmente mais difícil que para os modelos Markovianos e impõe limitações ao estudo e projeto relacionados ao atendimento de métricas de QoS.

Neste artigo, propõe-se uma equação para a função de autocorrelação baseada apenas nos parâmetros de um modelo multifractal para o processo de tráfego de rede. Por meio da função de autocorrelação obtida estima-se o parâmetro de *Hurst* [Cappe et al. 2002]. O estimador proposto é comparado a outros estimadores presentes na literatura. Por fim, apresenta-se um algoritmo de alocação dinâmica de recursos de rede baseado no parâmetro de *Hurst* encontrado.

2. Modelagem Multifractal do Tráfego de Redes

O modelo multifractal MWM (*Multifractal Wavelet Model*) é baseado em uma cascata multiplicativa no domínio *wavelet* [Riedi et al. 1999], tendo sido recentemente utilizado para a modelagem do tráfego de redes [Tuberquia-David et al. 2016, Anderson et al. 2017].

Para se gerar um processo segundo o modelo MWM é preciso aplicar a transformada *wavelet* discreta (DWT – *Discrete Wavelet Transform*) ao tráfego de rede, calcular os momentos de segunda ordem dos coeficientes *wavelet* em cada escala, a média e a variância dos coeficientes na escala de maior resolução e calcular ρ_j , variável usada para capturar o decaimento de energia dos coeficientes *wavelet* em escala.

No MWM, um sinal $C(t)$ positivo, estacionário e com longa dependência entre as amostras é representado no domínio *wavelet*. Assim, seja $C^{(n)}[k]$ um sinal discreto no tempo e uma aproximação de $C(t)$ na resolução de 2^{-n} , onde n é a variável relacionada à escala de tempo. Utilizando a *wavelet* de Haar, o processo discreto $C^{(n)}[k]$ assume valores que correspondem à integral de $C(t)$ no intervalo $[k2^{-n}, (k+1)2^{-n}[$, e pode ser matematicamente descrito pela seguinte equação:

$$C^{(n)}[k] = \int_{k2^{-n}}^{(k+1)2^{-n}} C(t)dt = 2^{-n/2}U_{n,k}. \quad (1)$$

Para a *wavelet* de Haar, os coeficientes de escala $U_{j,k}$ e os coeficientes da transformada *wavelet* $W_{j,k}$ podem ser recursivamente calculados, onde j é a escala e k o instante de tempo, e a simples restrição $|W_{j,k}| \leq U_{j,k}$ garante que o processo $U_{j,k}$ seja sempre positivo [Riedi et al. 1999, Tuberquia-David et al. 2016].

Neste modelo, pode-se relacionar o deslocamento k_j de um coeficiente de escala ao deslocamento de um de seus dois descendentes diretos k_{j+1} por meio de $k_{j+1} = 2k_j + k'_j$ onde $k'_j = 0$ corresponde ao descendente levogiro da cascata e $k'_j = 1$ ao descendente dextrogiro da cascata.

No MWM assume-se que, em cada escala j, os multiplicadores $A_{j,k}$, $k = 0, 1, ..., 2^{j-1}$, são independentes e identicamente distribuídos (i.i.d) de acordo com alguma variável aleatória $A_j \in [-1, 1]$, onde A_j representa o vetor de multiplicadores $A_{j,k}$ na escala j e para todos os valores de k. Quando os multiplicadores $A_{j,k}$ podem ser descritos por meio de uma distribuição beta, o modelo MWM é chamado de βMWM. O processo de tráfego sintético é definido pela seguinte equação:

$$C^{(n)}[k] = 2^{-n}U_{0,0}\prod_{j=0}^{n-1}[1 + (-1)^{k_j{}'}A_{j,k_j}]. \quad (2)$$

Assim, processos multifractais podem ser definidos por uma lei de escala para os momentos estatísticos dos processos de incrementos sobre intervalos de tempo finitos. Maiores detalhes podem ser obtidos em [Riedi et al. 1999].

3. Função de Autocorrelação do Modelo βMWM

A função de autocorrelação (ACF-*AutoCorrelation Function*) pode ser utilizada para expressar estatísticas de segunda ordem e o seu decaimento pode fornecer uma ideia a respeito da característica de longa-dependência entre as amostras de um processo de tráfego de rede. Nesta seção, propõe-se uma equação para a ACF do modelo βMWM.

Proposição 1. *A Função de Autocorrelação para o processo de tráfego sintético gerado a partir do modelo βMWM pode ser escrito em função dos parâmetros μ_c, σ_c e ρ, conforme a seguir:*

$$\hat{r}(l) = \frac{1}{N}\sum_{k=1}^{N-l} 2^{-2n}G(\mu_c,\sigma_c)^2\left[\prod_{j=0}^{n-1}(1+(-1)^{k_j{}'}\beta(\rho_j,\rho_j))-1\right]\left[\prod_{j=0}^{n-1}(1+(-1)^{k_j{}'}\bar{\beta}(\rho_j,\rho_j))-1\right]. \quad (3)$$

Seja l o deslocamento de tempo, onde, $l = 2^m$ para $m = 1, 2, ..., n$, N é o número de amostras, n significa o número de escalas, os parâmetros μ_c e σ_c correspondem a média e desvio padrão do coeficiente de escala $U_{0,0}$, respectivamente, gerado segundo uma distribuição Gaussiana e, ρ_j representa o parâmetro da distribuição β e $\bar{\beta}$ para cada escala j.

Demonstração. Uma Função de Autocorrelação pode ser escrita para uma série temporal X com média μ [Aguirre 2004], por meio seguinte equação:

$$r_{x,x}[l] = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} (X(i) - \mu)(X(i+l) - \mu). \tag{4}$$

Para séries temporais de duração finita, $r_{x,x}[l]$ pode ser estimado utilizando-se [Aguirre 2004]:

$$\hat{r}_{x,x}[l] = \frac{1}{N} \sum_{i=1}^{N-l} (X(i) - \mu_c)(X(i+l) - \mu_c), l = 0, 1, ..., N-1. \tag{5}$$

A função de autocorrelação amostral $\hat{r}_{x,x}[l]$ é um estimador consistente da função de autocorrelação $r_{x,x}[l]$ [Aguirre 2004]. Substituindo $X(i)$ em (5) por $C^n[k]$, tem-se :

$$\hat{r}[l] = \frac{1}{N} \sum_{k=1}^{N-l} (C^{(n)}[k] - \mu_c)(C^{(n)}[k+l] - \mu_c), \tag{6}$$

onde,

$$C^{(n)}[k] - \mu_c = 2^{-n} U_{0,0} \prod_{j=0}^{n-1} (1 + (-1)^{k_j'} A_{j,k_j}) - \mathbb{E}[C^{(n)}[k]], \tag{7}$$

Aplicando operador de esperança matemática $\mathbb{E}[.]$ em (2), tem-se:

$$\mathbb{E}[C^{(n)}[k]] = 2^{-n} \mathbb{E}[U_{0,0}] \mathbb{E} \left[\prod_{j=0}^{n-1} (1 + (-1)^{k_j'} A_{j,k_j}) \right]. \tag{8}$$

Sabendo que os multiplicadores são i.i.d (independentes e identicamente distribuídos) [Riedi et al. 1999], tem-se:

$$\mathbb{E}[C^{(n)}[k]] = \mathbb{E}[U_{0,0}] \prod_{j=0}^{n-1} \mathbb{E} \left[\frac{(1 + (-1)^{k_j'} A_{j,k_j})}{2} \right]. \tag{9}$$

Como A_{j,k_j} segue uma distribuição β simétrica no intervalo $[-1, 1]$, então da f.g.m (função geradora de momentos) da distribuição β tem-se:

$$\mathbb{E}[C^{(n)}[k]] = \mathbb{E}[U_{0,0}] \prod_{j=0}^{n-1} \frac{1}{2} = 2^{-n} \mathbb{E}[U_{0,0}], \tag{10}$$

sabendo que $E[U_{0,0}] = U_{0,0}$, onde assume-se uma cascata com apenas um elemento (amostra) na escala mais grosseira. Substituindo 10 em 7 tem-se:

$$C^{(n)}[k] - \mu_c = 2^{-n} U_{0,0} \left[\prod_{j=0}^{n-1} (1 + (-1)^{k_j'} A_{j,k_j}) - 1 \right]. \tag{11}$$

De forma análoga, tem-se:

$$C^{(n)}[k+l] - \mu_c = 2^{-n} U_{0,0} \left[\prod_{j=0}^{n-1} (1 + (-1)^{k'_j} A_{j,\bar{k}_j}) - 1 \right], \tag{12}$$

onde $\bar{k}_j = 2k_j + 2^{m-1}$ e $l = 2^m$, isto é, para $m \in \mathbb{N}$, l é diático. Substituindo (11) e (12) em (6) e sabendo que os multiplicadores $A_{j,k_j} \overset{\mathrm{d}}{=} \beta(\rho_j, \rho_j)$ e $A_{j,\bar{k}_j} \overset{\mathrm{d}}{=} \bar{\beta}(\rho_j, \rho_j)$ e que o coeficiente de escala $U_{0,0} \overset{\mathrm{d}}{=} G(\mu_c, \sigma_c)$, tem-se:

$$\hat{r}(l) = \frac{1}{N} \sum_{k=1}^{N-l} 2^{-2n} G(\mu_c, \sigma_c)^2 \left[\prod_{j=0}^{n-1} (1 + (-1)^{k'_j} \beta(\rho_j, \rho_j)) - 1 \right] \left[\prod_{j=0}^{n-1} (1 + (-1)^{k'_j} \bar{\beta}(\rho_j, \rho_j)) - 1 \right]. \tag{13}$$

□

4. Estimação do Parâmetro *Hurst*

O conceito de autossimilaridade foi discutido por [Mandelbrot 1965]. Posteriormente, esse conceito encontrou aplicações no contexto da modelagem do tráfego de redes de computadores [Riedi et al. 1999]. O parâmetro de *Hurst* pode ser utilizado para expressar o grau de autossimilaridade de um processo. Esta seção propõe-se a estimação do parâmetro de *Hurst* a partir da equação 6 e apresenta alguns métodos para estimação do parâmetro de *Hurst*.

4.1. Estimação do Parâmetro de *Hurst* baseado na Função de Autocorrelação Proposta

Seja $X = (X_t : t = 0, 1, ...)$ um processo estocástico cuja $r(k)$ (função de autocorrelação) pode ser escrita como: $r(k) = E[(X_t - \mu)(X_{t+k} - \mu)]$, que depende apenas de k. Em particular, vamos assumir que $r(k) \sim k^{-\beta} c$, com $0 < \beta < 1$ e c sendo uma constante.

Os parâmetros β e c são calculados pelo método dos mínimos quadrados ordinários (OLS-*Ordinary Least Squares*) que é um método para estimar os parâmetros desconhecidos em um modelo de regressão linear, com o objetivo de minimizar a soma dos quadrados das diferenças entre as respostas observadas em um conjunto de dados e aqueles preditos por uma função linear. Visualmente isso é visto como a soma das distâncias verticais quadradas entre cada ponto de dados no conjunto e o ponto correspondente na linha de regressão, quanto menor as diferenças, melhor o modelo se encaixa nos dados [Hayashi 2000].

Seja $X^{(m)} = (X^{(m)}(k) : t = 1, 2, ...)$ uma nova série de tempo agregado obtida pela média da série temporal original X sobre blocos não superpostos de tamanho m . O agregado $X^{(m)}$ define também um processo estacionário de covariância com função de autocorrelação $r^{(m)}$. O processo X é chamado (exatamente de segunda ordem) autossimilar com o parâmetro de *Hurst* que é dado pela equação $H = 1 - \beta/2$, se todos os processos agregados $X^{(m)}$ tiverem a mesma função de autocorrelação como X. Em outras palavras, X é exatamente autossimilar se for indistinguível dos seus processos agregados pelo menos em relação às suas propriedades de segunda ordem [Becchi 2008].

O processo X é chamado (assintoticamente de segunda ordem) autossimilar se, para grandes m, as séries temporais agregadas correspondentes tenham estrutura de correlação fixa (determinada por β) [Becchi 2008].

4.2. Métodos para estimação do parâmetro de *Hurst*

No presente artigo serão utilizados 3 métodos de estimação do parâmetro de *Hurst* para comparação com o valor de H obtido a partir da equação proposta para a Função de Autocorrelação.

O método AGGVAR é baseado na curva de um gráfico logaritmo que relaciona a variância das amostras dos dados com o tamanho do bloco de cada nível de agregação [Karagiannis et al. 2006].

O método R/S estima o parâmetro de *Hurst* baseado em redimensionamentos de tamanhos diferentes de seções do processo analisado. Os redimensionamentos ocorrem dividindo cada uma das seções anteriores ao meio e calculando a faixa redimensionada para cada nova seção. Cada seção é simbolizada pelo seu valor médio. O valor de R é obtido da diferença entre os valores máximos e mínimos dos desvios das seções subdivididas e S está relacionado ao desvio padrão dessas seções. O H calculado é proporcional ao quociente R/S[Gospodinov and Gospodinova 2005].

O *Whittle Estimator* é um estimador para o parâmetro de *Hurst* no domínio *wavelet* e é baseado na minimização de uma função de verossimilhança aplicada ao periodograma da série temporal. O método *Whittle* assume que o espectro de potência do processo subjacente do conjunto de dados seja conhecido [Karagiannis et al. 2006].

5. Método de Alocação Dinâmica de Recursos de Rede

O DRA (*Dynamic Resource Allocation*) é aplicável ao tráfego de redes com características autossimilares e baseia-se na predição do tráfego de entrada e na estimação de parâmetros como o *H* (parâmetro de *Hurst*), μ (a média da taxa do tráfego de entrada) e σ (o desvio padrão do tráfego de entrada). O algoritmo DRA é aplicado a um sistema de fila a fim de controlar os valores de taxas de transmissão a medida que um preditor de tráfego e um estimador de parâmetros fornece informações ao controlador de taxa, conforme representado pela Figura 1:

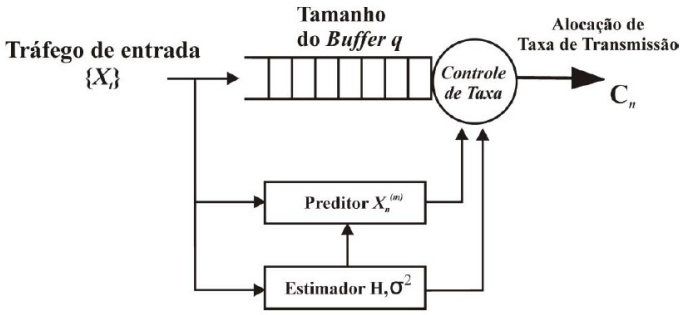


Figura 1. Esquema de alocação dinâmica de recursos [Trinh and Miki 1999]

No DRA, define-se uma variável $X_s^{(p)}$ chamada de média em blocos do processo de tráfego como:

$$X_s^{(p)} = \frac{1}{p}(X_{sp-p+1} + ... + X_{sp}), \tag{14}$$

onde $p, s = 1, 2, ...$, p é o tamanho do bloco e s o número do bloco.

Dessa forma, um preditor para $X_s^{(p)}$ pode ser dado por [Beran 1994]:

$$\hat{X}_s^{(p)} = aX_{s-1}^{(p)}(1-a)\mu, \tag{15}$$

onde $a = (2^{2H} - 2)/2$ e a está contido no intervalo aberto $(0, 1)$, ou seja, $a \in (0, 1)$. H é o parâmetro de *Hurst*.

A taxa alocada T para o bloco n é dada por :

$$T_n = \hat{X}_s^{(p)} + \Delta T. \tag{16}$$

O valor de ΔT é dado por:

$$\Delta T \approx \left\{ \frac{-2ln\varepsilon + 2ln\left(\frac{1}{\sqrt{2\pi}}y^{-1}\right)^{1/2}\sigma(1-a)H^H(1-H)^{1-H}}{b^{1-H}} \right\}^{1/H}, \tag{17}$$

onde ε é a probabilidade de violação do tamanho do *buffer*, σ é o desvio padrão do tráfego considerado e q é o tamanho do *buffer*, onde a variável y está relacionada com a seguinte aproximação:

$$\varepsilon \approx \frac{1}{\sqrt{2\pi}}y^{-1}e^{\left(-\frac{1}{2}y^2\right)}. \tag{18}$$

Em [Trinh and Miki 1999, Doty et al. 2016], os autores consideram o DRA como um método eficiente para alocação dinâmica de taxa, quando comparado ao serviço de taxa fixa, já que este esquema apresenta melhor utilização dos recursos enquanto provê as mesmas garantias de qualidade de serviço (QoS – *Quality of Service*).

6. Algoritmo de Alocação de Recursos baseado na Proposta Equação para a Função de Autocorrelação

Nesta seção, propõe-se que o DRA pode ser melhorado se estimativas mais precisas puderem ser obtidas para *H*. O Algoritmo 1, apresentado a seguir, aloca recursos baseado no parâmetro de *Hurst* (*H*) obtido a partir da Função de Autocorrelação proposta em (3).

Algoritmo 1: ALGORITMO PROPOSTO

início
  Passo 1 - Cálculo dos Parâmetros por meio do Método βMWM
  X(t) - Série Temporal do Tráfego de Rede
  Entrada: $(X(t))$
  início
    Executa $\rightarrow \beta MWM(X(t))$ [Riedi et al. 1999]
  fim
  Saída: (ρ,μ,σ)
  Passo 2 - Estimação da Autocorrelação a Partir dos Parâmetros βMWM
  Entrada: (ρ,μ,σ)
  início
    Calcula a Autocorrelação Através da Equação 3
  fim
  Saída: ($\hat{r}[l]$)
  Passo 3 - Cálculo do Parâmetro β e Estimação do Parâmetro de *Hurst* (*H*)
  Entrada: ($\hat{r}[l]$)
  início
    Calcula o β Utilizando o Método OLS [Hayashi 2000]
    β = ($log(\hat{r}[l])/log(k)$)
    Calcula o parâmetro de *Hurst* (*H*) [Becchi 2008]
    $H = 1-\beta/2$
  fim
  Saída: (H)
  Passo 4 - Algoritmo de Alocação Dinâmica de Recursos
  Entrada: (H,μ,σ)
  início
    Executa $\rightarrow DRA$(H,μ,σ) [Trinh and Miki 1999]
  fim
  Saída: Taxa de transmissão Alocada ao Servidor
fim

7. Resultados e Simulações

Foram conduzidas simulações com o objetivo de avaliar o desempenho do Algoritmo Proposto. Para tanto, são apresentados os seguintes resultados: processo de modelagem utilizando o βMWM; resultados de obtenção da Função de Autocorrelação utilizando a equação (3); estimação do parâmetro de *Hurst* a partir da Função de Autocorrelação e por fim, os resultados de alocação dinâmica de recursos utilizando o parâmetro de *Hurst* obtido.

Serão apresentados os resultados obtidos para as seguintes séries reais de tráfego de rede: dec-pkt-1, dec-pkt-2 que são séries de tráfego TCP/IP, disponíveis em [Dec 2000] e obtidas da LBL (*Lawrence Berkeley Laboratory*) e, WaikatoVIII-20111027-213205-5 (Waikato1) e WaikatoVIII-20110921-000000-0 (Waikato2) disponíveis em [Wand 2011], essas séries foram obtidas em 2011 no Departamento de Ciência da Computação da Universidade de Waikato.

7.1. Resultados de modelagem utilizando o βMWM

A Tabela 1 abaixo apresentam as estatísticas obtidas para a série dec-pkt-1. Foram estimados parâmetros estatísticos para as séries temporais reais e para a séries temporais sintéticas modeladas utilizando os modelos βMWM e VVGM (*Variable Variance Gaussian Model*) [Krishna et al. 2003].

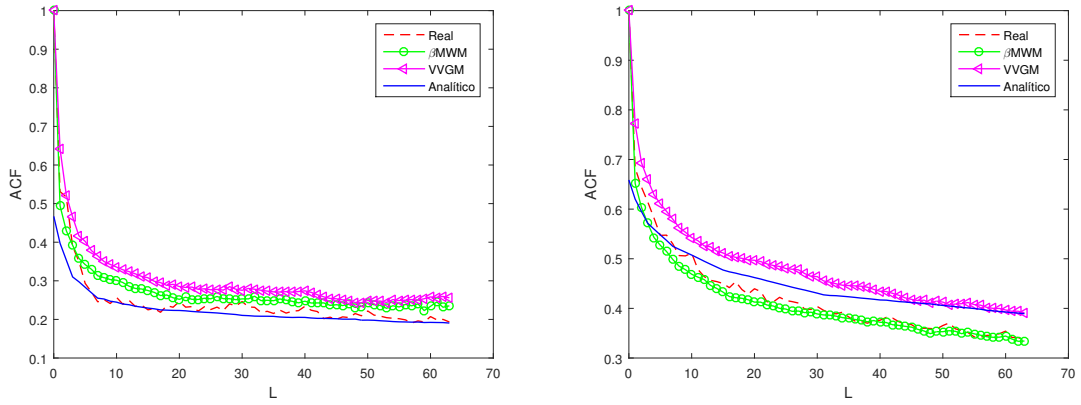
Tabela 1. Estatísticas dadas em *bytes* para a série de tráfego real dec-pkt-1 em comparação aos modelos utilizados.

dec-pkt-1	**MÉDIA**	**VARIÂNCIA**	**PICO/MÉDIA**
REAL	$1,26\text{x}10^4$	$6,51\text{x}10^7$	$5,76$
βMWM (E %)	$1,28\text{x}10^4$ $(1,58)$	$6,99\text{x}10^7$ $(7,37)$	$7,20$ $(25,00)$
VVGM (E %)	$1,43\text{x}10^4$ $(13,49)$	$2,90\text{x}10^7$ $(55,45)$	$5,34$ $(7,29)$

Analisando-se os resultados da Tabela 1 dos modelos βMWM e VVGM e comparando com os resultados obtidos com o tráfego real, percebe-se que, o modelo βMWM obteve resultados próximos aos da série real para a média e variância porém para o pico/média o VVGM obteve um resultado com um erro inferior ao βMWM.

7.2. Resultados com a Função de Autocorrelação (ACF-*AutoCorrelation Function*)

A seguir são apresentadas as ACFs obtidas para as séries de tráfego dec-pkt-2 e Waikato2 geradas a partir da série de tráfego real, tráfego sintético gerado por modelos multifractais βMWM, VVGM e por meio da equação 3 .



(a) Gráfico da ACF para a série de tráfego dec-pkt-2. (b) Gráfico da ACF para a série de tráfego Waikato2.

Figura 2. Gráfico da ACF das séries dec-pkt-2 e Waikato2.

Analisando os gráficos acima, percebe-se que a curva ACF gerada pela modelo proposto apresentou resultados similares ao da curva da ACF geradas pelo βMWM, e superior as curvas da ACF geradas pelo modelo VVGM.

Abaixo a Tabela 2 apresenta os EQM (Erro Quadrático Médio) para as ACFs do modelos sintético e proposto em relação a ACF do tráfego real para as séries dec-pkt-1, dec-pkt-2, Waikato1 e Waikato2. Os números da Tabela 2 mostram como o modelo

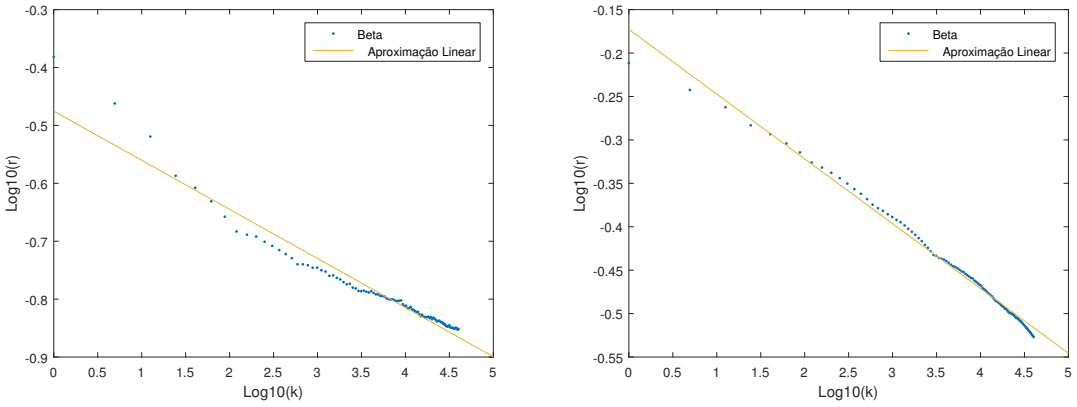
proposto apresentou resultados próximos ao modelo βMWM e superior ao VVGM para quase todas as séries.

Tabela 2. EQM das curvas da ACF.

Séries de Tráfego	Erro Quadrático Médio		
	MWM	VVGM	Modelo Proposto
dec-pkt-1	$6,34\text{x}10^{-4}$	$3,90\text{x}10^{-3}$	$4,50\text{x}10^{-3}$
dec-pkt-2	$6,80\text{x}10^{-3}$	$4,10\text{x}10^{-3}$	$3,50\text{x}10^{-3}$
Waikato1	$7,00\text{x}10^{-3}$	$3,30\text{x}10^{-3}$	$2,90\text{x}10^{-3}$
Waikato2	$6,23\text{x}10^{-4}$	$1,30\text{x}10^{-2}$	$1,60\text{x}10^{-3}$

7.3. Resultados de Estimação do Parâmetro de *Hurst* (*H*)

O cálculo do H foi feito por meio da equação $r(k) \sim k^{-\beta}c$, sendo, $c = 2$, onde o objetivo era encontrar o parâmetro β e, através do método OLS procura encontrar o melhor ajuste para um conjunto de dados β como ilustra a Figura 3.



(a) Gráfico do parâmetro β da série dec-pkt-2

(b) Gráfico do parâmetro β da série dec-pkt-2

Figura 3. Gráficos do parâmetro β séries de tráfego dec-pkt-2 e Waikato2.

Com o valor do expoente β pode-se encontrar o valor do parâmetro de *H* por meio da equação $H = 1 - \beta/2$, conforme proposto na seção 4.

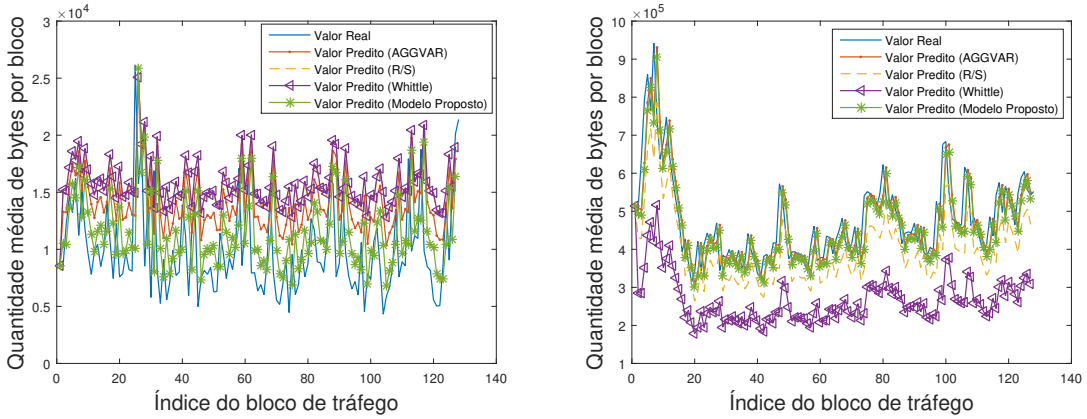
Tabela 3. Estimação do parâmetro de *Hurst*.

Séries de Tráfego	Modelos de Estimação do Parâmetro de *Hurst*			
	Aggvar	R/S	Whittle	Proposto
dec-pkt-1	0,8505	0,8857	0,7696	0,9200
dec-pkt-2	0,8785	0,8286	0,7731	0,9530
Waikato1	0,9250	0,9546	0,8229	0,9500
Waikato2	0,9955	0,9361	0,8094	0,9850

Analisando a Tabela 3 observa-se que o *H* calculado de forma analítica (modelo proposto) produziu valores coerentes quando comparados aos métodos AGGVAR, R/S e *Whittle*. O que garante ainda mais confiabilidade ao método proposto desenvolvido nesse trabalho.

7.4. Resultados com o DRA

Nesta subseção são apresentados os resultados de simulações utilizando as séries de tráfego dec-pkt-2 e Waikato2. As simulações foram conduzidas utilizando o Algoritmo Proposto, o DRA com diferentes estimadores para H e um esquema de taxa de transmissão fixa. Os desempenhos dos três métodos são avaliados em termos de capacidade de predição, utilização média do enlace e estimativas de taxa de perda e ocupação média do *buffer*.



(a) Predição da média do tráfego. Série - dec-pkt-2. (b) Predição da média do tráfego. Série - Waikato2.

Figura 4. Predição da média do tráfego para n blocos de tamanho m.

Para avaliar o desempenho desses métodos de predição, foi utilizado o conceito de Erro Quadrático Médio Normalizado (EQMN). A Tabela 4 apresenta os resultados de EQMN para o DRA e para o método proposto, referente aos resultados da Figura 4. É possível observar que o modelo proposto obteve o menor EQMN em comparação com os métodos de estimação (AGGVAR, R/S e *Whittle*) para as séries dec-pkt-2 e Waikato2.

Tabela 4. EQMN de predição do tráfego.

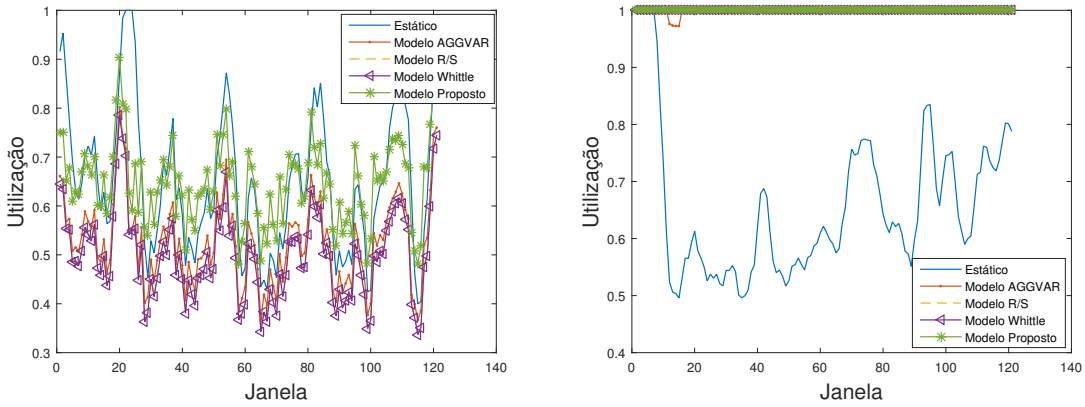
dec-pkt-2	EQMN de Predição
DRA (AGGVAR)	$1,4522\text{x}10^{-8}$
DRA (R/S)	$2,0884\text{x}10^{-8}$
DRA (Whittle)	$2,9877\text{x}10^{-8}$
DRA (Modelo Proposto)	$9,0766\text{x}10^{-9}$
Waikato2	EQMN de Predição
DRA (AGGVAR)	$3,6736\text{x}10^{-11}$
DRA (R/S)	$2,1522\text{x}10^{-11}$
DRA (R/S)	$1,5162\text{x}10^{-10}$
DRA (Modelo Proposto)	$1.8524\text{x}10^{-11}$

A interpretação do valor do Fator de Adequação Médio (AGF - *Average Goodness Factor*) é feita da seguinte forma: quanto maior o grau de superdimensionamento ou subdimensionamento do enlace, menor o valor de AGF. Quanto maior a acurácia do esquema de provisão de taxa de transmissão, mais próximo o valor de AGF será de 1.

Tabela 5. Resultados de utilização e AGF.

	dec-pkt-2		Waikato2	
Método	Utilização	AGF	Utilização	AGF
Estático	$0,6507$	$0,7037$	$0,6499$	$0,7059$
DRA (AGGVAR)	$0,5279$	$0,5835$	$0,9998$	$0,9034$
DRA (R/S)	$0,5014$	$0,5505$	$0,9976$	$0,9055$
DRA (Whittle)	$0,4843$	$0,5278$	$1,0000$	$0,9033$
DRA (Modelo Proposto)	$0,6630$	$0,7549$	$0,9900$	$0,9171$

Por meio da Tabela 5, observa-se que o algoritmo proposto não destoa dos demais modelos apresentando resultados próximos ou melhores que todos os métodos de estimação, destacando-se para o valor obtido da AGF para as duas séries de tráfego, onde o modelo proposto obteve o melhor resultado geral (mais próximo de 1) o que também pode ser observado por meio da Figura 5. Além disso salienta-se que para o valor obtido da utilização do enlace, o método proposto obteve o melhor valor para a série dec-pkt-2 e na série Waikato2 o valor de utilização ficou próximo dos métodos AGGVAR, R/S e *Whittle* acima do modelo estático.



(a) Utilização do enlace. Série - dec-pkt-2.

(b) Utilização do enlace. Série - Waikato2.

Figura 5. Utilização do enlace para as séries dec-pkt-2 e Waikato2.

O esquema de alocação estática de taxa de transmissão baseia-se na estimativa da menor taxa de transmissão necessária para o que o valor de perda média seja inferior ao valor de $\epsilon = 10^{-2}$. Além disso, foram aferidos os valores relativos à taxa de perda de *bytes* no enlace e a ocupação média do *buffer*. Por meio da Tabela 6, observa-se que não houve registro de perda para a série dec-pkt-2. Já analisando a séries de tráfego Waikato2 observa-se que houve perda de dados para todos os métodos, entretanto, vale ressaltar que o algoritmo proposto obteve um resultado para perda média e máxima inferior aos modelos AGGVAR, R/S e *Whittle*.

Tabela 6. Taxa de perda média (P_{medio}), máxima (P_{max}), e ocupação média do *buffer*.

	dec-pkt-2 $b = 5,0 \cdot 10^6 bytes$		
Método	P_{medio}	P_{max}	B_{medio}
Estático	$0,0000$	$0,0000$	$0,018$
DRA (AGGVAR)	$0,0000$	$0,0000$	$3,5269\text{x}10^{-4}$
DRA (R/S)	$0,0000$	$0,0000$	$2,7623\text{x}10^{-4}$
DRA (Whittle)	$0,0000$	$0,0000$	$2,3452\text{x}10^{-4}$
DRA (Modelo Proposto)	$0,0000$	$0,0000$	$9,8481\text{x}10^{-4}$
	Waikato2 $b = 5,0 \cdot 10^5 bytes$		
Método	P_{medio}	P_{max}	B_{medio}
Estático	$0,0033$	$0,1511$	$0,0360$
DRA (AGGVAR)	$0,1725$	$0,2836$	$0,9472$
DRA (R/S)	$0,1051$	$0,2236$	$0,8472$
DRA (Whittle)	$0,4979$	$0,5651$	$0,9999$
DRA (Modelo Proposto)	$0,0456$	$0,1541$	$0,6015$

8. Conclusão

Neste artigo, foi proposto uma equação para a função de autocorrelação baseada apenas nos parâmetros de um modelo multifractal para o processo de tráfego de rede. Através da equação proposta pode-se estimar o parâmetro de *Hurst* (*H*). Apesar dos diversos métodos existentes na literatura para estimar o valor de *H*, o método proposto estimou valores para *H* que quando utilizados em um algoritmo de controle dinâmico de taxas de transmissão resultou em melhores resultados de utilização dos recursos de rede quando comparado ao DRA. Dessa forma, conclui-se que o algoritmo proposto apresentado nesse artigo é eficiente, tanto para estimar a ACF de forma paramétrica, quanto para estimar o parâmetro de *Hurst* e alocar recursos dinamicamente para um servidor de rede. Comparações com outros métodos presentes na literatura comprovam os resultados analíticos obtidos.

Referências

Aguirre, L. A. (2004). *Introdução à Identificação de Sistemas – Técnicas Lineares e Não-Lineares Aplicadas a Sistemas Reais*. UFMG.

Anderson, D., Cleveland, W. S., and Xi, B. (2017). Multifractal and gaussian fractional sum–difference models for internet traffic. *AEU - International Journal of Electronics and Communications*, page 217–230.

Aquino, V. A. and Barria, J. A. (2006). Multiresolution fir neural-network-based learning algorithm applied to network traffic prediction. *IEEE Transactions on Systems, Man and Cybernetics-Part C: Applications and Reviews*, page 208 – 220.

Becchi, M. (2008). *From Poisson Processes to Self-Similarity: a Survey of Network Traffic Models*.

Cappe, O., Moulines, E., Pesquet, J.-C., Petropulu, A., and Yang, X. (2002). Long-range dependence and heavy-tail modeling for teletraffic data. *IEEE Signal Processing Magazine*, pages 14–27.

Dec (2000). The internet traffic archive. Último acesso em 01-10-2017.

Doty, K., Roy, S., and Pham, K. D. (2016). Dynamic resource allocation for military communications satellites with cognitive users. *Aerospace Conference, IEEE*.

Feldmann, A., Gilbert, A. C., and Willinger, W. (1998). Data networks as cascades: Investigating the multifractal nature of internet wan traffic. page 42–55.

Gospodinov, M. and Gospodinova, E. (2005). The graphical methods for estimating hurst parameter of self-similar network traffic. *International Conference on Computer Systems and Technologies*, pages 1–6.

Hayashi, F. (2000). *Econometrics*. Princeton University.

Karagiannis, T., Molle, M., and Faloutsos, M. (2006). Understanding the limitations of estimation methods for long-range dependence. pages 1–23.

Krishna, M. P., Gadre, V., and Desai, U. B. (2003). *Multifractal Based Network Traffic Modeling*. International Journal of Network Management.

Lee, I. W. C. and Fapojuwo, A. O. (2005). Stochastic processes for computer network traffic modeling. *Computer Communications*, page 1–23.

Liu, H.-H. and Hsu, P.-L. (2006). Design and simulation of adaptive fuzzy control on the traffic network. *SICE-ICASE. International Joint Conference*.

Mandelbrot, B. B. (March 1965). Self-similar error clusters in communication systems and the concept of conditional stationarity. *IEEE Transactions on Communications*, 13:71– 90.

Park, K. and Willinger, W. (2000). *Self-Similar Network Traffic and Performance Evaluation.* John Wiley Sons, Inc. New York, NY, USA, 1st edition.

Riedi, R., Crouse, M., Ribeiro, V., and Baraniuk, R. (1999). A multifractal wavelet model with application to network traffic. *IEEE Transaction on Information Theory*, page 992–1018.

Rocha, F. G. C. and Vieira, F. H. T. (2011). Modeling of mpeg-4 video traffic using a multifractal cascade with autoregressive multipliers. *IEEE Latin AmericaTransactions*, pages 860 – 867.

Trinh, N. C. and Miki, T. (1999). Dynamic resource allocation for self-similar trac in atm network. *IEEE*, 1:160–165.

Tuberquia-David, M., Vela-Vargas, F., Lópes-Chávez, H., and Hernández, C. (2016). A multifractal wavelet model for the generation of long-range dependency traffic traces with adjustable parameters. *Science Direct*, 62:373–384.

Vieira, F. H. T. and Lee, L. L. (2010). An admission control approach for multifractal network traffic flows using effective envelopes. *AEU - International Journal of Electronics and Communications*, page 629 – 639.

Wand (2011). Waikato university - computer science department - wand network research group. Último acesso em 01-10-2017.

Won, Y., Choi, M.-J., and Ki-Hong, J. W. (2003). *An approach for failure recognition in ip-based industrial control networks and systems*. Springer US.

Avaliação Comparativa de Tecnologias de Modulação para Redes 5G baseadas em Sistemas LTE

Dalton F. Souza[1], Vinicius C. M. Borges[2], Flávio H. T. Vieira[1], Diego C. Abrahão[1]

[1]Escola de Engenharia Elétrica, Mecânica e de Computação – Universidade Federal de Goiás (UFG) – Goiânia, GO, Brasil

[2]Instituto de Informática – Universidade Federal de Goiás (UFG) – Goiânia, GO, Brasil

daltonfoltran@gmail.com, vcmborges@gmail.com, flavio@eee.ufg.br, diegoabrahao@outlook.com

Abstract. *The growth in data traffic and in the number of connected devices has imposed challenging requirements to be followed during the development of the next generation of mobile wireless communication systems (also called 5G). In this sense, several technologies are being studied aiming to support this new demand. Among them, is the use of new modulations in substitution of traditional OFDM, existent on 4G. In this article, the application of the UFMC and F-OFDM modulations on downlink LTE is evaluated. For that, we compared performance considering parameters like relationship between BLER x SNR, throughput and spectral efficiency.*

Resumo. *O crescimento na demanda por tráfego de dados e no número de dispositivos conectados à internet tem imposto requisitos desafiadores a serem seguidos no desenvolvimento da próxima geração dos sistemas comunicação móvel sem fio (5G). Com isso, novas tecnologias estão sendo estudadas com o objetivo de atender essa demanda. Dentre elas, está a utilização de novas modulações em substituição à OFDM tradicional, existente no 4G. Neste artigo, avaliamos a aplicação no downlink LTE das modulações UFMC e F-OFDM, candidatas às redes 5G. Para tal, realizamos uma comparação de desempenho considerando parâmetros tais como, a relação BLER x SNR, vazão e eficiência espectral.*

1. Introdução

A demanda por tráfego de dados móveis tem crescido anualmente. De acordo com Cisco 2017, o volume de tráfego de dados móveis mensal cresceu de 4.4 exabytes ao final de 2015 para 7.2 exabytes em 2016, ou seja, um aumento de 63% no período de 1 ano. Comparando-se os últimos 5 anos, o aumento foi de 18 vezes. Paralelamente, quase meio bilhão (429 milhões) de dispositivos móveis foram adicionados à rede, totalizando 8 bilhões de conexões. Os maiores responsáveis por esse crescimento foram os smartphones, seguidos pelos dispositivos M2M [Cisco 2017].

Estima-se que, em 2021, o tráfego de dados móveis mensal alcançará 49 exabytes e haverá 1,5 dispositivos móveis conectados para cada habitante do planeta, totalizando 11,6 bilhões de dispositivos móveis [Cisco 2017].

Diante desse cenário, o sistema de comunicação móvel sem fio deverá ser capaz de prover conexão a uma quantidade massiva de usuários, atender a requisitos de altas taxas de transferência de dados e baixíssimos tempos de latência, o que é um grande desafio para qualquer tecnologia. Assim, uma das perguntas que é preciso responder é se os sistemas de comunicação existentes são capazes de suprir tais exigências para diferentes tipos de aplicações. Para responder a essa pergunta é preciso entender o estado atual das tecnologias de comunicação móveis sem fio.

Em 2008, o *3rd Generation Partnership Project* (3GPP) padronizou o LTE *(Long Term Evolution)* com taxa de dados teóricas de 300 Mbps de *downlink* e 75 Mbps de *uplink*. O tempo de latência mínima no plano do usuário é 5 ms. O *LTE Advanced Pro*, incorporou inúmeras novas tecnologias que também serão utilizadas na padronização do 5G, incluindo modulação 256-QAM, MIMO massivo, Internet das Coisas (IoT) e uso simultâneo de espectros licenciados e não licenciados. O *LTE Advanced Pro* é capaz de prover até 3 Gbps de *downlink* e 1,5 Gbps de *uplink* e 2ms de latência [Barreto et al 2016].

Apesar das constantes evoluções há algumas limitações presentes no LTE, tais como falta de suporte para um grande número de conexões simultâneas e tempo de latência relativamente alto, o que poderia impactar em aplicações em tempo real e suporte incompleto para conexão de diversos tipos de dispositivos (*smart meters*, veículos etc) [Barreto et al 2016].

Em 2015, o ITU *(International Telecomunication Union)* definiu alguns requisitos a serem preenchidos pela nova geração de comunicação móvel (5G), referenciada como *International Mobile Communications* (IMT-2020), tais como: taxa de dados de até 10 Gbps, taxa de dados experimentada pelo usuário de 100 Mbps a 1 Gbps e tempo de latência da ordem de 1 ms.

De acordo com o IMT-2020, algumas tendências que irão guiar a evolução tecnológica rumo ao 5G são: baixo tempo de latência, tanto para comunicação entre humanos quanto para comunicações entre máquinas; alta densidade de usuários; alta mobilidade; aplicações multimídia de alta qualidade; Internet das Coisas, tecnologias de softwarização das redes, dentre outras.

Atualmente, a modulação OFDM *(Orthogonal Frequency Division Multiplexing)* tem sido amplamente adotada nos padrões de comunicação sem fio, tendo como mais conhecidos o LTE e WiFi IEE 802.11. Não obstante suas vantagens, tais como robustez ao desvanecimento por multipercurso e a facilidade de implementação, a OFDM sofre com inúmeras desvantagens, dentre elas, alta taxa de pico de potência a média, do inglês *Power Peak to Average Power Ratio* (PAPR) e alto vazamento de potência no domínio da frequência para as bandas vizinhas ou *out-of-band emissions* (OOBE). Uma das principais consequências desse fato é a necessidade de se manter uma rigorosa sincronização entre o dispositivo transmissor e receptor com o objetivo de se manter a ortogonalidade entre as subportadoras.

Assim, para se atender às expectativas geradas por diferentes tipos de serviços em um cenário 5G faz-se necessário o desenvolvimento de novas tecnologias, tanto na arquitetura de rede atual quanto na interface aérea, também chamada de camada física. Neste sentido, várias modulações têm sido consideradas candidatas a serem adotadas como padrão para a nova geração de comunicação móvel sem fio, dentre elas a FBMC *(Filter Bank Multicarrier)*, F-OFDM *(Filtered – Orthogonal Frequency Division Multiplexing)* e a UFMC *(Universal Filtered Multicarrier)* [Abdoli et al 2015].

Nos estudos apresentados por Wunder et al 2014 a UFMC é admitida como uma generalização da F-OFDM. Segundo Wunder et al 2014, a UFMC é uma técnica que possui alta granularidade, capacidade de suprimir emissões OOBE, dispensa necessidade do prefixo cíclico e possui um comprimento de filtro menor que a F-OFDM, o que a torna interessante para aplicações IoT. Entretanto, não é apresentado uma avaliação do comportamento da BER com a variação da SNR, eficiência espectral e da vazão total alçançada em comparação a modulação atual, OFDM [Wunder et al 2014].

Em uma comparação entre F-OFDM e OFDM, Abdoli et al 2015 mostram que a F-OFDM possui uma menor emissão OOBE, BLER (Taxa de Erro de Bloco) similar e vazão total média superior a OFDM. Ficou demonstrado que a necessidade da banda de guarda para modulações de ordem mais altas e o uso do prefixo cíclico continuam incorrendo em uma perda de eficiência espectral. Apesar dos resultados serem promissores com relação a vazão total média, o trabalho não analisou o desempenho da modulação F-OFDM em uma rede baseada em sistemas LTE e não avaliou as modulações FBMC e UFMC [Abdoli et al 2015].

Ijaz et al 2016 analisaram e compararam as modulações FBMC, F-OFDM, W-OFDM, UFMC e GFDM sob a ótica do espectro de potência, BER, eficiência espectral, taxa de pico de potência a média (PAPR), impactos do atraso no tempo e frequência e complexidade computacional. Entretanto, não foram analisados aspectos referentes à eficiência energética, integração com MIMO massivo, estimativa do canal, equalização e vazão. Por fim, não foram feitas simulações aplicando-se as modulações ao enlace LTE [Ijaz et al 2016].

Gerzaguet et al 2017 realizou uma comparação entre as modulações OFDM, UFMC, FBMC e GFDM e avaliou a eficiência espectral, densidade espectral de potência, taxa de pico de potência a média (PAPR), robustez ao sincronismo na transmissão de *uplink* e complexidade computacional. A UFMC mostrou-se como uma opção interessante no que se refere a eficiência espectral. A FBMC e GFDM destacaram-se pela baixa emissão fora da banda com relevância para aplicações com necessidade proteção a interferências entre transmissões simultâneas. Entretanto, os autores em [Gerzaguet et al 2017] não realizaram implementação prática ou simulação do enlace LTE, assim como não consideraram parâmetros de vazão total e comportamento da BER com a variação da SNR.

Nos testes realizados por Doré et al 2017 as modulações OFDM, UFMC, e FBMC foram analisadas sob os aspectos de densidade espectral de potência, eficiência espectral, potência de pico a média e robustez a sincronismo na transmissão de multiusuários. A UFMC apresentou uma eficiência espectral semelhante a OFDM com uma maior robustez a um relaxamento das exigências de sincronismo, preservou a compatibilidade com algoritmos de MIMO e estimativa de canal. Já a FBMC apresentou dificuldades de adaptação a pacotes de tamanho pequeno, entretanto a ausência de banda de guarda proporciona um ganho de eficiência para aplicações de transmissão broadcast. Além disso, a FBMC apresentou a menor taxa de potência de pico a média dentre as tecnologias analisadas, entretanto a adaptação ao MIMO não se mostrou tão trivial quanto na OFDM. Restou, portanto, a comparação com a modulação F-OFDM e a análise das métricas de vazão total, comportamento da BER com a variação da SNR [Doré et al 2017].

Desta forma, percebe-se uma carência na avaliação comparativa das simulações OFDM, F-OFDM e UFMC aplicados a redes 5G baseadas em sistemas LTE. Por este motivo,

nosso principal objetivo neste trabalho é apresentar uma comparação entre OFDM, F-OFDM e UFMC como elemento integrante de uma rede 5G baseada no sistema LTE. Para isso, uma tabela comparativa e uma avaliação de desempenho por meio de simulação no Matlab® serão apresentadas neste artigo. As métricas avaliadas são o comportamento da BER com a variação da SNR, a vazão total média e a eficiência espectral.

Este artigo está divido conforme descrito a seguir; Seção 2: Modulações para o 5G, traz um resumo sobre os princípios, vantagens e desvantagens da OFDM e de algumas modulações candidatas ao 5G; Seção 3: Avaliação das tecnologias, na qual é feito um comparativo e avaliação das tecnologias avaliadas; Seção 4: Conclusões

2. Modulações para o 5G

Nesta seção é feito um breve resumo da modulação atualmente padronizada pelo 4G e algumas das modulações que mais fortemente estão sendo consideradas pela literatura como candidatas a serem adotadas na padronização do conjunto de tecnologias empregadas no 5G.

2.1 OFDM (Orthogonal Frequency Division Multiplexing)

Conforme mencionado anteriormente, o sistema LTE foi introduzido pelo 3GPP Release 8, em 2008, e utiliza a modulação OFDM na transmissão de *downlink*, o que permite maior liberdade no escalonamento de canal e na alocação de largura de banda flexível por meio da variação do número de subportadoras utilizadas na transmissão.

A OFDM é uma técnica de transmissão sobre múltiplas frequências, baseada na ideia de divisão do canal em várias subportadoras ortogonais entre si e com larguras de banda iguais. A ortogonalidade entre elas torna possível arranjá-las de tal maneira que suas larguras de banda se sobreponham e ainda se consiga recuperar a informação sem interferência entre subportadoras adjacentes.

Entre as principais vantagens do esquema de transmissão OFDM podemos citar: robustez aos efeitos de desvanecimento multipercurso na transmissão do sinal; simplicidade de implementação com mapeamento de bits feitos por uso de IFFT/FFT; flexibilidade de adaptação a diferentes disponibilidades de espectro ajustando o número de subportadoras à largura de banda; modulação adaptativa na qual é possível escolher a melhor modulação de acordo com o SINR de cada subcanal; fácil integração com sistemas MIMO *(Multiple-Input Multiple-Output)*.

O recurso físico básico do enlace de descida LTE pode ser visto como uma grade de recursos tempo-frequência, em que cada elemento de recurso corresponde a uma subportadora durante o intervalo de um símbolo OFDM. No domínio da frequência, as subportadoras do enlace de descida são agrupadas em blocos de recurso, onde cada um dos blocos de recurso é composto por 12 subportadoras consecutivas de 15 kHz durante uma janela de 0,5ms, totalizando uma largura de banda de 180 kHz, conforme ilustrado na Figura 1.

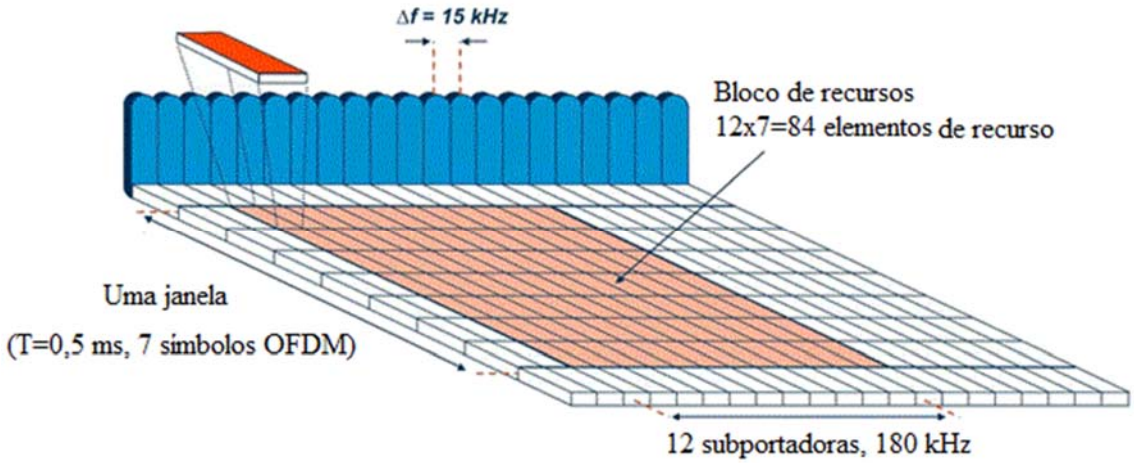


Figura 1. Estrutura de um bloco de recurso com prefixo cíclico normal.

No LTE, o canal compartilhado do *downlink* DLSCH *(Downlink Shared Channel)* é um canal de transporte usado para a transmissão de dados de usuário, para o controle dedicado e para informações do sistema do *downlink* e de camadas superiores específicas do usuário. No DLSCH está prevista a adição de bits de verificação de redundância cíclica CRC *(Cyclic Redundancy Check)* e a codificação de canal, baseada em códigos turbo. Já o canal compartilhado do *downlink* físico PDSCH *(Physical Downlink Shared Channel)* é o canal físico que transporta os dados codificados. No canal PDSCH do LTE *Release* 8, está previsto o embaralhamento (*scrambling*) e as modulações QPSK, 16-QAM e 64-QAM.

Baseado na estimativa da qualidade de canal dos usuários e tendo em vista os requisitos de parâmetros de QoS, os algoritmos de alocação de blocos de recurso tomam decisões de escalonamento para o atendimento às demandas dos usuários e operam com um esquema de modulação e codificação MCS *(Modulation and Coding Scheme*) definido.

Considerando a transmissão de *downlink* do sistema LTE com N blocos de escalonamentos (SB – *Scheduling Block*), onde cada SB é alocado com a mesma potência, assumindo que k usuários são servidos por uma estação base (BS - *Base Station*) e que a taxa de dados mínima exigida pelo k-ésimo usuário seja R_k bits/s. Um bloco de escalonamento (SB) com N_s símbolos OFDM consecutivos no domínio do tempo e N_{SC} subportadoras consecutivas no domínio da frequência. Considerando que existem sinais pilotos e de controle nos blocos de escalonamento, apenas ${N_{SC}}^{(d)}(s)$ das N_{SC} subportadoras podem ser utilizadas para transferência de dados no s-ésimo símbolo OFDM, onde $s \in \{1, 2,..., N_s\}$ e ${N_{SC}}^{(d)}(s) \leq N_{SC}$. Seja ${R_j}^{(c)}$ a taxa de código associada com o MCS $j \in \{1, 2, ..., J\}$, onde J é o número total de MCS suportados na transmissão, M_j é o tamanho da constelação do MCS j e T_S é a duração do símbolo OFDM. Então, a taxa de bits $r^{(j)}$ alcançada por um único SB com o MCS j é dada por:

$$r^{(j)} = \frac{{R_j}^{(c)} log_2(M_j)}{T_S N_S} \sum_{S=1}^{N_S} {N_{SC}}^{(d)}(s) \qquad (1)$$

Apesar de ter se consolidado como a técnica dominante utilizada nos atuais sistemas de comunicação sem fio, tais como LTE e WiFi 802.11, a OFDM possui algumas características indesejadas que impedem sua aplicação ao 5G. Dentre elas podemos citar

o fato de que, durante a transmissão, qualquer corrupção no sinal, por exemplo pela seletividade em frequência do canal de rádio, pode desfazer a estrutura em frequência dos recursos físicos e levar à interferência entre subportadoras.

Para lidar com esse problema, a OFDM faz a inserção de um intervalo de guarda, chamado prefixo cíclico, que consiste em copiar a última parte do símbolo OFDM e inserir no seu início, incrementando assim o tamanho do símbolo, com o objetivo de garantir que as réplicas atrasadas de um sinal OFDM sempre formarão um símbolo completo dentro da janela FFT. Entretanto, a aplicação desta técnica de inserção de prefixo cíclico causa uma perda de fração da potência do sinal utilizada na transmissão, pois parte da potência total é utilizada na transmissão do prefixo, e à perda na largura de banda, pois reduz a taxa total de transmissão sem redução da largura de banda.

Outra característica indesejada da OFDM é que seu sinal básico decai muito vagarosamente. Essa emissão fora de banda é causada pelo formato do sinal como um pulso retangular. Na prática, um simples filtro pode ser usado para suprimir essa emissão fora de banda, consumindo em torno de 10% da largura de banda total em forma de banda de guarda.

Além disso, a OFDM possui uma alta relação de potência de pico a média (PAPR), o que acaba por reduzir a eficiência e aumenta o consumo de energia nos dispositivos.

Por fim, para se obter um bom desempenho e fazer uso de seus benefícios, a OFDM possui a necessidade de se manter um rigoroso sincronismo entre o transmissor e o receptor. Para garantir esse sincronismo, o LTE faz uso de sinais de referência inseridos no sinal transmitido. O uso dessa sinalização pode se tornar cada vez mais sobrecarregada a medida que o número de usuários na rede aumenta, o que pode se tornar uma séria restrição ao 5G, onde se prevê cenários com conexões massivas e aplicações de comunicação orientadas à maquinas [Schaich e Wild 2014].

2.2 FBMC *(Filter Bank Multicarrier)*

Na FBMC, um conjunto de filtros é aplicado por subportadora. Assim, é possível se obter uma maior seletividade dentro do espectro de frequência e, com isso, minimizar a ocorrência de interferência interportadoras (ICI) adjacentes. Desta forma, a FBMC dispensa o uso de bandas de guarda entre subportadoras e de prefixo cíclico. Dentre as principais características da FBMC podemos citar o uso mais eficiente do espectro alocado, a possibilidade de se utilizar mesmo filtro no processamento do sinal recebido, uma sensibilidade de espectro com alta resolução e flexibilidade. Com isso, a FBMC tem sido considerada uma modulação promissora para as aplicações de Radio Cognitivo, Redes com Múltiplos Acessos, Comunicação em Linha Viva e MIMO [Tensubam e Singh 2014].

A FBMC possui duas variantes, denominadas FMT *(Filtered Multi-Tone)* e SMT *(Staggered Multi-Tone)* (Farhang-Boroujeny, 2011). Por possuir uma maior eficiência espectral e ser mais fortemente promovida [Wunder et al 2014] que o FMT nossa atenção se concentrará no SMT.

Para atingir uma eficiência espectral (tempo x frequência) igual a 1 o SMT precisa realizar o escalonamento dos símbolos multiportadoras no tempo e utilizar a modulação *Offset-QAM* (OQAM). Assim, os símbolos complexos a serem transmitidos são separados em parte real e imaginária e modulados em símbolos consecutivos. Com isso, é gerado uma

estrutura semelhante a um tabuleiro de xadrez com símbolos reais e imaginários alternando em tempo e frequência. Diferentemente do que ocorre na OFDM, a FBMC-SMT não é ortogonal no domínio complexo, pois cada subportadora introduz uma interferência na subportadora vizinha. Devido ao *Offset-QAM*, essa interferência é facilmente cancelada no receptor simplesmente ignorando-se a parte referente ao símbolo que não está transportando dados [Schaich e Wild 2014].

Conforme mencionado, a FBMC realiza uma filtragem por cada subportadora. Entretanto, a resposta em frequências desses filtros precisa ser bastante curta e, simultaneamente, possuir um comprimento muito longo quando comparados ao comprimento de um único símbolo (três ou quatro vezes). Assim, a área de aumento e diminuição dos dados durante uma rajada precisa ser aumentada. Esse fato, aliado à geração de interferência entre subportadoras vizinhas mencionadas anteriormente, tem ocasionado algumas dificuldades práticas na aplicação da FBMC [Schaich e Wild 2014].

2.3 F-OFDM *(Filtered Orthogonoal Frequency Division Multiplexing)*

Outra modulação candidata ao 5G é a F-OFDM. Como a operação de convolução no domínio do tempo é igual a uma operação de multiplicação no domínio da frequência, esse sistema pode ser visto como uma modulação multiportadora com janelas no domínio da frequência, em que a largura da janela é de fato a largura da banda do filtro, intencionalmente concebida para cobrir uma determinada sub-banda.

Nessa abordagem, a densidade espectral de potência de toda a sub-banda torna-se bem localizada sem alterar o impulso retangular do símbolo OFDM. Para esse fim, o sinal da OFDM é passado através de um filtro bem localizado em frequência, cuja largura de banda está próxima da largura de banda da sub-banda. Como resultado, apenas algumas subportadoras próximas as bordas são afetadas pelo filtro, uma vez que o filtro suprime os seus lóbulos laterais.

O sinal F-OFDM é composto de blocos de k sub-símbolos OFDM com comprimento $M + L_{cp}$, sendo L_{cp} o tamanho do prefixo cíclico e M a quantidade de subportadoras. Esses blocos são aplicados a um filtro f_i no domínio do tempo de comprimento L para cada sub-banda i, produzindo um sinal F-OFDM que pode ser definido para $k \in [-L_{cp}, KM + (K-1)L_{cp} + L - 2]$ é dado por:

$$s_{F-OFDM}[k] = \sum_{i=1}^{B} \sum_{n=0}^{K-1} \sum_{l=0}^{L-1} \sum_{m=0}^{M-1} c_{m,n}^{i} \, e^{j\frac{2\pi(k-l-nL_{cp})m}{M}} f_i[l] \qquad (2)$$

sendo $c_{m,n}^{i}$ o valor complexo do símbolo de uma subportadora m, n o sub-símbolo, i o índice da sub-banda e B a quantidade de sub-bandas do sinal F-OFDM [Luo et al 2016].

Nota-se que o intervalo de transição entre blocos consecutivos, resultante do filtro, desempenha o papel do intervalo de guarda e protege os símbolos transmitidos, desde que L seja maior ou igual a dispersão temporal (*delay spread*) máxima introduzida por um canal de múltiplos percursos. Em (3rd Generation Partnership Project (3GPP), 2016) é mostrado que o sinal F-OFDM, utilizando um filtro *sinc* (*Root-Raised Cosine window*) de ordem 256 e 512, pode-se dispensar a banda de guarda, mesmo para taxa de codificação de 64-QAM, em canais com *delay spread* de 30 ns e 1000 ns. Sem a banda de guarda é possível a utilização de uma maior quantidade de blocos de recurso em uma mesma portadora.

2.4 UFMC *(Universal Filtered Multicarrier)*

Enquanto a FBMC aplica o filtro por subportadora e a F-OFDM faz a filtragem de toda a banda de frequência, a UFMC a executa por sub-banda. Assim, a UFMC é considerada como uma generalização da F-OFDM e FBMC, conforme ilustrado na Figura 2.

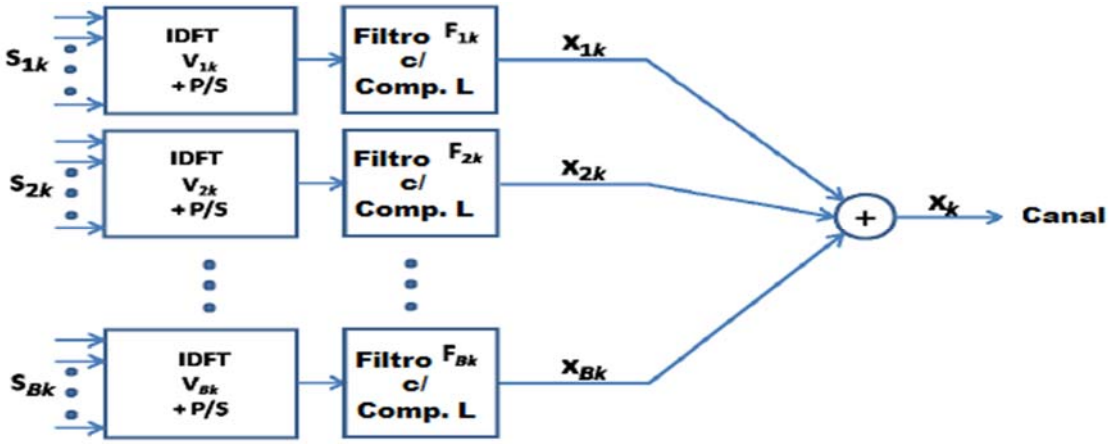


Figura 2. Arquitetura do transmissor UFMC [Schaich e Wild 2014].

No domínio do tempo, o vetor transmissão de um determinado símbolo multiportadoras de um usuário k é a superposição da largura de banda com o filtro de comprimento L e FFT de comprimento N, conforme (3.2):

$$\boldsymbol{X}_{k[(N+L-1)\times 1]} = \sum_{i=1}^{B} \boldsymbol{F}_{ik[(N+L-1)\times 1]}\, \boldsymbol{V}_{ik[N\times n_i]}\, \boldsymbol{S}_{ik[n_i\times 1]} \qquad (3)$$

Para cada sub-banda B, o enésimo símbolo QAM de índice i é convertido para o domínio do tempo por meio da matriz $\boldsymbol{V}_i$ da Transformada Inversa Rápida de Fourier (IDFT). $\boldsymbol{V}_i$ inclui as colunas relevantes da matriz inversa de Fourier de acordo com a posição da respectiva sub-banda dentro do espectro de frequência disponível. $\boldsymbol{F}_i$ é a matriz Toeplitz, composta pela resposta ao impulso do filtro. Pode-se notar que não há sobreposição entre símbolos adjacentes da UFMC.

A filtragem por bloco de recurso traz uma flexibilidade adicional e pode evitar a principal desvantagem da FBMC, ou seja, o comprimento do filtro. Com isso, a UFMC possui uma melhor resposta a rajada curta de dados e uma melhor operação em fragmentos de bandas. A supressão dos lobes laterais é feita entre blocos de recursos e não mais entre subportadoras. Assim é possível obter um símbolo com uma proteção suave a interferência intersimbólica, bem como uma maior robustez no acesso a múltiplos usuários que não estejam perfeitamente sincronizados no tempo. Desta forma, devido a possibilidade de se reduzir as bandas de guarda e evitar o uso do prefixo cíclico, a UFMC proporciona uma eficiência espectral superior a OFDM [Schaich e Wild 2014].

A UFMC pode ser utilizado com subportadoras de larguras de banda e comprimentos de filtros diferentes a cada sub-banda para cada usuário, possibilitando uma duração de símbolo personalizada. Essa característica permite que a UFMC seja altamente adaptativa a esquema de modulação e pode ser melhor ajustada a diferentes aspectos da

comunicação, tais como tempo de atraso, efeito Doopler e perfil do usuário [Schaich e Wild 2014].

A Tabela 1 apresenta um comparativo entre as modulações comentadas neste artigo referente a alguns dos principais aspectos relevantes para o 5G, tais como eficiência espectral, MIMO, compatibilidade com Rádio Cognitivo, cenários de aplicação 5G, eficiência energética, emissões OOBE.

Tabela 1. Comparativo modulações OFDM, FBMC, UFMC e F-OFDM

	Eficiência Espectral	**MIMO**	**Rádio Cognitivo**	**Cenários de Aplicação**	**Eficiência Energética**	**OOBE**
OFDM	Baixa	Sim	Não	Enhanced Mobile Broadband (eMBB)	Muito Baixa	Muito Alta
FBMC	Alta	Não trivial	Sim	Enhanced Mobile Broadband (eMBB)	Baixa	Muito Baixa
UFMC	Alta	Sim	Não	Enhanced Mobile Broadband (eMBB) Massive Machine Type Communications (mMTC)	Alta	Baixa
F-OFDM	Baixa	Sim	Sim	Enhanced Mobile Broadband (eMBB) Ultra-reliable and Low-latency Communications (uRLLC)	Alta	Média

Devido a sua característica de filtragem por subportadora, a FBMC possui uma emissão fora da banda (OOBE) muito baixa, tornando-a uma candidata a aplicações de Rádio Cognitivo em um cenário de evolução das redes móveis (eMBB). Entretanto, sua compatibilidade com MIMO não é trivial como a OFDM e o comprimento do filtro gera uma eficiência energética considerada baixa.

Dentre as modulações avaliadas, a UFMC é modulação com maior flexibilidade e maior eficiência energética devido a possibilidade de utilização de uma sub-banda bastante

estreita pelo fato de realizar uma filtragem por bloco de recurso. A compatibilidade com MIMO é outra vantagem. Por estes motivos a UFMC torna-se uma candidata a aplicação nos cenários de ambientes ultradensos e de evolução de redes móveis. Por outro lado, as emissões OOBE são maiores que na FBMC, o que a torna inadequada para aplicações de Rádio Cognitivo.

Por fim, a F-OFDM consegue manter uma emissão fora de banda (OOBE) melhor que a OFDM, com boa flexibilidade na alocação de recursos, possibilidade de aplicação a Rádio Cognitivo e fácil compatibilidade com MIMO. Apesar da redução da banda de guarda, a F-OFDM ainda não dispensa totalmente o uso do prefixo cíclico para esquemas de modulação de ordem mais alta, o que acaba por incorrer em uma perda da eficiência espectral. No caso da F-OFDM, o cenários mais promissores para aplicação são os de evolução de redes móveis (eMBB) e comunicações com latência ultra baixa (uRLLC).

3. Avaliação das Tecnologias

A fim de realizar uma análise comparativa entre as modulações estudadas, foram feitas simulações computacionais no Matlab® com as modulações F-OFDM e UFMC aplicadas ao *downlink* do sistema LTE considerando-se os canais PDSCH e DLSCH. Os resultados foram comparados com a modulação OFDM. Devido a redução da banda de guarda a níveis mínimos foi considerado um aumento de 10% na quantidade de blocos de recursos disponíveis para a modulação F-OFDM e um acréscimo de 16,67% para a modulação UFMC por dispensar também o prefixo cíclico.

As características do cenário da simulação são apresentadas na Tabela 2.

Tabela 2. Parâmetros de simulação adotados

	OFDM	F-OFDM	UFMC
Número de blocos de escalonamento	50	55	55
Prefixo Cíclico	Normal	Normal	Não
Banda de Guarda	Sim	Não	Não
Tempo de simulação	10 *subframes*		
Tamanho da subportadora	15kHz		
Duração de um slot	0,5ms		
Largura de banda	10 MHz		
Tamanho do *subframe*	1ms		
Frequência da portadora	2 GHz		
Modo de transmissão	Antena única		
Modelo de propagação multipercurso	EVA *(Extended Vehicular A model)*		

Com os resultados obtidos na simulação observa-se que as modulações OFDM e F-OFDM apresentaram um comportamento de BLER bastante parecidos em relação à variação do SNR, o que corrobora com os resultados obtidos por [Abdoli et al 2015]. A UFMC apresentou a maior taxa de BLER, sendo ligeiramente superior às anteriores. Como exemplo, a Figura 3 ilustra esse comportamento para o CQI 10. Percebe-se que,

quanto maior o SNR essa característica fica mais acentuada. Entretanto, tomando-se como referência o ponto em que a BLER é 10^{-1} essa diferença não chega a ser tão significativa a ponto de ser necessário diminuir a ordem do esquema de modulação QAM [Abdoli et al 2015].

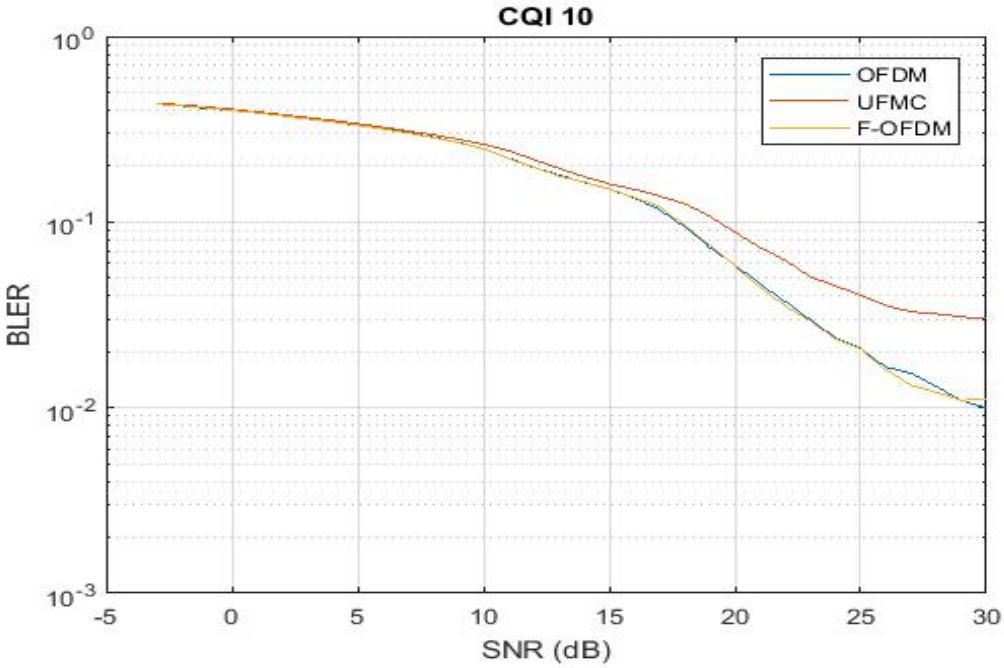


Figura 3. Comportamento da BLER com a variação do SNR para o CQI 10

Pelo fato da UFMC ter a possibilidade de utilizar mais blocos de recursos para a transmissão de dados é possível obter uma maior vazão ao custo de uma taxa de perda de blocos ligeiramente maior que as modulações OFDM e F-OFDM. A Tabela 3 mostra que a UFMC apresentou a maior taxa de vazão média dentre as modulações comparadas com ganho de 16% em relação a OFDM enquanto a F-OFDM, que havia apresentado uma curva de BLER similar a OFDM, apresentou uma vazão 10% maior.

Tabela 3. Vazão total média e eficiência espectral

	OFDM	F-OFDM	UFMC
Vazão total média	117,3 Mbps	129 Mbps	136,1 Mbps
Eficiência espectral	11,7 bps/Hz	12,9 bps/Hz	13,6 bps/Hz

Comparando-se a eficiência espectral das modulações analisadas, a UFMC apresentou a maior eficiência espectral dentre as modulações comparadas com 13,6 bps/Hz, convalidando com os resultados obtidos por Ijaz et al 2016. A F-OFDM apresentou uma eficiência espectral de 12,9 bps/Hz, com desempenho superior a OFDM e inferior ao UFMC. O pior desempenho, mais uma vez, ficou por conta da modulação OFDM, ressaltando uma de suas maiores deficiências apontadas na literatura, ou seja, baixa eficiência espectral.

Outra métrica comparada é a vazão alcançada por um único bloco de recurso calculada dado por (1). A Tabela 4 mostra a taxa de bits alcançada por cada CQI das modulações comparadas considerando o SNR no qual a BLER atinge o valor esperado de 10^{-1}.

Tabela 4. Vazão para cada CQI

Índice do CQI	Modulação	Taxa de código (x 1024)	OFDM	F-OFDM	UFMC
1	QPSK	78	7,8 Mbps	8,5 Mbps	9 Mbps
2	QPSK	120	12 Mbps	13,2 Mbps	13,9 Mbps
3	QPSK	193	19,3 Mbps	21,23 Mbps	22,4 Mbps
4	QPSK	308	30,8 Mbps	33,8 Mbps	35,7 Mbps
5	QPSK	449	44,9 Mbps	49,39 Mbps	52,1 Mbps
6	QPSK	602	60,2 Mbps	66,2 Mbps	69,8 Mbps
7	16-QAM	378	75,6 Mbps	83,2 Mbps	87,7 Mbps
8	16-QAM	490	98 Mbps	107,8 Mbps	113,7 Mbps
9	16-QAM	616	123,2 Mbps	135,5 Mbps	142,9 Mbps
10	64-QAM	466	139,8 Mbps	153,1 Mbps	162,2 Mbps
11	64-QAM	567	170,1 Mbps	187,1 Mbps	197,3 Mbps
12	64-QAM	666	199,8 Mbps	219,8 Mbps	231,8 Mbps
13	64-QAM	772	231,6 Mbps	254,8 Mbps	268,7 Mbps
14	64-QAM	873	261,9 Mbps	288,1 Mbps	303,8 Mbps
15	64-QAM	948	284,4 Mbps	312,8 Mbps	329,9 Mbps

4. Conclusões

Os desafios impostos pelos requisitos definidos para a futura geração de comunicação móvel sem fio trazem um cenário com uma rede bastante heterogênea, quantidade massiva de conexões, altas taxas de transferência, latência ultrabaixa, dentre outras. Por este motivo, uma modulação flexível e com alta eficiência espectral é tida como tecnologia chave para atender às novas demandas. Neste trabalho, apresentamos uma tabela comparativa com as principais formas de ondas candidatas ao 5G e simulamos as modulações F-OFDM e UFMC como parte integrante do enlace LTE. A F-OFDM apresentou um comportamento de BLER com a variação do SNR similar a OFDM com a vantagem de proporcionar uma maior flexibilidade na alocação dos recursos a diferentes tipos de usuários e serviços. Seu cenário de aplicação pode envolver tanto eMBB quanto uRLL. Já o UFMC apresentou um comportamento da BLER inferior a OFDM. Entretanto, essa desvantagem é compensada em outros parâmetros, tais como maior vazão, eficiência espectral e maior flexibilidade de alocação de recursos dentre as modulações avaliada. Como trabalhos futuros poderá ser considerado uma modulação adaptativa com estimativa do canal e a comparação com outras modulações, tais como GFDM e W-OFDM.

5. Referências

3rd Generation Partnership Project (3GPP). (2016). Link level evaluation results of waveform proposals. *3GPP TSG RAN WG1 Meeting #85bis R1-165426.* Nanjing, China.

André Noll Barreto, B. F. (2016). 5G – Wireless Communications for 2020. *JOURNAL OF COMMUNICATION AND INFORMATION SYSTEMS, 31*, 1-18.

Ayesha Ijaz, L. Z. (2016). *"Towards 5G Wireless Networks - A Physical Layer Perspective".* InTech.

Bidyalaxmi Devi Tensubam, S. S. (2014). A Review on FBMC: An Efficient Multicarrier Modulation System. *International Journal of Computer Applications, 98*, 06-09.

Cisco. (2017). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021.*

Farhang-Boroujeny, B. (2011). OFDM Versus Filter Bank Multicarrier. *IEEE Signal Processing Magazine*, 92-112.

Frank Schaich, T. W. (2014). Waveform contenders for 5G - OFDM vs. FBMC vs. UFMC. *IEEE ISCCSP.*

Gerhard Wunder, P. J. (2014). 5GNOW: Non-Orthogonal Asynchronous Waveforms for Future Mobile Applications. *IEEE Communications Magazine*, 97-105.

Javad Abdoli, Ming Jia, and Jianglei Ma. (2015). Filtered OFDM: A New Waveform for Future. *16th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC).* IEEE.

Jean Baptiste Doré, R. G. (2017). Waveform contenders for 5G: Description, analysis and comparison. *Physical Communication*.

Luo, F.-L., & Zhang, C. (2016). *Signal Processing for 5G: Algorithms and Implementations.* Chichester: John Wiley & Sons.

P.P.Vaidyanathan. (2001). Filter banks in digital communication. *IEEE.*

Robin Gerzaguet, N. B.-B.-B. (2017). The 5G candidate waveform race: a comparison of complexity and performance. *Journal on Wireless Communications and Networking*.

Algoritmo Adaptativo Baseado em Modelo Autorregressivo Transformado para Geração de Séries Sintéticas de Tráfego de Rede

Alisson A. Cardoso[1], **Flávio H. T. Vieira**[1]

[1]Escola de Engenharia Mecânica, Elétrica e Computação (EMC)
Universidade Federal de Goiás (UFG)
Goiânia, Goiás, Brasil

alsnac@gmail.com, flavio@emc.ufg.br

Resumo. *Neste trabalho, propomos um algoritmo para estimar adaptativamente os parâmetros de um modelo ARMA Transformado (TARMA - Transformed Autoregressive Moving Average) de tal forma a capturar a função de autocorrelação e a função de distribuição acumulada de séries de tráfego de redes reais. Comparamos o desempenho da modelagem proposta com os dos modelos autorregressivo e transformado tradicional em termos de média, variância, momentos, autocorrelações e funções de distribuição de probabilidade. Também realizamos simulações de um enlace de transmissão com servidor para verificar a eficiência da modelagem proposta em gerar e descrever séries de tráfego de redes. As simulações efetuadas mostram que os resultados da modelagem adaptativa proposta em geral são superiores aos dos modelos autorregressivos existentes na literatura, com uma complexidade computacional reduzida.*

Abstract. *In this paper, we propose an adaptive algorithm to estimate the Transformed Autoregressive Moving Average (TARMA) model parameters in order to capture the autocorrelation function and the cumulative density function of the desired network traffic trace. We compare the performance of the proposed model to those of the autoregressive model and of the transformed model in terms of mean, variance, moments, autorrelations and probability density function. A transmission link composed of a single server with buffer is also simulated, which proves the efficiency of the proposed model in describing real traffic traces. The simulations carried out in this work show that the adaptive model results are superior in general compared to those autoregressive models in the literature, presenting a lower computational complexity.*

1. Introdução

O estudo de modelos estatísticos para geração de dados sintéticos é de fundamental importância para a modelagem e simulação de fenômenos naturais. Alguns trabalhos destacam essa importância, como o estudo do tráfego de rede para o paradigma máquina-a-máquina [Laner et al. 2013], verificação de anomalias no tráfego de rede baseado em predição para identificação de anomalias e ataques maliciosos [Wang 2015], análise de desempenho de preditores de tráfego de rede na nuvem, o qual inclui o modelo ARMA (Autorregressive Moving Average) como preditor [Dalmazo et al. 2017].

O modelo autorregressivo de médias móveis transformado (TARMA-*Transformed* ARMA)) [Laner et al. 2014] consiste de uma variação do modelo autorregressivo de médias móveis (ARMA), o qual acrescenta a capacidade de modelar a função de autocorrelação (ACF) e a função de distribuição acumulada (FDA) da série temporal desejada, através de um filtro linear e uma transformação polinomial.

Os autores em [Laner et al. 2014] mostram que o modelo TARMA apresenta resultados em termos de Função de Distribuição de Probabilidade (FDP) e ACF próximos ao tráfego real. Entretanto, estes autores não apresentaram comparações com outros modelos da literatura, como por exemplo o modelo ARMA tradicional [Montgomery et al. 2015]. No modelo TARMA são efetuadas uma regressão polinomial e uma estimação dos parâmetros do modelo ARMA contendo todos os dados da série temporal. A regressão polinomial e a estimação de parâmetros podem demandar um alto custo computacional e inviabilizar seu uso em aplicações de tempo real. Devido a tais características, objetivamos neste trabalho propor um modelo adaptativo baseado no modelo TARMA.

Este trabalho está organizado da seguinte forma: Na seção 2, apresentamos o modelo TARMA. Na seção seguinte, apresentamos uma proposta de algoritmo para encontrar a função de distribuição acumulada aproximada. Na seção 4, apresentamos o algoritmo TARMA adaptativo proposto neste trabalho. Na sequência, apresentamos as simulações e os resultados obtidos. Por fim, concluímos este trabalho.

2. Modelo ARMA Transformado

O modelo TARMA (*Transformed* ARMA) é constituído por 3 processos independentes para modelar a série temporal. O primeiro é a geração de uma série temporal descorrelacionada no tempo, sendo uma variável aleatória normal com média 0 e variância 1. O segundo processo é um filtro linear invariante no tempo e o terceiro e último processo, é a transformação polinomial da saída do filtro linear. Ao final destas etapas espera-se gerar uma série sintética com as mesmas características da série real.

2.1. Transformação Polinomial

A transformação polinomial, última etapa realizada no algoritmo TARMA, realiza a conversão da Função de Distribuição Acumulada qualquer da variável $Y[k]$ na FDA da série real a ser sintetizada. Assim, podemos escrever a transformação polinomial como sendo [Allen 2014],

$$Z[k] = F_z^{-1}(F_y(Y[k])) = \mathfrak{p}_y(Y[k]) \tag{1}$$

onde F_z^{-1} é função inversa da FDA da variável Z e F_y a FDA da variável Y. Escrevendo a Equação 1 na forma polinomial, temos,

$$Z[k] = \sum_{i=0}^{P} p_i \cdot (Y[k])^i \tag{2}$$

Ainda de acordo com a Equação 1, para se realizar a transformação de FDA é necessário a obtenção da Função de Distribuição Acumulada da variável $Y[k]$, a qual representa um variável aleatória normal com média zero e variância 1 [Laner et al. 2014]. Assim, temos a seguinte expressão para $F_y(Y[k])$ [Montgomery et al. 2015]:

$$F_y(Y[k]) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{Y[k]}{\sqrt{2}}\right)\right] \tag{3}$$

3. Função da Distribuição Acumulada Inversa Discreta Adaptativa

Nesta seção, propomos um algoritmo para calcular adaptativamente a Função de Distribuição Acumulada aproximada através do cálculo do histograma da série a ser sintetizada.

Seja $Z[k]$ a série temporal normalizada a ser sintetizada, contida no intervalo $[0, 1]$. Ao fracionarmos esse intervalo em N partes iguais, obtemos,

$$[0,1] = \left[0, \frac{1}{N}\right), \left[\frac{1}{N}, \frac{2}{N}\right), \ldots, \left[\frac{N-1}{N}, \frac{N}{N}\right] \tag{4}$$

Com o intervalo dividido, podemos calcular a quantidade de ocorrências de valores da série temporal Z nestes intervalos, ou seja,

$$\mathcal{V}_Z = [v_1, \ \ v_2, \ \ \ldots, \ \ v_N] \tag{5}$$

onde, v_i representa quantos valores da série que se encontram no intervalo $[(i-1)/N, i/N]$. Como a probabilidade de um evento ocorrer pode ser vista como a quantidade de ocorrência de um determinado evento dividido pela quantidade total de ocorrência do conjunto universo em análise, podemos obter através da divisão do vetor $\mathcal{V}_Z$ pelo tamanho total da série temporal, uma aproximação da Função de Distribuição de Probabilidade, com $N \to \infty$.

Através de $\mathcal{V}_Z$, obtemos o histograma acumulado f_k, como sendo,

$$f_t = \sum_{i=0}^{t} v_i, \quad t = 1, \ldots, N \tag{6}$$

Dividindo a Equação 6 pela quantidade total de amostras, obtemos a Função de Distribuição Acumulada, ou seja,

$$\hat{\mathcal{F}} = \left(\frac{1}{L}\right) \cdot [f_1, \ \ f_2, \ \ \ldots, \ \ f_N] \tag{7}$$

onde L é a quantidade total de amostras da série temporal utilizadas para realizar o histograma.

Uma vez obtida a Equação 7, podemos formalizar o **Algoritmo da FDA Inversa Adaptativa**, como se segue,

- **Passo 1**: Inicialize todos os valores do vetor $\mathcal{V}$ iguais a zeros, $v_i = 0$ para $i = 1, \ldots, N$.
- **Passo 2**: Inicialize o vetor $\hat{\mathcal{F}}$, de tamanho N, sendo igual a zero, $f_i = 0$ para $i = 1, \ldots, N$.
- **Passo 3**: Inicialize a quantidade de amostras utilizadas L como sendo igual a zero, $L = 0$.
- **Passo 4**: Encontre o intervalo no qual a amostra $Z[k]$ se encontra.

$$r = \left\lfloor \frac{Z[k]}{1/N} \right\rfloor \tag{8}$$

- **Passo 5**: Atualize a quantidade de ocorrência do intervalo no histograma $\mathcal{V}_Z$,

$$v_r = v_r + 1 \tag{9}$$

- **Passo 6**: Atualize o histograma acumulado, através da seguinte equação:

$$f_i = f_i + 1 \qquad i = r, \ldots, N \tag{10}$$

- **Passo 7**: Atualize a quantidade de amostra utilizadas $L = L + 1$.
- **Passo 8**: Calcule uma estimativa (Ver Apêndice 1), para a FDA inversa, através da seguinte equação:

$$F_z^{-1}(d) = (e-1) + \frac{(L \cdot d - f_{e-1})}{(f_e - f_{e-1})} \tag{11}$$

onde d é o valor para o qual se deseja avaliar a função e e o valor inteiro da posição do intervalo para o qual d está contido.

- **Passo 9**: Retorne ao Passo 4 na próxima iteração, até o fim de todas iterações.

Com o algoritmo da FDA Inversa Adaptativa obtemos F_z^{-1} com poucos cálculos computacionais, sendo a complexidade computacional deste algoritmo da ordem de $\mathcal{O}(N)$ para cada iteração.

4. Modelo TARMA Adaptativo

O modelo autorregressivo de médias móveis, leva este nome por considerar como saída a média das entradas anteriores (médias móveis), uma constante μ e médias das saídas do modelo nos instantes anteriores (autorregressivo). Matematicamente, podemos representar o modelo ARMA pela seguinte equação:

$$\begin{aligned} y[t] + a_1 \cdot y[t-1] + \ldots + a_n \cdot y[t-n] = \mu + \\ b_1 \cdot x[t] + \ldots + b_m \cdot x[t-m] \end{aligned} \tag{12}$$

ou na forma vetorial,

$$AY = \mu + BX \tag{13}$$

Nesta seção, propomos um algoritmo para estimação dos parâmetros do Modelo TARMA. Para tal, inicialmente apresentamos uma equação para a média de um modelo ARMA.

4.1. Média do Modelo ARMA

Aplicando o operador esperança matemática (média) na Equação 12, obtemos:

$$\begin{aligned} &\mathbb{E}[y[t]] + a_1 \cdot \mathbb{E}[y[t-1]] + \ldots + a_n \cdot \mathbb{E}[y[t-n]] \\ &= \mathbb{E}[\mu] + b_1 \cdot \mathbb{E}[x[t]] + \ldots + b_m \cdot \mathbb{E}[x[t-m]] \end{aligned} \tag{14}$$

Considerando um filtro invariante no tempo temos $\mathbb{E}[y[t]] = \mathbb{E}[y[t-1]] = \ldots$. Sabendo que $X[k]$ representa um variável aleatória normal com média zero, logo, $\mathbb{E}[x[t]] = \mathbb{E}[x[t-1]] = \ldots = 0$. Reescrevendo a Equação 15, obtemos a expressão da média do modelo ARMA,

$$\mathbb{E}[y[t]] = \frac{\mu}{(1 + a_1 + \ldots + a_n)} \tag{15}$$

Neste trabalho, com o objetivo de se obter uma variável normal $Y[k]$ com média igual a zero, fazemos $\mu = 0$.

4.2. Estimação Adaptativa dos Parâmetros ARMA

Os autores em [Laner et al. 2014] apresentam algoritmos de regressão polinomial para estimação dos parâmetros do modelo ARMA utilizando a quantidade total de amostras do tráfego de rede real. Neste trabalho, com o objetivo de estimar adaptativamente os parâmetros do modelo ARMA, empregamos o algoritmo RLS (*Recursive Least Squares*) [Farhang-Boroujeny 2013].

Reescrevendo a Equação 12, temos:

$$y[t] = -a_1 \cdot y[t-1] - \ldots - a_n \cdot y[t-n] + \\ b_1 \cdot x[t] + \ldots + b_m \cdot x[t-m] \tag{16}$$

Nota-se, que a Equação 16 possui valores conhecidos (Y e X) e valores que desejamos descobrir (A e B) para obtermos a resposta desejada $y[t]$. Reescrevendo os parâmetros da Equação 16, temos,

$$\theta^T = [a_1, \ \ldots, \ a_n, \ b_1, \ \ldots, \ , b_m] \tag{17}$$

e

$$\varphi^T[t] = [-y[t-1], \ldots, -y[t-n], \ x[t], \ldots, \ x[t-m]] \tag{18}$$

Assim, obtemos a equação para a saída do modelo, ou seja,

$$y[k] = \varphi^T[t]\theta \tag{19}$$

Obtidos φ^T e $y[k]$, pelas Equações 18 e 19, aplicamos o algoritmo RLS para obter os valores dos parâmetros do modelo ARMA através da Equação 17.

4.3. Obtenção dos Parâmetros da Transformação Polinomial

Podemos aplicar o algoritmo RLS para obter também os parâmetros da transformação polinomial. A seguir, descrevemos como isso pode ser feito.

Ao expandirmos a Equação 2, obtemos:

$$Z[k] = p_0 \cdot (Y[k])^0 + p_1 \cdot (Y[k])^1 + \ldots + p_P \cdot (Y[k])^P \tag{20}$$

A Equação 20 possui valores conhecidos (Y e Z) e valores que desejamos descobrir (p_i para $i = 0, \ldots, P$) para obtermos a resposta desejada $Z[t]$. Reescrevendo os parâmetros da Equação 20, temos,

$$\theta_p^T = [p_0, \ \ldots, \ p_P] \tag{21}$$

e

$$\varphi_p^T[t] = [(Y[k])^0, \ (Y[k])^1, \ \ldots, \ (Y[k])^P] \tag{22}$$

Assim, obtemos a equação para a resposta estimada, ou seja,

$$\hat{Z}[n] = \varphi_p^T[t]\theta_p. \tag{23}$$

Através do Algorimo da FDA inversa Adaptativa e da Equação 3, obtemos a resposta desejada,

$$Z[k] = F_z^{-1}(F_y(Y[k])) \tag{24}$$

Com as Equações 21, 22 e 24, podemos aplicar o algoritmo RLS e a cada iteração obteremos a estimativa dos parâmetros do polinômio $\mathfrak{p}_y$ através da Equação 21.

Utilizando o algoritmo para obtenção do polinômio através do algoritmo da FDA Inversa Adaptativa e da regressão polinomial pelo algoritmo RLS, podemos formalizar o **algoritmo TARMA Adaptativo**.

- **Passo 1**: Efetuar o algoritmo RLS para obtenção dos parâmetros do filtro adaptativo ARMA cuja complexidade computacional tem ordem $\mathcal{O}(n^2)$ [Farhang-Boroujeny 2013].
- **Passo 2**: Efetuar o algoritmo da FDA Inversa Adaptativa para obter a inversa da FDA do fluxo de tráfego de rede Z cuja complexidade computacional tem ordem $\mathcal{O}(N)$.
- **Passo 3**: Efetuar o algoritmo RLS para obtenção dos parâmetros do polinômio através da Equação 3 e da inversa obtida no Passo 2. A complexidade computacional do algoritmo RLS tem ordem $\mathcal{O}(n^2)$ [Farhang-Boroujeny 2013].
- **Passo 4**: Gerar uma amostra da variável Normal $\mathcal{N}(0, 1)$, efetuar a filtragem e a transformação polinomial para gerar amostras do tráfego sintético cuja complexidade computacional tem ordem $\mathcal{O}(P)$.
- **Passo 5**: Retornar ao Passo 1, na próxima iteração, até o fim de todas iterações.

O algoritmo TARMA Adaptativo apresenta uma complexidade computacional de ordem $\mathcal{O}(2n^2 + N + P) = \mathcal{O}(n^2)$ para cada iteração, onde n é a ordem do filtro utilizado no algoritmo RLS. Na Figura 1, o fluxograma do algoritmo TARMA Adaptativo é apresentado.

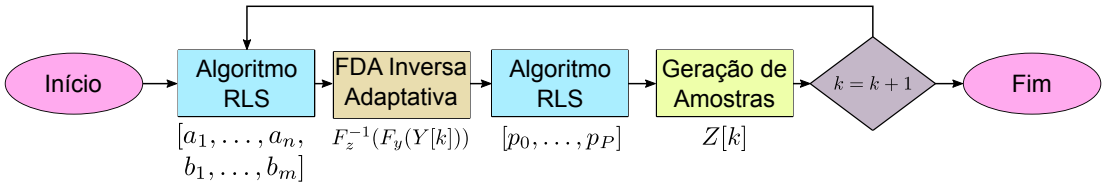


Figura 1. Fluxograma do Algoritmo TARMA adaptativo.

Na sequência, apresentamos os resultados obtidos dos modelos considerados para séries de tráfego reais e descrevemos como foram realizadas as simulações

5. Simulações e Resultados

As simulações foram realizadas levando em conta os modelos ARMA, TARMA e o modelo adaptativo TARMA proposto. As séries de tráfego utilizadas neste trabalho foram o tráfego de rede TCP/IP Waikato VIII 20110520-000000-0 [WITS 2014] coletado na comunicação entre a Universidade de Waikato e o resto do mundo com escala de agregação de $10ms$ e o tráfego de rede SNU20100315 [Kim 2011] obtido na rede sem fio WiMAX da Universidade Nacional de Seoul. Neste trabalho, nomearemos esses tráfegos de rede como Waikato e SNU, respectivamente.

Na comparação dos algoritmos, avaliamos os resultados em termos de média, variâncias, momentos até a quarta ordem, função de distribuição de probabilidade, autocorrelação normalizada e relação pico/média. Para comparar a FDP e a ACF das séries sintéticas em relação ao tráfego real, calculamos o erro quadrático médio (EQM).

Buscando-se obter o valor ótimo para a variável N do algoritmo da FDA Inversa Adaptativo, simulações para encontrar o EQM da FDA foram realizadas com N variando de 2 até 10000 como pode se verificar na Figura 2. Pode-se notar que os valores do EQM da FDA apresenta um decaimento em comparação ao aumento da quantidade N de intervalos, os quais, a partir do valor de N=100 não apresenta um significante decaimento. Com base neste comportamento de decaimento de EQM a quantidade de intervalos foi de $N = 100$.

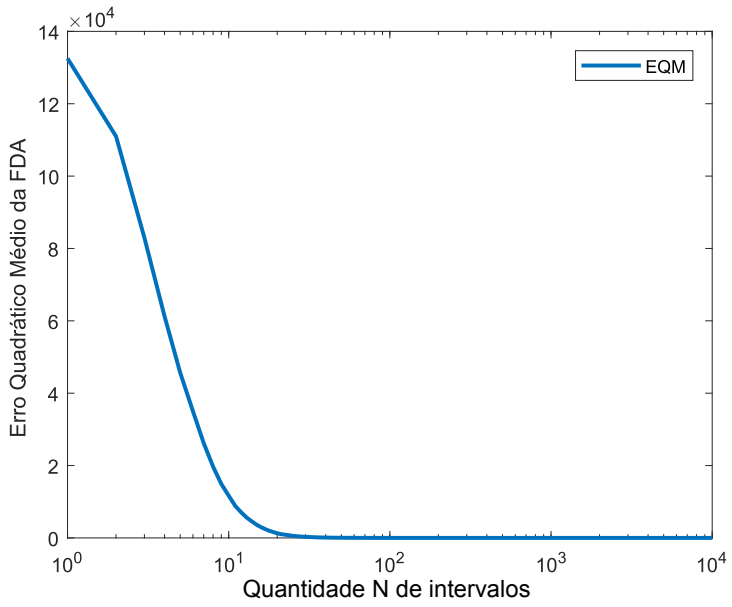


Figura 2. Erro Quadrático Médio da FDA em relação à quantidade N de intervalos.

A ordem P do Polinômio influencia diretamente o EQM da FDA, podendo impactar o desempenho da geração de amostras sintéticas do tráfego de rede [Laner et al. 2014]. Um polinômio com poucos parâmetros pode não ser suficiente para representar a FDA do tráfego de rede. Enquanto para valores grandes de P o modelo também pode não representar adequadamente a FDA devido ao surgimento de oscilações [Laner et al. 2014]. Assim, faz se necessário obter o valor P para o qual o algoritmo apresenta o menor EQM. Na Figura 3, pode-se verificar o EQM da FDA em relação à P. Nota-se que o menor EQM foi obtido para P igual a 10.

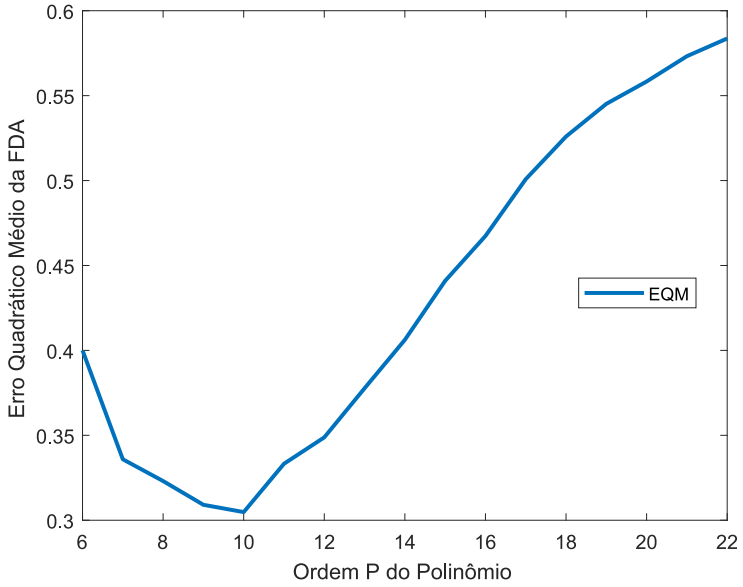


Figura 3. Erro Quadrático Médio da FDA em relação à ordem P do Polinômio, com N=100.

Na Figura 4, visualizamos os resultados obtidos para a FDP dos tráfegos de rede agregados em $10ms$. Os resultados indicam que todas as séries apresentaram maiores valores para a FDP no intervalo entre 0 e 0,4, com destaque para a FDP do modelo ARMA, o qual apresentou os valores mais dispersos em comparação à FDP do tráfego real. Destaca-se que as FDP do modelo TARMA e do modelo proposto mostram valores de probabilidade próximos ao desejado (do tráfego real).

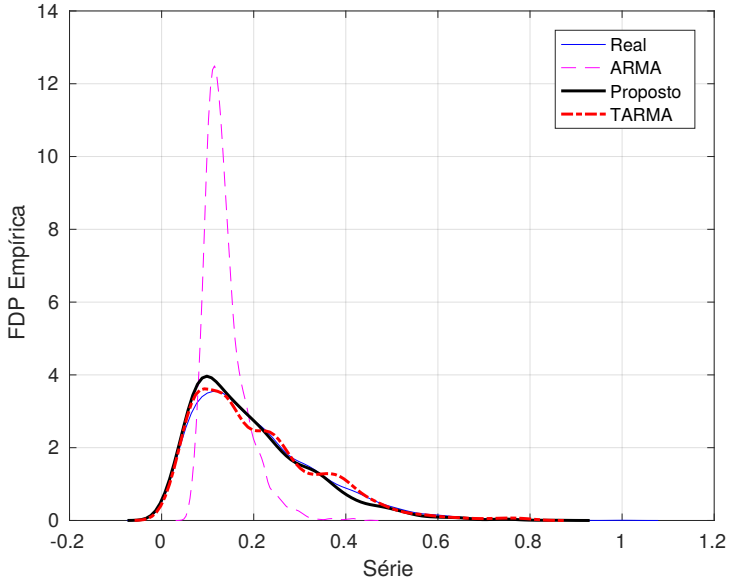


Figura 4. Função de Distribuição de Probabilidade para a Série Waikato, agregada na escala de 10ms.

Os resultados para a função de autocorrelação obtidos para o tráfego Waikato agregado na escala de $10ms$ podem ser visualizados na Figura 5. Nota-se que assim como os resultados da FDP, o modelo ARMA apresentou valores mais dispersos do desejado em comparação ao TARMA e o proposto. Destaca-se que para valores de atraso maiores que 20, o modelo proposto apresentou valores mais próximos da ACF da série real.

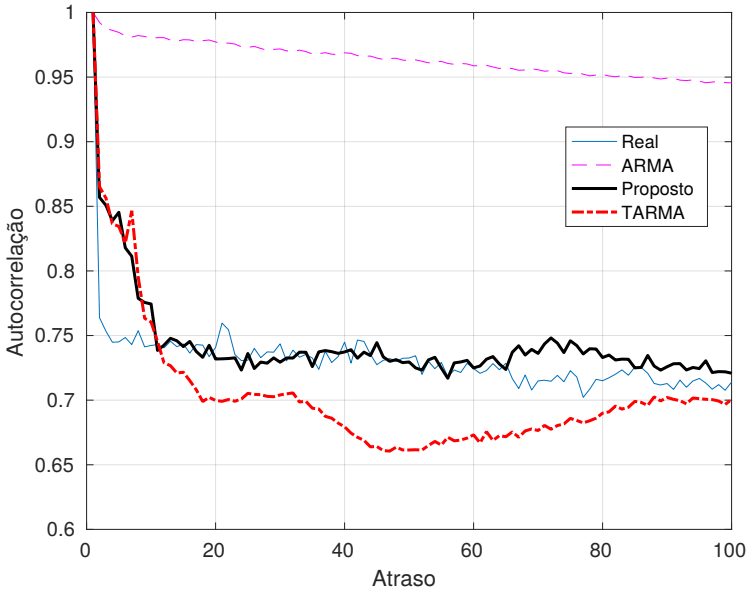


Figura 5. Autocorrelação com atraso até 100 amostras para a Série Waikato, agregada na escala de 10ms.

A Tabela 1 mostra os resultados estatísticos das séries sintéticas e real na escala de $10ms$. Os valores mais próximos em relação ao tráfego real para as estatísticas de média, de variância e de momentos foram obtidos pelo modelo TARMA, seguidos pelo modelo Proposto e pelo modelo ARMA. Destaca-se, que apesar do algoritmo TARMA utilizar todos os dados do fluxo de rede real para estimar seus parâmetros, o mesmo apresentou valores estatísticos semelhantes aos do algoritmo TARMA adaptativo.

Tabela 1. Dados Estatísticos para a Série Waikato, agregada na escala de 10ms.

	Real	ARMA	TARMA	Proposto
Média	0,2083	0,1572	0,2076	0,1938
Variância	0,0188	0,0051	0,0187	0,0162
2º Momento	0,0622	0,0409	0,0618	0,0537
3º Momento	0,0239	0,0076	0,0234	0,0191
4º Momento	0,0111	0,0024	0,0106	0,0081
EQM FDP	0	8,1634	0,0938	0,0646
EQM ACF	0	4.284,5	1.471,9	309,3
Pico/Média	4,7999	3,3037	3,8614	4,4148

Os resultados do erro quadrático médio da FDP e da ACF na Tabela 1 indicam valores mais próximos do desejado pelo algoritmo TARMA proposto, seguidos do TARMA e do ARMA.

Os resultados obtidos para a relação Pico/Média na Tabela 1 indicam valores próximos ao do tráfego de rede real para o modelo TARMA adaptativo, seguidos dos modelos TARMA e do ARMA.

5.1. Simulação de Enlace de Transmissão com *Buffer*

Para verificar se o comportamento das séries sintéticas é semelhante ao da série de tráfego real, realizamos a simulação de um enlace de transmissão de um servidor simples, onde o tráfego de rede é dado como entrada do servidor.

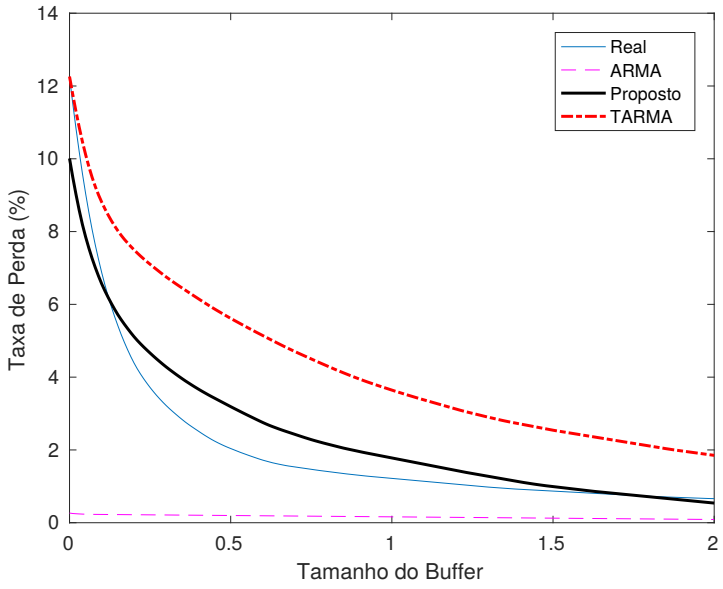


Figura 6. Taxa de Perda (%) no *buffer* em relação ao tamanho máximo da fila para o tráfego Waikato com escala de agregação de 10ms e $\eta = 0.6$

A Figura 6 apresenta os resultados para a Taxa de Perda (%) em relação à variação do tamanho máximo do *buffer* B, onde quanto maior o valor de B menor será a Taxa de Perda, explicada pela maior capacidade de armazenamento de dados. Nota-se que os modelos TARMA e o modelo proposto apresentaram valores semelhantes aos do tráfego de rede real, enquanto os valores mais dispersos foi obtido pelo modelo ARMA. Observa-se que o modelo proposto produziu valores mais próximos do real.

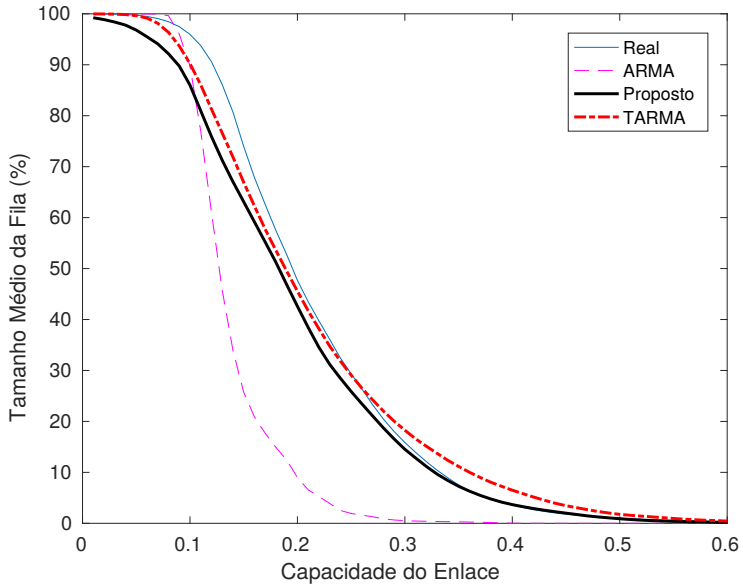


Figura 7. Tamanho Médio da fila no *buffer* em relação à capacidade de transmissão do enlace η para o tráfego Waikato com escala de agregação de 10ms e $B = 0.6$.

A Figura 7 apresenta os resultados para o Tamanho Médio da Fila (%) em relação à variação da capacidade do enlace de saída η. O comportamento apresentado pelo modelo proposto e pelo modelo TARMA foram os mais próximos da série real, enquanto o modelo ARMA apresentou valores inferiores para quase todos os valores de capacidade.

Tabela 2. Dados Estatísticos para a Série SNU agregada na escala de 100ms.

	Real	ARMA	TARMA	Proposto
Média	0,6352	0,6016	0,6398	0,6356
Variância	0,0231	0,0125	0,0217	0,0142
2º Momento	0,4266	0,3744	0,4310	0,4182
3º Momento	0,2949	0,2355	0,2818	0,2989
4º Momento	0,2088	0,1495	0,2122	0,1941
EQM FDP	0	3,2525	0,6019	0,1571
EQM ACF	0	1.507,2	1.502,2	1.333,2
Pico/Média	1.5742	1,4194	1,4194	1,4880

Os resultados estatísticos obtidos para o tráfego de rede SNU podem ser visualizados na Tabela 2. Os valores mais próximos de média, 4º momento e Pico/Média em relação ao tráfego real foram obtidos pelo algoritmo proposto, seguidos pelo algoritmo TARMA. Os valores mais próximos do desejado de variância, 2º e 3º momentos foram

obtidos pelo modelo TARMA. Nota-se que os valores mais dispersos foram obtidos pelo modelo ARMA. Destaca-se os menores erros quadráticos da FDP e ACF obtidos pelo modelo adaptativo.

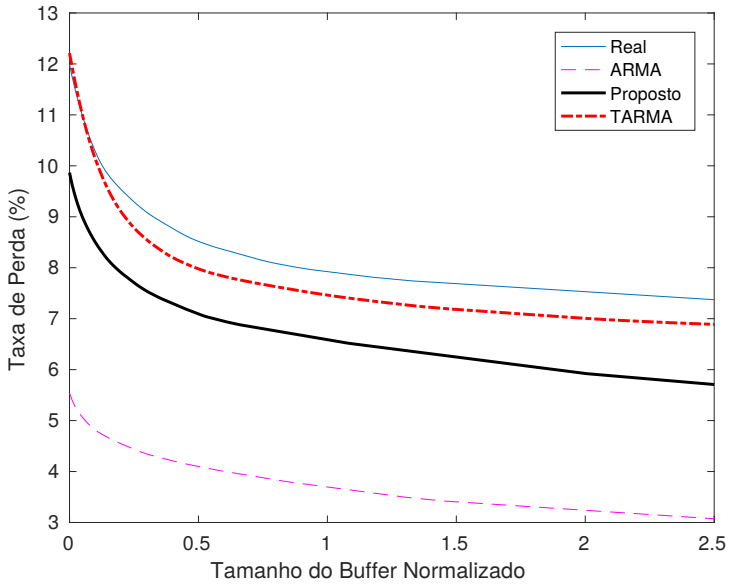


Figura 8. Taxa de Perda no *buffer* em relação ao tamanho do *Buffer* para o tráfego SNU com escala de agregação de 10ms e $\eta = 0.6$.

Na Figura 8 podemos visualizar os resultados da Taxa de Perda para o tráfego SNU. Observam-se comportamentos semelhantes obtidos pelos modelos TARMA e proposto, com valores mais distantes do real obtidos pelo modelo ARMA.

Ao analisarmos os resultados para a simulação do enlace de transmissão, notamos que de forma geral, o modelo proposto apresenta comportamento semelhante aos fluxos de tráfegos reais, indicando que as amostras sintéticas podem ser utilizadas em simulações para representar as séries desejadas.

Realizamos simulações com várias outras séries de tráfego de rede onde constatamos desempenho semelhante dos modelos considerados neste trabalho para essas séries. Entre essas séries de tráfego estão: tráfego de rede da *Bellcore* (*Bellcore-Aug89*) [SIGCOMM 1993], tráfego de rede da *Digital Equipment Corporation* (*Dec-Pkt-2*) [SIGCOMM 1993], etc.

6. Conclusão

Neste trabalho, propomos um algoritmo para estimar adaptativamente os parâmetros do modelo TARMA para geração de séries temporais sintéticas de forma a capturar a autocorrelação e função de distribuição acumulada independentemente de séries de tráfego reais.

Os resultados, de forma geral, mostraram que os modelos TARMA e o modelo TARMA adaptativo proposto apresentaram melhor desempenho para modelar a função de autocorrelação e função de distribuição de probabilidade em comparação ao modelo ARMA.

Observa-se na simulação do enlace de transmissão que as séries sintéticas do modelo TARMA e do modelo proposto apresentaram resultados semelhantes aos do tráfego de rede real, com alguns valores mais próximos do desejado para o modelo proposto. Destaca-se que o modelo TARMA necessita de todos as amostras da série real, enquanto o modelo proposto é realizado adaptativamente.

Para trabalhos futuros, objetivamos avaliar o uso de outros algoritmos para obtenção do polinômio e do filtro linear, com menores complexidades computacionais.

Referências

Allen, A. O. (2014). *Probability, statistics, and queueing theory*. Academic Press.

Dalmazo, B. L., Vilela, J. P., and Curado, M. (2017). Performance analysis of network traffic predictors in the cloud. *Journal of Network and Systems Management*, 25(2):290–320.

Farhang-Boroujeny, B. (2013). *Adaptive filters: theory and applications*. John Wiley & Sons.

Kim, S. (2011). Dataset of bittorrent traffic on korea telecom's mobile wimax network. http://crawdad.cs.dartmouth.edu/snu/ bittorrent/.

Laner, M., Svoboda, P., Nikaein, N., and Rupp, M. (2013). Traffic models for machine type communications. In *ISWCS 2013; The Tenth International Symposium on Wireless Communication Systems*, pages 1–5.

Laner, M., Svoboda, P., and Rupp, M. (2014). Parsimonious network traffic modeling by transformed arma models. *IEEE Access*, 2:40–55.

Montgomery, D. C., Jennings, C. L., and Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.

SIGCOMM, A. (1993). The acm/sigcomm internet trafficarchive. http://ita.ee.lbl.gov.

Wang, H. (2015). Anomaly detection of network traffic based on prediction and self-adaptive threshold. *International Journal of Future Generation Communication and Networking*, 8(6):205–214.

WITS (2014). Wits: Waikato internet traffic storage from university of waikato. http://wand.net.nz/wits/waikato/8/.

Apêndice 1: Prova da Equação (11)

Seja o vetor $[f_1, \ f_2, \ \ldots, \ f_N]$, o histograma acumulado e d o valor da FDA para o qual se deseja obter a inversa. Podemos estimar o valor de F_z^{-1}, através da média ponderada dos valores dos extremos no intervalo onde se encontra d, realizando a localização do intervalo de d, temos:

$$f_e = \sup_{d \leq f_i} \{f_1, \ f_2, \ \ldots, \ f_N\} \tag{25}$$

e

$$e = \arg \left\{ \sup_{d \leq f_i} \{f_1, \ f_2, \ \ldots, \ f_N\} \right\} \tag{26}$$

Sabendo que a FDA é aproximadamente o histograma acumulado dividido pela quantidade total de amostras L utilizadas da série e através de proporcionalidade no intervalo $[f_{e-1}, f_e]$, onde $d \in [f_{e-1}, f_e]$ temos:

$$\frac{(f_e/L - f_{e-1}/L)}{(d - f_{e-1}/L)} = \frac{(e - (e-1))}{(F_z^{-1}(d) - (e-1)} \tag{27}$$

Isolando $F_z^{-1}(d)$, obtemos a equação final para a inversa da FDA discreta,

$$F_z^{-1}(d) = (e-1) + \frac{(L \cdot d - f_{e-1})}{(f_e - f_{e-1})} \tag{28}$$

Estimação de Frequências em Sinais Eletroencefalográficos Utilizando Autovalores

Lucas F. Cruz[1], **Marcela G. Magalhães**[1], **Rodrigo P. Lemos**[1]

[1] Escola de Engenharia Elétrica, Mecânica e de Computação – Universidade Federal de Goiás (UFG) CEP 74.605-010 – Goiânia – GO – Brasil

fiorini.cruz@gmail.com, mar.cela.gm@hotmail.com, rodrigo.p.lemos@gmail.com

Abstract. *Studies to facilitate the mobility of people who, due to accidents or illnesses, are deprived of movement, has been one of the main focuses of the scientific community. A widely used instrument for peripheral commands is the electroencephalographic signal and the knowledge of their frequencies is essential. This article proposes a method based on eigenvalues for the knowledge of this signal characteristic and thus be compared with the traditional Fast Fourier Transform (FFT). The comparison was done as for the frequency spectrum and in terms of the Root Mean Square Error (RMSE) in function of Signal-to-Noise Ratio (SNR). The proposed method, even at low SNRs, has a statistical concentration of peaks in the frequencies of interest and the RMSE versus SNR curve remains below of FFT curve. The developed method, therefore, is applicable to frequency estimation.*

Resumo. *Estudos para facilitar a mobilidade de pessoas que, por motivo de acidentes ou enfermidades, são privadas de movimentos, têm sido um dos grandes focos da comunidade científica. Um meio muito utilizado para comandos de periféricos é o sinal eletroencefalográfico e o conhecimento de suas frequências é imprescindível. Este artigo propõe um método baseado em autovalores para o conhecimento de tal característica do sinal para, então, ser comparado com a tradicional Transformada Rápida de Fourier (FFT). A comparação é feita quantos aos espectros de frequências e em termos da Raiz do Erro Médio Quadrático (REMQ) em função da Relação Sinal Ruído (SNR). O método proposto, mesmo em baixas SNRs, tem uma concentração estatística de picos nas frequências de interesse e sua curva REMQ versus SNR mantém-se abaixo da curva referente à FFT. O método desenvolvido, portanto, mostra-se aplicável à estimação de frequência.*

1. Introdução

O corpo humano produz diversos tipos de biopotenciais e entre eles está os sinais eletroencefalográficos (EEG), provenientes da atividade elétrica do cérebro. Pelas mãos do alemão Hans Berger, em 1929, nasceu a eletroencefalografia dando início a diversos estudos e transformando o EEG em uma ferramenta para pesquisas e diagnósticos simples e ainda atual [Vaid et al. 2015]. A gama de frequências no EEG varia normalmente entre 1Hz-80Hz, sendo os sinais divididos em diferentes ritmos: delta, theta, alfa e beta, relacionados ao estado de atividade do indivíduo e coletados em diferentes áreas do escalpo.[Kaur and Kaur 2015]. Para essa coleta são distribuídos diversos eletrodos, com

a quantidade e posições variáveis conforme as características dos sinais que se busca. O conhecimento das frequências de um EEG traz várias vantagens pois, uma vez que um cérebro normal gera sinais com frequências de até 40Hz, qualquer alta amplitude em frequências acima deste valor pode ser considerada anômala.

Diferentes estudos têm utilizado o sinal EEG, pela gama de informações e usos que ele proporciona além de ser um método não invasivo para aquisição de dados cerebrais, entre eles, os envolvendo Interfaces Cérebro-Computador (ICC). A primeira ICC implantada em humanos e com boa qualidade de sinais data de 1998 [Brain-Chip Interfaces 2005], pelas mãos do pesquisador e neurologista Philip Kennedy e desde então as pesquisas na área vem crescendo a cada ano, muito devido à quantidade de problemas e doenças que tem tornado mais pessoas incapacitadas.

O funcionamento de interfaces é dependente das frequências captadas do EEG, já que diferentes regiões e atividades geram sinais com frequências e amplitudes diversificadas, assim o reconhecimento desses valores é tão importante. Uma precisa estimação de frequência, tradicionalmente realizada por meio da Transformada Rápida de Fourier (*Fast Fourier Transform* - FFT), principalmente nos casos de magnitudes próximas e com a presença maciça de ruído, possibilita o desenvolvimento de aplicações que necessitariam de rigorosa identificação de valores e alta confiabilidade, caso da maioria das ICCs.

Com esse intuito é proposto um método de estimação de frequências por meio de autovalores, sendo comparado a FFT, tanto em relação aos espectros de frequências quanto a Raiz do Erro Médio Quadrático (REMQ) em função da Relação Sinal Ruído (*Signal-to-Noise Ratio* - SNR). Para tal, são gerados sinais a partir de um modelo de sinal esférico com o espectro em diferentes SNRs assim como curvas de REMQ representando a degradação na estimação das duas técnicas. Os resultados obtidos são então comentados na conclusão do trabalho.

2. A Transformada Rápida de Fourier

A Transformada de Fourier de Discreta (*Discrete Fourier Transform*-DFT) é utilizada para processamento de sinais digitais, devido a sua dependência por variáveis em frequência discreta, necessária para a aplicação em métodos onde o trabalho no tempo é inviável, fazendo então a análise espectral de sinais e respostas em frequência. Diversos algoritmos fazem o cálculo da DFT, sendo um dos mais conhecidos a FFT, propostos por em 1965 por Cooley $\&$ Tukey [Diniz et al. 2014].

Partindo da Equação 1 que representa a DFT de um sinal $x(n)$ com duração N, o algoritmo FFT faz então a estimação e separação das senoides que constituem o sinal de diferentes frequências. A FFT é baseada na propriedade de convolução, sendo empregada por requerer um número menor de multiplicações complexas, possuindo cálculos de menor complexidade e consequentemente com menos esforço computacional.

$$X(k) = \sum_{n=0}^{N-1} x(n) W_N^{kn}, 0 \leq k \leq N-1 \tag{1}$$

Em que W_N é igual a $e^{-j(\frac{2\pi}{N})}$. Tem-se como exemplo rápido a visualização de um pico na frequência em um sinal senoidal quando é aplicado o algoritmo da FFT, apresen-

tado na Figura 1.

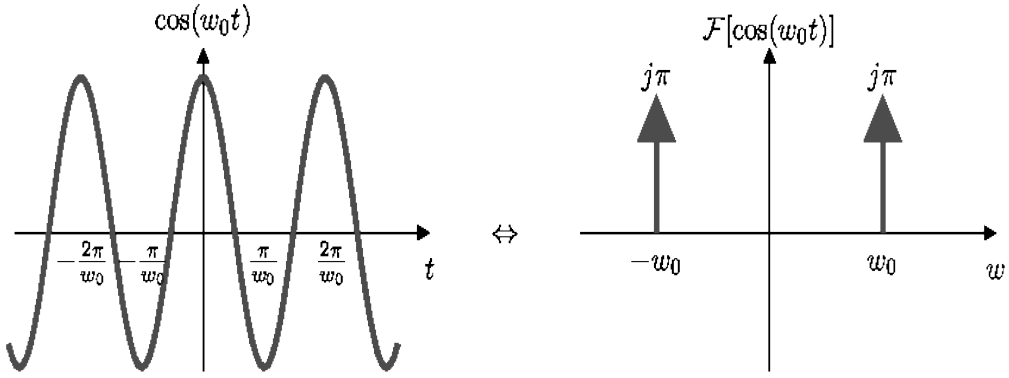


Figura 1. Pico de frequência estimado pela FFT

3. O Teorema dos Autovalores

A teoria para o método proposto nesse artigo baseia-se na interpretação física da Expansão de *Karhunen-Loéve*, em que são empregados vetores aleatórios num conjunto arbitrário de vetores ortonormais. Considerando o conjunto formado por M vetores ortonormais, autovetores $\vec{q_1}$, $\vec{q_2}$, ..., $\vec{q_M}$, regidos pela Equação 2, pode-se ter um $x(n)$ formado por $M \times 1$ vetores aleatórios como mostra a Equação 3, com projeções $x_1(n)$, $x_2(n)$, ..., $x_M(n)$ cada um representando uma variável aleatória.

$$\mathbf{q}_k^T \mathbf{q}_i = \begin{cases} 1, k = i \\ 0, k \neq i \end{cases} \tag{2}$$

$$x(n) = \sum_{i=1}^{M} x_i(n)\vec{q_i} \tag{3}$$

Com o produto de $x(n)$ por seu transposto é possível perceber que essa multiplicação corresponde a energia equivalente ao vetor $x(n)$ [Haykin 1989]. A equação encontrada para este cálculo é demonstrada por Equação 4.

$$x^T(n)x(n) = [x(n), x(n-1), ..., x(n-M+1)] \begin{bmatrix} x(n) \\ x(n-1) \\ ... \\ x(n-M+1) \end{bmatrix} = \sum_{k=0}^{M-1} x^2(n-k) \tag{4}$$

Com o resultado encontrado na Equação 4 pode-se aplicar a expansão dada em 3 provando que o somatório encontrado ao final as manipulações matemáticas em 4 iguala-se como mostrado na Equação 5.

$$\sum_{i=1}^{M} x_i^2(n) = \sum_{k=0}^{M-1} x^2(n-k) \tag{5}$$

É perceptível que o coeficiente $x_i^2(n)$ representa a energia no vetor aleatório $x(n)$ até a i-ésima coordenada o que consequentemente torna o autovalor correspondente ao exposto na Equação 6.

$$E[x_i^2(n)] = \lambda_i \tag{6}$$

Nota-se então que a energia esperada de $x(n)$, em caso prático o sinal avaliado, é equivalente a soma dos autovalores da matriz de correlação do próprio $x(n)$, como representado pela Equação 7.

$$E\left[\sum_{k=0}^{M-1} x^2(n-k)\right] = \sum_{i=1}^{M} E[x^2(n)] = \sum_{i=1}^{M} \lambda_i \tag{7}$$

4. Modelo de Sinal

No estudo das Atividades Neurais, vários modelos diferentes são adotados, por exemplo o modelo de Elementos Finitos. Para os propósitos deste artigo, o modelo adotado será o esférico, onde a massa cerebral, óssea e o escalpo são considerados como materiais condutores homogêneos e a fonte é modelada como um dipolo infinitesimal [Vergallo and Lay-Ekuakille 2013].

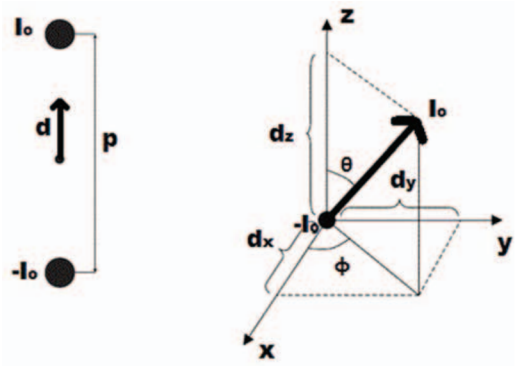


Figura 2. Modelo de Fonte Dipolo

A Figura 2 mostra um dipolo com momento $\vec{d}$, corrente I_0 e comprimento infinitesimal p com orientação $\vec{a_p}$ decomposto em três eixos ortogonais entre si, de acordo com as Equações 8 e 9.

$$\vec{d} = I_0 p \vec{a_p} \tag{8}$$

$$\vec{d} = d_x \vec{a_x} + d_y \vec{a_y} + d_z \vec{a_z} \tag{9}$$

Tal relação, portanto, torna possível calcular o Potencial V, apresentado pela Equação 10 advindo de uma fonte dipolo localizada em $\vec{r_f}$, em qualquer ponto $\vec{r_s}$ do espaço.

$$V(\vec{r_s}, \vec{r_f}, \vec{d}) = d_x V(\vec{r_s}, \vec{r_f}, \vec{a_x}) + d_y V(\vec{r_s}, \vec{r_f}, \vec{a_y}) + d_z V(\vec{r_s}, \vec{r_f}, \vec{a_z}) \tag{10}$$

O potencial gerado por um dipolo como o da Equação 8 em um condutor σ, em um modelo esférico é apresentado na Equação 11.

$$V(\vec{r_s}, \vec{r_f}, \vec{d}) = \frac{I_0 p}{4\pi\sigma|\vec{r_s} - \vec{r_f}|^2} \vec{a_p} \circ \left(\frac{\vec{r_s}}{|\vec{r_s}|} \right) \tag{11}$$

Como p é infinitesimal, temos que $p \to 0$ e $I_0 \to \propto$, mantendo o produto $I_0 p$ finito e igual a d. Conforme $p \to 0$, $\vec{a_p} \to \vec{a_z}$ [Malmivuo and Plonsey 1995] e a Equação 11 se torna a Equação 12.

$$V(\vec{r_s}, \vec{r_f}, d) = \frac{dcos(\theta)}{4\pi\sigma|\vec{r_s} - \vec{r_f}|^2} \tag{12}$$

Fazendo-se d uma função de tempo discreto $d_i(n)$ de uma i-ésima fonte localizada em $\vec{r_{f,i}}$ e $\vec{r_s}$ na posição de cada sensor k se tornar $\vec{r_{s,k}}$, V passa a ser uma matriz de ordem $(\mathbf{k}, \mathbf{N})$, sendo N o número de *snapshots* e com um número total de fontes m, de acordo com a Equação 13

$$V = \begin{bmatrix} \sum_{i=1}^{m} \frac{d_i(0)cos(\theta_1)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,1}}|^2} & \sum_{i=1}^{m} \frac{d_i(1)cos(\theta_1)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,1}}|^2} & \cdots & \sum_{i=1}^{m} \frac{d_i(N-1)cos(\theta_1)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,1}}|^2} \\ \sum_{i=1}^{m} \frac{d_i(0)cos(\theta_2)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,2}}|^2} & \sum_{i=1}^{m} \frac{d_i(1)cos(\theta_2)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,2}}|^2} & \cdots & \sum_{i=1}^{m} \frac{d_i(N-1)cos(\theta_2)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,2}}|^2} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} \frac{d_i(0)cos(\theta_k)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,k}}|^2} & \sum_{i=1}^{m} \frac{d_i(1)cos(\theta_k)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,k}}|^2} & \cdots & \sum_{i=1}^{m} \frac{d_i(N-1)cos(\theta_k)}{4\pi\sigma|\vec{r_{f,i}} - \vec{r_{s,k}}|^2} \end{bmatrix} \tag{13}$$

Na Figura 3 tem-se um exemplo da aplicação da Equação 13.

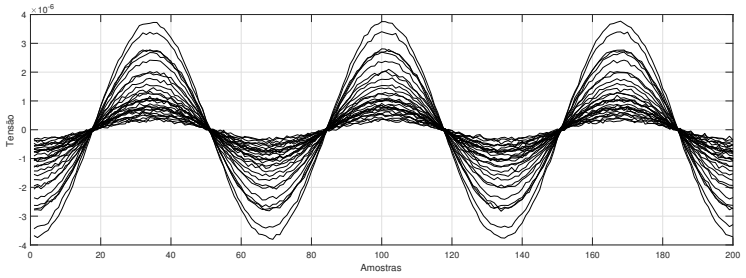


Figura 3. Tensões sobrepostas geradas por uma fonte com apenas uma frequência em $k = 30$ sensores, com $N = 200$ e $SNR = 30dB$.

Para efeitos de simulação e comparação de métodos, a localização dos sensores no escalpo e das fontes são ambas conhecidas.

5. Método de Autovalores

Os sinais usados para a determinação de frequências, normalmente ruidosos, serão chamado de sinais medidos ($\mathbf{V_{sm}}$). A autocorrelação temporal dos sinais medidos, $\mathbf{R_{sm(NxN)}}$, é dada pela Equação 14.

$$\mathbf{R_{sm}} = E\left\{\mathbf{V_{sm}^H V_{sm}}\right\} \tag{14}$$

Faz-se, então, uma matriz de autocorrelação com a mesma dimensão de $\mathbf{R_{sm}}$, contendo um sinal simulado e livre de ruídos com taxa de amostragem F_s, dependente de uma frequência de controle f, apresentada na Equação 15.

$$\mathbf{R_{ss}}(f) = E\left\{[cos(2\pi f\mathbf{t})]^H cos(2\pi f\mathbf{t})\right\}, \mathbf{t} = 0, \frac{1}{F_s}, \frac{2}{F_s}, ..., \frac{N-1}{F_s} \tag{15}$$

Somando as Equações 14 e 15, tem-se a matriz equacionada em 16 que, parte é referente às frequências das fontes do sinal medido e parte é referente à frequência do sinal simulado [Ferreira and Lemos 2005].

$$\mathbf{R_T}(f) = \mathbf{R_{sm}} + \mathbf{R_{ss}}(f) \tag{16}$$

Da matriz $\mathbf{R_T}(f)$ se extrai N autovalores, da forma mostrada pela Equação 17.

$$\lambda(f) = [\lambda_1(f), \lambda_2(f), \lambda_3(f), ..., \lambda_N(f)] \tag{17}$$

Cada autovalor é dependente da frequência do sinal utilizado em 15 e a soma de todos eles resulta na energia total dos sinais das matrizes de autocorrelação temporal medida e simulada [Haykin 1989].

Quando a frequência do sinal simulado diverge das frequências do sinal medido, a energia concentrada no primeiro autovalor é igual à energia total do sinal medido, enquanto o segundo autovalor concentra a energia do sinal simulado, pois os subespaços gerados por cada autovalor são ortogonais. No caso em que f do sinal simulado coincide com uma das frequências da fonte, o subespaço gerado pelo primeiro autovalor contém, além dos vetores geradores do sinal medido, os do sinal simulado, aumentando, portanto, o seu valor para a soma de suas energias.

Se fizermos $\mathbf{V} = 1,5cos(2\pi 20\mathbf{t})$, com frequência de amostragem $F_s = 160Hz$ e na ausência de ruído, teremos uma fonte com energia total $E_{fonteMedida} = 90$. Quando somamos sua matriz de correlação temporal com $\mathbf{R_{ss}}(60)$, que detém um sinal $cos(2\pi 60\mathbf{t})$ com energia $E_{fonteSimulada} = 40$, teremos os autovalores da Equação 18.

$$\lambda(60) = [90, 40, 0, ..., 0] \tag{18}$$

Ou seja, as energias das fontes medidas e simulada são alocadas em autovalores diferentes.

Para uma mesma fonte $\mathbf{V} = cos(2\pi 20\mathbf{t})$, porém com sua matriz de correlação somada a $\mathbf{R_{ss}}(20)$, teremos os da Equação 19.

$$\lambda(20) = [130, 0, 0, ..., 0] \tag{19}$$

Isto mostra que, quando as frequências coincidem, a amplitude do primeiro autovalor aumenta em $E_{fonteSimulada}$.

Fazendo-se uma varredura $(1 \leqslant f \leqslant F_s/2)$ teremos, então, um espectro de frequências.

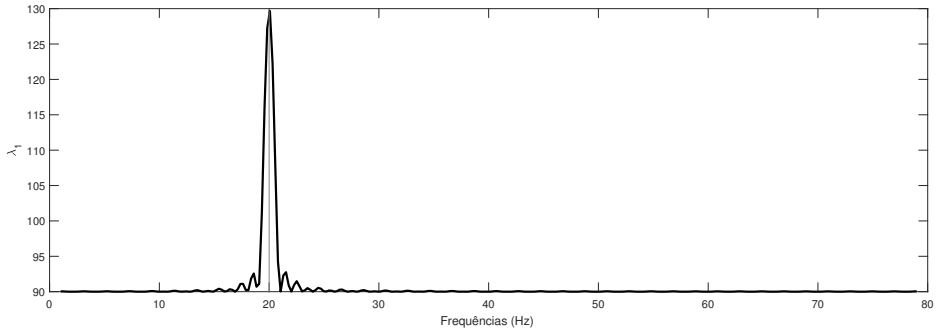


Figura 4. Espectro de λ_1 com uma linha vertical na posição da frequência conhecida do sinal medido

Na Figura 4, o mínimo valor que λ_1 atinge é o valor da energia da fonte medida e o máximo é a soma da energia da fonte medida e do sinal simulado.

Para uma fonte com frequências de valores 10, 20, 30, 40 e 50Hz:

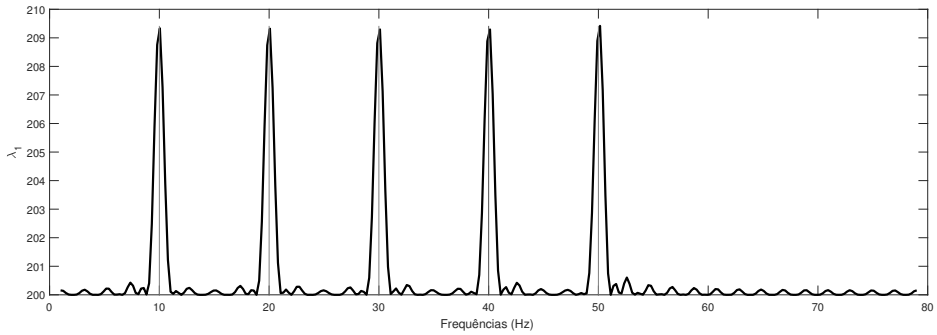


Figura 5. Espectro de λ_1

É importante ressaltar que o limite para a estimação das frequências se dá pelo critério de *Nyquist*, ou seja, só se é possível estimar frequências com valores menores que $F_s/2$.

6. Resultados

6.1. Parâmetros das Simulações

Em todo sistema que se envolve medição, o sinal medido possui uma parcela de ruído. Nesta análise é considerado o Ruído Branco aditivo, aleatório, Gaussiano e possuindo todas as frequências do espectro. Como tal perturbação possui natureza estocástica, todo experimento feito é repetido mil vezes para ter seu comportamento médio analisado.

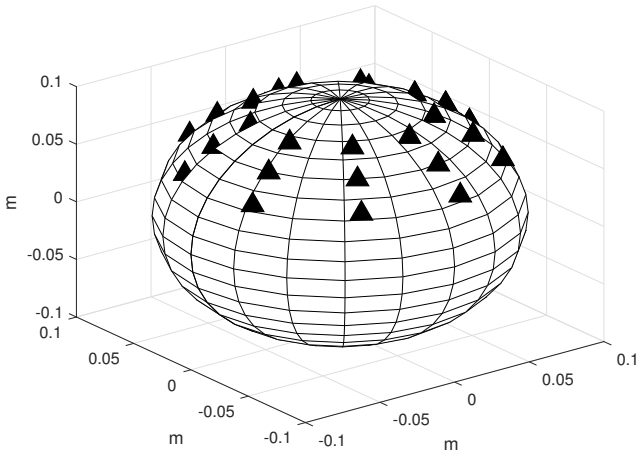
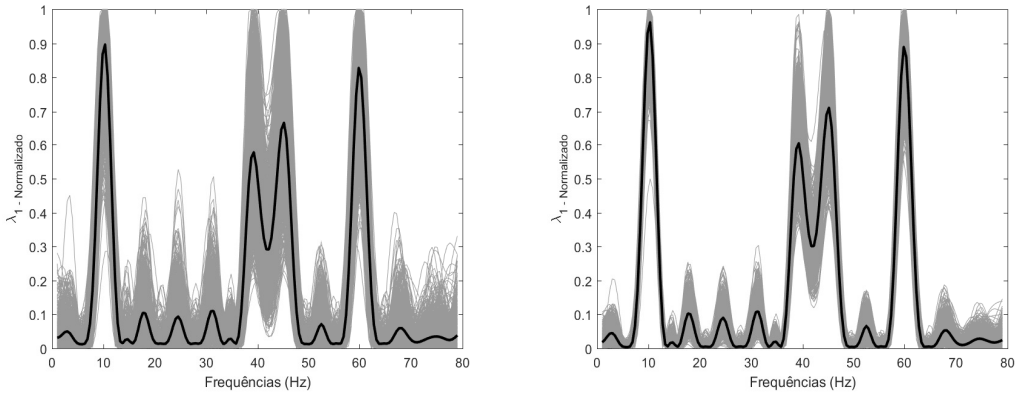


Figura 6. Posição dos sensores - Modelo esférico com raio de 0,1m

O número de sensores usados é $k = 30$ e nas posições da Figura 6. Todas as funções discretas possuem 24 amostras, ou seja $N = 24$.

6.2. Espectro

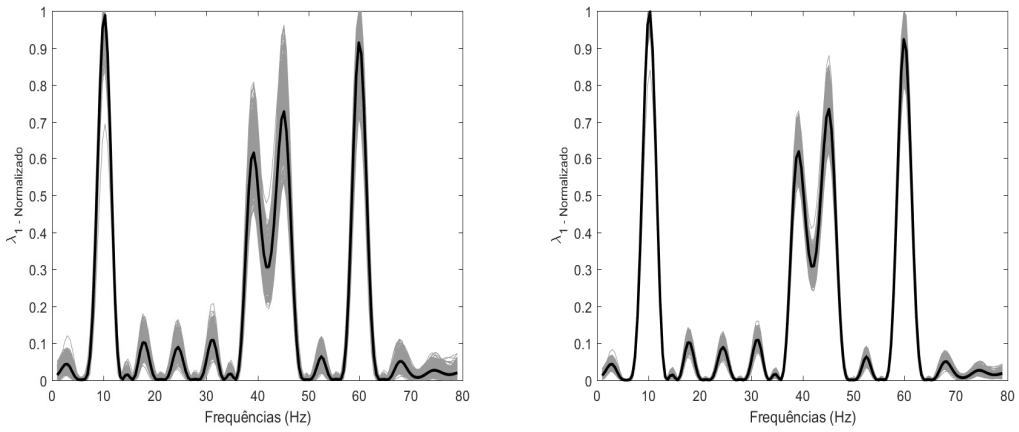
As Relações Sinal Ruído adotadas para a confecção dos espectros serão de $-5dB$, $0dB$, $5dB$ e $10dB$. A fonte possui quatro frequências: $10Hz$, $40Hz$, $45Hz$ e $60Hz$, sendo duas separadas apenas por $5Hz$, dificultando a distinção entre elas.



(a) Espectro a SNR de $-5dB$

(b) Espectro a SNR de $0dB$

Figura 7. Espectros - Curvas cinzas representando cada um dos 1000 experimentos sobrepostos e a curva preta sendo a média de todos os experimentos.



(a) Espectro a SNR de $5dB$ (b) Espectro a SNR de $10dB$

Figura 8. Espectros - Curvas cinzas representando cada um dos 1000 experimentos sobrepostos e a curva preta sendo a média de todos os experimentos.

Nas SNRs mais baixas (Figuras 7(a) e 7(b)) são observados picos espúrios no espectro devido à alta potência do ruído em relação ao sinal, o que pode perturbar a estimação da frequência, porém, em média, os picos se concentram nas frequências das fontes. Já nas SNRs mais altas (Figuras 8(b) e 8(a)) os picos nas frequências se destacam de maneira bastante perceptível, facilitando a estimação.

6.3. REMQ *versus* SNR

Como visto, SNRs baixas tendem a degradar a estimação e uma medida para tal degradação é a curva REMQ *versus* SNR [Boccato et al. 2012], dada pela Equação 20

$$REMQ(SNR) = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{Ne}\frac{(f_i - \hat{f}_{ij}(SNR))^2}{mNe}} \tag{20}$$

Em que Ne é o número de experimento, f_i as frequências a serem determinas e $\hat{f}_{ij}(SNR)$ as estimativas em cada SNR.

Nesse sentido, o método proposto é comparado com a FFT em dois casos: frequências em $10Hz$, $40Hz$, $45Hz$ e $60Hz$ (Figura 9); e frequências em $10Hz$, $20Hz$, $30Hz$, $40Hz$, $50Hz$ e $60Hz$ (Figura 10). Essa comparação tem como objetivo verificar a estabilidade dos métodos principalmente em taxas baixas de SNRs, quando os sinais estão com uma maior interferência pelos ruídos.

A avaliação se faz necessária devido, também, ao foco de aplicação em ICCs, haja visto que diferentes ruídos afetam o sinal, tanto originados do próprio sinal cerebral em frequências que não são necessárias para o controle do equipamento quanto pelos advindos da rede e de outros componentes eletrônicos dos circuitos implementados em ICCs.

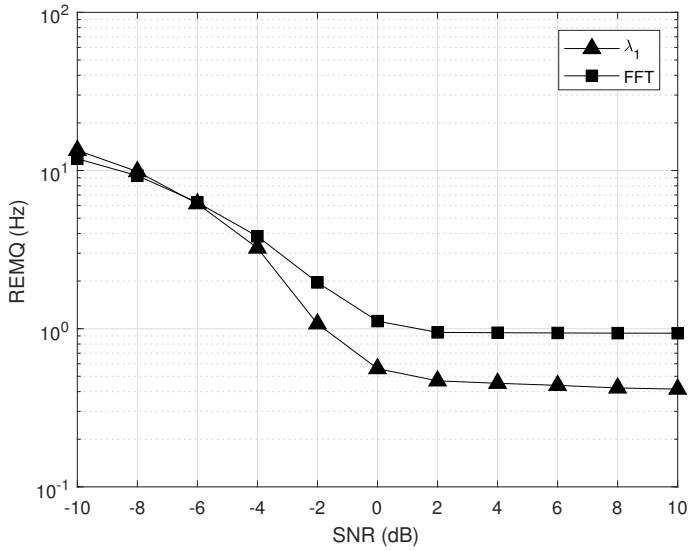


Figura 9. REMQ *verus* SNR - Frequências em $10Hz$, $40Hz$, $45Hz$ e $60Hz$.

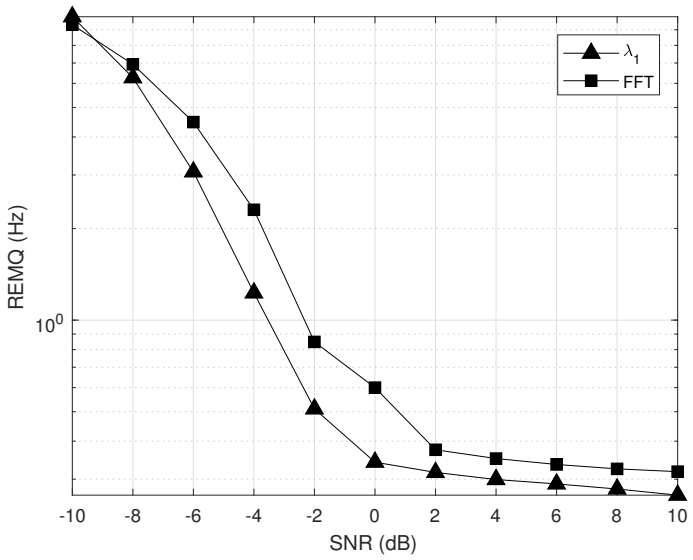


Figura 10. REMQ *verus* SNR - Frequências em $10Hz$, $20Hz$, $30Hz$, $40Hz$, $50Hz$ e $60Hz$.

7. Considerações Finais

O método de autovalores mostra-se viável na estimação de frequências, tendo a capacidade de distingui-las levando em conta a energia da fonte medida e do sinal simulado. Os espectros, mesmo em SNRs baixas, possuem concentrações de picos nas frequências da fonte e com um comportamento médio que favorece a estimação das frequências de interesse.

Quando comparado com a FFT quanto à curva REMQ *versus* SNR, o método proposto mostra-se mais robusto ao ruído e estabiliza em um erro menor, tornando-o mais viável em aplicações que requerem uma maior precisão.

Um método de busca local se faz necessário para manter o erro decrescente conforme a SNR aumenta, eliminando a dependência de uma varredura. Sendo esse, então, um problema de interesse em trabalhos futuros.

Referências

Boccato, L., Krummenauer, R., Attux, R., and Lopes, A. (2012). Application of natural computing algorithms to maximum likelihood estimation of direction of arrival. *Signal Processing*, 92(5):1338–1352.

Brain-Chip Interfaces (2005). History.

Diniz, P. S. R., da Silva, E. A. B., and Netto, S. L. (2014). *Processamento Digital de Sinais - Projeto e Análise de Sistemas*. Bookman Editora LTDA.

Ferreira, Y. R. and Lemos, R. P. (2005). A new doa estimation algorithm based on differential spectrum. 1:283–286.

Haykin, S. (1989). *Modern filters*. Macmillan Press.

Kaur, J. and Kaur, A. (2015). A review on analysis of eeg signals. pages 957–960.

Malmivuo, J. and Plonsey, R. (1995). *Bioelectromagnetism*. Oxford University Press.

Vaid, S., Singh, P., and Kaur, C. (2015). Eeg signal analysis for bci interface: A review. pages 143–147.

Vergallo, P. and Lay-Ekuakille, A. (2013). Brain source localization: A new method based on multiple signal classification algorithm and spatial sparsity of the field signal for electroencephalogram measurements. *AIP Publishing*.

Localização Usando Beacons em Ambientes Internos para Monitoramento Remoto de Pacientes

Ana Paula de Sousa Oliveira[1], Sérgio T. Carvalho[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74001-970 – Goiânia – GO – Brasil

{anaoliveira,sergio}@inf.ufg.br

Abstract. *Recently, indoor location has become an important research topic since it is fundamental to numerous Internet of Things (IoT) applications, such as human tracking. This paper presents an implementation of a prototype with 2 applications using beacons: (i) setup application, responsible for setup the environments that will be monitored, data collection and calculations to determine the patient's position, and (ii) monitoring application, responsible for displaying the position of the patient to an interested user. The experiments conducted using the implemented prototype with 3 beacons resulted in an error rate of 37%, with a localization error of approximately 1m.*

Resumo. *Recentemente, localização em ambientes internos se tornou um tópico importante de pesquisa já que é fundamental para inúmeras aplicações na área da Internet das Coisas (Internet of Things – IoT), como, por exemplo, rastreamento humano. Este artigo apresenta uma implementação de um protótipo com 2 aplicativos usando beacons: (i) aplicativo para configuração dos ambientes que serão monitorados e cálculos para determinar a posição do paciente, e (ii) aplicativo para exibição do posicionamento do paciente a um usuário interessado. Os experimentos conduzidos usando o protótipo implementado com 3 (três) beacons resultaram numa taxa de erro relativamente baixa, com erro de localização de aproximadamente 1m.*

1. Introdução

Nas últimas décadas o avanço nas tecnologias de localização foi grande e atualmente elas desempenham um papel crucial em diversos serviços de reconhecimento do contexto como navegação, robótica, sistemas de resposta emergencial e monitoramento de paciente [Golestiania et al. 2001]. Tradicionalmente, aplicações que fazem uso de reconhecimento do contexto estavam confinadas ao ambiente externo, e apesar de ser a mais conhecida, a tecnologia de localização por satélite - como o GPS (Sistema de Posicionamento Global), que é um sistema de navegação por satélite formado por 24 satélites que circundam a Terra [El-Rabbany, A 2002] - é eficiente em ambientes externos, porém não são bem sucedidas em ambientes internos, por conta da atenuação e reflexão dos sinais pelas paredes e outras construções. Com os novos avanços da tecnologia, pesquisas têm explorado cada vez mais esse reconhecimento em ambientes internos, especialmente no que se refere à localização de um dispositivo ou de uma pessoa em um ambiente interno complexo, como, por exemplo, prédios, aeroportos e residências, usando meios automatizados [Chawathe et al. 2009].

Este artigo lida com técnicas de localização em ambientes internos no contexto

de pacientes monitorados em suas residências. Percebe-se que os tradicionais sistemas de saúde estão sobrecarregados por terem altos custos devido ao envelhecimento das populações. No sentido de torná-los mais eficientes, cenários de cuidados domiciliares onde o empoderamento do paciente é o foco estão começando a ser propostos. Com essa mudança nos paradigmas de assistência médica, novas tecnologias emergem para viabilizar suas implementações, como, por exemplo, o UbiCare [Germano 2015][Germano et al. 2016][Ribeiro et al. 2016] – que é um sistema de monitoramento remoto baseado num plano de cuidados domiciliar ubíquo com o objetivo de estreitar a relação entre o paciente, os tratamentos e os profissionais de saúde.

Um ambiente de monitoramento remoto de pacientes - como o UbiCare [Germano 2015][Germano et al. 2016][Ribeiro et al. 2016] - possui funcionalidades para tornar o tratamento do paciente mais rápido, fácil e informativo para o médico, paciente e interessados no tratamento.

Nesse contexto o incremento de funções que permitem a localização de um dispositivo em ambientes interno e externo pode trazer diversas vantagens como: tomar conhecimento se o paciente - que está portando um smartphone - passou muito tempo em determinado cômodo da casa, e caso essa posição seja suspeita (por exemplo, mais de meia hora no banheiro), sejam disparadas notificações para um outro usuário que previamente se inscreveu para receber informações desse paciente e tomar quaisquer medidas cabíveis.

Neste artigo é proposta uma abordagem de localização em ambientes internos utilizando sinais Bluetooth por meio de *beacons* com a tecnologia Bluetooth de Baixa Energia, ou BLE (Bluetooth Low Energy) para a criação de um aplicativo que será integrado ao sistema de monitoramento remoto UbiCare [Germano 2015][Germano et al. 2016][Ribeiro et al. 2016]. A tecnologia BLE possui baixo consumo de energia, permitindo que os transmissores, no caso, *beacons,* se mantenham ligados continuamente por meses, e até anos.

Este artigo está organizado em mais 5 seções: a Seção 2 apresenta os conceitos fundamentais que envolvem as técnicas de localização em ambientes internos; na Seção 3 são apresentadas as equações matemáticas empregadas na abordagem proposta; a Seção 4 fornece uma visão completa do protótipo desenvolvido; a Seção 5 descreve os experimentos e resultados alcançados, e, por fim a Seção 6 traz as conclusões e possibilidades de trabalhos futuros.

2. Fundamentação Teórica

Essa seção apresenta os fundamentos teóricos utilizados na implementação do aplicativo para localização em ambientes internos.

2.1. Técnicas Conhecidas Utilizadas para Localização em Ambientes Internos

O GPS se tornou a tecnologia ubíqua mais utilizada para a identificação de dispositivos móveis em ambientes externos, porém ele não provê posicionamento preciso em ambientes internos.

Atualmente são conhecidas três técnicas principais que sistemas de rastreio de

localização para posicionamento em ambientes internos:

- Trilateração de Radiofrequência (RF): usa as propriedades geométricas dos triângulos através dos sinais RF que chegam a múltiplos receptores. A distância entre a localização do objeto de interesse e outras localizações são detectadas com base em análise de sinal. Vários objetos - como paredes, mobílias e pessoas - podem distorcer o intervalo e ângulo estimado levando a uma localização imprecisa [Hightower et al. 2001]. Um exemplo desse tipo de triangulação é aquela realizada com os sinais WiFi, que se aproveita da infraestrutura sem fio de um edifício para localizar dispositivos. O principal problema, porém, está em determinar a precisão da localização, a qual depende da quantidade de pontos de acesso WiFi visíveis para o dispositivo, e, devido a esse custo, não chega a ser uma solução tão satisfatória. Outro exemplo que está sendo investigado se refere às soluções usando Bluetooth para estimar as distâncias. Em relação à WiFi, havia a desvantagem de exigir muita energia e ter pouco tempo de vida. No entanto, com o surgimento do Bluetooth de Baixa Energia (BLE) essas desvantagens foram solucionadas, mostrando essa tecnologia como promissora [Wang 2015].
- Proximidade acústica de Radiofrequência: usa frequência gerada por um oscilador. Este sinal é alimentado por um transdutor que emite pulsos de ultra-som à várias frequências numa sequência codificada. Estes pulsos refletem de objetos próximos e são retornados a um outro transdutor, que converte o ultra-som de volta em pulsos de alta frequência e essas reflexões são usadas para estimar a distância. Como essa técnica também faz uso de radiofrequência, vários objetos também podem interferir nos cálculos e gerar uma localização imprecisa.
- Registro de impressões digitais (*fingerprinting*): usa aprendizado de máquina sendo composto de duas fases: na primeira fase - aprendizado - vetores compostos dos valores dos Indicadores de Força de Sinais Recebidos (RSSI - *Received Signal Strengh Indicator*) e componentes opcionais extras medidos por um dispositivo de medição numa localização conhecida, são coletados. Esses dados de referência - conjuntos de dados calibrados - são salvos juntos com as coordenadas de localização no banco de dados *fingerprint* usando um método cabível, e são treinados num modelo de algoritmo de aprendizado de máquina; na segunda fase o modelo criado anteriormente é usado para localização do dispositivo. O dispositivo coleta dados, envia os dados para o servidor, o qual faz a previsão da localização usando o modelo, e então recebe a localização prevista [Kriz et al. 2016].

Com isso, pode-se notar que o principal desafio na localização em ambientes internos está nas técnicas que utilizam sinais de radiofrequência, onde o sinal pode sofrer atenuação, refração e distorção, fazendo com que a distância calculada entre o emissor e receptor sofra alterações bruscas por conta do ambiente, dificultando a precisão da localização do receptor no espaço monitorado.

2.2. Beacons no Contexto de Localização sem Fio

As tecnologias sem fio vêm sendo amplamente utilizadas dentre as tecnologias de localização em ambientes internos, em específico a localização WiFi, tecnologia de localização sem fio mais comum [Zhuang et al. 2015]. Outro candidato importante para localização sem fio em dispositivos inteligentes para consumidores é o Bluetooth. O Bluetooth tradicional tem um tempo de varredura relativamente longo (aproximadamente 10 segundos), o que limita seu valor para localização. Diferentemente, o novo protocolo Bluetooth (BLE), suportado pela maioria dos dispositivos inteligentes atuais superou as limitações desse longo tempo de varredura. Além disso, os *beacons* BLE têm as vantagens de terem tamanho pequeno, serem leves, de baixo custo, maior economia de energia e ainda amplamente suportados por dispositivos inteligentes. Logo, os *beacons* BLE tem grande potencial de se tornarem uma tecnologia de localização sem fio dominante.

Um *beacon* consiste em um *chipset* Bluetooth (BLE), uma bateria e uma antena. *Beacons* BLE não precisam de conexão à Internet para funcionar, sendo dispositivos apenas de transmissão. Um *beacon* transmite informações de identificação que aplicações podem usar para identificar o tipo de espaço em que o *beacon* se encontra [Gast 2014].

O *beacon* é um "disseminador", um tipo de dispositivo que transmite periodicamente pacotes Bluetooth com identificação de dados que contêm informações usadas pelos receptores - chamados de "anúncios" - em intervalos regulares. Esses pacotes são projetados para serem transmitidos, mas os receptores não precisam respondê-los. Esse dispositivo não se comunica com os dispositivos em sua volta por nenhum outro meio sofisticado. Cada pacote contém as seguintes informações:

- Endereço MAC;
- Identificador Universal Único (UUID) - comum para ser único em um local;
- Número principal - designado para dividir os conjuntos de *beacons* em segmentos menores;
- Número secundário - designado para dividir os segmentos de *beacons* em subsegmentos menores;

A maioria dos casos de uso envolvendo *beacons* foca em detecção de proximidade e localização aproximada do usuário, porém já existem alguns casos envolvendo localização em ambientes internos, como estudos relativos à performance do sinal de rádio frequência dos *beacons* [Neburka et al. 2016], baseado em smartphone [Zhuang et al. 2016] e combinando triangulação e impressões digitais [Faragher et al. 2015]. Todas essas abordagens fazem uso do RSSI do *beacon*, onde o sinal é usado para calcular a distância entre o emissor e receptor e posteriormente são utilizadas abordagens de trilateração e/ou de impressões digitais [Jae et al. 2016].

3. Modelo Matemático

Essa seção apresenta as equações matemáticas usadas para determinar as coordenadas referentes à localização do dispositivo móvel no ambiente interno.

3.1. Distância entre *Beacon* e *Smartphone*

Para determinar as coordenadas referentes à localização do dispositivo móvel no ambiente interno é necessário primeiro que o dispositivo colete as informações (anúncios) transmitidas pelo *beacon*. Assim, o valor de RSSI é utilizado no algoritmo proposto para obter as informações de localização. O modelo de perda de trajeto que entrega a relação entre o RSSI e a distância é dada pela seguinte equação:

(1) $RSSI = -10n\log_{10}d + A + B$

Nesta equação, "n" representa o expoente de perda de trajeto que depende do ambiente, "d" é a distância entre o transmissor e receptor, "A" é o RSSI a 1m proporcionado pelo próprio anúncio recebido pelo *beacon* e "B" é uma variável aleatória normal com zero que caracteriza a atenuação causada por desvanecimento plano [Cho et al. 2015]. A equação pode ser reorganizada para obter a distância:

(2) $d = 10^{(A-RSSI)/10n}$

O valor de "n" pode ser encontrado por experimentos empíricos. A equação a seguir e seus coeficientes foram adotados na *Android Beacon Library*[1] e os resultados desse artigo também se basearam nela:

(3) $d = 0.89976*(RSSI \div A)^{7.7095} + 0.111$

3.2. Trilateração

Três *beacons* foram utilizados e dispostos em coordenadas X, Y conhecidas. Após ter as distâncias entre os *beacons* e o dispositivo móvel receptor, as coordenadas de cada *beacon* e as distâncias são utilizadas no método de trilateração (Figura 1).

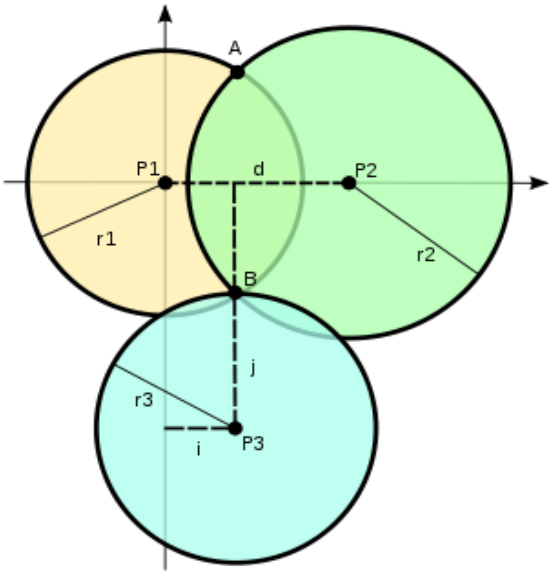


Figura 1. Ponto de intersecção entre 3 circunferências

Ao tratar as coordenadas do *smartphone* do paciente P(x,y) como ponto de

[1] https://altbeacon.github.io/android-beacon-library/.

intersecção das 3 circunferências cujos centros são as localizações dos 3 *beacons* $B_i(x_i, y_i, z_i)$ para i = 1, 2 e 3 e a distância entre o *smartphone* e cada *beacon* sendo r_i o raio individual das circunferências. As equações das circunferências são:

(4) $r_{P1}^2 = (x - x_{P1})^2 + (y - y_{P1})^2$

(5) $r_{P2}^2 = (x - x_{P2})^2 + (y - y_{P2})^2$

(6) $r_{P3}^2 = (x - x_{P3})^2 + (y - y_{P3})^2$

Uma vez que se supõe que todos os pontos se estendem no mesmo plano, considere os três pontos (P1, P2 e P3) que têm distância (r1, r2, r3) para o nó de interseção, como ilustrado na Figura 1. A fórmula para todas as circunferências em um plano é [Neburka et al. 2016]:

(7) $r_{P1}^2 = x^2 - 2x*x_{P1} + x_{P1}^2 + y^2 - 2y*y_{P1} + y_{P1}^2$

(8) $r_{P2}^2 = x^2 - 2x*x_{P2} + x_{P2}^2 + y^2 - 2y*y_{P2} + y_{P2}^2$

(9) $r_{P3}^2 = x^2 - 2x*x_{P3} + x_{P3}^2 + y^2 - 2y*y_{P3} + y_{P3}^2$

As três equações formadas são equações simultâneas não-lineares independentes que não podem ser resolvidas matematicamente. No entanto, utilizando o método proposto por Dixon[Dixon 2009] para obter o radical plano para intersecção, a equação (9) foi subtraída da equação (8) resultando na seguinte equação:

(10) $r_{P2}^2 - d_{P3}^2 = 2x(x_{P3} - x_{P2}) + x_{P2}^2 - x_{P3}^2 + 2y(y_{P3} - y_{P2}) + y_{P2}^2 - y_{P3}^3$

E subtraindo a equação (7) da (8) obtém-se:

(11) $r_{P2}^2 - d_{P1}^2 = 2x(x_{P1} - x_{P2}) + x_{P2}^2 - x_{P1}^2 + 2y(y_{P1} - y_{P2}) + y_{P2}^2 - y_{P1}^3$

Rearranjando a equação (10) para produzir a nova equação e nova variável:

(12) $x(x_{P2} - x_{P3}) + y(y_{P2} - y_{P3}) = ((d_{P2}^2 - d_{P3}^2) - (x_{P2}^2 - x_{P3}^2) - (y_{P2}^2 - y_{P3}^2)) \div 2 = v_{P1}$

Rearranjando a equação (11) para produzir a nova equação e nova variável:

(13) $x(x_{P1} - x_{P2}) + y(y_{P1} - y_2) = ((d_{P2}^2 - d_{P1}^2) - (x_{P2}^2 - x_{P1}^2) - (y_{P2}^2 - y_{P1}^2)) \div 2 = v_{P2}$

Resolvendo a equação (12) e a equação (13) para descobrir o ponto x e y de interseção:

(14) $y = (v_{P2}(x_{P3} - x_{P2}) - v_{P1}(x_{P1} - x_{P2})) \div (y_{P1} - y_{P2})(x_{P3} - x_{P2}) - (y_{P3} - y_{P2})(x_{P1} - x_{P2})$

(15) $x = (v_{P1} - y(y_{P3} - y_{P2})) \div (x_{P3} - x_{P2})$

Os valores para x e y fornecem a posição do ponto de intersecção das 3 circunferências.

4. Protótipo

Essa seção fornece uma visão completa do protótipo desenvolvido neste artigo.

4.1. Cenário de Distribuição dos *Beacons* numa Residência

A Figura 2 mostra um exemplo de cenário de distribuição dos *beacons* em uma residência, onde cada cômodo deve ter 3 *beacons*, sendo 2 centralizados nas

paredes de maior tamanho do cômodo e um centralizado em uma das paredes de menor tamanho. Cada *beacon* será um ponto de coordenadas X e Y que será usado no processo de trilateração citado na Seção 3.2 e transmitirá anúncios continuamente no ambiente.

Figura 2. Cenário de distribuição dos *beacons* em uma residência

4.2. Arquitetura

A localização do *smartphone* do paciente dentro de cada cômodo fazendo uso dos *beacons* com suas posições estáticas e suas distâncias até o *smartphone* (que é também a posição do paciente), será dada após cálculos usando equações apresentadas na Seção 3.

O Servidor (Firebase[2] com o banco de dados em tempo real) mantém as informações sobre cada cômodo enviadas a partir do aplicativo de configuração (coordenadas e dados de cada *beacon* num cômodo e coordenadas atualizadas do *smartphone* do paciente onde foi feita a configuração inicial). As informações são recuperadas pelo aplicativo de monitoramento junto ao mesmo Servidor para que o aplicativo de monitoramento possa então exibir a localização do *smartphone* do paciente que está sendo monitorado.

A Figura 3 mostra a Arquitetura do protótipo. Na arquitetura, como está sendo utilizado o método de trilateração, são necessários no mínimo 3 *beacons*, sendo que esses *beacons* enviam constantemente "anúncios" contendo as informações descritas

[2] https://firebase.google.com/

anteriormente (Seção 2) e o valor de intensidade do sinal de radiofrequência (RSSI) para o receptor (neste caso o *smartphone* do paciente executando o aplicativo onde é feita a configuração para localização em ambientes internos).

Dois aplicativos fazem parte do protótipo:

- O aplicativo onde é feita a configuração dos ambientes que serão monitorados e os cálculos para determinar a localização do *smartphone* do paciente onde esse aplicativo está instalado. Nele é mostrado para o paciente como dispor os *beacons* em um cômodo, e, posteriormente solicitado ao mesmo paciente que configure esse cômodo com nome e suas dimensões. Cálculos são realizados em segundo plano (usando Service[3]) para determinar a posição do paciente fazendo uso do modelo matemático mostrado na Seção 3. O aplicativo de configuração também faz o envio para o banco de dados das informações relativas aos cômodos e às coordenadas do paciente atualizadas, além do cômodo atual onde o *smartphone* do paciente está localizado.
- O aplicativo de monitoramento do paciente que é utilizado pelo interessado. Ele se comunica com o servidor para coletar e exibir esses dados para posteriormente exibir na tela o posicionamento do *smartphone* do paciente no cômodo em que ele se encontra.

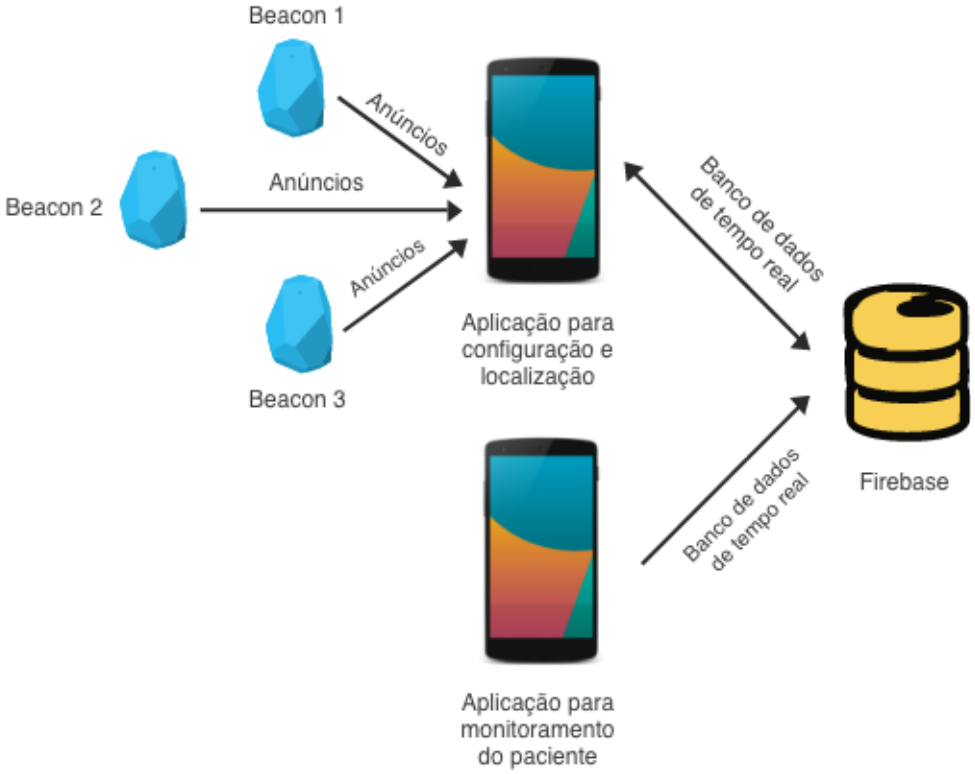


Figura 3. Arquitetura do ambiente para localização em ambientes internos

[3] https://developer.android.com/guide/components/services.html

4.3. Uso do Protótipo - Aplicativo de Configuração

A Figura 4 mostra a 1ª e 2ª tela do modo de configuração. Na 1ª tela o paciente deve adicionar todos os cômodos (com nome e dimensões) os quais deseja que o interessado monitore e cada cômodo deve possuir no mínimo 3 *beacons*. Na 2ª tela começa a configuração do primeiro cômodo adicionado na 1ª tela, é então solicitado ao usuário que ele disponha os 3 *beacons* necessários no cômodo para cálculo da trilateração de acordo com o desenho informado.

Figura 4. 1ª e 2ª telas do aplicativo de configuração

A Figura 5 mostra as 3ª e 4ª telas do modo de configuração. Na 3ª tela é dado o início da calibração do *smartphone* do paciente com os *beacons* através do reconhecimento de cada *beacon* correspondente ao cômodo correspondente. Nesse passo o aplicativo irá inserir no servidor os dados referente ao cômodo e a cada *beacon* (id, endereço MAC, posição X e Y) detectado daquele cômodo. Para realizar a detecção de cada *beacon*, o usuário deve colocar o dispositivo ao lado do *beacon*, pressionar o botão "detectar *beacon*", esperar que seja feita a detecção daquele *beacon* e repetir esse processo até que todos os *beacons* daquele cômodo sejam detectados. Uma vez que a parte de calibração tenha sido completada, caso não tenha adicionado nenhum outro cômodo a ser calibrado (do contrário a calibração para os próximos n - 1 cômodos é repetida n - 1 vezes), é exibida a 4ª tela - onde é fornecida para o usuário uma chave para que ele possa entrar no modo de monitoramento.

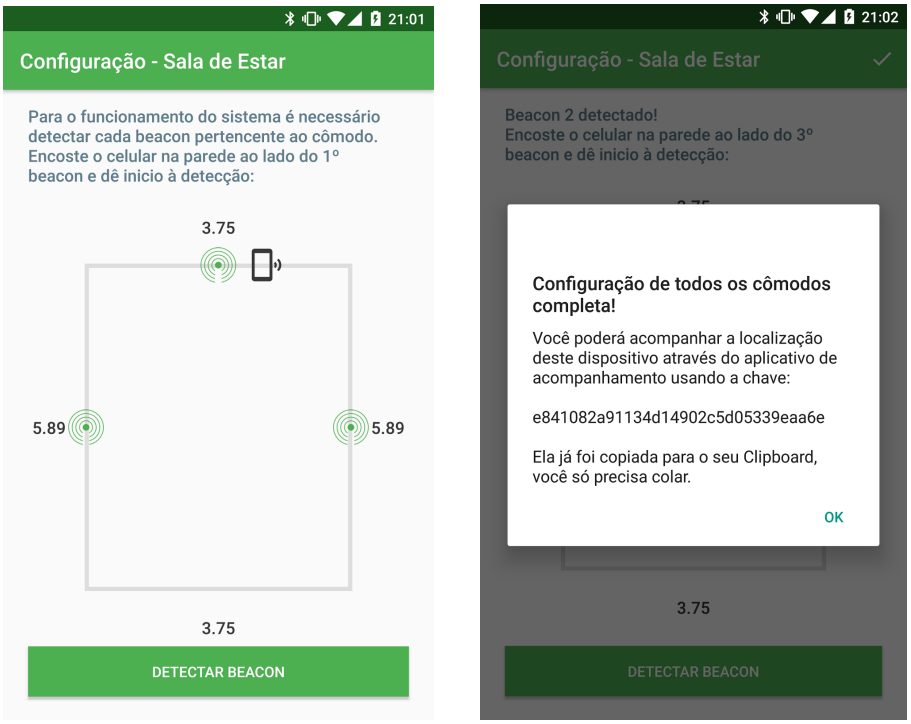


Figura 5. 3ª e 4ª telas do aplicativo de configuração

4.3.1. Uso do Protótipo - Aplicativo de Monitoramento

A Figura 6 mostra a 1ª tela do aplicativo de monitoramento, onde o interessado deve inserir a chave (apresentada na 4ª tela da Figura 5) fornecida ao final da configuração do ambiente usando o aplicativo de configuração.

Figura 6. 1ª tela do aplicativo de monitoramento

A Figura 7 mostra a tela do aplicativo de monitoramento que é exibida logo após o usuário escolher a opção de monitoramento. Nela é exibida a posição do paciente e em que cômodo (dentre os configurados inicialmente) ele se encontra.

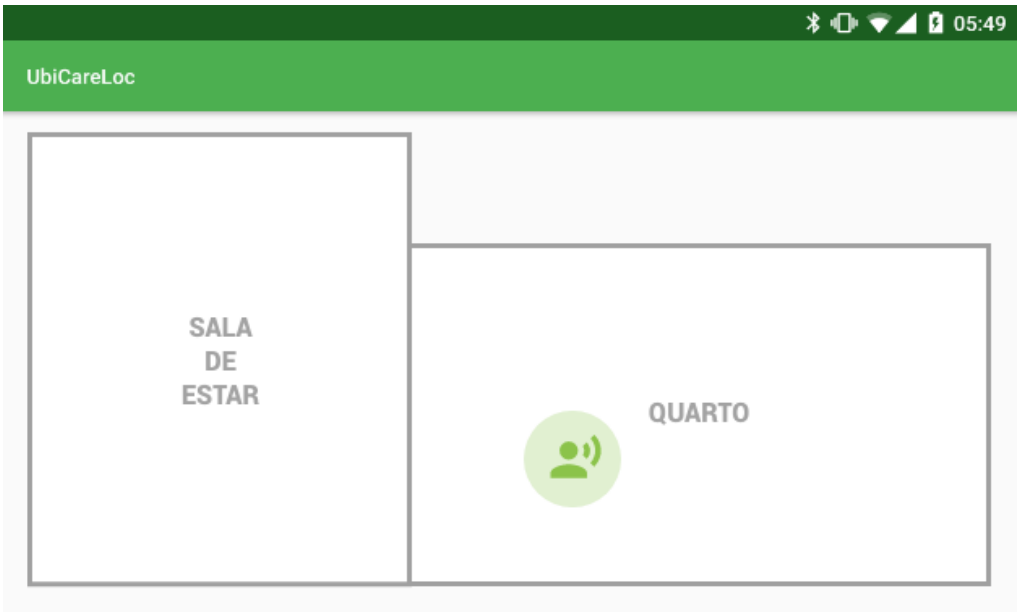


Figura 7. Tela do aplicativo de monitoramento mostrando o posicionamento do *smartphone*

5. Experimentos e Resultados

No experimento foi usado como *smartphone* do paciente um Nexus 6P com processador 2.0 GHz octa-core 64-bit, memória de 128GB e sistema operacional Android 8.0 (Oreo) e como *smartphone* do interessado foi usado um Galaxy S3 com processador quad-core 1.4 GHz, memória de 16GB e sistema operacional Android 6.0 (Marshmallow), e 3 *beacons* Estimote Localização[4] com processador 2.3 MHz 32-bit, memória flash de 256kB, acelerômetro, sensores de temperatura, luminosidade, pressão e magnetômetro e *Bluetooth* 4.0 2.4GHz rádio bidirecional.

A Tabela 1 mostra as coordenadas referentes à localização real e estimada presentes num experimento com 3 *beacons* seguindo o modelo de configuração e monitoramento mostrado na Seção 4. Conclui-se que a taxa de erro aumenta à medida que o *smartphone* se distancia dos *beacons* e que a média de erro máximo foi de 1m. No caso de saber em qual cômodo o *smartphone* está enquanto se move, a implementação foi um sucesso, pois sempre foi detectada a mudança de cômodo e na maioria das vezes chegava a mostrar em tempo real no aplicativo de monitoramento instalado no *smartphone* do interessado que está monitorando a localização do *smartphone* do paciente.

[4] https://estimote.com/

Localização real (x, y)	Localização estimada (x, y)	Taxa de erro
(0.79, 0.95)	(1.18, 1.11)	~32.5%
(1.35, 1.67)	(0.97, 0.92)	~37%
(1.64, 2.62)	(1.36, 1.57)	~44.5%
(0.85, 3.24)	(1.81, 4.33)	~22.5%
(1.64, 3.90)	(1.72, 2.32)	~54%

Tabela 1. Experimento gerado num ambiente interno

A Figura 8 mostra visualmente as localizações reais e estimadas encontradas no experimento, assim como o posicionamento dos *beacons* num plano cartesiano.

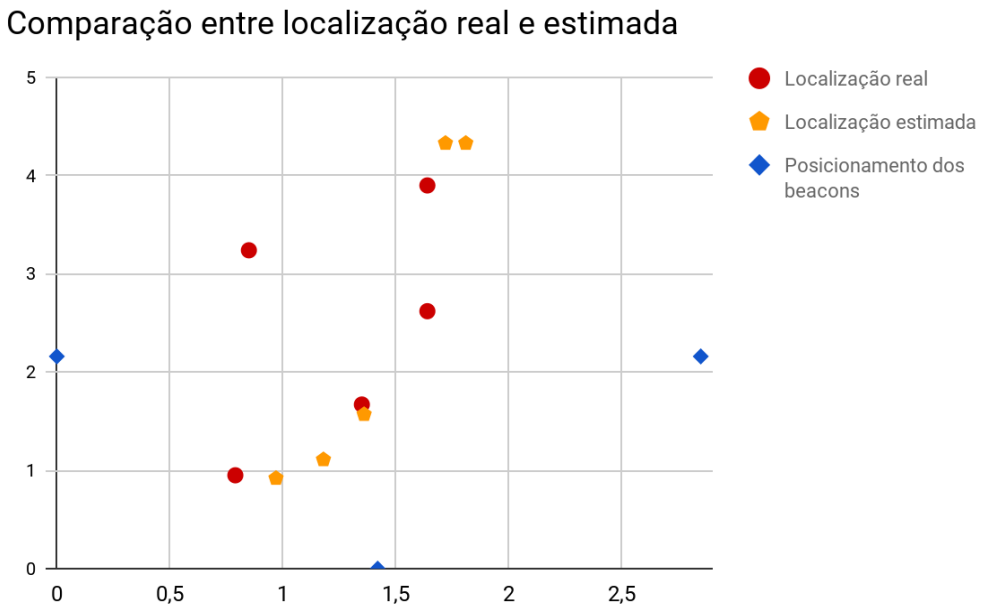


Figura 8. Disposição dos dados do experimento citado na Tabela 1

6. Conclusão

O protótipo apresentado neste artigo tem como objetivo auxiliar pacientes em ambientes domiciliares que possuem necessidade de serem monitorados constantemente. Alguns cenários foram propostos onde essa solução de monitoramento remoto se encaixaria assim como na integração do protótipo ao UbiCare.

Foi discutido anteriormente que vários fatores podem interferir no sinal de radiofrequência emitido pelos *beacons* (incluindo o design da antena, do hardware, os drivers e o ambiente onde o experimento foi realizado), o que faz com que a acurácia da

localização estimada não seja ótima.

Ao comparar os resultados obtidos neste artigo com outros trabalhos relacionados que também utilizaram métodos de trilateração e distância baseada no RSSI, temos que: em [Golestiania et al. 2001] os resultados foram similares pois a média de erro foi de 1 metro, e à medida que o receptor se distanciou dos *beacons* essa média aumentou, já em [Aman et al. 2016] foram gerados erros de distância menores (média de 0.89 m) do que os apresentados nesse artigo.

Para trabalhos futuros, alguns pontos poderiam ser explorados: testes de outros métodos matemáticos de trilateração para calcular o posicionamento do usuário; o alto consumo de bateria, dado pelos constantes cálculos realizados pelo próprio *smartphone* em segundo plano, que poderia ser resolvido pela transferência desses cálculos para um servidor; a inviabilidade do paciente estar sempre com seu *smartphone* em mãos, e nesse caso poderia ser construído um dispositivo vestível que faria o papel do *smartphone* e utilização dos dados gerados pelo paciente para aprender sua rotina e sugerir um plano de cuidados personalizado.

Referências

Germano, E. S. (2015) Plano de Cuidados Ubíquo com Sistema de Notificações voltado a Pacientes Domiciliares. III Escola Regional de Informática de Goiás.

Germano, E., Battisti, D., Ribeiro, H. A., Carvalho, S. T. (2016) Plano de Cuidados Ubíquo para Acompanhamento Domiciliar de Pacientes. IV Escola Regional de Informática de Goiás.

Ribeiro, H. A., E., Battisti, D., Germano, Carvalho, S. T. (2016) Notificações de Monitoramento Remoto de Pacientes Usando Redes Sociais. XV Congresso Brasileiro em Informática e Saúd*e.*

Golestiania, M. e Poellabauer, C. (2001) Indoor Localization using Multi-Range Beaconing. 17th ACM International Symposium on Mobile Ad Hoc Networking and Computing.

El-Rabbany, A. (2002) Introduction to GPS: The Global Positioning System, Artech House, 1ª Edição.

Chawathe, S. S. (2009) Low-Latency Indoor Localization Using Bluetooth Beacons. 12th International IEEE Conference on Intelligent Transportation Systems.

Hightower, J., Borriello G. (2001) Location systems for ubiquitous computing. On Journal Computer, Volume 34, Issue 8, Page 57-66.

Wang, M. (2015) Indoor Navigation Systems Based on iBeacon fingerprinting. Master of Science Thesis.

Kriz, P., Maly, F., Kozel, T. (2016) Improving Indoor Localization Using Bluetooth Low Energy Beacons. On Hindawi. In Mobile Information Systems, Volume 2016, Article ID 2083094.

Zhuang, Y., Syed, Z., Georgy, J., El-Sheimy, N. (2015) Autonomous smartphone-based WiFi positioning system by using access points localization and crowdsourcing. In

Pervasive and Mobile Computing archive, Volume 18, Issue C, Pages 118-136.

Gast, S. M. (2014) Building Applications with iBeacon, O'Reilly Media, 1th Edition.

Neburka, J., Tlamsa, Z., Benes, V., Polak, L., Kallert, O., Bolecek, L., Sebesta, J., Kratochvil, T. (2016) Study of the Performance of RSSI based Bluetooth Smart Indoor Positioning. 26th International Conference Radioelektronika.

Zhuang, Y., Yang, J., Li, Y., Qi, L., El-Sheimy, N. (2016) Smarthphone-Based Indoor Localization with Bluetooth Low Energy Beacons. Sensors 2016, Volume 16, Page 596.

Faragher, R., Harle, R. (2015) Location fingerprinting with bluetooth low energy beacons. IEEE Journal on Selected Areas in Communications.

Jae, E., Kim, C. (2016) An Indoor Positioning Method Using a Combination of the Triangulation and Fingerprinting Techniques, Volume 33, Issue 11, Pages 2418-2428.

Cho, H., Ji, J., Chen, Z., Park, H. and Lee, W. (2015) Measuring a Distance between Things with Improved Accuracy. The 5th International Symposium on Internet of Ubiquitous and Pervasive Things, Pages 443-742.

Dixon, J. (2009) Suspension Geometry and Computation, John Wiley, John C.Dixon, 2ª Edição

Aman, M. S., Jiang H., Quint C., Yelamarthi K., Abdelgawad, A. (2016) reliability evaluation of ibeacon for Micro-Localization. IEEE 7th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON , pages. 1-5.

Análise, Implementação e Simulação de Propriedades da Cicloide em Software Livre de Geometria Dinâmica para o Ensino de Matemática

Rosário Ribeiro[1], **Antonio Oliveira-Jr**[2]

[1]FacUnicamps – Faculdade Unida de Campinas
Goiânia – GO – Brasil

[2]Instituto de Informática (INF) – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brasil

rosariocamposribeiro@gmail.com, antonio@inf.ufg.br

Abstract. *The use of technologies in the classroom environment can add value to student such as autonomy, knowledge building and learning achievement, since these technologies are used properly, creatively and critically. The cycloid curve, due to its properties, have been the one of the most studied curves over many centuries. In this work, we have analyzed, implemented and simulated the main properties of the cycloid curve in dynamic geometric open source software to assist the mathematics professors.*

Resumo. *A utilização das tecnologias, em ambiente de sala de aula, pode constituir uma mais-valia para a autonomia dos alunos, a construção de conhecimentos e a conquista de aprendizagens, desde que estas tecnologias sejam utilizadas de forma adequada, criativa e crítica. A cicloide, devido as suas propriedades, foi uma das curvas mais estudadas ao longo de muitos séculos. Neste trabalho, o objetivo é analisar, implementar e simular as principais propriedades da cicloide em software livre de Geometria Dinâmica para auxiliar o ensino de matemática.*

1. Introdução

Numa época em que as tecnologias estão cada vez mais presentes na sociedade torna-se pertinente estarem também presente em ambiente de sala de aula. Por isso compete, também, à disciplina de matemática inserir as tecnologias em sala de aula e estas são uma mais-valia para o professor desenvolver um bom trabalho [Hohenwarter et al. 2009].

Consideramos que a utilização da tecnologia no ensino contribui para o desenvolvimento das capacidades do aluno de forma diferenciada e respeita a sua individualidade. Contudo sabemos que é necessário ter cuidado na sua utilização pois pode ser bem ou mal usada e um ensino pouco cuidado, repleto de tecnologia, não promove a aprendizagem.

Uma das formas de utilizar a tecnologia é através da utilização de softwares de Geometria Dinâmica, tais como o *Cabri Geométre*, o *C.a.R*, o *Cinderella*, o *The Geometer's Sketchpad* e o *GeoGebra*. Neste trabalho, optamos pelo *GeoGebra* por ser um software livre amplamente utilizado, intuitivo, de fácil aprendizagem e ainda possui construções pré-definidas.

O estudo das curvas e das suas propriedades, por vezes é complexo para alguns níveis de ensino, pode ser administrado com recurso a um software de Geometria Dinâmica para minimizar a sua complexidade. Devido às suas propriedades a cicloide foi uma das curvas mais estudadas ao longo de muitos séculos. Neste trabalho, o objetivo é analisar, implementar e simular as principais propriedades da Cicloide em software livre de Geometria Dinâmica para auxiliar o ensino de matemática.

Portanto, analisou-se (i) o Paradoxo da Roda de Aristóteles, (ii) que a área sob uma cicloide é três vezes a área do seu círculo gerador e (iii) resolução do problema da braquistócrona. A solução do problema da braquistócrona é a cicloide invertida, portanto, implementamos duas simulações computacionais (i.e. uma discreta e outra contínua) para este problema. Dessa forma, compararamos o tempo que uma partícula demora a percorrer uma cicloide invertida com o tempo que demora a percorrer um plano inclinado considerando ambas implementações.

Para além desta Introdução, apresentamos a definição de cicloide na seção 2. Na seção 3 analisamos e implementamos o paradoxo da roda de Aristóteles e na seção 4 descrevemos uma construção em que os alunos podem verificar que a área sob uma cicloide é três vezes a área do seu círculo gerador. A seção 5 apresenta a análise, implementação e simulação do problema da braquistócrona e na seção 6 a comparação entre o movimento de uma partícula sobre a cicloide invertida e sobre o plano inclinado. As conclusões e trabalhos futuros é discutido na seção 7.

2. A Cicloide

Na Figura 1 está representada uma cicloide definida na página 617 de [Katz 2010] como «a curva traçada por um ponto no aro de uma roda a rolar ao longo de uma reta».

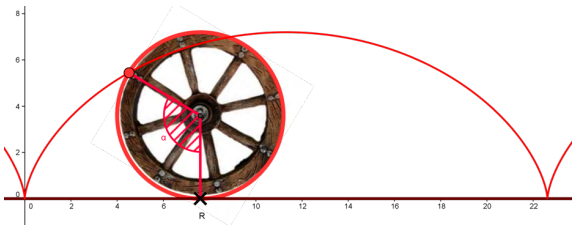


Figura 1. Cicloide

Definição 1: Seja C uma circunferência de raio r, s uma reta e P um ponto de C. Denominamos por cicloide a curva descrita pelo ponto P quando C rola sobra a reta s, sem deslizar, dada pela equação paramétrica (1).

$$\begin{aligned} x &= r\left(\alpha - \sin\alpha\right) \\ y &= r\left(1 - \cos\alpha\right) \end{aligned} \tag{1}$$

A equação cartesiana da cicloide é dada pela equação (2).

$$x = r\arccos\left(1 - \frac{y}{r}\right) - \sqrt{2ry - y^2} \tag{2}$$

3. O Paradoxo da Roda de Aristóteles

Um paradoxo é uma declaração aparentemente verdadeira que leva a uma contradição lógica, ou a uma situação que contradiz a intuição comum. Desde sempre que o infinito está associado a diversos paradoxos. Aos poucos, alguns dos paradoxos foram sendo explicados, mas ainda assim, não se trata de um assunto sem controvérsia.

Nesta seção vamos descrever o Paradoxo da Roda de Aristóteles [Drabkin 1950] e refletir sobre uma explicação possível para este conhecido paradoxo.

Consideremos duas rodas com centro no mesmo ponto e raios diferentes e a rolar em simultâneo. Quando a roda maior rola de A para A', dando uma volta completa, o aro da roda menor roda de B para B', ao longo de uma linha paralela como podemos ver na Figura 2. Temos que em cada instante, um único ponto do aro da roda grande toca no segmento AA' e um único ponto do aro da roda pequena toca no segmento BB'. Não se excluiu ponto algum de qualquer das duas circunferências. Isto leva-nos ao paradoxo da Roda de Aristóteles, uma vez que, em primeira análise podemos concluir que os perímetros das duas circunferências são iguais. E sabemos que esta conclusão não é verdadeira.

Na Figura 2 está representada uma construção que realizamos, em *GeoGebra*, do paradoxo de roda. Nesta construção os alunos devem deslocar o ponto R de forma às rodas rolarem uma volta completa e analisar a imagem. Na análise desta construção os alunos devem de ser confrontados com a possibilidade de os perímetros das duas rodas serem iguais e refletirem sobre como podem explicar o facto de:

- o perímetro ser o mesmo se os raios são diferentes;
- o perímetro ser diferente se, tal como observam, em cada instante, um único ponto do aro da roda grande toca no segmento AA' e um único ponto do aro da roda pequena toca no segmento BB', logo o perímetro tem de ser igual.

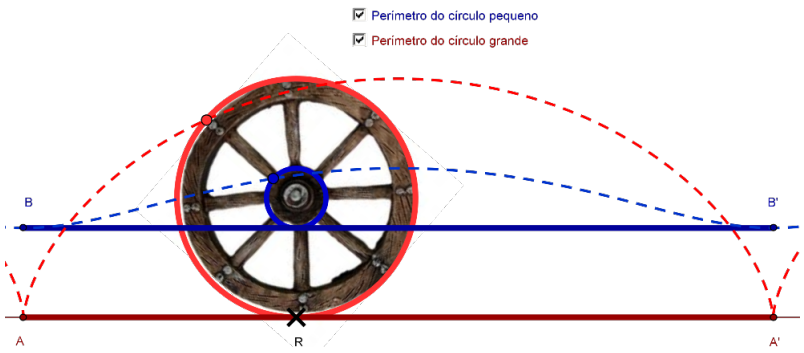


Figura 2. Paradoxo da Roda

Após a reflexão dos alunos sobre o paradoxo poderá ser apresentada uma ideia de uma possível explicação deste paradoxo. Para tal utilizamos um texto de apoio ao aluno escrito em `HTML` e realizamos algumas construções em *GeoGebra*.

A definição de circunferência utilizada mais frequentemente é que esta é o lugar geométrico de todos os pontos do plano que estão a uma dada distância (raio) de um ponto fixo (centro da circunferência). Mas a circunferência pode ser encarada como um polígono regular com um número infinito de lados.

Arquimedes usou este conceito para determinar o perímetro de um círculo. Este calculou o valor do perímetro de um círculo a partir da sucessão dos perímetros dos polígonos regulares inscritos nele. À medida que o número de lados aumenta, o perímetro do polígono aproxima-se cada vez mais do perímetro do círculo. Assim, o perímetro do círculo corresponde ao limite da sucessão dos perímetros correspondentes ao limite da sucessão dos perímetros dos polígonos inscritos, quando os números de lados tende para infinito conforme apresentado na Figura 3.

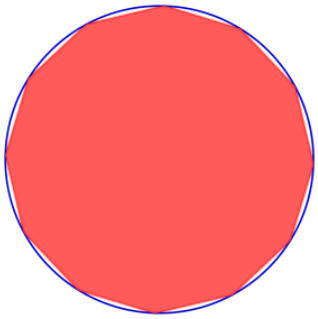


Figura 3. Perímetro do círculo corresponde ao limite da sucessão dos perímetros dos polígonos inscritos.

Para dar a ideia da explicação do paradoxo da roda fizemos uma construção, em *GeoGebra*, onde recorremos à noção de circunferência usada por Arquimedes. Nesta construção os alunos podiam observar o que acontece quando fazemos rolar dois polígonos regulares com um número de lados finitos e concêntricos (utilizamos um polígono regular com seis lados, Figura 4). Essa construção será analisada com os alunos de forma a deduzirem o que acontece quando o número de lados do polígono regular é infinito.

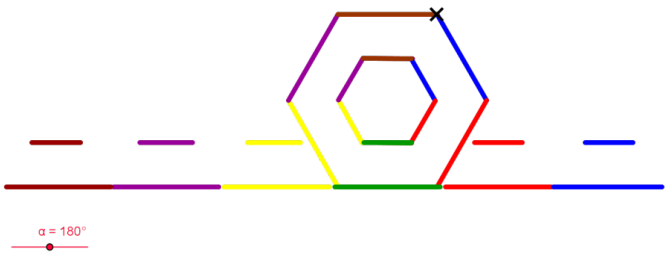


Figura 4. Explicação do Paradoxo

4. Área sob um Arco de Cicloide

O matemático Gilles Roberval, por volta de 1637, demonstrou que a área sob uma cicloide é três vezes a área do círculo gerador. Para tal, este matemático utilizou o que hoje é conhecido pelo *Princípio de Cavalieri*, mas que já havia sido usado por matemáticos como Herão e Galileu.

> **Princípio de Cavalieri:** ≪Se duas Figuras planas têm alturas iguais e se secções feitas por linha paralelas às bases e a distâncias iguais delas são sempre na mesma proporção, então as Figuras planas também estão nesta proporção.≫ [Katz 2010]

Gilles Roberval definiu ainda uma nova curva à qual chamou *companheira da cicloide*. Esta era definida como sendo o lugar geométrico definido pelo ponto $(x(R),\ y(R'))$ da Figura 5.

Na Figura 5 está representada uma construção feita em *GeoGebra*. Nesta construção os alunos podem:

- deslocar o ponto R e verificar que os segmentos a verde têm o mesmo comprimento;
- deslocar o ponto I e verificar que os segmentos a roxo têm o mesmo comprimento.

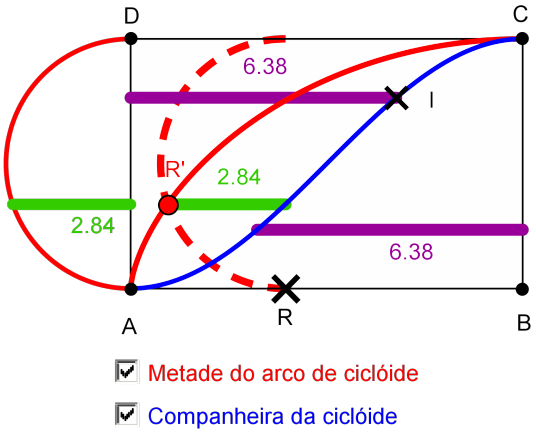


Figura 5. Área sob um arco de cicloide

Dependendo do nível de ensino ao qual este atividade seja aplicada os alunos mais empenhados conseguirão concluir, após a análise da construção, que a área sob uma cicloide é três vezes a área do círculo gerador.

No seu raciocínio os alunos deverão ter em conta que:

- a área entre metade do arco de cicloide e a companheira da cicloide é igual a metade da área do círculo gerador;
- a área do retângulo $[ABCD]$ é igual ao produto de metade do perímetro do círculo pelo o seu diâmetro, isto é, $2\pi r^2$, onde r é o raio do círculo gerador;
- a área sob a curva companheira da cicloide é metade da área do retângulo $[ABCD]$ logo é igual à área do círculo gerador, ou seja, πr^2;
- a área sob metade do arco de cicloide é igual a $\frac{3}{2}$ da área do círculo;
- a área sob o arco inteiro é três vezes a área do círculo.

5. O Problema da Braquistócrona

O problema da Braquistócrona foi apresentado pela primeira vez aos leitores da revista científica *Acta Eruditorum* pelo matemático suíço Johann Bernoulli, em 1696, como um problema que já havia resolvido. Na publicação Johann Bernoulli desafiava os filósofos da época a apresentarem soluções para o problema. O enunciado original é o seguinte [Ross 2009]:

> ≪*Datis in plano verticali duobus punctis A et B assignare mobili M, viam AMB per quam gravitate sua descends et moveri incipiens a puncto A, brevissimo tempore perveniat ad alterum punctum B.*≫

Em português contemporâneo pode ser formulado da seguinte forma:

> ≪Sejam A e B dois pontos de um plano vertical. Encontre a curva na qual uma partícula M, sujeita somente a ação da gravidade, descreve a trajetória mais rápida entre os pontos A e B.≫

O Problema da Braquistócrona consiste, portanto, em encontrar a curva ao longo da qual uma partícula se desloca entre dois pontos no menor intervalo de tempo, sendo que esta partícula está sujeita a um campo gravitacional constante, sem atrito e com velocidade inicial nula. Para além disso, é suposto que os dois pontos não se encontrem sobre a mesma reta horizontal ou sobre a mesma reta vertical. Uma vez que, se os pontos estiverem sobre a mesma reta horizontal não haverá movimento e se os pontos estiverem sobre a mesma reta vertical a solução do problema é a própria reta.

A curva que resolve este problema denomina-se Braquistócrona, do grego brachistos (mínimo) e chronos (tempo). A solução do Problema da Braquistócrona é uma cicloide invertida.

Efetuamos diversas tentativas para fazer uma simulação computacional do Problema da Braquistócrona no caso particular em que pretendemos encontrar a curva ao longo da qual uma partícula, com forma de bola e de massa 1, se desloca entre os pontos $A = (0,\, 0)$ e $B = (\pi,\, -2)$ no menor intervalo de tempo. Neste caso a solução do Problema da Braquistócrona é a cicloide invertida cujo o círculo gerador tem raio 1. Das várias tentativas que realizamos decidimos descrever neste documento uma simulação em que consideramos o tempo discreto e outra simulação em que consideramos o tempo continuo. Embora, neste caso, a discretização não seja fundamental, é importante que se conheça a técnica para o caso de se estar a lidar com uma curva para a qual não se consigam fazer a simulação de outro modo.

Vamos começar por descrever a construção que realizamos com base na discretização do tempo.

Para construir esta simulação do Problema da Braquistócrona foi necessário utilizar a folha de cálculo do programa de geometria dinâmica *GeoGebra*. Começamos esta construção com a definição das quantidades $g = 9,8$ e $t_1 = 0,01$ e com a criação de uma ferramenta que, aplicada a um ponto P com ordenada entre -2 e 0 (inclusive), dá a interseção do arco de cicloide com a reta horizontal que passa pelo ponto P. Mais precisamente, aplicando a ferramenta ao ponto de coordenadas $(x,\, y)$, obtém-se o ponto

$$(\arccos(y+1) - \sin(\arccos(y+1)),\, y) = \left(\arccos(y+1) - \sqrt{-y^2-2y},\, y\right)$$

determinado com base nas equações, paramétrica (1) e cartesiana (2), da cicloide de raio 1.

Inicialmente pensamos que seria impossível, com auxilio do *software* de geometria dinâmica utilizado, resolver o problema de forma continua. Tivemos dificuldades na realização da simulação em que consideramos o tempo discreto. Estas dificuldades devem-se aos factos do *software* de Geometria Dinâmica utilizado não efetuar interseções entre uma curva e outros tipos de representações (como por exemplo, retas, circunferências, gráficos, curvas) e não representar graficamente funções escritas em função de y (isto é, como a função dada pela equação cartesiana (2)). Para resolver esta dificuldade, optamos por na execução desta simulação discretizar o tempo em intervalos de

tempo ($t_1 = 0,01$) pequenos e prever, por cálculos efetuados na Folha de Cálculo, qual a posição da partícula nos instantes seguinte, como podemos ver na Figura 6.

Folha de Cálculo

	A	B	C	D
1	(0, 0)	0	0	$(x - 0.08)^2 + y^2 = 0.0064$
2	(0.00001, -0.00049)	0.03131	0.01	$(x - 0.08)^2 + (y - 0.00076)^2 = 0.0064$
3	(0.00003, -0.00147)	0.05423	0.02	$(x - 0.08)^2 + (y - 0.0007)^2 = 0.0064$
4	(0.00008, -0.00317)	0.0796	0.03	$x - 0.08002)^2 + (y - 0.00002)^2 = 0.0064$
5	(0.0002, -0.00566)	0.10641	0.04	$(x - 0.08009)^2 + (y + 0.0014)^2 = 0.0064$
6	(0.0004, -0.00898)	0.13412	0.05	$x - 0.08022)^2 + (y + 0.00362)^2 = 0.0064$

Figura 6. Folha de Cálculo Utilizada na Simulação

De seguida vamos descrever a função de cada uma das quatro colunas da Folha de Cálculo utilizadas na construção.

- **Coluna A:** É uma coluna de pontos, dos quais o primeiro é $(0,\ 0)$e o segundo é o resultado de se aplicar a ferramenta a $\left(0,\ -\frac{gt_1^2}{2}\right)$. A partir da terceira linha, cada ponto A_{i+1} é obtido de A_i pelo seguinte processo:
  - soma-se a A_i o vetor tangente unitário à cicloide no ponto onde A_i toca na cicloide, multiplicado pela velocidade escalar da bola nesse momento (que é o produto de t_1 pela raiz quadrada do produto de $2g$ pelo simétrico da ordenada de A_i , em que g representa a aceleração da gravidade e o simétrico da ordenada de A_i representa o deslocamento vertical da partícula, como é demonstrado mais abaixo);
  - aplica-se a ferramenta ao ponto anterior.

Ou seja,

$$A_{i+1} = Ferramenta\left[A_i + t_1\sqrt{-2gy(A_i)} \times \left(\sin\left(\frac{B_i}{2}\right),\ -\cos\left(\frac{B_i}{2}\right)\right)\right]$$

Este método não poderia ser aplicado logo no primeiro passo devido à velocidade inicial ser nula.

- **Coluna B:** Se $A_i = (\theta - \sin(\theta),\ -1 + \cos(\theta))$ então $B_i = \theta$, isto é, $B_i = \arccos(1 + y(A_i))$. Esta coluna é necessária para calcular A_{i+1} uma vez que o vetor tangente unitário é $\left(\sin\left(\frac{\theta}{2}\right), -\cos\left(\frac{\theta}{2}\right)\right)$.
- **Coluna C:** $C_i = t_1 \times i$**.**
- **Coluna D:** Equação do círculo de raio r (partícula) tangente à cicloide no ponto A_i, isto é, $D_i = Circunferência\left[A_i + r\left(\cos\left(\frac{B_i}{2}\right),\ \sin\left(\frac{B_i}{2}\right)\right),\ A_i\right]$. O valor de r é controlável por um seletor. Além disso, o círculo só é visível quando t (que toma valores de 0 a $1,5$ com saltos de t_1) for igual a C_i.

Obtivemos destas forma a animação do movimento de uma partícula, com forma de bola, sobre a cicloide representado na Figura 7. Nesta construção, em *GeoGebra*, os alunos podem alterar o raio da partícula entre 0 e $0,2$, visualizar a posição da partícula num dado

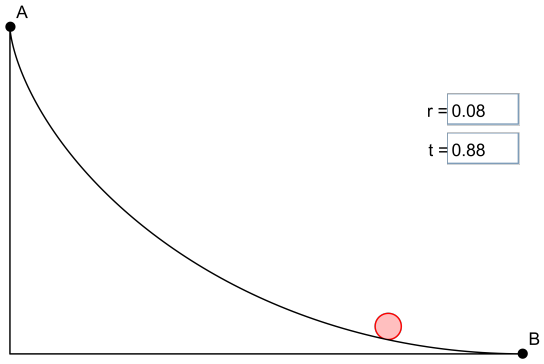


Figura 7. Braquistócrona

instante de tempo entre 0 e $1,5$ e verificar o tempo que esta leva a percorrer a cicloide invertida.

Para elaboramos a simulação computacional na qual consideramos o tempo continuo, para o nosso caso particular do Problema da Braquistócrona, tivemos de encontrar uma expressão analítica da curva Braquistócrona que satisfaça uma duas propriedades seguintes:

1ª Propriedade: a velocidade escalar (v) de um corpo (de massa m) que se desloca sobre uma curva (partindo da altura 0 e com velocidade nula) num campo sujeito apenas à força gravítica é $\sqrt{2gy}$ em que g é a aceleração da gravidade e y representa o deslocamento vertical da partícula;

2ª Propriedade: a aceleração tangencial tangencial ($\vec{a}_T(t)$) de um corpo (de massa m) que se desloca sobre uma curva (partindo da altura 0 e com velocidade nula) num campo sujeito apenas à força gravítica é, a cada instante t, igual ao produto de $-g\sin(\theta)$ pelo vetor tangente unitário $\vec{T}(t)$ em que g é a aceleração da gravidade e θ é o ângulo que este último vetor faz com a horizontal.

Apesar de não apresentarmos uma demonstração neste artigo pode-se demonstrar que estas duas propriedades são equivalentes.

Vamos começar por demonstrar que a expressão analítica da curva Braquistócrona que resolve o nosso problema tem que respeitar a **1ª Propriedade**, obter uma expressão analítica que respeite a **1ª Propriedade** e verificar que esta expressão analítica também respeita a **2ª Propriedade**.

Pelo princípio da conservação da energia sabemos que a energia total (E) de um sistema isolado é conservada. Isto é, sabemos que a soma das duas componentes da energia total (energia cinética e energia potencial) respeita a seguinte igualdade

$$E = mgy + \frac{1}{2}mv^2 = k$$

em que m é a massa do corpo, g a aceleração da gravidade, y a altura a que o corpo se encontra, v a velocidade escalar e k uma constante.

Consideremos um corpo cuja altura inicial é 0 e este corpo começa-se a mover logo a sua velocidade inicial é nula e, por conseguinte, a energia potencial e a energia

cinética são nulas, isto é, obtemos a igualdade 3.

$$mgy + \frac{1}{2}mv^2 = 0 \tag{3}$$

Quando este corpo atinge a altura $-y$ $(y \geq 0)$ a igualdade 3 equivale a afirmar que

$$v = \sqrt{2gy}. \tag{4}$$

Demonstramos assim que um corpo nas condições acima respeita a **1ª Propriedade**. Vamos agora encontrar a expressão analítica da curva Braquistócrona que respeita esta propriedade.

Uma expressão analítica da curva que resolve o nosso problema é dada pela função $c : [0,\, \pi] \longrightarrow R$ definida pela igualdade 5.

$$c\,(t) = (t - \sin\,(t)\,,\, -1 + \cos\,(t)) \tag{5}$$

Mas como esta expressão analítica da curva Braquistócrona (c) não envolve a aceleração da gravidade não conseguimos obter, através desta expressão, a velocidade escalar dada pela equação 4.

Portanto, vamos re-parametrizar c de modo à velocidade escalar ser dada pela equação 4. Pretendemos então substituir c pela função $c \circ s$, onde s é alguma bijeção derivável de algum intervalo de tempo $[0,\, T]$ em $[0,\, \pi]$. Então, como

$$\begin{aligned} c'\,(t) &= (1 - \cos\,(t)\,,\, -\sin\,(t)) \\ &= \left(1 - \left(1 - 2\sin^2\left(\tfrac{t}{2}\right)\right),\, -2\sin\left(\tfrac{t}{2}\right)\cos\left(\tfrac{t}{2}\right)\right) \\ &= 2\sin\left(\tfrac{t}{2}\right)\left(\sin\left(\tfrac{t}{2}\right),\, -\cos\left(\tfrac{t}{2}\right)\right), \end{aligned}$$

como $\left\|\left(\sin\left(\frac{t}{2}\right),\, -\cos\left(\frac{t}{2}\right)\right)\right\| = 1$ e como $2\sin\left(\frac{t}{2}\right) \geq 0$, a velocidade escalar no instante t é $2\sin\left(\frac{t}{2}\right)$. Logo, se se considerar $c \circ s$, a velocidade escalar no instante t será $2\sin\left(\frac{s(t)}{2}\right) s'\,(t)$ e afirmar que se tem 4 é afirmar que:

$$\begin{aligned} 2\sin\left(\tfrac{s(t)}{2}\right) s'\,(t) &= \sqrt{g\,(1 - \cos\,(s\,(t)))} \\ &= 2\sqrt{g}\sin\left(\tfrac{s(t)}{2}\right), \end{aligned}$$

ou seja, que $s'\,(t) = \sqrt{g}$. Logo, $s\,(t) = \sqrt{g}t$. Assim sendo, podemos definir a expressão analítica da curva Braquistócrona que resolve o nosso problema, pois respeita a **1ª Propriedade,** pela função:

$$\begin{array}{rccc} c_1 : & \left[0,\, \frac{\pi}{\sqrt{g}}\right] & \longrightarrow & R \\ & t & \longmapsto & \left(\sqrt{g}t - \sin\left(\sqrt{g}t\right),\, -1 + \cos\left(\sqrt{g}t\right)\right). \end{array}$$

Com esta definição c_1 da curva Braquistócrona, o ponto inicial é $(0,\, 0)$ e a velocidade inicial é nula como pretendíamos.

Quando um corpo (de massa m) que se desloca sobre uma curva num campo sujeito apenas à ação da gravidade (partindo da altura 0 e com velocidade nula) temos que:

$$\vec{a} = -g\,\vec{e}_y = \vec{a}_N\,(t) + \vec{a}_T\,(t) = -g\cos(\theta)\,\vec{N}\,(t) - g\sin(\theta)\,\vec{T}\,(t)$$

em que g é a aceleração da gravidade, $\vec{e}_y$ o vetor unitário segundo o eixo dos yy, $\vec{a}_N\,(t)$ a aceleração normal, $\vec{N}\,(t)$ vetor normal unitário, $\vec{a}_T\,(t)$ aceleração tangencial, $\vec{T}\,(t)$vetor tangente unitário e θ é o ângulo que este último vetor faz com a horizontal.

Vamos agora demonstrar que c_1 respeita a **2ª Propriedade**, isto é, que a aceleração tangencial $\vec{a}_T\,(t)$ é, a cada instante t, igual ao produto de $-g\sin(\theta)$ pelo vetor tangente unitário $\vec{T}\,(t)$. Um cálculo direto revela que:

$$\vec{a}_T\,(t) = g\cos\left(\frac{\sqrt{g}t}{2}\right)\underbrace{\left(\sin\left(\frac{\sqrt{g}t}{2}\right),\,-\cos\left(\frac{\sqrt{g}t}{2}\right)\right)}. \tag{6}$$

Mas

$$\begin{aligned}\vec{T}\,(t)=\left(\sin\left(\tfrac{\sqrt{g}t}{2}\right),\,-\cos\left(\tfrac{\sqrt{g}t}{2}\right)\right) &= \left(\cos\left(\tfrac{\pi}{2}-\tfrac{\sqrt{g}t}{2}\right),\,-\sin\left(\tfrac{\pi}{2}-\tfrac{\sqrt{g}t}{2}\right)\right)\\ &= \left(\cos\left(\tfrac{\sqrt{g}t}{2}-\tfrac{\pi}{2}\right),\,\sin\left(\tfrac{\sqrt{g}t}{2}-\tfrac{\pi}{2}\right)\right).\end{aligned}$$

Logo, $\theta = \frac{\sqrt{g}t}{2} - \frac{\pi}{2}$ e, portanto,

$$\begin{aligned}-g\sin(\theta) &= -g\sin\left(\tfrac{\sqrt{g}t}{2}-\tfrac{\pi}{2}\right)\\ &= g\sin\left(\tfrac{\pi}{2}-\tfrac{\sqrt{g}t}{2}\right)\\ &= g\cos\left(\tfrac{\sqrt{g}t}{2}\right)\end{aligned}$$

o que, juntamente com 6, confirma que $\vec{a}_T\,(t) = -g\sin(\theta)\,\vec{T}\,(t)$.

Com base expressão analítica de c_1 construimos uma simulação onde os alunos podem alterar o raio da partícula entre 0 e $0,2$, visualizar a posição da partícula num dado instante de tempo entre 0 e $1,2$ e verificar o tempo exato que a partícula leva a percorrer a cicloide invertida $\left(t = \frac{\pi}{\sqrt{g}} \simeq 1,00354\right)$. O aspecto visual desta construção é igual ao da Figura 7.

6. O Problema da Braquistócrona vs. Movimento Sobre o Plano Inclinado

Nesta seção vamos comparar o tempo que uma partícula demora a percorrer uma cicloide invertida e um plano inclinado, nas nossas simulações, com o tempo previsto analiticamente.

Para tal fizemos duas construções em *GeoGebra*, com aspecto visual igual, retratadas na Figura 8. Nestas simulações os estudantes podem comparar o movimento de uma partícula, com forma de bola e de massa 1, que se desloca entre os pontos $A = (0,\,0)$ e $B = (\pi,\,-2)$ sobre a cicloide invertida (numa considerando o tempo discreto e na outra continuo) e sobre um plano inclinado. Os alunos nestas simulações podem constatar que, realmente, a partícula chega mais rapidamente a $B = (\pi,\,-2)$ quando se desloca sobre a cicloide invertida.

As simulações do movimento de uma partícula sobre a cicloide invertida desta seção foi construida da mesma forma das simulações descrita na seção 5. Tendo por

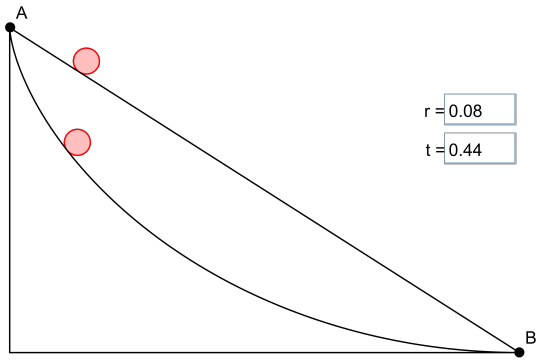


Figura 8. Comparação entre o movimento de uma partícula sobre a cicloide invertida e sobre o plano inclinado.

base estes apontamentos a equação posição de um corpo de massa 1 que desliza desde a posição inicial $x_0 = 0$, sem rolar, sem atrito, com velocidade inicial nula ($v_0 = 0$), num plano inclinado que faz um ângulo θ com a horizontal é dada pela equação (7).

$$x\left(t\right) = \frac{1}{2}t^2 g \sin\theta \tag{7}$$

Para calcular analiticamente o tempo que uma partícula demora a percorrer, de A para B, uma cicloide invertida e um plano inclinado para determinar o tempo total ($t\left(x\right)$) que a partícula demora a se deslocar de A para B pode ser determinado pelo seguinte integral

$$t\left(x\right) = \frac{1}{\sqrt{2g}} \int_0^{\pi} \sqrt{\frac{1 + y'^2\left(s\right)}{-y\left(s\right)}} ds$$

no qual

$$s\left(x\right) = \int_0^{x} \sqrt{1 + y'^2\left(\xi\right)} d\xi.$$

Curva	x	y	$t\left(x\right)$	$t\left(x\right)$(simulação discreta)
Cicloide invertida	$\theta - \sin\left(\theta\right)$	$-1 + \cos\left(\theta\right)$	$1,00354$	$]0,99,\ 1]$
Plano inclinado	x	$-\frac{2}{\pi}x$	$1,18965$	$]1,18;\ 1,19]$

Tabela 1. Tempo que uma partícula demora a se deslocar de A para B sobre uma dada curva.

Com base nestas integrais obtivemos os valores da Tabela 1. No caso da simulação que contem a cicloide invertida em que consideramos o tempo continuo o tempo que a partícula demora a chegar a B é igual ao valor obtido analiticamente, por isso, não os diferenciamos na tabela. Tal como era de esperar constatamos que, realmente, a partícula chega mais rapidamente a $B = (\pi,\ -2)$ quando se desloca sobre a cicloide invertida.

Quando comparamos e analisamos o que a partícula leva a chegar ao ponto B obtido analiticamente com o da simulação da Braquistócrona em que consideramos o tempo discreto verificamos que:

- no caso da cicloide invertida o tempo da nossa simulação é ligeiramente inferior;
- no caso do plano inclinado o tempo é igual.

Pensamos que a ligeira diferença, no caso da cicloide invertida, se deve à discretização do tempo em intervalos de tempo ($t_1 = 0,01$) e aos arredondamentos.

7. Conclusão

A inserção da tecnologia no ensino contribui para o desenvolvimento das capacidades do aluno de forma diferenciada e respeita a sua individualidade. Compete ao professor, para maximizar as potencialidades da tecnologia, selecionar ou criar tarefas matemáticas, recorrendo a esta ferramenta digital de forma correta e eficiente, no sentido de melhorar as oportunidades de aprendizagem dos seus alunos.

Neste trabalho, utilizamos o software livre GeoGebra para analisar, implementar e simular as principais propriedades da Cicloide. Apresentou-se o Paradoxo da Roda de Aristóteles, que a área sob uma cicloide é três vezes a área do seu círculo gerador e a resolução do problema da braquistócrona. Apresentamos os resultados da comparação entre o movimento de uma partícula sobre cicloide invertida e o plano inclinado. Os resultados demonstram que para ambas as simulações discreta e contínua, que o movimento é mais rápido na cicloide invertida.

Podemos concluir que o GeoGebra é um software livre que contribui de forma eficaz para melhorar o processo de aprendizagem. Acreditamos que sem o recurso a esta tecnologia dificilmente o discente compreenderia as propriedades apresentadas neste artigo.

Como trabalhos futuros, pretendemos analisar outras propriedades da cicloide bem como abordar os três problemas clássicos da geometria grega. Posteriormente, utilizar as implementações em um estudo de caso real.

Referências

Drabkin, I. E. (1950). Aristotle's wheel: Notes on the history of a paradox. *Osiris*, 9:162–198.

Hohenwarter, J., Hohenwarter, M., and Lavicza, Z. (2009). Introducing dynamic mathematics software to secondary school teachers: the case of geogebra. *Journal of Computers in Mathematics and Science Teaching*, 28(2):135–146.

Katz, V. J. (2010). *Historia da Matematica*. Fundacao Calouste Gulbenkian, 15th edition.

Ross, M. (2009). *A Primer on Pontryagin's Principle in Optimal Control*. Collegiate Publishers.

Uso de Softwares Educacionais como Recurso Didático-Pedagógico para o Ensino da Geografia

Raimundo Nonato de Araujo Soares Neto[1], **Sônia Regina Gouvêa Rezende**[1], **Tiago do Carmo Nogueira**[2]

[1]Universidade Estadual de Goiás (UEG)

[2]Instituto Federal de Mato Grosso

raimundonsneto@gmail.com, profsoniarez@gmail.com, tiago.nogueira@bag.ifmt.edu.br

Resumo. *Nesta pesquisa, realizada em duas etapas, buscou-se mensurar, a partir de um estudo de caso, o impacto proporcionado pela utilização de softwares educativos, KGeography e Marble do sistema Linux Educacional, no 8º ano do ensino fundamental. Primeiramente, realizaram-se testes comparando o método tradicional de ensino da Geografia com o método sob a mediação do KGeography e Marble, tendo como objetivo aferir o efeito desses softwares sobre o processo de aprendizagem dos alunos. Posteriormente, a partir desta experiência, investigou-se a percepção dos estudantes em relação aos softwares utilizados, buscando conhecer suas opiniões sobre tais instrumentos de apoio didático-pedagógico. Estatisticamente, comprovou-se elevação no nível de aprendizado dos alunos após utilização das ferramentas tecnológicos propostas, bem como satisfação a partir da interação dos discentes com tais softwares educacionais. Assim, considerou-se que os recursos computacionais potencializaram o aprendizado e, despertaram interesses desses alunos na obtenção do conhecimento, melhorando a relação ensino-aprendizagem.*

Abstract. *In this research, carried out in two stages, we tried to measure, from a case study, the impact of educational software, KGeography and Marble of the Linux Educacional system, in the 8th year of elementary school. Firstly, tests were performed comparing the traditional method of teaching Geography with the method under the mediation of KGeography and Marble, aiming to assess the effect of these software on the learning process of students. Subsequently, from this experience, we investigated the students' perceptions regarding the software used, seeking to know their opinions about such didactic-pedagogical support instruments. Statistically, there was an increase in the student's level of learning after using the proposed technological tools, as well as satisfaction from the interaction of the students with such educational software. Thus, it was considered that the computational resources increased the learning and awakened interests of these students in obtaining knowledge, improving the teaching-learning relationship.*

1. Introdução

Desde o final da década de 1990, tem-se tornado comum a utilização de recursos computacionais no processo ensino-aprendizagem, dado o crescente desenvolvimento de softwares

educativos, aliados a disseminação e expansão da internet. O decurso histórico tem revelado como o avanço das Tecnologias da Informação e Comunicação (TIC) influenciam diversas esferas da sociedade, inclusive a educação, dada as suas potencialidades. Por ser capaz de digitalizar todo e qualquer tipo de informação, estas tecnologias auxiliam desde ações corriqueiras até as mais complexas atividades humanas, tornando-se indispensável para a realização das tarefas.

Na educação, as tecnologias computacionais se colocam como instrumento que potencializa e enriquece o aprendizado em virtude de suas múltiplas facetas, produzindo perspectivas inimagináveis e auxiliando as práticas pedagógicas. Utilizando os recursos computacionais de forma adequada professores podem tornar suas aulas mais atrativas, dinâmicas, criativas e produtivas, pois seu uso auxiliará a explicação e exploração do conteúdo ministrado; colaborando assim, para maior progresso dos alunos. Ademais, os estudantes necessitam cada vez mais de informações concisas, convincentes e próximas da realidade; possíveis por meio da inserção das tecnologias computacionais com instrumento didático; aliado aos já existentes.

As tecnologias fazem parte do contexto da maioria dos estudantes, o que torna interessante sua aplicação no âmbito escolar dada sua capacidade de estimular ideias, de desenvolver o aprendizado e produzir maior interação e diálogo entre alunos e professores. Estratégias inovadoras para o aperfeiçoamento do processo educacional são perfeitamente possíveis a partir da inserção das tecnologias computacionais nos planejamentos e currículos escolares. Destarte, trata-se de ferramentas pedagógicas digitais capazes de, não apenas tornar as aulas mais atraentes e inovadoras, mas também estimular os alunos a aprenderem; despertar a curiosidade e as novas descobertas; auxiliar na melhoria do desempenho dos alunos; e possivelmente, contribuir para a diminuição das reprovações e da evasão escolar e aprimorar a qualidade da educação daqueles que entenderam o potencial transformador dessas tecnologias.

[Lima Filho 2014] chama atenção para o continuo e crescente incorporação do componente tecnológico à educação, como um recurso didático. Segundo este autor, recursos computacionais caracterizam-se como valioso instrumento direcionado para o desenvolvimento do ato de ensinar e de aprender, reunindo características que permitem novos olhares e práticas nos percursos da construção do conhecimento.

Nesse sentido, buscou-se mensurar os impactos proporcionados pela utilização de softwares educativos (KGeography[1] e Marble - Linux Educacional[2]) no estudo de Geografia, especificamente Cartografia, dos alunos do 8º ano do ensino fundamental, de uma escola estadual em Goiânia – Goiás.

Para tanto, por meio de uma abordagem quali-quantitativa, na modalidade estudo de caso, esta pesquisa constitui-se em duas etapas. Incialmente foi realizada uma comparação entre o método tradicional de ensino da Geografia e o método de ensino com a mediação dos softwares propostos, tendo como objetivo mensurar o impacto desses softwares sobre o nível de aprendizado dos alunos. Posteriormente, na segunda etapa o objetivo foi conhecer a percepção dos alunos em relação a essa experiência; ou seja, o uso do recurso computacional (softwares do Linux Educacional), como estratégias para o

[1]https://edu.kde.org/kgeography/
[2]https://marble.kde.org/

ensino da Geografia.

Os resultados revelaram significativa melhora dos estudantes na absorção dos conteúdos em Cartografia a partir da utilização de tais softwares, conforme comprovou-se estatisticamente. Além disso, foi possível conhecer as impressões dos estudantes sobre o processo ensino-aprendizagem com mediação tecnológica. Verificou-se interesse superior a 90% dos alunos por esta estratégia de ensino, demonstrando-se favoráveis a sua aplicação e consolidação como instrumento didático-pedagógico.

2. A Informática como Recurso Didático-Pedagógico: Ensino da Geografia a partir do Viés Tecnológico

A adoção da informática na educação não apenas potencializa o processo de aprendizagem e a disseminação do conhecimento; mas também, acarreta melhoras quanto à qualidade pedagógica, tornando-se instrumentos didáticos de ensino que podem e devem estar a serviço do processo de construção e assimilação do conhecimento dos aprendizes. [Werthein 2000], antecedendo os apontamentos de [Valente 2009], destaca as novas potencialidades possíveis, a partir do uso das novas tecnologias em sala de aula.

A motivação para aplicação dos recursos midiático-tecnológicos se dá justamente pela as inúmeras possibilidades que tais recursos proporcionam para os docentes e discentes, contribuindo bastante para a exploração do conteúdo estudado e elevação do saber. Segundo [Maciel 2009] é importante que o professor repense sua prática pedagógica, pois é necessário tirar proveito das tecnologias computacionais em sala de aula. Para este autor é necessário organizar a prática de ensino a ser usada, estruturando-a enquanto técnica, selecionando as tecnologias físicas a serem usadas (tal como computadores e capacidade de acesso a Internet) e estabelecendo a comunicação entre os envolvidos durante o desenvolvimento da prática pedagógica.

Para [Tajra 2011] o professor deve ser capaz de elaborar um processo de ensino-aprendizagem de forma mais aberta, flexível, inovadora e contínua, a partir das novas tecnologias computacionais exigindo de si uma melhor formação teórica e comunicacional; uma vez que, quanto maior for o número de informações com as quais se depara mais complexo se tornará todo esse processo, possível de ser dinamizado com ferramentas e softwares educacionais. Obviamente que não se pretende aqui substituir o tradicional sistema educativo existente, mas aliá-lo às inúmeras ferramentas computacionais para a melhoria da qualidade de ensino; e distribuição da informação e conhecimento na educação.

Nesse sentido, acredita-se que a informática e o conjunto de tecnologias informacionais passaram a permear a sociedade, contribuindo para o acesso a uma enorme e diversificada quantidade de informação, corroborando para a constituição de inauditas maneiras de interação, organização e atividades sociais, proporcionando novas formas de socialização; em que as relações virtuais passaram a completar as relações pessoais, fazendo com que até mesmo o processo de ensino e aprendizagem pudesse usufruir de maior interatividade na atividade pedagógica.

[Mendes 2011] destaca que ao pensar a relação entre escola e sociedade da informação é preciso considerar a existência de novas formas de aprender, propiciadas pelas Tecnologias da Informação e Comunicação (TIC). [Mendes 2011] ainda afirma que

há possibilidade da escola inserir em seu contexto, novas formas de ensinar, apontando que a escola necessita estar inserida no mundo contemporâneo; ou seja, inserir no seu ambiente os avanços tecnológicos. O sucesso da educação com um todo não está atrelado à obrigatoriedade de submissão persuasiva proporcionada pelas mudanças tecnológicas.

No entanto, o uso apropriado das Tecnologias da Informação e Comunicação (TIC) pode enriquecer ainda mais o magistério, dado justamente ao caráter interativo que o professor do século XXI pode fazer uso, visando o enriquecimento de sua didática para a capacitação dos alunos; isto é, por meio da utilização de inúmeros recursos proporcionados pelas novas tecnologias. O papel e a importância do professor continua o mesmo, ele é o principal agente no processo educacional; contudo, agora este agente de transformação possui mais recursos para o desenvolvimento do seu trabalho.

Segundo [Almeida 2007], o uso das TIC proporciona ao docente, concepções de aprendizagem e pressupostos pedagógicos de elevado potencial e que estão em sintonia a conceitos e conteúdos abordados a cada nível de aprendizagem. Deve-se vislumbrar as tecnologias na educação em diversas vertentes e modalidades de ensino, desde apoio ao ensino presencial quanto ao ensino a distancia e à autoaprendizagem. É importante evidenciar que o uso de softwares aplicados como ferramenta e metodologia de ensino permite certa autonomia na forma de pensar do aluno, dada à interatividade e hipertextualidade das Tecnologias da Informação e Comunicação em virtude de suas potencialidades. O uso de softwares educativos aperfeiçoa de maneira impressionante a produção pedagógica; entretanto, é importante salientar que sua utilização deve ser contextualizada e em sintonia com os conteúdos administrados pelo professor em sala de aula. Além disso, devem-se possuir objetivos estabelecidos de forma clara e bem definidos por parte dos professores e a escola para o correto uso das tecnologias computacionais, para se atingir o fim desejado; isto é, melhorar a relação ensino-aprendizagem.

Sendo assim, notavelmente, tornou-se prática a utilização das Tecnologias da Informação e Comunicação (TIC) em Geografia, especificamente, na Cartografia. Percebe-se vertiginoso crescimento de softwares e ferramentas com fins geográficos, via internet ou para instalação, que passaram a ser usados não apenas em atividades de ensino e pesquisa, mas em meras ações cotidianas. Tornou-se comum a utilização de aplicativos e sites com foco em Geografia, seja para uma finalidade simples como localização espacial, seja em atividades mais complexas como as voltadas para o geoprocessamento ou para o estudo de áreas espaciais e do ambiente terrestre.

É bastante ampla a relação informática e Geografia dada à infinidade de recursos e possibilidades para realização de diversas atividades. Na educação básica tem-se destacado a utilização de atlas digitais, livros eletrônicos, enciclopédias geográficas, jogos e softwares educacionais focados principalmente em Cartografia, possibilitando uma pluralidade de opções para o ensino e aprendizagem.

Levantamentos teóricos indicam uma significativa relação entre Geografia e as TIC, bem como sua relevante aplicação na educação. [Bezerra 2014] afirma que tais tecnologias possibilitam a atratividade como também proporcionam a interatividade entre educando e o objeto de estudo, transmitida ou mediada pelo professor". [Rodrigues and Sousa 2012], sobre tal perspectiva, afirmam que, as TIC, voltadas para o ensino da Geografia podem, proporcionar aos alunos uma visão mais ampla e crítica so-

bre o estudo geográfico saindo do tradicionalismo dos livros e ingressando em um mundo com possibilidades tridimensionais, e de informações em tempo real seja em escala local ou em escala global.

Dentre os inúmeros recursos tecnológicos disponíveis para a educação, o software livre Linux Educacional chama atenção como ferramenta no processo de ensino-aprendizagem, dado seu pacote de aplicativos educativos para várias disciplinas; inclusive a Geografia, onde se destacam o KGeography e Marble, que foram utilizados nesta pesquisa. Desenvolvido especificamente para fins didáticos, o Linux Educacional é a distribuição Linux sobre iniciativa do governo federal, instalada em espaços informatizados em escolas de todo as regiões do Brasil. Dada suas potencialidades e visando a qualificação docente, segundo [Lemos 2011] é preciso que o Linux Educacional seja valorizado como dispositivo transformador das práticas pedagógicas.

A inovação tecnológica pode proporcionar a ampliação do campo da docência e a melhoria da qualidade da educação desde que utilizados adequadamente e sobre conhecimentos específicos. A prática e a formação do professor deve ser mais abrangente, dinâmica e permitir o desenvolvimento de habilidades necessárias para utilizar recursos tecnológicos, de acordo com [Xavier 2009]. Nesse sentido, também o ensino da Geografia aproveita-se do viés tecnológico da atualidade e se insere neste contexto as novas estratégias de aprendizado.

3. Procedimentos Metodológicos

Tratou-se de uma pesquisa aplicada, de abordagem quali-quantitativa, de caráter exploratório, cujos objetivos visaram analisar e descrever as características dos sujeitos investigados considerando suas interações com os recursos computacionais; isto é, os softwares educacionais da plataforma Linux Educacional (KGeography e Marble) no laboratório de informática aplicados a partir de projeto de intervenção para aprendizagem escolar. Visando compreender o comportamento de alunos da 8ª série de uma escola estadual em Goiânia - Goiás, antes e posterior a intervenção com os softwares propostos, esta pesquisa se enquadrou na modalidade estudo de caso. Segundo [Gil 2002], um estudo de caso visa conhecer em profundidade uma determinada situação, procurando descobrir o que há nela de mais essencial e característico.

Esta pesquisa consistiu em duas etapas; inicialmente foi realizada uma comparação entre o método tradicional de ensino da Geografia e o método de ensino com a mediação dos softwares propostos. O objetivo aqui foi aferir o impacto desses softwares sobre o nível de aprendizado dos alunos. Para tanto, nesta primeira etapa, com auxilio do professor regente, foi aplicado aos estudantes um conjunto de 10 (dez) questões de Cartografia baseado apenas no conhecimento adquirido pelos discentes no método tradicional de ensino. Posteriormente, após os alunos terem vivenciado algumas aulas de apresentação (disponibilização de material didático), manuseio e familiaridade dos softwares propostos, que passaram a ser utilizados como recursos didáticos, foram aplicados mais 10 (dez) questões de Cartografia, visando conhecer se os alunos elevaram ou não seu nível de aprendizado.

É válido enfatizar que, os softwares livres KGeography e Marble são plenamente adequados a ementa estudada; isto é, o ensino da Cartografia. Também, além de acesso livre e gratuito, possuem amplas possibilidades de utilização e elevada quantidade de

informação, capazes de proporcionar tanto ao docente quanto aos discentes inúmeras alternativas para fixação do conteúdo e obtenção de conhecimento. Ainda, podem proporcionar aos estudantes formas mais significativas, criativas, dinâmicas e ativas no processo ensino/aprendizagem. Além do mais, essas ferramentas podem tornar as aulas mais atrativas e chamativas, se utilizadas adequadamente.

Já na segunda etapa desta pesquisa buscou-se conhecer a percepção dos alunos em relação a essa experiência; isto é, a utilização do recurso computacional (softwares do Linux Educacional), como estratégias didáticas para o ensino da Geografia, algo diferente do que eles vinham vivenciando em sala de aula até então. Nesta etapa buscou-se conhecer fatores como: o cotidiano dos alunos em meio as TIC, suas opiniões sobre a relação informática e educação e; principalmente, suas impressões sobre a utilização dos softwares propostos como instrumento de aprendizado. Para tanto, foi aplicado aos alunos um questionário com perguntas sobre estes temas, visando ainda analisar o relacionamento/interação dos discentes com softwares propostos, considerando a ótica do aprendizado e também enquanto ferramenta de apoio didático. 70 (setenta) estudantes participaram das duas etapas desta pesquisa.

A abordagem quali-quantitativa permitiu organizar, analisar, interpretar e apresentar os resultados por meio de análises estatísticas e de comparação que se considerou como adequados. Pretendeu-se, então mensurar, a partir dos questionamentos aos alunos e por meio da observação e análise dos documentos (notas, plano de aulas dos professores), como os softwares educacionais sugeridos interagiram nos assuntos e conteúdos ministrados. Segundo [Cervo and Bervian 2002], a pesquisa é uma atividade voltada para a solução de problemas com emprego do método científico. Corroborando com essas ideias, [Silva 2010] relata que a pesquisa atende à necessidade de se conhecer a natureza dos problemas ou fenômenos, uma vez que trata de validar ou refutar as hipóteses lançadas sobre eles.

Para análise dos dados da etapa 1 (comparação entre o método tradicional e o método com mediação dos softwares propostos) foi utilizado o software (*Statistical Package for the Social Sciences*) – Pacote Estatístico para as Ciências Sociais – SPSS Statistics, que proporcionou a análise de dados por meio de medidas básicas, em que foi explorada principalmente, média, mediana, moda, variação, desvio padrão, teste de significância estatística; dentre outros. Os dados analisados por meio do SPSS foram bastante úteis e permitiram a mensuração de algumas categorias relacionadas à utilização das ferramentas computacionais propostas, contribuindo satisfatoriamente para a análise dos dados quantitativos.

Já para analise dos dados da etapa 2 (percepção dos alunos em relação a experiência de utilização dos softwares propostos) foi realizado principalmente estatística descritiva. Os dados coletados pelo questionário acessados via Google forms, por serem questões fechadas e como respostas definidas em intervalos, foram analisadas inicialmente pela própria ferramenta do Google, pela sua gratuidade, pelo seu potencial e por proporcionar o processo de avaliação ao criar ferramentas de coleta, organização de dados e análises comparativas. Concomitantemente, realizou-se uma análise mais detalhada, explorando tabulação cruzada e buscando entender as variáveis em situações diferentes.

4. O Ensino da Geografia com Recurso Computacional: Análise dos Resultados Obtidos a partir do Estudo de Caso

Esta pesquisa contou com a participação de 70 alunos da 8ª série, sendo 2 (duas) turmas de 35 alunos cada, de uma escola estadual na cidade de Goiânia - Goiás. A média de idade dos estudantes era de 13 anos, sendo 39 alunos do sexo feminino e 31 alunos do sexo masculino; 55,7% e 44,3% respectivamente. Inicialmente buscou-se aferir o nível de aprendizado dos alunos a partir da utilização dos softwares propostos da plataforma Linux Educacional (Marble e KGeography); para tanto, realizou-se comparação entre o método tradicional de ensino da Geografia e o método com mediação tecnológica. Posteriormente, buscou-se conhecer a percepção dos alunos em relação aos softwares utilizados, aferindo suas experiências vivenciadas; na tentativa de apreciar suas opiniões sobre a relação informática e educação.

Conforme demonstrado a seguir, na sessão 4.1 apresentam-se os resultados comparativos antes e após a intervenção com os softwares propostos. Considera-se este ponto do trabalho de grande relevância para a compreensão dos dados obtidos. Já na sessão 4.2 apresentam-se os resultados sobre o relacionamento cotidiano dos estudantes com as Tecnologias da Informação e Comunicação (TIC); e especialmente, a impressão que os alunos tiveram na prática com a utilização dos recursos computacionais propostos enquanto instrumento didático pedagógico; bem como suas expectativas, caso esta metodologia de ensino torna-se realidade na escola. É importante ressaltar que anterior a intervenção, os únicos recursos tecnológicos utilizados eram filmes e apresentação de slides.

4.1. Etapa 1: Comparação Método Tradicional e Softwares Propostos

Nesta fase da pesquisa, com auxilio do professor regente, foram aplicados 10 (dez) questões de Cartografia aos alunos antes da intervenção com os softwares propostos; isto é, somente com o conhecimento adquirido pelos alunos a partir do método tradicional de ensino; outras 10 (dez) questões, posteriormente a familiaridade dos alunos com o Marble e KGeography. Deve-se ressaltar que os alunos passaram por um período de treinamento com demonstração, testes, utilização e familiarização dos softwares propostos, com auxilio do material didático elaborado para esta finalidade. Após prática dos alunos com os softwares pesquisados, 10 (dez) novas questões de Cartografia foram propostas, visando aferir o nível de aprendizado. A avaliação 1 demonstra as notas obtidas por cada aluno durante o método tradicional de ensino da Geografia e a avaliação 2 as notas posterior a utilizações dos recursos computacionais, Figura 1.

Percebe-se que todos os estudantes obtiveram notas superiores na avaliação 2, quando utilizou-se os softwares propostos como recurso didático pedagógicos para o ensino-aprendizado. Houve uma significativa melhora no desempenho da aprendizagem da Cartografia dos alunos em relação ao conjunto de questões propostas. Se enquanto na avaliação 1 os resultados obtidos durante o método tradicional variaram entre 1(um) a 7 (sete) acertos, em um conjunto de 10 (dez) questões; a partir da utilização dos softwares propostos, na avaliação 2 o quantitativo de acertos variou entre 5(cinco) a 10 (dez), revelando uma variação média de 3,45 a mais de acertos. Alguns alunos melhoram seus desempenhos com até 5 (cinco) acertos a mais do que em relação a avaliação 1; 13 (treze) alunos obtiveram o valor máximo de acertos 10 (dez) na avaliação 2, diferentemente da avaliação 1, em que o número máximo de acerto foi 7 (sete), por apenas 2 alunos.

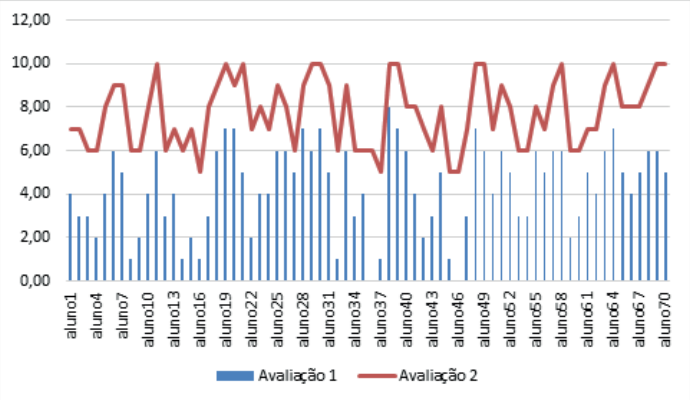


Figura 1. Comparações entre o Método tradicional (avaliação 1) e Método com recursos tecnológicos (avaliação 2)

Estes resultados comparativos, conforme [Chiofi and Furlan 2014] afirmam "o uso de tecnologias educacionais liga-se à qualidade do ensino; uma vez que novas tecnologias permitem aplicabilidades pedagógicas inovadoras que podem contribuir para resultados diferenciados". Comprovou-se a partir dos resultados da avaliação 2 que recursos computacionais potencializaram o aprendizado dos alunos, despertando o interesses dos discentes na obtenção do conhecimento. Obviamente o computador não é solução para problemas pedagógicos da sala de aula, conforme destaca [Rosalen and Mazzilli 2005]; porém, a formação de professores para a utilização de computadores na Educação "pode vir a contribuir para o aprimoramento da prática educativa se pautada pela compreensão das possibilidades e limites deste instrumento na concretização do papel educativo da escola" [Rosalen and Mazzilli 2005].

No decurso da análise dos dados comparativos entre a avaliação 1 e avaliação 2 foi possível ainda extrair medidas estatísticas essenciais para compreensão dos resultados, conforme Tabela 1.

Tabela 1. Dados estatísticos para comparação do uso de softwares na educação

Situação	Média	Mediana	Desvio Padrão	Valor-p
Antes da inserção dos softwares	4,29	4,50	1,96	$< 0{,}005$
Depois da inserção dos softwares	7,74	8,00	1,56	

Observa-se que houve elevação considerável na média dos acertos na avaliação 2 em comparação a avaliação 1; mesmo comportamento observado para a mediana. Também foi possível visualizar o desvio padrão nos dois testes, na avaliação 2 seu valor foi menor que na avaliação 1, reforçando a eficácia do método com utilização dos softwares propostos uma vez que a medida de dispersão foi inferior. Verificou-se que houve diferença significativa entre as duas avaliações, visto que o valor-p foi menor que 0,005; assegurando que o teste de significância comprova o elevado nível de confiança dos resultados, e ainda que a amostra (sujeitos) participantes da pesquisa fosse aumentada, os resultados seriam significantemente parecidos. O valor-p ≤ 0.005 comprovou a mudança comportamental dos alunos, aumentando o número de acertos a partir da inserção do recurso computacional utilizado (softwares educacionais).

A moda nas avaliações conforme Tabela 2 comprovou que na avaliação 2 o número de acertos dos estudantes incidiram com maior frequência nos valores máximos. Verifica-se que enquanto na avaliação 1 a maioria dos alunos tiveram nota $\leq$ 6,00 totalizando 88,57%; na avaliação 2 os estudantes tiveram nota $\geq$ 6,00 em 94,28% dos casos.

Tabela 2. Dados estatísticos para comparação do uso de softwares na educação

Nota	0,0	1,00	2,0	3,00	4,0	5,00	6,0	7,00	8,0	9,00	10,0
AVALIAÇÃO 1	2	6	6	10	11	11	16	7	1	-	-
AVALIAÇÃO 2	-	-	-	-	-	4	16	12	13	12	13

Assim, na etapa 1 denominada comparação do método tradicional x softwares propostos comprovou-se estatisticamente que a utilização de recursos computacionais não apenas deixam as aulas mais dinâmicas e interativas, mas efetivamente contribuem para o melhor aprendizados dos estudantes se utilizadas adequadamente. As tendências reconhecidas pela literatura na relação informática e educação, comprovaram nesta pesquisa que as novas tecnologias são poderosos instrumentos didáticos.

4.2. Etapa 2: Percepção dos Alunos em Relação aos Softwares Propostos (Experiências Vivenciadas)

Esta etapa da pesquisa buscou analisar o relacionamento e interação dos discentes com softwares propostos, considerando a ótica do aprendizado e também enquanto ferramenta de apoio didático. Tendo em vista que na educação contemporânea as novas tecnologias buscam enriquecer o processo ensino-aprendizagem, foi importante conhecer também como os alunos perceberam as aulas com o uso dos softwares propostos, a partir desta experiência. Além disso, tornou-se objeto de estudo compreender o relacionamento cotidiano dos estudantes com as Tecnologias da Informação e Comunicação (TIC), bem como suas opiniões sobre a relação informática e educação. Para tanto, foi aplicado questionário; e cujos resultados são apresentados em seguida.

Inicialmente, considerou-se importante mensurar a quanto tempo os estudantes conhecem as TIC e seu período atual de utilização, uma vez que os resultados foram bastante satisfatórios quando comparando o método tradicional x softwares propostos. Verificou-se que 62,90% dos estudantes já faziam uso de recursos tecnológicos a mais de 4 (quatro) anos e apenas 2,80% a menos de 1 (um) ano. Ainda considerou-se como habitual sua utilização, uma vez que 31,40% dos alunos declararam fazer uso dessas tecnologias entre 2 a 5 horas semanais, 24,30% entre 5 a 20 horas e 27,10% mais de 10 horas por semana. Deduziu-se assim, que a boa interação dos alunos com os recursos computacionais utilizados ocorreu em virtude de uma habitual familiaridade com tecnologias informacionais, uma vez que os resultados obtidos comprovam tal premissa.

Buscou-se conhecer para qual finalidade os estudantes utilizavam as TIC, visando mensurar o nível de interesse dos mesmos em relação aos estudos anteriormente a interação com os softwares propostos; seja como meio de pesquisa e/ou visualização de conteúdos ou mesmo como instrumentos (site e softwares) educacionais. Verificou-se que, a opção "estudos" foi a menos assinalada pelos estudantes, enquanto prática cotidiana, correspondendo a apenas 28% das respostas. Jogos, redes sociais e entretenimento; obtiveram 72% . Pode-se inferir que esse resultado seja porque os recursos tecnológicos

não eram utilizados na escola anterior a intervenção, e, também aliado a outros fatores sociais como ao desconhecimento do potencial destas tecnologias para a educação, priorizando sua utilização para outros fins. Acredita-se que a partir desta intervenção os estudantes atentaram para as possibilidades de tais tecnologias quando aplicadas adequadamente com a finalidade educacional.

[Silva 2013] afirma que os recursos tecnológicos e midiáticos fazem parte do contexto dos alunos, sendo papel do professor e da escola conhecer melhor como o estudante utiliza a tecnologia em seu cotidiano e a partir daí buscar usá-la pedagogicamente na sala de aula. De acordo com este autor, são as concepções que o professor tem sobre o que é ensinar que direcionam sua prática, como cabe a ele articular as mudanças sociais, bem como tecnológicas no espaço de sala de aula [Silva 2013]. Ademais, Investimentos em recursos tecnologias e capacitação docente são fatores determinantes para a obtenção de bons resultados. A Tabela 3 revela as opiniões dos estudantes sobre a utilização de recursos tecnológicos como instrumento didático-pedagógico e demonstra que eles reconhecem a contribuição de tais instrumentos para favorecer o ensino-aprendizado.

Tabela 3. Opiniões dos alunos ao uso de tecnologias midiáticas na educação

	Sim	Não	Indiferente
Favorável a utilização de tecnologias midiáticas	87,1%	11,4%	1,5%
Os softwares propostos contribuíram para o melhor aprendizado	95,7%	4,3%	-
As aulas se tornaram mais interessantes e o conteúdo foi ensinado com mais facilidade	94,1%	5,9%	-

Pode-se deduzir pelos resultados anteriores que os softwares propostos melhoraram o aprendizado e tornaram as aulas mais interessantes, visto os alunos estarem em concordância com essa estratégia pedagógica.

Verificou-se ainda, os motivos que impedem a utilização de recursos computacionais e/ou tecnológicos na unidade escolar. Nesta questão os estudantes puderam assinalar mais de uma alternativa. A maior incidência das respostas, 78,60%; 32,90% e 31,40% ocorreu em aspectos relacionados a problemas com infraestrutura de redes, ausência de softwares e computadores respectivamente, fato comprovado pela precariedade do laboratório de informática e pelo não investimento em recursos computacionais e de infraestrutura lógica. Apesar do Projeto Político Pedagógico – PPP dessa unidade escolar sinalizar que prioriza ações de motivação e inovação que ofereçam possibilidades de desenvolvimento amplo para o ensino médio, fundamental e ensino integral; não foi este o cenário observado e refletido nas respostas dos alunos, no que tange a utilização de tecnologias. Em menor proporção, mas também relevante, a percepção do desinteresse da direção da escola em 30% das respostas, ausência de um intermediário (estagiário e/ou professor de informática) 25% e falta de conhecimento tecnológico do professor regente 24,30%.

Os alunos consideram que os recursos tecnológicos são instrumento didático-pedagógico capazes de despertar maior interesse e, consequentemente, tornar as aulas mais produtivas. Ao mesmo tempo, revelaram com que frequência gostaria de utilizar tais recursos. A Tabela 4 ilustra os resultados.

Tabela 4. Interesse e frequência dos alunos por instrumento didático-pedagógico

As ferramentas tecnológicas despertaram maior interesse pelo aprendizado	Sim Não Indiferente	88,60% 8,60% 2,80%
Frequência de utilização	Diariamente	61,40%
	Semanalmente	22,90%
	Mensalmente	10%
	Eventualmente	5,70%

Conforme se pode observar, 61,40% estudantes gostariam de utilizar recursos tecnológicos como instrumento didático-pedagógico diariamente e 22,90% semanalmente, totalizando 84,30% de respostas afirmativas em período de tempo relativamente curto. A Tabela 4 ainda revela que 88,60% dos alunos afirmam que o computador e a internet poderiam ser utilizados para fins educacionais porque despertam maior interesse dos alunos e do professor em sala de aula e consequentemente tornam as aulas mais produtivas. Tais resultados corroboram com a literatura pesquisada reconhecendo as potencialidades das TIC no processo de disseminação do conhecimento. [Perrenound 2015] reforça esse parecer quando assegura que "as novas tecnologias podem reforçar a contribuição dos trabalhos pedagógicos e didáticos contemporâneos, pois permitem que sejam criadas situações de aprendizagens ricas, complexas, diversificadas".

Por fim, ainda foi possível extrair resultados interessantes por meio de tabulação cruzada de variáveis. Verificou-se que todos os estudantes cuja resposta foi "Estudos" para a pergunta: "em que ocupa a maior parte do tempo quando está utilizando o computador e internet" e responderam "sim" para as questões: "se favorável a utilização de tecnologias midiáticas"; "se as ferramentas utilizadas para o auxilio do ensino em Geografia contribuíram para um melhor aprendizado"; e "se as aulas se tornaram mais interessantes e o conteúdo foi ensinado com mais facilidade", os resultados mostram o amplo interesses dos mesmos pela metodologia de ensino com uso de recursos tecnológicos.

Contrariamente, os alunos cujas respostas foram "não" ou "não sabe dizer" para a pergunta: "se favorável a utilização de tecnologias midiáticas" e "não" para o questionamento: "se as ferramentas utilizadas para o auxilio do ensino em Geografia contribuíram para um melhor aprendizado" e "não" para a indagação: "se as aulas se tornaram mais interessantes e o conteúdo foi ensinado com mais facilidade", afirmaram que ocupam a maior parte de seus tempos em jogos, redes sociais e entretenimento quando estão utilizando computador e a internet.

Assim, nesta segunda fase da pesquisa verificou-se amplo interesse dos alunos pela utilização de recursos tecnológicos e computacionais como instrumentos de mediação para o aprendizado. Considerou-se que os estudantes não apenas experimentaram algo novo e inovador, diferente do que estavam acostumados, mas também experimentaram uma perspectiva significativa para obtenção de conhecimento, que contribui para a melhoria da qualificação do ensino.

5. Considerações Finais

Os resultados desta pesquisa demonstraram que a partir da intervenção com os softwares propostos, KGeography e Marble - Linux Educacional, houveram contribuições significativas para a aprendizagem dos estudantes pesquisado no estudo da Cartografia. Comparando-se o método tradicional de ensino com o método com inserção dos recursos tecnológicos, comprovou-se estatisticamente a importância dessas ferramentas no processo ensino-aprendizagem. Obviamente que, esses recursos somente possuem sua efetividade quando utilizados adequadamente, o que requer capacitação do docente e investimento em infraestrutura, especialmente em equipamentos computacionais. Acredita-se que o sucesso apresentado neste estudo de caso foi resultado do uso de recurso tecnológico didático-pedagógico durante o processo de intermediação.

A partir da observação, coleta e análise dos dados vislumbrou-se ainda as percepções e expectativas dos estudantes em relação à experiência vivenciada na utilização destes recursos computacionais, que apesar de comum ao cotidiano dos estudantes até então, era algo incomum e inédito como instrumento didático-pedagógico para a unidade escolar pesquisada.

Desse modo, evidenciou-se satisfação superior a 90% dos estudantes, desejo na continuidade de sua utilização; e anseio em possível expansão desta metodologia para as demais disciplinas. No entanto, na contramão desses interesses, a unidade escolar demonstra problemas como: sucateamento do laboratório de informática (eminente desativação), infraestrutura de rede limitada e inadequada; e despreparo docente no manuseio das Tecnologias da Informação e Comunicação (TIC) sob a perspectiva educativa.

A construção de uma educação em que a inserção de meios tecnológicos se torne realidade nas instituições públicas de ensino requer esforço conjunto de todos envolvidos nesse processo; sejam alunos, professores, direção escolar e até mesmo agentes políticos. Esta pesquisa demonstrou que, quando aplicados corretamente, os recursos computacionais tornam-se bastante eficazes para a construção e disseminação do conhecimento. Igualmente, ainda é necessário um despertamento da sociedade e especialmente da comunidade acadêmica, visando incentivar a disseminação dessas novas tecnologias no ambiente educacional, dada suas potencialidades.

Referências

Almeida, F. d. (2007). Computador, escola e vida: aprendizagem e tecnologias dirigidas ao conhecimento. *São Paulo: Cubzac*.

Bezerra, M. d. S. (2014). O uso das tecnologias digitais como recurso no ensino de geografia.

Cervo, A. and Bervian, P. A. (2002). *Metodologia científica*. México.

Chiofi, L. C. and Furlan, M. R. (2014). O uso das tecnologias educacionais como ferramenta didática no processo de ensino e aprendizagem. *III Jornada de Didática - Jornada de Didática: Desafios para a Docência e II Seminário de Pesquisa do CEMAD*.

Gil, A. C. (2002). Como elaborar projetos de pesquisa. *São Paulo*, 5(61):16–17.

Lemos, C. D. (2011). Linux educacional: desafio para o professor. *Revista Novas Tecnologias na Educação*, pages 1–11.

Lima Filho, J. F. d. (2014). A geografia escolar e o uso de softwares educacionais como recursos didáticos. *VII Congresso Brasileiro de Geográfos*, 1(1):45–60.

Maciel, C. (2009). A internet como ferramenta educacional.

Mendes, J. (2011). *TICs –Tecnologias da Informação e da Comunicação Educativa*. Universidade Federal do Paraná.

Perrenound, P. (2015). *Dez novas competências para ensinar*. Artmed editora.

Rodrigues, A. and Sousa, N. (2012). A internet e o ensino de geografia. *PROJEÇÃO E DOCÊNCIA*, 3(1):37–55.

Rosalen, M. and Mazzilli, S. (2005). Formação de professores para o uso da informática nas escolas: evidências da prática. *REUNIÃO ANUAL DA ASSOCIAÇÃO NACIONAL DE PÓS-GRADUAÇÃO E PESQUISA EM EDUCAÇÃO*, 28:1–17.

Silva, A. M. d. (2010). *Metodologia da Pesquisa*. 2ª Edição. Universidade Estadual do Ceará, RDS Editora.

Silva, L. A. (2013). O uso pedagógico de mídias na escola: práticas inovadoras. *Revista Eletrônica de Educação de Alagoas*, 1(01).

Tajra, S. F. (2011). *Informática na Educação: novas ferramentas pedagógicas para o professor na atualidade*. Érica.

Valente, J. A. (2009). Análise dos diferentes tipos de softwares usados na educação. *O computador na sociedade do conhecimento*, page 71.

Werthein, J. (2000). A sociedade da informação e seus desafios. *Ciência da informação, Brasília*, 29(2):71–77.

Xavier, K. (2009). *Proinfo Integrado: A Formação em Linux Educacional no DF*. IX Congresso Nacional de Educação - EDUCERE, PUCPR.

Ambientes Virtuais Tridimensionais no Processo de Ensino-Aprendizagem de Libras: Uma Revisão Sistemática da Literatura

Tiago do Carmo Nogueira[1], **Emily Vanessa Vasconcelos Mendonça**[1], **Júlia Pereira dos Santos e Araújo**[1], **Julia Batista Alves Moura**[1], **Deller James Ferreira**[2]

[1]Departamento de Informática – Instituto Federal de Mato Grosso (IFMT)
Caixa Postal 78.600-000 – Barra do Garças – MT – Brazil

[2]Instituto de Informática – Universidade Federal de Goiás (UFG)

tiago.nogueira@bag.ifmt.edu.br, deller@inf.ufg.br

Abstract. *Three-dimensional virtual environments have become common tools to stimulate students in learning, and they introduce innovative technologies in the teaching-learning process. However, it is essential to identify which are the main three-dimensional tools used in the teaching-learning process, specifically in the teaching of Brazilian Sign Language (Libras). Therefore, a Literature Systematic Revision (RSL) was performed in order to identify the possible Three-Dimensional Virtual Environments of Learning, specifically the ones applied in the Sign Language teaching-learning process. Thus, 172 scientific articles were identified that correlated to Sign Language teaching with Three-Dimensional Virtual Environments. It was noticed that the Learning Three-Dimensional Environments, identified through this RSL, used the main techniques of Virtual Reality for conversion and interpretation from the Portuguese Language to Sign Language, supported by the three-dimensional humanoids construction. Lastly, with these mechanisms, the use of technology as an educational support may be stimulated, helping the learning of deaf people in a significant way. Thus, media resources through Teaching Three-Dimensional Environments favor the learning, minimizing the differences between deaf students and those who can hear.*

Resumo. *Os Ambientes Virtuais Tridimensionais se tornaram ferramentas comuns para incentivar a aprendizagem dos alunos e introduzir tecnologias inovadoras no processo de ensino-aprendizagem. Entretanto, é imperativo identificar quais são as principais ferramentas tridimensionais empregadas em processo de ensino-aprendizagem, especificamente no ensino da Língua Brasileira de Sinais (Libras). Para tanto, foi realizada uma Revisão Sistemática da Literatura (RSL), com intuito de identificar os possíveis Ambientes Virtuais Tridimensionais de Aprendizagem, especificamente o aplicado nos processos de ensino-aprendizagem de Libras. Assim, 172 artigos científicos que correlacionavam o ensino de Libras com Ambientes Virtuais Tridimensionais foram identificados. Percebeu-se que os Ambientes Virtuais Tridimensionais de Aprendizagem identificados por meio desta RSL utilizavam-se das principais técnicas de Realidade Virtual para conversão ou interpretação da Língua Portuguesa para Libras,*

apoiados nas construções de humanoides tridimensionais. Por fim, com estes mecanismos, o uso das tecnologias como apoio educacional deve ser incentivado, ajudando de forma significativa no aprendizado dos deficientes auditivos. Dessa forma, recursos midiáticos, por meio de Ambientes Tridimensionais de ensino, favorecem o aprendizado, minimizando as diferenças entre alunos com deficiência auditivas e ouvintes.

1. Introdução

Os gestos são movimentos elementares das mãos, tornando-as componentes de comunicação atômica, representando os pensamentos das pessoas surdas [Aggarwal and Ryoo 2011]. Assim, a Linguagem Brasileira de Sinais (Libras), constitui-se formalmente de regras fonológicas, morfológicas, semânticas, sintáticas e pragmáticas. A Libras é um instrumento substancial no processo de inclusão social dos deficientes auditivos, constituída por aspectos linguísticos e culturais, de forma singular [Gonçalves 2013].

Neste sentido, torna-se imperativa a aplicabilidade das principais técnicas computacionais nas implementações de ambientes de aprendizagem que corroboram para a construção do conhecimento dos deficientes auditivos. Para [Kawakami et al. 2015], há uma evolução nos Ambientes Virtuais de Aprendizagem para a utilização de tecnologias tridimensionais, os quais possuem as características de identificar as necessidades dos alunos com deficiência auditiva, isto é, Ambientes Virtuais tridimensionais de ensino que se adaptam às experiências vivenciadas pelos deficientes auditivos.

Segundo [Gonçalves 2013], por meio da implementação de sistemas de sínteses com modelos formais computacionais paramétricos, utilizando avatares tridimensionais para auxiliar no ensino de Libras, pode-se gerar movimentos automáticos através de chamadas externas, contribuindo dessa forma no processo de ensino-aprendizagem dos deficientes auditivos.

No entanto, faz-se necessária uma investigação mais profunda de quais ambientes tridimensionais são aplicados nos processos de ensino-aprendizagem de Libras, tornando-se essencial a sua identificação e classificação. Dessa forma, este artigo propõe uma Revisão Sistemática da Literatura a fim de identificar os principais Ambientes Virtuais Tridimensionais de Ensino dedicados aos processos de ensino-aprendizagem de Libras.

Assim, foram identificados 172 artigos científicos que endereçavam assuntos relacionados aos Ambientes Virtuais Tridimensionais de Aprendizagem e Libras. Utilizando-se dos critérios de inclusão e exclusão, foram aceitos e analisados 12 artigos, identificando-os e classificando-os de forma qualitativa e quantitativa.

Este artigo apresentado está organizado da seguinte forma: na seção 2 são realizados os trabalhos relacionados sobre Ambientes Virtuais Tridimensionais de Aprendizagem; na seção 3 é realizada a Revisão Sistemática da Literatura; na seção 4 são realizados os resultados e discussões; e na seção 5 as conclusões.

2. Ambientes Virtuais Tridimensionais de Aprendizagem

As tecnologias vêm se destacando como um recurso eficaz, dando oportunidade a alunos com necessidades especiais em realizar exercícios básicos e práticos, adaptando-os às necessidades especiais individuais [Bertacchini et al. 2013].

O uso dos computadores, nos últimos anos, contribuiu de forma significativa para a construção de novas formas de aprendizagem [Scott et al. 2016]. No entanto, antes da informatização, a educação era uma ação informatizada e transmitida de geração a geração, utilizando-se do mecanismo da oralidade [Alves et al. 2014].

Neste sentido, a informatização gerou grandes mudanças no comportamento e na maneira com que as pessoas se comunicam [Alves et al. 2014], corroborando de forma direta para as transformações conceituais da educação. Por meio dessas modificações, Ambientes Virtuais de Aprendizagens têm sido amplamente adotados por educadores, obtendo resultados promissores [Scott et al. 2016].

Recentemente, os Ambientes Virtuais de Aprendizagens evoluíram para meios mais avançados, utilizando-se de tecnologias tridimensionais (3D), levando em conta às necessidades e preferências individuais do aluno. Esta nova mudança exige que os Ambientes Virtuais de Aprendizagens se adaptem ao aluno, aplicando abordagens de aprendizagem personalizadas.

A Realidade Virtual (RV) é outro mecanismo bastante empregado em processos de elaboração de elementos tridimensionais. A RV como ferramenta mediática aplicada na educação não busca apenas deslocar os métodos de educação atuais, mas aumentá-los. Sistemas de RV são amplamente utilizados em treinamento virtual tridimensional, nos mais diversos campos do conhecimento. Isso ocorre porque os Ambientes Virtuais Tridimensionais baseados em RV podem simular situações reais em diferentes contextos, inclusive no campo educacional, como instrumento de apoio pedagógico, tanto no ensino presencial quanto a distância [Kawakami et al. 2015].

Com a fomentação de Ambientes Virtuais Tridimensionais, novas perspectivas relacionadas à aprendizagem poderão ser aplicadas com mais eficiência [Campbell et al. 2016]. Na educação, os mundos virtuais tridimensionais permitem o emprego flexível da educação a distância, proporcionando uma experiência imersiva e sensorial, além de oferecer um conjunto de mecanismos que permitem fazer coisas não possíveis no mundo real [Alves et al. 2014]. Segundo [Muñoz-Cristóbal et al. 2017], Ambientes Virtuais Tridimensionais são ótimos locais para implementar ideias por meio de ferramentas de comunicação, permitindo melhorias contínuas nos processos de ensino-aprendizagem em diferentes espaços físicos. Para [Scott et al. 2016], Ambientes Virtuais Tridimensionais exploram recursos adaptativos para criar novas e aprimoradas experiências de aprendizado em diferentes contextos.

No entanto, pouco se sabe sobre os fatores envolvidos em Ambientes Tridimensionais Adaptativos para obter-se benefícios de aprendizado e quais fatores de avaliação estão presentes nos estudos atuais, isto ocorre por falta de pesquisas em Ambientes Virtuais Tridimensionais.

Portanto, Ambientes Virtuais Tridimensionais tornaram-se ferramentas comuns para incentivar a aprendizagem dos alunos e introduzir tecnologias inovadoras no processo de ensino-aprendizagem. Entretanto, é imperativo identificar quais são as principais ferramentas tridimensionais empregadas em processo de ensino-aprendizagem dos deficientes auditivos. Assim, na seção 3 será apresentada uma revisão sistemática desses instrumentos aplicados ao ensino e aprendizado de Libras.

3. Revisão Sistemática da Literatura

Para [Wholin et al. 2012], Revisão Sistemática da Literatura (RSL) é um meio de identificar, analisar e interpretar todas as evidências disponíveis relacionadas a uma pesquisa específica.

Dessa forma, com intuito de identificar os possíveis ambientes virtuais tridimensionais de aprendizagem, especificamente, aplicadas nos processos de ensino-aprendizagem de Libras, torna-se essencial uma pesquisa mais profunda sobre a utilização desses ambientes. Esta RSL incidiu sobre os artigos indexados entre jan./2010 a dez./2016, utilizando-se a abordagem proposta por [Keele et al. 2007].

Assim, na subseção 3.1 será apresentado o planejamento desta RSL e na subseção 3.2 serão apresentadas a execução e a extração dos dados.

3.1. Planejamento

Segundo [Wazlawick 2014], as indagações sobre determinados problemas elencado pela literatura são essenciais para a construção de uma pesquisa. Neste sentido, a partir da importância dos ambientes tridimensionais de ensino e dos processos de ensino-aprendizagem das pessoas surdas, elenca-se a seguinte indagação:

- Q01 – Quais os principais Ambientes Virtuais Tridimensionais de Aprendizagem aplicados nos processos de ensino-aprendizagem de Libras?

Para realização desta RSL, foram selecionadas as seguintes bases: IEEE Digital Library; ACM Digital Library; Springer Digital Library; Science Direct; Google Scholar. Assim, na subseção 3.1.1 será apresentada a construção do protocolo de pesquisa e na subseção 3.1.2 a construção dos critérios de inclusão e exclusão.

3.1.1. Construção do Protocolo de Pesquisa

O processo de construção do protocolo de pesquisa desta RSL deu-se pela extração das principais palavras-chaves da questão de pesquisa (Q01). Dessa forma, foram extraídas as seguintes palavras-chaves: “ambientes virtuais”, “ensino-aprendizagem”, “Libras” e “tridimensionais”.

Para identificar o maior número de trabalhos relevantes para pesquisa, foram construídos dois protocolos, um em português/Brasil e outro em Inglês:

- P01 – (“ambientes” AND “aprendizagem” AND “surdos” AND (“tridimensional” OR “tridimensional”) AND “libras”); P02 – (“environments” AND “Learning” AND “deaf” AND (“Three-dimensional” OR “Three dimensional”) AND “libras”).

Com intuito de realizar testes primários dos protocolos P01 e P02, foram submetidos na base de pesquisa Google Scholar. Dessa forma, no primeiro protocolo (P01), foi necessário realizar uma pequena mudança na *String* de busca. Assim, trocou-se a palavra-chave “tri-dimensional” por “tridimensional”, retornando 150 artigos. No segundo protocolo de pesquisa (P02), foram identificados 22 artigos que endereçavam Ambientes Virtuais Tridimensionais de Aprendizagem. Porém, um dos artigos não continha a data de publicação. Em função disso, optamos por excluí-lo desta RSL.

3.1.2. Construção dos Critérios de Inclusão e Exclusão

Para a mineração dos artigos científicos, deu-se a construção dos critérios de inclusão e exclusão. Estes critérios são primordiais para o processo de seleção dos artigos, isto é, a aceitação ou rejeição dos mesmos, por esta RSL.

Os critérios de inclusão dos artigos identificados, foram definidos com intuito de responder à questão de pesquisa (Q01). Entretanto, os critérios de exclusão possuem a finalidade de eliminar artigos científicos que não contribuem ou não respondem à questão de pesquisa (Q01). A Tabela 1 apresenta os critérios de inclusão e exclusão adotados por esta RSL.

Tabela 1. Critérios de inclusão e exclusão adotados por esta RSL.

Critérios de Inclusão	Critérios de Exclusão
a) Artigos que abordam Ambientes Virtuais Tridimensionais de Aprendizagem aplicados no processo de ensino-aprendizagem de Libras.	a) Artigos que abordam Ambientes Virtuais de Aprendizagem, mas não estão relacionados com Ambientes Virtuais Tridimensionais de Aprendizagem; b) Artigos que abordam Ambientes Virtuais Tridimensionais de Aprendizagem não empregados nos processos de ensino-aprendizagem de Libras; c) Artigos que não estão relacionados ao ensino de Libras; d) Artigos que não estão relacionados a Ambientes Virtuais Tridimensionais de Aprendizagem; e) Artigos que não possuem resumos ou data de publicação.

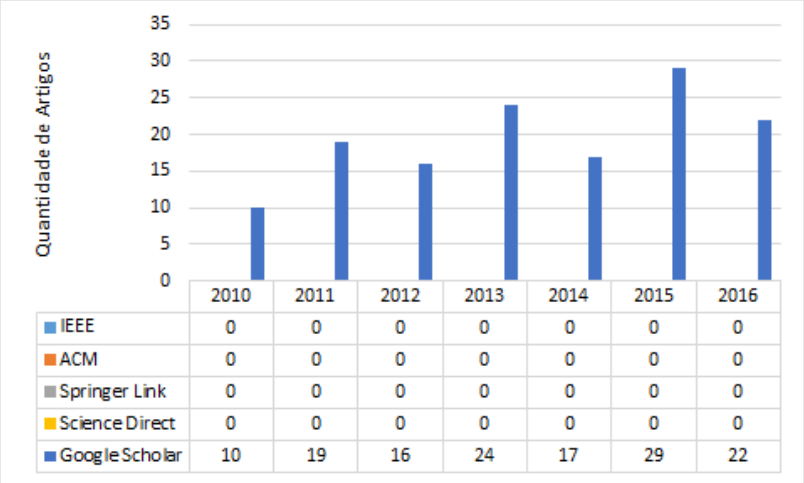
3.2. Execução e Extração

Nesta subseção é apresentado o processo de identificação dos artigos científicos que endereçam Ambientes Virtuais Tridimensionais de Aprendizagem e Libras (subseção 3.2.1) e o processo de execução e extração dos artigos, aplicando os critérios de inclusão e exclusão (subseção 3.2.2).

3.2.1. Identificação de Artigos

Nesta fase, a identificação dos artigos foi realizada em duas execuções, E01 e E02. A primeira execução (E01) foi efetuada à pesquisa por meio do protocolo P01, com intuito de identificar artigos científicos publicados em Português. A Figura 1 apresenta o quantitativo de artigos identificados por meio da execução do protocolo P01.

Por meio da execução do protocolo P01, verificou-se que não houve publicações nas bases IEEE Digital Library, ACM Digital Library, Springer Digital Library e Science

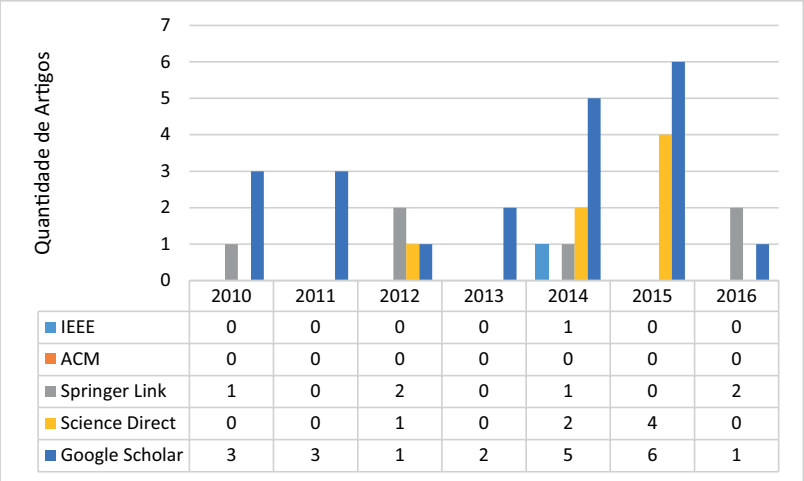


	2010	2011	2012	2013	2014	2015	2016
IEEE	0	0	0	0	0	0	0
ACM	0	0	0	0	0	0	0
Springer Link	0	0	0	0	0	0	0
Science Direct	0	0	0	0	0	0	0
Google Scholar	10	19	16	24	17	29	22

Figura 1. Quantitativo de artigos identificados por meio da execução do protocolo P01

Digital Library. No entanto, foram identificados 137 artigos científicos publicados e indexados na base Google Scholar. Observa-se que o intuito da execução do protocolo P01 é a inserção nesta RSL de artigos científicos publicados em português, visto que o tema abordado trata-se da Língua Brasileira de Sinais (Libras).

A segunda execução (E02) foi efetuada à pesquisa por meio do protocolo P02, com intuito de identificar artigos científicos publicados em bases nacionais e internacionais. A Figura 2 apresenta o quantitativo de artigos identificados por meio da execução do protocolo P02.



	2010	2011	2012	2013	2014	2015	2016
IEEE	0	0	0	0	1	0	0
ACM	0	0	0	0	0	0	0
Springer Link	1	0	2	0	1	0	2
Science Direct	0	0	1	0	2	4	0
Google Scholar	3	3	1	2	5	6	1

Figura 2. Quantitativo de artigos identificados por meio da execução do protocolo P02

A Figura 2 apresenta o quantitativo de artigos identificados por esta RSL, por meio da execução do protocolo P02. Observa-se que foram identificados 21 artigos na base Google Scholar, 7 artigos na base Science Direct, 6 artigos na base Springer Link e 1 artigo na base IEEE Digital Library.

Ressalta-se a importância da alteração do mecanismo de busca em algumas bases,

com intuito de identificar o maior número de artigos que sejam relevantes para a extração dos dados, por esta RSL. Dessa forma, foi utilizado o mecanismo de busca avançada nas bases ACM Digital Library e Science Direct.

3.2.2. Processo de Seleção dos Artigos

O processo de seleção dos artigos está divido em três etapas: exclusão dos artigos duplicados; rejeição dos artigos por critérios de exclusão; e aceitação de artigos por critérios de inclusão. Este processo contou com dois revisores e um pesquisador supervisor.

A leitura era realizada em pares de revisores e supervisionada pelo pesquisador/especialista. Dessa forma, os critérios de inclusão e exclusão eram cuidadosamente discutidos entre os revisores na tentativa de minimizar os riscos de exclusões de estudos relevantes e as inclusões de estudos irrelevantes.

Para auxiliar nas leituras e revisões dos artigos, os quais eram classificados como duplicados, rejeitados e aceitos, foi utilizada a ferramenta *Start Tool*[1]. Neste sentido, de forma automática, foram identificados 8 artigos científicos duplicados, isto é, indexados em diferentes bases. Observa-se que, por meio da Figura 1, 152 artigos científicos foram rejeitados, aplicando os critérios de exclusão e 12 artigos foram aceitos, aplicando os critérios de inclusão.

Tabela 2. Quantitativo de artigos rejeitados pelos critérios de exclusão

Bases	**Período/Ano**							**Total**
	2010	2011	2012	2013	2014	2015	2016	
IEEE	0	0	0	0	0	0	0	0
ACM	0	0	0	0	0	0	0	0
Springer Link	1	0	0	0	0	0	0	1
Science Direct	0	0	1	0	2	4	0	7
Google Scholar	12	17	18	23	20	32	21	144

A Tabela 2 apresenta a quantidade de artigos que foram rejeitados por esta RSL, por meio da aplicação dos critérios de exclusão. Observa-se que, foram rejeitados 144 artigos advindos da base Google Scholar, 7 artigos advindos da base Science Direct e 1 artigo da base Springer Link.

Aplicando-se os critérios de inclusão no processo de seleção dos artigos, foram aceitos 12 artigos, os quais endereçavam ambientes virtuais de aprendizagem tridimensionais.

A Tabela 3 apresenta a quantidade de artigos que foram aceitos por esta RSL, por meio da aplicação dos critérios de inclusão. Observa-se que foram aceitos apenas 11 artigos científicos, 1 artigo científico foi excluído desta RSL por não ter a data de sua publicação. Assim, nota-se que, foram aceitos 9 artigos advindos da base Google Scholar, 1 artigo da base Springer Link e 1 artigo da base IEEE Digital Library.

[1]Dispnível em: http://lapes.dc.ufscar.br/tools/

Tabela 3. Quantitativo de artigos aceitos pelos critérios de inclusão

Bases	Período/Ano							Total
	2010	2011	2012	2013	2014	2015	2016	
IEEE	0	0	0	0	1	0	0	1
ACM	0	0	0	0	0	0	0	0
Springer Link	0	0	0	0	0	0	1	1
Science Direct	0	0	0	0	0	0	0	0
Google Scholar	1	1	1	3	0	3	0	9

4. Resultados e Discussões

Os artigos aceitos no processo de seleção (subseção 3.2.2) foram analisados com intuito de identificar e discutir quais ferramentas tridimensionais são aplicadas nos processos de ensino-aprendizagem de Libras.

Neste sentido, [Brega et al. 2014] apresentam o desenvolvimento de um software capaz de receber como entrada, sentenças e palavras, e transformá-las em uma simulação tridimensional, com um *avatar* modelado, utilizando-se da Realidade Virtual. Assim, de acordo com [La Marca et al. 2010], por meio da aplicação da ferramenta *The Virtual Humans* (VH), o qual é classificada como um Ambiente Virtual Tridimensional, pode-se processar humanoides que executam movimentos em Libras, a partir de palavras fornecidas pelos usuários.

Ainda, utilizando-se das técnicas da Realidade Virtual, para [Brega et al. 2013], pode-se, por meio da aplicação de Dicionários Temáticos, proporcionar a interpretação e a edição de movimentos através de dispositivos móveis. Dessa forma, o dicionário temático é capaz de interpretar, editar e gravar movimentos em Libras. Desse modo, utilizando-se da tecnologia de captura de movimento, a ferramenta *Signing Avatars* traduz de forma automática as frases ou palavras da Língua Portuguesa para Libras, propiciando os resultados através de um corpo humano virtual [Martins and Fadel 2015, de Farias Klimsa and Klimsa 2013, De Martino et al. 2016].

A implementação dos sistemas visuais-gestuais pode proporcionar aos usuários uma experiência imersiva por meio de interações com dicionário virtual temático em Libras com multiprojeções [Brega et al. 2013]. Dessa forma, os assuntos escolhidos pelos usuários são representados por modelos 3D, permitindo que estes possam explorá-los, de acordo com a navegação realizada no ambiente tridimensional. Assim, os sinais em Libras são interpretados por um ator (*avatar*), possibilitando aos usuários a interpretação e a edição de gestos. Deste modo, as seleções das palavras interpretadas podem ser feitas através de uma interface gráfica de controle, ou pela seleção dos objetos virtuais modelado em 3D [Brega et al. 2013].

Outra atividade importante em Ambientes Tridimensionais são as criações do design de personagens e animações em 3D, direcionadas ao ensino de Libras para crianças ouvintes. [Martins and Fadel 2015] apresentam o desenvolvimento de um personagem para a interpretação em libras, formatado em 3D. Neste sentido, utilizando-se de modelos computacionais, por exemplos, *avatares* 3D, sistemas de síntese, entrada externa de dados, pode-se por meio de vetores paramétricos e configuráveis, promover ambientes tridimensionais para processos de ensino-aprendizagem em Libras [Gonçalves 2013].

[Stiehl 2015] apresenta uma ferramenta tridimensional conhecida como *SignWriting*. Esta ferramenta possui um mecanismo de escrita para representação de línguas de sinais para deficientes auditivos, baseando-se em técnicas de reconhecimento de padrões, com intuito de identificar símbolos de configuração de mãos.

Para [BAQUETA 2012], ao construirmos um design de uma interface que ofereça suporte à agregação de recursos, com imagens e vídeos em Libras, pode-se auxiliar na interação dos deficientes auditivos por meio da gamificação. Esta abordagem pode ser aplicada no processo de ensino-aprendizagem de crianças surdas.

Segundo [Paschuini 2015], o *Hand Talk* é uma das ferramentas utilizadas para traduções da língua portuguesa para Libras. Por meio deste aplicativo, pode-se enfatizar às necessidades do letramento digital, como auxílio nos processos de ensino-aprendizagem de Libras, aplicando-se na Educação de Jovens e Adultos (EJA).

Por meio da aplicação destas ferramentas, os estudos de Libras na educação a distância podem ser aprimorados, levando-a à um grande avanço na educação para deficientes auditivos, contribuindo para o processo de formação, capacitação e inclusão social dos deficientes auditivos [Jesus et al. 2013].

5. Conclusões

Esta Revisão Sistemática da Literatura identificou 172 artigos científicos. Com intuito de identificar os principais Ambientes Virtuais Tridimensionais de Aprendizagem, aplicado nos processos de ensino-aprendizagem de Libras, foram selecionados, por meio da aplicação dos critérios de inclusão e exclusão, 11 artigos científicos.

Desse modo, observou-se a aplicabilidade da modelagem 3D para construção de *avatares* ou humanoides, os quais possuíam a função de estabelecer a interação, interpretando e convertendo, através de gestos, a Língua Portuguesa para Libras. Nota-se também que as técnicas de reconhecimento de padrões são bastante empregadas para identificação de símbolos de configurações de mãos no processo da escrita em Libras.

Percebeu-se que os Ambientes Virtuais Tridimensionais de Aprendizagem identificados por meio desta Revisão Sistemática da Literatura utilizava-se das principais técnicas de Realidade Virtual para conversão ou interpretação da Língua Portuguesa para Libras.

Portanto, com estes mecanismos, o uso das tecnologias como apoio educacional deve ser incentivado, ajudando de forma significativa no aprendizado dos deficientes auditivos. Assim, recursos midiáticos, por meio de Ambientes Tridimensionais de ensino, favorecem o aprendizado, minimizando as diferenças entre alunos com deficiência auditiva e ouvintes.

Referências

Aggarwal, J. K. and Ryoo, M. S. (2011). Human activity analysis: A review. *ACM Computing Surveys (CSUR)*, 43(3):16.

Alves, J. B. D. M., Formanski, F. N., and Da Silva, J. B. (2014). New technologies applied to education: A new concept of education. In *Remote Engineering and Virtual Instrumentation (REV), 2014 11th International Conference on*, pages 416–419. IEEE.

BAQUETA, J. J. (2012). Especificação e protótipo de um jogo educativo para aprendizagem de conceitos por crianças surdas. In *Universidade Estadual do Oeste do Paraná*.

Bertacchini, F., Bilotta, E., Gabriele, L., Vizueta, D. E. O., Pantano, P., Rosa, F., Tavernise, A., Vena, S., and Valenti, A. (2013). An emotional learning environment for subjects with autism spectrum disorder. In *Interactive Collaborative Learning (ICL), 2013 International Conference on*, pages 653–659. IEEE.

Brega, J. R. F., Lamarca, A. F., Neto, M. P., Dias, D. R. C., and Trevelin, L. C. (2013). Sistema gerador de apoio a um dicionário temático visual-gestual baseado em realidade virtual. In *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*.

Brega, J. R. F., Rodello, I. A., Dias, D. R. C., Martins, V. F., and de Paiva Guimarães, M. (2014). A virtual reality environment to support chat rooms for hearing impaired and to teach brazilian sign language (libras). In *Computer Systems and Applications (AICCSA), 2014 IEEE/ACS 11th International Conference on*, pages 433–440. IEEE.

Campbell, A. G., Santiago, K., Hoo, D., and Mangina, E. (2016). Future mixed reality educational spaces. In *Future Technologies Conference (FTC)*, pages 1088–1093. IEEE.

de Farias Klimsa, S. B. and Klimsa, B. L. T. (2013). Sistema signwriting de grafia das línguas de sinais: Desafios do ensino-aprendizagem em ambientes virtuais. In *Anais eletrônicos do 5 Simpósio Hipertexto e tecnologias na Educação*.

De Martino, J. M., Silva, I. R., Bolognini, C. Z., Costa, P. D. P., Kumada, K. M. O., Coradine, L. C., da Silva Brito, P. H., do Amaral, W. M., Benetti, Â. B., Poeta, E. T., et al. (2016). Signing avatars: making education more inclusive. *Universal Access in the Information Society*, pages 1–16.

Gonçalves, D. A. (2013). Avatar 3d para síntese automática de sinais da língua de sinais brasileira. In *Programa de Pós-Graduação em Informática, Universidade Federal do Paraná*.

Jesus, L. M. d. et al. (2013). Motion graphic design como ferramenta de educação a distância em libras. In *Universidade Federal de Santa Catarina*.

Kawakami, R., dos Santos Nunes, E. P., and Maciel, C. (2015). Three-dimensional virtual environments that support hospitalized children's learning: A systematic review. In *Virtual and Augmented Reality (SVR), 2015 XVII Symposium on*, pages 190–199. IEEE.

Keele, S. et al. (2007). Guidelines for performing systematic literature reviews in software engineering. In *Technical report, Ver. 2.3 EBSE Technical Report. EBSE*. sn.

La Marca, A. F., Dias, D. R. C., Ricci, C. C. P., and Brega, J. R. F. (2010). Integrating interaction resources in the graphics clusters. In *Virtual Systems and Multimedia (VSMM), 2010 16th International Conference on*, pages 310–313. IEEE.

Martins, G. and Fadel, L. M. (2015). Personagem 3d para atuar como interprete de libras. In *Anais do Congresso Nacional de Ambientes Hipermídia para Aprendizagem*.

Muñoz-Cristóbal, J. A., Gallego-Lema, V., Arribas-Cubero, H. F., Martínez-Monés, A., and Asensio-Pérez, J. I. (2017). Using virtual learning environments in bricolage mode

for orchestrating learning situations across physical and virtual spaces. *Computers & Education*, 109:233–252.

Paschuini, E. A. (2015). A infoinclusão de alunos surdos na educação de jovens e adultos utilizando o aplicativo hand talk em sala de aula. In *Programa de Pós-Graduação em Educação, Universidade Federal do Paraná*.

Scott, E., Soria, A., and Campo, M. (2016). Adaptive 3d virtual learning environments–a review of the literature. *IEEE Transactions on Learning Technologies*.

Stiehl, D. (2015). Reconhecimento de símbolos de configuração de mão do signwriting. In *(Dissertação de Mestrado) Programa de Pós-Graduação de Informática, Universidade Federal do Paraná*.

Wazlawick, R. (2014). *Metodologia de pesquisa para ciência da computação, 2a edição*, volume 2. Elsevier Brasil.

Wholin, C., Runeson, P., Host, M., and Ohlsson, M. R. (2012). B. e wesslén, a. *Experimentation in Software Engineering. Springer Berlin Heidelberg*.

Esteganografia, A Escrita Oculta: Um Mapeamento Sistemático

Hayanna Silva Oliveira [1], Leonardo de Jesus Silva[1], Rafael Battaglin de Almeida Alves[1] Gilton José Ferreira da Silva[1]

[1]Departamento de Computação
Universidade Federal de Sergipe (UFS) – São Cristóvão, SE – Brazil

{hayanna.oliveira,leonardojs,rafaelbaa,gilton}@dcomp.ufs.br

Abstract. *Steganography is the art of hidden writing. And although it has been used since ancient times, it now finds special applications in the digital environment, where information is of great value. In this field,it acts on media files, such as: images, videos and audios. The main objective of this paper is to show a systematic mapping to identify the main ways of applying steganography in digital archives today.Therefore, it is carried out all the methods covered, the media used and the attributes manipulated in the productions of the last 10 years on the theme. Among the 13 papers selected according to the inclusion criteria of this systematic mapping, it was concluded that the most used method is the LSB, the least significant bit. Being the main digital media that steganography acts, verified in this work was the image.*

Resumo. *Esteganografia é a arte da escrita escondida. E, embora seja usada desde os antigos, hoje encontra especiais aplicações no meio digital, onde informação é objeto de grande valia. Neste meio, ela atua sobre arquivos de mídias, como: imagens, vídeos e áudios. Este artigo objetiva apresentar um mapeamento sistemático para identificar as principais formas de aplicação de Esteganografia em arquivos digitais atualmente. Para tanto, é realizado levantamento dos métodos abordados, das mídias utilizadas e dos atributos manipulados nas produções dos últimos 10 anos sobre o tema. Dentre os 13 trabalhos selecionados segundo os critérios de inclusão deste mapeamento sistemático, chegou-se a conclusão que o método mais utilizado é o LSB, o menor bit significante. Sendo a principal mídia digital que a esteganografia atua é a imagem.*

1. Introdução

Esteganografia é uma palavra de origem grega (*estegano* - esconder - e *grafia* - escrita) referente à arte da escrita escondida [Julio et al. 2007]. Apesar de sua história se iniciar na Grécia, e ter sido utilizada por gregos, egípcios e outros povos antigos, foi na Idade Média que a esteganografia foi mais estudada e desenvolvida. O monge Tritheimius, em 1499, teria escrito uma série de livros chamados "*Steganographia*" contendo a descrição de várias técnicas.

Dentre elas, a mais relevante para o entendimento das aplicações de esteganografia atuais seria a Grade de Cardano. Nela, o remetente utilizava uma grade de retângulos definidos randomicamente para escrever as palavras da mensagem secreta. Após a escrita, retirava a grade e preenchia os espaços restantes para formar uma mensagem de cobertura

(o disfarce para a mensagem secreta). O segredo estava, portanto, na quantidade e no posicionamento dos retângulos. Então, o destinatário deveria possuir uma grade semelhante à do emissor, pela qual poderia identificar as palavras dentro dos retângulos e conhecer a mensagem secreta.

Essa técnica é análoga ao tipo de esteganografia de chave secreta, onde o receptor pode reverter o processo e extrair a mensagem secreta se possuir conhecimento da chave utilizada no processo de incorporação. Nesse caso, [Coelho and Bento 2004, p. 15] sugere a seguinte descrição de esteganografia:

> O dado embutido (*embedded data*) é a informação que alguém deseja enviar em segredo. Este dado geralmente fica escondido em uma mensagem aparentemente inocente, chamada de recipiente ou de objeto cobertura (*container* ou *cover-object*), produzindo um estego-objeto (*stego-object*) ou estego-recipiente (*stego-carrier*), ou seja, um arquivo com uma mensagem embutida. Uma estego-key (*stego-key*) ou simplesmente chave é utilizada para controlar o processo de esconder, assim como, para restringir detecção e/ou recuperação do dado embutido, somente para quem a conhece, ou conheça parte dela. Uma possível fórmula deste processo pode ser representada da seguinte forma:
>
> $$recipiente + mensagemembutida + chave = estego - objeto$$

Entretanto, a chave esteganográfica é um item opcional, dispensável nos outros dois tipos de protocolos esteganográficos: esteganografia pura, que não exige troca prévia de informações secretas e o processo de incorporação pode ser descrito como um mapeamento funcional; e esteganografia de chave pública, que não depende da troca de uma chave secreta, pois exige o uso de duas chaves (privada e pública), onde a pública é usada no processo de incorporação e a privada para reconstruir a mensagem secreta [Sokół and Yarmolik 2005].

Portanto, a esteganografia é definida pelo ato de troca de mensagens secretas, onde apenas o remetente e o destinatário estão cientes da mensagem presente no meio de comunicação, sem necessariamente haver a ocultação a mensagem, pois esta está disfarçada em outras mídias ou mensagens [Cox et al. 2002].

No caso das mídias digitais, a esteganografia pode ser aplicada em diversas formas, especialmente na ocultação de informações em imagens, arquivos de texto, áudio, vídeo e transmissão de dados [Sokół and Yarmolik 2005]. Tem sido amplamente utilizada no meio artístico para proteção de direitos autorais diante da pirataria e do compartilhamento ilegal de obras (como filmes, músicas e vídeos) na Internet. Porém, também se mostra sujeita ao uso malicioso por parte de pedófilos, para divulgação de imagens de pornografia infantil na rede, e até de terroristas, havendo indícios de propagação de mensagens de redes terroristas como a *Al-Qaeda* na Internet.

Este mapeamento sistemático tem como objetivo identificar e sistematizar qual o principal método utilizado de esteganografia, bem como sobre quais tipos de mídia ele atua. Para isso, foram mapeados trabalhos em diversas bases de dados na área da Ciência da Computação.

A Seção 2 deste artigo aborda as aplicações da Esteganografia em mídias digitais, sejam elas imagem, video ou som. Na Seção 3, é descrito o método de pesquisa adotado

neste trabalho para realização do mapeamento sistemático. Na Seção 4, é realizada a análise dos resultados encontrados. Finalmente, na Seção 5, constam as considerações finais dos autores.

2. Aplicações da Esteganografia

Em decorrência do advento da Internet e das vulnerabilidades que esta apresenta à informação, é crescente a quantidade de aplicações de esteganografia digital. Esta consiste em um conjunto de técnicas e algoritmos capazes de permitir uma comunicação digital mais segura [Rocha 2003].

Podendo prover desde a inserção de imagens em outras, de forma que uma imagem irrelevante oculte uma de maior importância, até a escrita de textos aleatórios que escondam outro de conteúdo confidencial em sua estrutura. Oferecendo segurança, por exemplo, a dados que transitem entre *e-mails* e computadores pessoais, possivelmente sujeitos a invasões.

Também se mostrou uma grande ferramenta para a proteção dos direitos autorais, oferecendo técnicas eficientes e eficazes para inibir a realização de cópias não-autorizadas. Pois, uma vez que obras dispostas em arquivos digitais podem ser facilmente copiadas e reproduzidas, técnicas avançadas de "marcas-d'água" (*watermarking*) e de identificação por digitais (*fingerprinting*) foram desenvolvidas como tentativas de restringir a pirataria indiscriminada.

Além disso, a esteganografia digital tem ainda aplicação no meio militar, em modernos aparelhos que fazem transmissões de rádio e codificam em ondas-curtas mensagens mais importantes.

2.1. Esteganografia em Imagem

As abordagens mais comuns de inserção de mensagens em imagens incluem técnicas de: modificação dos bits menos significativos (LSB — *Least Significant Bits*), filtragem e mascaramento, e algoritmos e transformações [Julio et al. 2007, Rocha 2003]. Explicadas a seguir.

LSB: Pela técnica de LSB, os bits menos significativos dos valores de pixel no domínio espacial são modificados. Nas implementações mais básicas, o plano LSB inteiro é substituído por dados esteganográficos. Porém, a inclusão de LSB simples é suscetível a processamento de imagem, especialmente à técnica de compressão sem perda [Julio et al. 2007].

É possível aplicar a técnica a cada *pixel* de uma imagem codificada em 32*bits* por *pixel*. Nessas imagens, há uma codificação em quatro *bytes* (aRGB) que representam: alfa (*Alpha transparency*), vermelho (*Red*), verde (*Green*) e azul (*Blue*) [Rocha 2003]. De forma que o bit menos significativo de cada *byte* do *pixel* pode ser modificado para representar um *bit* da informação a ser escondida sem comprometer a imagem.

Filtragem e Mascaramento: Mais robustas que a inserção LSB [Julio et al. 2007, Rocha 2003], as técnicas baseadas em filtragem e mascaramento geram imagens esteganográficas imunes à compressão e ao recorte. Entretanto, por atuarem sobre os bits mais significativos das imagens, são mais propensas à detecção,

pois geram muitos artefatos. Também é por isso que atuam exclusivamente sobre imagens em tons de cinza (*grayscale*).

A informação é escondida como uma marca d'água (técnica de *watermarking*), onde valores de pixel são aumentados ou diminuidos por uma porcentagem. Esta técnica garante a permanência da marca ainda que a imagem sofra compressão.

Algoritmos e Transformações: Ao contrário das técnicas de LSB, que são vulneráveis à compressão, as técnicas de algoritmos e transformações fazem uso dela para realizar um processo mais sofisticado de manuseio das imagens. Também podem atuar sobre o brilho e a saturação da imagem, por exemplo [Rocha 2003].

O resultado é maior resistência contra processamento de sinal, uma vez que os dados escondidos são espalhados pelas áreas mais robustas de toda a imagem, no domínio de transformação [Julio et al. 2007].

2.2. Esteganografia em Vídeo

Vídeos combinam as dimensões de espaço e tempo, entretanto, dentro do tempo eles são exatamente imagens (*frames*) [Böhme 2010, Julio et al. 2007]. Assim, o processo de esteganografia em vídeo se dá de forma similar ao de imagens, porém atuando sobre os *frames*.

O método normalmente utilizado é a transformada de cosseno discreta (DCT - *Discrete Cosine Transform*). Vantajosa na compressão de dados por transferir a maior parte da informação contida para os primeiros elementos do vetor, de modo que, para compressões sem perda, o armazenamento é otimizado e, para compressões com perda, a quantização dos valores é facilitada.

2.3. Esteganografia em Áudio

A esteganografia em áudio considera o ambiente em que o som deve trafegar, levando em conta a representação digital do sinal que será usado e o caminho de transmissão do sinal. Sob análise dessas condições, podem ser adotadas técnicas como: Codificação *Low-Bit*, Codificação em Fase, *Spread Spectrum* e Eco.

Codificação Low-Bit: Análoga ao LSB, substitui o bit menos significativo de cada amostra por um codificador binário. Tal processo gera ruídos, mas a percepção é relativa a depender do conteúdo original.

Codificação em Fase: Neste tipo de codificação, a fase de um segmento inicial de áudio é substituída por uma fase de referência que representa os dados a serem escondidos, e a dos subsequentes é ajustada para preservar a fase relativa.

Assim, a dispersão de fase só é notada quando há uma mudança drástica entre a relação de fase entre cada componente de frequência; do contrário, a codificação é inaudível, pois os componentes de fase de um som são mais imperceptíveis que o ruído ao ouvido humano.

Spread Spectrum: Esta técnica espalha informações secretas sobre o espectro de frequências de áudio utilizando um código independente do sinal. Onde o sinal resultante ocupa uma banda superior à utilizada para a transmissão do sinal original.

Apesar de introduzir ruído, ela é capaz de oferecer uma taxa de transmissão moderada enquanto mantém alto nível de robustez contra técnicas de remoção.

Eco: Informações secretas podem ser escondidas em um arquivo de áudio pela introdução de eco, variando parâmetros do sinal como: amplitude, taxa de deterioração e variação do sinal original (*offset*). Eles são configurados abaixo dos limites que o ouvido humano pode perceber, e no *offset* é representada a mensagem binária codificada.

3. Método de Pesquisa

Com objetivo de estudar a técnica de ocultação da escrita, esteganografia, foi escolhido o método de mapeamento sistemático (SLM, *Systematic Literature Mapping*). Para [Petersen et al. 2008], o levantamento sistemático é um estudo secundário produto da análise dos estudos primários sobre determinada temática. Esse estudo tem como finalidade prover um conhecimento geral sobre uma determinada área, questão de pesquisa e assunto.

Nesta seção, é descrito o processo de seleção dos estudos primários, bem como as questões de pesquisas discutidas neste trabalho para produzir o estudo secundário. Nas subseções seguintes, são apresentadas as questões elaboradas e, posteriormente, os processos de seleção dos trabalhos, mediante critérios de inclusão e exclusão.

3.1. Questões de Pesquisa

Este mapeamento sistemático se concentra na identificação das principais formas de aplicação de Esteganografia em arquivos digitais nos últimos 10 anos. Para tanto, foram elaboradas quatro questões a serem respondidas mediante a análise dos trabalhos selecionados:

Q1. Qual o método de Esteganografia abordado na publicação?
Q2. Qual o ano na publicação do trabalho?
Q3. Sobre que tipo de mídia o método atua (áudio, vídeo, imagem)?
Q4. Que atributo é manipulado?

3.2. Estratégia de Busca

As bases de dados na área de Ciência da Computação selecionadas para a busca foram: *Scopus*[1] , *IEEE Xplore*[2] e a *ScienceDirect*[3]. O serviço CAFE disponibilizado no portal de periódicos da CAPES foi utilizado para o *download* dos trabalhos sem restrições nas três bases.

Foi elaborada uma *string* genérica para padronizar a busca nas bases: *steganography* AND *media* AND (*"data-hiding"* OR *"information hiding"*) *algorithm* AND *"secret message"*. Entretanto, a *string* necessitou ser alterada para reconhecimento da composição com a palavra *algorithm* na ScienceDirect e na IEEE. Na Scopus, a busca se deu apenas em títulos, resumos e palavras chaves. Também foi utilizado refinamento em todas para obtenção de publicações somente entre os anos de 2007 e 2017.

Conforme as peculiaridades de cada base, as *strings* de busca foram definidas da seguinte maneira:

[1]https://www.scopus.com/
[2]http://ieeexplore.ieee.org/Xplore/home.jsp
[3]http://www.sciencedirect.com/

- **Scopus:** TITLE-ABS-KEY (*steganography* AND *media* AND (*"data-hiding"* OR *"information hiding"*) *algorithm* AND *"secret message"*) AND PUBYEAR > 2006
- **IEEE:** (*steganography* AND *media* AND (*"data-hiding algorithm"* OR *"information hiding algorithm"*) AND *"secret message"*)
- **ScienceDirect:** *steganography* AND *media* AND (*"data-hiding algorithm"* OR *"information hiding algorithm"*) AND *"secret message*"AND PUB-DATE > 2006

As pesquisas com as *strings* de busca foram realizadas no mês de julho de 2017 e retornaram um total de 168 trabalhos encontrados, seguindo a distribuição da Tabela 1.

Tabela 1. Quantidade de trabalhos encontrados em cada base

Scopus	*IEEE*	*ScienceDirect*
54	102	12

Sobre tais trabalhos foi iniciada a fase de seleção, obedecendo aos critérios e procedimentos apresentados na subseção seguinte.

3.3. Critérios de Seleção

A busca por trabalhos relevantes para este mapeamento incluiu a realização de um processo de seleção entre os trabalhos encontrados nas bases sob os critérios de inclusão (CI) e de exclusão (CE) apresentados na Tabela 2.

Tabela 2. Critérios de inclusão (CI) e de exclusão (CE) para seleção de Trabalhos

Critérios de Inclusão	Critérios de Exclusão
CI1- Trabalhos com foco em Esteganografia. CI2 - Trabalhos que descrevem aplicações/algoritmos de ocultação de dados/informações em arquivos de mídia.	CE1 - Trabalhos duplicados. CE2 - Trabalhos que não estejam escritos em inglês. CE3 - Trabalhos publicados há mais de 10 anos. CE4 - Trabalhos que não estavam disponíveis para visualização do texto completo. CE5 - Trabalhos que não descrevem métodos e/ou técnicas de Esteganografia.

A Figura 1, mostra o processo de seleção dos trabalhos, de acordo com os critérios de inclusão e exclusão, para analise deste trabalho. Entende-se por processo, segundo o [CBOK 2013], a seguinte definição: “é um conjunto definidos de atividades ou comportamentos executados por humanos ou máquinas para alcançar uma ou mais metas”. Para demonstrar o processo de seleção de trabalhos para este estudo, se fez uso do BPMN

[Silver 2011], uma linguagem de diagramação para processos de negócio, conforme apresenta a Figura 1.

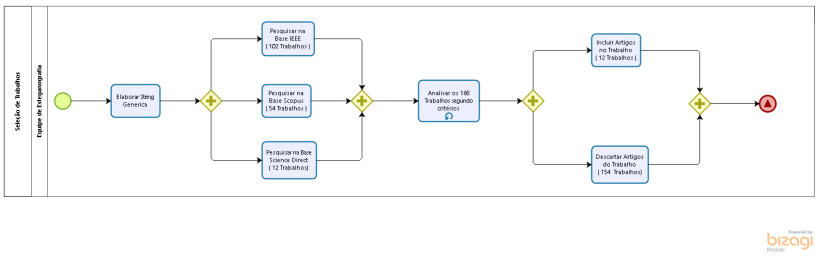


Figura 1. Processo de seleção dos trabalhos de acordo com os critérios.

O processo é inicializado, então é elaborada a *string* de busca. Realiza-se uma ação paralela de busca nas 3 bases (*Scopus*, *IEEE*, *ScienceDirect*), para retornarem cada uma seus respectivos resultados. Este *output* entra numa tarefa de avaliar cada artigo retornado, segundo os critérios de inclusão e exclusão. Consequentemente, duas tarefas são realizadas paralelamente: uma separação dos trabalhos pelos critérios de inclusão e outra pelos critérios de exclusão. Por fim, segue um evento de finalização por sinalização, indicando o fim deste processo para que outro processo se inicie ao receber esse sinal.

Após esse processo, dos 168 trabalhos das bases encontrados com a *string* de busca, 12 trabalhos foram selecionados: [Shanableh 2012] [Mstafa et al. 2017][Debnath et al. 2017][Li et al. 2017][Huang et al. 2016][Shi et al. 2016] [Kuo-Chen Wu and Wang Chung-Ming 2015] [Gue et al. 2014] [Weiqi Luo et al. 2010] [Lin et al. 2017] [Amirtharaj et al. 2013] [Thanikaiselvan et al. 2016], distribuídos como indica a Tabela 3.

Tabela 3. Quantidade de trabalhos selecionados em cada base

Scopus	IEEE	ScienceDirect
0	10	2

É dado início ao processo de extração das informações, a fim de responder as questões de pesquisa levantadas neste estudo. A Figura 2 demonstra o processo que compõe o conjunto de atividades realizadas, desde elaboração de questões de pesquisa até a tabulação das respostas extraídas de cada artigo selecionado.

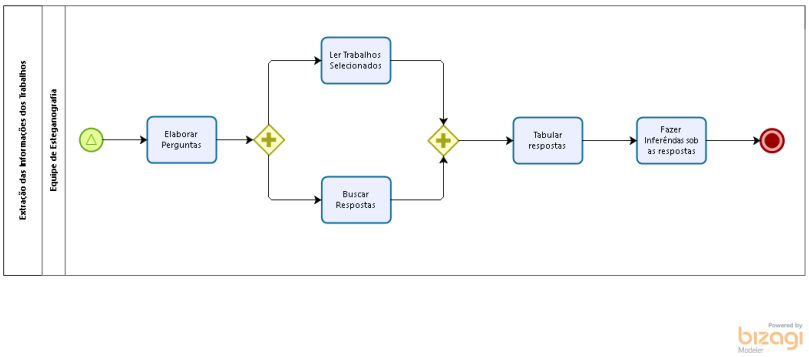


Figura 2. Processo de extração de informações dos trabalhos selecionados

Este processo é inicializado pelo recebimento de um sinal de finalização do processo da Figura 1, então as questões de pesquisas são elaboradas. Em seguida, é realizada uma ação paralela de leitura dos trabalhos selecionados segundo os critérios de inclusão e de busca pelas respostas em cada artigo de acordo com as questões de pesquisa elaboradas. Consequentemente, ocorre a tabulação destas respostas para realização das inferências. Só então é realizada a finalização deste processo.

4. Análise dos Resultados

Nesta seção, são apresentados os resultados obtidos após a execução dos processos definidos na Seção 3 aplicados aos estudos primários extraídos das bases.

A resposta para a questão de pesquisa Q1. Qual o método de Esteganografia abordado na publicação?, está ilustrada na Figura 3. Ela indica os tipos de métodos encontrados nos estudos primários e a frequência de cada um.

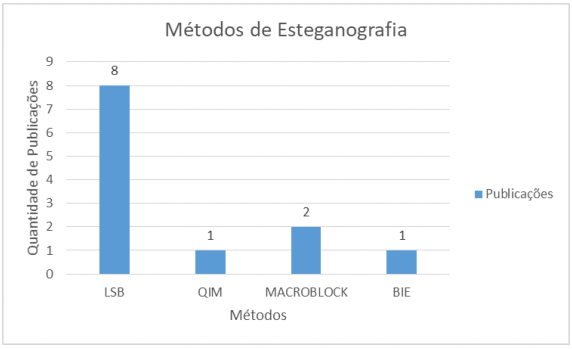


Figura 3. Métodos de esteganografia abordados nos trabalhos

Como demonstra o gráfico da Figura 3, o método mais recorrente nos trabalhos selecionados foi o LSB (*Least Significant Bit*). Abordado na Seção 2 como o método que realiza substituição de *bits* menos significativos para Esteganografia em imagens. Seguido por *Macroblock*, que utiliza transformações de bloco linear. E, em menor incidência, por QIM (*Quantization Index Modulation*), uma abordagem por ajuste que usa diferentes quantizadores para transportar diferentes *bits* dos dados secretos, e por BIE (*Bit Inversion Embedding*), um esquema de incorporação de inversão de *bits*.

A resposta para a questão de pesquisa Q2. Qual o ano de publicação do artigo?, está ilustrada na Figura 4, que mostra a distribuição de publicações por ano, entre os estudos primários selecionados para análise neste trabalho.

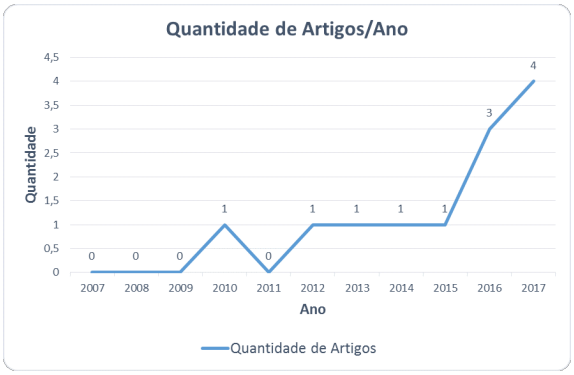


Figura 4. Publicações por ano

O gráfico indica crescente produção de trabalhos na área de Esteganografia nos últimos anos. Com um salto significativo entre os anos de 2016 e 2017.

A resposta para a questão de pesquisa Q3. Sobre que tipo de mídia o método atua (áudio, vídeo, imagem)?, é ilustrada pela Figura 5, que indica a quais tipos de arquivos de mídias estavam relacionadas as práticas de Esteganografia abordadas nos trabalhos selecionados.

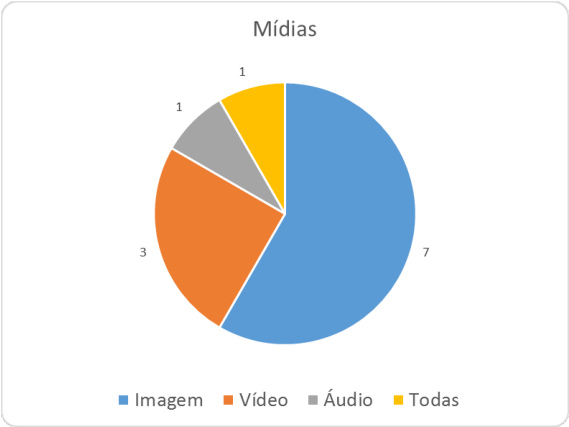


Figura 5. Mídias tratadas nos trabalhos

Sendo verificada a (previsível) predominância de produções sobre técnicas de Esteganografia em imagens, seguida por Esteganografia em arquivos de vídeo e, em menor frequência, de áudio. Também foi considerado um artigo que abordou práticas sobre todos estes tipos de arquivos (imagem, vídeo e áudio).

A resposta para a questão de pesquisa Q4. Que atributo é manipulado?, é ilustrada na Figura 6, que indica quais atributos são manipulados nas práticas descritas nos trabalhos encontrados nos estudos primários.

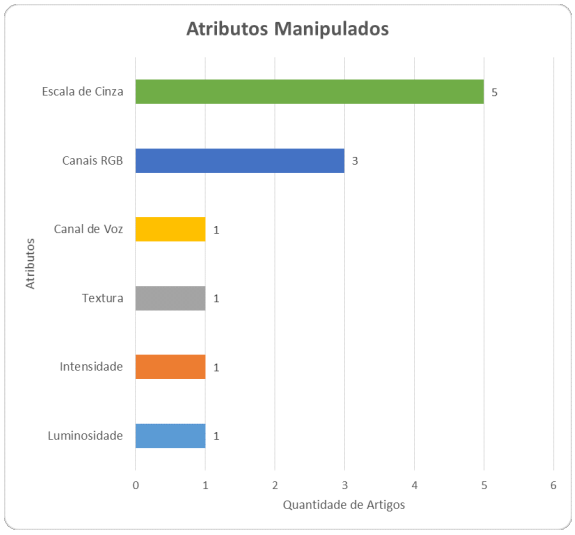


Figura 6. Atributos manipulados

Por este gráfico, nota-se que o atributo mais manipulado foi a escala de cinza (*grayscale*) das imagens. O que é justificado pelo fato de que, ao utilizar a técnica LSB (*Least Significant Bit*) sobre imagens em escala de cinza, a mudança é mais sutil, dificultando a descoberta da ocultação de informação.

5. Considerações Finais

Neste trabalho foi realizado um mapeamento sistemático com o objetivo de analisar a utilização da Esteganografia em mídias digitais e quais métodos e/ou técnicas são mais utilizadas. Inicialmente, foram obtidos 168 resultados e, após a aplicação de critérios de inclusão e exclusão, foram selecionados 12 trabalhos relevantes.

Nesse mapeamento, foi identificada a predominância da técnica LSB (*Least Significant Bit*) entre os trabalhos estudados, enquanto houve pouca ocorrência das outras (*Macroblock*, QIM e BIE). Predominando também a abordagem de Esteganografia em imagens, com baixa frequência de uso em vídeo e áudio.

Entre os anos pesquisados, houve um aumento na quantidade de trabalhos realizados na área de Esteganografia, com uma diferença significativa no número de publicações durante os anos de 2016 e 2017 em relação à quantidade de publicações nos anos anteriores.

Por fim, foram analisados os atributos manipulados em cada técnica abordada nos trabalhos. Com destaque para manipulação de tons em escala de cinza, seguida pela modificação de canais RGB e outros menos relevantes em quantidade: canal de voz, textura, intensidade e luminosidade.

Acredita-se que este trabalho possui resultado relevante ao estudo da Segurança da Informação ao relacionar as práticas vigentes de ocultamento de mensagens e seus modos de abordagem. Também faz saber o estado da arte no que se refere a Esteganografia em mídias digitais.

Como trabalhos futuros, é esperada a realização de estudos detalhados sobre Esteganografia em imagens por meio da aplicação da técnica de LSB. Pesquisando formas de

implementação, casos de uso e, possivelmente, técnicas de Esteganálise aplicáveis, bem como formas de otimização das técnicas vigentes na literatura.

Referências

Amirtharaj, R., Anushiadev, R., Meena, V., Kalpana, V., and Rayappan, J. (2013). Concealed to Protect and Protect to Conceal: A Conserved Stego Image. *Research Journal of Information Technology*, 5(2):209–217.

Böhme, R. (2010). Principles of modern steganography and steganalysis. *Advanced Statistical Steganalysis*, pages 11–77.

CBOK, B. (2013). Guia para o gerenciamento de processos de negócio corpo comum de conhecimento. *Association of Business Process Management Professionals. ABPMP BPM CBOK*, 3.

Coelho, L. C. M. and Bento, R. J. (2004). Ferramentas de esteganografia e seu uso na Infowar. *ICCyber2004*, page 14.

Cox, I. J., Miller, M. L., Bloom, J. A., and Honsinger, C. (2002). *Digital watermarking*, volume 1558607145. Springer.

Debnath, B., Das, J. C., and De, D. (2017). Reversible logic-based image steganography using quantum dot cellular automata for secure nanocommunication. *IET Circuits, Devices & Systems*, 11(1):58–67.

Gue, S.-T., Yeh, H.-L., Shih, W.-K., and Tsai, P. (2014). Reversible video data hiding using neighbouring similarity. *IET Signal Processing*, 8(6):579–587.

Huang, F., Huang, J., and Shi, Y. Q. (2016). New framework for reversible data hiding in encrypted domain. *IEEE Transactions on Information Forensics and Security*, 11(12):2777–2789.

Julio, E. P., Brazil, W. G., and Neves, C. V. (2007). Esteganografia e suas Aplicações. *Livro de minicursos do SBSEG*, pages 54–102.

Kuo-Chen Wu and Wang Chung-Ming (2015). Steganography Using Reversible Texture Synthesis. *IEEE Transactions on Image Processing*, 24(1):130–139.

Li, S., Jia, Y., and Kuo, C.-C. J. (2017). Steganalysis of QIM Steganography in Low-Bit-Rate Speech Signals. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(5):1011–1022.

Lin, Y.-T., Wang, C.-M., Chen, W.-S., Lin, F.-P., and Lin, W. (2017). A Novel Data Hiding Algorithm for High Dynamic Range Images. *IEEE Transactions on Multimedia*, 19(1):196–211.

Mstafa, R. J., Elleithy, K. M., and Abdelfattah, E. (2017). A Robust and Secure Video Steganography Method in DWT-DCT Domains Based on Multiple Object Tracking and ECC. *IEEE Access*, pages 1–1.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. In *EASE*, volume 8, pages 68–77.

Rocha, A. (2003). Camaleão: um software para segurança digital utilizando esteganografia, 2003. *Monografia. Depto. de Ciência da Computação, Universidade Federal de Lavras*.

Shanableh, T. (2012). Data Hiding in MPEG Video Files Using Multivariate Regression and Flexible Macroblock Ordering. *IEEE Transactions on Information Forensics and Security*, 7(2):455–464.

Shi, Y.-Q., Li, X., Zhang, X., Wu, H.-T., and Ma, B. (2016). Reversible data hiding: Advances in the past two decades. *IEEE Access*, 4:3210–3237.

Silver, B. (2011). Bpmn method and style, with bpmn implementer's guide: A structured approach for business process modeling and implementation using bpmn 2.0. *Cody-Cassidy Press, Aptos, CA*, 450.

Sokół, B. and Yarmolik, V. N. (2005). Cryptography and Steganography: teaching experience. In *Enhanced Methods in Computer Security, Biometric and Artificial Intelligence Systems*, pages 83–92. Kluwer Academic Publishers, Boston.

Thanikaiselvan, V., Bansal, T., Jain, P., and Shastri, S. (2016). 9/7 IWT Domain Data Hiding in Image using Adaptive and Non Adaptive Methods. *Indian Journal of Science and Technology*, 9(5).

Weiqi Luo, Fangjun Huang, and Jiwu Huang (2010). Edge Adaptive Image Steganography Based on LSB Matching Revisited. *IEEE Transactions on Information Forensics and Security*, 5(2):201–214.

Gestão de Continuidade do Negócio (GCN): Normas, Recomendações e Modelos

Mariane Moreira Batista[1], Edison Andrade Martins Morais[2], Ana Cláudia Bastos Loureiro Monção [3]

[1]Escola de Engenharia Elétrica, Mecânica e da Computação – Universidade Federal de Goiás (UFG)
CEP – 74.605-010 – Goiânia – GO – Brasil

[2,3]Instituto de Informática - Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.0001-970 – Goiânia, GO – Brasil

mbatista1711@gmail.com, edison@inf.ufg.br, anaclaudia@inf.ufg.br

Abstract. *The business operations of companies are becoming increasingly dependent on resources and services in the area of Information Technology, and it's necessary that these resources are available in an appropriate way. In this context, the area of Business Continuity Management (BCM) emerge, which aims to ensure that the activities of the companies continue to be realized out with the minimum possible impact in the event of unexpected incidents and interruptions. The objective of this article is to present the main norms, recommendations and models that define BCM as part of its processes, showing the relevance of this topic to the IT area.*

Resumo. *As operações de negócio das empresas estão se tornando cada vez mais dependentes de recursos e serviços da área de Tecnologia da Informação (TI), sendo necessário que estes recursos estejam disponíveis de forma adequada. Neste contexto surge a área de Gestão de Continuidade do Negócio (GCN), que visa garantir que as atividades das empresas continuem sendo realizadas com o mínimo de impacto possível diante de incidentes e interrupções inesperadas. O objetivo deste artigo é apresentar quais são as principais normas, recomendações e modelos que definem a GCN como parte de seus processos, mostrando a relevância deste tema para a área de TI.*

1. Introdução

A disponibilidade crescente das tecnologias tem gerado um grande volume de informação e também de serviços disponibilizados. Desta forma, as operações de negócio das empresas estão se tornando cada vez mais dependentes de recursos e serviços da área de TI, sendo necessário que a mesma garanta não só a segurança destas informações, mas também a disponibilidade dos sistemas suportados por estas organizações.

Para que determinados acontecimentos, tidos como incidentes, contingências ou desastres, causem o mínimo de impacto possível nas operações do negócio, tem sido cada vez mais necessário a implantação da Gestão de Continuidade do Negócio (GCN).

A GCN é definida como o processo cujo objetivo é garantir que os serviços essenciais de uma empresa, identificados previamente, não sejam impactados durante a ocorrência de um desastre, até o retorno à situação normal de funcionamento [ABNT 2015].

O objetivo deste artigo é apresentar quais são as principais normas, recomendações e modelos que definem a GCN com parte de seus processos. Neste sentido o artigo está dividido em quatro seções. A seção 2 apresenta os conceitos relacionados à GCN. A seção 3 apresenta as principais normas, recomendações e modelos e como elas se relacionam com a GCN. A seção 4 é a conclusão do trabalho.

2. Gestão de Continuidade do Negócio

Antes de compreender o conceito de Gestão da Continuidade do Negócio (GCN), é importante entender o que é a Continuidade do Negócio. Trata-se da capacidade de uma empresa em identificar e se preparar para continuar suas principais funções de negócios, mesmo em situações extremas, alinhando processos, pessoas e tecnologias, melhorando sua resiliência com objetivo de proteger seu patrimônio, de seus acionistas e investidores.

Já o termo gestão está relacionado à garantia de que as atividades desta empresa continuem sendo realizadas com o mínimo de impacto possível diante de incidentes e interrupções inesperadas, principalmente através da identificação prévia das ameaças e riscos, e da análise dos impactos no negócio caso estes ocorram. Ou seja, gerir a continuidade do negócio significa estar preparado para minimizar o impacto na empresa caso as ameaças e riscos ocorram.

A norma ISO 22301 [ABNT 2013], principal norma de GCN disponível atualmente, especifica os requisitos para planejar, estabelecer, implementar, operar, monitorar, analisar criticamente, manter e melhorar continuamente um sistema de gestão documentado para se proteger, reduzir a possibilidade de ocorrência, preparar-se, responder e recuperar-se de incidentes de interrupção quando estes ocorrerem. Ela trata da capacidade de uma organização de continuar com a entrega de seus produtos ou serviços em um nível aceitável, previamente definido, após um incidente de interrupção.

O principal produto gerado a partir da GCN é o Plano de Continuidade do Negócio (PCN). Trata-se de um conjunto de documentos que contém as estratégias e planos de ação a serem adotados em uma possível interrupção das atividades consideradas críticas à organização.

Além do PCN, a GCN também é responsável por estabelecer uma política de continuidade de negócios, que define as diretrizes que serão seguidas na execução do PCN. O objetivo desta política é garantir maior eficiência na recuperação das atividades críticas que fazem parte do escopo da GCN.

O PCN geralmente é criado através de uma ferramenta conhecida como Análise do Impacto do Negócio (BIA – *Business Impact Analysis*). A partir da BIA é registrado no PCN o produto da análise dos riscos realizada na organização, através de um relatório que contém a descrição dos riscos potenciais caso uma interrupção das operações de negócio ocorra, além das atividades que devem ser mantidas prioritariamente [ABNT 2015].

Este relatório também pode conter outras informações importantes, tais como Objetivo de Tempo de Recuperação (RTO – *Recovery Time Objective*) e Objetivo do Ponto de Recuperação (RPO – *Recovery Point Objective*). O RTO é o tempo necessário para a restauração do nível de serviço de TI de forma a não comprometer a continuidade dos negócios. O RPO é o período máximo tolerado de perda de dados em um incidente sem a possibilidade de recuperação através de um backup de dados [ABNT 2015].

Além dos itens definidos a partir da BIA, o PCN também pode documentar outros planos, tais como: Plano de Contingência (CP – *Contingency Plan*), Plano de Gerenciamento de Incidentes (IM *Plan – Incident Management Plan*) e Plano de Recuperação de Desastres (DRP – *Disaster Recovery Plan*).

O Plano de Contingência[1] documenta as atividades que devem ser realizadas para minimizar os impactos causados em uma situação de contingência, através do registro das estratégias que devem ser seguidas nestas situações, e da descrição de todas as condições para a retomada da operação normal da organização em caso de desastres.

O Plano de Gerenciamento de Incidentes contém o registro das ações necessárias para minimizar os impactos causados pela ocorrência de incidentes. Um incidente é um é uma interrupção não planejada de um serviço de tecnologia da informação [FREITAS, 2011, p. 187].

Na prática, toda vez que ocorre um incidente que cause a interrupção de um serviço, a primeira ação é a restauração da operação normal deste serviço, para que o impacto do negócio seja minimizado. O ciclo de vida de tratamento de um incidente pode ser visualizado na figura 1, a seguir:

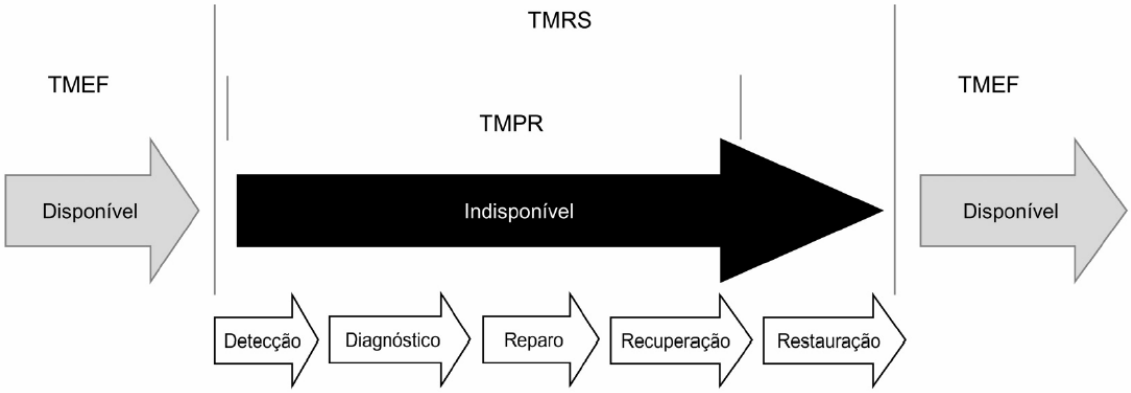


Figura 1. Ciclo de Vida Expandido do Incidente [FREITAS 2011, p. 194]

Detecção é a medição do tempo que gasto na identificação do incidente. Diagnóstico é o tempo gasto para determinar a origem e causa do incidente. Reparo é tempo gasto para reparar o serviço. Recuperação é o tempo gasto para recuperar o serviço. Restauração é o tempo gasto para restaurar o serviço.

Já o Plano de Recuperação de Desastres deve conter a descrição das atividades necessárias para que uma organização mantenha suas atividades, e restabeleça suas

[1] Contingência é uma eventualidade, um acaso, um acontecimento que tem como fundamento a incerteza de que pode ou não acontecer. É a característica daquilo que é contingente, ou seja, que é duvidoso, possível, mas incerto, que pode ocorrer, mas não necessariamente.

operações de negócios, caso ocorra um desastre. Este plano deve atender pelo menos os seguintes requisitos: prover um ambiente seguro e pessoas preparadas para um desastre; reduzir as perdas financeiras em casos de desastres; identificar linhas de negócios críticas que requeiram suporte em situações de desastres; identificar as fraquezas e executar um programa da prevenção de desastre; minimizar a duração de uma paralisação das operações de negócio; facilitar a coordenação eficaz de tarefas da recuperação; reduzir a complexidade do esforço de recuperação.

Para garantir a efetividade destes planos, é importante que sejam realizados testes de recuperação de desastre, através da simulação controlada de um incidente que requisite o acionamento do plano. Os resultados devem ser cuidadosamente avaliados, e podem gerar a necessidade de alterações nos planos registrados do PCN. Estes testes devem ser realizados periodicamente, ficando a cargo da empresa definir a periodicidade de execução dos mesmos. A figura 2, a seguir, mostra a evolução da atividade de GCN ao longo do tempo.

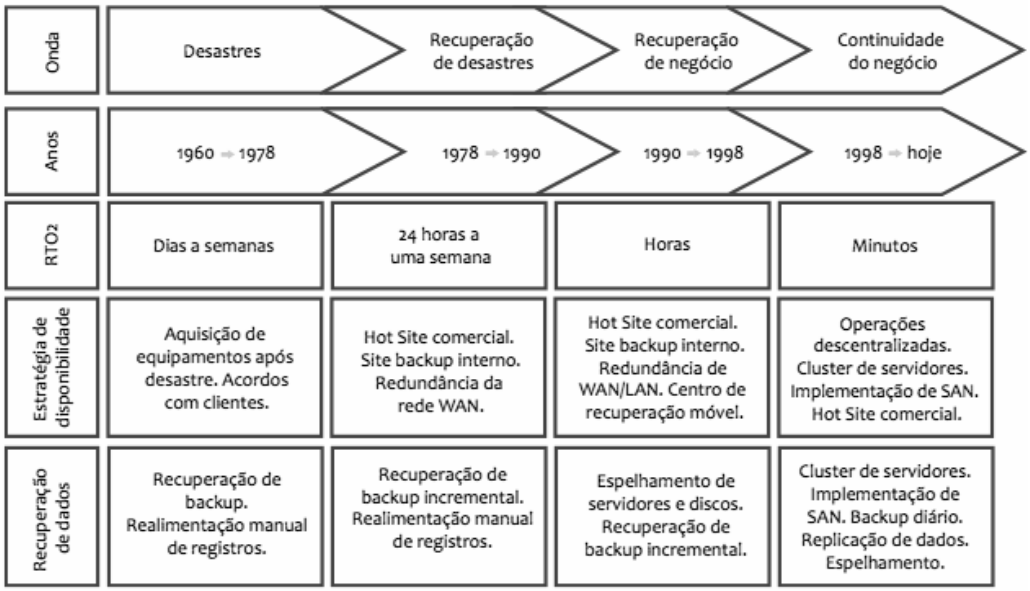


Figura 2: Evolução da Continuidade do Negócio [GUINDANI, 2008]

Ao analisar a figura 2 é possível observar que procedimentos de recuperação de desastres que, entre os anos de 1960 e 1978, levavam dias ou semanas para serem realizados, atualmente são cumpridos em curto espaço de tempo através de mecanismos automatizados.

Também é possível inferir que o foco na identificação de desastres evoluiu para a preocupação com a continuidade do negócio, principalmente através da descentralização das operações organizacionais, evolução das políticas de backup, replicação de dados e estratégias de espelhamento.

Considerando o contexto da GCN, a próxima seção apresenta as principais normas, recomendações e modelos que definem a GCN com parte de seus processos.

3. Normas, Recomendações e Modelos para GCN

Antes de apresentar as principais normas, recomendações e modelos que definem a GCN com parte de seus processos, é necessário fazer a distinção conceitual destes três elementos.

Uma Norma é o documento estabelecido por consenso e aprovado por um organismo reconhecido, que fornece regras, diretrizes ou características mínimas para atividades ou para seus resultados, visando à obtenção de um grau ótimo de ordenação em um dado contexto. A norma é, por princípio, de uso voluntário, mas quase sempre é usada por representar o consenso sobre o estado da arte de determinado assunto, obtido entre especialistas das partes interessadas [ABNT 2017].

Uma Recomendação é um conjunto de boas práticas de mercado que orientam e aconselham os processos de gestão e operacionais das empresas. Geralmente são desenvolvidas por centros de pesquisa ou instituições sem fins lucrativos que agrupam profissionais de determinadas áreas de conhecimento para trocarem experiências sobre o que pode dar certo ou errado no mundo corporativo. O resultado desse trabalho em grupo resulta na publicação de guias ou recomendações que podem servir como referência para as empresas buscarem a excelência nos seus processos. [FREITAS 2011, p. 49].

Já um Modelo (ou *Framework*) é uma forma de descrever um determinado conjunto de componentes e como estes componentes se relacionam entre si para descrever as principais funções de um objeto, sistema ou conceito [ISACA 2012, p. 95].

Ao contrário das normas, recomendações e modelos não certificam. Certificação é um processo no qual uma entidade independente (3ª parte) avalia se determinados processos atendem àquilo que foi definido em suas respectivas normas técnicas. Esta avaliação se baseia em auditorias nos processos. O resultado satisfatório destas atividades leva à concessão da certificação [ABNT 2017].

Nesta seção serão apresentadas as normas ISO/IEC 27001 [ABNT 2013] e ISO/IEC 20000 [ABNT 2011], a recomendação ITIL [ITIL 2007] e o modelo COBIT [ISACA, 2012].

A norma ISO 27001 especifica os requisitos necessários para a implantação de um Sistema de Gerenciamento de Segurança da Informação (SGSI) em uma organização, que visa garantir a proteção dos seus dados de forma eficiente.

Os princípios da norma 27001 garantem a implantação de um modelo SGSI que faça uma abordagem completa à segurança da informação, tratando de múltiplos temas relacionados, dentre elas: as telecomunicações, a segurança aplicacional, a proteção do meio físico, os recursos humanos, o licenciamento e a continuidade do negócio [GOMES, 2016]. Já a norma ISO 27002, representa uma evolução da 27001, complementando esta norma, através da recomendação de boas práticas.

A norma ISO 20000 especifica requisitos para o provedor de serviço planejar, estabelecer, implementar, operar, monitorar, analisar criticamente, manter e melhorar um Sistema de Gerenciamento de Serviços (SGS). A norma inclui termos e definições, requisitos gerais para o sistema de gestão de serviços, desenho e transição de serviços

novos e modificados e os processos de fornecimento de serviço, de relacionamento, de resolução e de controle.

A Biblioteca de Infraestrutura de Tecnologia da Informação (ITIL – *Information Technology Infrastructure Library*) é um conjunto de recomendações que definem boas práticas que podem ser aplicadas na infraestrutura, operação e Gerenciamento de Serviços de Tecnologia da Informação (ITSM – *Information Technology Service Management*). Trata-se de um conjunto de recomendações que contém boas práticas para o gerenciamento dos serviços de TI.

A biblioteca ITIL provê um abrangente e consistente conjunto de melhores práticas para a identificação de processos da área de TI e o alinhamento dos seus serviços às necessidades da organização, tendo como objetivo obter vantagens para a organização tanto em termos de redução de custos pelo aumento da eficiência na entrega e suporte dos serviços de TI, quanto no aumento da capacidade da organização de gerar receita. Desta forma, a área de TI pode concentrar os seus esforços em novos projetos para o atendimento à estratégia de negócio da organização. [MAGALHÃES, 2007].

Em sua versão atual, a ITIL passou a adotar um modelo denominado ciclo de vida dos serviços de TI, onde o gerenciamento dos serviços de TI é dividido em cinco fases, que podem ser observadas através da figura 3, a seguir.

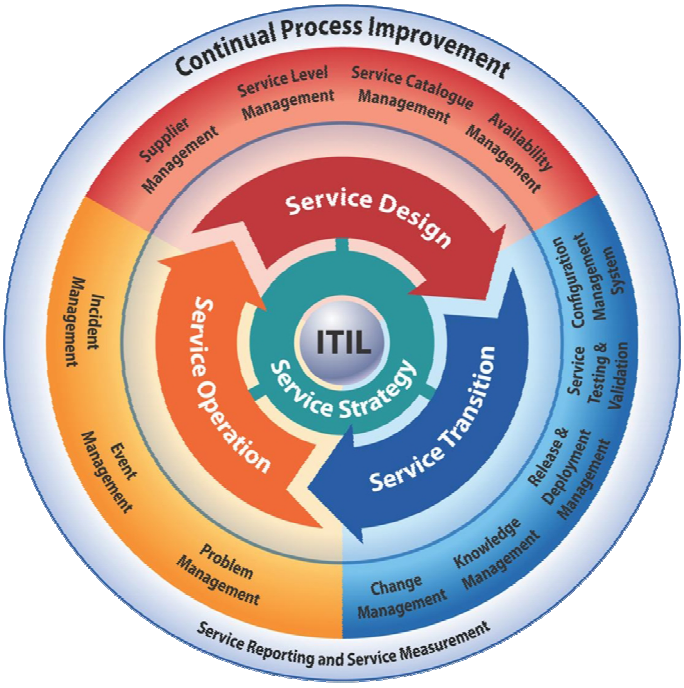


Figura 3. Ciclo de Vida dos Serviços de TI [ITIL 2007, p.19]

A fase Estratégia de Serviço (*Service Estrategy*) capacita a TI no desenvolvimento de suas habilidades necessárias para melhorar e garantir a entrega dos seus serviços aos clientes [ITIL 2007]. O objetivo deste ciclo é compreender qual será a estratégia de serviço a ser adotada e definir como os serviços de TI irão entregar valor para o negócio [FREITAS 2011].

Os processos desta fase são: Gerenciamento da Estratégia para os Serviços de TI, Gerenciamento do Portfólio de Serviço, Gerenciamento Financeiro para os Serviços de TI, Gerenciamento da Demanda e Gerenciamento do Relacionamento com o negócio.

A fase Desenho de Serviço (*Service Design*) realiza o planejamento e desenho dos serviços de TI. O principal objetivo deste ciclo é garantir o alinhamento dos serviços de TI com as necessidades do negócio. [ITIL 2007].

Os processos desta fase são: Coordenação do Desenho, Gerenciamento de Nível de Serviço, Gerenciamento de Catálogo de Serviço, Gerenciamento da Capacidade, Gerenciamento da Disponibilidade, Gerenciamento da Continuidade de Serviço de TI, Gerenciamento de Segurança da Informação e Gerenciamento de Fornecedores.

A fase Transição de Serviço (*Service Transition*) define arquiteturas, processos, políticas e documentação que devem atender aos requisitos identificados na fase Estratégia do Serviço. Seu objetivo é garantir que os serviços novos ou alterados atendam às necessidades do negócio de acordo com as fases de Estratégia e Desenho do Serviço [FREITAS 2011, p. 241].

Os processos desta fase são: Planejamento e Suporte da Transição, Gerenciamento de Mudanças, Gerenciamento de Liberação e Implantação, Gerenciamento da Configuração e Ativo de Serviço, Validação e Testes de Serviço, Avaliação da Mudança, Gerenciamento do Conhecimento.

A fase Operação de Serviço (*Service Operation*) contempla as atividades operacionais do dia a dia de TI necessárias para manter os serviços de TI funcionando de acordo com os níveis de serviço acordados [FREITAS 2011, p. 295]. Seu objetivo é manter a satisfação dos usuários dos serviços.

Os processos desta fase são: Gerenciamento de Eventos, Gerenciamento de Incidentes, Cumprimento de Requisição, Gerenciamento de Problemas e Gerenciamento de Acesso.

Finalmente a fase Melhoria Continuada de Serviço (C*ontinual Process Improvement*) é necessária para a melhoria dos processos e gerenciamento dos serviços, com foco no seu custo. O principal objetivo desta fase é a melhoria contínua (ou continuada) dos serviços de TI de acordo com os requisitos do negócio [FREITAS 2011].

Já o modelo Objetivo de Controle para Tecnologia da Informação e Áreas Relacionadas (COBIT – *Control Objectives for Information and Related Technology*), atualmente em sua versão 5, é um *framework* que contém boas práticas para governança e gestão de TI. Este modelo auxilia as organizações na criação de valor por meio da TI, mantendo o equilíbrio entre a realização de benefícios, a otimização dos níveis de risco e de utilização dos recursos [ISACA, 2012].

A figura 4, a seguir, mostra os cinco princípios definidos pelo COBIT 5:

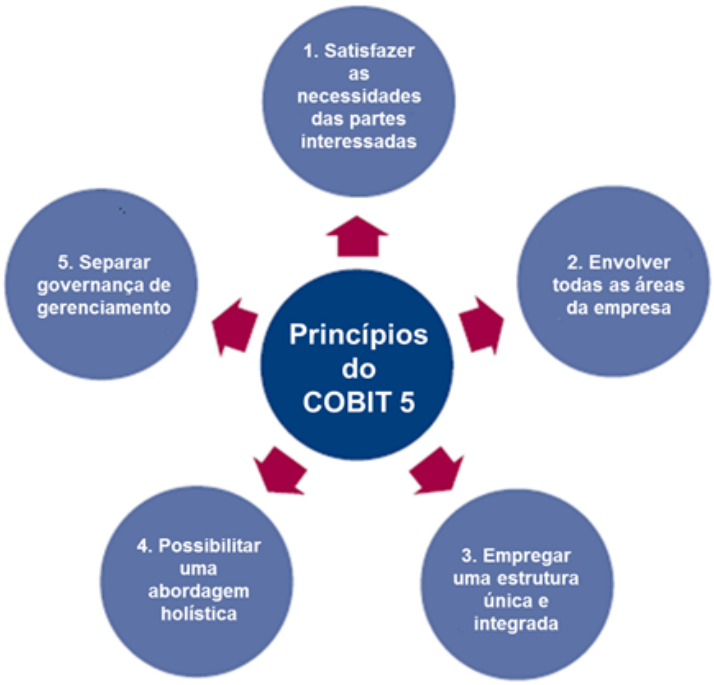


Figura 4. Os cinco princípios do COBIT [ISACA, 2012, p.15]

Atender às necessidades das partes interessadas significa que organizações existem para criar valor para suas partes. Neste sentido o COBIT fornece todos os processos necessários e demais habilitadores para apoiar a criação de valor para a organização com o uso de TI.

Cobrir a organização de ponta a ponta está relacionado à integração da governança corporativa de TI à governança corporativa, através da cobertura de todas as funções e processos corporativos. Isso significa que o COBIT 5 não se concentra somente na "função de TI", mas considera a tecnologia da informação e tecnologias relacionadas como ativos que devem ser tratados como qualquer outro ativo por todos na organização.

Aplicar um modelo único integrado está relacionado à existência de diversas normas e boas práticas relacionadas a TI, cada qual provê orientações para um conjunto específico de atividades de TI. Neste contexto o COBIT se alinha a outros padrões e modelos importantes em um alto nível e, portanto, pode servir como o um modelo unificado para a governança e gestão de TI da organização.

Permitir uma abordagem holística significa que O COBIT define um conjunto de habilitadores para apoiar a implementação de um sistema abrangente de gestão e governança de TI da organização, através de definição de um conjunto de habilitadores, que são definidos como qualquer coisa que possa ajudar a atingir os objetivos corporativos. Os sete habilitadores definidos neste princípio são: Princípios, Políticas e Modelos; Processos; Estruturas Organizacionais; Cultura, Ética e Comportamento; Informação; Serviços, Infraestrutura e Aplicativos; Pessoas, Habilidades e Competências.

Finalmente, distinguir a governança da gestão significa que o modelo do COBIT faz uma clara distinção entre governança e gestão. Essas duas disciplinas compreendem

diferentes tipos de atividades, exigem modelos organizacionais diferenciados e servem a propósitos diferentes.

Além de definir cinco princípios, o COBIT divide os processos de governança e gestão de TI da organização em dois domínios de processos principais: Governança, que contém os processos Avaliar, Dirigir e Monitorar (*Evaluate, Direct and Monitor* - EDM), dentro destes processos são definidos as suas práticas; e Gestão, que contém quatro domínios que estão em conformidade com as áreas responsáveis por planejar, construir, executar e monitorar *(Plan, Build, Run and Monitor* - PBRM)

Os processos de Gestão, por sua vez, são subdivididos em quatro domínios: Alinhar, Planejar e Organizar *(Align, Plan and Organise* – (APO); Construir, Adquirir e Implementar *(Build, Acquire and Implement* – (BAI); Entregar, Serviços e Suporte *(Deliver, Service and Support* - (DSS); Monitorar, Avaliar e Analisar *(Monitor, Evaluate and Assess* – (MEA).

As próximas subseções mostram como as normas ISO/IEC 27001 e 20000, o conjunto de recomendações ITIL e o modelo COBIT definem a GCN como parte de seus processos.

3.1. ISO 27001

A norma ISO 27001 especifica em seu conteúdo o controle A.17, que trata dos aspectos da segurança da informação na gestão da continuidade do negócio. Este item contém os controles necessários para a GCN na visão da área de segurança da informação, conforme pode ser visualizado nas tabelas 1 e 2, a seguir:

Tabela 1: Os objetivos de controle e controles da Continuidade da Segurança da Informação [ABNT 2013, p.27]

A.17 Aspectos da segurança da informação na gestão da continuidade do negócio		
A.17.1 Continuidade da segurança da informação Objetivo: A continuidade da segurança da informação deve ser contemplada nos sistemas de gestão da continuidade do negócio da organização.		
A.17.1.1	Planejando a continuidade da segurança da informação	Controle A organização deve determinar seus requisitos para a segurança da informação e a continuidade da gestão da segurança da informação em situações adversas, por exemplo, durante uma crise ou desastre.
A.17.1.2	Implementando a continuidade da segurança da informação	Controle A organização deve estabelecer, documentar, implementar e manter processos, procedimentos e controles para assegurar o nível requerido de continuidade para a segurança da informação durante uma situação adversa.
A.17.1.3	Verificação, análise crítica e avaliação da continuidade da segurança da informação	Controle A organização deve verificar os controles de continuidade de segurança da informação, estabelecidos e implementados, a intervalos regulares, para garantir que eles são válidos e

		eficazes em situações adversas.

A tabela 1 especifica os aspectos da segurança da informação na gestão da continuidade do negócio. O controle A.17.1 estabelece que a continuidade da segurança da informação deve ser contemplada nos sistemas de gestão da continuidade do negócio da organização e define três sub-itens de controle: A.17.1.1; A.17.1.2; A.17.1.3.

Estes três subitens definem, respectivamente: o processo de planejamento da continuidade da segurança da informação; a implementação da continuidade da segurança da informação; a verificação, análise crítica e avaliação da continuidade da segurança da informação.

Tabela 2: Os objetivos de controle e controles das Redundâncias [ABNT 2013, p.28]

A.17.2 Redundâncias		
Objetivo: Assegurar a disponibilidade dos recursos de processamento da informação		
A.17.2.1	Disponibilidade dos recursos de processamento da informação	Controle Os recursos de processamento devem ser implementados com redundância suficiente para atender aos requisitos de disponibilidade.

A tabela 2 especifica através do subitem A.17.2 o controle de redundâncias, que visa assegurar a disponibilidade dos recursos de processamento da informação. Isso é definido através do item de controle A.17.2.1 que estabelece que os recursos de processamento devem ser implementados com redundância suficiente para atender aos requisitos de disponibilidade.

Acerca do princípio de redundância, a norma também especifica como boa prática que a organização faça o levantamento dos requisitos de negócio levando em consideração a disponibilidade, em maior ou menor nível, de seus sistemas de informação. Neste caso, a norma define que quando a disponibilidade não puder ser assegurada usando a arquitetura de sistemas existentes, convém que componentes redundantes ou novas arquiteturas sejam considerados.

Finalmente, ela recomenda ainda que sejam realizados testes nestes sistemas de informação para assegurar que a transferência de um componente para seu substituto, em casos de ocorrências de falha no primeiro, funcione conforme esperado.

3.2. ISO 20000

A norma ISO 20000 possui um requisito voltado para o Gerenciamento de Continuidade dos Serviços de Tecnologia da Informação (TI) que trata das interrupções inesperadas nos serviços de TI que oferecem suporte ao negócio. Também é especificado neste requisito que, para o tratamento destas interrupções, devem ser mantidos Planos de Contingência e Desastres.

Isto implica que uma empresa que busca se certificar nesta norma, ao aplicar os requisitos tratados, também deve implantar formalmente o processo de GCN contido na norma, visto que o Gerenciamento dos Serviços de TI está contido no escopo da GCN.

O Gerenciamento de Continuidade dos Serviços de TI é parte do processo geral da organização da GCN, e depende de informação derivada desse processo. O Gerenciamento de Continuidade dos Serviços de TI foca na garantia da continuidade dos serviços de TI para o negócio, incorporando todos os serviços dos quais o negócio depende mais os serviços de TI [MAGALHÃES et al 2007].

A ISO 20000 ainda define que eventos inesperados que tenham impacto na disponibilidade dos serviços críticos devem ser investigados, e ações adequadas devem ser realizadas, buscando sempre a excelência e disponibilidade dos serviços prestados.

3.3. ITIL

O conjunto de recomendações ITIL define na fase Desenho de Serviço o processo de Gerenciamento da Continuidade dos Serviços de TI (SCM – *Service Continuity Management*). Este processo é base para a implementação da GCN e seu objetivo é suportar o PCN, principalmente através do gerenciamento de riscos que possam afetar os serviços de TI [FREITAS 2011, p. 204].

O ITIL define a continuidade como a capacidade da empresa em prever eventos de riscos que possam afetar seriamente o negócio, e com a habilidade da empresa em estar preparada para reagir caso esses eventos ocorram (resiliência).

Para que o gerenciamento da continuidade ocorra de fato é necessária a criação de um PCN, que é o documento resultante de um planejamento maior de continuidade do negócio, voltado para toda empresa, que descreve, dentre outros itens, como TI responde por serviços críticos de negócio que são entregues através de serviços de TI. Além disso, também é necessário que haja um Plano de Continuidade de Serviços de TI alinhado ao Plano de Continuidade do Negócio, já que é a área de negócio que deve definir quais os serviços críticos para o negócio e qual deve ser a disponibilidade esperada.

É importante ainda que o Gerenciamento da Continuidade esteja alinhando ao processo de processo de Gerenciamento de Disponibilidade, que tem como objetivo garantir que os níveis de disponibilidade entregues para os serviços de TI atendam às necessidades de disponibilidade acordadas nos Acordos de Nível de Serviço (ANS).

Um ANS é acordo firmado entre o Provedor de Serviço de TI e um cliente onde é descrito o Serviço de TI, as Metas de Nível de Serviço e a especificação das responsabilidades do Provedor de Serviços de TI e do cliente.

Como atividade proativa, o Gerenciamento de Disponibilidade prevê a realização de análises de falhas e riscos de disponibilidade dos serviços e componentes de TI, e tem como objetivo principal garantir que a duração e os impactos causados por falhas na disponibilidade destes serviços de TI sejam minimizados, se relacionando também com o processo de Gerenciamento de Incidentes, cujo objetivo é restaurar a operação normal dos serviços o mais rápido possível para minimizar o impacto no negócio, garantindo que os níveis de qualidade de serviço acordados sejam mantidos.

O processo de Gerenciamento de Mudanças, que faz parte da fase Transição de Serviços, também possui relação com GCN, uma vez que este processo permite que as mudanças sejam registradas, avaliadas, autorizadas, priorizadas, planejadas, testadas, implementadas, documentadas e revisadas de maneira controlada. Neste contexto, todas

as mudanças realizadas no ambiente de TI precisam passar por um processo de planejamento, visando justamente minimizar os impactos na continuidade dos negócios que estas podem causar.

Já na fase Operação de Serviço, o principal processo que se relaciona com a GCN é o Gerenciamento de Incidentes, que é o processo responsável pelo tratamento e pela resolução de todos os incidentes observados nos serviços de TI.

Os incidentes podem receber classificações de acordo com o grau de impacto que podem causar ao negócio, definindo sua prioridade de tratamento de acordo com a sua classificação. Isto quer dizer que incidentes classificados como graves, por exemplo, devem ser atendidos com urgência, o que minimiza a probabilidade de ocorrências relacionadas à GCN.

Além disso, caso haja algum incidente em algum serviço que tenha sido mapeado como crítico ao negócio há a necessidade de utilizar planos de recuperação de desastre, que devem ser documentados no processo de Gerenciamento de Continuidade dos Serviços de TI.

3.4. COBIT

No modelo COBIT [ISACA, 2012], a implantação da Continuidade do Negócio é tratada através do domínio "Entregar, Atender e dar Suporte" (DSS – *Deliver, Service and Support*), no quarto processo, Gerenciar Continuidade (DSS04 – *Manage Continuity*), conforme pode ser visualizado na figura 5.

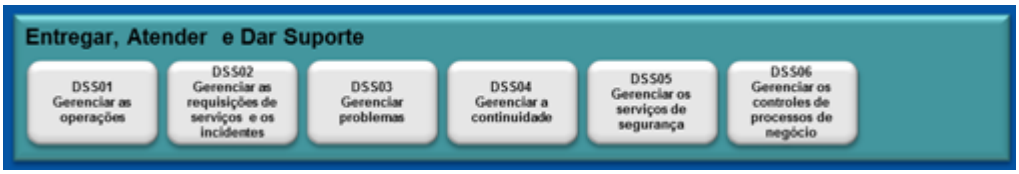


Figura 5: Modelo do Processo DSS04

O objetivo do processo DSS04 é estabelecer e manter um plano para permitir que o negócio e a TI possam responder aos incidentes e interrupções não programadas dos serviços, visando garantir a continuidade da operação de processos de negócios críticos de TI necessários, mantendo a disponibilidade de informações em um nível aceitável para a empresa.

Também trata da entrega dos serviços de TI necessários para atender ao planejamento estratégico, incluindo processos para gerenciar operações, requisições de serviços e incidentes, gerenciamento de problemas, segurança e controle de processos de negócio, além do gerenciamento de continuidade. Seus subprocessos são descritos na tabela 3, a seguir:

Tabela 3: Subprocessos DSS04

Subprocesso	**Descrição**
Definir os Objetivos e o Escopo da Política de Continuidade do Negócio (DSS 04.01 – *Define the Business Continuity Policy Objectives and Scope*)	Tem como premissa a definição da política de continuidade de negócios e o escopo que deve ser alinhado com os objetivos da empresa e partes interessadas.

Manter uma Estratégia de Continuidade (DSS 04.02 – *Maintain a Continuity Strategy*)	Avalia as estratégias existentes para o gerenciamento da continuidade do negócio para definir qual será a estratégia adotada, levando em consideração a viabilidade e custo da mesma, de forma que haja a garantia de recuperação da empresa e a continuidade de suas operações diante de um incidente, interrupção não programada ou desastre.
Desenvolver e Implementar uma Resposta de Continuidade do Negócio (DSS 04.03 – *Develop and Implement a Business Continuity Response*)	Tem como objetivo a documentação de um PCN efetivo para a Organização, capacitando-a continuar com as suas atividades críticas sempre que houver um incidente.
Exercitar, Testar e Rever a PCN (DSS 04.04 – *Exercise, Test and Review the BCP*)	Prevê a execução de testes de continuidade pra exercitar a eficiência dos de continuidade da empresa, desenvolvendo recomendações de melhoria a partir da análise de seus resultados.
Revisar, Manter e Melhorar o Plano de Continuidade (DSS 04.05 – *Review, Maintain and Improve the Continuity Plan*)	Define a realização de uma revisão periódica do plano de continuidade da empresa, para garantir que ele seja eficaz ao ser acionado. As alterações no plano devem ser gerenciadas em acordo com o processo de controle de mudanças, para garantir que o plano de continuidade seja atualizado sempre refletindo os requisitos reais de negócio.
Conduzir Treinamento do Plano de Continuidade (DSS 04.06 – *Conduct Continuity Plan Training*)	Recomenda que todas as partes envolvidas no processo de continuidade devem ser capacitadas, através de treinamentos periódicos.
Gerenciar Arranjos de Backup (DSS 04.07 – *Manage Backup Arrangements*)	Gerenciar a disponibilidade dos recursos e informações críticas de negócio.
Conduzir Revisão de Pós-Reinício (DSS 04.08 – *Conduct Post-Resumption Review*)	Define que o PCN deve ser avaliado após a recuperação de um desastre, garantindo a sua eficiência.

De acordo com Gomes [2016], o COBIT inclui todos os processos e subprocessos DSS04 alinhados com o restante dos processos de gestão e governança da TI da organização, garantindo assim a GCN.

4. Conclusão

Este trabalho realizou o levantamento e estudo de normas, recomendações e modelos que contribuem para a implantação da Continuidade do Negócio nas empresas, mostrando a relação da GCN com os seus processos.

Com a TI assumindo um papel fundamental no ambiente de negócio e também na competitividade das organizações, é preciso garantir que os recursos da TI necessários para o andamento das atividades do negócio estejam disponíveis ou que a sua indisponibilidade cause um impacto mínimo. Isto pode ser possível se a TI implementar a GCN com um processo.

Além disso, as normas, recomendações e modelos, mostram como pode ser feita a implementação da GCN na TI. Afinal, qualquer empresa está sujeita a interrupções não programadas de seus serviços de TI e situações adversas que podem causar

prejuízos significativos, que podem ser minimizados, ou até mesmo evitados, com um PCN bem elaborado e estruturado.

Referências

ABNT, Associação Brasileira de Normas Técnicas. Normalização - Definição. Disponível em http://www.abnt.org.br/normalizacao/o-que-e/o-que-e. Acessado em Set/2017.

ABNT, Associação Brasileira de Normas Técnicas. ABNT NBR 22301: Segurança da sociedade — Sistema de gestão de continuidade de negócios — Requisitos. São Paulo (2013).

ABNT, Associação Brasileira de Normas Técnicas. ABNT NBR 22313: Segurança da sociedade — Sistema de gestão de continuidade de negócios — Orientações. São Paulo (2015).

ABNT, Associação Brasileira de Normas Técnicas. ABNT NBR 27001: Tecnologia da informação — Técnicas de segurança — Sistemas de gestão da segurança da informação — Requisitos. Rio de Janeiro (2013).

ABNT, Associação Brasileira de Normas Técnicas. ABNT NBR 27002: Tecnologia da informação — Técnicas de segurança — Código de prática para controles de segurança da informação. Rio de Janeiro (2013).

ABNT, Associação Brasileira de Normas Técnicas. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS (ABNT). ABNT NBR 20000: Tecnologia da informação — Gestão de serviços Parte 1: Requisitos do sistema de gestão de serviços. São Paulo (2011).

FREITAS, Marcos André dos Santos. Fundamentos do Gerenciamento de Serviços de TI. São Paulo: Brasport (2011).

GOMES, Pedro Emanuel de Campos Copeto. Modelação da Arquitetura Empresarial no Planejamento da Continuidade. Lisboa (2016). Disponível em: https://repositorioaberto.uab.pt/bitstream/10400.2/6271/1/TMISE_PedroGomes.pdf. Acessado em Set/2017.

GUINDANI, Alexandre. Gestão da continuidade dos negócios. São Paulo (2008).

ISACA. COBIT 5: Modelo Corporativo para Governança e Gestão de TI da Organização. Brasília (2012).

ISACA. COBIT 5: Habilitando Processos. Brasília (2012).

ITIL. The Official Introduction to the ITIL Service Lifecycle. London (2007).

ITIL. Service Strategy. London (2007).

ITIL. Service Design. London (2007).

MAGALHÃES, Ivan Luizio e PINHEIRO, Walfrido Brito. Gerenciamento de Serviços na Prática: uma abordagem com base na ITIL. São Paulo: Novatec (2007).

Modelagem ontológica de informações meteorológicas para aplicações de IoT

Murilo Barcelos da Silva, Renato de Freitas Bulcão Neto[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74.690-900 – Goiânia – GO – Brazil

murilobarcelos95@gmail.com, renato@inf.ufg.br

Abstract. *The Internet of Things (IoT) is an emergent computing paradigm in which a plenty of sensor-enabled things (e.g. smartphones) are capable of interacting among themselves towards a common goal. In most IoT scenarios, those things should be capable of exchange a diverse type of information under a common understanding, which is commonly known as semantic interoperability. This paper presents the modeling of meteorological information (atmospheric pressure, humidity, precipitation, temperature and wind speed) in an interoperable and expressive way using ontologies. A simple case study is used to show the ontological representation of real information from meteorological stations of several Brazilian regions. The main contribution is a semantic interoperability solution for IoT applications that make use of meteorological information.*

Resumo. *A Internet das Coisas (*Internet of Things - *IoT) prevê a proliferação de objetos simples do cotidiano (p.ex. smartphones) munidos de sensores capazes de interagir entre si para atingirem objetivos em comum. Logo, é de grande importância que informações obtidas de sensores sejam interoperáveis, para que quando dados de sensores forem trocados entre aplicações, esses tenham um significado não ambíguo para ambas aplicações, sendo interpretados de forma que tenham semântica única. Nesse sentido, este trabalho propõe serviços de representação baseada em ontologias para informações meteorológicas, como pressão atmosférica, umidade relativa do ar, precipitação pluviométrica, temperatura ambiente e velocidade do vento. Um simples estudo de caso foi utilizado com informações reais de estações meteorológicas de várias regiões do Brasil com o objetivo de demonstrar como informações meteorológicas podem ser modeladas de forma interoperável. A principal contribuição deste trabalho é uma solução de interoperabilidade semântica de informações meteorológicas para cenários de IoT.*

1. Introdução

Uma empresa de monitoração de ambientes aviários estuda e acompanha o ambiente que aves habitam para que possa identificar as condições ao longo do crescimento das aves, melhorar a qualidade de vida desses animais e, consequentemente, aumentar a produtividade destas. Para que esse processo seja realizado, um profissional da empresa deve visitar semanalmente o ambiente no qual as aves habitam, colher informações necessárias para poder realizar um estudo, e assim chegar a uma solução satisfatória para um dado cliente. Entretanto, ao longo do tempo identificou-se que o sistema de visitas semanais não

conseguia produzir de forma eficaz um estudo de qualidade, por não ser possível coletar tantas informações quanto era necessário.

Para apoiar o processo de coleta de dados, a empresa adotou o Sistema de Monitoramento Aviário (SMA), que comprende uma rede de sensores que adquire informações diretamente dos ambientes aviários (p.ex. umidade relativa do ar e temperatura ambiente). Essas informações coletadas são enviadas na modalidade 24/7 para um segundo sistema, chamado Sistema de Gerenciamento Aviário (SGA), por meio do qual usuários na empresa podem realizar um estudo mais eficaz e eficiente sobre a produtividade aviária.

Em função da importância dos dados fornecidos pela rede de sensores dos ambientes aviários, outros sistemas podem ser desenvolvidos de forma a poderem fazer uso dessa massa de dados, mas, para isso, é salutar que os dados coletados estejam representados de modo que qualquer aplicação a ser integrada a esse ambiente tenha a mesma interpretação desses dados. Este desafio, chamado interoperabilidade semântica [Perera et al. 2014], é recorrente em cenários de Internet das Coisas (do inglês, IoT) como o aqui descrito, em que existem coisas com capacidade de sensoriamento (p.ex. rede de sensores) que fornecem dados a outras coisas (p.ex. SGA) para que provejam um serviço relevante a usuários.

A literatura é unânime ao afirmar que ontologias são a melhor técnica para modelagem de dados obtidos de sensores para aplicações de IoT [Bettini et al. 2010, Perera et al. 2014]. Isto se deve ao fato de ontologias oferecerem construtores formais (p.ex. classes, propriedades, relações, axiomas, etc.) para modelar um domínio de informação com alta expressividade semântica de modo não ambíguo, e também por permitirem o reúso e a extensão de suas definições por outras ontologias.

Para viabilizar o intercâmbio de dados entre aplicações IoT, a literatura também aponta a importância dos sistemas gerenciadores de contexto (do inglês, *Context Management System* - CMS), que devem fornecer um conjunto de serviços comuns a aplicações dessa natureza. Nesse cenário, os autores desenvolveram o CMS *Hermes*, que reúne um conjunto de componentes que fornecem os serviços de modelagem ontológica de dados, agregação, filtragem, interpretação, armazenamento e disseminação de dados entre aplicações de IoT [Veiga et al. 2014, Veiga and Bulcão-Neto 2016, Maranhão and Bulcão-Neto 2016]. O componente de *Hermes* chamado *Hermes Widget* (HW) é essencial para que os demais serviços de *Hermes* e aplicações possam utilizar o contexto de maneira interoperável, recebendo as informações de sensores do ambiente (p.ex. dados meteorológicos do cenário fictício supracitado) e representando-as com base em ontologias.

Entretanto, os componentes HW precisam ser especializados para representar de modo interoperável as informações adquiridas por sensores, assim como o fizeram [Veiga and Bulcão-Neto 2016], ao criarem um HW para cada sinal vital monitorado em UTIs: pressão arterial, temperatura, frequência de pulso, saturação de oxigênio e frequência respiratória.

O objetivo desta pesquisa é demonstrar que dados meteorológicos, importantes em cenários de aplicação de IoT, podem ser modelados de forma consensual e interoperável por meio do reúso de ontologias existentes na literatura, como *IoT-Lite* [Bermudez-Edo et al. 2016] e *SSN* [Compton et al. 2012]. Cinco novos serviços HW são desenvolvidos para a representação ontológica dos seguintes dados meteorológicos:

pressão atmosférica, umidade relativa do ar, precipitação pluviométrica, temperatura ambiente e velocidade do vento. Todos esses serviços tem como base o reúso e a extensão do componente *Hermes Widget* original [Veiga et al. 2016].

Este artigo tem a seguinte estrutura: a Seção 2 apresenta as informações relacionadas ao *Hermes Widget* e as ontologias utilizadas; a Seção 3 apresenta requisitos, arquitetura, implementação e validação dos componentes implementados; a Seção 4 descreve trabalhos relacionados; e a Seção 5 apresenta as conclusões da pesquisa.

2. Fundamentação Teórica

Esta seção apresenta o componente *Hermes Widget* do CMS *Hermes*, bem como as ontologias *SSN*, *QU* e *IoT-Lite* utilizadas neste trabalho.

2.1. Hermes Widget

O *Hermes Widget* (HW), ilustrado na Figura 1, é o componente responsável pelo serviço de representação de dados no CMS *Hermes*. Essencial para atribuir interoperabilidade às informações do ambiente, ele recebe dados de leitura de sensores como entrada, e realiza sua representação com base nas ontologias do domínio. Dessa forma, uma informação que possui inicialmente baixa expressividade e diferentes possibilidades de interpretação, ganha semântica e pode ser utilizada por diferentes aplicações, de maneira uniforme [Veiga et al. 2016].

A arquitetura do HW mostra a ligação entre a camada de representação e os modelos ontológicos, pré-definidos de acordo com o domínio de aplicação. As representações realizadas por este serviço são gerenciadas pelas camadas de *configuração*, *comunicação* e *notificação*, sendo as duas últimas responsáveis por entregar o contexto representado aos demais componentes de *Hermes* (para interpretação, histórico, etc.) e às aplicações.

Figura 1. Arquitetura do *Hermes Widget* [Veiga and Bulcão-Neto 2016].

Para que aplicações e componentes interessados possam receber as informações de contexto representadas por HW, este componente possui um arquivo de configuração, que está definido na camada de nome análogo, e contém os parâmetros para assinatura de tópicos criados para as instâncias do componente.

2.2. Ontologia SSN (*Semantic Sensor Network*)

Com objetivo de promover a integração das informações obtidas de diferentes sensores, e também a interoperabilidade entre sistemas e aplicações, o W3C *Incubator Group* propõe

a ontologia *Semantic Sensor Network* (SSN) [Compton et al. 2012]. A SSN possui construtores que permitem descrever observações, precisão, capacidade, e métodos que os sensores utilizam para detectar o ambiente em que estão inseridos. Como elemento central para representação semântica de dados de sensoriamento, a SSN utiliza o padrão Estímulo-Sensor-Observação (SSO), apresentado na Figura 2.

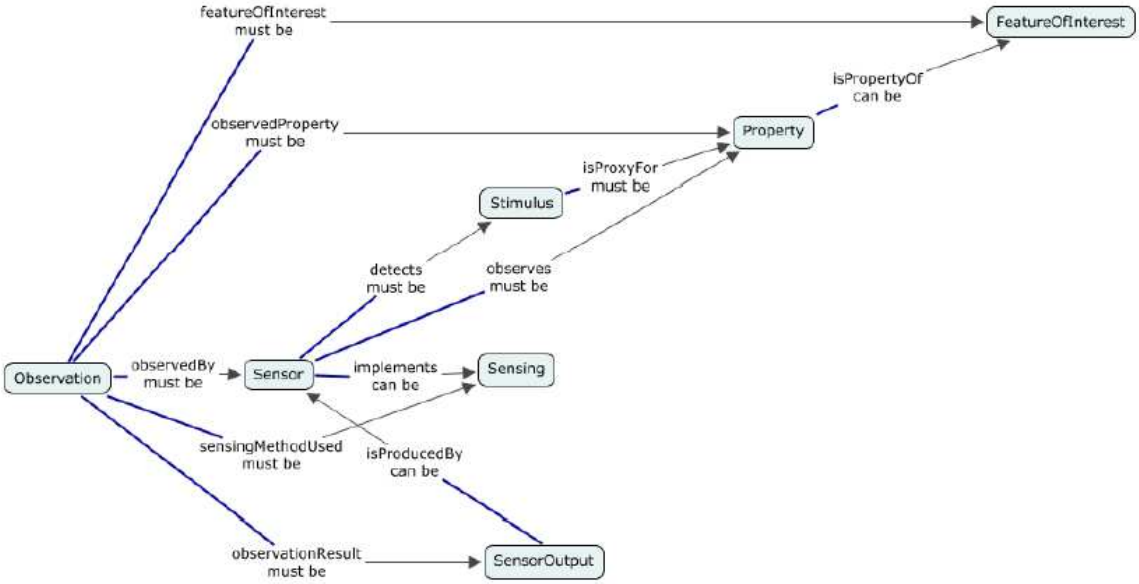


Figura 2. O Padrão Estímulo-Sensor-Observação adotado na ontologia SSN [Compton et al. 2012].

Com o uso do padrão SSO, a ontologia SSN pode ser vista por quatro diferentes perspectivas: *i)* do sensor, com foco sobre o que é sensoreado, como é sensoreado e o que é detectado por ele; *ii)* da observação, com foco em aferições de dados e os metadados correspondentes; *iii)* do sistema, com foco em sistemas de sensores e sua implantação; e *iv)* das características e propriedades de interesse, que o sensor captura [Compton et al. 2012].

Dessa forma, o padrão SSO é o ponto central para a representação e realização das ligações entre o sensor, a propriedade sensoreada, a entidade de interesse, e o dado aferido, abrangendo as perspectivas acima descritas [Compton et al. 2012].

2.3. Ontologia Quantity Kinds and Units (QU)

No nosso dia a dia utilizamos termos de quantidade e frequentemente utilizamos não apenas para definirmos o tipo de quantidade, mas também para definirmos valor e unidade, dimensões e valores, por exemplo uma quantidade de 3 metros.

Com essa necessidade de descrevermos a quantidade e unidade das informações foi proposta uma ontologia também desenvolvida pelo W3C *Incubator Group*, com o objetivo de ter uma ampla base de conhecimento, fornecendo conceitos de tipos de quantidades e unidades de medidas para que representações semânticas tenham um nível de contexto maior [W3C 2010]. As propriedades da ontologia que foram utilizadas são:

- ***qu:QuantityKind***: representa o tipo de medição (p.ex. temperatura ambiente);
- ***qu:Unit***: representa a unidade associada à medição (p.ex. graus Celsius).

2.4. Ontologia IoT-Lite

A ontologia *IoT-Lite* é uma instanciação da ontologia SSN, porém voltada para ambientes IoT [Bermudez-Edo et al. 2016]. Embora ela instancie a SSN, a *IoT-Lite* se consolida como uma ontologia que tem um modelo semântico leve em termos de de representação ontológica e de processamento. Ela adota um modelo simples, representando apenas os conceitos mais utilizados para análise de dados nos sistemas de IoT.

Essa ontologia permite também adicionar diferentes módulos semânticos, caso seja necessário utilizar diferentes conceitos e relacionamentos de um domínio específico. Sendo assim, a *IoT-Lite* é uma ontologia totalmente reutilizável e interoperável. Também é feito o uso de um vocabulário com termos conhecidos utilizados por ontologias de destaque para ser possível compartilhar e combinar dados de diferentes fontes [Bermudez-Edo et al. 2016]. A Figura 3 apresenta as principais propriedades e relacionamentos entre essas classes.

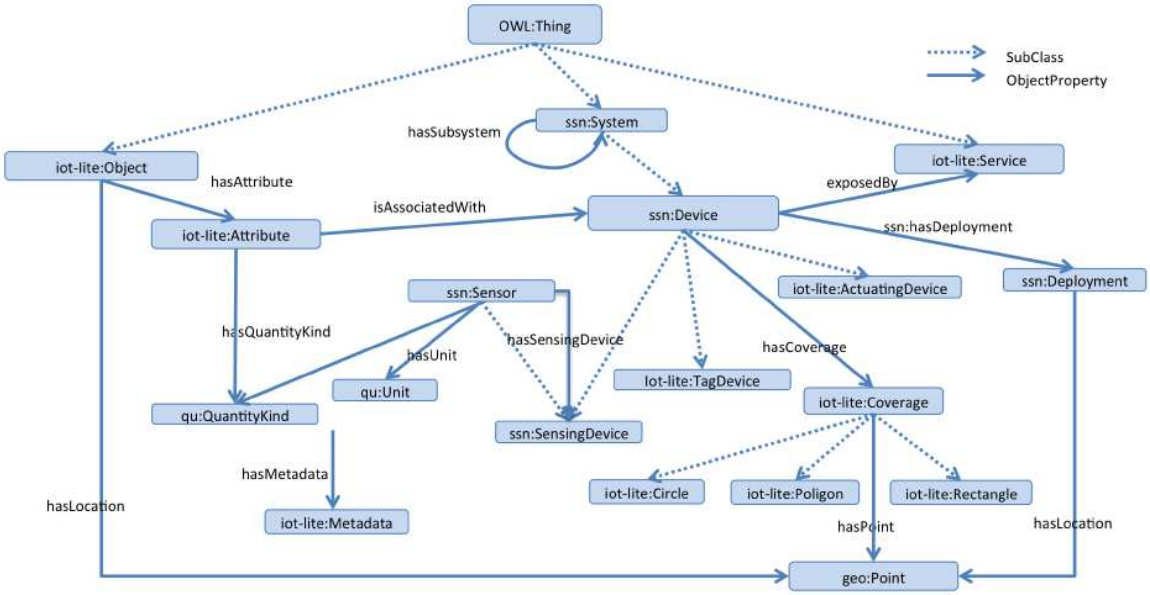


Figura 3. Classes e relações da ontologia *IoT-Lite* [Bermudez-Edo et al. 2016].

3. Representação semântica de dados meteorológicos

Esta seção apresenta os requisitos necessários para o desenvolvimento do trabalho, a arquitetura utilizada, a implementação dos serviços de representação ontológica de dados meteorológicos e a validação dessas representações.

3.1. Requisitos

Os HWs desenvolvidos são baseados no *Hermes Widget* original, portanto os requisitos utilizados para a construção das representações são os mesmos utilizados no componente original [Veiga and Bulcão-Neto 2016]: permitir a especificação de um sensor e do processo de sensoreamento, a representação das informações sensoreadas utilizando ontologias, a padronização do modelo de representação independente de domínio de aplicação, e a disseminação de informações aos demais componentes do CMS *Hermes* ou a aplicações interessadas.

Os requisitos supracitados são comuns a todos os HWs desenvolvidos. Complementando esses requisitos, temos que cada HW desenvolvido tem um requisito específico, quanto à representação ontológica de dados meteorológicos:

- **HW AtmosphericPressure**: representação ontológica de pressão atmosférica;
- **HW Humidity**: representação ontológica de umidade do ar;
- **HW Precipitation**: representação ontológica de precipitação pluviométrica;
- **HW Temperature**: representação ontológica de temperatura; e
- **HW WindSpeed**: representação ontológica de velocidade do vento.

3.2. Arquitetura

A Figura 4 apresenta a arquitetura dos HWs desenvolvidos neste trabalho. Têm-se que as ontologias ligadas à camada de representação que compõem os modelos ontológicos são *SSN*, *IoT-Lite* e *QU*. Utilizando essas ontologias serão gerados dados com semântica e interoperabilidade. Os demais serviços fornecidos, configuração, comunicação e notificação, são todos herdados do projeto original do *Hermes Widget*.

É importante destacar que, embora todos os HWs utilizem a mesma estrutura, não quer dizer que suas implementações são iguais. Temos que cada HW gerado tem particularidades em sua implementação. Essas particularidades são exploradas na ontologia *QU*, que a partir dessa ontologia é possível representar dados meteorológicos diferentes, podendo ser modificados tipos de medição e unidades na implementação de cada HW com o uso dessa ontologia.

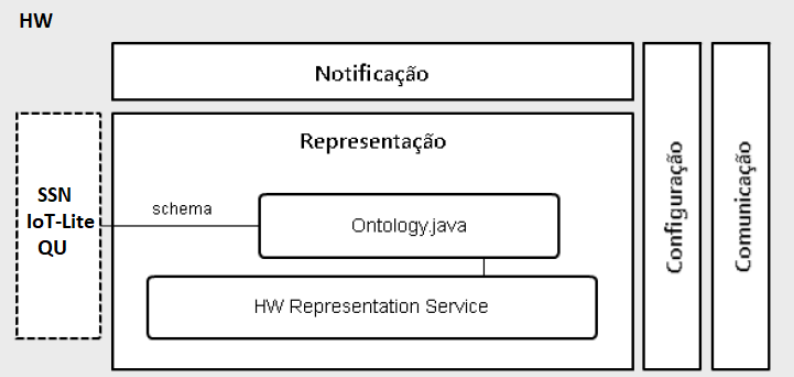


Figura 4. Arquitetura comum utilizada pelos HWs de dados meteorológicos.

3.3. Representação ontológica de dados meteorológicos

Os serviços desenvolvidos neste trabalho incluem a representação ontológica dos seguintes dados meteorológicos: pressão atmosférica, umidade do ar, precipitação pluviométrica, temperatura e velocidade do vento.

Todos esses serviços são baseados na ontologia *IoT-Lite*, cujos conceitos e relacionamentos são complementados pelas ontologias *SSN* e *QU*. Vale destacar que embora a ontologia *IoT-Lite* forneça vários conceitos básicos de modelagem para a criação dessas representações, podendo ser considerada a ontologia principal nesse contexto, não devemos deixar de lado a expressividade fornecida pelas demais ontologias. Como as demais

ontologias são instanciações da *IoT-Lite*, as representações realizadas neste trabalho não teriam a mesma expressividade sem as demais ontologias utilizadas.

É importante destacar que, como ilustrado na Figura 3, a ontologia *IoT-Lite* é repleta de propriedades e relacionamentos, que permitem descrever sensores e relacionamentos entre esses sensores. Essas propriedades e relacionamentos fazem uma descrição completa de um sensor que pertence ao mundo da IoT.

As propriedades e relacionamentos aplicados sobre os dados meteorológicos foram para que aplicações IoT recebam esses dados e consigam reconhecer que se trata de informações meteorológicas de determinado tipo de medição e unidade. Foram utilizadas as seguintes características básicas para isso: (i) os dados são transmitidos por um sensor; (ii) esses dados têm um tipo de medição; e (iii) há uma unidade relacionada a esses dados.

Há também uma descrição do atributo sensoreado associado a sua medição, assim como um *web service* onde estão expostos os dados aferidos representados e uma interface de como realizar o acesso a esses dados. Dessa maneira, esse tipo de representação atende à ideia do HW, gerando uma representação formal das informações de contexto tratada.

Para complementar a formalidade que deve existir nas informações, é necessária também a interoperabilidade semântica. Nesse sentido, os dados são modelados e gerados no formato *Resource Description Framework* (RDF), que fornece uma estrutura de grafos padrão para representação de informações na Web Semântica. Com o RDF é possível criar uma identificação única para entidades e relacionamentos entre essas entidades. Relações das quais são denominadas triplas, onde cada tripla é composta de um sujeito, um predicado e um objeto.

Para cada um dos serviços a serem apresentados a seguir, serão descritas nas listagens 1 a 5 as respectivas representações semânticas de dados meteorológicos, nesta ordem: pressão atmosférica, umidade relativa do ar, precipitação pluviométrica, temperatura ambiente e velocidade do vento. Todas essas representações no formato de triplas RDF seguem uma mesma estrutura e, portanto, podem ser descritas de forma generalizada da seguinte maneira:

- **Linhas 6 a 8**: o dispositivo é modelado como um sensor de detecção. E o serviço de medição para cada tipo de dado meteorológico (pressão atmosférica, umidade relativa do ar, precipitação pluviométrica, temperatura ambiente e velocidade do vento) é exposto em um web service particular.
- **Linhas 10 a 14**: as informações obtidas pelo sensor são obtidas do dispositivo de detecção. Esse sensor produz os dados meteorológicos com as respectivas unidades (milibar, percentual, milímetro, grau Celsius e metro por segundo), modelados de acordo com as propriedades da ontologia *QU*.
- **Linhas 16 a 19**: o valor medido pelo sensor é modelado como um *iot-lite:Attribute* e está associado ao dispositivo de medição de dado meteorológico, o qual gera os respectivos dados.
- **Linhas 21 a 25**: o serviço de medição de dado meteorológico é configurado para acesso.

3.3.1. Pressão atmosférica

A listagem 1 descreve a representação da estrutura de triplas RDF para o sensor e os respectivos dados de pressão atmosférica produzidos pelo componente **HW AtmosphericPressure**. A unidade adotada para representar dados de pressão atmosférica é o milibar.

```
###           LISTAGEM 1 - Representação RDF de Pressão Atmosférica           ###
1  @prefix qu: <http://purl.org/NET/ssnx/qu/qu#> .
2  @prefix ssn: <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#> .
3  @prefix iot-lite: <http://purl.oclc.org/NET/UNIS/fiware/iot-lite#> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5
6  <http://meteorep/atmospheric_pressure/sensingDevice>  a  ssn:SensingDevice ;
7        iot-lite:exposedBy
8                 <http://meteorep/atmospheric_pressure/webService> .
9
10 <http://meteorep/atmospheric_pressure/sensor>  a  ssn:Sensor ;
11       iot-lite:hasSensingDevice
12                <http://meteorep/atmospheric_pressure/sensingDevice> ;
13       iot-lite:hasQuantityKind  qu:quantity#pressure ;
14       iot-lite:hasUnit  qu:unit#millibar .
15
16 <http://meteorep/atmospheric_pressure/attribute>  a  iot-lite:Attribute ;
17       iot-lite:isAssociatedWith
18                <http://meteorep/atmospheric_pressure/sensingDevice> ;
19       iot-lite:hasQuantityKind  qu:quantity#pressure> .
20
21 <http://meteorep/atmospheric_pressure/webService>  a  iot-lite:Service ;
22       iot-lite:endpoint
23                <http://meteorep/atmosphericPressure/sensor/data> ;
24       iot-lite:interfaceDescription
25                <http://meteorep/atmosphericPressure/sensor/interface> .
```

3.3.2. Umidade relativa do ar

A listagem 2 descreve a representação na estrutura de triplas RDF para o sensor e os respectivos dados de umidade relativa do ar produzidos pelo componente **HW Humidity**. A unidade adotada para representar os dados de umidade do ar é o percentual.

```
###           LISTAGEM 2 - Representação RDF de Umidade Relativa do Ar           ###
1  @prefix qu: <http://purl.org/NET/ssnx/qu/qu#> .
2  @prefix ssn: <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#> .
3  @prefix iot-lite: <http://purl.oclc.org/NET/UNIS/fiware/iot-lite#> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5
6  <http://meteorep/humidity/sensingDevice>  a  ssn:SensingDevice ;
7        iot-lite:exposedBy
8                 <http://meteorep/humidity/webService> .
9
10 <http://meteorep/humidity/sensor>  a  ssn:Sensor ;
11       iot-lite:hasSensingDevice
12                <http://meteorep/humidity/sensingDevice> ;
13       iot-lite:hasQuantityKind  qu:quantity#humidity ;
14       iot-lite:hasUnit  qu:unit#percent .
15
16 <http://meteorep/humidity/attribute>  a  iot-lite:Attribute ;
17       iot-lite:isAssociatedWith
18                <http://meteorep/humidity/sensingDevice> ;
19       iot-lite:hasQuantityKind  qu:quantity#humidity> .
20
21 <http://meteorep/humidity/webService>  a  iot-lite:Service ;
22       iot-lite:endpoint
23                <http://meteorep/humidity/sensor/data> ;
24       iot-lite:interfaceDescription
25                <http://meteorep/humidity/sensor/interface> .
```

3.3.3. Precipitação pluviométrica

Na listagem 3 são descritos a representação na estrutura de triplas RDF para o sensor e os respectivos dados de precipitação pluviométrica produzidos pelo componente **HW Precipitation**. A unidade para dados de precipitação pluviométrica é o milímetro.

```
###          LISTAGEM 3 - Representação RDF de Precipitação Pluviométrica          ###
1  @prefix qu: <http://purl.org/NET/ssnx/qu/qu#> .
2  @prefix ssn: <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#> .
3  @prefix iot-lite: <http://purl.oclc.org/NET/UNIS/fiware/iot-lite#> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5
6  <http://meteorep/precipitation/sensingDevice>  a  ssn:SensingDevice ;
7        iot-lite:exposedBy
8                 <http://meteorep/precipitation/webService> .
9
10 <http://meteorep/precipitation/sensor>  a  ssn:Sensor ;
11       iot-lite:hasSensingDevice
12                <http://meteorep/precipitation/sensingDevice> ;
13       iot-lite:hasQuantityKind  qu:quantity#precipitation ;
14       iot-lite:hasUnit  qu:unit#millimeter .
15
16 <http://meteorep/precipitation/attribute>  a  iot-lite:Attribute ;
17       iot-lite:isAssociatedWith
18                <http://meteorep/precipitation/sensingDevice> ;
19       iot-lite:hasQuantityKind  qu:quantity#precipitation> .
20
21 <http://meteorep/precipitation/webService>  a  iot-lite:Service ;
22       iot-lite:endpoint
23                <http://meteorep/precipitation/sensor/data> ;
24       iot-lite:interfaceDescription
25                <http://meteorep/precipitation/sensor/interface> .
```

3.3.4. Temperatura ambiente

A listagem 4 descreve a representação na estrutura de triplas RDF para o sensor e os respectivos dados de temperatura ambiente produzidos pelo componente **HW Temperature**. A unidade adotada para representar os dados de temperatura ambiente é o grau Celsius.

```
###          LISTAGEM 4 - Representação RDF de Temperatura Ambiente          ###
1  @prefix qu: <http://purl.org/NET/ssnx/qu/qu#> .
2  @prefix ssn: <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#> .
3  @prefix iot-lite: <http://purl.oclc.org/NET/UNIS/fiware/iot-lite#> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5
6  <http://meteorep/temperature/sensingDevice>  a  ssn:SensingDevice ;
7        iot-lite:exposedBy
8                 <http://meteorep/temperature/webService> .
9
10 <http://meteorep/temperature/sensor>  a  ssn:Sensor ;
11       iot-lite:hasSensingDevice
12                <http://meteorep/temperature/sensingDevice> ;
13       iot-lite:hasQuantityKind  qu:quantity#temperature ;
14       iot-lite:hasUnit  qu:unit#degree_Celsius .
15
16 <http://meteorep/temperature/attribute>  a  iot-lite:Attribute ;
17       iot-lite:isAssociatedWith
18                <http://meteorep/temperature/sensingDevice> ;
19       iot-lite:hasQuantityKind  qu:quantity#temperature> .
20
21 <http://meteorep/temperature/webService>  a  iot-lite:Service ;
22       iot-lite:endpoint
23                <http://meteorep/temperature/sensor/data> ;
24       iot-lite:interfaceDescription
25                <http://meteorep/temperature/sensor/interface> .
```

3.3.5. Velocidade do vento

A listagem 5 descreve a representação na estrutura de triplas RDF para o sensor e os respectivos dados de temperatura ambiente produzidos pelo componente **HW WindSpeed**. A unidade adotada para representar os dados de temperatura ambiente é o metro por segundo (m/s).

```
###          LISTAGEM 5 - Representação RDF de Velocidade do Vento          ###
1  @prefix qu: <http://purl.org/NET/ssnx/qu/qu#> .
2  @prefix ssn: <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn#> .
3  @prefix iot-lite: <http://purl.oclc.org/NET/UNIS/fiware/iot-lite#> .
4  @prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
5
6  <http://meteorep/wind_speed/sensingDevice>  a  ssn:SensingDevice ;
7        iot-lite:exposedBy
8                 <http://meteorep/wind_speed/webService> .
9
10 <http://meteorep/wind_speed/sensor>  a  ssn:Sensor ;
11       iot-lite:hasSensingDevice
12                <http://meteorep/wind_speed/sensingDevice> ;
13       iot-lite:hasQuantityKind  qu:quantity#wind_speed ;
14       iot-lite:hasUnit  qu:unit#meter_per_second .
15
16 <http://meteorep/wind_speed/attribute>  a  iot-lite:Attribute ;
17       iot-lite:isAssociatedWith
18                <http://meteorep/wind_speed/sensingDevice> ;
19       iot-lite:hasQuantityKind  qu:quantity#wind_speed> .
20
21 <http://meteorep/wind_speed/webService>  a  iot-lite:Service ;
22       iot-lite:endpoint
23                <http://meteorep/wind_speed/sensor/data> ;
24       iot-lite:interfaceDescription
25                <http://meteorep/wind_speed/sensor/interface> .
```

3.4. Validação dos Serviços

Uma vez que a representações ontológicas dos dados meteorológicos encontram-se no formato RDF, é necessária uma ferramenta que manipule esses dados. Nesse sentido, foi utilizado o *framework Apache Jena*, que fornece uma API para a criação de informações instanciadas de ontologias, como as que foram utilizadas nos modelos de representação de dados meteorológicos.

Por exemplo, existe uma classe Java chamada *Model*, que cria um grafo vazio no qual as informações referentes a um determinado sensor são inseridas de acordo com a estrutura e semântica das ontologias utilizadas, expressas no formato de triplas RDF. Tal como representadas por cada componente HW de dados meteorológicos, essas informações estão prontas para serem disseminadas para outros componentes do CMS (p.ex. armazenamento) ou aplicações interessadas nessas informações.

O objetivo dessa validação é representar dados reais de acordo com toda uma estrutura e semântica baseado nas ontologias escolhidas e utilizadas neste trabalho. Tal que, nesse caso foram utilizadas ontologias que se adequem as características apresentadas por sensores de dados meteorológicos.

Para realizar a validação de cada serviço de representação semântica apresentado anteriormente, foram utilizados dados de estações meteorológicas convencionais coletados do Banco de Dados Meteorológicos para Ensino e Pesquisa (BDMEP)[1], que registra aferições de várias regiões do Brasil.

[1]Disponível em `www.inmet.gov.br/projetos/rede/pesquisa/`, acesso em 29/09/2017.

Inicialmente foram construídas classes que geram as representações semânticas referente a cada tipo de sensor meteorológico. Após isso, é disponibilizado cada representação semântica. Além disso é disponibilizado um *web service* para cada sensor. Sendo que através desses *web services*, qualquer aplicação será capaz de acessar a atual medição(*endpoint*) de um sensor, ou seja, o *web service* lê o próximo valor de um banco de dados e o apresenta através de um *endpoint*. E com uso da representação semântica gerada, qualquer aplicação será capaz também de compreender o dado obtido do *web service*.

Com esse método utilizado, a representação ontológica, que é um processo caro computacionalmente, só sera realizada uma única vez, em vez de realizar uma representação ontológica para cada dado aferido. Logo, como para um mesmo sensor, a única coisa que muda de uma representação para outra é somente o valor medido, é vantajoso realizar toda representação ontológica uma única vez e o valor medido poder ser disponibilizado na web, através de um *web service*.

Para validar o conhecimento representado foram utilizadas consultas na linguagem SPARQL (*Protocol and RDF Query Language*) [W3C 2008]. Segue uma consulta SPARQL relacionada à questão "Qual o tipo de quantidade está associado a um determinado valor aferido de um *endpoint*, em que tal se encontra em um *web service* de um determinado sensor? ". Sendo que tipo de quantidade é o tipo de medição de um sensor, ou seja, nesse contexto pode ser temperatura ambiente, velocidade do vento, pressão atmosférica, umidade relativa do ar ou precipitação pluviométrica.

```
1 SELECT ?sensingDevice ?webService ?endpoint ?sensor ?quantityKind
2 WHERE
  { ?sensor iot-lite:hasSensingDevice ?sensingDevice .
   ?sensor iot-lite:hasQuantityKind ?quantityKind .
   ?sensingDevice iot-lite:exposedBy ?webService .
   ?webService iot-lite:endpoint ?endpoint .
3 } ORDER BY (?sensor);
```

A linha 1 representa as variáveis que armazenarão o resultado da consulta, neste caso, a identificação de um dispositivo de detecção, de um sensor, de um *web service* e o tipo de quantidade relacionado a um valor aferido de um *endpoint*. Na linha 2 é descrito um conjunto de triplas que deverá fazer o casamento com as triplas de dados. Na linha 3, é aplicado um filtro para ordenar o resultado baseado pelo sensor.

Essa consulta então retornará valores aferidos de *endpoints* com seus respectivos dispositivos de detecção, sensores, *web services* e tipos de quantidade associados a eles. Essa consulta é importante, pois, para aplicações realizarem a troca de dados, é importante saber qual tipo de quantidade está relacionado a uma determinada informação, para que essa informação seja tratada de acordo com sua semântica.

Temos outra consulta SPARQL relacionada com a questão "Qual o tipo de unidade está associado a um determinado valor aferido de um *endpoint*, em que tal se encontra em um *web service* de um determinado sensor? ". Sendo que o tipo de unidade representa a unidade que está associada a medição, ou seja, nesse contexto pode ser milibar, percentual, milímetro, grau Celsius ou metro por segundos.

A linha 1 representa as variáveis que armazenarão o resultado da consulta, neste caso, a identificação de um dispositivo de detecção, de um sensor, de um *web service* e o tipo de unidade relacionado a um valor aferido de um *endpoint*. Na linha 2 é descrito um

conjunto de triplas que deverá fazer o casamento com as triplas de dados. Na linha 3, é aplicado um filtro para ordenar o resultado baseado pelo sensor.

```
1 SELECT ?endpoint ?webService ?sensingDevice ?sensor ?unit
2 WHERE
  { ?sensor iot-lite:hasSensingDevice ?sensingDevice .
   ?sensor iot-lite:hasUnit ?unit .
   ?sensingDevice iot-lite:exposedBy ?webService .
   ?webService iot-lite:endpoint ?endpoint .
3 } ORDER BY (?sensor);
```

Essa consulta então retornará valores aferidos de *endpoints* com seus respectivos dispositivos de detecção, sensores, *web services* e tipos de unidade associados a eles. Essa consulta é importante pelo fato que uma aplicação que recebe dados de outra aplicação, pode ter a necessidade de saber qual tipo de unidade está relacionada a esses dados, para saber qual tratamento vai ser realizado em cima destes dados.

4. Trabalhos Relacionados

O *Hermes Widget* original é um componente responsável por realizar a representação ontológica de informações e, consequentemente, abstrair o serviço de representação ontológica do informações em aplicações IoT [Veiga and Bulcão-Neto 2016]. A arquitetura do *Hermes Widget* foi construída de tal forma a facilitar o desenvolvimento de novos HWs, específicos para cada tipo de informação utilizada em cenários de IoT. Essa expressividade e interoperabilidade das informações é baseada na ontologia *SSN*, a partir da qual foram construídos HWs específicos para a representação de sinais vitais de pacientes.

Batista Neto et al. realizaram trabalho similar a este ao desenvolverem um serviço de representação ontológica de informações de localização *outdoor* para aplicações IoT [Batista Neto et al. 2016]. A expressividade das informações de localização é obtida por meio do uso das ontologias *SSN* e *WGS84*, que descreve latitude, longitude e altitude. O objetivo do serviço de Batista Neto et al. é permitir que aplicações consigam trocar informações de localização de forma consensual e com expressividade semântica.

Embora haja grande semelhança entre os trabalhos citados, este trabalho se destaca ao propor serviços de representação ontológica para informação meteorológica.

5. Conclusões

Este trabalho apresentou uma solução para modelagem de dados meteorológicos em cenários de IoT por meio de ontologias-padrão, como a *IoT-Lite*, *SSN* e *QU*. Foram desenvolvidos componentes HW para representação semântica de cada um dos tipos de dados citados, a saber: pressão atmosférica, umidade relativa do ar, precipitação pluviométrica, temperatura ambiente e velocidade do vento.

Um simples estudo de caso foi utilizado com informações reais de estações meteorológicas de várias regiões do Brasil com o objetivo de demonstrar como informações meteorológicas podem ser modeladas de forma interoperável. A principal contribuição deste trabalho é uma solução de interoperabilidade semântica de informações meteorológicas para cenários de IoT.

Como trabalho futuro, pretende-se desenvolver aplicações de IoT que façam uso dos HWs desenvolvidos para representação de dados meteorológicos.

Referências

Batista Neto, J. A., Veiga, E. F., and Bulcão-Neto, R. d. F. (2016). Um serviço de representação ontológica de informações de localização para aplicações em Internet das Coisas. In *IX Escola Regional de Goiás (ERI-GO)*, pages 1–14.

Bermudez-Edo, M., Elsaleh, T., Barnaghi, P., and Taylor, K. (2016). IoT-Lite: A lightweight semantic model for the Internet of Things. In *Ubiquitous Intelligence & Computing, Advanced and Trusted Computing, Scalable Computing and Communications, Cloud and Big Data Computing, Internet of People, and Smart World Congress (UIC/ATC/ScalCom/CBDCom/IoP/SmartWorld), 2016 Intl IEEE Conferences*, pages 90–97. IEEE.

Bettini, C., Brdiczka, O., Henricksen, K., Indulska, J., Nicklas, D., Ranganathan, A., and Riboni, D. (2010). A survey of context modelling and reasoning techniques. *Pervasive Mob. Comput.*, 6(2):161–180.

Compton, M., Barnaghi, P., Bermudez, L., García-Castro, R., Corcho, O., Cox, S., Graybeal, J., Hauswirth, M., Henson, C., Herzog, A., et al. (2012). The ssn ontology of the w3c semantic sensor network incubator group. *Web semantics: science, services and agents on the World Wide Web*, 17:25–32.

Maranhão, G. M. e. and Bulcão-Neto, R. d. F. (2016). A semantic filtering mechanism geared towards context dissemination in ubiquitous environments. *J. UCS*, 22(8):1123–1147.

Perera, C., Zaslavsky, A., Christen, P., and Georgakopoulos, D. (2014). Context aware computing for the Internet of Things: A survey. *IEEE Communications Surveys & Tutorials*, 16(1):414–454.

Veiga, E. F. and Bulcão-Neto, R. d. F. (2016). Um serviço de representação ontológica de contexto baseada no padrão de projeto estímulo-sensor-observação. *VIII Simpósio Brasileiro de Computação Ubíqua e Pervasiva (SBCUP)*, pages 1–10.

Veiga, E. F., Maranhão, G. M. e., and Bulcão-Neto, R. d. F. (2014). Apoio ao desenvolvimento de aplicações de tempo real sensíveis a contexto semântico. In *IX Workshop de Teses e Dissertações do XX Simpósio Brasileiro de Sistemas Multimídia e Web (WebMedia)*, pages 1–4.

Veiga, E. F., Maranhão, G. M. e., and Bulcão-Neto, R. d. F. (2016). Development and scalability evaluation of an ontology-based context representation service. *IEEE Latin America Transactions*, 14(3):1372–1379.

W3C (2008). SPARQL query language for RDF. `https://www.w3.org/TR/rdf-sparql-query/`. Último acesso em Maio de 2017.

W3C (2010). Ontology for quantity kinds and units: Units and quantities definitions. `http://purl.oclc.org/NET/ssnx/qu/qu-rec20`. Último acesso em Maio de 2017.

Uma Abordagem para Garantia da Consistência de *Middleware* Específico de Domínio Dirigido a Modelos

Marcelo R. Fortes[1], **Fábio M. Costa**[2]

[1]Programa de Pós-Graduação em Ciência da Computação
Universidade Federal de Goiás (UFG)
Goiânia, GO – Brasil

marcelorfortes@inf.ufg.br

[2]Instituto de informática – Universidade Federal de Goiás (UFG)
Goiânia, GO – Brasil

fmc@inf.ufg.br

Abstract. *In the context of Model Driven Engineering (MDE), model interpreters are used to process models that represent software systems. Model-Driven Domain-Specific Middleware (MD-DSM) is an approach that employs models to construct a type of model interpreter, called domain-specific middleware. In this work, an approach will be presented to systematically verify the consistency of the models used in the construction of this middelware category. In addition, a methodology will be presented to verify the conformity between modeling languages used in the definition of application based in models and the middleware systems constructed to process such applications.*

Resumo. *No contexto de Model Driven Engineering (MDE), model interpreters são utilizados para processar modelos que representam sistemas de software. Model-Driven Domain-Specific Middleware (MD-DSM) é uma abordagem que emprega modelos para construir um tipo de model interpreter, denominado middleware específico de domínio. Neste trabalho, será apresentada uma abordagem para realizar, de forma sistemática, a verificação da consistência de modelos utilizados na construção dessa categoria de middleware. Além disso, será apresentada uma metodologia para verificar a conformidade entre as linguagens de modelagem utilizadas na definição de modelos de aplicações e os sistemas de middleware construídos para processar tais aplicações.*

1. Introdução

Devido aos avanços nas tecnologias de hardware e de rede, a complexidade dos sistemas de software tem aumentado rapidamente. Mesmo com o avanço das linguagens de programação de alto nível e dos ambientes de desenvolvimento, bem como a existência de numerosas bibliotecas de rede (por exemplo, baseadas em sockets TCP/IP), o desenvolvimento de aplicações complexas continua a ser um processo complicado e dispendioso [France and Rumpe 2007].

Atualmente, a maior parte dos sistemas de software é construída utilizando linguagens de programação de propósito geral, que se baseiam em conceitos relacionados à plataforma de implementação e não ao problema a ser resolvido [Selic 2003]. No entanto,

a diferença semântica entre as abstrações oferecidas pelas linguagens de programação e o domínio do problema que se deseja resolver pode ser minimizada por meio do uso de modelos [France and Rumpe 2007]. Exemplo disso é a área de pesquisa denominada Engenharia Orientada a Modelos (*Model-Driven Engineering* - MDE), que tem como objetivo principal a redução da lacuna entre o domínio do problema e o domínio da implementação de software, utilizando tecnologias que realizam, de maneira sistemática, a transformação das abstrações em nível do problema para implementações de software. [Favre 2004, Kent 2002].

Abordagens convencionais de MDE realizam a tradução de modelos, instanciados a partir de DSMLs (*Domain Specific Modeling Language* - DSML), para códigos de linguagens de propósito geral, que posteriormente são compilados e executados. Entretanto, existem abordagens alternativas, que evitam a etapa de transformação de modelos em códigos [Deng et al. 2008, Riehle et al. 2001, Kirshin et al. 2006]. Nesse tipo de abordagem, os próprios mecanismos que executam os modelos das aplicações são responsáveis por executar os serviços requisitados.

As DSMLs que oferecem suporte à execução de modelos são chamadas de Linguagens de Modelagem Específicas de Domínio Interpretadas (*Interpreted Domain Specific Modeling Languages*, i-DSMLs) [Clarke et al. 2012]. As i-DSMLs são executadas por *model interpreters*, chamados de Máquinas Virtuais de Domínio Específico (*Domain-Specific Virtual Machines - DSVMs*).

Apesar de DSVMs conseguirem diminuir o tempo e o custo relacionados à construção de sistemas complexos, ainda necessita-se de grande esforço na implementação das próprias DSVMs, uma vez que elas são tipicamente desenvolvidas utilizando linguagens de programação de propósito geral. Além disso, cada DSVM possui características intrínsecas a seu domínio de aplicação, o que leva à necessidade de construir uma nova DSVM para cada domínio diferente.

Com o intuito de solucionar esse problema, foi proposta uma abordagem, denominada *Model-Driven Domain-Specific Middleware* (MD-DSM) [Costa et al. 2017], que emprega o uso de modelos para construção de sistemas de *middleware* específicos de domínio que possuem suporte para execução de aplicações também baseadas e modelos. Para instanciação dos modelos utilizados na construção desse tipo de *middleware*, MD-DSM utiliza um metamodelo que contém conceitos comuns a diferentes DSVMs, independentes do domínio de aplicação de tais plataformas. A arquitetura de referência adotada pela abordagem MD-DSM é uma generalização da arquitetura em camadas empregada por dois exemplos de DSVM, CVM [Deng et al. 2008] e MGridVM [Allison et al. 2012], composta por quatro camadas: **User Interface**, **Synthesis**, **Controller** e **Broker**.

Como cada camada dessa categoria de *middleware* é instanciada de forma independente, e possivelmente por diferentes especialistas de *middleware*, surge a necessidade de verificação da conformidade entre os modelos utilizados na instanciação de cada camada. Essa verificação é importante visto que, para que uma camada funcione de maneira correta, é necessário garantir que as camadas adjacentes a ela implementem todos os serviços dos quais ela necessita. Outro problema é a conformidade entre uma DSML e um *middleware* construídos para um determinado domínio. Ou seja, dada uma DSML,

é necessário verificar se as construções do *middleware* são adequadas para executar tal DSML.

Assim sendo, o presente trabalho apresenta uma abordagem para verificar a consistência entre modelos utilizados na construção de sistemas de *middleware* específicos de domínio, e a conformidade entre uma DSML e o *middleware* construído para executá-la. Para verificar a consistência entre modelos de *middleware*, propomos o uso de restrições inter modelos, implementadas utilizando uma linguagem denominada EVL (*Epsilon Validation Language*). Para verificar a conformidade entre DSML e *middleware*, propomos uma metodologia a ser seguida pelo especialista responsável pela construção do *middleware*.

O restante deste artigo está estruturado como segue. Na Seção 2, serão apresentados os principais conceitos sobre os quais este trabalho foi desenvolvido. A Seção 3 apresenta a abordagem proposta neste trabalho para verificação da consistência de *middleware* específico de domínio. Na Seção 4, são apresentadas as considerações finais.

2. Fundamentação teórica

2.1. Engenharia Dirigida a Modelos

É cada vez mais recorrente a necessidade de usuários finais construírem suas próprias aplicações, ou personalizar aplicações já existentes para que elas atendam suas necessidades específicas. No entanto, esses usuários não possuem conhecimento técnico necessário para isso. Como alternativa para resolver esse problema tem-se utilizado a abordagem de Engenharia Dirigida por Modelos (*Model-Driven Engeneering*) [Schmidt 2006], que fornece níveis mais altos de abstração para permitir que esses usuários se concentrem no problema, ao invés de se preocuparem com uma maneira específica de solucioná-lo por meio de plataformas tecnológicas que possuem baixo nível de abstração.

Grande parte do sucesso no emprego de abordagens dirigidas por modelos está associada ao uso de Linguagens de Modelagem Específicas de Domínio (*Domain-Specific Modeling Languages*, DSMLs) [Fishwick 2007, Kelly and Tolvanen 2008]. A construção das linguagens de modelagem é realizada por meio da definição de sua sintaxe abstrata, sintaxe concreta, semântica dinâmica e semântica estática. A sintaxe abstrata é a definição dos conceitos de um domínio de problema e seus relacionamentos. A sintaxe concreta é a representação desses conceitos. Normalmente utiliza-se representação gráfica ou textual. A semântica dinâmica atribui significado aos elementos da linguagem e a semântica estática define restrições sobre os elementos representados pela sintaxe abstrata.

A sintaxe das linguagens de modelagem é comumente especificada por metamodelos, que descrevem conceitos em um domínio de problema e seus relacionamentos. Para especificar a semântica dinâmica (comportamento) das linguagens de modelagem ainda não existe um formalismo definido [Bryant et al. 2011]. Para essa finalidade, uma das alternativas é a construção de interpretadores de modelos (*model interpreters*), que agregam comportamento operacional às construções das DSMLs.

Em abordagens tradicionais, os interpretadores de modelos transformam modelos em código fonte de linguagens de propósito geral, que depois é compilado e executado. Contudo, existe um problema com essa abordagem, que é o atraso entre a mudança realizada nos modelos e sua propagação para o sistema.

Como alternativa para solucionar esse problema foram propostas as máquinas de execução de modelos [Riehle et al. 2001, Kirshin et al. 2006], que evitam a etapa de geração de código. Elas capturam o comportamento do sistema por meio de diagramas comportamentais, por exemplo, diagramas de estado ou diagramas de sequência. Além dessas abordagens, estudos vêm sendo feitos para desenvolver máquinas de execução de modelos [Allison et al. 2011, Clarke et al. 2012, Deng et al. 2008], que possuem o comportamento implantado em sua própria construção, o que torna a modelagem do sistema mais simples por não existir a necessidade de definir diferentes modelos para diferentes situações de um mesmo sistema. Isto é, nesse tipo de abordagem, não é necessário instanciar modelos que definem o comportamento do sistema, uma vez que o comportamento é embutido na própria máquina. O único modelo que precisa ser criado é o modelo que define os componentes estruturais da aplicação requisitada.

Esse tipo de máquina é chamada de Máquina Virtual de Domínio Específico (*Domain Specific Virtual Machine* - DSVM). Elas possuem uma arquitetura dividida em camadas, cada uma com um papel específico na execução de modelos. Para construir os modelos que serão interpretados pelas DSVMs, os usuários utilizam um tipo específico de linguagem de modelagem, denominada *Interpreted Domain Specific Modeling language* - i-DSML.

2.2. *Middleware* específico de domínio dirigido a modelos

Como introduzido anteriormente, uma alternativa para diminuir a complexidade relacionada ao desenvolvimento de sistemas de software é a utilização de modelos como artefatos primários ao invés de linguagens de programação de propósito geral. Para viabilizar o uso de modelos é necessário construir linguagens de modelagem específicas de domínio e mecanismos capazes de interpretá-las.

Embora essa abordagem diminua a complexidade e o tempo relacionados ao desenvolvimento de aplicações, a construção dos próprios mecanismos responsáveis pela interpretação dos modelos é uma atividade que envolve grande esforço de codificação, uma vez que eles são construídos por meio de linguagens de programação de propósito geral, as quais possuem construções muito primitivas e que não estão relacionadas aos conceitos referentes à implementação de mecanismos de execução de modelos.

Baseado nisso, foi proposta uma nova abordagem, denominada *Model-Driven Domain-Specific Middleware* (MD-DSM) [Costa et al. 2017]. Essa abordagem, ilustrada na Figura 1, propõe um metamodelo que captura os princípios e construções independentes de domínio usados na implementação de DSVMs, por meio do qual são instanciados modelos utilizados na construção de plataformas de *middleware* específicas de domínio que suportam a execução de aplicações baseadas em modelos. Nesse caso, o significado de "*model-driven*" é sobrecarregado, denotando a ideia de que o *middleware* é construído utilizando modelos, e que as aplicações executadas pelo *middleware* também são definidas por modelos.

A diferença entre essa categoria de *middleware* e os sistemas de *middleware* tradicionais está no nível de abstração das interfaces utilizadas por desenvolvedores de aplicações, que passa de um nível de API tradicional para o nível de modelos definidos por usuários.

A arquitetura de referência adotada pela abordagem MD-DSM é uma

generalização da arquitetura em camadas empregada pela CVM [Deng et al. 2008] e pela MGridVM [Allison et al. 2012], e inclui as seguintes camadas: (1) **User Interface** provê uma interface externa para utilização da plataforma e para definição e gerenciamento de modelos; (2) **Synthesis** recebe o modelo definido na camada User Interface e o transforma em um *script* que é executado pela camada inferior; (3) **Controller** executa os *scripts* repassados pela camada Synthesis, gerando comandos para serem executados na camada Broker. Essa camada também é responsável pela aplicação de requisitos não funcionais aos *scripts*, como restrições de segurança, qualidade de serviço, entre outros; (4) **Broker** provê uma interface de acesso aos recursos gerenciados por ela. Além disso, seleciona os recursos apropriados para aplicar os comandos recebidos pela camada *Controller* e monitora eventos lançados pelos recursos.

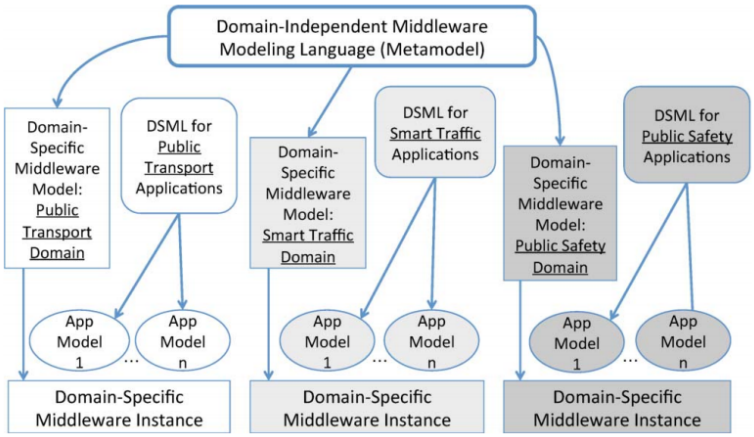


Figura 1. Abordagem para construção de *middleware* específico de domínio dirigido a modelos [Costa et al. 2017].

O processo de construção de *middleware* e aplicações em MD-DSM é ilustrado na Figura 2. Primeiramente, a plataforma de *middleware* é construída utilizando um modelo que representa sua estrutura e um modelo que contém o conhecimento de domínio (*domain knowledge*), responsável pela descrição da semântica operacional do *middleware*.

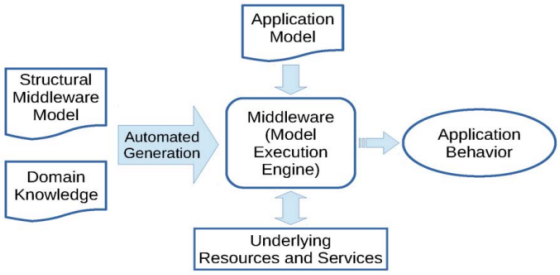


Figura 2. Construção de *middleware* específico de domínio [Costa et al. 2017].

Após sua instanciação, o *middleware* pode ser utilizado para interpretar modelos de aplicações definidos por usuários. Como resultado, ele realiza o gerenciamento de recursos e serviços subjacentes para produzir o comportamento da aplicação descrita pelo modelo dado como entrada.

Até o presente momento, a abordagem MD-DSM foi empregada em dois domínios de aplicação: comunicação [Sousa et al. 2012] e *microgrids* [Junior et al.]. A macroestrutura do metamodelo proposto por tal abordagem engloba (sub-) metamodelos que definem cada uma das camadas da arquitetura de referência adotada, como mostrado na Figura 3.

Atualmente, existem metamodelos totalmente funcionais para as camadas Broker e Controller. O metamodelo para a camada Synthesis está atualmente em desenvolvimento. Quanto à camada UI, são utilizadas ferramentas existentes, como por exemplo as que compõem o *Eclipse Modeling Framework* [Steinberg et al. 2008], que fornecem editores e demais ferramentas para manipulação de modelos, de forma que os usuários consigam definir modelos de aplicações por meio de DSMLs.

Na próxima subseção, será apresentado o (sub-) metamodelo proposto em [Sousa 2012] utilizado para instanciar modelos que correspondem à camada Broker de sistemas de *middleware* específicos de domínio. Após algumas modificações, esse mesmo metamodelo também foi utilizado em [de Sousa 2015], para construir modelos referentes à camada Controller.

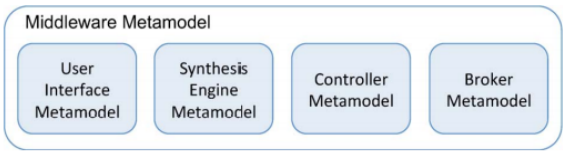


Figura 3. Macroestrutura do metamodelo [Costa et al. 2017].

2.3. Metamodelo da camada Broker

Nesta seção, será apresentado o metamodelo referente à camada Broker da arquitetura de referência adotada pela abordagem MD-DSM. Essa camada é responsável pelo gerenciamento dos recursos usados para prover serviços solicitados. Além disso, a camada Broker provê para a camada sobrejacente, Controller, uma interface uniforme que abstrai todos os detalhes envolvidos na seleção, monitoramento e configuração de recursos.

Para construção do metamodelo da camada Broker, foi utilizado o metamodelo Ecore, que nesse caso atuou como um meta-metamodelo. O metamodelo Ecore integra o *Eclipse Modeling Framework*, que disponibiliza uma série de ferramentas para construção, armazenamento e processamento de modelos.

A classe principal do metamodelo é a `Manager`, cuja instância é responsável por definir um escopo para o gerenciamento de recursos e agrupar outros elementos que definem as atribuições específicas da camada Broker. A Figura 4 apresenta o diagrama de classes com os principais elementos do metamodelo.

Instâncias da classe `Manager` podem utilizar outras instâncias dessa classe como um recurso. Dessa forma, é possível construir uma hierarquia de gerenciadores, cada um fornecendo serviços por meio de seus recursos e gerenciadores mais básicos.

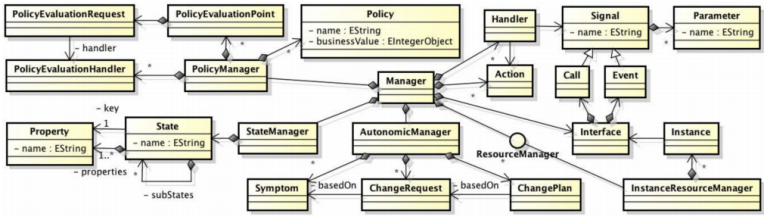


Figura 4. Metamodelo da camada *Broker* [Sousa et al. 2012].

A seguir, serão apresentados os componentes do metamodelo que são importantes para o entendimento da abordagem proposta neste trabalho. Assim sendo, serão enfatizadas as estruturas responsáveis pela definição da interface e das funcionalidades que determinam o comportamento da camada Broker.

2.3.1. Interface

Cada instância da classe `Manager` possui uma interface associada, responsável por definir os sinais (`Signal`) que podem ser enviados ou recebidos pelo gerenciador. Os sinais podem ser chamadas (`Call`) ou eventos (`Events`). A classe `Interface` também é usada para definir a interface dos recursos gerenciados pela camada Broker.

2.3.2. Tratamento de sinais

O comportamento da camada Broker é caracterizado pela forma como ela reage a chamadas feitas pela camada superior e eventos provenientes dos recursos por ela gerenciados.

As principais classes do metamodelo que permitem definir o comportamento da camada são: `Action`, `Handler`, e `Signal`. A classe `Signal` é utilizada para definir os sinais que serão tratados pela camada. Uma instância da classe `Action` define uma operação a ser realizada quando um determinado sinal é identificado. Para associar um sinal a uma ação, é utilizada uma instância da classe `Handler`.

3. Verificação da consistência de *middleware* específico de domínio

Como apresentado na subseção 2.2, MD-DSM é uma abordagem proposta para construção de plataformas de *middleware* específicas de domínio que executam aplicações baseadas em modelos construídos usando DSMLs.

Essa abordagem utiliza um metamodelo que captura conceitos e construções independentes de domínio, inerentes à uma categoria de mecanismos de execução de modelos. Esse metamodelo engloba (sub-) metamodelos referentes a cada camada da arquitetura de referência adotada pela abordagem MD-DSM.

Como as camadas dessa categoria de *middleware* são instanciadas de forma independente, utilizando diferentes metamodelos (um para cada camada), e possivelmente por diferentes especialistas de *middleware*, a possibilidade de inconsistências entre as camadas é muito alta. Isso acontece porque cada camada da pilha possui dependências com as

camadas adjacentes a ela, de forma que o bom funcionamento de uma camada depende de funcionalidades fornecidas por camadas vizinhas.

Atualmente, toda a verificação de consistência entre as camadas é realizada de forma manual, e esse processo se repete para cada novo *middleware* construído e para cada modificação realizada nessas plataformas, resultando em um processo demorado, tedioso e propenso a erros. Baseado nesse problema, nesta seção será apresentada uma abordagem sistemática para verificação de consistência entre as camadas, que utiliza uma ferramenta para verificação de consistência de forma automatizada, e que pode ser empregada para qualquer *middleware* construído utilizando a abordagem MD-DSM.

Outro problema decorrente da abordagem MD-DSM é a necessidade de verificação da conformidade entre um *middleware* e a DSML para a qual ele foi projetado. Isto é, dada uma DSML construída para um determinado domínio, é necessário verificar se o *middleware* implementado para processar modelos instanciados por tal DSML, possui todas as construções necessárias para realizar tal processamento.

Diante disso, na Subseção 3.1, será apresentada uma abordagem para verificar a consistência dos modelos de sistemas de *middleware* específicos de domínio. Na Subseção 3.2, será apresentada uma metodologia para verificar a conformidade entre DSMLs e sistemas de *middleware*.

3.1. Consistência de modelos de *middleware*

Uma vez que os conceitos relacionados a plataformas de *middleware* específicos de domínio foram formalizados na forma de um metamodelo, torna-se possível a utilização de ferramentas que realizam, de forma automatizada, a verificação da consistência dos modelos estruturais dessa categoria de *middleware*. Com base nisso, propomos o uso de restrições inter modelos, construídas usando uma linguagem denominada *Epsilon Validation Language* (EVL) [Paige et al. 2009], como forma de verificar a consistência da pilha de camadas.

A Figura 5 ilustra todas as dependências (simbolizadas pelas setas azuis) identificadas entre as camadas da arquitetura de referência adotada pela abordagem MD-DSM. Ou seja, todo *middleware* construído utilizando MD-DSM possui essas mesmas dependências entre suas camadas. Para cada uma das dependências foi implementada uma restrição inter modelo, cuja responsabilidade é verificar as interfaces dos modelos das camadas a fim de constatar a existência (ou inexistência) de instâncias de determinadas construções definidas nos metamodelos das camadas e que possuem algum tipo de dependência.

Basicamente as restrições realizam a seguinte verificação nos modelos das camadas de um *middleware*: caso uma instância de uma construção X, definida no metamodelo da camada A, seja encontrada no modelo dessa camada, então a restrição EVL busca, no modelo da camada B (adjacente à camada A), uma instância de um elemento Y (que possui dependência com o elemento X), instanciado a partir de uma construção definida no metamodelo da camada B. Dessa forma, caso todas as restrições aplicadas nos modelos de um *middleware* sejam satisfeitas, pode-se considerar que a pilha de camadas está consistente.

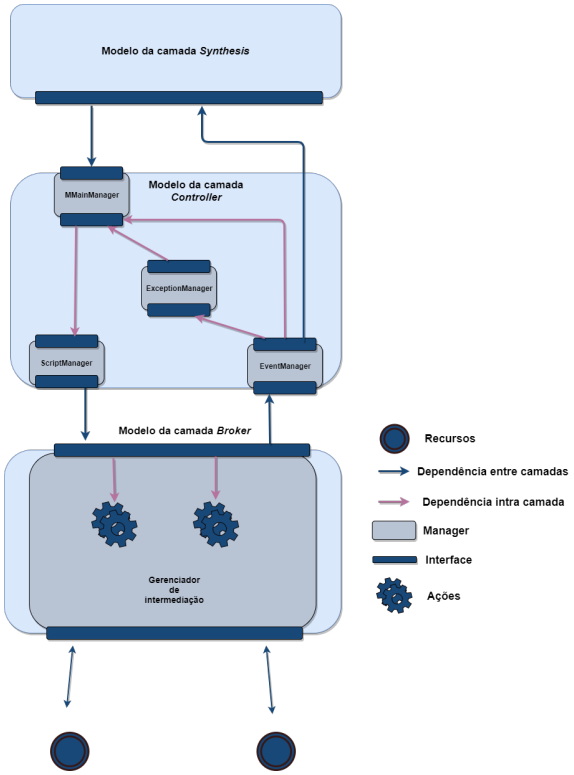


Figura 5. Dependências inter e intra camada

A Tabela 1 apresenta as dependências existentes entre instâncias de construções definidas nos metamodelos das camadas Synthesis, Controller e Broker, e está estruturada da seguinte forma: nas colunas 1 e 3 são apresentadas as construções dos metamodelos cujas instâncias possuem dependências entre si, enquanto as colunas 2 e 4 apresentam as camadas as quais essas instâncias pertencem. Note que as dependências inter camada só existem entre camadas adjacentes, uma vez que cada camada da arquitetura de referência se comunica apenas com a camada imediatamente sobrejacente ou subjacente à ela.

Nas linhas 1 e 2 da Tabela 1, são apresentados os elementos dos modelos das camadas Controller e Synthesis que possuem dependências. Para que a camada Synthesis consiga enviar um *script* à camada Controller, é necessário utilizar uma ação (`Action`), que envia o *script* para uma chamada (`Call`) pertencente à interface da camada Controller (linha 1 da tabela). A camada Synthesis deve instanciar uma ação responsável por receber os eventos provenientes da camada subjacente, que são enviados por meio de uma instância da classe `Event`, pertencente à interface da camada Controller (linha 2 da tabela).

Ao receber um *script*, a camada Controller realiza o mapeamento de cada um dos comandos do *script* para chamadas que compõem a interface da camada Broker. Entre as linhas 3 e 5 da tabela, são apresentados os elementos que necessariamente devem ser instanciados nas camadas Controller e Broker para possibilitar o mapeamento de comandos. Para cada comando contido no *script*, deve ser implementada uma chamada na interface do componente ScriptManager da camada Controller responsável por realizar uma chamada a um ou mais métodos da camada Broker. Os métodos implementados na camada Broker são disponibilizados à camada superior por meio de chamadas que compõem sua

interface.

Além de chamadas, a interface da camada Broker contém elementos do tipo `Event` que sinalizam para a camada superior os tipos de eventos que podem ser disparados, como apresentado entre as linhas 6 a 8 da tabela. A camada Controller, por sua vez, deve conter uma ação responsável pelo tratamento dos eventos lançados pela camada subjacente.

Tabela 1. Dependências inter camada

	Instância	**Camada**	**Instância**	**Camada**
1	`Action`	*Synthesis*	`Call`	*Controller*
2	`Event`	*Controller*	`Action`	*Synthesis*
3	`Call(1)`	*Controller*	`Call(1)`	*Broker*
4	`Call(2)`	*Controller*	`Call(2)`	*Broker*
5	`Call(n)`	*Controller*	`Call(n)`	*Broker*
6	`Event(1)`	*Broker*	`Action`	*Controller*
7	`Event(2)`	*Broker*	`Action`	*Controller*
8	`Event(n)`	*Broker*	`Action`	*Controller*

Além da necessidade de verificar as interfaces das camadas, também é necessário verificar se as funcionalidades expostas nas interfaces estão realmente implementadas no *core* da camada, bem como a consistência entre componentes internos de uma camada. Para essa finalidade, foram construídas restrições intra modelo. As dependências existentes entre os elementos internos dos modelos das camadas são representadas na Figura 5 pelas setas lilás.

A Tabela 2 apresenta as dependências entre componentes internos das camadas Controller e Broker, e possui a mesta estrutura da Tabela 1. Entre as linhas 1 e 4 da Tabela 2, são apresentadas as dependências existentes entre os componentes da camada Controller. O componente MMainManager é responsável por receber os *scripts* da camada Synthesis. Ao receber um *script*, o MMainManager o repassa, por meio de uma ação, para o componente ScriptManager, responsável por realizar todas as operações de processamento relacionadas à *scripts* (linha 1 da tabela). O ScriptManager também precisa instanciar uma ação responsável por receber *scripts*.

Nas linhas 2 e 3, são apresentadas as dependências existentes entre o componente EventManager, responsável pela recepção de eventos provenientes da camada Broker, e os componentes ExceptionManager e MMainManager. Quando um evento é classificado como uma exceção pelo EventManager, ele é repassado, por meio de uma ação, para o componente ExceptionManager. Caso o evento seja classificado como um evento de rotina, ele é repassado para o componente MMainManager. Ambos componentes, MMainManager e ExceptionManager, precisam definir ações para possibilitar a recepção de eventos.

O componente ExceptionManager também deve instanciar uma ação responsável por receber exceções na forma de eventos provenientes da falha de execução de *scripts* no componente ScriptManager (linha 4 da tabela). Como resultado do tratamento de uma exceção, o componente ExceptionManager gera um plano de execução que é enviado ao

MMainManager (linha 5).

Entre as linhas 6 e 8 da tabela, são apresentadas as dependências entre chamadas que compõem a interface da camada Broker e as ações responsáveis pela implementação das funcionalidades expostas na interface.

Tabela 2. Dependências intra camada

	Instância	Camada/Componente	Instância	Camada/Componente
1	`Action`	MMainManager	`Action`	ScriptManager
2	`Action`	EventManager	`Action`	ExceptionManager
3	`Action`	EventManager	`Action`	MMainManager
4	`Action`	ScriptManager	`Action`	ExceptionManager
5	`Action`	ExceptionManager	`Action`	MMainManager
6	`Call(1)`	*Broker*	`Action(1)`	*Broker*
7	`Call(2)`	*Broker*	`Action(2)`	*Broker*
8	`Call(n)`	*Broker*	`Action(n)`	*Broker*

Como as restrições EVL são aplicadas sobre instâncias de construções dos metamodelos das camadas de *middleware*, elas se tornam genéricas o bastante para realizar a verificação de consistência estrutural de qualquer *middleware* construído usando a abordagem MD-DSM, independentemente do domínio de aplicação para o qual ele é utilizado.

3.2. Conformidade entre DSML e *middleware*

Nesta seção será apresentada a segunda parte da abordagem proposta neste trabalho, referente à verificação de conformidade entre DSMLs e sistemas de *middleware* construídos utilizando a abordagem MD-DSM.

Como apresentado na subseção 2.2, a abordagem MD-DSM é empregada para a construção de sistemas de *middleware* para diferentes domínios. Para cada domínio distinto é necessário criar uma nova DSML que contêm conceitos específicos de tal domínio e um *middleware* capaz de interpretá-la. Nesse processo é importante observar a necessidade de conformidade entre o *middleware* e a DSML para a qual ele foi projetado.

Nesse sentido, propomos uma metodologia que sistematiza todo o processo de verificação de conformidade entre *middleware* e DSML. Para aplicação da metodologia assumimos a existência de uma DSML para um determinado domínio e do modelo do *middleware* que será utilizado para executar tal DSML. Quando nos referimos ao "modelo do *middleware*" implicitamente estamos nos referindo aos modelos que representam cada uma das camadas da arquitetura de referência adotada por essa categoria de *middleware* tratada neste trabalho.

A Figura 6 apresenta uma visão geral da metodologia proposta. Uma vez que o processo de construção da DSML esteja finalizado, o primeiro passo da metodologia pode ser aplicado. Nessa etapa propomos a construção de um *feature model* da linguagem de modelagem, que tem como objetivo representar todas as funcionalidades suportadas pela DSML. Esse *feature model* da linguagem é construído com base no *feature model* de domínio, que é um dos produtos da primeira etapa (análise de domínio) do ciclo de desenvolvimento de DSMLs.

A responsabilidade de construção do *feature model* da linguagem de modelagem é do mesmo especialista responsável pela construção da DSML. Uma vez construídos, a DSML e o *feature model* são repassados para o especialista responsável pela construção do middleware que irá executar tal DSML.

O especialista de *middleware* é responsável pela execução da segunda etapa da metodologia, que tem como objetivo expressar, por meio de um modelo de relacionamentos, todos os relacionamentos existentes entre *features* do *feature model* da DSML e construções definidas no metamodelo de *middleware*. Ou seja, cada relacionamento expressa quais construções do metamodelo precisam, obrigatoriamente, serem instanciadas no modelo do *middleware* para que ele ofereça suporte à execução de tal *feature* da DSML.

Para construção do modelo de relacionamentos, fornecemos um metamodelo que possui todas as construções necessárias para representar relacionamentos entre construções definidas no metamodelo de *middleware* MD-DSM e *features* de qualquer *feature model* de DSML. Para o estabelecimento dos relacionamentos propomos o uso de uma ferramenta, denominada ModeLink, que compõe uma família de ferramentas e linguagens para manipulação de modelos chamada Epsilon. ModeLink permite carregar o metamodelo de *middleware*, o modelo de relacionametto e o *feature model* da DSML, e estabelecer os relacionamentos de forma gráfica.

Uma vez construído, o modelo de relacionamento, juntamente com um modelo de *middleware* podem ser dados como entrada para uma ferramenta de verificação, construída a partir da linguagem EVL, que verifica, baseada no modelo de relacionamento, se o modelo do *middleware* está em conformidade com uma determinada DSML.

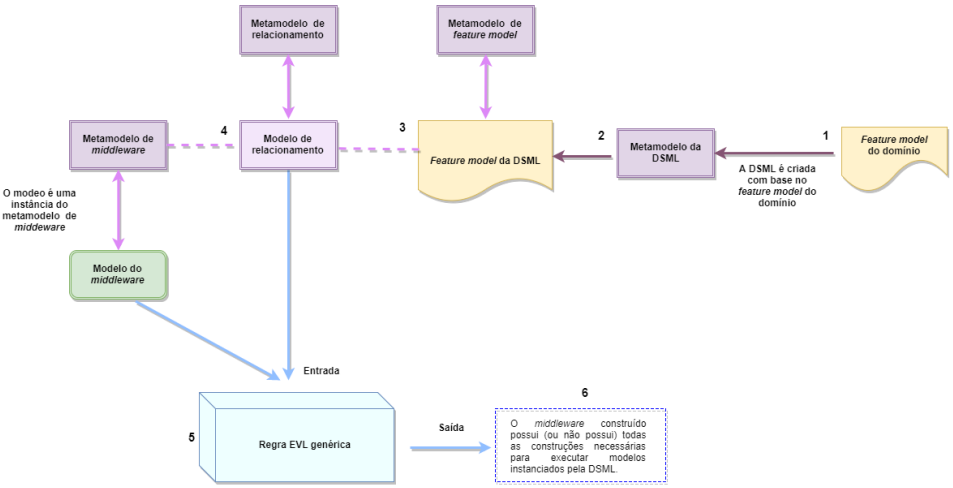


Figura 6. Abordagem para verificação de conformidade entre sistemas de *middleware* específicos de domínio e DSMLs.

Note que todo o processo de verificação de conformidade pode ser realizado de forma manual, sem a necessidade de passar por essas etapas intermediárias. Entretanto, esse processo manual se repetiria novamente à cada novo *middleware* construído para uma determinada DSML e a cada evolução que um *middleware* sofresse. Ou seja, em um mesmo domínio podem existir diferentes plataformas de *middleware* que podem ser cons-

truídas para uma mesma DSML. Assim sendo, a metodologia apresentada nesta subseção sistematiza e semi-automatiza o processo de verificação entre uma DSML e os diferentes sistemas de *middleware* que podem ser construídos para processá-la.

Vale ressaltar que para cada novo domínio (e consequentemente, para cada nova DSML) é necessário construir um *feature model* da DSML e um modelo de relacionamento. Entretanto, a ferramenta de verificação EVL, não precisa ser modificada, pois ela é independente de domínio e pode ser empregada para verificar a conformidade de qualquer modelo de *middleware* e qualquer DSML.

4. Considerações Finais

Técnicas de MDE têm sido utilizadas como alternativa para diminuir o tempo, custo e esforço empregados na construção de sistemas de software. Essas abordagens empregam modelos como artefatos primários no processo de desenvolvimento de software. Para instanciar modelos utilizados por abordagens de MDE, são construídas DSMLs. A transformação dos modelos em artefatos de *software* é realizada por mecanismos denominados *model interpreters*.

MD-DSM é uma abordagem que emprega modelos para a construção de uma categoria de *model interpreter*, chamada de *middleware* específico de domínio, utilizados para processar aplicações descritas por DSMLs. Neste artigo, apresentamos uma abordagem para verificar a consistência dos modelos utilizados na construção dessa categoria de *middleware*, bem como uma metodologia a ser seguida por especialistas de *middleware* para verificar a conformidade entre uma DSML e o *middleware* construído para processá-la.

Atualmente, já temos uma implementação totalmente funcional da abordagem proposta para verificação de consistência da pilha de camadas de middleware específico de domínio. Uma validação inicial dessa abordagem já foi realizada, onde empregamos as restrições EVL em dois domínios: comunicação [Sousa 2012] e *microgrids* [Junior et al.]. Já a abordagem para verificação de conformidade entre DSML e *middleware* ainda está em desenvolvimento.

Referências

Allison, M., Allen, A. A., Yang, Z., and Clarke, P. J. (2011). A software engineering approach to user-driven control of the microgrid. In *SEKE*, pages 59–64.

Allison, M., Morris, K. A., Yang, Z., Clarke, P. J., and Costa, F. M. (2012). Towards reliable smart microgrid behavior using runtime model synthesis. In *High-Assurance Systems Engineering (HASE), 2012 IEEE 14th International Symposium on*, pages 185–192. IEEE.

Bryant, B. R., Gray, J., Mernik, M., Clarke, P. J., France, R. B., and Karsai, G. (2011). Challenges and directions in formalizing the semantics of modeling languages. *Computer Science and Information Systems*, 8(2):225–253.

Clarke, P. J., Wu, Y., Allen, A. A., Hernandez, F., Allison, M., and France, R. (2012). Towards dynamic semantics for synthesizing interpreted dsmls. *Formal and Practical Aspects of Domain-Specific Languages: Recent Developments: Recent Developments*, page 242.

Costa, F. M., Morris, K. A., Kon, F., and Clarke, P. J. (2017). Model-driven domain-specific middleware. In *Distributed Computing Systems (ICDCS), 2017 IEEE 37th International Conference on*, pages 1961–1971. IEEE.

de Sousa, A. B. (2015). Middleware Dirigido por Modelos para Microgrids.

Deng, Y., Sadjadi, S. M., Clarke, P. J., Hristidis, V., Rangaswami, R., and Wang, Y. (2008). Cvm–a communication virtual machine. *Journal of Systems and Software*, 81(10):1640–1662.

Favre, J.-M. (2004). Towards a basic theory to model model driven engineering. In *3rd Workshop in Software Model Engineering, WiSME*, pages 262–271. Citeseer.

Fishwick, P. A. (2007). *Handbook of dynamic system modeling*. CRC Press.

France, R. and Rumpe, B. (2007). Model-driven development of complex software: A research roadmap. In *2007 Future of Software Engineering*, pages 37–54. IEEE Computer Society.

Junior, A. R. S., Costa, F. M., and Clarke, P. A model-driven approach to develop and manage cyber-physical systems.

Kelly, S. and Tolvanen, J.-P. (2008). *Domain-specific modeling: enabling full code generation*. John Wiley & Sons.

Kent, S. (2002). Model driven engineering. In *International Conference on Integrated Formal Methods*, pages 286–298. Springer.

Kirshin, A., Dotan, D., and Hartman, A. (2006). A uml simulator based on a generic model execution engine. In *MoDELS Workshops*, volume 4364, pages 324–326.

Paige, R. F., Kolovos, D. S., Rose, L. M., Drivalos, N., and Polack, F. A. (2009). The design of a conceptual framework and technical infrastructure for model management language engineering. In *Engineering of Complex Computer Systems, 2009 14th IEEE International Conference on*, pages 162–171. IEEE.

Riehle, D., Fraleigh, S., Bucka-Lassen, D., and Omorogbe, N. (2001). The architecture of a uml virtual machine. *ACM SIGPLAN Notices*, 36(11):327–341.

Schmidt, D. C. (2006). Model-driven engineering. *COMPUTER-IEEE COMPUTER SOCIETY-*, 39(2):25.

Selic, B. (2003). The pragmatics of model-driven development. *IEEE software*, 20(5):19.

Sousa, G., Costa, F. M., Clarke, P. J., and Allen, A. A. (2012). Model-driven development of dsml execution engines. In *Proceedings of the 7th Workshop on Models@ run. time*, pages 10–15. ACM.

Sousa, G. C. M. (2012). Desenvolvimento de máquinas de execução para linguagens de modelagem específicas de domínio: uma estratégia baseada em engenharia dirigida por modelos. Master's thesis, Universidade Federal de Goiás.

Steinberg, D., Budinsky, F., Merks, E., and Paternostro, M. (2008). *EMF: eclipse modeling framework*. Pearson Education.

IFLivros – Sistema de Gerenciamento de Livros para o IFG Câmpus Luziânia

Jorge Tawnan Leles Vilela[1], Reinoldo Oliveira Sommer[1], Ernane Rosa Martins[1]

[1]Instituto Federal de Goiás (IFG)
Caixa Postal 72.811-580 – Luziânia – GO – Brasil

jorgetawnan@gmail.com, reinoldo@outlook.com, ernane.martins@ifg.edu.br

Abstract. *The book is part of the educational foundation in any educational institution, so making it more accessible to students is an excellent contribution to the academic community. There are social networking groups that bring together people interested in swapping books, but a system would make it easier to identify which books a person needs and which ones they have and want to get rid of. In this work a web system is developed with the purpose of managing the personal books of teachers and students of the Federal Institute of Goiás (IFG). The applied methodology was a qualitative approach characterized, mainly, by its descriptive nature, with bibliographical research and development of software design.*

Resumo. *O livro faz parte da base educacional em qualquer instituição de ensino, por esse motivo torná-lo mais acessível aos alunos é uma excelente contribuição para a comunidade acadêmica. Existem grupos em redes sociais que reúnem o pessoal interessado em trocar livros, mas um sistema facilitaria identificar quais livros uma pessoa precisa e quais ela tem e quer se desfazer. Neste trabalho é desenvolvido um sistema web com a finalidade de gerenciar os livros pessoais dos docentes e discentes do Instituto Federal de Goiás (IFG). A metodologia aplicada foi uma abordagem qualitativa caracterizada, sobretudo, pela sua natureza descritiva, com pesquisa bibliográfica e desenvolvimento de projeto de software.*

1. Introdução

O livro é uma das formas mais antigas de se passar e adquirir conhecimento, querendo ou não ele faz parte de toda a vida acadêmica, do jardim de infância com os livros de colorir ao nível superior com os complexos livros de engenharia de *software*. A cidade de Luziânia conta com apenas uma livraria e uma biblioteca pública que não tem um acervo atualizado constantemente e nem tantos livros como deveria, para uma cidade que segundo os dados de 2015 do Instituto Brasileiro de Geografia e Estatística (IBGE) têm 1.823 docentes e 43.778 alunos matriculados na rede escolar.

O Instituto Federal de Goiás (IFG) Câmpus Luziânia têm disponível para os 536 alunos matriculados no primeiro semestre de 2017, 44 servidores do administrativo, 68 professores e aproximadamente 3.700 livros. A instituição possui um sistema de empréstimo desse acervo bibliográfico, no entanto, é notória a escassez de exemplares em todos os cursos que o instituto oferece, sendo o aluno obrigado a recorrer à compra de livros pessoais ou a buscar o recurso da pirataria. Para isso, será implantado um sistema web exclusivo para o IFG – Luziânia que gerenciará os livros dos alunos e servidores, proporcionando alguns benefícios como: ganho de tempo, venda de material inutilizado e conforto.

A vida acadêmica exige muito estudo e dedicação, e os estudantes sempre estão à procura de bons livros, seja para adquirir conhecimento ou apenas para passar o tempo. Tendo isso em vista, muitos livros são adquiridos no decorrer dos anos e depois não se sabe o que fazer com eles. Assim, vê se a necessidade de criação de um sistema único capaz de gerenciar e unificar o acervo pessoal de todos os envolvidos na instituição. Atualmente, não existe um serviço de automação do acervo pessoal de docentes e discentes do IFG. Os livros (digitais e físicos) são armazenados da forma que lhes convêm e muitas vezes inutilizados com o passar do tempo. Além do mais, livros são caros (em sua maioria) para alunos, fazendo com que o mesmo opte por adquirir cópias pirateadas.

O objetivo deste trabalho é desenvolver um sistema *web* para gerenciar o acervo individual de livros dos docentes e discentes do Instituto Federal de Goiás Câmpus Luziânia. Automatizar as atividades acadêmicas que envolva estoque pessoal de livros proporcionaria o gerenciamento e integração de estoque e desejos de cada usuário. Um sistema possibilitaria a unificação de todos os usuários, assim existirá uma forma de empréstimo, troca ou venda do material. Com todos estes recursos disponíveis, a procura por falsificações na Internet diminuiria, possibilitando a oportunidade de vender, trocar ou emprestar seus livros.

Nesta introdução foi apresentada além da definição da problemática de pesquisa, o objetivo, a justificativa, a importância do estudo e a estrutura da presente pesquisa. A sessão 2 traz o referencial teórico, com a formação de uma base conceitual e teórica que fornecem subsídios para o desenvolvimento deste estudo, bem como seu desenvolvimento prático. Na sessão 3 são apresentados os materiais, os métodos empregados e as técnicas e procedimentos metodológicos utilizados para a realização do estudo. Na sessão 4 são mostrados os artefatos do projeto necessários para a criação do mesmo, como: regras de negócio, modelos e diagramas. Na sessão 5 são apresentadas as análises e discussões dos resultados obtidos na pesquisa sobre o desenvolvimento do sistema *web*. Finalmente na A sessão 6, é realizada uma abordagem conclusiva do projeto e apresentadas às recomendações de trabalhos futuros.

2. Referencial Teórico

O mundo atual é altamente dependente de software. No século atual, a maioria dos produtos elétricos fabricados incluem um *hardware* e um *software* embutido. Uma grande parte da indústria e manufatura já é informatizada, assim como o sistema financeiro. Os sistemas de *software* são abstratos e intangíveis, isso significa que eles não são restringidos pelas propriedades dos materiais, nem governados pelas leis da física, mas devido a essa falta de restrições físicas, os sistemas de *software* podem se tornar extremamente complexos de modo muito rápido, difíceis de entender e caros para alterar (SERVILHEIRO, 2011).

As aplicações para web são projetadas para serem executados em um navegador através da internet. Essas aplicações se conectam com um servidor (pode ser uma conexão local, ou seja, na própria máquina ou em um servidor na internet) e seu principal objetivo é interagir com esse servidor através de requisições para obter dados/respostas. As tecnologias disponíveis, como *Hyper Text Markup Language* (HTML), *JavaScript* e uma linguagem de programação do *server-side* eram suficientes para atender a demanda. Ao longo dos anos a *internet* passou por grandes avanços e se tornou um ambiente também de negócios. Com isso o usuário além de ler suas notícias, passou a consultar preços de produtos, reservar passagens aéreas, solicitar orçamentos

para serviços e inúmeras outras coisas (SANCHEZ & ALTHMANN, 2013). Uma aplicação para *Web* (“*WebApp”*) é uma categoria de *software* centralizada em redes que está em constante evolução. Possui ambientes computacionais sofisticados que podem estar integrados a bancos corporativos e aplicações comerciais. Uma WebApp além de oferecer recursos especializados, fornece funções computacionais e conteúdo para o usuário final (PRESSMAN, 2011).

A arquitetura web é projetada em camadas e existem vários modelos disponíveis. Os mais conhecidos são em duas camadas, três camadas e quatro camadas. Os mais utilizados são em três e quatro camadas, pois separam melhor o comportamento da aplicação (PRESSMAN, 2011). Neste projeto faz-se a utilização do conceito *Thin Client* (“Cliente Magro”). Fundamentalmente significa que o cliente não precisa instalar nada, além de um navegador para utilizar os recursos de uma aplicação web, e o processamento dos dados é feito pelos servidores. O padrão de projeto arquitetônico Model, View e Controller (MVC) separa a aplicação em três camadas trazendo várias vantagens na infraestrutura do projeto, a principal vantagem é justamente desacoplar a interface do usuário (*view*) da funcionalidade (*controller*) e do conteúdo (*model*) da aplicação *web (PRESSMAN, 2011)*.

O objetivo da engenharia de *software* visa à construção de um *software*, em todos os aspectos. Os requisitos, projeto, desenvolvimento, gerenciamento, teste (SOMMERVILLE, 2007). A grande parte dos desenvolvedores comete o erro de achar que uma análise mais detalhada dos requisitos seria perda de tempo e começam o *software* pelo desenvolvimento. A engenharia de requisitos constrói uma ponte para o projeto, permitindo uma apreciação mais detalhada do *software* a ser realizado (PRESSMAN, 2011). As necessidades específicas do projeto e a construção devem atender as prioridades que orientam a ordem na qual o trabalho deve ser completado e as informações, funções e comportamentos que terão um impacto profundo no projeto resultante (PRESSMAN, 2011).

Os requisitos funcionais de um sistema descrevem o que ele deve fazer, como deve reagir a entradas específicas e de como deve se comportar em determinadas situações. Dependendo do caso, os requisitos funcionais podem dizer o que o sistema não deve fazer (SOMMERVILLE, 2011). Segundo Sommerville (2011), os requisitos funcionais podem ser divididos em duas categorias: Requisitos de usuário: são os requisitos elaborados de uma forma mais abstrata para que o usuário possa entender. Requisitos de sistema: são requisitos mais específicos que descrevem com detalhes as funcionalidades do sistema. Os requisitos não funcionais descrevem os requisitos que não estão diretamente relacionados com os serviços específicos do sistema. Os requisitos surgem por meio das necessidades dos usuários, devido a restrições de orçamento, políticas organizacionais, necessidade de interoperabilidade com outros sistemas de *software* ou *hardware*, ou a partir de fatores externos, como regulamentos de segurança ou legislações de privacidade (SOMMERVILLE, 2011).

Modelagem de sistemas é um processo de desenvolvimento de modelos abstratos, em que cada modelo apresenta uma visão ou perspectiva, diferente do sistema. Esses modelos são utilizados na engenharia para ajudar a extrair os requisitos do sistema. Na fase de projeto, os modelos são utilizados para auxiliar os engenheiros na criação do *software*, e servem de documentação para a estrutura do sistema (SOMMERVILLE, 2011).

A *Unified Modeling Language* (UML) é um conjunto de notações gráficas, apoiada por modelos que ajudam na descrição e no projeto de *software*. A UML é um

padrão relativamente aberto, nasceu da unificação das muitas linguagens gráficas de modelagem orientadas a objetos que floresceram no final dos anos oitenta, início dos noventa (FOWLER, 2005). No presente trabalho, a linguagem UML foi aplicada no desenvolvimento dos modelos de caso de uso e diagrama de classes. Os diagramas de casos de uso representam as funcionalidades e elementos externos que podem ser observadas em um sistema. Modelo de análise que deve ser feito com base nos requisitos funcionais. Cada caso de uso é definido por uma descrição narrativa das interações que ocorrem entre os elementos externos e o sistema (BEZERRA, 2007). O diagrama de classes descreve as variações de objetos presentes no sistema e os tipos de relacionamento entre eles. Mostram as propriedades e as operações, mais conhecidas como métodos, de uma classe e a maneira que os objetos se aplicam as restrições conectadas (FOWLER, 2005).

3. Materiais e Métodos

O projeto está na categoria de Aplicações *Web*. Foram utilizadas as metodologias da engenharia de *software* descritas a seguir, para atender as necessidades do projeto. É retratado o processo de engenharia de requisitos utilizado no sistema. Na fase de estudo da viabilidade, foi feito uma análise com os professores e servidores da biblioteca do IFG Luziânia sobre a viabilidade do projeto e se poderia agregar valor à instituição. Na fase de elicitação e análise de requisitos, utilizando dados do IBGE, realizou-se o levantamento de quantos professores e alunos existem em Luziânia. Utilizando dados do IFG Luziânia, foi feito o levantamento da quantidade de livros na biblioteca, de alunos e de servidores. Assim, foi definido o problema e a necessidade do sistema. Na fase de especificação de requisitos, foi feito um documento de requisitos com base no problema e nas necessidades encontradas na fase de elicitação e de viabilidade. O documento é composto do escopo, requisitos funcionais, requisitos não funcionais e regras de negócio do sistema. Na fase de validação de requisitos foi realizada uma análise dos requisitos para garantir que todos os requisitos foram declarados de forma não ambígua. Procuramos detectar inconsistências, erros e se os requisitos estavam dentro do padrão estabelecido para o projeto. Em seguida foi feita uma análise dos requisitos junto com o orientador e entramos em comum acordo com as mudanças sugeridas.

Para nivelar a análise de requisitos a ser tomada prestou-se atenção nos seguintes fatores: Tamanho e complexidade do aplicativo; Tamanho da equipe de desenvolvedores; Nível em que os membros da equipe de desenvolvimento trabalham juntos; Nível de êxito da organização é diretamente dependente do êxito da aplicação. Analisando o tamanho da equipe e a complexidade do projeto, o mesmo foi mensurado apenas para atender a demanda do IFG. A pesquisa envolveu uma abordagem qualitativa caracterizada, sobretudo, pela sua natureza descritiva, com pesquisa bibliográfica, desenvolvimento de projeto de software e realização de questionário de usabilidade com discentes e docentes do Instituto Federal de Goiás Câmpus Luziânia, de modo a avaliar os resultados do desenvolvimento do sistema *web*.

Foram utilizadas as seguintes tecnologias na construção do projeto:

1. ASP.NET MVC (*Microsoft*, 2017b): *framework* para criação de aplicativos *web* que utiliza o padrão arquitetônico MVC. Foi optada essa tecnologia pela boa integração com as ferramentas da *Microsoft* utilizadas nesse projeto.
2. *Visual Studio Community (Microsoft, 2015):* uma IDE gratuita, completa e extensível para a criação de aplicativos modernos para *Windows*, *Android* e iOS, bem como aplicativos *web* e serviços de nuvem. O *Visual*

Studio é o ambiente de desenvolvimento mais indicado para trabalhar com a linguagem de programação C# e possui integração com as ferramentas da *Microsoft.*

3. Linguagem C# versão 7.0: O C# (leia-se C-Sharp) é uma linguagem de programação bastante popular desenvolvida pela *Microsoft*. É uma tecnologia gratuita e disponível para download no site da *Microsoft*. Foi utilizada por ser a linguagem padrão do *Visual Studio*.
4. *Entity Framework Code First* 6.1.3 (Dykstra, 2015): ferramenta de mapeamento objeto relacional (ORM – *Object Relational Management*). Utilizada para fazer a persistência com o banco de dados. Foi utilizada por ser uma ferramenta já integrada no projeto ASP.NET MVC, isso garante o seu bom funcionamento e suporte.
5. *Code First Migrations (dos Santos, 2012):* permite gerar atualizações gerenciáveis no banco de dados ou deixar que o próprio *Migrations* cuide de tudo de forma automática, mantendo o banco de dados sempre atualizado com suas classes. Utilizado para manter o banco atualizado com as classes de modelo.
6. ASP.NET Identity (Rastogi, Anderson, Dykstra, & Galloway, 2013): é responsável por toda a autenticação e facilmente integrado com as redes sociais. O Identity é um componente indispensável neste trabalho. Tornou-se *open-source* (código aberto) a partir da segunda versão.
7. MySql versão 5.7 (Oracle, 2017): Sistema gerenciador de banco de dados. Foi utilizado por ser gratuito.
8. Linguagem HTML, CSS, JavaScript: Utilizado na camada de visão. Essas linguagens são padrões universais, HTML por ser de marcação na web, CSS por ser de estilização das páginas e JavaScript por ser a linguagem de programação utilizada nos navegadores.
9. Bootstrap versão 3.3.7 (Otto & Thornton, 2017): *framework front-end* para desenvolvimento de projetos web. Foi utilizado neste trabalho por reunir de forma simples e fácil as linguagens utilizadas.
10. JQuery versão 1.12.1 (The jQuery Foundation, 2017): conjunto de interações de interface de usuário, efeitos, *widgets* e temas construídos em cima da biblioteca jQuery *JavaScript*. Utilizado na camada de visão.
11. Astah *Community (Astah, 2017):* ferramenta para criação de modelos, como casos de uso e diagramas de classe. Foi utilizada por ser uma ferramenta gratuita e que atende as necessidades desse projeto.
12. *Team Foundation Server* – TFS (Microsoft, 2017a): ferramenta para controle de versão. Integrada no *Visual Studio*, é bastante eficiente e atende as necessidades desse projeto.

4. IFLivros

IFLivros surgiu a partir da ideia de reunir todo material obtido pelas pessoas durante a fase acadêmica, sejam livros, revistas, *e-books,* e colocá-los num sistema onde essas possam trocar, vender e emprestar seu acervo. Esse sistema foi desenvolvido inicialmente para o IFG, mas pretende ser expandido a diversas instituições a nível mundial.

A Tabela 1 apresenta a especificação dos requisitos funcionais com um identificador, descrição e os requisitos relacionados.

Tabela 1 – Requisitos funcionais.

Identificador	Descrição	Requisitos Relacionados
RF01	O sistema deve permitir o cadastramento e manutenção de tipos de grupo.	
RF02	O sistema deve permitir o cadastramento e manutenção de tipos de integrantes dos grupos.	
RF03	O sistema deve permitir o cadastramento e manutenção de tipos de livros.	
RF04	O sistema deve permitir o cadastramento e manutenção de tipos de oferta.	
RF05	O sistema deve permitir o cadastramento e manutenção de tipos de pessoa.	
RF06	O sistema deve permitir o cadastramento e manutenção de pessoas.	RF05, RF19
RF07	O sistema deve permitir o cadastramento e manutenção de livros.	
RF08	O sistema deve permitir o cadastramento e manutenção de desejos.	RF07, RF06, RF04
RF09	O sistema deve permitir o cadastramento e manutenção de grupos.	RF01
RF10	O sistema deve permitir o cadastramento e manutenção de estoque.	RF03, RF06, RF07
RF11	O sistema deve permitir o cadastramento e manutenção de gênero dos livros.	
RF12	O sistema deve permitir o cadastramento e manutenção de sugestão de livros.	RF06, RF11
RF13	O sistema deve permitir o cadastramento e manutenção de ligação entre livro e gênero.	RF07, RF11
RF14	O sistema deve permitir o cadastramento e manutenção de ofertas.	RF10, RF04
RF15	O sistema deve permitir o cadastramento e manutenção de transações.	RF06, RF10, RF14
RF16	O sistema deve permitir o cadastramento e manutenção de entradas no grupo.	RF06, RF09
RF17	O sistema deve permitir o cadastramento e manutenção de integrantes do grupo	RF16, RF02
RF18	O sistema deve permitir o cadastramento e manutenção de ofertas para grupo.	RF04, RF10, RF09
RF19	O sistema deve permitir o cadastramento e manutenção de usuários.	
RF20	O sistema deve conter tela de login.	RF19
RF21	O sistema deve ter uma tela onde o usuário posso pesquisar ofertas.	

A Tabela 2 apresenta a especificação dos requisitos não funcionais com um identificador, descrição e os requisitos relacionados.

Tabela 2 – Requisitos não funcionais.

Identificador	Descrição	Requisitos Relacionados
RNF01	O sistema deve ser desenvolvido para *web*.	
RNF02	O sistema deve ser exclusivo para pessoas com algum vínculo ao IFG – Luziânia.	
RFN03	O sistema deve estar online 24 h/ 7 dias por semana.	
RFN04	O sistema deve ter uma interface de fácil utilização.	RNF01
RFN05	O sistema deve ser responsivo, para que rode em mobile.	RFN04

A Tabela 3 apresenta a especificação das regras de negócio com um identificador, descrição e os requisitos relacionados.

Tabela 3 – Regras de negócio.

Identificador	Descrição	Requisitos Relacionados
RN01	A tela de cadastro, visualização, edição e exclusão de tipos de grupo deve conter o campo de: Nome do tipo (Será obrigatório, do tipo string e com tamanho de até 80 caracteres); Quantidade máxima de participantes do grupo (Será obrigatório, do tipo inteiro).	RF01
RN02	A tela de cadastro, visualização, edição e exclusão de tipos de integrantes dos grupos deve conter o campo de: Nome do tipo (Será obrigatório, do tipo string e com tamanho de até 80 caracteres).	RF02
RN03	A tela de cadastro, visualização, edição e exclusão de tipos de livros deve conter o campo de: Nome do tipo (Será obrigatório, do tipo string e com tamanho de até 80 caracteres).	RF03
RN04	A tela de cadastro, visualização, edição e exclusão de tipos de oferta deve conter o campo de: Nome do tipo (Será obrigatório, do tipo string e com tamanho de até 80 caracteres).	RF04
RN05	A tela de cadastro, visualização, edição e exclusão de tipos de pessoas deve conter o campo de: Nome do tipo (Será obrigatório, do tipo string e com tamanho de até 80 caracteres).	RF05
RN06	A tela de cadastro, visualização, edição e exclusão de gêneros deve conter o campo de: Nome (Será obrigatório, do tipo string e com tamanho de até 150 caracteres).	RF11
RN07	A tela de cadastro de pessoas deve conter os seguintes campos: Nome (Será obrigatório, do tipo string e com tamanho de até 80 caracteres). Matrícula (Será obrigatório, do tipo string e com tamanho de até 50 caracteres). CPF (Será obrigatório, do tipo string e com tamanho fixo de 15 caracteres). Sexo (Será obrigatório, do tipo string e com tamanho de 1 caractere). RG (Será obrigatório, do tipo string e com tamanho de até 20 caracteres). Órgão expedidor (Será obrigatório, do tipo string e com tamanho de até 50 caracteres). Cidade (Será obrigatório, do tipo string e com tamanho de até 150 caracteres). UF (Será obrigatório, do tipo string e com tamanho fixo de 2 caracteres). E-mail (Será obrigatório, do tipo string e com tamanho de até 150 caracteres). Endereço (Será obrigatório, do tipo string e com tamanho de até 150 caracteres). Telefone (Será obrigatório, do tipo string e com tamanho entre 13 e 14 caracteres). Celular (Será obrigatório, do tipo string e com tamanho entre 13 e 14 caracteres). Tipo Pessoa, da pessoa por padrão será um usuário comum, deve fazer referência com a tabela de tipopessoa.	RF06

	Usuário, recebera o código do usuário da pessoa cadastrada.	
RN08	O cadastramento de usuário deve conter os seguintes campos: E-mail (Será obrigatório, do tipo string e com tamanho de até 150 caracteres). Senha (Será obrigatório, do tipo string e com tamanho de até 50 caracteres).	RF19
RN09	O cadastro de pessoas e usuário devem ser feitos na mesma tela e sincronizados.	RN07, RN08
RN10	A tela de cadastro, visualização, edição e exclusão de livros deve conter o campo de: Autor (Será obrigatório, do tipo string e com tamanho entre 2 e 254 caracteres). Autor (Será obrigatório, do tipo string e com tamanho entre 2 e 50 caracteres). Número da edição (Será obrigatório e do tipo inteiro) Páginas (Será obrigatório e do tipo inteiro) Título (Será obrigatório, do tipo string e com tamanho entre 2 e 100 caracteres). Subtítulo (Será obrigatório, do tipo string e com tamanho entre 2 e 100 caracteres). Editora (Será obrigatório, do tipo string e com tamanho entre 2 e 100 caracteres). Ano (Será obrigatório, do tipo string e com tamanho fixo de 4 caracteres). Url (Não obrigatório do tipo string com tamanho máximo permitido pelo banco).	RF07
RN11	O cadastramento de Ligar Livro Gênero deve conter os seguintes campos: IdDoLivro que deve fazer referência ao livro. IdDoGenero que deve fazer referência ao gênero. Ativo, deve ser do tipo boolean e obrigatório.	RF13
RN12	O cadastramento de ligar livro gênero deve ser feito junto com o cadastro de livros, dando ao usuário a opção de selecionar quais gêneros ele quiser para o livro.	RN11, RN10
RN13	A tela de cadastro, visualização, edição e exclusão de grupos deve conter os campos de: IdTipoGrupo (Faz referência ao tipo do grupo). Nome (Será obrigatório, do tipo string com tamanho 100). DataCriacao (Será obrigatório, do tipo DateTime). Ativo (Será obrigatório, do tipo boolean).	RF09
RN14	O cadastro de entrada grupo deve conter os seguintes campos: IdPessoaConvite (Faz referência a pessoa que fez o convite). IdPessoaConvidado (Faz referência a pessoa que foi convidada). IdDogrupo (Faz referência ao grupo). FoiConvidado (Será obrigatório, do tipo boolean). Cadastrado (Será obrigatório, do tipo boolean).	RF16
RN15	O cadastro de integrantes do grupo deve conter os seguintes campos: IdTipoIntegranteGrupo (Faz referência ao tipo de integrante grupo). IdEntradaGrupo (Faz referência a entrada grupo). DataEntrada (Será obrigatório, do tipo DateTime). Bloqueado (Será obrigatório, do tipo boolean).	RF17
RN16	O cadastramento de entrada no grupo deve ser feito assim que o usuário confirma sua entrada no grupo.	RN14
RN17	Caso o usuário insira o mesmo livro o sistema deve contabilizar a quantidade inserida no livro que já existe em estoque.	RN10
RN18	A tela de cadastro, visualização, edição e exclusão de estoque deve conter o campo de: IdDapessoa (Faz referência a pessoa). IdLivro (Faz referência ao livro). IdTipoLivro (Faz referência ao tipo livro). Quantidade (Será obrigatório, do tipo inteiro).	RF10
RN19	Caso a quantidade do estoque seja zero o livro não deve aparecer no estoque do usuário, porém deve estar cadastrado.	RN18
RN20	A tela de cadastro, visualização, edição e exclusão de oferta deve conter o campo de: IdEstoque (Referente ao estoque). IdTipoOferta (Referente ao tipo oferta). Valor (Não será obrigatório, do tipo decimal). Quantidade (Será obrigatório, do tipo inteiro).	RF14
RN21	Caso a quantidade da oferta seja zero o livro não deve aparecer no estoque do usuário, porém deve estar cadastrado.	RN20
RN22	Caso o usuário insira o mesmo estoque e mesmo tipo de oferta o sistema deve contabilizar a quantidade inserida no livro que já existe em estoque.	RN18
RN23	A tela de cadastro, visualização, edição e exclusão de desejo deve conter o campo de: IdLivro (Referente ao livro). Quantidade (Será obrigatório, do tipo inteiro). IdDapessoa (Faz referência a pessoa). DataCriacao (Será obrigatório, do tipo DateTime).	RF08
RN24	Caso a quantidade do desejo seja zero o livro não deve aparecer no estoque do usuário, porém deve estar cadastrado.	RN23
RN25	Caso o usuário insira o mesmo desejo o sistema deve contabilizar a quantidade inserida no livro que já existe em estoque.	RN23
RN26	A tela de cadastro, visualização, edição e exclusão de oferta do grupo deve conter o campo de: IdEstoque (Referente ao estoque). IdTipoOferta (Referente ao tipo oferta). Valor (Não será obrigatório, do tipo decimal). Quantidade (Será obrigatório, do tipo inteiro). IdGrupo (Referente ao grupo).	RF18

RN27	A tela de pesquisa deve ser feita por nome do livro e deve conter um filtro por tipo do livro.	RF21
RN28	O cadastro de sugestão de livros deve conter os campos de: IdPessoa (Referente a pessoa). IdGenero (Referente ao gênero). Quantidade (Será obrigatório, do tipo inteiro).	RF12
RN29	O cadastro de sugestão de livro deve ser feito quando o usuário cadastra um livro em estoque, o sistema soma a quantidade mais um de acordo com o gênero do livro. Deve ser feito também na pesquisa, o sistema soma a quantidade de acordo com o gênero dos 5 primeiros registros encontrados na busca.	RN28
RN30	O Sistema deve permitir login com facebook	RF20
RN31	O Sistema deve conter uma tela de recuperação de senha.	RF19

O modelo relacional (Figura 2) é utilizado no BD, contendo informações importantes que são traduzidas em coleção de relações (tabelas). E o diagrama de classes (Figura 1) é um documento imprescindível para o desenvolvedor de código, pois esse artefato ilustra quais atributos e métodos devem conter nas classes, além de mostrar o relacionamento entre elas.

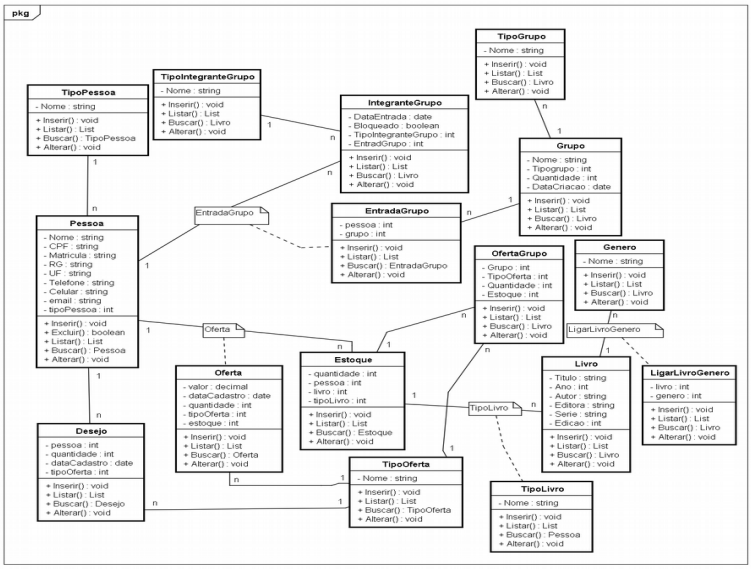


Figura 1 – Diagrama de classes do IFLivros.

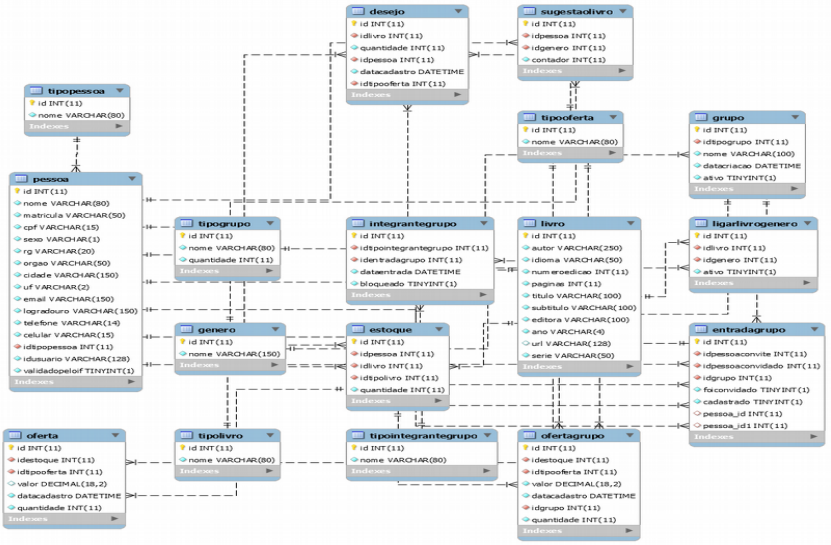


Figura 2 – Modelo relacional do IFLivros.

Diagramas de caso uso documentam a interação do sistema com o usuário do modo geral, sem se preocupar muito com a parte técnica. Um caso de uso requer um cenário, onde os eventos acontecem, um usuário (pode ser um administrador ou um

usuário comum) e a comunicação entre os eventos e o usuário. O caso de uso referente aos usuários *off-line* é apresentado na Figura 3. As funções disponíveis no momento em que o usuário acessa o sistema são: cadastrar, recuperar senha e *logar* no sistema. Quando se cadastra uma pessoa, o sistema inclui um usuário no banco de dados.

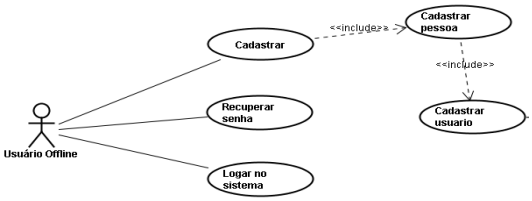


Figura 3 – Caso de uso *off-line*

A Figura 4 mostra os casos de uso referente aos usuários cadastrados. Depois de cadastrado, o usuário tem alguns recursos disponíveis: manter perfil, manter estoque, manter desejo, manter ofertas, manter grupo e pesquisar. Manter significa alterar, criar, excluir e adicionar algo, igualmente conhecido como CRUD (*Create, Read, Update, Delete*).

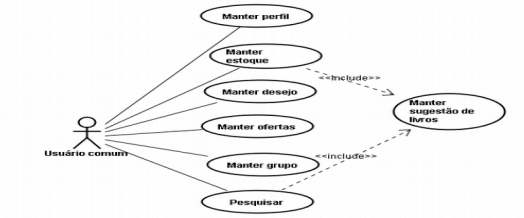


Figura 4 – Caso de uso: usuário comum

Por fim, a Figura 5 mostra os casos de uso referentes aos usuários administradores do sistema. Um administrador tem as permissões de usuário comum com algumas vantagens que são: manter tipo pessoa, manter tipo livro, manter gênero e manter tipo grupo. Isso significa que somente um administrador poderá cadastrar um livro ou o gênero de um, por exemplo, essa regra de negócio traz mais organização e segurança para o sistema, pois impossibilita que aja gêneros e livros inexistentes.

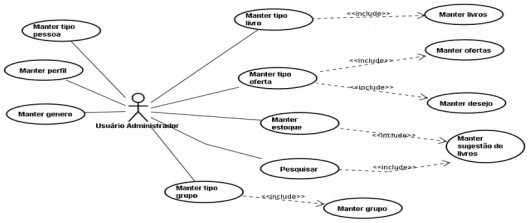


Figura 5 – Caso de uso: Usuário Administrador

5. Resultados

Baseado no conhecimento adquirido no presente estudo foi possível construir o sistema proposto. Sendo assim, este capítulo apresenta às telas do sistema seguindo as regras de negócio. O sistema encontra-se *online* no endereço www.iflivros.com.br.

A Figura 6 representa o resultado da regra de negócio RN01 e do requisito funcional RF01.

Figura 6 – Tipo dos Grupos.

A Figura 7 representa o resultado da regra de negócio RN02 e do requisito funcional RF02.

Figura 7 – Tipo de integrantes do grupo.

A Figura 8 representa o resultado da regra de negócio RN03 e do requisito funcional RF03.

Figura 8 – Tipo dos livros.

A Figura 9 representa o resultado da regra de negócio RN04 e do requisito funcional RF04.

Figura 9 – Tipos de oferta.

A Figura 10 representa o resultado da regra de negócio RN05 e do requisito funcional RF05.

Figura 10 – Tipos de pessoa.

A Figura 11 representa o resultado da regra de negócio RN06 e do requisito funcional RF11.

Figura 11 – Gêneros.

A Figura 12 representa o resultado das regras de negócio RN07, RN08, RN9 e dos requisitos funcionais RF06, RF07, RF08 e RF19.

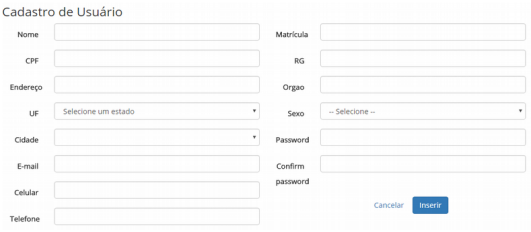


Figura 12 – Pessoas e usuário.

A Figura 13 representa o resultado da regra de negócio RF10 e do requisito funcional RF07.

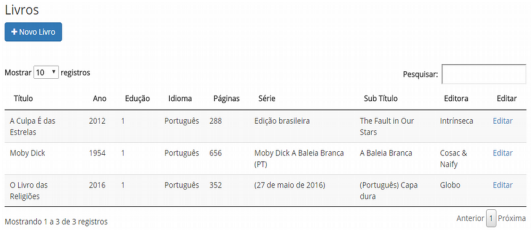


Figura 13 – Livros.

A Figura 14 representa o resultado das regras de negócio RN11 es RN12.

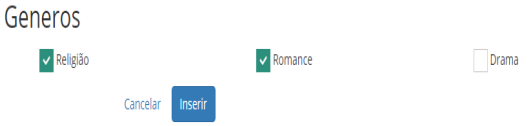


Figura 14 – Ligar livro gênero.

A Figura 15 representa o resultado da regra de negócio RN13 e requisito funcional RF09.

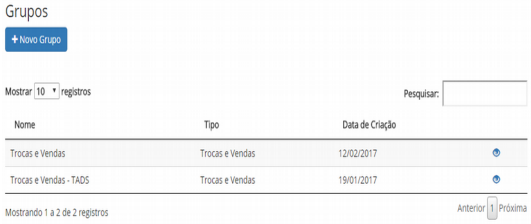


Figura 15 – Grupos.

A Figura 16 representa o resultado das regras de negócio RN14, RN16 e do requisito funcional RF16.

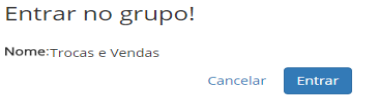


Figura 16 – Entrada no grupo.

A Figura 17 representa o resultado da regra de negócio RN15 e do requisito funcional RF17.

Figura 17 – Integrantes do grupo.

A Figura 18 representa o resultado da regra de negócio RN17, RN18, RN19 e do requisito funcional RF10

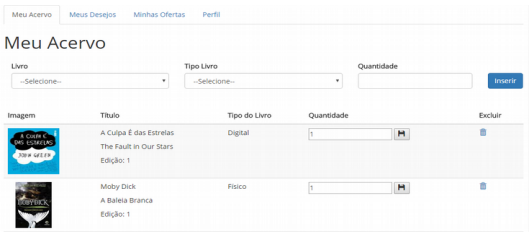


Figura 18 – Estoque.

A Figura 19 representa o resultado das regras de negócio RN20, RN21, RN22 e do requisito funcional RF14.

Figura 19 – Ofertas.

A Figura 20 representa o resultado das regras de negócio RN23, RN24, RN25 e do requisito funcional RF08.

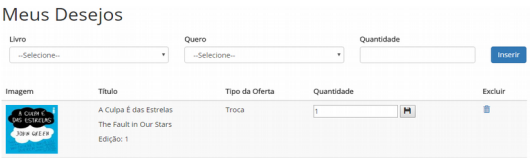


Figura 20 – Desejos.

A Figura 21 é referente ao resultado da regra de negócio RN26 e do requisito funcional RF18.

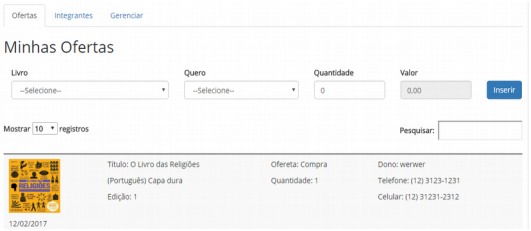


Figura 21 – Ofertas grupos.

A Figura 22 é referente ao resultado da regra de negócio RN27 e do requisito funcional RF21.

Figura 22 – Pesquisa.

A Figura 23 é referente ao resultado da regra de negócio RN30 e RF20.

Figura 23 – Login.

A Figura 24 é referente ao resultado da regra de negócio RN31 e RF19.

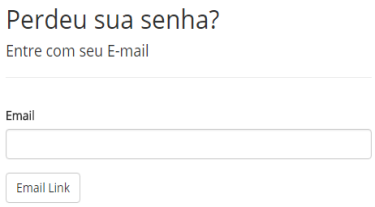


Figura 24 – Recuperação de senha.

6. Conclusão

Neste trabalho foi desenvolvido um sistema *web* que visa atender os alunos e professores do Instituto Federal de Goiás Câmpus Luziânia. O sistema se chama IFLivros e desempenha o papel de gerenciar a coletânea de livros dos usuários. O Sistema de Gerenciamento de Livros – IFLivros atendeu aos objetivos propostos, que foram: ser disponibilizado *online* para os discentes e docentes da instituição, ter as funcionalidades de cadastramento, catalogação e consultas a livros, além de facilitar a venda, troca ou empréstimo de livros por docentes e discentes, diminuindo a busca por cópia ilegal de material didático

Com a criação desse sistema, a acumulação de livros desnecessária pode ser resolvida, tendo em vista que todos os alunos e servidores do IFG Câmpus Luziânia dispõem de uma plataforma *online* gratuita capaz de facilitar a venda, troca ou empréstimo de livros pessoais, e consequentemente diminuir a pirataria. O sistema foi pensado para atender as necessidades do IFG Câmpus Luziânia, porém ele é ampliável para outras instituições e até mesmo para o público em geral, no entanto, seria inviável no momento, pois o projeto não possui investimento financeiro para expandi-lo a nível mundial.

Para trabalhos futuros sugerimos a ampliação do sistema para uso em outras instituições a nível global. Sendo necessárias algumas modificações, como adaptações de linguagem caso seja expandido para fora do Brasil (e nesse caso, a adição de

exemplares pelo administrador do sistema seria muito mais vasta), contratação de servidor mais robusto para atender um publico maior sem perder desempenho, oferecer opções de pagamento e de frete e funcionalidades novas como *chat online,* adicionar relatórios estatísticos para utilização dos dados armazenados pelo sistema, como por exemplo, qual o livro ou gênero de livro mais procurado e compartilhado.

Referencias

ASTAH. (2017). *Astah*. Disponível em: http://astah.net/editions/community. Acesso em: 10 dez. 2016

DOS SANTOS, C. (2012). *Entity Framework Code First–Migrations*. Disponível em: https://msdn.*Microsoft*.com/pt-br/library/jj591330.aspx. Acesso em: 03 fev 2017

DYKSTRA, T. (2015). *Getting Started with Entity Framework 6 Code First using MVC 5*. Fonte: asp.net: https://www.asp.net/mvc/overview/getting-started/getting-started-with-ef-using-mvc/creating-an-entity-framework-data-model-for-an-asp-net-mvc-application. Acesso em: 10 out. 2016

FOWLER, M. (2005). *UML Essencial – Um breve guia para a linguagem-padrão de modelagem de objetos 3a Edição.*

MICROSOFT. (2015). *Microsoft*. VisualStudio. Disponível em: https://www.visualstudio.com/pt-br/vs/community/. Acesso em: 15 dez. 2016

MICROSOFT. (2017a). *TFS*. Fonte: TFS. Disponível em: https://www.visualstudio.com/pt-br/tfs/. Acesso em: 10 jan. 2017

MICROSOFT. (2017b). Visão geral do ASP.NET MVC. Disponível em: https://msdn.microsoft.com/pt-br/library/dd381412(v=vs.108). Acesso em: 10 fev. 2017.

ORACLE. (2017). *MySQL 5.7 Reference Manual*. Disponível em: https://dev.mysql.com/doc/refman/5.7/en/. Acesso em: 10 out. 2016

OTTO, M., & THORNTON, J. (2017). *Bootstrap*. Disponível em: http://getbootstrap.com.br/. Acesso em: 10 out. 2016

PRESSMAN. (2011). *Engenharia de Software uma abordagem profissional 7ª Edição.*

RASTOGI, P., ANDERSON, R., DYKSTRA, T., & GALLOWAY, J. (2013). *Introduction to ASP.NET Identity*. Disponível em: https://www.asp.net/identity/overview/getting-started/introduction-to-aspnet-identity. Acesso em: 15 dez. 2016

SANCHEZ, F., & ALTHMANN, M. (2013). *Desenvolvimento Web com Asp.Net MVC.*

SOMMERVILLE. (2007). *Engenharia de Software 8ª Edição.* São Paulo: Pearson.

SOMMERVILLE. (2011). *Engenharia de Software – 9ª Edição.*

THE JQUERY FOUNDATION. (2017). *jqueryui*. Disponível em: https://jqueryui.com/ Acesso em: 10 jan. 20

Uma abordagem de entrega contínua para empresas de desenvolvimento de *software*

Rodolfo A. Bueno[1], **Adailton F. Araújo**[2]

[1]Pontifícia Universidade Católica de Goiás (PUC Goiás) – Goiânia – GO

[2]Instituto de Informática – Universidade Federal de Goiás (UFG)– Goiânia – GO

rodolfo.bueno13@gmail.com, adailton@inf.ufg.br

Abstract. *The growing demand for software products and high market competition are causing software developers to deploy to be able to deliver faster and faster deliveries to their customers. The present study proposes an approach to work with continuous deliveries, consisting of a software development process and the association of yours activities with a set of tools frameworks that will serve as a model for the implementation of this process . The approach was evaluated through a maturity model and deployed in a small company of software development. It was concluded that the proposed approach has an intermediate level of maturity, Level 1 of a scale that can vary between -1 and 3, ideal for organizations that wish to change their culture of manual deliveries. The implementation of the approach in organization that had a costly software delivery process generate an effective gain of 99 % with time spent with software delivery and this increasing significantly the team productivity. With short delivery scopes, there was a reduction in risk and an increase in product quality.*

Resumo. *A crescente demanda por produtos de software e a alta concorrência do mercado estão fazendo com que as empresas de desenvolvimento de software se desdobrem para conseguir realizar entregas cada vez mais rápidas para seus clientes. O presente estudo propõe uma abordagem para se trabalhar com entregas contínuas, composta por um processo de desenvolvimento de software e a associação de suas atividades com um conjunto de ferramentas/frameworks que servirão de base para implantação desse processo. A abordagem foi avaliada por meio de um modelo de maturidade e implantada em uma empresa de pequeno porte da área de desenvolvimento software. Concluíu-se que a abordagem proposta possui um nível de maturidade intermediário, Nível 1 de uma escala que pode variar entre -1 e 3, ideal para o contexto de organizações que desejam mudar sua cultura de entregas manuais. Com a implantação da abordagem em uma organização que possuía um processo oneroso de entrega de software foi constatado um ganho efetivo de 99% no tempo gasto com entrega de software, aumentando significamente a produtividade da equipe. Com escopos curtos de entregas, houve uma redução dos riscos e aumento da qualidade do produto.*

1. Introdução

Com o avanço da tecnologia, *software* tornou-se algo essencial na vida das pessoas, aumentando a demanda de produtos e fazendo com que novas empresas sejam cri-

adas todos os dias, resultando em um rápido crescimento da indústria de *software* [Caffery et al. 2007, Argento and Hamilton 2009, Sato 2013].

O alto dinamismo das necessidades do mercado exige que as empresas se desdobrem para descobrir como transformar uma boa ideia em um sistema e entregá-lo aos usuários o quanto antes [Humble and Farley 2014].

Mesmo com a evolução das metodologias de desenvolvimento de *software* o processo de transformação de ideias em código ainda envolve diversas atividades como: levantamento de requisitos, *design*, arquitetura, implementação, teste e entrega [Sato 2013].

Os métodos ágeis de desenvolvimento de *software* surgiram com o objetivo de organizar tais atividades para que as mesmas fossem realizadas no menor tempo possível, sem perder qualidade, gerando um *feedback* rápido do cliente e tornando a sua satisfação como uma das prioridades da empresa de desenvolvimento de *software* [Rasmusson 2010].

É observado que, em muitos projetos, o processo de entrega de uma versão de *software* ainda é bastante manual e complexo, cada atividade deve ser executada perfeitamente para que o *software* funcione corretamente [Humble and Farley 2014].

Com um processo complexo, a tendência é que a frequência de entregas diminua pois elas estarão sempre sobrecarregadas com uma grande quantidade de mudanças. Em consequência disso, aumenta-se o risco de que a versão entregue não funcione conforme o esperado pelo cliente[Sato 2013].

A entrega contínua surgiu como uma alternativa para minimizar esses problemas. Pode-se definir entrega contínua como uma disciplina de desenvolvimento que sugere a construção de um *software* pela premissa de que ele possa ser liberado para produção, com segurança, a qualquer momento [Fowler 2013a]. Isso será possível a partir do momento em que toda mudança realizada no código-fonte for submetida em um ambiente que seja possível executar testes automatizados, analisar qualidade do código e gerar um indicador de status do produto [Humble and Farley 2014].

Vários autores citam os benefícios de se implantar um ambiente de entregas contínuas, são alguns deles: aumento da produtividade e da autonomia da equipe; redução de riscos; aumento na qualidade do produto; entregas rápidas e estáveis; *feedback* constante e aumento da satisfação do cliente [Humble and Farley 2014, Chen 2015, Fowler 2013a, Sato 2013].

Humble and Farley (2014) introduzem em seu livro o conceito de entrega contínua, mostrando todo o seu ecossistema e explicando de forma detalhada como unificar as atividades de desenvolvimento, entrega e operações em um contexto único, a fim de alcançar os benefícios dessa metodologia. Sato (2013) realizou uma abordagem mais técnica, demonstrando como criar e configurar todo um ambiente de entregas contínuas do zero através de um conjunto de ferramentas e *frameworks* pré-estabelecidos.

Diversos artigos discorrem sobre entregas contínuas, exibindo seus benefícios e também as dificuldades em sua implantação [Chen 2015, Gmeiner et al. 2015, Neely and Stolt 2013], entretanto, eles não apresentam orientações estruturadas que possam auxiliar, de maneira objetiva, os profissionais de desenvolvimento de *software*, na aplicação prática dos conceitos de entregas contínuas.

Diante desse fato, este trabalho apresentará uma abordagem de entregas contínuas de *software* que é composta pela especificação detalhada de um processo de desenvolvimento de *software* focado nas atividades de entrega contínua e a associação das atividades do processo com algumas ferramentas/*frameworks* necessários para a sua execução. Gerando uma base de conhecimento para as empresas de desenvolvimento de *software* que desejam implantar esses conceitos.

Assim, o presente estudo está organizado da seguinte forma. A Seção 2 descreve a metodologia de pesquisa. A Seção 3 apresenta uma revisão não sistematizada da literatura. Com essa base, um modelo de processo de entrega contínua de *software* será apresentado na Seção 4 e disposto para avaliação na Seção 5. Por fim, a Seção 7 sintetiza as contribuições e conclusões deste trabalho e indica direções para trabalhos futuros.

2. Metodologia de Pesquisa

Os resultados descritos neste artigo originaram-se a partir das quatro etapas a seguir:

Etapa 1 - Revisão não sistematizada da literatura: nesta etapa foram realizadas buscas nas bases de dados eletrônicas IEEE Xplore Digital Library, ACM Digital Library, Biblioteca Digital Brasileira de Teses e Dissertações com os descritores: Entrega de Versão de *software*, Entrega Contínua, Integração Contínua, *Pipeline* de implantação, DevOps, Desenvolvimento ágil e seus correspondentes em inglês. Para serem selecionados, os artigos deveriam referenciar em algum ponto o processo de entrega de versão de *software*. Para identificar o potencial de relevância dos artigos obtidos foi realizada uma leitura do título e dos resumos e, no caso de dúvidas sobre a possível utilização, o artigo foi lido integralmente e então incluído ou excluído.

Etapa 2 - Desenvolvimento do processo: nesta etapa, as informações coletadas nos trabalhos selecionados na Etapa 1, foram utilizadas para desenvolver um processo de entrega de *software* contínuo (Seção 4), utilizando a notação BPMN (*Business Process Modeling Notation*). Além de um fluxograma que representa a dinâmica das atividades que devem ser executadas para atingir os objetivos da entrega contínua, foram levantados os papeis reponsáveis pela execução das atividades, a dependência entre as atividades e um conjunto de ferramentas/*frameworks* de apoio a execução das atividades.

Etapa 3 - Avaliação da maturidade do processo: nesta etapa, o nível de maturidade do processo desenvolvido na Etapa 2 foi avaliado (Seção 5) com base no modelo desenvolvido por Humble and Farley (2014).

Etapa 4 - Implantação do processo: nesta etapa, o processo desenvolvido na Etapa 2 foi implantado (Seção 6) em uma empresa de desenvolvimento de *software*, o processo foi utilizado em um projeto real e foram coletados resultados da implantação.

3. Entrega contínua

Para alcançar os benefícios da entrega contínua de *software* é necessário que o processo torne-se automatizado e frequente. Isso possibilitará que as versões do *software* fiquem mais enxutas, os riscos associados à entrega diminuam e um rápido *feedback* do resultado possa ser recebido [Duvall et al. 2007, Humble and Farley 2014]. Ao mesmo tempo exigirá que a equipe encontre uma forma de garantir que o *software* esteja funcionando corretamente a todo momento e com o máximo de qualidade possível [Humble and Farley 2014].

Uma das formas para conseguir essa garantia é através da integração contínua (IC) do *software*, essa pratica consiste na integração do código-fonte, com o processo de *build* de executáveis e a execução de testes automatizados durante o processo de construção do *software* [Fowler 2013a].

A IC originou-se como uma das práticas da metodologia ágil XP (*extreme programming*) [Meyer 2014], representando uma mudança de paradigma pois exige que toda a equipe envolvida na construção do *software* tenha ciência de que quanto mais tempo eles ficarem sem integrar suas mudanças, maior o risco de gerarem conflitos [Sato 2013]. [Beck 1999] reforça que se a integração frequente do código é boa, então a mesma deverá ser realizada todo o tempo.

Fowler (2013a) frisa que não existe uma receita fixa para a implantação de integração contínua, muitas coisas dependerão da natureza do ambiente organizacional e também das pessoas envolvidas, porém, Humble and Farley (2014) estabeleceram algumas premissas básicas para a criação do ambiente:

O **controle de versão** é uma das atividades do processo de gerenciamento de liberação e entrega de *software* que por sua vez faz parte da Gerência de Configuração de *software* (GCS) [IEEE 2005]. A GCS refere-se ao processo pelo qual é possível armazenar, identificar, modificar e recuperar artefatos relevantes ao produto [Humble and Farley 2014].

Através do controle de versão é possivel guardar múltiplas versões de seus arquivos, de modo que, mesmo modificando algum deles, ainda será possível o acesso a versões anteriores. Esse é um dos principais mecanismos de colaboração entre os envolvidos na construção e entrega de um *software* [Humble and Farley 2014].

Além da atividade de controle de versão, o processo de gerenciamento de liberação e entrega de *software* tambem é composto pelo controle de modificações e pelo gerenciamento de construção [Murta 2006].

O objetivo do controle de mudança é o de manter a integridade dos artefatos do produto em todos os seus estados [IEEE 2005]. Possibilitando assim, o acompanhamento e o gerenciamento das modificações que serão realizadas durante a construção/manutenção do *software* [Sales et al. 2009].

O **processo automatizado de *build*** é muito mais do que simplesmente a compilação de arquivos [Duvall et al. 2007]. O *build* consiste também no gerenciamento de dependências, execução de testes automatizados, análise estática do código, análise de cobertura de código, geração de documentação e implantação. [Humble and Farley 2014, Sato 2013].

Todo o processo de *build* deve ser realizado de forma automática, seja por linha de comando, por *scripts* ou por um *software* específico de IC [Humble and Farley 2014].

Duvall et al. (2007) apresenta um cenário (Figura 1) contendo os componentes envolvidos no processo de IC através de um servidor.

Um servidor de IC é responsável por monitorar/observar o servidor de versionamento de código e iniciar um *build* do projeto toda vez que detectar um novo *commit* [Sato 2013].

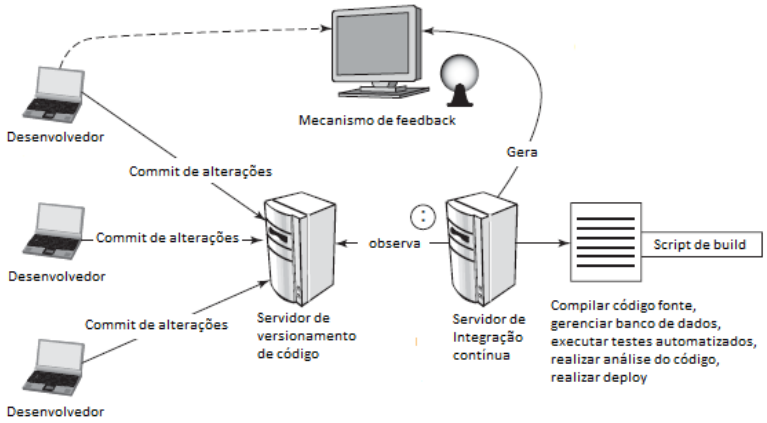


Figura 1. Componentes de Integração Contínua

Existem duas saídas para um *build* executado pelo servidor de IC, a primeira é de falha na execução, resultando no aviso imediato do problema aos responsáveis pelo código; e a segunda é de sucesso na execução, tornando aquele *commit* uma versão candidata a ser implantada em um servidor de testes, homologação ou produção [Fowler 2013b, Humble and Farley 2014].

Ao realizar a automação do processo de compilação, testes e implantação, pode-se dizer que foi criado um *pipeline* de implantação, ou seja, foi criada uma maneira automatizada do processo de levar um *software* do controle de versão até os usuários através de uma forma tão simples quanto apertar um botão [Humble and Farley 2014].

A automatização dos testes garante o *feedback* instantâneo a respeito da saúde do sistema, deixando assim a equipe mais segura para realizar alterações no código [Bernardo and Kon 2008]. Os testes deverão ser desenvolvidos para serem de rápida execução, confiáveis e totalmente acessíveis [Neely and Stolt 2013].

Para atingir uma alta qualidade do *software* as atividades de testes devem se tornar uma prioridade e de responsabilidade de todos os envolvidos na construção do *software*, sendo praticadas durante todo o seu ciclo de vida [Humble and Farley 2014].

Com os testes automatizados e o rápido *feedback* proprocionado pelo ambiente de IC, a quantidade de tempo gasto com a identificação e correção de erros diminuirá [Bernardo and Kon 2008]. O custo com correção cresce exponencialmente dez vezes para cada estágio que o projeto avança. Ou seja, defeitos encontrados na fase de desenvolvimento são mais baratos de serem corrigidos do que aqueles encontrados com o *software* em produção [Myers et al. 2011]. Porém, mesmo com este aumento de custo, muitas empresas ainda negligenciam a fase de teste [de Souza and Gasparotto 2013].

É importante lembrar que mesmo que se tenha um ambiente automatizado, os testes manuais também têm o seu valor no processo. Com uma automação de teste rigorosa cenários típicos estarão garantidos, deixando os testadores com mais tempo para poder aplicar a sua experiência e criatividade em testes que buscam caminhos raros/incomuns de execução e reações suspeitas do sistema [Gmeiner et al. 2015].

A **aceitação da equipe** inicia-se com a compreensão de que IC é uma prática

e não uma ferramenta e que o seu sucesso depende muito do conjunto de ferramentas selecionadas mas também exige muita disciplina da equipe. É de extrema importância o envolvimento de todos desde o início do processo. Uma boa comunicação entre a equipe é essencial para a entrega rápida de *software* de qualidade [Abdul and Fhang 2012, Humble and Farley 2014].

O time deverá incorporar essa nova filosofia de trabalho e tornar o monitoramento das *builds* do sistema uma prioridade, caso algum erro aconteça os responsáveis devem parar tudo que estão fazendo para realizar a correção [Meyer 2014, Humble and Farley 2014]. Possibilitando assim, que o restante da equipe continue trabalhando e integrando as suas mudanças no ambiente de desenvolvimento normalmente [Abdul and Fhang 2012].

A implantação de IC garante que as pessoas envolvidas no desenvolvimento de um sistema possam atingir seus objetivos em conjunto e com um nível de qualidade/controle muito mais alto do que seria possível sem ela [Humble and Farley 2014].

Com a aplicação dessas três premissas será possível alcançar os benefícios da implantação de um ambiente de IC. Humble and Farley (2014), Duvall et al. (2007) e Fowler (2006) destacam alguns deles: redução de erros e riscos; garantia de qualidade do produto; flexibilidade de implantação; redução de processos manuais repetitivos; entregas confiáveis e aumento na visibilidade do projeto.

4. Processo de entrega contínua

A entrega de *software* tornou-se algo bastante discutido nos dias atuais devido ao alto dinamismo e concorrência do mercado de desenvolvimento de *software*. Um processo bem definido e estruturado pode garantir que a empresa esteja sempre entregando valor ao seu cliente.

Os benefícios em se implantar práticas de entregas contínuas são de consenso entre os diversos autores que escrevem sobre o assunto [Humble and Farley 2014, Chen 2015, Abdul and Fhang 2012, Sato 2013]. Visando estabelecer um forma de alcançar esses benefícios o presente trabalho desenvolveu uma abordagem de entregas contínuas de *software* que será apresentada nesta Seção.

Para que seja possível a implantação de um ambiente de entregas contínuas, é de extrema relevância que as seguintes premissas sejam atendidas. Primeiramente, todos os envolvidos na criação do produto devem ter ciência que são responsáveis pelo processo e que a qualidade do produto é uma prioridade desde o início do projeto[Humble and Farley 2014]. Por fim, tudo o que for necessário para compilar, configurar, implantar, testar e entregar o *software* deverá ser centralizado em um sistema de controle de versão [Humble and Farley 2014, Fowler 2006].

As Figuras 2, 3 e 4 ilustram o processo proposto, especificando as atividades que devem ser executadas e suas dinâmicas de execução. As descrições detalhadas de cada atividade também serão apresentadas ao longo desta Seção e terão o seu nome destacado em negrito para servir de referência com o processo proposto. Ao final são apresentadas ferramentas de apoio ao processo proposto.

O processo de entrega de *software* inicia-se através de uma **nova demanda de desenvolvimento**, que pode conter novas implementações e/ou correções de defeitos. Com

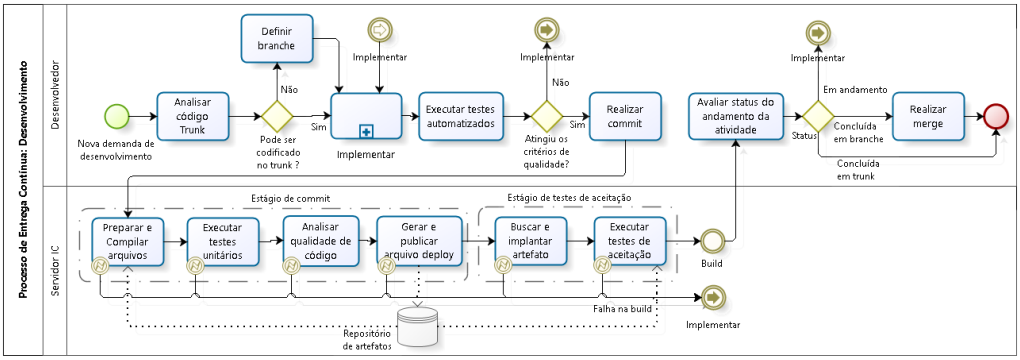


Figura 2. Processo de desenvolvimento

o pacote de necessidades em mãos a equipe deve **Analisar o código do *trunk/master*** do projeto para saber se não está sendo desenvolvido algo que impacte diretamente em uma possível entrega dessa demanda.

Humble and Farley (2014) reforçam que o ideal é que tudo seja desenvolvido no *trunk/master* e definam estratégias para alcançar esse objetivo, como: esconder a funcionalidade até sua finalização; realizar mudanças incrementalmente deixando o código sempre apto a ir para produção; criar uma abstração do código que será alterado. Caso não seja possível realizar nenhuma das estratégias será necessário **Definir uma *branche*** (cópia indepentente de um determinado ponto do *trunk/master*) para a implementação da demanda.

Com a definição do local de implementação (*trunk/master* ou *branche*) inicia-se o subprocesso de implementação. Na Figura 3 é possível ver que ele é composto de quatro atividades que são executadas em paralelo.

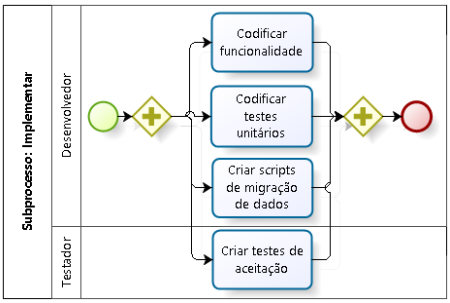


Figura 3. Subprocesso: Implementação

Na atividade de **Codificar funcionalidade** o desenvolvedor irá codificar a demanda. Por meio da atividade **Criar *scripts* de migração de dados** deverão ser criados *scripts* que controlarão o estado da base de dados do produto. Esses *scripts* deverão ser configurados para que seja possível criar uma base do zero ou atualizar uma base existente para versões superiores de forma automática.

As atividades de **Codificar testes unitários** e **Criar testes de aceitação** são de extrema importância para o restante do processo. Com essa automação de testes o *build* do projeto se tornará auto-testável garantindo assim maior qualidade do produto [Fowler 2006, Duvall et al. 2007].

Os desenvolvedores são responsáveis por codificarem os testes unitários e os testadores pelos testes de aceitação. Mesmo com essa separação, deve ser de consenso de toda a equipe que todos devem trabalhar em conjunto para alcançar uma boa qualidade do produto. Deverão ser realizados o máximo de testes automatizados possíveis para garantir a confiabilidade e a repetitividade do processo [Humble and Farley 2014].

Durante a execução dessas quatro atividades os envolvidos deverão realizar *commit* e *update* do código com frequência, encurtando assim possíveis problemas com integração dos seus trabalhos [Fowler 2006].

Antes de se realizar um *commit* é necessário que seja feito um *update* primeiro no repositório de código. Com o código atualizado deve-se **Executar os testes automatizados** a fim de validar se a *build* permanece estável e se ela alcançou os critérios de qualidade estabelecidos ao projeto.

Diversos critérios podem ser estabelecidos para um projeto, como por exemplo a definição de um percentual mínimo de cobertura de código pelos testes automatizados. Com o uso de ferramentas de análise estática de código diversos outros critérios podem ser aplicados, como: duplicidade de código, complexidade, atributos não utilizados, violação de boas práticas e etc.

Após ter a certeza que a *build* permanece estável o *commit* poderá ser realizado. O servidor de IC estará observando o repositório de código e assim que enxergar a mudança iniciará um processo de *build* da aplicação.

O servidor de IC executará uma série de atividades por meio da ferramenta de automação de *build* e é de responsabilidade dos desenvolvedores manter esse processo o mais rápido possível [Humble and Farley 2014].

Esse processo de *build* é o primeiro passo de um *pipeline* de implantação. O servidor de IC é responsável por disponibilizar uma forma de configuração de *pipeline* totalmente dinâmica e adaptável as necessidades do projeto.

Os passos **Estágio de *commit*** e **Estágio de testes de aceitação** são obrigatórios para alcançar resultados satisfatórios. Os demais passos referentes a testes específicos serão configurados de acordo com a necessidade e as características da empresa.

O primeiro passo do *pipeline* de implantação é o **Estágio de *commit***, ele é iniciado através da atividade de **Preparar e Compilar arquivos**, essa preparação consiste na sincronização do código com o servidor de versionamento e na gerência das dependências de bibliotecas de terceiros com o repósitorio de artefatos. Após a compilação, o servidor de IC deverá **Executar testes unitários**, **Analisar a qualidade do código** e por fim **Gerar e publicar o arquivo *deploy*** no repositório de artefatos.

Ao realizar a atividade de **Analisar a qualidade do código** o servidor de IC deve gerar um relatório de sua execução e este deverá estar sempre disponível aos envolvidos do projeto, possibilitando assim que todos vejam como está a qualidade do produto.

O segundo passo é o **Estágio de testes de aceitação**, nele o servidor de IC deverá **Buscar e implantar o artefato** gerado no estágio anterior e em seguida **Executar os testes de aceitação**.

Caso aconteça alguma **falha na *build*** a equipe deverá ser comunicada, voltando a

execução para o subpresso **Implementar** e a partir desse momento a prioridade de todos se torna a estabilização da *build*, ou seja, todos os erros reportados deverão ser corrigidos imediatamente [Duvall et al. 2007, Fowler 2006].

Caso o ***build* esteja OK** o artefato gerado no estágio de *commit* será marcado com status de versão candidata. Isso quer dizer que a versão em questão possui um código funcional e apto a ser implantado em um servidor de testes, homologação ou produção [Humble and Farley 2014].

Com esse informativo de sucesso o desenvolvedor deverá **Avaliar status do andamento da atividade** para definir seus próximos passos. Caso a atividade ainda esteja em andamento então ele voltará para o subpresso **Implementar**. Caso tenha sido concluída, o processo de desenvolvimento da demanda será finalizado. Lembrando que se a codificação tiver sido realizada em um *branche* o responsável deverá **Realizar *merge*** com o *trunk/master* para então finalizar.

Com a execução do processo de desenvolvimento, as fases de compilação e testes de um *pipeline* de implatanção foram finalizadas, faltando somente a fase de implantação do sistema. Para a sua execução, é necessário que o servidor de IC seja configurado para tornar visível e de fácil acesso o histórico de *builds* realizados durante a codificação do produto [Fowler 2006].

Como o processo de implantação de versão nos servidores é muito parecido, um processo genérico foi criado para solucionar essa etapa (Figura 4).

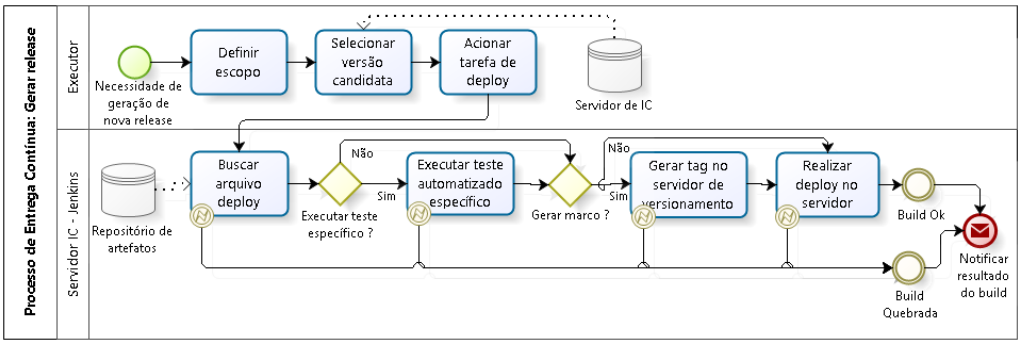


Figura 4. Processo de geração de *release*

O processo inicia-se a partir de uma **Necessidade de geração de *release*** em um dos servidores do projeto. O responsável deverá **Definir o escopo** da entrega de acordo com as políticas da empresa e em seguida **Selecionar a versão candidata** na lista de *builds* do servidor de IC, por fim, deve-se **Acionar tarefa de *deploy*** para o servidor desejado.

O primeiro passo do servidor de IC é **Buscar o arquivo de *deploy*** (versão candidata) no repositório de artefatos. Em seguida serão realizadas duas análises: se será necessário a execução de algum teste automatizado específico para aquele servidor (*smoke tests*, testes de integração e etc.) e se será necessário ou não **Gerar *tag* (marcação) no servidor de versionamento** dessa *release*.

Após as análises, o servior de IC deverá **Realizar o *deploy* no servidor** e independente do resultado (**Build OK** ou **Build Quebrada**) o servidor de IC deverá **Notificar**

o resultado do *build* aos responsáveis pelo processo.

As gerações de *tags* e *branches* deverão ser realizadas de acordo com o padrão de nomenclatura de versões da empresa. Um exemplo que pode ser seguido é o de determinar uma lógica do número da versão do seu produto (1.0.0) e no final acrescentar o identificador da *build* gerada pelo servidor de IC, assim mantendo a rastreabilidade.

A execução do processo deve ser realizada com o apoio de ferramentas e *frameworks* que melhor se encaixem no ambiente de trabalho da empresa. Como existem diversas soluções que resolvem um mesmo problema foi criada a Tabela 1 que faz uma relação de conceito/atividade com ferramenta/*framework* para auxiliar nas escolhas.

Tabela 1. Relação de atividades com ferramentas/*frameworks*

Conceito/Atividade	**Ferramenta/*Framework***
Linguagem de programação	Java, PHP, Ruby, Python, Erlang
Ambiente de desenvolvimento Integrado (IDE)	Eclipse, NetBeans, WebStorm
Banco de dados	PostgreSQL, MySQL, MongoDB, Oracle
Servidor *web*/aplicação	Apache, Jetty, JBoss, Tomcat
Servidor de versionamento	SVN, GIT, Mercurial
Servidor IC	Jenkins, Cruise Control, Bamboo
Automação de *build* e gerenciamento de dependências	Maven, Ant, Gradle
Repositório de artefatos	Artifactory, Archiva, Nexus
Gerência de dados	FlywayDB, dbMigration .NET, Liquibase
Análise de qualidade	SonarQube, Coverage, NCover, PMD, Checkstyle, CodeRush, Findbugs
Testes unitários	Junit, NUnit, Mockito, RSpec
Testes de performance e carga	Jmeter, Ginder, Octoperf
Testes de aceitação	Selenium, Sikuli, CasperJS
Gerência de defeitos	Mantis, TRAC, Jira

5. Avaliação da maturidade do processo

Humble and Farley (2014) criaram um modelo de maturidade para gerência de configuração e entrega de versão de *software* para servir de guia para as empresas que desejam alcançar os benefícios de um ambiente de entregas contínuas.

O modelo analisa seis processos e práticas - gerência de compilação e integração contínua; ambientes e implantação; gestão de entrega de versão e observância; testes; gestão de dados; gerência de configuração - para então, estabelecer cinco níveis que definem o estado de maturidade da empresa. São eles:

- Nível 3 - Otimização: Foco em melhoria de processo
- Nível 2 - Gerido quantitativamente: Processos medidos e controlados
- Nível 1 - Consistente: Processos automatizados aplicados ao longo de todo o ciclo de vida da aplicação.

- Nível 0 - Repetível: Processo documentado e parcialmente automatizado
- Nível -1 - Regressivo: Processos que não podem ser repetidos, pouco controlados e reativos

Com a adoção do processo proposto será possível alcançar o Nível 1 (consistente) de maturidade. Na Tabela 2 são exibidos os itens necessários para alcançar este nível e a relação de cada item com o processo proposto.

Tabela 2. Relação das práticas de maturidade com as atividades do processo proposto

Prática	Processo Proposto
Compilação e testes automáticos são executados para cada check-in.	O processo de build que compõe essas atividades é disparado pelo servidor IC após o desenvolvedor Realizar commit.
Processos completamente automatizados para implantação apertando-se um botão. O mesmo processo é usado para implantar em todos os ambientes.	Servidor de IC configurado para executar a implantação (deploy) nos servidores que a empresa possuir (vide processo de geração de release).
Há reuso de *scripts* e ferramentas.	Configuração das tarefas do servidor de IC, utilização do servidor de versionamento, criação e manutenção de *scripts*.
Existem testes unitários e de aceitação; estes últimos são escritos por testadores. Teste é parte do processo de desenvolvimento.	Atividades do subprocesso Implementar: Codificar testes unitários e Criar testes de aceitação.
Mudanças no banco de dados são executadas automaticamente como parte do processo de implantação.	Os *scripts* são criados na atividade Criar *scripts* de migração de dados e deverão ser executados na incialização do sistema.
Gestão de mudança e processo de aprovação são definidos e seguidos. Políticas de controle de versão são determinadas pelo processo de gestão de mudança.	Foi referenciado pelas atividades de Definir escopo e Gerar tag no servidor no processo de geração de release.
Bibliotecas e dependências são gerenciadas.	Nas atividades do subprocesso Implementar e nos builds executados pelo servidor de IC.

6. Implantação do Processo

A abordagem desenvolvida na Seção 4 foi implantada em uma empresa de desenvolvimento e manutenção de *software* de pequeno porte. Com 28 anos de existência, a empresa possuía naquele momento 15 pessoas atuando diretamente na construção e manutenção dos seus softwares. Desde 2008 várias iniciativas de melhoria de processo e qualidade de software estavam sendo elaboradas e implantadas na área de desenvolvimento.

A empresa havia iniciado um novo projeto cujo objetivo era desenvolver uma plataforma web. A equipe técnica envolvida no projeto era composta de três desenvolvedores, um analista de teste e um analista de infraestrutura, todos com algum tipo de responsabilidade na geração de novas *releases*.

O projeto, que estava no quinto mês de execução, possuia uma rotina onerosa de entrega de *software*. Ao surgir a necessidade de gerar *release* em algum servidor da empresa, um dos desenvolvedores, geralmente o mais sênior, iniciava o processo de entrega de versão. Primeiramente, era realizada a atualização do código-fonte da sua máquina com tudo o que foi desenvolvido pelo time. Em seguida, era criado um script com todas as alterações de banco de dados e era gerado um relatório contendo as atualizações de *frameworks*/bibliotecas. Por último, o arquivo de *deploy* da aplicação era gerado e todos esses artefatos eram enviados para o analista de infraestrutura.

A solicitação entrava em uma fila de atendimento e quando chegava a hora de sua execução o analista de infraestrutura informava a todos os envolvidos que seria necessário pausar a aplicação por um tempo, para que a mesma pudesse ser atualizada. Com a permissão dos envolvidos, o processo era executado seguindo os seguintes passos: pausa do servidor; atualização de ferramentas e bibliotecas; execução dos *scripts* de banco; troca do arquivo de *deploy*; e início do servidor.

Com a liberação da versão, o analista de teste realizava um teste de fumaça no sistema, onde era verificado o funcionamento correto das principais funcionalidades da aplicação. Por fim, o testador emitia alerta a equipe de desenvolvimento caso fosse necessário voltar a versão.

Como pode ser visto, o processo envolvia toda a equipe técnica do projeto, exigindo a execução perfeita de suas atividades e obrigando que todos ficassem em alerta durante o processo, se comunicando muito bem, pois qualquer falha na execução de alguma atividade o sistema ficaria indisponível, prejudicando o andamento do projeto.

Considerando um cenário feliz, onde todas as atividades eram executadas com sucesso, uma nova versão do *software* estaria disponível no servidor da empresa com cerca de 2h de trabalho do desenvolvedor, 1h do analista de infraestrutura e 30 minutos do analista de teste, totalizando em média 3:30h de trabalho.

Para implantação do modelo proposto por este trabalho, os diretores da empresa solicitaram que essa nova abordagem não gerasse nenhum novo custo para o projeto com aquisição de *softwares* de apoio, ferramentas, frameworks e etc. Diante disso, primeiramente, toda uma estratégia de capacitação e mudança cultural foi desenvolvida pelo gerente do projeto, com o intuíto de mostrar a equipe todos os benefícios dessa mudança. Em seguida, a equipe de desenvolvedores realizou diversos estudos e experimentos para determinar quais das ferramentas/*frameworks* propostas pela abordagem melhor se encaixariam ao cenário/premissas do projeto e da empresa.

O ambiente da empresa ficou estruturado da seguinte forma: SVN como servidor de versionamento, Jenkins para servidor de Integração Contínua; Mavem para Gerenciamento de Dependências e Automação de builds; Nexus para Servidor de Repositórios; FlywayDB para Gerência de Dados; SonarQube ficou responsável pela análise de qualidade; JUnit e Mockito para os testes unitários e o Selenium para os testes de aceitação.

Com a implantação do processo e do suporte ferramental, o processo de entrega de *software* que era executado em 3:30 horas passou a ser executado em 2 minutos, sem a necessidade de realização de testes manuais e somente com um clique no Servidor de IC. Foi gerado um ganho efetivo de 99% no tempo gasto com entrega de *software*.

Durante um ano executando esta abordagem, foram executados 970 builds de monitoramento, 95 *deploys* em ambiente de testes, 12 *deploys* em ambiente de homologação e 8 *deploys* em ambiente de produção. A implantação dessa nova abordagem gerou um trabalho inicial de 96h para a adaptação do ambiente, técnicas e processos, porém ao mesmo tempo trouxe a economia de aproximadamente 402 horas que seriam gastas com os 115 *deploys* realizados em um ano. Com a retirada das atividades manuais da equipe técnica a produtividade aumentou de forma significativa. Com a facilidade obtida no processo de entrega foi possível planejar escopos menores de entregas, gerando assim *feedbacks* constantes e consequentemente aumentando a qualidade do produto.

7. Conclusões

As empresas de desenvolvimento de *software* se deparam com ambientes cada vez mais dinâmicos e por isso necessitam realizar entregas cada vez mais frequentes. A maioria das empresas ainda realizam entregas de versões de *software* de maneira manual, onde se perde muito tempo com atividades repetitivas, complexas e que exigem bastante atenção dos envolvidos para que o *software* chegue até o usuário final sem erros. As entregas contínuas de *software* surgiram como alternativa para tornar o processo de entrega de versões confiável e automatizado.

Como contribuição o presente trabalho apresentou uma sistematização dos conhecimentos de entregas contínuas e propôs uma abordagem de desenvolvimento de *software* com foco em entregas contínuas. Esta abordagem é composta pela especificação detalhada de um processo e a indicação de ferramentas de apoio a execução deste processo.

A avaliação empírica mostrou que a abordagem conseguiu atingir um nível consistente de processo, ou seja, Nível 1 de uma escala de -1 a 3. Com isso, pode-se constatar que a abordagem ainda não contempla todas as boas práticas indicadas para este cenário, mas pode ser uma boa alternativa para empresas que pretendem iniciar a mudança de cultura em relação a entrega de *software*.

A implantação da abordagem proposta constatou um ganho efetivo de 99% no tempo gasto com entrega de *software*, saindo de um processo que demandava cerca de 3:30h para subir uma nova versão para um que consiste no clique de um botão e na espera de 2 minutos para a conclusão do processo. Com a eliminação das atividades manuais e repetitivas foi evidenciando um significativo aumento da produtividade da equipe. A segurança e simplicidade do novo processo possibilitou entregas cada vez mais curtas, reduzindo riscos, gerando *feedbacks* constantes e aumentando a qualidade do produto.

Como trabalho futuro pretende-se aplicar a abordagem em outras empreas, com o objetivo de comparar os resultados alcançados e também melhorar a maturidade da abordagem, trabalhando com tópicos que não foram abordados e que levarão a um nível de otimização/melhoria contínua do processo.

Referências

Abdul, F. A. and Fhang, M. C. S. (2012). *Implementing Continuous Integration towards Rapid Application Development*. International Conference on Innovation, Management and Technology Research.

Argento, R. and Hamilton, V. (2009). *Lowering business costs: Mitigating risk in the software delivery lifecycle*. IBM.

Beck, K. (1999). *Extreme Programming Explained*. Addison-Wesley Professional.

Bernardo, P. C. and Kon, F. (2008). *A Importância dos Testes Automatizados: Controle ágil, rápido e confiável de qualidade*. Engenharia de Software Magazine.

Caffery, F., Taylor, P., and Coleman, G. (2007). *Adept: A Unified Assessment Method for Small Software Companies*. IEEE Computer Society.

Chen, L. (2015). *Continuous Delivery: Huge Benefits, but Challenges Too*. IEEE Software.

de Souza, K. P. and Gasparotto, A. M. S. (2013). *A importância da atividade de teste no desenvolvimento de software*. Abepro.

Duvall, P., Matyas, S., and Glover, A. (2007). *Continuous Integration: Improving Software Quality and Reducing Risk*. Addison-Wesley Professional.

Fowler, M. (2006). *Continuous Integration*. http://martinfowler.com/articles/continuousIntegration.html. Acessado em 10/04/2016.

Fowler, M. (2013a). *Continuous Delivery*. http://martinfowler.com/bliki/ContinuousDelivery.html. Acessado em 20/03/2016.

Fowler, M. (2013b). *DeploymentPipeline*. http://martinfowler.com/bliki/DeploymentPipeline.html. Acessado em 12/04/2016.

Gmeiner, J., Ramler, R., and Haslinger, J. (2015). *Automated Testing in the Continuous Delivery Pipeline: A Case Study of an Online Company*. User Symposium on Software Quality, Test and Innovation.

Humble, J. and Farley, D. (2014). *Entrega Contínua: Como entregar software de forma rápida e confiável*. Bookman.

IEEE (2005). *IEEE Standard for Software Configuration Management Plans*. IEEE Computer Society.

Meyer, M. (2014). *Continuous Integration and Its Tools*. IEEE Software.

Murta, L. G. P. (2006). *Gerência de Configuração no Desenvolvimento Baseado em Componentes*. Universidade Federal do Rio de Janeiro.

Myers, G. J., Sandler, C., and Badgett, T. (2011). *The art of software testing*. Wiley, 3 edition.

Neely, S. and Stolt, S. (2013). *Continuous delivery? easy! just change everything (well, maybe it is not that easy)*. Agile Conference.

Rasmusson, J. (2010). *The Agile Samurai: How Agile Masters Deliver Great Software*. Pragmatic Bookshelf.

Sales, E., Reis, C. L., and Reis, R. Q. (2009). *Apoio a Gerência de Configuração de Artefatos de Software integrado a Execução de Processos de Software*. Universidade Federal do Pará.

Sato, D. (2013). *DevOps na prática: entrega de software confiável e automatizada*. Casa do código.

Balanceamento de Carga de Rede: Alto Desempenho e Escalabilidade para Serviços Computacionais

Raimundo Nonato de Araújo Soares Neto[1], Tiago do Carmo Nogueira[2], Cláudio Antônio Amoury Siva[1]

[1]Universidade Federal de Goiás (UFG)
Avenida Esperança s/n, Campus Samambaia
CEP 74690-900 Goiânia – Goiás – Brazil.

[2]Departamento de Informática - Instituto Federal de Mato Grosso (IFMT)

raimundonsneto@gmail.com

Abstract. *This work was developed with the purpose of conducting a study on Network Load Balancing, presenting its general characteristics, benefits, techniques, resources and solutions for mission critical computing systems. In order to do so, a bibliographical research was carried out, of a basic nature, with a qualitative approach, with an exploratory purpose. Initially, it dealt with issues necessary to understand Load Balancing, addressing general concepts and their basic operating principles. Afterwards, a discussion was held on the most relevant points on the subject, presenting techniques and main algorithms implemented for Load Balancing in the LVS - Linux Virtual Server tool. It was concluded that Network Load Balancing is of great relevance for the availability of computational services, in order to maximize the resource utilization in order to allow a better performance avoiding congestion in the network.*

Resumo. *Este trabalho foi desenvolvido com o objetivo de realizar um estudo sobre Balanceamento de Carga de Rede, apresentando suas características gerais, benefícios, técnicas, recursos e soluções para sistemas computacionais de missão crítica. Para tanto, realizou-se uma pesquisa bibliográfica, de natureza básica, de abordagem qualitativa, com finalidade exploratória. Inicialmente tratou-se de assuntos necessários para compreensão do Balanceamento de Carga, abordando conceitos gerais e seus princípios básicos de funcionamento. Posteriormente, realizou-se uma discussão sobre os pontos mais relevantes sobre tema, apresentando técnicas e principais algoritmos implementados para Balanceamento de Carga na ferramenta LVS - Linux Virtual Server. Concluiu-se queo Balanceamento de Carga de Rede é de grande relevância para a disponibilidade de serviços computacionais, visando maximizar a utilização dos recursos de forma a possibilitar um melhor desempenho evitando-se o congestionamento na rede.*

1. Introdução

É notório o crescimento de organizações, tanto públicas como privadas que trabalham com grandes volumes de informações e demais recursos envolvidos; contudo, os problemas de congestionamento da rede tem sido um reflexo deste crescimento. Pode-se dizer que Balanceamento de Carga de Rede pode proporcionar melhoria de desempenho em um

ambiente computacional crítico. Existem diversos benefícios empregados pelas técnicas de Balanceamento de Carga, dentre eles podemos destacar: possibilidade de escalabilidade sem a necessidade de renovação de hardware, melhor utilização de recursos, alta disponibilidade e tolerância a falhas.

Essas vantagens contribuem para um ambiente capaz de remodelar um espaço até então centralizado e crítico, para um novo sistema computacional capaz de absorver todas as requisições de clientes, distribuindo-as de forma coordenada para o conjunto de máquinas que compõem o sistema, visando melhorar o desempenho da prestação dos serviços envolvidos no processo [Costa et al. 2009]. Garantir qualidade e disponibilidade de serviços computacionais tornou-se tarefa essencial, em virtude do grande número de usuários utilizando a rede de computadores simultaneamente [Kopparapu 2002].

Nesse contexto, o objetivo deste trabalho é realizar por meio de uma pesquisa bibliográfica, um estudo sobre Balanceamento de Carga; apresentando suas principais características e funcionalidades; técnicas utilizadas e algoritmos de escalonamento. Para tanto, além da introdução, este artigo está estruturado em mais 4 seções. Na seção 2, apresentam-se informações e características do Balanceamento de Carga de rede. Na seção 3, denominada de Materiais e Métodos, apresentam-se os procedimentos metodológicos adotodados nesta pesquisa. Na seção 4, denominada de resultados e discussões, destacam-se pontos relevantes sobre soluções disponiveis para implantação de Balanceamento de Carga; e ainda, seus principais algoritmos. Finalmente, na seção 5, são realizadas as conclusões e perspectivas de trabalhos futuros.

2. Contextualização Teórica

Nesta seção apresentam-se informações e características básicas do Balanceamento de Carga de Rede para compreensão deste estudo. Para tanto, discute-se: Conceitos gerais (2.1) e Princípios de Funcionamentos Básicos do Balanceamento de Carga (2.2).

2.1. Balanceamento de Carga - Conceitos Gerais

Com o aumento da dependência das organizações por sistemas automatizados e informatizados; disponibilidade e confiabilidade se tornaram cada vez mais desejáveis e necessárias em sistemas de computação [Árabe and Murta 1996]. Um sistema tolerante a falhas e com alta disponibilidade está intrinsecamente ligado a mecanismos de redundância em níveis de hardware e software. Nesse sentido, as técnicas de balanceamento de carga funcionam como um mecanismo preparado tanto para substituir hosts defeituosos em ocorrência de falhas, como também para balancear a carga de trabalho, o que caracteriza a alta disponibilidade de um serviço [Kopparapu 2002].

No balanceamento de carga cada nó integra o cluster de maneira ativa, uma vez que passa a fornecer diversas informações acerca dos recursos existentes; tais informações são transmitidas para o dispositivo de balanceamento, que pode ser uma unidade única ou um grupo de unidades trabalhando em paralelo, responsável pelo monitoramento do status de todos os nós envolvidos, escalonando; dessa forma, a tarefa seguinte que deverá ser executada para um nó do cluster por meio de uma espécie de política de consumo de recursos.

Entre os principais objetivos do balanceamento de carga destacam-se a disponibilidade do sistema e a diminuição do tempo de resposta às requisições dos clientes, pela

diminuição do tempo de execução da aplicação [da Silva Leandro 2005]. Para tanto, as requisições devem ser distribuídas entre os nós para prover aos usuários a melhor qualidade disponível de serviço. Para se conseguir a tarefa de equilibrar o tráfego para um bom balanceamento de carga existem dois integrantes essenciais: a entidade que realiza a tarefa de balanceamento e o algoritmo usado para tomar as decisões necessárias de como será distribuída as requisições entre os servidores que fazem parte do cluster utilizado.

2.2. Princípios de Funcionamentos Básicos do Balanceamento de Carga

Na atualidade, o balanceador de carga normalmente fica alinhado entre o cliente e os hosts que fornecem os serviços desejados pelos clientes; como isto não é uma regra, mas uma boa prática em uma implementação, considera-se que o balanceador de carga está configurado como servidor virtual que aponta para um cluster formado por dois pontos de serviço [Santos et al. 2010]. O servidor virtual é o responsável por receber as requisições, ele é o nó balanceador de carga, que detém uma política de escalonamento configurada.

Para a percepção de um sistema único, faz-se necessário definir na estrutura do ambiente, o responsável por receber as requisições; papel este, desempenhado pelo nó balanceador de carga do cluster. Nesse sentido, para os clientes, suas requisições foram enviadas para um sistema único; contudo, suas requisições serão direcionadas para um ambiente de cluster que integram o sistema de balanceamento de carga. Existem vários métodos para balanceamento de carga; entretanto, a transação básica de balanceamento de carga funciona da seguinte forma e conforme detalhado na figura 1.

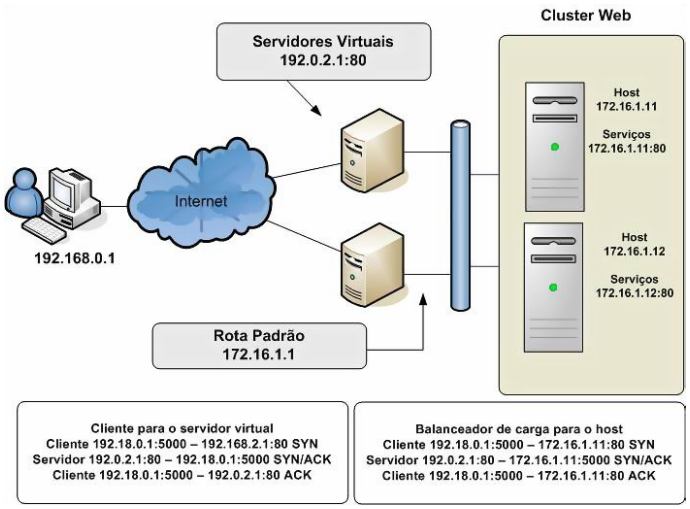


Figura 1. Funcionamento básico - Balanceamento de Carga

De forma geral, quando o nó balanceador de carga e/ou Director recebe uma solicitação do cliente, um dos servidores do grupo processa o pedido. Cada servidor é capaz de lidar com o pedido de forma independente. Se algum servidor não está disponível devido a erro ou devido à manutenção, outros servidores podem ainda servir os pedidos sem serem afetados, garantindo a disponibilidade global do serviço.

Devido a grande flexibilidade, os serviços destinados aos nós podem ser diferentes. Segundo [Batista 2007 2007], enquanto um ou mais nós podem ser destinados

a requisições de processos batch2 , os outros nós podem ser configurados para receber requisições somente de consulta e de curta duração. Além disso, pode-se ter o processamento distribuído uniformemente entre os membros do cluster.

3. Materiais e Métodos

Trata-se de uma pesquisa bibliográfica, de natureza básica, de abordagem qualitativa, com finalidade exploratória. Este tipo de pesquisa ocorre quando elaborada a partir de material já publicado, constituído principalmente de: livros, revistas, publicações em periódicos e artigos científicos, jornais, boletins, monografias, dissertações, teses e demais materias científicos disponíveis na internet [Prodanov and de Freitas 2013].

Para a realização desta pesquisa foram utilizadas como apoio as ferramentas de busca Google Scholar, Springer Link, ACM Digital Library, IEEE Digital Library, a partir de terminologias em Português e Inglês e utilizando strings de busca das seguintes palavras-chave: Balanceamento de Carga de Rede / Network Load Balancing, Clusters de Balanceamento de Carga / Clusters Load Balancing e; Algoritmos para Balanceamento de Carga / Algorithms for Load Balancing.

O intervalo temporal desta pesquisa considerou trabalhos publicados sobre o tema a partir do anos 1990 até a atualidade, visto que entendeu-se como relevante o aspecto histórico e também porque justamente no fim da década de 1990 que tal tema começou a se expadir, dada a preocupação com sistemas de missão crítica.

Já como critério de seleção de artigos considerou-se, para inclusão: relevância, comprovação científica, disponibilidade nas línguas definidas, disponibilidade integral, trabalhos com qualidade acadêmica e em respeito as normas em vigor; e ainda, cujos resultados tenham no título e no resumo relação com o tema deste trabalho. Já como critérios de exclusão: cujos resultados não contemplem ao critério definido nas palavras-chave, disponibilidade parcial, trabalhos não científicos ou sem comprovação; e trabalhos sem a qualidade esperada.

4. Resultados e Discussões

Nesta seção são abordados os principais pontos sobre Balanceamento de Carga, a partir da pesquisa bibliográfica realizada, sendo dividida da seguinte forma: Técnicas de Balanceamento de Carga de Rede (4.1) e Algoritmos para Balanceamento de Carga (4.2).

4.1. Técnicas de Balanceamento de Carga de Rede

As técnicas para Balanceamento de Carga de Rede podem ser feitas de várias maneiras em relação a diferentes recursos que o sistema provém, entre eles: armazenamento, banda de rede, CPU - Central Processing Unit (Unidade Central de Processamento), e demais outros recursos estabelecidos. O balanceamento de carga pode ser dividido em duas grandes categorias:

4.1.1. Balanceamento de Carga via Software

Balanceamento de carga via software é composto por um software especial com a função de balanceamento de carga integrado diretamente no sistema operacional dos servidores

que compõem o cluster. As vantagens desse método de balanceamento de carga incluem velocidade, capacidade de configuração, confiabilidade e redução de custos.

Entre as soluções para técnicas de balanceamento implementadas em software, é importante destacar: NLB - Microsoft Network Load Balancing, esta ferramenta usa uma tecnologia de software que suporta uma arquitetura cluster web, tendo como características principais a segurança de comunicação, roteamento e serviço de acesso remoto [Sato 2007]. A solução LVS - Linux Virtual Server, um projeto open-source que tem como principal objetivo construir um servidor de alto desempenho e altamente disponível para Linux usando a tecnologia de clustering, que fornece altos níveis de escalabilidade, confiabilidade e usabilidade.

4.1.2. Balanceamento de Carga via Hardware

Balanceamento de carga de via hardware especializado, também conhecido como roteador de camada 4/7, é um dispositivo físico que redireciona as conexões para um conjunto de maquinas em uma rede interna [Meffe et al. 2006]. É uma solução que integra comutação e Balanceamento de Carga em um único dispositivo, o que reduz a quantidade de hardware extra que é necessário para implementar o balanceamento de carga combinando as duas funções. A decisão para o Balanceamento de Carga é baseada na análise de uma série de fatores como carga do processador, conexões ativas no servidor, entre outros. Como esses dispositivos são neutros em relação a aplicativos e residem fora do servidor de aplicativos, eles podem fornecer o balanceamento de carga usando técnicas diretas de rede [Meffe et al. 2006].

4.2. Algoritmos para Balanceamento de Carga

Balanceamento de Carga via software é composto por um software especial com a função de balanceamento de carga integrado diretamente no sistema operacional dos balanceadores de carga, que usam diferentes algoritmos para controle de tráfego. O objetivo destes algoritmos é a inteligência em distribuir a carga e/ou maximizar a utilização de todos os servidores no cluster utilizando-se de um método. Os algoritmos de escalonamento decidem qual servidor real será usado para atender uma nova requisição.

Esses algoritmos são um dos fatores mais importantes, pois determinam os modos de escalonamento; ou seja, regulam a maneira como a carga é distribuída entre os nós que compõe o sistema. Existem muitas estratégias disponíveis, dependendo do fornecedor do projeto a ser utilizado na rede. Estes métodos são pautados conforme a evolução da tecnologia pois o mercado oferece vários tipos de opções.

Cada algoritmo tem a sua peculiaridade, tal como, a forma que estes fazem à distribuição de tarefa, qual o critério de atribuir ou não o serviço a um nó do cluster que já tem vários serviços sendo executados, conforme discutido. É valido enfatizarmos que existem também diversos algoritmos utilizados para balanceamento de carga dependendo da ferramenta e/ou solução utilizada para este fim; contudo, a seguir serão apresentados os principais algoritmos implementados na ferramenta LVS - Linux Virtual Server.

4.2.1. RR - Round-Robin

O Round-Robin é um algoritmo de escalonamento mais simples e mais antigo dentre os existentes. É um exemplo típico de algoritmos aleatório. Considerado um algoritmo estático, distribui as requisições que chegam ao Director uniformemente, com igual probabilidade de entrega em qualquer host servidor, independentemente do número atual de conexões ou o tempo de resposta; todos os servidores reais são tratados como equivalentes sem considerar a capacidade ou carga; são atribuídas frações de tempo para cada processo em partes iguais e de forma circular e possuem a mesma prioridade.

Possui uma unidade de tempo, denominada quantum, que é definida pelo sistema operacional, onde determina o período de tempo entre cada sinal de interrupção, com isso, o escalonador da unidade central de processamento (CPU) percorre a fila, alocando a CPU para cada processo durante certo tempo e retira o primeiro processo da fila procedendo à execução. Se o processo não termina a execução no quantum determinado, o processo é inserido no final da fila, e se termina antes de um quantum, a CPU é liberada para a execução de novos processos [Cavalcante et al. 2010] [Revah and Kadosh 2016].Neste método, o balanceamento funciona de maneira cíclica; ou seja, o algoritmo utiliza a técnica de sempre direcionar as requisições através de uma lista circular e um apontador para o último host servidor selecionado, obedecendo à sequencialidade de cada item da lista; dessa forma quando um cliente requisita o serviço, o mais antigo na fila responde [Sabo 2006] [Revah and Kadosh 2016].

4.2.2. WRR - Weighted Round-Robin (Round-Robin com Peso)

O algoritmo Weighted Round-Robin é uma variação do algoritmo Round-Robin, contudo; as requisições respeitam a um peso atribuído para cada servidor. As requisições são distribuídas seqüencialmente entre os servidores, direcionando maior número de conexões para o servidor de maior peso. O Director irá distribuir as conexões com base nestes pesos. É a escolha indicada quando há diferença de capacidades entre os servidores reais. Ao atribuir um peso para um determinado servidor, este receberá um número maior de conexões que outro servidor [Cavalcante et al. 2010].

Como a sequência de escalonamento é gerada de acordo com os valores dos pesos dos servidores [Zhang and Zhang 2003], desta forma, assumindo-se que os servidores S1, S2 e S3, tem valores de pesos 3, 2 e 1 respectivamente. Então a sequência de escalonamento pode ser S1S1S2S1S2S3. O servidor real com peso 3 irá receber duas vezes mais conexões que um servidor real com peso 2, que irá receber conexões duas vezes mais que um servidor real com peso 1 [Garcia 2010] [Mesbahi and Rahmani 2016].

4.2.3. Dynamic Weighted Round-Robin (Round-Robin com Atribuição Dinâmica de Pesos)

O algoritmo Round-Robin com a atribuição dinâmica de pesos é uma variação da versão estática. Esse algoritmo associa para cada host servidor um peso dinamicamente estimado, proporcional ao estado de cada host. São políticas baseadas em informações do servidor. Periodicamente, o Director coleta informações dos servidores que são dinami-

camente incrementados para cada nova conexão atribuída. O algoritmo Round-Robin com atribuição de pesos é freqüentemente referenciado na literatura sem ser especificado se é uma versão estática ou dinâmica [Cardellini et al. 1999] [Chen et al. 2017] [Sroya and Singh 2017].

4.2.4. LC - Least Connections (Menos Conexões)

É um Algoritmo que utiliza conexões mínimas. As conexões são redirecionadas para o servidor real com menos conexões ativas, no momento da escolha. São políticas baseadas em informações do servidor. Utilizam informações de índices de estados dos hosts servidores para tomada de decisão. Esse é um algoritmo de escalonamento dinâmico, pois precisa checar quantas conexões estão ativas em determinado instante. É um bom algoritmo para quando a quantidade de conexões recebidas varia bastante [Garcia 2010].

Com este método, quando uma nova conexão chega, o Director verifica o número de conexões ativas e inativas para determinar para qual nó irá enviar a requisição. Para realizar esta escolha, o Director multiplica o número de conexões ativas do nó por 256 (atribuição interna do algoritmo em sua implementação) e adiciona ao resultado o número de conexões inativas resultando num overhead para cada nó. O nó com menor overhead receberá a conexão. Caso haja nós com mesmo overhead, o primeiro nó encontrado na tabela do código IPVS5 será selecionado [Meffe et al. 2006].

4.2.5. WLC - Weighted Least Connection (Menos Conexões com Peso)

Similar ao Least Connection, este método combina o Least Connection com um peso para cada nó. Distribui requisições ao servidor com menor quantidade de conexões no momento da escolha, levando em consideração também o peso atribuído a cada servidor real. O Director determina para qual nó uma requisição será enviada, calculando o overhead, como no método anterior, dividindo este número pelo peso associado ao nó. O nó com menor valor associado após esta operação receberá a conexão. Caso haja o mesmo valor associado, o primeiro da tabela do IPVS será selecionado [Meffe et al. 2006] [Zhao et al. 2016b].

4.2.6. LBLC - Locality Based Least Connection (Menos Conexões Baseada na Localidade)

O algoritmo Locality Based Least Connection define o servidor real com base no endereço destino da requisição, desta forma o algoritmo distribui mais solicitações aos servidores com poucas conexões ativas referentes a seus IPs de destino. Este algoritmo é projetado para uso em um cluster de servidores proxy cache. Exemplo para este cenário: o servidor real (proxy) X possui 100 (cem) conexões ativas com o servidor web www.exemplo.com.br. Novas requisições de clientes com destino ao servidor www.exemplo.com.br continuarão a ser encaminhadas para o servidor real X, até que esse esgote seus recursos. A partir daí o servidor real (proxy) Y será eleito como o novo servidor preferencial [Meffe et al. 2006] [Zhao et al. 2016a].

4.2.7. LBLCR - Locality Based Least Connection With Replication (Menos Conexões Baseada na Localidade com Replicação)

Este algoritmo funciona de maneira similar ao Locality Based Least Connection, com uma funcionalidade suplementar. As diferenças incluem a eleição de uma lista de servidores que podem responder por solicitações a um determinado destino. O Director mantém um mapeamento de um cliente para um conjunto de servidores que podem atendê-lo. Se todos os nós do conjunto estiverem sobrecarregados, o Director apanha um servidor com menos conexões e o adiciona ao conjunto. Caso este conjunto não se modificar em um intervalo de tempo específico (seis minutos, intervalo padrão com definido no código IPVS), o servidor com maior carga será excluído [Oktavianus 2013].

4.2.8. DH - Destination Hashing - (Hashing Destino)

Este algoritmo utiliza o recurso hash, ou seja, o Director monta uma tabela hash e redireciona as conexões para os servidores reais baseado nesta tabela, com ou sem pesos. Desta forma as requisições são repassadas através de uma lista estática, tendo como índice o endereço de destino das requisições [Ricarte 2010]. O método é útil quando o servidor real é um servidor proxy.

A ferramenta LVS - Linux Virtual Server ainda oferece para balanceamento de carga os algoritmos: SH - Source Hashing (Hashing Origem) - que também distribui as conexões de rede para os servidores reais através de uma tabela hash, com ou sem pesos, estaticamente atribuído pelos respectivos endereços IP de origem; ou seja, repassa os pedidos para o conjunto de servidores reais, através de um lista estática, tendo como índice o endereço de origem das requisições [Xu and Yang 2017]. SED - Shortest Expected Delay (Menor Atraso Esperado) - cuja função é repassar requisições para o servidor real que, presumidamente, responderá mais rapidamente. [Documentation 2011] [Kumar and Kalra 2016]. NQ - Never Queue (Nunca Enfileirar) - que adota um modelo de duas velocidades. Quando há um servidor ocioso disponível, o trabalho será enviado para o servidor inativo, em vez de esperar um rápido. Quando não há nenhum servidor ocioso disponível, o trabalho será enviado para o servidor que minimizar o seu atraso esperado [Server 2012].

5. Conclusões

Balancear carga entre servidores é a maneira mais eficaz e inteligente para a disponibilidade de serviços computacionais. Nesse sentido, o objetivo deste estudo foi apresentar o Balanceamento de Carga de Rede, destacando suas principais características, funcionalidades, técnicas utilizadas, recursos, soluções e algoritmos de escalonamento. Entendeu-se que o balanceamento de carga quando utilizado oferece escalabilidade e um poder maior de processamento, buscando maximizar o desempenho, para evitar a sobrecarga de serviços computacionais.

Conforme evidenciado, este trabalho tratou-se de uma pesquisa bibliográfica, onde a partir de estudos realizados sobre o tema, apresentou-se os pontos considerados mais relevantes. Verificou-se que o Balanceamento de Carga de Rede pode ser adotada visando evitar falhas ou a interrupções de serviços computacionais, garantindo disponibili-

dade elevada em um ambiente corporativo. Conforme pôde-se observar, existem diversas estratégias para balanceamento de carga, dependendo do fornecedor do projeto, porém optou-se por aprofundar nos estudos dos principais algoritmos implementados na ferramenta LVS - Linux Virtual Server, por ser um projeto open-source.

Considerou-se esta pesquisa relevante para a literatura, visto que o Balanceamento de Carga de Rede necessita ser mais difundida, especialmente para as instituições que enfrentam dificuldades relacionadas a sobrecarga de seus serviços computacionais. Como proposta para trabalhos futuros sugere-se a realização de um estudo de caso utilizando-se software livre, com o objetivo de comprovar a viabilidade e eficácia na utilização do Balanceamento de Carga em uma arquitetura em cluster. Para tanto, pretende-se realizar a implantação de um ambiente de rede virtualizado, onde seja possível testar alguns algoritmos de escalonamento utilizando a solução LVS - Linux Virtual Server (Servidor Virtual Linux), demonstrando quais os resultados encontrados a partir de um comparativo de desempenho e indicando aqueles que se mostrem mais eficientes.

Referências

Árabe, J. N. C. and Murta, C. D. (1996). Auto-balanceamento de carga em programas paralelos. In *Proceedings of the VIII Brazilian Symposium on Computer Architecture and High Performance Computing*, pages 161–171.

Batista 2007, A. C. (2007). Estudo teórico sobre cluster linux.

Cardellini, V., Colajanni, M., and Yu, P. S. (1999). Dynamic load balancing on web-server systems. *IEEE Internet computing*, 3(3):28–39.

Cavalcante, M. W., Alves, J. B. d. M., and Filho, P. J. F. (2010). Análise de desempenho de modelos para a distribuição de tráfego internet a um cluster de servidores web.

Chen, S.-L., Chen, Y.-Y., and Kuo, S.-H. (2017). Clb: A novel load balancing architecture and algorithm for cloud services. *Computers & Electrical Engineering*, 58:154–160.

Costa, H. L. A. et al. (2009). Alta disponibilidade e balanceamento de carga para a melhoria de sistemas computacionais críticos usando software livre: um estudo de caso.

da Silva Leandro, S. (2005). Balanceamento de carga em web services. *Universidade Federal de Santa Catarina*.

Documentation, R. H. (2011). Cluster suite overview. Disponível em: https://access.redhat.com/documentation/pt/. Acesso em: 06 mai. 2017.

Garcia, S. (2010). Cluster de balanceamento de carga em linux. Disponível em: https://sulamita.wordpress.com/2007/01/31/clusters-de-balanceamento-de-carga-em-linux/. Acesso em: 03 ago. 2017.

Kopparapu, C. (2002). *Load balancing servers, firewalls, and caches*. John Wiley & Sons.

Kumar, A. and Kalra, M. (2016). Load balancing in cloud data center using modified active monitoring load balancer. In *Advances in Computing, Communication, & Automation (ICACCA)(Spring), International Conference on*, pages 1–5. IEEE.

Meffe, C., Mussi, E. d. P., and Mello, L. d. (2006). Guia de estruturação e administração do ambiente de cluster e grid. *Guia Cluster*.

Mesbahi, M. and Rahmani, A. M. (2016). Load balancing in cloud computing: a state of the art survey. *International Journal of Modern Education and Computer Science*, 8(3):64.

Oktavianus, Y. L. (2013). Membangun sistem cloud computing dengan implementasi load balancing dan pengujian algoritma penjadwalan linux virtual server pada ftp server. *Jurnal Nasional Teknik Elektro*, 2(1).

Prodanov, C. C. and de Freitas, E. C. (2013). *Metodologia do Trabalho Científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico-2ª Edição*. Editora Feevale.

Revah, Y. and Kadosh, A. (2016). Dynamic load balancing switch architecture. US Patent 9,455,918.

Ricarte, I. L. M. (2010). Tabelas hash. Disponível em: http://www.dca.fee.unicamp.br/cursos/EA876/apostila/HTML/node26.html. Acesso em: 03 ago. 2017.

Sabo, C. P. (2006). *Avaliaçao de desempenho com algoritmos de escalonamento em clusters de servidores web*. PhD thesis, Universidade de São Paulo.

Santos, J., Lúcia, A., Marques, A., Augusto, M., Dórea, R., Carlos, R., and Rodrigo, É. (2010). Hs–clusters para balanceamento de carga. Disponível em: https://pt.scribd.com/document/97282862/HS-Clusters-Para-Balanceamento. Acesso em: 06 ago. 2017.

Sato, J. (2007). Balanceamento de requisições em cluster de servidores web: uma extensão para o mod-proxy-balancer do apache. Disponível em: http://www.din.uem.br/ mestrado/diss/2007/sato.pdf. Acesso em: 05 ago. 2017.

Server, L. V. (2012). Linux virtual server. Disponível em: http://www.linuxvirtualserver.org/. Acesso em: 17 mai. 2017.

Sroya, M. S. and Singh, V. (2017). Lddwrr: Least delay dynamic weighted round-robin load balancing in software defined networking. *International Journal*, 8(5).

Xu, J. and Yang, X. (2017). Research on key technologies of technological service and management based on cluster load balancing. *Cluster Computing*, pages 1–7.

Zhang, W. and Zhang, W. (2003). Linux virtual server clusters. *Linux Magazine*, 5(11).

Zhao, J., Liu, F., Tang, T., and Fu, J. (2016a). Method and device for accessing application server. US Patent 9,450,912.

Zhao, J., Yang, K., Wei, X., Ding, Y., Hu, L., and Xu, G. (2016b). A heuristic clustering-based task deployment approach for load balancing using bayes theorem in cloud environment. *IEEE Transactions on Parallel and Distributed Systems*, 27(2):305–316.

Extração De Edificações Em Imagens de Baixa Resolução Para Atualização De Mapa Urbano Básico

Paulo V. Santos[1], Dyego Almeida[1], Ronaldo Costa[1], Leandro Oliveira[1], Cristhiane Gonçalves[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

{paulovsantos,dyegoalmeida,leandroluis,ronaldocosta}@inf.ufg.br
cristhiane.goncalves@ufg.br

Abstract. *The high cost of geo-referenced technologies capable of keeping up to date the data of the geographic information system of a power distribution company, which has the basic urban map as the main data layer, may imply the inability to acquire this product, causing this information to be outdated or incomplete. This proposal to support the power distribution concessionaires will use computational vision (CV) methods and Digital Image Processing (DPI), enabling low-cost and efficient maintenance in the layers of buildings present in the basic urban maps From the distributors.*

Resumo. *O alto custo de tecnologias georreferenciadas capazes de manterem atualizados os dados do sistema de informação geográfico de uma concessionária de distribuição, que tenha o mapa urbano básico como a principal camada de dados, pode implicar na impossibilidade de aquisição deste produto, fazendo com que estas informações estejam desatualizadas ou incompletas. Esta proposta para apoio às concessionárias de distribuição de energia utilizará métodos de Visão Computacional (VC) e Processamento Digital de Imagens (PDI), possibilitando uma manutenção de baixo custo e eficiente nas camadas de edificações presentes nos mapas urbanos básicos das distribuidoras.*

1. Introdução

O Sistema de Informação Geográfica (SIG) é composto por um conjunto de conceitos, elementos e procedimentos que realizam tratamento computacional de dados geográficos através de aquisição, armazenamento, recuperação, transformação e geração de informações espaciais. Do acrônimo, *GIS - Geographic Information System*, os SIGs resultam da composição de três tecnologias: O sensoriamento remoto, GPS e geoprocessamento [Chang 2006].

As tecnologias envolvidas no Sensoriamento Remoto (RS) permitem a captação de informações e imagens da superfície terrestre através de ferramentas como satélites, radares e aerofotogrametrias. Juntamente com GPS e topografia, as informações obtidas poderão ser sintetizadas em mapas de uma determinada área de estudo, esta etapa consiste na fase de geoprocessamento.

Em [Câmara et al. 2001] , os autores afirmam que sistemas de informações geográficas são aplicados principalmente nos planejamentos e ordenamentos territoriais. Para que isto

seja possível, a geometria e os atributos dos dados de um SIG devem estar georreferenciados, isto é, localizados na superfície terrestre e representados numa projeção cartográfica computadorizada e armazenado juntamente com os dados informatizados.

Outra característica encontrada na maioria dos SIG's é a separação das informações em diferentes camadas agrupadas por assuntos similares, permitindo trabalhar em conjunto e relacionando as informações. Nota-se que as empresas privadas e órgãos públicos enfatizam o setor de planejamento urbano, havendo camadas separadas para cada item que compõe a área urbana, tais como ruas, quadras, edificações e outras conforme a área de interesse.

O trabalho de [de Aguiar and Medeiros] enfatiza a importância dos SIG's para o planejamento urbano citando várias vantagens, entre elas a facilidade da reprodução automática dos mapas temáticos a todos os usuários e obtenção das informações urbanas como: lotes, quadras, bairros e qualquer outra área de forma imediata e precisa, principalmente a redução de gastos na manutenção das informações vinculadas à representação física do espaço urbano. Neste contexto, o planejamento urbano representa uma das principais áreas de utilização dos SIG's e o alicerce da implantação do georreferenciamento neste planejamento é o mapeamento urbano básico (MUB).

O mapa urbano básico (MUB) constitui um importante conjunto de feições e informações gráficas representativas da organização territorial para uso nos SIG's, quando agrupadas e visualizadas simultaneamente formam uma base geográfica digital de grande escala e alta precisão, podendo apenas representar os níveis necessários a cada empresa conforme objetivos básicos. Esta base pode conter detalhes que são importantes para o negócio da empresa proporcionando uma visão única da cidade para os usuários como: lotes, logradouros, quadras, ruas e etc. [Fernandes et al. 2006].

Com base nos conhecimentos sobre mapas urbanos e sistemas de informações geográficas, [Singhal and Radhika 2014] mostraram que a detecção de edificações em imagens de alta resolução foi um problema por mais de duas décadas, e seu processo de detecção manual é muito complexo. Eles propõem um método não supervisionado no processo de detecção, mantendo assim os resultados satisfatórios.

[Singhal and Radhika 2014] demonstram ainda um método baseado em modelo que objetiva detecção de casas retilíneas, uma vez que as casas vistas em imagens aéreas são basicamente estruturas em 2D formadas por uma região contendo bordas em linhas retas. A dificuldade envolvida na pesquisa foi solucionada utilizando-se as propriedades de invariância das cores e o algoritmo de detecção de bordas de *Canny* [Canny 1986].

[Alex 2013] aborda que além das edificações as ruas também são importantes para o planejamento urbano e sua identificação poderá ser de grande interesse. Além de identificar edificações isoladas para atualização de dados cadastrais, demonstra também que há vários outros métodos de extração. O maior obstáculo da técnica é providenciar os melhores parâmetros para cada imagem em particular, porém a principal vantagem dessa abordagem é que a extração depende unicamente da coloração das ruas.

[Yuan and Cheriyadat 2014] abordam a importância em determinar a quantidade de edifícios para estimar o crescimento populacional, detecção de mudanças, monitoramento urbano e gestão de desastres. Esta metodologia enfatiza a utilização de imagens de

satélite por serem de baixo custo e efetiva em relação a outras tecnologias como *LIDAR* [Asner et al. 2007].

Com a aplicação de técnicas de processamento digital de imagens, visão computacional e aprendizado de máquina, é possível criar ferramentas que podem facilitar a identificação e extração de informações contidas nas imagens.

A segmentação é um processo de subdivisão de regiões baseada em cores ou conexões dos pixels, facilitando a localização de objetos e formas presentes nas imagens. Nesta abordagem segmentamos imagens de satélite de baixa resolução para atualização de mapa urbano básico (MUB), garantindo economia na realização do trabalho sem a necessidade de verificação in-loco, atividade com alto custo que demanda tempo dadas as proporções das áreas a serem verificadas em grandes centros urbanos.

2. Descrição do Problema

Há vários trabalhos publicados que utilizam imagens aéreas de alta resolução para identificação e reconhecimento de objetos, como edifícios, ruas, lotes e outros [Yeom and Kim 2015, Alshehhi and Marpu 2017, Xiao et al. 2017]. Essa função de reconhecimento é realizada naturalmente pelo ser humano através do processo cognitivo visual ou visão biológica. A visão computacional estuda como extrair informações das imagens através de sistemas artificiais baseados em modelos biológicos. O processo de identificação manual de objetos em imagens aéreas é dispendioso e demorado, podendo ser inviável. A partir de modelos naturais, geométricos e cores, esta abordagem utilizará segmentação em imagens de satélites, buscando formatos geométricos que se assemelham a edificações para possíveis identificações.

3. Método Proposto

O principal objetivo foi desenvolver uma metodologia capaz de criar mapas urbanos de edificações para serem utilizados pelas concessionárias de distribuição de energia, a partir do processamento de imagens georreferenciadas de baixa resolução. Os resultados obtidos pela extração da segmentação podem ser utilizadas em qualquer base de dados geoespacial, inclusive em softwares particulares ou livres independentemente de sua configuração. A partir das extrações georreferenciadas das edificações é possível realizar consultas, cruzamento de objetos espaciais e a visualização em qualquer escala desejada dentro dos sistemas de informação geográficas, nos quais receberão os dados atualizados após a aplicação desta metodologia.

4. Metodologia

Nesta seção será tomado como objeto de estudo imagens aéreas extraídas do *Google Earth* de dimensões variadas, desde que seja possível visualizar edificações, ruas e quadras.

A metodologia terá sua operação simplificada, obedecendo aos seguintes passos [Bhanu 1986, Sonka et al. 2014]: captura da imagem aérea, aplicação de filtros para remoção de ruídos, aplicação de filtros de baixo nível, preenchimento e remoção de falhas na imagem, binarização a partir um limiar ajustável, busca por componentes conectados, busca por padrões de objetos e geometrias.

Cada fase será detalhada, de modo que seja possível obter as edificações propostas nesta abordagem com a devida localização georreferenciada, para que tais dados possam ser utilizados futuramente em possíveis atualizações de qualquer camada espacial de softwares presentes em qualquer concessionária de distribuição. A abordagem proposta se assemelha ao trabalho de [Li et al. 2017], por utilizar as características clássicas da imagem e trabalhar com regiões conectadas para designar os possíveis candidatos, diferente de propostas como de [Chai 2017] que usam modelos de representação mais avançados do que imagens 2D, como representações 3D.

4.1. Captura de Imagens Aéreas Georreferenciadas

Primeiramente obtém-se uma imagem georreferenciada, que obedeça à precisão geográfica de suas coordenadas, facilitando pois assim a exata conversão de *pixels* para coordenadas de latitude e longitude. Após a obtenção da imagem, esta deverá conter indícios de construções: casas, ruas, quadras. Na imagem da Figura 1 é apresentado um loteamento que poderá ser analisado caso não exista no mapa urbano básico (MUB) da concessionária de distribuição será atualizado no SIG com as novas informações do resultado. Trata-se de uma imagem da região metropolitana de Goiânia-GO, município de Goianira, resolução de 1500 x 970 *pixels* com intensidade de 24 *bits* no formato do sistema de cores RGB.

(a) Imagem Extraída do *Google Earth*. (b) Imagem em níveis de cinza

Figura 1. Imagem extraída transformada em tons de cinza

4.2. Remoção de Ruídos

Esta etapa consiste em transformar a Figura 1(a) em tons de cinza e eliminar possíveis ruídos com ajuda de um filtro mediano [Van Kempen and Van Vliet 2017], que suaviza a imagem transformando *pixels* fora do padrão de uma determinada vizinhança em valores em uma escala aceitável aos seus vizinhos. O resultado será uma imagem com detalhes irrelevantes reduzidos conforme a Figura 1(b) e visualmente melhor.

4.3. Detector de Bordas e Binarização

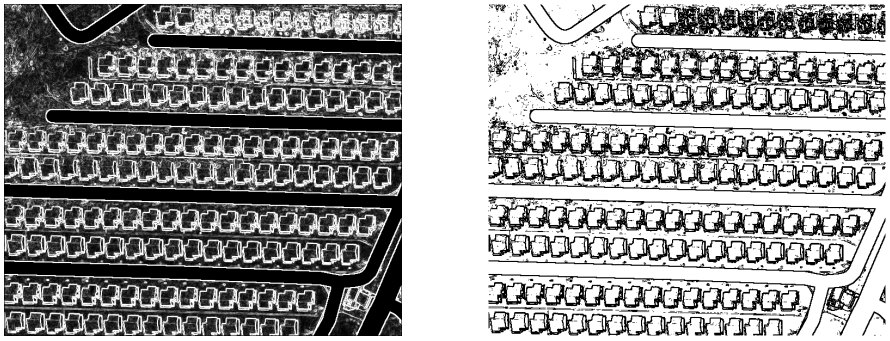
Utilizando-se do filtro de características de baixo nível, detector de bordas de *Sobel* [Gao et al. 2010], será possível realçar as linhas através de um operador que calcula as variações no sentido horizontal e vertical, importante por demarcar o limite entre diferentes regiões da imagem.

Também será aplicado um filtro de binarização [Otsu 1979], com limiar constante no valor de 120, tornando qualquer *pixel* abaixo desse valor como zero (0) e acima como um (1).

Este limiar constante foi definido após analisar o histograma gerado pela imagem resultante da aplicação do filtro de *Sobel*, é possível observar uma distribuição heterogênea de tons de cinza. O filtro gera *pixels* de diferentes intensidades luminosas na faixa de 0 a 255, o limiar no valor selecionado elimina os níveis de cinza e branco superiores, destacando o fundo e as bordas geradas pelas fortes mudanças de cores das imagens e tornando a distribuição mais homogênea.

O resultado da aplicação poderá ser observado na Figura 2 , onde as bordas são evidenciadas separando possíveis objetos do fundo da imagem. Há uma falha que deverá ser tratada posteriormente, que é a relação entre construções e suas sombras conforme evidenciado no trabalho de [Irvin and McKeown 1989], as sombras podem obstruir partes de outras áreas de interesse.

É utilizada a implementação baseada na transformada de *Hough* para localizar o segmento de linha que melhor esteja alinhado com a extremidade da região, conforme demonstra a proposta de [Tsenga and Hungb 2016]. O mesmo comparou diversas imagens, e concluiu que há uma relação linear entre a quantidade de linhas com a quantidade de edificações (casas), por isso a semelhança das estruturas das edificações que tendem a exibir números iguais de linhas em imagens aéreas. Esta forte relação evidenciada entre bordas e extremidades das edificações enfatizou a escolha desta técnica para aplicação nesta metodologia.



(a) Aplicação Filtro *Sobel*

(b) Aplicação Binarização com limiar de 120

Figura 2. Resultado da aplicação do operador de *Sobel* e Binarização através de um limiar

Na Figura 2 a imagem binária é composta por *pixels* desconexos, não se tratando de ruídos, mas pontos e pequenos blocos que se desconectaram de objetos maiores por meio da aplicação do limiar. Será realizado novamente o processo de suavização da imagem através de filtro da mediana, para que ruídos sejam eliminados.

4.4. Algoritmo de Componentes Conectados

Na maioria das construções estabelece um padrão no formato de polígono equiangular [Cui et al. 2008], ou seja, todos seus ângulos são congruentes [Benarchid et al. 2013]. o algoritmo de componentes conectados é aplicado após a binarização para executar a extração de regiões rotuladas segmentadas baseadas em heurísticas. Dado um *pixel*(x,y)

todos os *pixels* que tocam horizontalmente, verticalmente e diagonalmente fazendo parte da mesma região. Este tipo de conectividade é conhecido como 8-conectados.

Esta técnica é bastante utilizada em reconhecimento de edifícios, rodovias e outros objetos urbanos [Jin and Davis 2005, Cleveland et al. 2017]. Conforme a Figura 3, nota-se que é possível visualizar as regiões conectadas em diferentes classificação de cores.

Figura 3. imagem após aplicação do algoritmo de rótulo de regiões conectadas.

4.5. Extração de Componentes Conectados

O método utiliza uma abordagem para extrair os componentes conectados, transformando-os em geometrias semelhantes aos contornos das edificações. Conforme apresentado na Figura 4, evidencia-se que esta abordagem poderá extrair possíveis características. A proposta poderá ser não supervisionada para extração em massa e os resultados poderão ser utilizados, por exemplo, em redes neurais para localização de outros imóveis similares. Na literatura são descritos vários trabalhos em extração, reconhecimento e identificação de objetos em imagens aéreas de satélite já utiliza redes neurais ou aprendizado de máquina para tais propósitos [Huang et al. 2017].

A proposta é eficiente computacionalmente, por processar principalmente características de baixo nível, mais performáticos e bastante utilizado na extração de objetos em imagens e vídeos [Xian et al. 2017, Konstantinidis et al. 2017]. No entanto, constata-se que há falhas nesta metodologia, assim mesmo esta poderá ser usada como geradora de candidatos, a partir da extração de construções e outros objetos, e em trabalhos futuros poderá acontecer o amadurecimento da metodologia, partindo para métodos probabilísticos [Sirmacek and Unsalan 2011] ou aprendizado de máquina [Bellman and Shortis 2002, Maloof et al. 2003, Dornaika et al. 2016].

5. Avaliação

O método de avaliação utilizado nesta proposta é o índice de *Rand* [Rand 971], pois a abordagem emprega segmentação como sendo o principal meio de reconhecimento das formas dos objetos nas imagens. A avaliação de algoritmos de segmentação é subjetiva, pois trata-se de algoritmos não supervisionados. O índice de *Rand* representa a medida da similaridade entre dois grupo e está relacionado à precisão matemática (acurácia). Em propostas que utilizam classificadores é comum métricas como taxas de precisão, exatidão, sensibilidade e especificidade, pois o algoritmo é treinado e depois verificado em uma base de teste, ou seja, supervisionado.

(a) Edificações Similares (b) Edificações Distintas

Figura 4. Imagem com a extração de possíveis hipóteses de construção em uma imagem de baixa qualidade

Sendo utilizado como método de avaliação em outras propostas como [Cmu-ri tr 2005], o ajuste do índice de *Rand* mede a qualidade global da solução e está relacionado a precisão, podendo variar ente 0 e 1 de acordo com a expectativa.

A partir dos dados de uma matriz de confusão, como por exemplo a Tabela 1 tem-se a Equação 1.

$$Rand = (a + d)/(a + b + c + d) \tag{1}$$

Na prática, o ajuste do índice de *Rand* é descrito como a equação 2.

$$IndexAdj = \frac{index - ExpectedIndex}{MaxIndex - ExpectedIndex} \tag{2}$$

Tabela 1. Rand e Matriz de confusão dos resultados da Figura 4(a) e Figura 4(b)

Rand

	Sim	Não
Sim	a	b
Não	c	d

(a)

	Sim	Não
Sim	118	7
Não	23	0

(b)

	Sim	Não
Sim	9	1
Não	13	0

Para o caso da Figura 4(a) a metodologia para extração de edificações obteve os seguintes valores conforme a Tabela 1(a), tendo o índice de *Rand* no valor de 0,8 e para a Figura 4(b) conforme a Tabela 1(b) o valor de 0,4.

6. Conclusão e Trabalhos Futuros

O objetivo do trabalho foi alcançado, uma vez que é possível uma abordagem para reconhecimento, extração e identificação de edifícios, ruas e objetos urbanos a partir de imagens aéreas de baixa resolução, conforme evidenciados nos resultados da proposta.

O processo de extração desses objetos para uma estrutura de dados no qual possam ser utilizados em outras aplicações é completamente viável, reduzindo a aferição in-loco, diminuindo o tempo e os custos envolvidos no processo atualização do MUB por empresas especialistas.

O grande diferencial desta abordagem é utilizar técnicas modernas avaliadas em propostas de imagens de alta resolução em imagens de baixa qualidade. A partir da extração de possíveis candidatos evidenciados nesta metodologia, estes poderão ser utilizados em abordagens de aprendizado de máquinas em trabalhos futuros, bem como a evolução da própria abordagem implementando algoritmos que possam melhorar as etapas propostas neste trabalho. Esta metodologia também destaca que a extração de componentes conectados poderá ser evoluído a partir de algoritmos que possam fazer a união de bordas [Yeom and Kim 2015] e linhas que foram omitidas nos passos anteriores, permitindo assim a real exibição de estruturas geométricas nas imagens aéreas analisadas.

Referências

Alex, D. M. (2013). Robust and efficient method to extract roads from satellite images. *Int. J. Latest Trends Eng. Technol*, 2:26–29.

Alshehhi, R. and Marpu, P. R. (2017). Hierarchical graph-based segmentation for extracting road networks from high-resolution satellite images. *ISPRS Journal of Photogrammetry and Remote Sensing*, 126:245–260.

Asner, G. P., Knapp, D. E., Kennedy-Bowdoin, T., Jones, M. O., Martin, R. E., Boardman, J., and Field, C. B. (2007). Carnegie airborne observatory: in-flight fusion of hyperspectral imaging and waveform light detection and ranging for three-dimensional studies of ecosystems. *Journal of Applied Remote Sensing*, 1(1):013536–013536.

Bellman, C. and Shortis, M. (2002). A machine learning approach to building recognition in aerial photographs. *INTERNATIONAL ARCHIVES OF PHOTOGRAMMETRY REMOTE SENSING AND SPATIAL INFORMATION SCIENCES*, 34(3/A):50–54.

Benarchid, O., Raissouni, N., El Adib, S., Abbous, A., Azyat, A., Achhab, N. B., Lahraoua, M., and Chahboun, A. (2013). Building extraction using object-based classification and shadow information in very high resolution multispectral images, a case study: Tetuan, morocco. *Canadian Journal on Image Processing and Computer Vision*, 4(1):1–8.

Bhanu, B. (1986). Automatic target recognition: State of the art survey. *IEEE transactions on aerospace and electronic systems*, (4):364–379.

Câmara, G., Davis, C., and Monteiro, A. M. V. (2001). Introdução à ciência da geoinformação.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 8(6):679–698.

Chai, D. (2017). A probabilistic framework for building extraction from airborne color image and dsm. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(3):948–959.

Chang, K.-T. (2006). *Geographic information system*. Wiley Online Library.

Cleveland, J., Thakur, D., Dames, P., Phillips, C., Kientz, T., Daniilidis, K., Bergstrom, J., and Kumar, V. (2017). Automated system for semantic object labeling with soft-object recognition and dynamic programming segmentation. *IEEE Transactions on Automation Science and Engineering*, 14(2):820–833.

Cmu-ri tr, C. P. M. H. (2005). A Comparison of Image Segmentation Algorithms. *Robotics*, page 336.

Cui, S., Yam, Q., Liu, Z., and Li, M. (2008). Building detection and recognition from high resolution remotely sensed imagery. In *Proceedings of the XXIst ISPRS Congress*, pages 411–416.

de Aguiar, C. D. and Medeiros, C. B. Construção de um modelo básico unificado a partir de sistemas stand-alone.

Dornaika, F., Moujahid, A., El Merabet, Y., and Ruichek, Y. (2016). Building detection from orthophotos using a machine learning approach: An empirical study on image segmentation and descriptors. *Expert Systems with Applications*, 58:130–142.

Fernandes, V. d. O. et al. (2006). Análise das cartas do mapeamento cadastral urbano no brasil: proposta para normatização da simbologia.

Gao, W., Zhang, X., Yang, L., and Liu, H. (2010). An improved sobel edge detection. In *Computer Science and Information Technology (ICCSIT), 2010 3rd IEEE International Conference on*, volume 5, pages 67–71. IEEE.

Huang, X., Yuan, W., Li, J., and Zhang, L. (2017). A new building extraction postprocessing framework for high-spatial-resolution remote-sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(2):654–668.

Irvin, R. B. and McKeown, D. M. (1989). Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1564–1575.

Jin, X. and Davis, C. H. (2005). Automated building extraction from high-resolution satellite imagery in urban areas using structural, contextual, and spectral information. *EURASIP Journal on Advances in Signal Processing*, 2005(14):745309.

Konstantinidis, D., Stathaki, T., Argyriou, V., and Grammalidis, N. (2017). Building detection using enhanced hog–lbp features and region refinement processes. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(3):888–905.

Li, E., Xu, S., Meng, W., and Zhang, X. (2017). Building extraction from remotely sensed images by integrating saliency cue. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 10(3):906–919.

Maloof, M. A., Langley, P., Binford, T. O., Nevatia, R., and Sage, S. (2003). Improved rooftop detection in aerial images with machine learning. *Machine Learning*, 53(1):157–191.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Transactions on systems, man, and cybernetics*, 9(1):62–66.

Rand, W. M. (971). Objective Criteria for the Evaluation of Clustering Methods Author : William M . Rand Source : Journal of the American Statistical Association , Vol . 66 , No . 336 (Dec ., 1971), pp . 846- Published by : American Statistical Association Stable URL. *Journal of American Statistical Association1*, 66(336):846–850.

Singhal, S. and Radhika, S. (2014). Automatic detection of buildings from aerial images using color invariant features and canny edge detection. *International Journal of Engineering Trends and Technology (IJETT)*, 11(8).

Sirmacek, B. and Unsalan, C. (2011). A Probabilistic Framework to Detect Buildings in Aerial and Satellite Images. *IEEE Transactions on Geoscience and Remote Sensing*, 49:211–221.

Sonka, M., Hlavac, V., and Boyle, R. (2014). *Image processing, analysis, and machine vision*. Cengage Learning.

Tsenga, Y.-H. and Hungb, H.-C. (2016). Extraction of building boundary lines from airborne lidar point clouds. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, 40.

Van Kempen, G. and Van Vliet, L. (2017). Prefiltering: Reducing the noise sensitivity of non-linear image restoration algorithms. *Fourth Quinquennial Review 1996-2001 Dutch Society for Pattern Recognition and Image Processing*.

Xian, Y., Rong, X., Yang, X., and Tian, Y. (2017). Evaluation of low-level features for real-world surveillance event detection. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(3):624–634.

Xiao, P., Yuan, M., Zhang, X., Feng, X., and Guo, Y. (2017). Cosegmentation for object-based building change detection from high-resolution remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing*.

Yeom, J. and Kim, Y. (2015). A regular grid-based hough transform for the extraction of urban features using high-resolution satellite images. *Remote Sensing Letters*, 6(5):409–417.

Yuan, J. and Cheriyadat, A. M. (2014). Learning to count buildings in diverse aerial scenes. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 271–280. ACM.

Aprimorando a Plataforma UbiCare por meio da Integração do Plano de Cuidados Ubíquo e do Coletor de dados Vitais via Serviços Web

Natália R. Junqueira, Sergio T. de Carvalho

[1] Instituto de Informática (INF) – Universidade Federal de Goiás (UFG)
Câmpus Samambaia – Goiânia – GO – Brasil

nattaliarj2016@gmail.com, sergio@inf.ufg.br

Abstract. *The project "Application of Ubiquitous Computing Techniques, Software Architecture and Computational Intelligence in the Context of Home Patient Monitoring"of the Institute of Informatics - UFG - developed applications and techniques aimed at designing a Care Plan in the context of remote monitoring of elderly patients with chronic diseases. Aiming at this goal and developing modules such as the Ubiquitous Care Plan and the Vital Data Collector. Also under development is a platform called UbiCare, an architectural solution aimed at operations without context of home health care. The objective of this work is to integrate the modules of the Plan of Care Located and to make Collector to the platform UbiCare, unifying like databases of the modules and platform, and creating an access control layer. To enable an integration with a code reference for a selection and validity of the functionalities of each of them. This work has contributed to the evolution of the UbiCare platform, as well as creating a new access control layer, allowing each module not to worry about its own access control and business rules.*

Resumo. *O projeto "Aplicação de Técnicas de Computação Ubíqua, Arquitetura de Software e Inteligência Computacional no Contexto do Monitoramento de Pacientes Domiciliares" do Instituto de Informática - UFG - tem desenvolvido aplicações e técnicas voltadas para a concepção de um Plano de Cuidados no contexto de monitoramento remoto de pacientes idosos com doenças crônicas. Visando atingir esse objetivo foram desenvolvidos módulos como o Plano de Cuidados Ubíquo e o Coletor de Dados Vitais. Também encontra-se em desenvolvimento uma plataforma denominada UbiCare, uma solução arquitetural voltada para as operações no contexto da assistência domiciliar à saúde. O objetivo deste trabalho é integrar os módulos do Plano de Cuidados Ubíquo e do Coletor à plataforma UbiCare, unificando as bases de dados dos módulos e da plataforma, e criando uma camada de controle de acesso. Para possibilitar a integração foi feita a refatoração dos códigos fazendo a verificação e a validação das funcionalidades de cada um deles. Este trabalho contribuiu para evoluir a plataforma UbiCare, além de criar uma nova camada voltada ao controle de acesso, permitindo que cada módulo não precise se preocupar em desenvolver seu próprio controle de acesso e regras de negócio.*

1. Introdução

A assistência domiciliar à saúde tem se tornado cada vez mais importante na redução do número de internações hospitalares. Além disso, pacientes idosos ou com doenças crônicas necessitam de acompanhamento contínuo, o que nem sempre é viável. Soluções de monitoramento remoto de pacientes viabilizam o acompanhamento contínuo, aprimorando a qualidade de vida dos mesmos [Germano et al. 2016, Ribeiro et al. 2016, Battisti and Carvalho 2016b, Germano et al. 2015, de Carvalho et al. 2011].

No Instituto de Informática - UFG - há um Projeto de Pesquisa e Desenvolvimento intitulado "Aplicação de Técnicas de Computação Ubíqua, Arquitetura de Software e Inteligência Computacional no Contexto do Monitoramento de Pacientes Domiciliares". Este projeto tem como um de seus objetivos a proposição e a aplicação de técnicas para a concepção de um Plano de Cuidados no contexto de monitoramento remoto de pacientes idosos com doenças crônicas. Visando atingir esse objetivo foram desenvolvidos módulos como o Plano de Cuidados Ubíquo e o Coletor de Dados Vitais. Também encontra-se em desenvolvimento uma plataforma denominada UbiCare. O UbiCare foi idealizado como uma solução arquitetural que pretende realizar de forma transparente operações no contexto da assistência domiciliar à saúde. Além disso, o UbiCare tem o intuito de fornecer os serviços para os módulos como um Web Service que utiliza chamadas HTTP RESTful para realizar a comunicação por troca de mensagens.

Nesse contexto, o objetivo geral deste trabalho é dar continuidade ao desenvolvimento do UbiCare, seja em relação a novas funcionalidades, seja em direção ao aprimoramento do protótipo. Foram identificados durante os estudos dos modelos e da implementação do UbiCare, alguns pontos de melhoria: os módulos do Plano de Cuidados Ubíquo e do Coletor de Dados Vitais não se encontram integrados à plataforma; devido aos módulos não se encontrarem integrados, ambos os módulos e a plataforma possuem bases de dados distintas; foi identificada a necessidade de se desenvolver serviços que impossibilitem o acesso indevido à base de dados da plataforma UbiCare, de forma que só usuários autorizados possam acessar os serviços.

O objetivo desse artigo é contribuir para o aperfeiçoamento da Plataforma UbiCare de modo que sejam realizadas: a integração dos módulos do Plano de Cuidados Ubíquo e do Coletor de Dados Vitais à plataforma UbiCare, a unificação das bases de dados dos módulos e da plataforma, e, por fim a criação de uma camada de controle de acesso. Para possibilitar a devida integração dos módulos à plataforma, foi feita a refatoração dos códigos fazendo a verificação e validação das funcionalidades de cada um deles.

O artigo está organizado em mais 4 seções. A Seção 2 traz detalhes sobre a plataforma UbiCare. A Seção 3 apresenta as tecnologias e os fundamentos necessários para o desenvolvimento do trabalho. A Seção 4 apresenta os resultados, e, por fim, a Seção 5 aborda as conclusões e as possibilidades de trabalhos futuros.

2. Plataforma UbiCare

A plataforma UbiCare teve seu desenvolvimento iniciado em 2015 com o objetivo de investigar e propor a aplicação de técnicas de computação ubíqua e pervasiva, de algoritmos inteligentes, de padrões de Informática em Saúde, e de arquitetura de software, no contexto de uma solução computacional para o monitoramento remoto de pacientes domiciliares.

Visando atingir esses objetivos, a plataforma UbiCare foi idealizada como uma solução arquitetural com o propósito de realizar de forma transparente operações no contexto da assistência domiciliar à saúde e monitoramento remoto de pacientes, tanto para os pacientes e os cuidadores, quanto para os profissionais de saúde. Ela tem como base a arquitetura orientada a serviços, de modo a coordenar as operações das aplicações, garantir a interoperabilidade, manter um baixo acoplamento e permitir o reuso de informações comuns entre elas.[Germano et al. 2016, Germano et al. 2015]

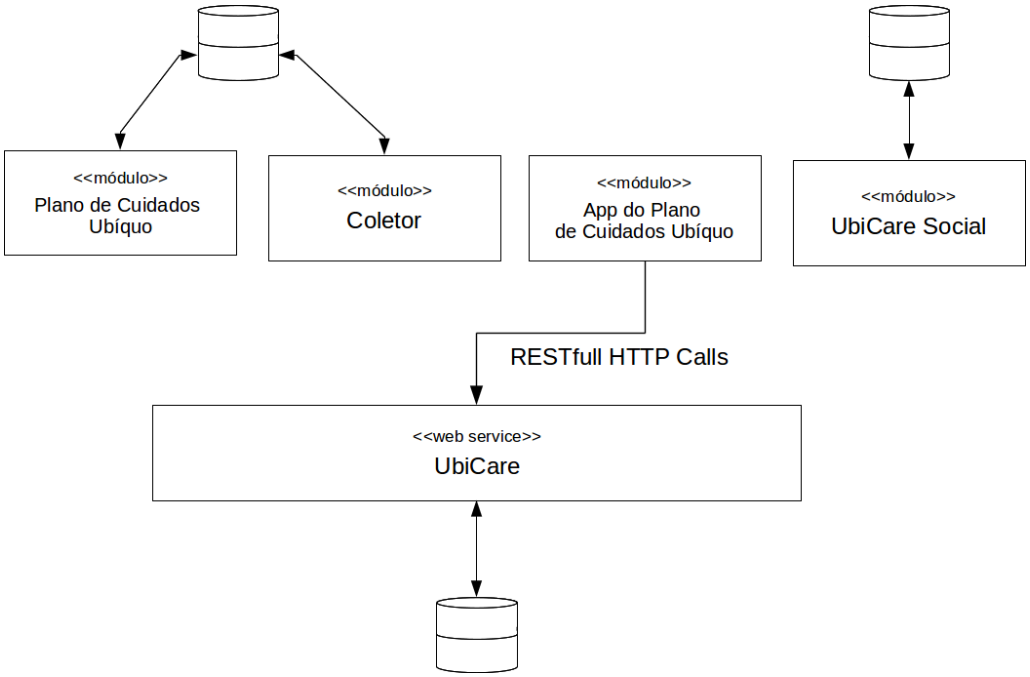


Figura 1. Arquitetura da Plataforma UbiCare

A Figura 1 apresenta a arquitetura da plataforma UbiCare e os módulos até então desenvolvidos que a compõem.

O módulo App do Plano de Cuidados Ubíquo foi construído para dispositivos móveis com sistema operacional Android, em linguagem de programação Java. Esse módulo se integra ao UbiCare via Internet por meio de Web Services, via HTTP, trazendo seus dados e suas funcionalidades às mãos do usuário final, em seu dispositivo celular ou tablet, atribuindo mais mobilidade ao sistema como um todo, permitindo acesso remoto ao usuário, para que consulte seus dados ou receba notificações sobre seu tratamento. Mesmo sem conexão à Internet essas funcionalidades continuam ativas, graças ao banco de dados local da aplicação que se mantém em sincronia com os servidores [Batalha de Carvalho and Carvalho 2016].

O módulo do UbiCare Social foi desenvolvido para permitir o encaminhamento de notificações, relacionadas ao monitoramento contínuo de sinais vitais de pacientes, para profissionais de saúde e familiares utilizando redes sociais online. O UbiCare Social realiza a integração destas redes sociais a um sistema de monitoramento remoto de pacientes permitindo que os usuários demonstrem interesse em receber notificações por meio da manutenção de relacionamentos, permitindo assim que complicações e anomalias sejam identificadas e comunicadas com antecipação, ocasionando redução na taxa de

readmissão hospitalar. [Ribeiro et al. 2016].

As próximas subseções apresentam os módulos Plano de Cuidados Ubíquo e Coletor de Dados Vitais, uma vez que foram o alvo da integração proposta e implementada neste trabalho.

2.1. Plano de Cuidados Ubíquo

Um plano de cuidados é um conjunto de prescrições voltadas ao paciente. Elas podem ser desde quais medições o paciente deve realizar e com qual frequência, quais remédios deve tomar e quando devem ser ingeridos, até recomendações sobre a dieta e exercícios físicos, entre outros.

O Plano de Cuidados Ubíquo é uma solução realizada com o objetivo de substituir a forma tradicional de elaboração de um plano de cuidados por uma implementação de *software*, visando o aumento da adesão dos pacientes aos tratamentos domiciliares. O paciente pode ser notificado de suas prescrições através de alguma tecnologia (como *smartphones,tablets*, TVs, entre outros). Além disso, o próprio cuidador pode utilizar o *software* para monitorar o cumprimento das prescrições [Germano et al. 2016, Battisti and Carvalho 2016b, Germano et al. 2015, Germano and Carvalho 2015].

2.2. Coletor de Dados Vitais

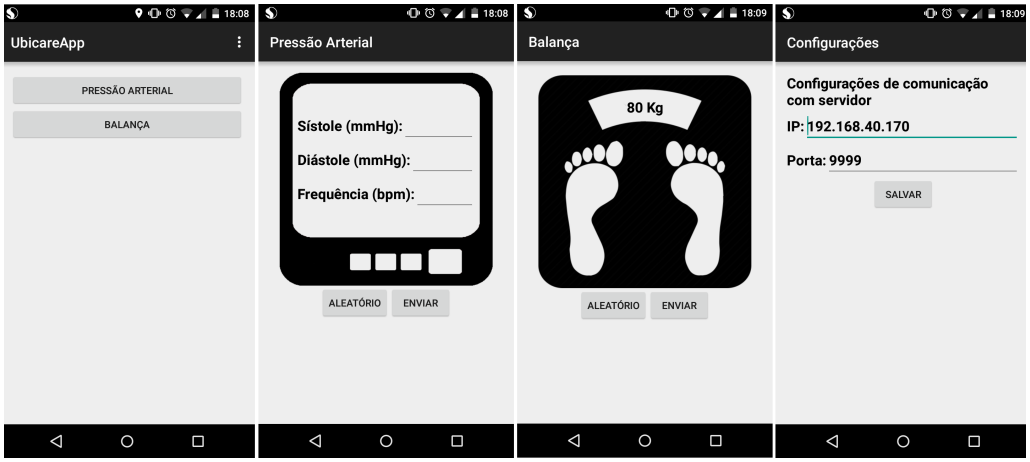


Figura 2. Interface do aplicativo de entrada de dados

O Coletor de Dados Vitais consiste em uma solução de computação ubíqua desenvolvida simulando sensores fisiológicos em um ambiente domiciliar com o objetivo de coletar os dados do paciente. Foi realizado por meio da investigação dos requisitos de interoperabilidade em sistemas de assistência domiciliar à saúde, estudo do padrão ISO/IEEE 11073 e implementação de um software que captura e manipula dois tipos sensores - pressão arterial e balança.

A validação do módulo do Coletor de Dados Vitais foi realizada por meio de um simulador de sensores. Esse simulador, um aplicativo para a plataforma Android, permite a geração de dados de sensores do tipo balança e medidor de pressão arterial. A Figura 2 mostra a interface do aplicativo que simula os sensores. O Coletor de Dados Vitais recebe os dados gerados pelo simulador, os encapsula em uma mensagem e os envia para o banco de dados [Battisti and Carvalho 2016b, Battisti and Carvalho 2016a].

3. Tecnologias e Fundamentos

Dada a heterogeneidade dos módulos e suas dependências, a realização deste trabalho se apoia em um conjunto de tecnologias e fundamentos tais como: refatoração de software, interoperabilidade entre dispositivos médicos pessoais e Web Services. As seções a seguir tratam desses métodos e materiais, explicitando o seu impacto na solução construída.

3.1. Refatoração

Segundo Martin Fowler (2004) a refatoração é uma alteração feita na estrutura interna do software para torná-lo mais fácil de ser entendido e menos custoso de ser modificado sem alterar seu comportamento observável [Fowler 2004].

Em outros termos, a refatoração tem como objetivo principal melhorar a estrutura interna do código, sem alterar a estrutura externa do mesmo, de modo que os usuários do sistema não percebam que qualquer alteração foi realizada. Para garantir que isso ocorra, toda modificação realizada precisa ser validada por meio da realização de testes para verificar se não surgiram novos bugs e que o comportamento continua o mesmo do esperado.

Neste projeto a refatoração foi realizada nos módulos do Plano de Cuidados Ubíquo e do Coletor de Dados Vitais com o intuito de possibilitar a integração deles ao barramento de forma que os mesmos consumissem os serviços do UbiCare.

3.2. Biblioteca Antidote

A interoperabilidade entre dispositivos médicos pessoais e os sistemas de saúde, é o que se refere o padrão ISO/IEEE 11073 (X73) [Signove 2012]. Este padrão define um protocolo de troca de dados eficiente, bem como os modelos de dados necessários para a comunicação entre dispositivos.

O Antidote é uma biblioteca open source do padrão X73, cujo principal objetivo é manipular a comunicação entre Agent e Manager. O Agent é o lado da comunicação que produz os dados, o qual poderia ser uma balança ou um medidor de pressão arterial, por exemplo. Nessa solução, porém, trata-se de uma aplicação desenvolvida em linguagem C que espera receber os dados do aplicativo Android para transmiti-los para o Manager. Por sua vez, o Manager é o lado da comunicação que recebe as mensagens, além de interpretá-las e disponibilizá-las ao sistema de assistência domiciliar à saúde.

Essa biblioteca foi utilizada como base para o desenvolvimento do módulo do Coletor de Dados Vitais, por isso foi tão importante a compreensão da mesma para realizar a integração deste módulo com o UbiCare.

3.3. Web Service

Web Service é um conjunto de serviços Web que visa facilitar o processamento distribuído em sistemas heterogêneos, apresentando funcionalidades para os usuários através de protocolos padrão. Pode-se dizer também que um Web Service é a maneira mais comum de se integrar diferentes sistemas.

O UbiCare é um Web Service que utiliza o protocolo de comunicação REST, do inglês, Representational State Transfer [FIELDING 2000]. O REST é baseado no protocolo de hipermídia HTTP e utiliza seus métodos para se comunicar, pois o HTTP

possui diversos mecanismos embutidos para representar recursos como cabeçalhos, tipos de conteúdo, código de status, entre outros. Uma deficiência do protocolo HTTP é que ele é stateless, ou seja, ele não é capaz por si só de reter informações entre requisições diferentes. Dessa forma para persistir informações é necessário utilizar sessões, cookies, campos de formulário ou variáveis na própria URL.

Os principais mecanismos nesta arquitetura são as URLs do sistema e os recursos. Os métodos mais utilizados são:

- GET: Solicita a representação de um determinado recurso. É definido como um método seguro e não deve ser usado para disparar uma ação (remover um usuário, por exemplo).
- POST: As informações enviadas no corpo (body) da requisição são utilizadas para criar um novo recurso. Também é responsável por fazer processamentos que não são diretamente relacionados a um recurso.
- DELETE: Remove um recurso. Deve retornar o status 204 caso não exista nenhum recurso para a URL especificada.
- PUT: Atualiza um recurso na URL especificada. Caso o recurso não exista, ele pode criar um. A principal diferença entre POST e PUT é que o primeiro pode lidar não somente com recursos, mas pode fazer processamento de informações, por exemplo.

4. Resultados

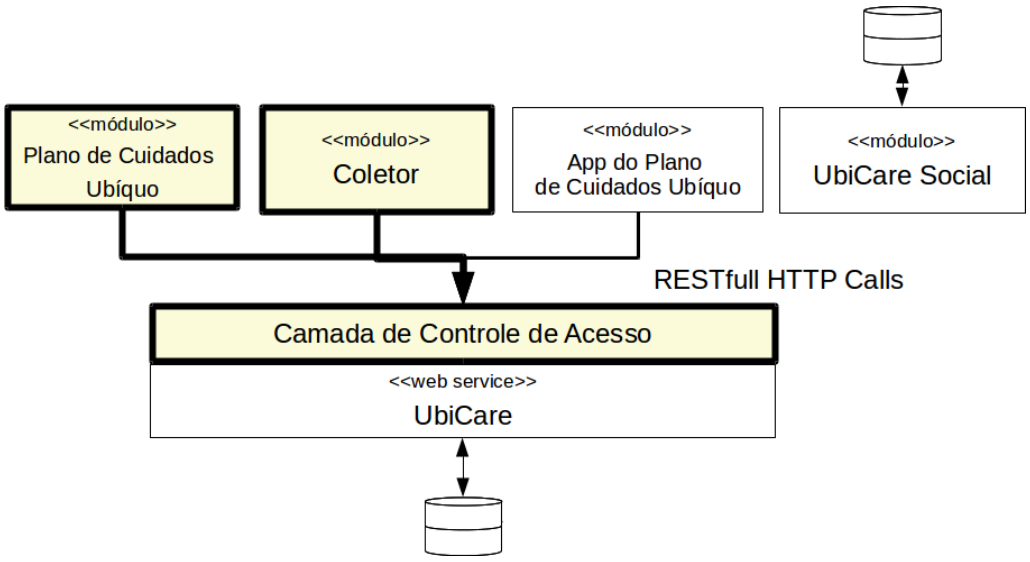


Figura 3. Arquitetura atual da PI UbiCare

A Figura 3 apresenta a arquitetura da Plataforma UbiCare após as ações de implementação realizadas neste trabalho. As ações destacadas em negrito são: A unificação da base de dados do Plano de Cuidados Ubíquo e do Coletor de Dados Vitais à do UbiCare, a integração com o barramento de serviços e o desenvolvimento de uma camada de controle de acesso. As seções seguintes trazem cada um desses resultados separadamente.

4.1. Unificação da Base de Dados

A unificação foi realizada com base na análise das duas bases de dados e na refatoração dos códigos do Plano de Cuidados Ubíquo. Foram criadas na base de dados do UbiCare todas as tabelas necessárias. Também foram criados novos serviços para realizar consultas, cadastros, alterações e exclusões de dados para cada uma das tabelas. Os novos serviços implementados são:

- login: serviço que realiza a autenticação do login.
- relogin: serviço para efetuar login automaticamente, depois te ter realizado o login pelo menos uma vez.
- logout: serviço para efetuar logout.
- atividade física: serviço de cadastro, consulta, alteração e exclusão das prescrições de atividade físicas.
- exames médicos: serviço de cadastro, consulta, alteração e exclusão das prescrições de exames médicos.
- medicamentos: serviço de cadastro, consulta, alteração e exclusão das prescrições de medicamentos.
- fatores de risco: serviço de cadastro, consulta, alteração e exclusão das prescrições de fatores de risco.
- dados fisiológicos: serviço de cadastro, consulta, alteração e exclusão dos dados fisiológicos do paciente.
- notificações da semana: serviço responsável por fornecer as notificações da semana.

4.2. Camada de Controle de Acesso

A camada de controle de acesso do UbiCare é composta por dois middlewares de autenticação e seis de controle de acesso. Os de autenticação são um para os módulos os quais possuem usuários que realizam login, e outro específico para o Coletor de Dados Vitais, pois ele não efetua login.

O primeiro utiliza um token identificador de sessão, o qual é concedido ao usuário quando o mesmo realiza login. Neste trabalho ele foi utilizado como uma alternativa para não ser necessário enviar em toda requisição o login e a senha do usuário. O token é representado por uma cadeia longa de caracteres, gerada aleatoriamente por uma função hash, e que geralmente possui um prazo de validade. Os tokens são bloqueados para só serem utilizados a partir do endereço IP em que foi realizado o login.

Quando é realizada uma requisição a um serviço que utiliza o middleware de token, este é verificado quanto à sua validade, observando três condições: se o token ainda não expirou, se o IP é o mesmo do esperado (o IP esperado é o de quem realizou login) e se o token existe na base de dados. Se válido, o serviço é chamado, se não, é retornada uma mensagem de token inválido.

O segundo middleware usa o conceito de chave assimétrica e funciona como um intermediador entre os serviços do UbiCare e o módulo do Coletor de Dados Vitais, sendo então atribuída ao Coletor de Dados Vitais uma chave pública a qual ele utiliza para criptografar os dados que o middleware descriptografa com a chave privada, e, na sequência, chama o serviço do barramento transmitindo os dados para serem processados e armazenados na base de dados.

Os middlewares de controle de acesso foram desenvolvidos para controlar quais tipos de usuários podem acessar cada serviço. Neste projeto existem três tipos de usuários: o paciente, o familiar e o profissional que acompanha o paciente. São três tipos de situações que precisam, portanto, serem tratadas no controle de acesso.

A primeira ocorre quando todos os usuários podem acessar um determinado serviço; neste caso o serviço não utiliza nenhum middleware de controle de acesso. Quando apenas um tipo de usuário pode acessar, utiliza-se o middleware especifico daquele perfil de usuário. Quando só um dos tipos de usuário não pode acessar utiliza-se o middleware que proíbe usuários daquele perfil de acessarem o serviço. O código mostrado na Figura 4 refere-se a um dos serviços que compõem o UbiCare, utilizado para listar todos os usuários. Este serviço utiliza o middleware proibe_familia, pois os usuários do perfil familiar não podem ter acesso este serviço.

```
Route::get('',['uses'=>'UsuarioController@allUsuarios',
               'middleware'=>'proibe_familia']);
```

Figura 4. Serviço do UbiCare utilizando um middleware da Camada de Controle de Acesso

4.3. Integração dos módulos ao UbiCare

Devido à unificação da base de dados e à integração dos módulos do Plano de Cuidados Ubíquo e do Coletor de Dados Vitais ao UbiCare, a comunicação direta entre os módulos e a base de dados não mais pode ser realizada. Para um módulo armazenar, buscar ou consultar algum dado, ele precisa chamar um serviço do barramento através de uma requisição HTTP.

Para realizar a integração do módulo do Plano de Cuidados Ubíquo, foi realizada a refatoração de seu código, de modo que, depois de integrado, o comportamento do Plano de Cuidados Ubíquo continuou inalterado do ponto de vista do usuário. No lugar das classes de comunicação com o banco de dados, existentes anteriormente no módulo, foram criadas duas classes, utilizando a linguagem PHP, para facilitar a comunicação com o UbiCare, a APIRequest responsável por encapsular os métodos HTTP para facilitar a chamada de uma requisição, e a UbicareAPI que é a classe onde são criadas as funções para as requisições de serviços que o plano de cuidados deseja acessar no UbiCare.

A Figura 5 mostra o código de uma função que faz uma requisição a um serviço do UbiCare. Uma requisição é composta por três partes principais, o método HTTP, os headers que são os cabeçalhos e o body que é o corpo da requisição. Neste caso a requisição foi a consulta de todos os usuários cadastrados, usando o método get, passando no cabeçalho o token.

```
protected function GetUsuarios() {
    return APIRequest::get( url: $this->domain.'/gerencia/usuario',
        ['token' => $this->token],[]);
}
```

Figura 5. Função que realiza uma requisição a um serviço do UbiCare

A integração do módulo do Coletor de Dados Vitais foi realizada de modo similar. Foi feita a refatoração do código do Coletor de Dados Vitais, onde, no lugar das classes de comunicação com o banco, foram criadas classes em Java para facilitar a comunicação com o UbiCare, seguindo o Padrão ISO/IEEE 11073. Na Figura 6 é apresentada uma função que faz uma requisição POST ao serviço de gerenciamento de sensores do UbiCare, com o intuito de gravar os dados da medição realizada pela balança na base de dados.

```
public String addMedicaoBalanca(Balanca med) throws Exception {
    final String rota_medicao_balanca = "balanca";

    return conn.enviar(rota_medicao_balanca,
            "{\"peso\":" + med.getPeso()
                    + ",\"data\":\"" + med.getData().toString()
                    + "\",\"hora\":\"" + med.getHora().toString() + "\"}");
}
```

Figura 6. Função que realiza uma requisição a um serviço do UbiCare

O esforço na integração do Coletor de Dados Vitais levou ao aprimoramento do UbiCare em termos de sua aderência ao Padrão ISO/IEEE 11073.

5. Conclusão

Este trabalho contribuiu para evoluir a plataforma UbiCare, possibilitando a integração entre ela, o Plano de Cuidados Ubíquo e o Coletor de Dados Vitais por meio da refatoração dos códigos. Foi realizada a verificação e validação da integração com cada um dos módulos por meio de testes que visam garantir que não foram gerados novos bugs ou mudanças no comportamento externo destes módulos.

Também foi realizada a unificação da base de dados que agora é acessada de forma direta somente pelo barramento de serviços. Os módulos quando desejam usar alguma funcionalidade que anteriormente era feita diretamente no banco de dados, agora, realizam uma requisição a um serviço do barramento que é responsável por realizar essa tarefa, seja ela cadastro, consulta, alteração ou exclusão.

Além disso foi criada uma camada de controle de acesso para aumentar a segurança do sistema, permitindo que cada módulo não precise se preocupar em desenvolver seu próprio controle de acesso e regras de negócio. Um trabalho futuro a ser desenvolvido é a integração do módulo do UbiCare Social, juntamente com a unificação das suas bases de dados com a do UbiCare.

Referências

Batalha de Carvalho, L. and Carvalho, S. (2016). Ubicareapp: Aplicativo para dispositivos móveis para um plano de cuidados ubíquo voltado ao monitoramento de pacientes. Trabalho de Conclusão de Curso (Graduação em Ciências da Computação) - Universidade Federal de Goiás.

Battisti, D. and Carvalho, S. (2016a). Aplicação do padrão ISO/IEEE 11073 no contexto da assistência domiciliar à saúde. In: Escola Regional de Informática (ERIGO), 2016, Goiânia. Anais da IV Escola Regional de Informática (ERIGO), 2016.

Battisti, D. and Carvalho, S. (2016b). Uso do padrão ISO/IEEE 11073 integrado a técnicas de computação ubíqua no contexto do monitoramento remoto de pacientes. Trabalho de Conclusão de Curso (Graduação em Ciências da Computação) - Universidade Federal de Goiás.

de Carvalho, S. T., Copetti, A., and Loques Filho, O. G. (2011). Sistema de computação ubíqua na assistência domiciliar à saúde. *Journal Of Health Informatics*, 3(2).

FIELDING, R. T. (2000). Architectural styles and the design of network-based software architectures. Tese (Doutorado em Informação e Ciência da Computação) – Universidade da Califórnia, Califórnia, 2000.

Fowler, M. (2004). *Refatoração: Aperfeiçoando o projeto de código existente.* Bookman Editora.

Germano, E., Battisti, D., Ribeiro, H., and Carvalho, S. (2016). Plano de cuidados ubíquo para acompanhamento domiciliar de pacientes. In: Congresso Brasileiro de Informática em Saúde (CBIS), 2016. Congresso Brasileiro de Informática em Saúde (CBIS), 2016. p. 849-858.

Germano, E. and Carvalho, S. (2015). Implementação de um plano de cuidados ubíquo para o monitoramento de pacientes domiciliares. Trabalho de Conclusão de Curso (Graduação em Ciências da Computação) - Universidade Federal de Goiás.

Germano, E., Carvalho, S., and De Souza-Zinader, J. (2015). Plano de cuidados ubíquo com sistema de notificações voltado a pacientes domiciliares. *III Escola Regional de Informática de Goiás.*, pages 33–44.

Ribeiro, H., Battisti, D., Germano, E., and Carvalho, S. (2016). Notificações de monitoramento remoto de pacientes usando redes sociais. Congresso Brasileiro de Informática em Saúde (CBIS), 2016, Goiânia. Anais do XV Congresso Brasileiro de Informática em Saúde (CBIS), 2016. p. 859-868.

Signove (2012). *Antidote: Program Guide. Documentation for developers of applications based on Antidote IEEE 11073 library*. Signove, 2.12 edition. http://oss.signove.com/index.php/File:AntidoteProgramGuide.pdf, acessado no dia 15/07/2017.

Uma Linguagem de Modelagem para o Desenvolvimento de Aplicações Ubíquas em um Contexto de Assistência Domiciliar à Saúde

Helberth Borelli, Sérgio Teixeira de Carvalho

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 - 74690-900 – Goiânia – GO – Brazil

hborello@gmail.com, sergio@inf.ufg.br

Resumo. *Este artigo apresenta o desenvolvimento de um Metamodelo baseado no paradigma de Linha de Produto de software (do inglês, Software Product Line - SPL), o qual tem como objetivo a modelagem de Aplicações Ubíquas para Assistência Domiciliar à Saúde. O metamodelo desenvolvido serve como base para a construção de uma Linguagem de Modelagem de Domínio Específico (do inglês, Domain Specific Modeling Language - DSML), permitindo assim a modelagem de sistemas baseados em SPL que produzam como resultado um esqueleto de códigos que buscam mitigar o desafio de desenvolver um sistema ubíquo. Ao final do artigo um cenário é apresentado com o objeto de validar a linguagem desenvolvida. Este cenário mostra um sistema ubíquo de notificações que visa assistir pessoas na adesão de tratamentos em domicílio.*

Abstract. *This paper presents the development of a Metamodel based on Software Product Line (SPL) paradigm, which aims to modeling Ubiquitous Applications to Home Care. The metamodel developed serves as the bases to build a Domain Specific Modeling Language (DSML), thus allowing the modeling of SPL-based systems that produce with result a skeleton of codes, with this, the study carried out seeks to mitigate the challenge of developing a ubiquitous system. At the end of the paper a scenario is presented with the purpose of validating the developed language. This scenario shows a ubiquitous system of notifications that aims to assist people in adhering to home-based treatments.*

1. Introdução

A cada dia tem-se realizado mais esforços em torno de trabalhos que promovam serviços de saúde, os quais garantam acesso e qualidade de vida às pessoas, especialmente em assuntos que dizem respeito a doenças crônicas. Deste modo, estudos buscam desenvolver estratégias para o cuidado de pessoas com doenças crônicas, tais estratégias envolvem um tratamento relacionado a mudanças de estilo de vida e um processo de cuidado contínuo. Essas estratégias visam, por exemplo, mudanças de hábitos alimentares, inclusão de atividades físicas e apoio a adesão de planos de cuidados [Ministério da Saúde do Brasil 2014]. A computação tem se mostrado bastante eficiente ao auxiliar e facilitar a rotina de tratamentos de saúde. Como exemplo temos sistmas computacionais que podem coletar e compartilhar informações de saúde referentes aos pacientes, oferecendo históricos relevantes aos profissionais de saúde.

A Computação Ubíqua consiste em uma abordagem que visa basicamente a inclusão da computação ao cotidiano do ser-humano de forma transparente ou invisível [Weiser 1999]. A intenção é que a computação potencialize o auxílio ao tratamento de saúde, uma vez que o uso de recursos computacionais possa coletar dados do paciente, bem como auxiliá-lo na adesão de planos de cuidados [Carvalho et al. 2011]. Por exemplo, com o uso da computação ubíqua é possível que sensores coletem informações como a presença do paciente por determinado tempo em um ambiente, níveis de batimento cardíaco, bem como, avisar o paciente de prescrições medicamentosas em seu tempo correto por meio de dispositivos, tais como: *smartphones*, *smartwatches*, *tablets*, TV etc.

Dada a complexidade da construção de sistemas que envolvem uma grande quantidade de recursos, técnicas de engenharia e arquitetura de software buscam mitigar algumas complexidades deste desenvolvimento. O conceito de SPL [Northrop 2008] tem, neste contexto, o objetivo de criar artifícios de gerenciamento dos recursos que poderão ser utilizados na montagem dos produtos. Neste estudo é utilizado o conceito de Metamoldelagem [OMG 2014a] para orientar a construção de Modelos de Característica [Kang et al. 1990] baseados no conceito de SPL. Este metamodelo é, então, transformado em uma DSML [Van Deursen et al. 2000] que visa a construção e derivação de modelos de aplicações ubíquas para um cenário de Assistência Domiciliar à Saúde.

Este artigo está organizado em seções, com a seguinte organização: a Seção 2 apresenta o Referencial Teórico que envolve todo o desenvolvimento do projeto; a Seção 3 apresenta cada processo para o desenvolvimento da linguagem de modelagem (DSML); a Seção 4 apresenta os resultados, ou a utilização da linguagem produzida em um teste de validação; e por fim, a Seção 5 apresenta as considerações finais do trabalho.

2. Fundamentação Teórica

Esta seção traz os principais conceitos utilizados no estudo que promoveu o desenvolvimento da DSML.

2.1. Linha de Produto de Software

Como definição, uma SPL tem como foco a reutilização planejada de códigos ou artefatos que visam a diminuir os custos de desenvolvimento, bem como promover a melhoria de qualidade dos produtos desenvolvidos. O conceito SPL vem sendo utilizado com sucesso em vários e diferentes domínios, segundo o Instituto de Engenharia de Software (SEI) [Northrop 2008].

Uma SPL pode ser definida como uma família de produtos de software, pois traz conceitos e características semelhantes entre várias aplicações [Carvalho 2013, Parnas 1976]. Dadas tais semelhanças, estas famílias de produtos podem compartilhar características que são reutilizadas a cada derivação ou criação de novos produtos. As características são usadas em SPL para especificar semelhanças e diferenças entre os produtos, assim como orientar a estrutura, a reutilização e as variações entre os produtos em seu ciclo de vida [Kang et al. 2002].

Pode-se definir, para melhor entendimento, que uma SPL é formada por conjuntos de características, divididas entre mandatórias e variáveis. As características mandatórias sempre estão presentes em todos os produtos derivados a partir da linha, ficando assim, a cargo das características variáveis a personalização entre os produtos.

2.2. Modelo de Características

Um modelo de características tem como objetivo representar, em alto nível de abstração, as partes significantes e seus relacionamentos dentro de um sistema [Laakko and Mäntylä 1993]. O modelo de caracteristicas amplamente utilizado para a representação de sistemas em SPLs, as características representam as propriedades de sistemas de um mesmo negócio que detém pontos em comum, bem como as suas variabilidades [Kang et al. 1990].

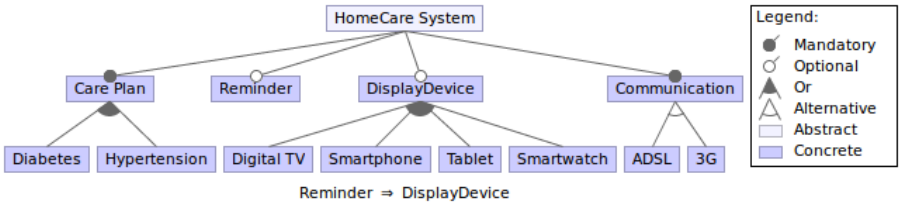


Figura 1. Modelo de Características da SPL.

O modelo de característica é usualmente representado por uma árvore de características. As características podem ser divididas em mandatórias, opcionais e alternativas. No caso das características mandatórias, toda derivação de produto proveniente da SPL deve contê-las, ou seja, devem fazer parte de qualquer derivação. As demais características determinam as variações de cada produto derivado [Carvalho 2013, Kang et al. 1990].

A Figura 1 traz a representação de um cenário de sistemas que compõe um Modelo de Características para uma SPL ou DSPL. Pode-se observar que apesar de o modelo de características destacar o tipo de cada característica, este não é determinante para definir se a linha de produto é ou não dinâmica. Deste modo, cabe ao trabalho de Metamodelagem e construção da DSML implementar mais restrições para a representação.

2.3. Metamodelos e Linguagens de Domínio Específico

Uma abordagem para lidar com a complexidade do desenvolvimento de uma SPL, bem como tratar a escalabilidade de uma família de produtos é a utilização de Metamodelos e Modelos de Características.

A utilização dos conceitos de construção de modelos pode propor a representação em alto nível de um sistema por meio do uso de metamodelos. A organização internacional OMG (*Object Management Group*), que tem como missão o desenvolvimento de padrões de tecnologia, apresenta como proposta de especificação para metamodelagem o MOF (*MetaObject Facility*) [OMG 2014a]. Deste modo, é possível definir o MOF, como uma proposta de linguagem para definição de metamodelos [Jouault and Bézivin 2006], e diante dessa especificação, uma solução de representação desta linguagem é a própria UML.

Os metamodelos são ainda considerados parte integrante do *software*, não se restringindo a apenas artefatos de documentação [Schmidt 2006]. A ideia central no desenvolvimento de metamodelos é a de estabelecer que o arquiteto mantenha seu esforço em modelos de alto nível, abstraindo-se das complexidades provenientes da implementação em plataformas distintas.

Uma DSL consiste em uma linguagem de programação usualmente declarativa, de elementos ou cláusulas de código reduzidos, justamente por determinar um domínio específico de problema [Van Deursen et al. 2000, Mernik et al. 2005].

Pode-se considerar como um ponto de partida para a construção de uma DSL o uso de um metamodelo. Neste sentido, pode-se definir que DSL trata-se de um conjunto coordenado de modelos [Jouault and Bézivin 2006]. Uma vez definido o domínio da aplicação por meio de um metamodelo, representando entidades e suas relações, obtém-se o primeiro passo para o desenvolvimento de uma DSL, ou melhor, uma DSML, visto que aqui o propósito é a modelagem. Mais especificamente, cada elemento (nós e arestas) do metamodelo produz a gramática central da DSML.

A construção de uma DSML, a partir de um metamodelo, não determina que a linguagem tenha apenas as restrições impostas pelo metamodelo. Dadas algumas limitações de representação de regras e restrições dentro de metamodelos representados por linguagens de modelagem (UML), cabe à linguagem, em determinado momento, especificar ou ampliar regras e restrições para os modelos por meio de uma Linguagem para Representar Restrições em Objetos (OCL do inglês *Object Constraint Language*) [OMG 2014b].

No contexto deste trabalho, pretende-se dar à DSML maior poder de restrição à modelagem, de modo que ela possibilite a descrição mais complexa de regras e restrições que descrevam os contratos dentro da DSPL.

2.4. Cenário

Este trabalho leva em consideração um cenário que tem como proposta o desenvolvimento de um sistema que apresente notificações com base no modelo de característica apresentado na Figura 2, situada na Seção 2.2. O sistema tem como objetivo notificar o paciente de suas tarefas diárias especificadas em seu plano de cuidados. O plano de cuidados contém especificações médicas que visam ao tratamento do paciente e o objetivo do sistema desenvolvido é auxiliar o paciente na adesão ao tratamento.

O objetivo é que os dispositivos que possuam *display* (smartTV, smartphone etc) possam notificar o paciente de tarefas prescritas em seu plano de cuidados. Outros dispositivos como, por exemplo, os sensores de presença, viabilizam tarefas que incluem a possibilidade de notificar o paciente baseado em sua localização dentro da casa.

3. Proposta

Esta seção apresenta o desenvolvimento da DSML, partindo da construção dos metamodelos baseados no conceito de SPL, os quais ,posteriormente, são transformados em uma Linguagem (DSML) que deverá proporcionar a modelagem de SPLs para aplicações ubíquas no contexto de Assistência Domiciliar à Saúde.

3.1. Desenvolvimento dos Metamodelos

A proposta é iniciada com a construção do Metamodelo de Características, o qual deve proporcionar que o modelo de negócio seja mapeado em forma de características, segundo a abordagem de Modelo de Características. Como resultado, a modelagem por intermédio da DSML deverá disponibilizar um esqueleto de códigos para montagem da aplicação derivada.

3.1.1. Metamodelo de Características

O metamodelo de características, primeiro passo para a modelagem da SPL, é representado pela Figura 2 que tem como raiz a metaclasse **ModelagemCaracteristicas**, que ,por sua vez, deve ser associada a ao menos uma característica modelada pela metaclasse **Características**.

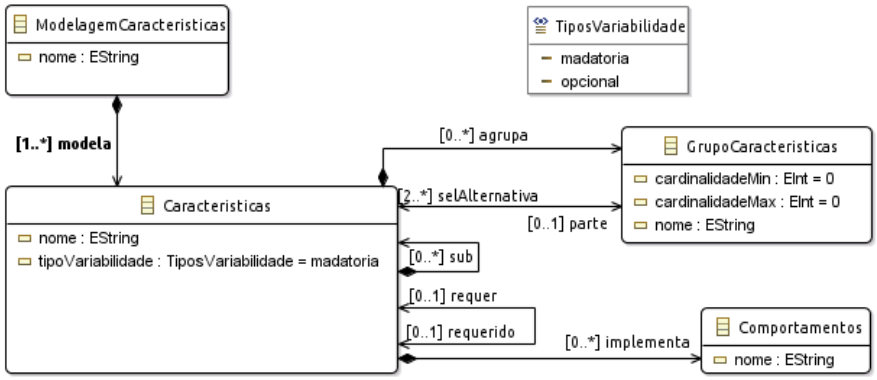


Figura 2. Metamodelo de Características.

Dado que o Modelo de Características possibilita uma hierarquização de características, o metamodelo de características possui a associação denominada **sub** que oferece este tipo de associação. Por exemplo, a Figura 1, da Subseção 2.2, apresenta um modelo de características no qual a característica *CarePlan* possui duas subcaracterísticas denominadas *Diabetes* e *Hypertension*. A metaclasse **Características** possui ainda propriedades que definem se uma característica é mandatória ou opcional, além das associações que definem que uma característica poderá requerer ou ser requerida por outra característica.

Outra metaclasse definida é denominada **GrupoCaracteristicas**, que tem como objetivo agrupar características que possuem finalidades semelhantes. Por exemplo, uma característica Conexão poderá agrupar um conjunto de características, tais como ADSL e 3G. Este tipo de modelagem permite que uma aplicação possa ser derivada com a seleção de um grupo de características. O **GrupoCaracteristicas** tem ainda como propriedades as cardinalidades mínima e máxima, que têm como objetivo definir a quantidade mínima e máxima de características ativas na utilização do grupo. Utilizando o grupo mencionado, as características ADSL e 3G poderão trabalhar em conjunto ou de forma excludente, uma característica ou outra.

A metaclasse **Comportamentos** visa dar à modelagem a flexibilidade de inclusão de novos comportamentos às características. Assim, ao modelar uma característica, esta poderá ser especificada com comportamentos específicos e inerentes as propriedades a esta característica. Assim, o metamodelo consegue cumprir com especificações determinadas pela abordagem de modelo de característica, que por sua vez pode ser utilizada para a especificação de SPLs.

Finalmente, por intermédio do Metamodelo de Características é possível estabelecer o mapeamento com propriedades, comportamentos e associações das características de uma SPL. Como exemplo, um sistema de *Healthcare* pode ser modelado com inúmeras

Características, tais como: *smartTVs, smartphones, tablets,* computadores, sensores etc. Em seguida, estas Características poderão ser selecionados para compor diversos produtos oriundos da SPL.

3.1.2. Metamodelo de Derivação

O Metamodelo de Derivação de aplicações permite a modelagem de aplicações a partir de um metamodelo SPL definido por meio do Metamodelo de Características. A derivação tem como meta a seleção das características ou grupo de características que devem compor uma aplicação nos moldes de necessidade de negócios de um determinado cliente.

A Figura 3 apresenta o Metamodelo de Derivação de Aplicações, metamodelo este que permite a derivação de aplicação oriundas do Modelo de Características definido pela modelagem de características de uma SPL.

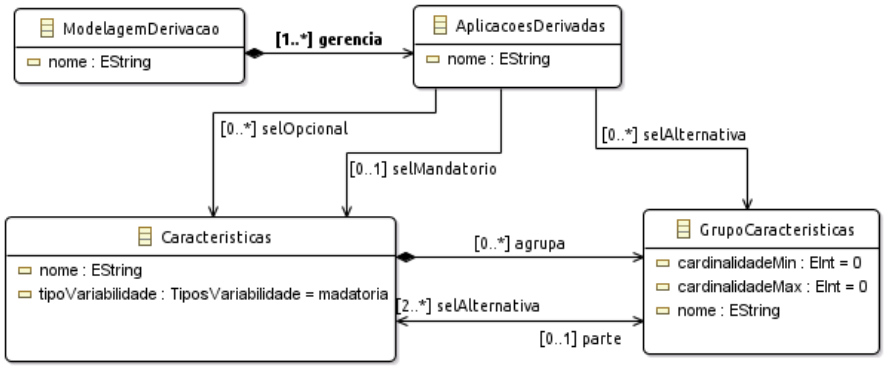


Figura 3. Metamodelo de Derivação de Aplicações.

A modelagem de derivações é iniciada a partir da metaclasse **ModelagemDerivacao** que é a raiz de modelagem de todas as aplicações derivadas da SPL. Em seguida, a metaclasse **AplicacoesDerivadas** é modelada a partir de associações com características ou grupos de características. A associação de características pode ser feita por meio de suas relações, sendo elas, a relação **selMandatorio** que deve selecionar todas as características mandatórias da SPL, e a relação **selOpcional** que seleciona todas as características opcionais relativas às necessidades do cliente.

Deste modo, o Metamodelo de Derivações permite compor produtos oriundos da SPL modelada pelo Metamodelo de Características. A modelagem de derivações deve ser considerada uma fase posterior a fase de modelagem de características, visto que as características ou grupo de características deverão ser selecionados para compor produtos, os quais são considerados derivações de uma SPL que possui uma variabilidade de características selecionáveis.

3.2. Desenvolvimento da DSML

A partir do metamodelo desenvolvido, tem início a transição deste para uma linguagem descritiva de modelagem. Para este processo, é utilizada uma ferramenta denominada Xtext [Efftinge and Völter 2006] que é instalada no ambiente de desenvolvimento Eclipse. O Xtext faz a leitura dos metamodelos e os transformam em uma gramática que

os representa. O Código 3.2 apresenta a transformação de uma meta-classe na gramática que a representa.

```
1  Caracteristicas returns Caracteristicas:
2      {Caracteristicas}
3      'Caracteristicas'
4      '{'
5          ('nome' nome=EString)?
6          ('tipoVariabilidade' tipoVariabilidade=TiposVariabilidade)?
7          ('tipoRequisicao' tipoRequisicao=TiposRequisicao)?
8          ('parte' parte=[GrupoCaracteristicas|EString])?
9          ('requerido' requerido=[Caracteristicas|EString])?
10         ('requer' requer=[Caracteristicas|EString])?
11         ('agrupa' '{' agrupa+=GrupoCaracteristicas ( "," agrupa+=
               GrupoCaracteristicas)* '}' )?
12         ('sub' '{' sub+=Caracteristicas ( "," sub+=Caracteristicas)* '}' )?
13         ('implementa' '{' implementa+=Comportamentos ( "," implementa+=
               Comportamentos)* '}' )?
14     '}';
```

Código 1. Trecho da gramática da DSML.

Com a gramática pronta, começa o desenvolvimento das classes que deverão fazer a leitura dos componentes da gramática (meta-classes, associações etc). Assim, o objetivo é que, ao modelar uma aplicação pela linguagem, uma máquina de estados faça a leitura do código e implemente um esqueleto de códigos em linguagem de propósito geral, neste caso o Java. Tal esqueleto de códigos deve apresentar propriedades, comportamentos e associações modeladas pelo modelo de características. Como exemplo, ao modelar a característica *smartphone*, uma classe é automaticamente gerada em linguagem Java, sendo que esta classe deverá possuir um esqueleto de códigos a ser entregue para um desenvolvedor.

A ferramenta XText permite a escrita de métodos baseados em máquina de estados, em outras palavras, ao utilizar a DSML para a modelagem de um modelo de características, cada objeto descrito poderá acionar um evento que deverá executar o método descrito para o objeto. Por exemplo, ao modelar a característica *smartphone*, será executado um método que deverá implementar um esqueleto de códigos que apresente propriedades, comportamentos e associações desta característica.

Outra fase do desenvolvimento da DSML é a inclusão de novas restrições de modelagem. O metamodelo desenvolvido já estabelece uma série de restrições de modelagem, por exemplo, determinando propriedades que poderão ser modeladas e associações que poderão ser realizadas. Contudo, tais restrições muitas vezes não são suficientes para manter a corretude da modelagem. A ferramenta XText possibilita o desenvolvimento de métodos, o quais podem ser considerados como uma prática da abordagem OCL [OMG 2014b], que tem por objetivo o desenvolvimento de restrições para objetos. Com base nesta abordagem foram desenvolvidos métodos que introduzem mais regras a modelagem. Para exemplificar, novas restrições foram desenvolvidas com os seguintes propósitos: uma delas obriga a modelagem de características a possuir ao menos uma característica mandatória; outra restrição obriga todas as características de um grupo a pertencer a mesma super característica; além dessa, uma restrição pode determinar que um grupo de características possua ao menos duas características.

Ao final da modelagem da SPL, por intermédio da DSML desenvolvida, o resultado deverá ser um conjunto de classes que representam a SPL, com estrutura inicial que deverá ser complementada por desenvolvimento manual. Tal desenvolvimento manual

deve-se principalmente ao fato de que as características modeladas deverão ser ligadas aos artefatos já desenvolvidos, por exemplo, uma interface que permite o sistema fazer uma notificação em um *smartphone*.

4. Validação

Com o propósito de validar a DSML, um Sistema de Healthcare (baseado na Figura 1) foi modelado. Todos os artefatos desenvolvidos nestes testes foram implantados em uma plataforma *OSGi*, mais especificamente na ferramenta *Apache Felix*.

4.1. Modelagem de Características

O primeiro passo consiste na atividade de modelagem de características, que neste artigo implementa uma parte do modelo apresentado na Figura 1 apresentado na Seção 2.2. O modelo de características apresentado tem como objetivo gerenciar um plano de cuidados para um paciente. Um plano de cuidados consiste em um conjunto de prescrições médicas que devem orientar o tratamento de um paciente.

Deste modo, a SPL deve possibilitar a derivação de produtos pertencentes a uma família de produtos, mas que tenham em sua configuração somente as necessidades específicas de um determinado usuário. Como exemplo, um produto pode ser derivado para notificar (*Reminder*) o paciente sobre os horários de medicação e medições em um dispositivo como um *smartphone*.

O Código 4.1 apresenta um trecho de código da modelagem do Sistema de *Healthcare* desenvolvido por meio da DSML construída. Deste modo, código da DSML é interpretado toda vez que é gravado, sendo assim ao gravar uma característica no modelo, os primeiros códigos em Java são gerados.

```
1  Caracteristicas 'DisplayDevice' {
2   tipoVariabilidade opcional
3    sub { Caracteristicas 'TV'{},
4     Caracteristicas 'SmartPhone'{},
5     Caracteristicas 'Tablet'{}}
6    agrupa { GrupoCaracteristicas 'DisplayDeviceGroup' {
7     cardinalidadeMin 1 cardinalidadeMax 4
8     selAlternativa (TV, SmartPhone, Tablet ...)}}
9    implementa {Comportamento 'display'} ...
```

Código 2. Trecho da Modelagem de Características (DSML).

A classe *Feature* tem como uma de suas funcionalidades estabelecer relações com as classes correspondentes a componentes armazenados em um repositório. Como já se considera que estas classes já existam em um repositório, para efeitos de testes foram desenvolvidas classes fictícias com o objetivo somente de reproduzir a conexão das *Features* com o serviço de componentes.

4.2. Derivação

O Código 4.2 apresenta parte da derivação de produto que promove o uso do modelo de características apresentado na subseção anterior. A modelagem define o desenvolvimento de um produto derivado do Modelo de Características previamente desenvolvido. O código ainda apresenta uma derivação de produto que consiste em uma aplicação de planos de cuidados para um paciente.

```
1  ProjectDSPL 'HealthCare' {
2    ModelagemDerivacao {
3        gerencia {
4      AplicacoesDerivadas 'App01' {
5        selMandatorio (CarePlan, Reminder)
6        selOpcional (smartTV, ADSL)
7        selAlternativa (DisplayDeviceGroup)
8  }}}}
```

Código 3. Trecho de Derivação de Produto (DSML).

Essa modelagem tem como principal resultado a geração de um documento XML, produzido pelo sistema gerenciador de projetos Apache Maven, o qual deverá possibilitar a compilação do projeto para que este seja instalado e executado na plataforma Apache Felix OSGi. Terminado o processo de modelagem e implementação manual do sistema, o projeto estará pronto para ser compilado pelo software de gerenciamento de projetos Maven, instalado e executado em ambiente de testes.

A validação foi implementada seguindo um roteiro que passa pela modelagem de características, derivação de produto e implementação manual. O passo seguinte é composto pela compilação do projeto, resultado de uma derivação (Sistema de *Healthcare*), instalação no ambiente de execução (Apache Feliz OSGi) e execução de notificações com a finalidade de proceder testes funcionais.

5. Considerações Finais

O desenvolvimento de sistemas ubíquos, os quais geralmente possuem a necessidade de utilização de diversos recursos ou dispositivos, demandam necessidades de implementação suscetíveis a várias mudanças para vários contextos. Este fato é decorrente de constantes mudanças nos paradigmas de utilização de tecnologias, tais como o volume de recursos tecnológicos disponíveis em um ambiente, onde tais recursos podem fazer parte de um sistema complexo e distribuído.

Diante destas necessidades, neste estudo foi desenvolvido uma DSML para implementação de determinados tipos de sistemas que possuem grande demanda de mudanças devido às necessidades de cada cliente ou usuário. Para isso, a abordagem de SPL foi utilizada por proporcionar uma variabilidade de produtos mediante um conjunto amplo de características.

Assim, o projeto teve seu início com a definição de um processo de metamodelagem, definido por conceitos de Arquitetura Dirigida a Modelos, proposta pela OMG em sua especificação MOF. Somado ao conceito de metamodelagem, o paradigma de SPL e Modelos de Características auxiliaram na construção do Metamodelo.

A linguagem desenvolvida visa mitigar a complexidade do desenvolvimento de famílias de produtos que possuem como propriedade a adaptação, por meio da derivação, às necessidades de cada usuário. Em outras palavras, um conjunto de características pode ser utilizado ou reutilizado na derivação de cada produto.

A linguagem ainda foi validada mediante a definição de cenários, como o descrito na seção de validação deste artigo. Deste modo, todos os passos para o desenvolvimento de uma SPL com a utilização da linguagem foram implementados. Tais passos são identificados como a modelagem de características, derivação de produtos, implementação manual, instalação e procedimentos de testes.

No que concerne a limitações e trabalhos futuros, este trabalho possui como principal limitação, a possibilidade do desenvolvimento de produtos que possam contar com a derivação em tempo de execução, o que seria bastante propício a ambientes com alta suscetibilidade à mudanças. Assim, como trabalho futuro, uma importante inclusão é a possibilidade de implementação de recursos que possibilitem a derivação de produtos de forma dinâmica e em tempo de execução.

Referências

Carvalho, S. T. (2013). *Modelagem de Linha de Produto de Software Dinâmica para Aplicações Ubíquas.* Tese (doutorado em computação), Instituto de Computação, Universidade Federal Fluminense, Niterói, Rio de Janeiro.

Carvalho, S. T., Copetti, A., and Loques Filho, O. G. (2011). Sistema de computação ubíqua na assistência domiciliar à saúde. *Journal Of Health Informatics*, 3(2).

Efftinge, S. and Völter, M. (2006). oAW xText: A framework for textual DSLs. In *Workshop on Modeling Symposium at Eclipse Summit*, volume 32, page 118.

Jouault, F. and Bézivin, J. (2006). KM3: a DSL for Metamodel Specification. In *Formal Methods for Open Object-Based Distributed Systems*, pages 171–185. Springer.

Kang, K. C., Cohen, S. G., Hess, J. A., Novak, W. E., and Peterson, A. S. (1990). Feature-oriented domain analysis (foda) feasibility study. Technical report, DTIC Document.

Kang, K. C., Lee, J., and Donohoe, P. (2002). Feature-oriented product line engineering. *IEEE Software*, 19(4):58–65.

Laakko, T. and Mäntylä, M. (1993). Feature modelling by incremental feature recognition. *Computer-Aided Design*, 25(8):479–492.

Mernik, M., Heering, J., and Sloane, A. M. (2005). When and how to develop domain-specific languages. *ACM computing surveys (CSUR)*, 37(4):316–344.

Ministério da Saúde do Brasil (2014). Estratégias para o cuidado da pessoa com doença crônica. Cadernos de Atenção Básica, nº35.

Northrop, L. (2008). Software product lines essentials. *Software Engineering Institute y Carnegie Mellon University. Pittsburgh*.

OMG (2014a). OMG Meta Object Facility (MOF) Core Specification, Version 2.4.2. `http://www.omg.org/spec/MOF`.

OMG (2014b). OMG Object Constraint Language (OCL) Standard document, Version 2.4. `http://www.omg.org/spec/OCL/2.4`.

Parnas, D. L. (1976). On the design and development of program families. *Software Engineering, IEEE Transactions on*, (1):1–9.

Schmidt, D. C. (2006). Guest editor's introduction: Model-driven engineering. *Computer*, 39(2):0025–31.

Van Deursen, A., Klint, P., and Visser, J. (2000). Domain-Specific Languages: An Annotated Bibliography. *Sigplan Notices*, 35(6):26–36.

Weiser, M. (1999). The computer for the 21st century. *SIGMOBILE Mob. Comput. Commun. Rev.*, 3(3):3–11.

Aplicação de mineração de dados para identificação de evasão de clientes

Ygor E.C Santana[1], **Nádia F. F. da Silva**[1], **Walison C. Moreira**[1]

[1]Instituto de Informática – Universidade Federal de Goias (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia – 74690-900 – Goiânia – GO – Brazil

ygorescatolim,nadia,walison@inf.ufg.br

Abstract. *The customer evasion rate, also known as* ***churning****, represents the amount of customers that stops doing business with a company overtime. There are plenty of reasons that drives a customer to leave a company: prices, market competition, new technologies, customer relationship, etc. And by that it becomes hard to identify which variables drives a customer churning and even harder to prevent that turnover in time. This article demonstrate and proposes the use of data mining techniques to identify variables and relationships that drives churning, allowing the use of that information in the decision making process.*

Resumo. *A taxa de evasão de clientes, conhecida também como* ***churning****, representa a quantidade de consumidores que deixam de fazer negócio com uma empresa após um certo período de tempo. Existem várias razões que levam um cliente a fazer negócio com uma empresa: Preços, competição no mercado, novas tecnologias, relacionamento com o cliente, e etc. Tornando-se complexo identificar quais dessas variáveis levam a saída de um cliente, e ainda mais difícil prevenir tal saída em tempo hábil. Este artigo irá demonstrar e propor o uso de técnicas de mineração de dados para identificar variáveis e relações que levam à evasão, permitindo o uso dessa informação no processo de tomada de decisão.*

1. Introdução

A taxa de evasão de clientes, conhecida também como ***churning***, representa a quantidade de consumidores que deixam de fazer negócio com uma empresa após um certo período de tempo. Isso é um problema enfrentado por todas empresas no mercado, seja uma pequena *startup* ou uma grande multinacional.

Evitar a saída de um consumidor é muito mais importante do que parece, visto que tal ação pode afetar a percepção de qualidade da marca, faturamento total e possíveis indicações orgânicas [Reichheld 1990], outra importante observação a ser feita é que o custo de se fidelizar um novo cliente pode ser até cinco vezes maior que de se manter um cliente fiel, além de possuírem menor lucratividade [Weber E 2008] [Toporek 2013]. Estudos ainda indicam que uma melhoria de 5% na retenção de clientes pode ser responsável pela manutenção de cerca de 35% no valor líquido médio de uma empresa [Reichheld 1996] [Fornell 1987].

Observa-se então que manter um cliente fidelizado é uma tarefa extremamente importante, e que devido à grande quantidade de variáveis e indicadores a predição e

categorização de potenciais evasões comumente só é percebida tarde demais, dificultando a possibilidade de reversão da situação [Mutanen 2006]. Partindo desse problema, a utilização de ferramentas e técnicas de mineração de dados é uma alternativa a ser considerada [Bhambri 2012][Chiu 2002].

Especificamente, como contribuições deste trabalho é possível enumerar: (i) Um estudo comparativo entre alguns algoritmos de classificação encontrados na literatura; (ii) A aplicação de algoritmos de seleção de atributos como alternativa para a redução de dimensionalidade dos conjuntos de dados envolvidos, bem como um estudo comparativo entre as técnicas utilizadas; (iii) Uma análise da distribuição de valores dos atributos e seu impacto nos resultados alcançados pelos algoritmos de classificação utilizados pela literatura relacionada.

Este trabalho está organizado conforme segue: Na Seção 2 são apresentados os principais trabalhos relacionados. Na Seção 3 é apresentada a formalização do problema de detecção de evasão de clientes sob o ponto de vista computacional. Na Seção 4 são apresentados os experimentos realizados, resultados e análises. A Seção 5 apresenta as considerações finais e indicações de trabalhos futuros.

2. Trabalhos Relacionados

Obter um modelo de evasão de clientes com boa precisão pode acarretar em uma boa vantagem estratégica na hora da tomada de uma decisão, sendo assim, naturalmente a utilização de técnicas de mineração de dados para esse tipo de problema já é um assunto abordado em estudos anteriores. Na literatura levantada percebe-se três pontos principais a serem atacados, o pré-processamento e obtenção de dados relacionados a clientes, a seleção de atributos, e a classificação supervisionada de dados.

No quesito pré-processamento e seleção de atributos, os estudos indicam uma alta correlação entre se obter um bom conjunto de dados, selecionando atributos importantes e transformando valores para formatos capazes de fornecer maior informação, e uma significativa melhora na taxa de acerto e confiabilidade dos algoritmos de classificação [Coussement 2016] [Niculescu-Mizil et al. 2009].

Em relação às técnicas de classificação, diversas foram estudadas, sendo identificado o uso de algoritmos baseados em *Decision Tree* em boa parte dos estudos levantados [Ying 2009] [Idris 2012] [Wang 2006]. Foram encontrados também estudos utilizando redes neurais [Lu 2009], algoritmos genéticos [Wai-Ho Au; Chan 2003] e *Support Vector Machines* [en XIA 2008]. A utilização de técnicas de agregação ou comitês de classificadores também pode ser observada nos estudos levantados [Niculescu-Mizil et al. 2009] [Bock and den Poel 2011].

3. Identificação de Evasão de Clientes

Existem diversas formas de se atacar o problema de evasão de clientes utilizando mineração de dados, a abordagem utilizada neste artigo consiste de duas etapas: O pré-processamento e seleção de atributos, e a aplicação de algoritmos de classificação.

A primeira etapa consiste em reduzir valores discrepantes, computar valores ausentes, e identificar quais são os melhores atributos do conjunto em relação à classe na qual as instâncias serão classificadas, tendo como objetivo preparar o conjunto de dados

para a aplicação das técnicas de classificação, buscando melhorar a precisão, acurácia e confiabilidade do modelo classificatório.

A segunda etapa consiste na aplicação de diversos algoritmos de classificação para categorizar automaticamente as instâncias de um conjunto de dados em clientes ativos e clientes que sofreram evasão.

4. Experimentos

Nesta seção será descrito o processo utilizado no artigo, os conjunto utilizados, metodologias, métricas de avaliação e resultados.

4.1. Conjuntos de Dados

Para os experimentos dois conjuntos de dados foram utilizados, ambos referentes a empresas de telecomunicação. O [Conjunto 1 2017] consiste de 109 atributos, sendo 4 atributos nominais, 4 atributos numéricos, 10 atributos temporais contendo informações dos últimos 10 meses do cliente, totalizando assim 100 atributos, e uma classe alvo indicando se o cliente deixou de realizar negócios com a empresa ou não. A base possui 15.310 instâncias, sendo 13.197 de clientes ativos e 2.113 de clientes que evadiram da empresa.

O [Conjunto 2 2017] consiste de 21 atributos, sendo 4 atributos nominais, 16 atributos numéricos e uma classe alvo indicando a evasão ou não do cliente, a base possui 3.333 instâncias, sendo 2.850 de clientes ativos e 483 que realizaram evasão.

4.2. Métricas de Avaliação

Para avaliação da qualidade, acurácia e precisão dos algoritmos de classificação foram utilizados três indicadores: A área sob a curva *ROC* [Bradley 1997] para avaliar a confiabilidade, a taxa de classificações corretas para definir a acurácia e o coeficiente *Kappa* para avaliar a precisão [Nichols et al. 2010].

Para avaliar os modelos de seleção de atributos e pré-processamento, os algoritmos de classificação foram aplicados em ambos os conjuntos de dados, antes e após o processo de seleção e pré-processamento, comparando assim o ganho ou perda em cada um dos três indicadores utilizados. Por questão de comparação, os mesmos algoritmos de classificação serão utilizados em todos os conjuntos de dados.

4.3. Pré-processamento

Um dos maiores problemas encontrados nos conjuntos de dados de telecomunicações é o desbalanceamento das classes, que pode afetar negativamente o funcionamento de diversos algoritmos de classificação [Niculescu-Mizil et al. 2009]. Para diminuir o impacto do desbalanceamento técnicas de *oversampling* e *undersampling* foram aplicadas aos conjuntos de dados utilizados. Para o conjunto de dados 1, que possui uma maior quantidade de instancias, foi aplicada a técnica de *Random Undersampling* [Idris 2012], reduzindo o número de instâncias do conjunto de dados, buscando balancear a quantidade de instâncias entre clientes ativos e clientes que evadiram. Já para o conjunto de dados 2, devido a baixa quantidade de instâncias foi aplicado o algoritmo *SMOTE* [Chawla et al. 2002], aumentando o conjunto e seguindo o mesmo objetivo de balancear as classes.

É natural que um conjunto de dados extraído de uma aplicação real possua uma quantidade mediana de *Outliers*, instâncias com valores muito discrepantes em relação ao conjunto e que também afetam negativamente o resultado obtido pelas técnicas de classificação [Deng 2004]. Os valores extremos foram identificados utilizando amplitude interquartil e removidos dos conjuntos de treinamento.

Para alguns atributos numéricos é importante a aplicação de alguma técnica de normalização, uma vez que muitos classificadores são afetados pela escala dos valores, sendo assim, alguns atributos numéricos foram normalizados no intervalo -1.0 a 1.0 [Niculescu-Mizil et al. 2009] [Pentreath 2015], evitando que os classificadores sejam afetados pela ordem de magnitude dos dados. Para os atributos numéricos que podem ser representados através de uma sequência de valores nominais tal tratamento foi realizado.

Com o objetivo de identificar quais atributos possuem mais peso ao classificar as instâncias, removendo aqueles que ou são irrelevantes ou afetam negativamente os resultados da classificação, foram aplicadas técnicas de seleção e criação de atributos [Chen 2010] [Niculescu-Mizil et al. 2009]. Para o conjunto de dados 1 foi aplicada uma técnica de redução nas séries temporais, gerando novos atributos contendo a média e o desvio padrão de cada série, tal tratamento foi necessário pois os algoritmos de classificação utilizados não são capazes de identificar se um atributo faz parte ou não de uma série temporal. Após a redução foi aplicada correlação de *Pearson* entre cada atributo e a classe alvo, descartando os atributos com pouca correlação. Para o conjunto de dados 2 além da correlação de *Pearson* foi aplicada a seleção de atributos através de *Decision Tree* [Blum and Langley 1997], gerando assim dois conjuntos com técnicas de seleção de atributos distintas para cada conjunto de dados inicial.

4.4. Resultados

Para o experimento foram selecionados cinco algoritmos de classificação, sendo eles: *Naive Bayes* (NB)[Hua 2011], *Decision Trees* (DT) [Wang 2006], *K-Nearest Neighbors* (KNN) [Cao 2007], *Convolutional Neural Networks* (CNN) [Lu 2009] e *Random Forest* (RF) [Idris 2012].

Os algoritmos foram executados utilizando a ferramenta *Weka*[Hall et al. 2009], utilizando *10-Fold cross validation* como mecanismo para validação dos testes. Para o algoritmo KNN foram utilizados 3 vizinhos próximos, para os demais foram mantidas as configurações padrão da ferramenta.

Tabela 1. Conjunto de dados 1 - % de Classificações corretas (Acurácia)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	77.1737	86.1986	83.9164	84.9517	87.4585
(2) Reduzido	77.1902	87.2840	84.4887	86.5528	**87.5686**
(3) Pearson	77.1659	86.1986	83.4024	85.7616	86.4331

No quesito acurácia o algoritmo *Random Forest* obteve o melhor resultado, classificando corretamente 87.56% das instâncias no conjunto de dados reduzido, um ganho de $0,1101$ em relação ao conjunto inicial. Em todos os algoritmos o conjunto reduzido aumentou levemente a acurácia, sendo o maior ganho encontrado no algoritmo *Decision Tree*, obtendo um aumento de $1,6011\%$. Os resultados podem ser vistos na Tabela 1.

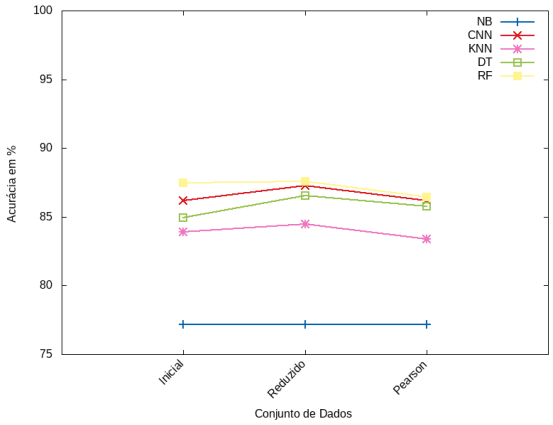


Figura 1. Acurácia - Conjunto de dados 1

Os atributos selecionados pela correlação de *Pearson* reduziram a acurácia levemente em todos os algoritmos, o que indica que algum dos atributos descartados afetavam levemente a classe alvo, como pode ser observado na Figura 1.

Tabela 2. Conjunto de dados 1 - Coeficiente Kappa (Precisão)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	0.0724	0.0000	0.0895	0.1991	**0.2063**
(2) Reduzido	0.0710	-0.0000	0.0786	0.0851	0.0825
(3) Pearson	0.0704	0.0000	0.0822	0.0723	0.1280

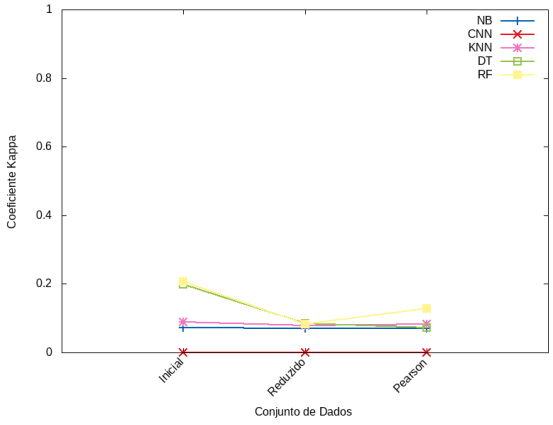


Figura 2. Precisão - Conjunto de dados 1

O coeficiente *Kappa* indica que mesmo no conjunto de dados inicial a concordância é extremamente baixa ($<= 0.20$), pressupondo que as instâncias informadas oferecem pouco ou nenhum valor ao modelo de categorização, isso pode acontecer devido ao fato de que apesar do conjunto possuir uma grande quantidade de instâncias não há variedade o suficiente de valores.

Na Tabela 2 pode se ver que para todas as aplicações do algoritmo *Convolutional Neural Networks* o índice de concordância foi próximo a zero, o que indica que para o

conjunto de dados em questão a concordância foi equivalente ao que se é esperado quando se classifica as instâncias de modo aleatório.

A Figura 2 demonstra que o pré-processamento obteve impacto negativo em todos os algoritmos testados, indicando que a redução de atributos não atingiu o objetivo de melhorar a precisão do modelo.

Tabela 3. Conjunto de dados 1 - AUC (Confiabilidade)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	0.6264	0.6915	0.5835	0.6118	**0.7460**
(2) Reduzido	0.6407	0.6962	0.5819	0.5778	0.7021
(3) Pearson	0.6319	0.6718	0.6047	0.6303	0.7105

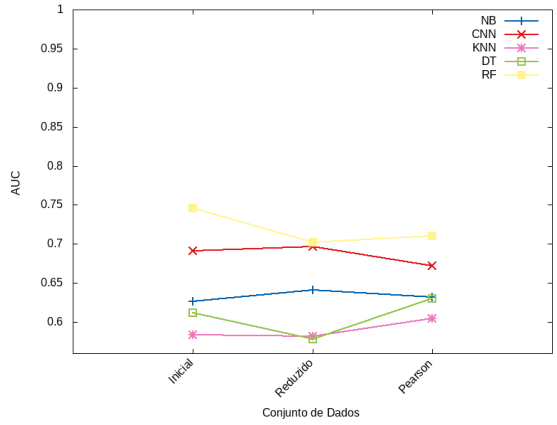


Figura 3. Confiabilidade - Conjunto de dados 1

No quesito confiabilidade o algoritmo *Random Forest* obteve novamente o melhor resultado, alcançando uma razão de $0,7460$ em relação à curva *AUC* no conjunto inicial. Ambas as técnicas de pré-processamento afetaram os algoritmos de forma distinta, onde o conjunto com os atributos selecionados pela correlação de *Pearson* acarretou em uma leve melhora em três dos cinco algoritmos, *Naive Bayes*, *KNN* e *Decision Tree*. Os resultados podem ser observados na Tabela 3.

A curva *AUC* indica que a maioria dos algoritmos é levemente mais confiável que apenas uma seleção ao acaso ($AUC = 0.5$), concordando assim com o coeficiente *Kappa*, isso indica uma baixa qualidade nos dados do conjunto, o que dificulta a classificação das instâncias, não alcançando assim resultados satisfatórios nos testes realizados. (Ver Figuras 1, 2 e 3).

Tabela 4. Conjunto de Dados 2 - % de Classificações corretas (Acurácia)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	88.1456	85.8716	88.1939	94.4794	95.5834
(2) DT	87.1166	85.7007	90.9276	94.5665	**95.8654**
(3) Pearson	88.1937	85.9077	91.0022	94.6235	95.1933

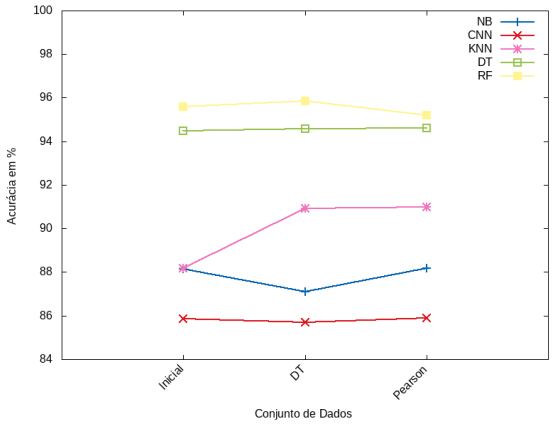


Figura 4. Acurácia - Conjunto de dados 2

Em todos os conjuntos de dados o algoritmo *Random Forest* alcançou ótimos resultados, tendo maior precisão no conjunto pré-processado utilizando *Decision Trees*, um ganho de 0,28% em relação ao conjunto inicial. Os resultados obtidos se encontram na Tabela 4.

Observa-se que diferentes algoritmos reagem de modo distinto aos pré-processamentos aplicados, e que no geral o pré-processamento foi benéfico para a precisão da maioria dos classificadores utilizados (ver Figura 4).

Tabela 5. Conjunto de dados 2 - Coeficiente Kappa (Precisão)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	0.4499	0.0574	0.3924	0.7579	0.8045
(2) DT	0.3249	0.0315	0.5674	0.7616	**0.8191**
(3) Pearson	0.4525	0.0675	0.5829	0.7647	0.7905

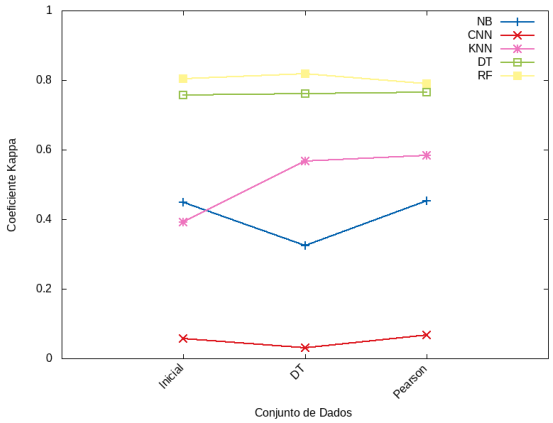


Figura 5. Precisão - Conjunto de dados 2

Na Tabela 5 observa se que para o conjunto de dados inicial os algoritmos *Decision Tree* e *Random Forest* identificaram respectivamente uma concordância substancial ($>= 0.6$) e excelente ($>= 0.81$), o que demonstra que os atributos possuem uma boa correlação

no modelo de classificação. Todos os algoritmos com exceção do *Convolutional Neural Network* identificaram ao menos uma boa correlação entre os atributos, atestando assim uma boa precisão do modelo de classificação.

Em geral, o pré-processamento melhorou o coeficiente *Kappa*, o que demonstra que alguns dos atributos descartados afetavam negativamente a precisão dos modelos utilizados (ver Figura 5).

Tabela 6. Conjunto de dados 2 - AUC (Confiabilidade)

Conjunto de dados	(NB)	(CNN)	(KNN)	(DT)	(RF)
(1) Inicial	0.8348	0.8022	0.7725	0.8667	0.9125
(2) DT	0.8507	0.8158	0.8401	0.8663	**0.9159**
(3) Pearson	0.8349	0.8085	0.8351	0.8832	0.9102

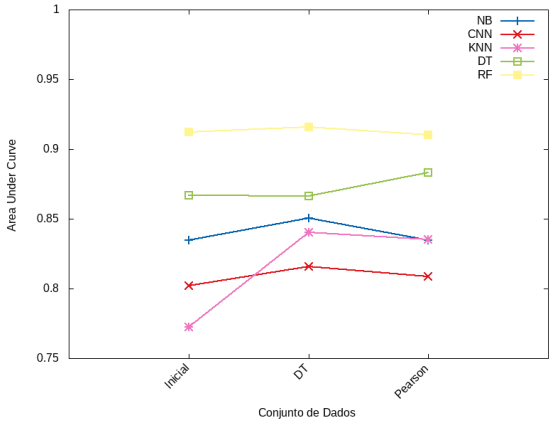


Figura 6. Confiabilidade - Conjunto de dados 2

Novamente os algoritmos *Decision Tree* e *Random Forest* alcançaram os melhores resultados, tendo o último alcançado uma confiabilidade de 0.9159 no conjunto de dados pré-processado utilizando *Decision Trees*. Ocorreram ganhos consideráveis em dois dos algoritmos de classificação, sendo 0.676 no *KNN* e $0,159$ no *Naive Bayes*. Os resultados podem ser vistos na Tabela 6.

A correlação entre o coeficiente *Kappa* e a *AUC*, ambos próximos a 1 indicam que o algoritmo *Random Forest* apresenta uma interessante combinação de precisão e confiabilidade, sendo assim um modelo classificatório adequado para o conjunto de dados trabalhado.

5. Conclusão e Trabalhos Futuros

Os experimentos indicam que algoritmos baseados em *Desision Tree* alcançam interessantes resultados nos conjuntos de dados propostos, onde o algoritmo *Random Forest* alcançou os melhores resultados em ambos os conjuntos testados, resultados que coincidem com o observado em alguns dos trabalhos relacionados [Niculescu-Mizil et al. 2009] [Idris 2012][Ying 2009].

No quesito qualidade dos dados, pode se observar que mesmo utilizando um conjunto de dados que possui uma boa quantidade de instâncias não é garantida a precisão do

algoritmo. No caso do conjunto 1 pode se observar que a distribuição de dados em todos os atributos está concentrada em uma variedade muito baixa de valores, o que dificulta o trabalho de classificação, onde mesmo após um severo trabalho de pré-processamento não foram alcançados resultados satisfatórios (Ver Tabelas 1, 2 e 3).

Para trabalhos futuros pretende-se utilizar a tarefa de regressão, buscando identificar melhor os clientes em grupos de risco, não só em ativos e com risco de evasão, mas também na geração de um *Scoring* permitindo melhor identificação dos riscos de evasão. Também será estudado um modelo preditivo, capaz de indicar não só a chance de um cliente evadir da empresa, mas também quando isso pode acontecer, permitindo assim o planejamento de uma estratégia antes do fato se consumar, gerando com isso uma ferramenta de apoio a decisão.

Referências

Bhambri, V. (2012). Data mining as a tool to predict churn behavior of customers. *International Journal of Computer & Organization Trends*, 2.

Blum, A. L. and Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97.

Bock, K. W. D. and den Poel, D. V. (2011). An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications*, 38.

Bradley, A. P. (1997). The use of the area under the roc curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30.

Cao, Y. Z. J. Q. H. S. J. (2007). A hybrid knn-lr classifier and its application in customer churn prediction.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Inteligence Research*.

Chen, C.-F. T. M.-Y. (2010). Variable selection by association rules for customer churn prediction of multimedia on demand. *Expert Systems with Applications*, 37.

Chiu, C.-P. W. I.-T. (2002). Turning telecommunications call details to churn prediction: a data mining approach. *Expert Systems with Applications*, 23.

Conjunto 1 (2017). Churning in telecom industry. `https://bigml.com/user/francisco/gallery/dataset/5163ad540c0b5e5b22000383`. Acessado em: 2017-10-27.

Conjunto 2 (2017). Data for telephone company customers. `http://azzalini.stat.unipd.it/Book-DM/telekom-www.zip`. Acessado em: 2017-10-27.

Coussement, Kristof; Lessmann, S. V. G. (2016). A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry. *Decision Support Systems*.

Deng, Z. H. X. X. J. Z. H. S. (2004). Mining class outliers: concepts, algorithms and applications in crm. *Expert Systems with Applications*, 27.

en XIA, G. (2008). Model of customer churn prediction on support vector machine. *Systems Engineering - Theory and Practice*, 28.

Fornell, C., . W. B. (1987). Defensive marketing strategy by customer complaint management: A theoretical anlaysis. *Journal of Marketing Research*, 24.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations*, 11(1):10–18.

Hua, G. (2011). Customer relationship management based on data mining technique — naive bayesian classifier. *IEEE 2011 International Conference on E-Business and E-Government (ICEE)*.

Idris, Adnan; Rizwan, M. K. A. (2012). Churn prediction in telecom using random forest and pso based data balancing in combination with various feature selection strategies. *Computers and Electrical Engineering*, 38.

Lu, C.-F. T. Y.-H. (2009). Customer churn prediction by hybrid neural networks. *Expert Systems with Applications*, 36.

Mutanen, T. (2006). Customer churn analysis – a case study. *Journal of Product and Brand Management*.

Nichols, T. R., Wisner, P. M., Cripe, G., and Gulabchand, L. (2010). Putting the kappa statistic to use. *The Quality Assurance Journal*, 13.

Niculescu-Mizil, A., Perlich, C., Swirszcz, G., Sindhwani, V., Liu, Y., Melville, P., Wang, D., Xiao, J., Hu, J., Singh, M., Shang, W. X., and Zhu, Y. F. (2009). Winning the kdd cup orange challenge with ensemble selection. *Proceedings of KDD-Cup 2009 Competition*, 7:23–34.

Pentreath, N. (2015). *Machine Learning with Spark - Tackle Big Data with Powerful Spark Machine Learning Algorithms*. Packt Publishing, 1st edition.

Reichheld, F. F. (1996). The loyalty effect: The hidden force behind growth, profits and lasting value. *Harvard Economic Press*.

Reichheld, F. F., S.-W. E. (1990). Zero defection: Quality comes to services. *Harvard Business Review*, 68.

Toporek, A. (2013). When customers stick: Customer retention by the numbers (infographic). *Customers that Stick*.

Wai-Ho Au; Chan, K. X. Y. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, 7.

Wang, S.-Y. H. D. C. Y. H.-Y. (2006). Applying data mining to telecom churn management. *Expert Systems with Applications*, 31.

Weber E, A. (2008). B2b customer experience priorities in an economic downturn: Key customer usability initiatives in a soft economy. *Forrester Research*.

Ying, Y. X. X. L. E. N. W. (2009). Customer churn prediction using improved balanced random forests. *Expert Systems with Applications*, 36.

Computação Evolucionária Aplicada à Otimização de Trajetos e Alocação de Torres de Linhas de Transmissão de Energia Elétrica

Caio José F. Póvoa, Leonardo da Cunha Brito, Pedro H. da Silva Palhares

Escola de Engenharia Elétrica, Mecânica e da Computação – Universidade Federal de Goiás (UFG)

CEP 74605-010 – Goiânia – GO – Brazil

caiopovoa@gmail.com, brito.lc@gmail.com, phpalhares@gmail.com

Abstract. *This paper presents a new method for electrical transmission line routing optimization. The proposed algorithm is based on Evolutionary Computation and the concept of speciation by mutation. The objectives to be achieved here are, as a matter of priority, to meet project constraints, and, then, reduce costs. This process uses an elevation map as a reference, from which the routing optimization takes place along with tower allocation, respecting imposed constraints. The results show the efficiency of the proposed approach.*

Resumo. *Este artigo apresenta um novo método para otimização de trajetos de linhas de transmissão de energia elétrica. O algoritmo proposto é baseado na computação evolucionária e no conceito de especiação por mutação. O objetivo a ser atingido é, prioritariamente, o atendimento a restrições de projeto, e, num seguindo momento, a redução de custos. O processo toma como base o mapa de relevo da região de interesse, a partir do qual se faz a otimização do trajeto juntamente com a alocação das torres, respeitando as restrições impostas. Os resultados evidenciam a eficiência da abordagem proposta.*

1. Introdução

Objetivando-se o atendimento a novos centros de carga e o aumento da capacidade do sistema elétrico de potência, é recorrente a necessidade de obras de expansão da malha já existente. Para isso, são realizados planejamentos específicos detalhados, por se tratar de um tipo de intervenção de alto custo financeiro e elevado impacto socioambiental.

O projeto de uma nova linha de transmissão, se feito de maneira manual, é um processo complexo que demanda tempo e, de maneira geral, não leva a uma solução ótima. Deste modo, a utilização de ferramentas de otimização computacional surge como elemento facilitador.

A literatura existente relacionada ao problema de encaminhamento de linhas de transmissão mostra que essa questão é tratada, na maioria dos casos, tomando-se como base mapas provenientes de Sistema de Informação Geográfica (SIG), a partir dos quais aplicam-se algoritmos de otimização para buscar um trajeto ótimo, minimizando-se os

custos associados. Monteiro et al. (2005) utiliza a Programação Dinâmica para o processo de encaminhamento de linhas de transmissão juntamente com a divisão do mapa em regiões com custos associados. Shu et al. (2012) realiza o processo de otimização com base em Programação Inteira Mista. Depreende-se de pesquisa bibliográfica relacionada ao assunto que os processos de otimização propostos têm como objetivo a busca pelo melhor trajeto, porém sem que seja tratado o problema de distribuição e alocação de estruturas ao longo do caminho escolhido.

O presente trabalho propõe um método que permite encontrar soluções para o problema de alocação de torres de transmissão, respeitando as restrições técnicas, estruturais e construtivas e buscando a minimização de custos. Associado a essa abordagem, o algoritmo implementa processo simultâneo de melhoria no traçado da linha, tendo em vista os obstáculos de relevo da região de interesse.

A base do método proposto é a Computação Evolucionária, que faz uso de modelos computacionais de processos evolutivos, tais como a adaptação de indivíduos ao longo das gerações, de modo a melhorar a capacidade de sobrevivência, a partir de processos como a seleção natural, sobrevivência do mais adaptado e reprodução, Engelbrecht (2007).

O restante do artigo organiza-se da seguinte forma. A seção 2 trata da formulação do problema, detalhando a abordagem escolhida, as funções de avaliação e apresenta um pseudocódigo resumindo o funcionamento do algoritmo. Na seção 3 são apresentados resultados computacionais para um exemplo de projeto. As conclusões estão presentes na seção 4.

2. Formulação do Problema

O problema abordado é a otimização de novos trechos de linhas de transmissão, objetivando a redução de custos e respeitando restrições. Como ponto de partida, tem-se um mapa topográfico da região de interesse, as características estruturais das torres a serem instaladas, restrições a serem respeitadas e um trajeto dado inicialmente pelo usuário, interligando os pontos inicial e final.

2.1. Abordagem

Sob o ponto de vista da Computação Evolucionária aplicada ao estudo em questão, temos o conjunto de torres de sustentação dos alimentadores (cabos) que transportam energia elétrica entre a origem e o destino como sendo os indivíduos da população de determinada geração. As características desses indivíduos são representadas na forma de cromossomo e a posição das torres ao longo do trajeto, excluindo-se as torres inicial e final, cujas posições são fixas, são os genes que o compõem. Por sua vez, cada um dos valores que as variáveis otimizáveis do problema podem assumir corresponde aos chamados alelos.

O método descrito parte de uma solução inicialmente gerada tomando como base o trajeto traçado pelo usuário e o atendimento a algumas das restrições impostas. A partir de perturbações dos genes do indivíduo inicial, gera-se a população da primeira geração, que passa a ser composta por indivíduos diferentes entre si, porém de mesma espécie, isto é, com a mesma quantidade de genes. As etapas posteriores dão sequência ao processo evolucionário, com a aplicação de operadores de reprodução e de seleção.

A geração de novas populações envolve a ocorrência de dois fenômenos evolutivos: a mutação e a recombinação gênica. O primeiro é o processo de mudança aleatória de genes de um cromossomo, com a finalidade de introduzir novo material genético na população. O último, por sua vez, consiste em criar um ou mais novos indivíduos a partir da combinação de materiais genéticos aleatoriamente selecionados de dois ou mais pais, Engelbrecht (2007). No algoritmo apresentado, a mutação se manifesta a partir da perturbação da posição de uma determinada torre de um indivíduo selecionado. Já a recombinação gênica é aqui representada como a permuta de trechos contínuos de materiais de iguais tamanhos entre indivíduos diferentes de uma mesma espécie. Visando ao aperfeiçoamento do processo, a faixa de valores dentro da qual as variáveis podem oscilar é dinâmica e vai ser alterando conforme a quantidade de iterações bem-sucedidas, isto é, que conseguiram melhora em relação à anterior.

Tendo em vista a natureza do problema, deve-se otimizar não apenas o posicionamento das torres do trajeto, mas também a quantidade delas, levando assim à necessidade de se gerar uma população com indivíduos de diferentes espécies. Para contemplar essa possibilidade, inseriu-se no algoritmo o conceito de especiação por mutação, que trata da criação de novas espécies a partir das existentes, Berlocher (1998). Com isso, a inserção ou retirada aleatória de genes do cromossomo no momento da reprodução dá origem a novas espécies na população, gerando propostas de solução com quantidades diferentes de torres. Estabelecida a variabilidade de soluções, constrói-se um cenário de competição interespecífica, Townsend et al. (2000), em que espécies diferentes disputam entre si pela sobrevivência em um mesmo ambiente, sobrevivendo na próxima geração os indivíduos mais bem adaptados, uma vez que o tamanho da população é constante. Com isso, o processo evolutivo abarca todo o espaço de busca das variáveis de interesse.

A Figura 1 representa os fenômenos reprodutivos considerados durante o processo evolutivo.

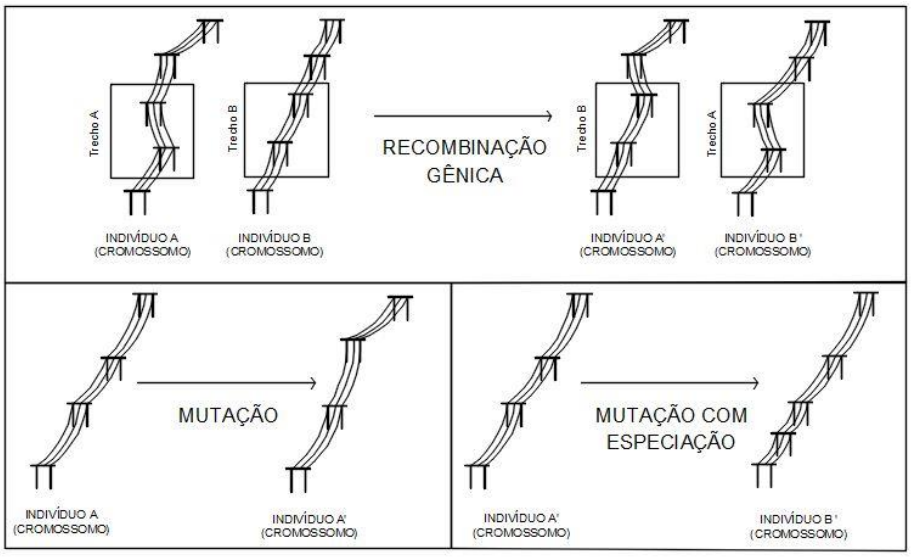


Figura 1. Operadores reprodutivos

2.2. Função de Avaliação

A definição do grau de adaptação de uma espécie é feita no problema por duas funções-objetivo que devem ser minimizadas de maneira simultânea. A cada iteração, os indivíduos são avaliados a partir delas, e aqueles que possuem melhores resultados passam para a próxima geração. Para aumentar a variabilidade de indivíduos e contornar a situação de estagnação em ótimos locais, admite-se, com determinada probabilidade, que alguns indivíduos que não estão entre os melhores passem para a próxima geração.

A primeira função está vinculada ao objetivo de minimização de custos e consiste no levantamento direto de valores monetários para a instalação da linha. A segunda função avalia o fator de qualidade do projeto, que leva em consideração as restrições do problema, e apresenta um caráter restritivo que a torna prioritária em relação à função custo.

Os parâmetros restritivos consideram limitações técnicas, tais como a extensão do vão entre torres adjacentes, a deflexão angular (mudança de direção) dos cabos no ponto de fixação em cada torre e a distância mínima entre o cabo e o solo. Para avaliar o último parâmetro, dadas as posições de cada torre na solução em análise, é simulada a catenária formada pelo cabo baseada nas características do cabo e na tração à qual ele está submetido. Adicionalmente, é feita uma simulação das torres a partir das características estruturais fornecidas pelo usuário, para avaliação dos esforços mecânicos aos quais cada uma delas é submetida. Dessa forma, calcula-se a deflexão resultante em cada estrutura, comparando-se o valor obtido com um limite máximo pré-determinado pelo projetista.

Por fim, o aspecto construtivo da linha de transmissão é levado em conta, procurando-se minimizar os obstáculos de acesso à região, através da limitação do encaminhamento da rede por trechos com inclinações muito acentuadas. Em termos práticos, muitas vezes é inviável encontrar um caminho que obedeça integralmente a esse tipo de condição ao longo de toda a sua extensão. Para isso, deve ser estabelecida uma tolerância, permitindo-se que uma determinada parcela da linha possa se localizar em regiões mais inclinadas.

Cada uma das restrições é normalizada dentro da geração e, só então, são avaliadas conjuntamente de maneira ponderada para cada indivíduo, de modo que eventual violação de uma ou mais é punida com o aumento de um fator de qualidade. A minimização desse fator, com a anulação de seu valor, representa o atendimento a todos os critérios restritivos. Tendo em vista o fato de que o atendimento a todas as restrições é imprescindível para a exequibilidade do projeto, a avaliação do custo é feita apenas para os indivíduos com fator de qualidade nulo, isto é, que atendem a todas as restrições.

2.3. O algoritmo de otimização

O procedimento para a resolução do problema pode ser resumido no pseudocódigo a seguir:

1) Criar e inicializar os parâmetros gerais de projeto e de reprodução (mutações e recombinação gênica);
2) Definir o trajeto inicial da linha de transmissão no mapa de relevo da região;

3) Aplicar rotina de geração de solução inicial;
4) Inicializar geração $g = 0$;
5) Perturbar a solução inicial para criar os demais membros da população da primeira geração *Pop(g)*, de tamanho *N*;
6) Avaliar e classificar indivíduos da população *Pop(g)*;
7) Aplicar reprodução à população da geração atual;
8) Avaliar e classificar pais e filhos gerados;
9) Selecionar os melhores indivíduos para a próxima geração;
10) Incluir indivíduos que não estão entre os melhores na próxima geração, com probabilidade *pi*;
11) Avançar para a próxima geração $g = g+1$;
12) Retornar a 7 até que o critério de parada seja atingido;

3. Resultados computacionais

O algoritmo de otimização desenvolvido no MATLAB foi testado utilizando um mapa de relevo com base no qual deseja-se interligar dois pontos com uma linha de transmissão de energia. Para isso, inicialmente, é definido pelo projetista um trajeto prévio entre os pontos A e B, a partir do qual gera-se uma solução inicial para o problema, a qual será otimizada para a busca de melhores resultados.

Para a execução do programa, utilizou-se uma quantidade de 1500 gerações, cada qual com 50 indivíduos. Deve-se reavaliar os parâmetros de geração de acordo com a extensão da rede, uma vez que trechos maiores tendem a ter uma convergência mais lenta para a resposta ótima.

A partir da observação da evolução do processo a cada geração, indicada na Figura 2, verifica-se que, no início da execução, há um aumento do fator de qualidade absoluto. Essa elevação é motivada pelo fato de os componentes do fator de qualidade serem normalizados individualmente e somados de maneira ponderada a cada geração. Sendo assim, a melhoria das restrições é um comparativo de valores normalizados a cada iteração e não tem uma relação direta com a tendência da curva de fator de qualidade absoluto.

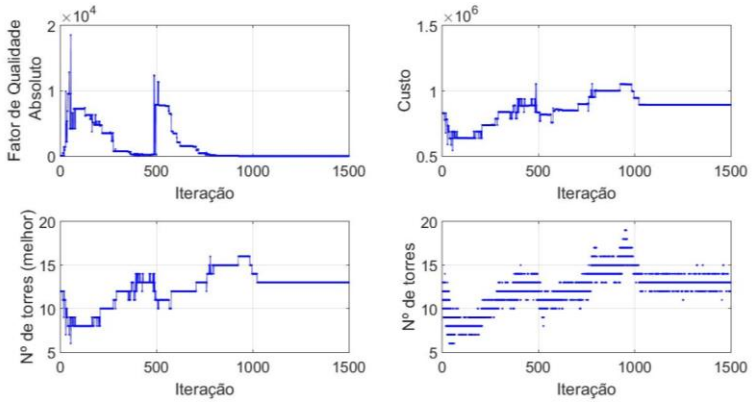


Figura 2. Desenvolvimento do processo de evolução

Inicialmente, fica evidente uma elevação de custos, que tem como razão a busca prioritária pelo atendimento das restrições. No entanto, na geração 941, todas as restrições de projeto passaram a ser atendidas pelo indivíduo mais adaptado da população e, a partir desse momento, a otimização de custos é executada, havendo uma queda de seus valores

O último gráfico evidencia a ocorrência da especiação por mutação, em paralelo com o estabelecimento da competição interespecífica no meio. Na geração de origem, tem-se apenas soluções com a quantidade de torres do indivíduo inicialmente gerado, porém, nas demais iterações passam a surgir novas espécies com quantidades de torres diversas e que passam a competir entre si. Nesse processo, espécies são extintas e geradas e sobrevivem as mais adaptadas.

Por fim, é possível observar na Figura 3 que, ao fim do processo, houve um desvio em relação à linha reta originalmente traçada entre os pontos A e B que permitiu o encaminhamento da rede por regiões de relevo mais suave. É importante destacar também que a visão de perfil da rede permite visualizar a alocação das torres no terreno e as catenárias formadas pelos cabos. A linha superior presente no perfil representa os cabos, a intermediária determina a altura mínima que o cabo pode estar do solo, a inferior representa o solo e as verticais representam as torres. É possível observar que, em todos os trechos, a catenária encontra-se acima da altura mínima, respeitando-se assim essa restrição técnica.

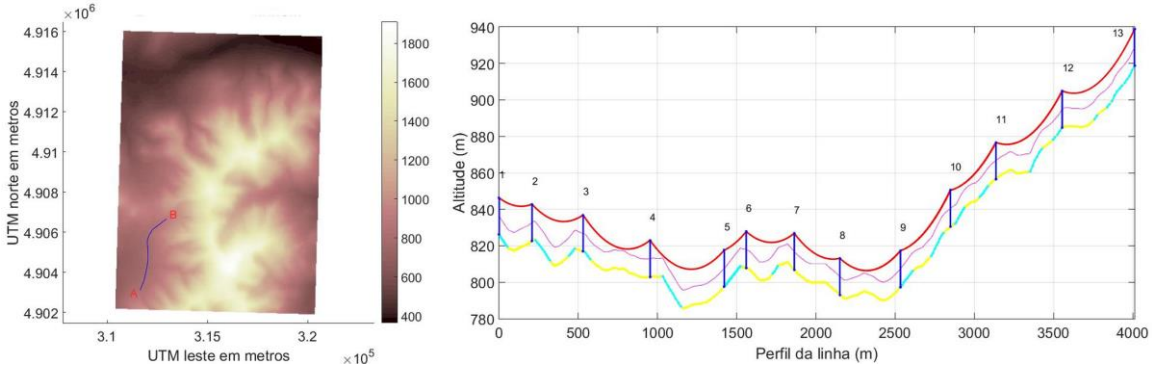


Figura 3. Mapa de relevo e perfil da linha otimizada

Cabe ressaltar que o método foi aplicado a outros casos de encaminhamentos de linhas de transmissão, tendo-se obtido resultados satisfatórios.

4. Conclusões

O objetivo principal deste artigo foi mostrar uma nova abordagem para o problema de encaminhamento de linhas de transmissão de energia elétrica, visando a redução de custos e o atendimento a restrições existentes, fazendo uso da Computação Evolucionária. Para isso, foi utilizado como referência o mapa de relevo da região em que se deseja construir a linha.

Devido às particularidades do tipo de projeto em questão, implementou-se o conceito de especiação por mutação, associado à ocorrência de competição interespecífica, para que os indivíduos de cada geração pudessem abranger todo o espaço de busca de interesse. Deste modo, foram gerados projetos de linhas com

quantidades de torres, encaminhamentos e posicionamento de estruturas diversas, que foram comparadas entre si para que ocorresse o processo de evolução.

O algoritmo implementado inova ao realizar uma otimização automatizada e simultânea do trajeto e da alocação de torres ao longo dele. Para isso são levados em conta os aspectos de economicidade e restrições de caráter estrutural, técnico e construtivo. Como as restrições são condições estritamente necessárias para a possibilidade de execução do projeto, sua avaliação é priorizada, devendo todas elas serem atendidas antes de se iniciar a busca pelas opções mais econômicas.

O artigo tratou de um processo baseado no mapa topográfico de uma região, partindo-se da premissa de que não há outras limitações, como as de caráter socioeconômico, ambiental ou de segurança, no local. Deste modo, futuros trabalhos podem utilizar o método de otimização proposto e implementar a previsão de punições a regiões do mapa em que haja quaisquer outros tipos de impedimentos à passagem da linha de transmissão, tais como os abordados por Gill et al. (2006).

5. Agradecimentos

Os autores agradecem à Fundação de Amparo à Pesquisa do Estado de Goiás (FAPEG), à Agência Nacional de Energia Elétrica (ANEEL) e à Companhia Energética do Estado de Goiás (CELG Distribuição).

6. Referências

Engelbrecht, A. P. (2007) "Computational Intelligence: An Introduction", 2ª edição.

Berlocher, S. H. (1998) "Origins: a Brief History of Research on Speciation". Em Howard, D. J., Berlocher, S. H. *Endless Forms: Species and Speciation*, New York, Oxford: Oxford University Press.

Townsend, C. R., Begon, M.; Harper, J. L (2000). "Fundamentos em Ecologia", 3ª edição.

Gill, R. S., Jewell, W. T., Grossardt, T., Bailey, K. (2006) "Landscape Features in Transmission Line Routing", em *IEEE Transmission and Distribution Conference and Exhibition*, 2006.

Monteiro, M., Ramírez-Rosado, I. J., Miranda, V., Zorzano-Santamaría, P. J., García-Garrido, E. and Fernandéz-Jiménez, L. A. (2005) "GIS Spatial Analysis Applied to Electric Line Routing Optimization", em *IEEE Transaction on Power Delivery*, vol.20, no. 2, páginas 934-942, Abril 2005.

Shu, J., Wu, L., Shahidehpour, M., Zhang L. and Han, B. (2012) "A New Method for Spatial Power Network Planning in Complicated Environments", em *IEEE Transactions on Power Systems*, vol. 27, no.1, páginas 381-389, Fevereiro 2012.

Identificação de surtos de altas frequências em sinais de Eletroencefalografia utilizando Transformada de StockWell e Decomposição em Valores Singulares

Marcela G. Magalhães[1], **Lucas F. Cruz**[1], **Rodrigo P. Lemos**[1]

[1]Escola de Engenharia Elétrica, Mecânica e de Computação – Universidade Federal de Goiás (UFG) CEP 74.605-010 – Goiânia – GO – Brasil

fiorini.cruz@gmail.com, mar.cela.gm@hotmail.com, rodrigo.p.lemos@gmail.com

Abstract. *The amount of studies on the development of fast and accurate diagnostic methods grows every year due to the large number of diseases that affect the human being, one of them being Epilepsy that affects about 1% of world's population. Several methods are used to aid clinical evaluation, among them those using Electroencephalography signals. In order to improve the analysis of those signals and find the areas related to epileptic seizures, this work proposes the use of the Stockwell Transform combined with the Singular Value Decomposition for the identification of frequency peaks above 80Hz, which characterize high frequency oscillations, a new biomarker in the study of Epilepsy. First, this article brings some basic concepts of Stockwell Transform and Singular Value Decomposition. Then, the proposed algorithm is introduced and used to build a time-frequency spectrum whose peaks allow locating those events in the EEG signal. Finally, some results evidence the effectiveness of the proposed algorithm.*

Resumo. *Estudos para o desenvolvimento de métodos diagnósticos rápidos e precisos crescem a cada ano pela grande quantidade de doenças que acometem o ser humano, sendo uma delas a Epilepsia, afetando cerca de 1% da população mundial. Diversos métodos são utilizados para auxiliar na avaliação clínica entre eles os sinais obtidos pela Eletroencefalografia. Para melhorar a análise desses sinais e encontrar as áreas relacionadas aos surtos epiléticos, este trabalho propõe o uso da Transformada de Stockwell combinada com a Decomposição em Valores Singulares para a identificação de picos de frequência acima de 80 Hz, que caracterizam as oscilações em alta frequência, um novo biomarcador no estudo da Epilepsia. Primeiro, este artigo apresenta os conceitos básicos da Transformada de Stockwell e da Decomposição em Valores Singulares, em seguida ele introduz o algoritmo proposto que constrói um espectro tempo-frequência para localização desses eventos no sinal de EEG. Os resultados obtidos evidenciam a efetividade do algoritmo proposto.*

1. Introdução

A necessidade de diagnósticos mais precisos está presente em uma variedade de doenças e uma dessas é a Epilepsia, relatada desde a Roma Antiga. Caracterizada como uma desordem neurológica, que atinge aproximadamente 1% da população mundial[Xia et al. 2015], ela é diagnosticada normalmente por meios clínicos e com

ajuda de exames de ressonância magnética, tomografia computadorizada ou Eletroencefalografia (EEG).

O eletroencefalograma data de 1929, quando o psiquiatra alemão Hans Berger apresentou os registros de sinais elétricos cerebrais captados de forma não-invasiva e constatou que eles mudavam dependendo do estado funcional do cérebro. Várias pesquisas em diversos setores utilizam os sinais captados por EEG, e especialmente no diagnóstico de epilepsia descobriu-se o grande potencial do registro das oscilações em alta frequência (HFO), acima de 80 Hz, para caracterização de surtos e marcação de áreas cerebrais epileptogênicas [Zijlmans et al. 2012].

No âmbito de análises desses sinais, Xia et al. apresenta um método para detecção de surtos epiléticos por meio da Transformada de StockWell (Transformada S) combinada com a Decomposição em Valores Singulares (SVD) em frequências abaixo de 80 Hz. Este artigo propõe, então, um campo de busca voltado às frequências acima de 80 Hz, visto que ocorrem diversas oscilações em frequências superiores a tal limite, podendo estas estarem ligadas à surtos, oferecendo uma forma mais simples de identificação de picos de alta frequência do sinal captado.

A Transformada S realiza um mapeamento tempo-frequência derivada da Transformada Wavelet Contínua (CWT) com uma janela Gaussiana. Ela foi apresentada em 1996 no artigo dos pesquisadores R. G. Stockwell, L. Mansinha e R. P. Lowe para análise de dados geofísicos e é muito utilizada na detecção de distúrbios em redes de distribuição de energia. A SVD aparece como uma ferramenta para auxiliar na manipulação das matrizes associadas ao mapeamento tempo-frequência produzidas pela Transformada S e atua também na redução do ruído [Xia et al. 2015] [Stockwell et al. 1996]

Com essa união de conceitos foi desenvolvido um algoritmo em *Matlab*© para demonstrar sua aplicação a um banco de dados de sinais de *EEG* público do *Massachusetts Institute of Technology* (MIT) [Shoeb 2009]. Foram traçados gráficos cujos picos delimitam os trechos das oscilações de alta frequência do EEG, permitindo localizar os surtos de epilepsia. Os resultados evidenciaram a efetividade da proposta.

2. Surtos em sinais de *EEG*

Os sinais captados por meio de *EEG* são aplicados em estudos de diferentes áreas incluindo a biomédica. Pesquisas sobre epilepsia observaram que os sinais de *EEG*, já largamente usados em diagnósticos, possuíam componentes ainda não aproveitadas em frequências maiores que 80 Hz, as oscilações em altas frequências (HFO), que poderiam ser usadas como biomarcadores para a localização de áreas epileptogênicas [Zijlmans et al. 2012]

As *HFO* podem ser divididas em faixas, as oscilações entre 80-250 Hz, as oscilações rápidas entre 250-600 Hz e, mais recentemente, as oscilações muito rápidas acima de 1000 Hz, requerendo que os sinais de *EEG* sejam normalmente adquiridos em frequências de amostragem de 500 Hz a 2000 Hz.

A relação entre as *HFO* e os surtos de epilepsia se dá pois, a cada momento individualizado em que ocorre esse tipo de oscilação, existe a ação de grupos de células patologicamente interligadas. Em portadores de epilepsia, o tecido epiléptico possui modificações em sua estrutura que ocasionam uma resposta atípica dos neurônios ou sua

ativação espontânea. O disparo de um grupo de células neurais conectadas pode provocar uma resposta em outro grupo de neurônios, ocasionando um outro disparo de potencial de ação síncrona, sendo todo esse processo ilustrado numa gravação de *EEG* como uma *HFO*.

No quesito de visualização do sinal, normalmente é aplicado um filtro passa-altas e em seguida amplificada sua amplitude, mas é preciso atenção pois o uso de filtragem pode produzir falsas oscilações. Por este motivo, outro método utilizado é a construção de um espectro de potência/frequência usando a transformada de Fourier ou a transformada Wavelet[Zijlmans et al. 2012].

3. A Transformada de StockWell e a SVD

3.1. Transformada S – Análise tempo-frequência

A Transformada de StockWell se caracteriza como uma extensão das ideias da transformada wavelet contínua (CWT) e se baseia no deslocamento de uma janela gaussiana escalável para promover a análise localizada de sinais não-estacionários a um custo computacional baixo. A Transformada S de um sinal **x(t)** dada pela Equação 1.

$$S(\tau, f) = \int_{-\infty}^{\infty} x(t) \frac{|f|}{\sqrt{2\pi}} e^{-\frac{(\tau-t)^2 f^2}{2}} e^{-j2\pi ft} dt \tag{1}$$

A transformada S resulta em um espectrograma, ou seja, um mapeamento tempo-frequência do sinal, no qual a resolução depende da frequência, mantendo uma relação direta com o espectro de Fourier. O cálculo numérico desse espectrograma resulta em uma matriz cujas linhas identificam as posições de frequências e as colunas, os instantes de amostragem. Essa matriz carrega importantes informações como amplitude, atraso e frequência do sinal[Wang 2010].

Como a ocorrência de surtos em sinais de *EEG* é relacionada à epilepsia quando há oscilações de frequências superiores a 80Hz, portanto apenas as linhas dessa matriz correspondentes às frequências maiores que 80Hz devem ser levadas em consideração na presente análise [Zijlmans et al. 2012].

3.2. SVD – Extração de características

A Decomposição em Valores Singulares permite diagonalizar uma matriz retangular $\mathbf{A}_{m\times n}$, decompondo-a em valores e vetores singulares, conforme indicado pela Equação 2:

$$\mathbf{A} = \mathbf{USV^T} \tag{2}$$

Em que a matriz **S** é uma matriz diagonal $m \times n$ cujos elementos de sua diagonal correspondem aos valores singulares σ_i da matriz $\mathbf{A}_{m\times n}$, as colunas $\mathbf{u}_i$ da matriz ortogonal $\mathbf{U}_{m\times m}$ são os vetores singulares à esquerda e as colunas $\mathbf{v}_i$ da matriz ortogonal $\mathbf{V}_{n\times n}$ são os vetores singulares à direita, onde $i = 1, 2, \ldots, min(m, n)$ [Haykin 1989], logo:

$$\mathbf{A} = \sum_{i=1}^{min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \tag{3}$$

Valores singulares mostram as características da matriz e refletem a energia contida em cada subespaço gerado pelo produto externo $\mathbf{u}_i\mathbf{v}_i^T$. O subespaço que contém mais energia e, portanto, o maior acúmulo de características do sinal, é aquele associado ao maior valor singular.

4. Metodologia

Como os sinais disponíveis na base de dados do MIT foram amostrados à taxa de 256Hz, suas informações permitem identificar as oscilações de interesse deste trabalho, com frequências logo acima de 80 Hz. Para a análise foi empregada uma janela de tempo com duração de 100 segundos, totalizando 25600 amostras. O tempo total de uma coleta para cada indivíduo disponibilizada na base de dados é de uma hora.

Assim, valores do espectrograma produzidos pela Transformada S são armazenados em uma matriz tempo-frequência em que cada coluna representa um instante de amostragem entre 0 e 100 segundos e as linhas representam o eixo de frequências entre 0 e 130 Hz, indicado pela Figura 1. Como a faixa de interesse para diagnóstico corresponde às frequências acima de 80Hz, basta utilizar apenas as linhas 80 a 130, resultando em uma matriz de dimensão 50x25600.

Figura 1. Subdivisão temporal e em frequência do Espectrograma gerado pela Transformada S

Essa matriz é, então, dividida temporalmente em 100 matrizes de dimensão 50x256, ou seja, uma submatriz por segundo de amostragem. A Figura 2 apresenta o espectrograma de um trecho de 100 segundos do sinal de *EEG* contendo surtos de epilepsia.

As linhas escuras contínuas sobre o espectrograma indicam as posições de seus picos nas frequências que identificam os surtos. Para fins de correlação, inclui-se na Figura 2 o traçado do trecho correspondente do sinal de EEG, no qual as linhas escuras verticais assinalam os instantes de detecção dos surtos epilépticos.

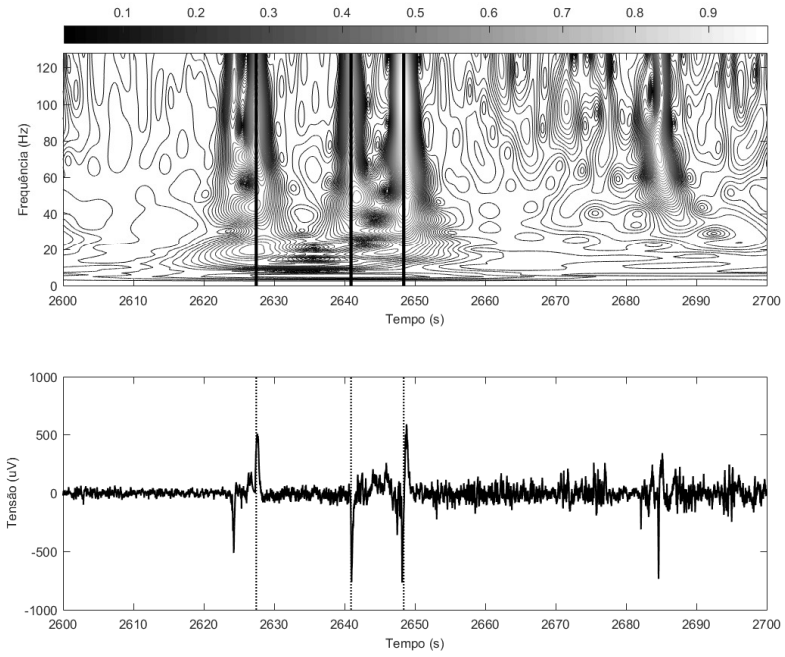


Figura 2. Módulo da Transformada S - Espectrograma de um sinal de EEG com Surtos

Como a *SVD* retorna valores relacionados às características de energia da matriz, pode-se inferir que distúrbios, neste caso surtos, interferem nas amplitudes dos valores singulares. Essa diferença entre amplitudes de sinais com e sem surtos pode ser observada na Figura 3. Como o máximo valor singular é o que representa a maior energia da matriz, ele é o mais sensível à presença de perturbações. No estudo para se detectar surtos, por tanto, utiliza-se o máximo valor singular de cada submatriz temporal.

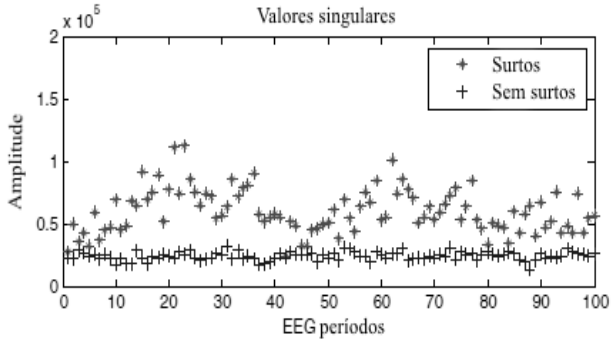


Figura 3. Valores singulares de EEG com e sem surtos

Observa-se que a amplitude do maior valor singular se mostra consideravelmente maior na presença de surtos, causando picos em seu espectro, portanto pode-se relacionar tais picos e suas respectivas larguras com a ocorrência ou não do surto[Xia et al. 2015].

5. Resultados

Todos os procedimentos computacionais foram desenvolvidos utilizando a linguagem computacional matricial matemática *MATLAB*©. A aplicação do método proposto para

uma janela de um sinal de *EEG* livre de surtos com valores singulares normalizados entre 0 e 1 é apresentado na Figura 4.

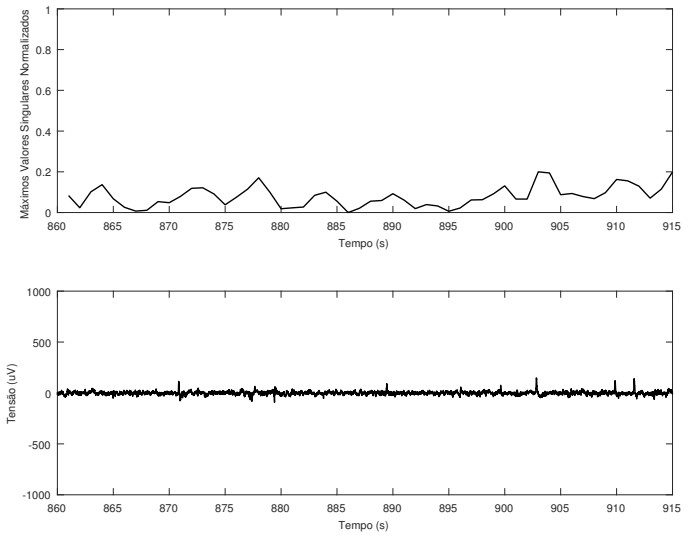


Figura 4. Valores singulares normalizados de um EEG sem surtos

Como não ocorreu nenhum pico considerável, não ocorrendo nenhuma marcação no espectro de máximos valores singulares, infere-se que o sinal de *EEG* não possui surto. Em um caso com surtos diagnosticados, os efeitos do método se mostram mais representativos, mostrados pela Figura 5.

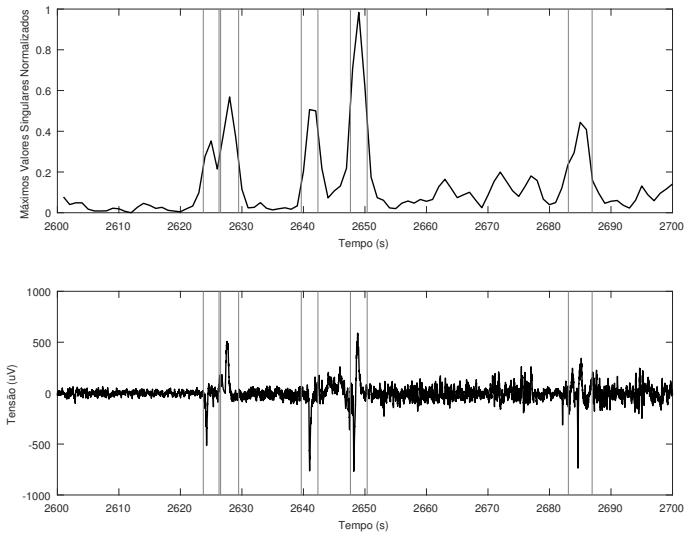


Figura 5. Valores Singulares normalizados de um EEG com surtos

Neste caso, é visível a ocorrência de surtos no sinal temporal, sendo determinados pelos picos no espectro de máximos valores singulares normalizados. A marcação auxilia na rápida visualização e interpretação do sinal estudado.

6. Conclusão

Com a aplicação da Transformada S juntamente da *SVD* constata-se a facilidade de identificação de surtos epilépticos por meio das *HFO*, e também a possibilidade de melhoria do método. O desenvolvimento de diagnósticos desse tipo auxilia o conhecimento rápido de características necessárias para a ajuda e continuidade de novos estudos na área da Epilepsia. Observa-se que o espaço para novas pesquisas ainda é vasto e importante, podendo beneficiar um número representativo de pessoas.

Referências

Haykin, S. (1989). *Modern filters*. Macmillan Press.

Shoeb, A. (2009). Application of machine learning to epileptic seizure onset detection and treatment. *Massachusetts Institute of Technology*.

Stockwell, R. G., Mansinha, L., and Lowe, R. P. (1996). Localization of the complex spectrum: the s transform. *IEEE Transactions on Signal Processing*, 44(4):998–1001.

Wang, Y.-H. (2010). The tutorial: S transform. *National Taiwan University*.

Xia, Y. et al. (2015). Seizure detection approach using s-transform and singular value decomposition. *Epilepsy Behavior*, 52:187–193.

Zijlmans, M. et al. (2012). High-frequency oscillations as a new biomarker in epilepsy. *Annals of Neurology*, 71:169–178.

Using Sentiment Analysis to Real-time Forecast Box Office Revenue

Rodrigo Sasse David[1], Deborah Silva Alves Fernandes[2], Márcio Giovane Cunha Fernandes[3]

[1]Escola de Engenharia Elétrica, Mecânica e de Computação
Universidade Federal de Goiás (UFG)
74.605-010 – Goiânia – GO – Brasil

sasserodrigo1@gmail.com

[2]Instituto de Informática
Universidade Federal de Goiás (UFG)
74690-900 – Goiânia – GO – Brasil

deborah@inf.ufg.br

[3]Instituto de Informática
Universidade Estadual de Goiás (UEG)
75132-903 – Goiânia – GO – Brasil

marcio.giovane@ueg.br

Abstract. *In a widely connected world, people are sharing their opinions and thoughts like never before. Social media plays an important role in the communication process nowadays, companies can not ignore the massive volume of its user-generated content. This paper presents the challenges and uses of sentiment analysis in the modern age. Our study uses sentiment analysis on the buzz generated after a trailer is released on social media in order to predict the box office revenue of the top 10 movies released in the first semester of 2017.*

1. Introduction

Before the Web 2.0[1], whenever someone needed an opinion regarding a product or service, for example, it probably meant reaching out to family or friends. For a company, getting to know their clients' opinions, probably meant to conduct a survey. Presently, both consumers and companies have a cheaper and easier to reach alternative: the enormous amount of opinions shared in social media.

Over the last two decades, the propagation of broadband Internet, and most recently, the ubiquity of the smartphones led to an ever increasing social networking. Due to its ease-of-use and reach, online social media became an important tool on the day-to-day communication process, it has been establishing trends and topics ranging from the environment and politics to technology and entertainment [Liu 2010].

Recently, people have been constantly and voluntarily sharing their opinions, feelings, and desires on blogs, forums and social media. Not only these information are readily available, they can also be more trustworthy than information provided by the sellers

[1]http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html

or vendors [Bickart and Schindler 2001]. For companies, this information is a resource to be harnessed just like the traditional word-of-mouth [Schouten and Frasincar 2016].

Over the last decade, social media content has been successfully used to forecast a wide range of outcomes: stock market volatility [Arnold and Vrugt 2008, Bollen et al. 2011, Arias et al. 2013], public opinion on elections [O'Connor et al. 2010, Zhou et al. 2013], movie box-office revenues [Asur and Huberman 2010, Arias et al. 2013, Liu et al. 2016], TV viewing rates [Wakamiya et al. 2011, Scharl et al. 2016], and tracking epidemics [Lampos et al. 2010].

In 2010, [Asur and Huberman 2010] tried to test the hypothesis that "movies that are well talked about will be well-watched". They collected a total of 2.89 million tweets for 24 movies they selected for the study. From that data set they created a time series and used linear regression to correlate the number of tweets, obtaining a p-value of $9.14e^{-14}$ as their best result. In 2013, [Arias et al. 2013] did a similar study. They collected more than 100 million tweets for 121 different movies. They used linear regression and SVM for a p-value of $2.94e^{-7}$.

In 2016, [Liu et al. 2016] also tried to forecast box office revenue.They performed three different analysis on their data set - purchase intention, attention and popularity, and sentiment analysis - before employing linear regression and support vector regression on it for an adjusted R^2 of 0.94.

The goal of this paper is to use sentiment analysis on social media content in order to predict the box-office revenue of the top 10 grossing movies released in the first semester of 2017 in a matter of hours with a small data set, close a real-time analysis. The paper is organized as follow: in Section 2 we briefly discuss Online Social Networks. Section 3 describes Sentiment Analysis, its approaches and challenges. Section 4 discuss the conducted experiment. We discuss and compare the acquired results in Section 6.

2. Online Social Networks

An Online Social Network (OSN), also known as Online Social Service or Social Media, is an internet-based and user-generated content-based platform which people use to build social relations with other people who share similar personal or professional interests [Obar and Wildman 2015].

OSNs are reaching more and more people. In 2010, 970 million people made use of an OSN worldwide, in 2017 the number jumped to 2.46 billion people[2]. The same study suggests that by 2021 the number of people using OSNs will exceed 3 billion. Just in the United States, in 2010, 48% of the population had an account on at least one OSN; in 2017, this number is 81%[2].

Among all the OSNs the microblogging ones are specially interesting to sentiment analysis. These platforms limit the length of the message to a rather short number, which forces their users to be more concise [Thakkar and Patel 2015]. This trait makes them to be known as "what's happening now" networks [Zhou et al. 2013]. Twitter, one of the most famous and used microblogging network nowadays, limits its messages, called tweets, to 140 characters. Founded in 2006, it has rapidly grown, reaching enormous

[2]https://www.statista.com/statistics/273476/percentage-of-us-population-with-a-social-network-profile/

popularity worldwide. In 2012, more than 100 million users posted an average of 340 million tweets daily[3]. In the first quarter of 2017, 328 million users posted an average of 500 million tweets daily[3].

3. Sentiment Analysis

Sentiment analysis (also known as Opinion Mining) is a computational science branch that works on the intersection of other three branches: information retrieval, natural language processing, and artificial intelligence [Schouten and Frasincar 2016]. Together, they give the tools for SA to rise to its challenges: identification, classification, and agglutination [Schouten and Frasincar 2016].

Given a sentence, sentiment analysis should identify: which part of the sentence carries opinion, who wrote it, its subject, and the opinion itself [King et al. 2009]. Note the following example:

> *"(1) I bought a MacBook Pro a couple of days ago. (2) It's a nice notebook. (3) The screen is gorgeous. (4) The battery life is great. (5) Although the performance is not the best, that's ok to me. (6) However, my dad was mad at me as I did not tell him before I bought it. (7) He thought the notebook was too expensive, and wanted me to return it."*

Based on the example above, sentiment analysis must be able to identify that sentences (2), (3), and (4) express positive opinions, while sentences (5), (6), and (7), negative opinions. It also should identify that sentences (1) and (2) reference the notebook as a whole, while sentences (3), (5), (6), and (7), some of its attributes - screen, battery life, performance, and price. At last, it should identify that sentences (2), (3), (4), and (5) express the opinions of the author of the text, while sentences (6) and (7), the opinions of his father.

Generally, opinion can be expressed on anything - products, services, people, organizations, events, topics etc. The term "object" is commonly used to refer to the entity target of the opinion. The object may have a set of components and a set of attributes. The union of these two sets is the set of features [Liu 2010]. On the example on figure 1, "MacBook" is the object; "screen" and "battery life" are components, and "performance" and "price", attributes.

From a given document d, a set of objects O = $\{o_1, o_2, ..., o_p\}$, which has a set of features F = $\{f_{i1}, f_{i2}, ..., f_{iq}\}$, and a set of authors H = $\{h_1, h_2, ..., h_r\}$, the opinions in d can be represented by the quintuple (o_i, f_{ij}, p_{ijkl}, h_i, t_l) [Liu 2010], on which o_i represents the object; f_{ij}, the set of features; p_{ijkl}, the polarity of the opinion, h_i, the opinion holder; and t_l, the moment the opinion was shared.

Since most websites display information on the author of the text or sentence, as well as its timestamp, by analyzing texts referencing a single object and feature at a time, the quintuple mentioned above has only the polarity of the opinion p_{ijkl} left undiscovered.

3.1. Sentiment Classification and Subjectivity

Textual information can be broadly classified in two main categories: facts and opinions [Liu 2010]. While facts are objective expressions about something, opinions are usually

[3]https://blog.twitter.com/official/en_us/a/2012/twitter-turns-six.html

subjective expressions about people's sentiments or feelings [Liu 2010]. Sentiment analysis is only interested in textual information containing opinions. The task of identifying those sentences is called subjectivity classification [Liu 2010].

Two analysis are possible: document-level or sentence-level [Liu 2010]. The first takes the whole text or document into account, and tries to estimate the polarity of the opinion carried by it; the second, sentence-level, tries to estimate the polarity of the opinion carried by each sentence individually [Liu 2010].

The problem of classification, regardless of the length of text, can be tackled by both supervised learning and unsupervised learning methods [Liu 2010]. Sentiment classification can be easily formulated as a supervised learning problem since there will be only labels, positive and negative. Not only that, review-like sentences usually have an assigned rating, like 1 to 5 stars or 0 to 100 score, which makes training and test data readily available. In unsupervised learning, opinion words and phrases are the dominating indicators for classification [Turney 2002]. In sentiment analysis, these methods are usually referred as lexicon-based analysis. They use a dictionary, called a lexicon of words, in which nouns and adjectives are classified in positive or negative. If a sentence matches a word in the lexicon, its score is incremented if the word is classified as positive, and vice versa. If the final score is positive the sentence as a whole is labeled positive.

4. Experiment

4.1. Dataset

The experiment in this paper tries to predict the opening box office revenue of the top 10 movies released in the first semester of 2017[4] (see Table 1) from the buzz generated on Twitter during the next seven days after the last trailer for each movie was released.

Movie	Hashtag	Box Office Revenue
Beauty and the Beast	#BeautyAndTheBeast	$174,750,616.00
Guardians of the Galaxy, Vol. 2	#GotGVol2	$146,510,104.00
Spider-Man: Homecoming	#SpiderManHomecoming	$117,027,503.00
Wonder Woman	#WonderWoman	$103,251,471.00
The Fate of the Furious	#FateOfTheFurious	$98,786,705.00
Despicable Me 3	#DespicableMe3	$72,434,025.00
Dead Men Tell No Tales	#DeadMenTellNoTales	$62,983,253.00
Kong: Skull Island	#KongSkullIsland	$61,025,472.00
Cars 3	#Cars3	$53,688,680.00
The LEGO Batman Movie	#LEGOBatmanMovie	$53,003,468.00

Table 1. Movies selected for the studies with their respective hashtags and box office revenue.

Twitter's API does not retrieve information older than two weeks. In order to collect data as older as January 2017, we used Twitter's Advanced search, which allows anyone to query for a search using filters like keywords, hashtags, language, period etc.

[4]http://www.boxofficemojo.com/yearly/chart/?yr=2017

[5]https://github.com/Jefferson-Henrique/GetOldTweets-python

For this study, we used a third-party open-source Python code[5] developed to implement this search and return JSON objects.

Each query was composed by keywords, initial date and final date. For keywords we used the official hashtag (see Table 1) for each movie and the movie title. For initial date, the day the trailer was released; and for final date, six days after that. To make the search more broader, we removed special characters such as hyphens, dots and colons from the keywords (e.g.: "Spider-Man: Homecoming" becomes "spiderman homecoming"). Lastly, we used Python's langdetect library to make sure we have only collected tweets written in English.

Each tweet was stored in the database containing an unique id, its message, its author, and the movie it makes reference to.

4.2. Tools for Sentiment Analysis

For sentiment analysis, we used Python's NLTK library. This is a handful toolkit for sentiment analysis and for some other problems of natural language processing since it implements machine learning techniques and routinely used tools and methods like tokenizer (to extract the tokens from a sentence), stemming (to extract the stem from a word), lemmatizer (to extract the lemma from a word) etc.

The comparative study [Asur and Huberman 2010] constructed a sentiment analysis classifier using the LingPipe[6] linguistic analysis package. They used Amazon Mechanical Turk[7] to obtain labeled training data for their classifier. Their results are shown on Table 2.

For this study, we used an online pratform[6] that implements NLTK for sentiment analysis using text classification to determine a sentence's polarity and already has multiple trained classifiers. Through its API, we can feed one of its endpoint with a sentence and it will return a JSON object containing the probability for positive, neutral or negative, and a label for the polarity with the higher probability.

5. Analysis and Comparison

A study from 2010 [Asur and Huberman 2010] used the buzz generated on Twitter to predict the opening box office revenue of 24 movies released that year. Its hypothesis was that "movies that are well talked about will be well-watched". This studied collected 2.89 million tweets over the week before the movie opening, then correlated the number of tweets for each movie to its box office through linear regression. The same study used sentiment analysis for the same prediction, but instead of the number of tweets, they used a ratio between the number of positive to negative tweets (the study only presented the p-value for the sentiment analysis part). Its results are shown on Table 2.

Type of analysis	R^2	p-value
Number of tweets	0.80	3.65e-9
Sentiment analysis	-	9.43e-6

Table 2. Results acquired by [Asur and Huberman 2010].

[6]http://www.alias-i.com/lingpipe

[7]http://www.mturk.com

[8]http://text-processing.com/docs/sentiment.html

In this study we also performed a linear regression between the tweet rate and the box office revenue, and ratio between the number of positive to negative tweets, but with a set of 50,000 tweets, collected over a period of two hours on September 4, 2017; that's 0.017% and 1.191%, respectively compared to [Asur and Huberman 2010] (0.041% and 2.858%, respectively, when adjusted to the number of movies).

Type of analysis	R^2	p-value
Number of tweets	0.744	0.0015
Sentiment analysis	0.522	0.0011

Table 3. Results acquired by this study.

In terms of number os tweets, we acquired results close to the ones in [Asur and Huberman 2010]. In terms of sentiment analysis, we managed to produce satisfactory results with a fraction of data and time when compared to [Asur and Huberman 2010]. This was actually the goal of this study, to demonstrate that sentiment analysis can achieve good results with less expanse. While the comparative study required a whole week to collect its data, it was possible for us to get a glimpse of the behaviour in a couple of hours. This shows how sentiment analysis can rise to the challenges of real-time applications in many areas.

6. Conclusion and Future Work

We had two overcome a couple of challenges in this study. The first limitation we encountered was the inability to collect tweets older than two weeks using Twitter's API. An alternative to that is to use Twitter's Streaming API, which can be used to collect tweets in real-time as they are posted in the platform.

The second one was regarding the sentiment analysis, the free service we used only allows a small number of requests per day, which forced us to only process a small percentage of our data set. Because of that, methods that only require a small amount of data are preferable.

For future work we are looking for to improve our model and yield better results with the addition of some techniques to the analysis such: frequency of terms, part of speech tags, syntactic dependency and negation [Liu 2010].

References

Arias, M., Arratia, A., and Xuriguera, R. (2013). Forecasting with twitter data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(1):8.

Arnold, I. J. and Vrugt, E. B. (2008). Fundamental uncertainty and stock market volatility. *Applied Financial Economics*, 18(17):1425–1440.

Asur, S. and Huberman, B. A. (2010). Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 492–499. IEEE.

Bickart, B. and Schindler, R. M. (2001). Internet forums as influential sources of consumer information. *Journal of interactive marketing*, 15(3):31–40.

Bollen, J., Mao, H., and Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1):1–8.

King, I., Li, J., and Chan, K. T. (2009). A brief survey of computational approaches in social computing. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 1625–1632. IEEE.

Lampos, V., De Bie, T., and Cristianini, N. (2010). Flu detector-tracking epidemics on twitter. *Machine Learning and Knowledge Discovery in Databases*, pages 599–602.

Liu, B. (2010). Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666.

Liu, T., Ding, X., Chen, Y., Chen, H., and Guo, M. (2016). Predicting movie box-office revenues by exploiting large-scale social media content. *Multimedia Tools and Applications*, 75(3):1509–1528.

Obar, J. A. and Wildman, S. S. (2015). Social media definition and the governance challenge: An introduction to the special issue.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., and Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM*, 11(122-129):1–2.

Scharl, A., Hubmann-Haidvogel, A., Jones, A., Fischl, D., Kamolov, R., Weichselbraun, A., and Rafelsberger, W. (2016). Analyzing the public discourse on works of fiction–detection and visualization of emotion in online coverage about hbo's game of thrones. *Information processing & management*, 52(1):129–138.

Schouten, K. and Frasincar, F. (2016). Survey on aspect-level sentiment analysis. *IEEE Transactions on Knowledge and Data Engineering*, 28(3):813–830.

Thakkar, H. and Patel, D. (2015). Approaches for sentiment analysis on twitter: A state-of-art study. *arXiv preprint arXiv:1512.01043*.

Turney, P. D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th annual meeting on association for computational linguistics*, pages 417–424. Association for Computational Linguistics.

Wakamiya, S., Lee, R., and Sumiya, K. (2011). Crowd-powered tv viewing rates: measuring relevancy between tweets and tv programs. *Database Systems for Adanced Applications*, pages 390–401.

Zhou, X., Tao, X., Yong, J., and Yang, Z. (2013). Sentiment analysis on tweets for social events. In *Computer Supported Cooperative Work in Design (CSCWD), 2013 IEEE 17th International Conference on*, pages 557–562. IEEE.

Requisitos Essenciais de Usuários de um Sistema de Informação em Saúde Bucal para a Atenção Básica

Fábio Nogueira de Lucena[1], Beatriz Nogueira Carvalho da Silveira[1], Renata Dutra Braga[1], Rejane Faria Ribeiro-Rotta[2]

[1]Instituto de Informática, [2]Faculdade de Odontologia
Universidade Federal de Goiás (UFG)
Alameda Palmeiras, Quadra D, Câmpus Samambaia– 74.690-900 – Goiânia – GO – Brazil

fabio@inf.ufg.br, beatrizncsilveira@gmail.com, renata@inf.ufg.br, rejanefrr@gmail.com

Abstract. *The use of Information and Communication Technology (ICT) in health is called e-Health. There is an expectation that e-Health will continue to grow as an instrument capable of improving the quality of health in Brazil, according to actions of the Ministry of Health. In the current context where e-Health is not present in many municipalities, it becomes relevant to guide possible software acquisitions. In this sense, the present work presents the result of a compilation of user requirements for a Oral Health Information System considered "essential". The requirements and the method used provide a direction for the adoption of e-Health and, in particular, for implementation through the identification of the subdomains adherent to a development based on microservices.*

Resumo. *O uso da Tecnologia da Informação e Comunicação (TIC) na saúde é denominado de e-Saúde. Há uma expectativa de que a e-Saúde continue crescendo como um instrumento capaz de melhorar a qualidade da saúde no Brasil, conforme ações do Ministério da Saúde. No contexto atual onde a e-Saúde não está presente em muitos municípios, torna-se relevante orientar possíveis aquisições de software. Nesse sentido, o presente trabalho apresenta o resultado de uma compilação "essencial" de requisitos de usuário para um Sistema de Informação em Saúde (SIS) bucal. Os requisitos e o método utilizado fornecem um direcionamento para a adoção da e-Saúde e, em particular, para implementação por meio da identificação dos subdomínios aderentes a um desenvolvimento baseado em microsserviços.*

1. Introdução

O Brasil possui várias ações em andamento para fomentar a e-Saúde com o propósito de estender o acesso e melhorar a qualidade da atenção básica. A Política Nacional de Informação e Informática em Saúde (PNIIS) [BRASIL 2016] e a estratégia de e-Saúde para o Brasil [BRASIL 2017], a ser implementada até 2020, são algumas das ações. A adoção de e-Saúde, contudo, é um processo complexo no qual a caracterização de um novo sistema ou adaptação de um sistema é um desafio. Sistema aqui é uma coleção de partes que cooperam entre si para a obtenção de um determinado objetivo e do qual o software, em geral, é um de seus componentes.

Quando um sistema inclui software, situação frequente e estimulada nos dias atuais no escopo da saúde [BRASIL 2017], o software em questão precisa ser desenvolvido ou adquirido. O desenvolvimento de software é uma alternativa quando se possui recursos para tal e há especificidades em abundância para serem atendidas no contexto em que o software será utilizado. Quando se observa a situação corrente da saúde nos municípios brasileiros, o

sentimento é de que tais recursos não estão disponíveis. Contudo, independente da disponibilidade ou não destes recursos, uma alternativa para o desenvolvimento de software, atividade geralmente onerosa, é a aquisição de software.

A aquisição de software demanda a execução de um processo de aquisição [SOFTEX 2013]. Se tal processo não é executado de forma adequada, então os riscos de fracasso são maiores. Visando contribuir com o processo de aquisição de software e, dessa forma, minimizar tais riscos, o presente texto relata o esforço e os resultados obtidos na identificação de um conjunto de requisitos de usuário para um Sistema de Informação em Saúde (SIS) bucal no escopo da atenção básica.

A expectativa é de que o conjunto identificado seja "essencial", isto é, contenha os requisitos mínimos para adequadamente atender as necessidades em saúde bucal no escopo da atenção básica. Entende-se requisito de usuário como um "objetivo ou tarefa que uma classe específica de usuários deve ser capaz de realizar com o software, ou um atributo desejado para o software" [Wiegers 2013].

O uso esperado desse conjunto essencial de requisitos de usuário produzido não é oferecer um mecanismo para avaliar se um SIS em saúde bucal é "bom" ou "ruim". O objetivo é facilitar a verificação que indique ou não se as necessidades consideradas indispensáveis são satisfeitas e, consequentemente, contribuir com o processo de aquisição de software. Naturalmente, tal conjunto também é útil ao desenvolvimento de um novo SIS.

Convém esclarecer que a noção de indispensável ou essencial não é necessariamente consensual, o que não contribui com a identificação de um único conjunto. Contudo, é razoável a busca por um conjunto de requisitos de usuário que sirva de referência para esta verificação.

Requisitos de usuário são relevantes porque são escritos na perspectiva dos usuários, o que facilita a participação dos profissionais de saúde tanto no desenvolvimento quanto na aquisição de um SIS. Tais requisitos, contudo, são insuficientes para caracterizar um SIS em sua totalidade. Além dos requisitos de usuário, há também os requisitos de qualidade, que incluem, por exemplo, interoperabilidade, escalabilidade, segurança e outros [ISO 25010 2011]. Embora não possam ser ignorados em um processo de desenvolvimento ou de aquisição, os requisitos de qualidade não foram contemplados no conjunto "essencial" obtido por dois motivos: (a) recebem influência direta do contexto de uso do SIS (por exemplo, quantidade de usuários e frequência de uso, dentre outros); e (b) são contemplados por outros artefatos como o Manual de Certificação de SISs produzido pela SBIS/CFM [SBIS 2016].

A Seção 2 apresenta a Metodologia empregada no trabalho. A Seção 3 os Resultados e a Seção 4 fornece as Considerações finais e trabalhos futuros.

2. Metodologia

As subseções seguintes apresentam as atividades realizadas e os resultados produzidos por cada uma delas. Em boa parte a execução ao longo do tempo seguiu a ordem abaixo.

2.1. Levantamento bibliográfico

Um levantamento bibliográfico foi realizado, mas não resultou na identificação de trabalhos específicos para o presente escopo, apesar da existência de trabalhos que fornecem insumos relevantes como o que se concentra nas informações necessárias para um atendimento adequado em saúde bucal [Song et al 2010], ou o que apresenta normas para a prescrição

produzida por odontólogos [Castro 2009]. A conclusão, contudo, é que uma revisão sistemática é indicada.

2.2. Análise de sistemas correlatos

Uma fonte valiosa para a identificação de requisitos de usuário são os SISs em saúde bucal disponíveis no mercado, dentre os quais 11 (onze) foram localizados e analisados. A escolha baseou-se naqueles com presença na internet e que, ao mesmo tempo, possuíam documentação suficiente e disponível para a análise. Em particular, nenhum possuía a certificação da SBIS/CFM[1].

A maior contribuição desta análise foi a "sugestão" de que requisitos de usuário, nesse domínio, estão geralmente agrupados, apesar das divergências de terminologia encontradas na documentação. Esta "sugestão" foi posteriormente o principal indutor da definição de contextos (*bounded-contexts*), comentado em seção posterior.

2.3. Observação

Etnografia é uma técnica de observação na qual um analista faz uma imersão no ambiente de trabalho em que o sistema será usado e, através disso, compreende os processos operacionais deste ambiente, possibilitando extrair requisitos de apoio para esses processos [Sommerville 2000].

Seguindo essa técnica foram realizados encontros de observação na Faculdade de Odontologia (FO) da Universidade Federal de Goiás (UFG). As ações dos usuários dos serviços de saúde oferecidos pela FO/UFG e dos profissionais de saúde que os assistiam foram registradas. Convém esclarecer que não foram feitas observações sobre atendimentos clínicos, mas de processos de ordem administrativa. O principal objetivo desse esforço foi a ambientação com o contexto de operação de um SIS em saúde bucal no escopo da atenção básica. Naturalmente, esse esforço também permitiu compreender funções indispensáveis para que os serviços sejam oferecidos com qualidade.

2.4. Análise dos requisitos

As ações anteriores, Levantamento bibliográfico, Análise de sistemas correlatos e Observação, produziram dezenas de requisitos de usuários que foram compilados, filtrados e depois classificados.

A compilação foi realizada por pessoal técnico em computação, ou seja, os mesmos que executaram as ações anteriores.

A filtragem foi realizada por profissional do domínio da saúde bucal e permitiu que requisitos "não essenciais" fossem eliminados, em particular, pela aplicação restrita a cenários específicos e, portanto, não considerados gerais o suficiente para uma classificação como "indispensável".

Os requisitos de usuários que restaram após a filtragem foram classificados conforme a "semelhança de propósito", o que deu origem a subdomínios, ou seja, a contextos (bounded-context) [Evans 2003]. Tais contextos também foram inspirados pelo resultado da Análise de sistemas correlatos. O resultado é exibido na Figura 1 abaixo por meio de um mapa de contexto [Evans 2003], no qual os contextos e os relacionamentos entre eles são

[1] SBIS. Lista de Sistemas Certificados da SBIS CFM. Sociedade Brasileira de Informática em Saúde (SBIS). 2016. Disponível em: http://www.sbis.org.br/lista-de-sistemas-certificados

apresentados. Por simplicidade, os relacionamentos são parcialmente apresentados tendo em vista o escopo do presente trabalho.

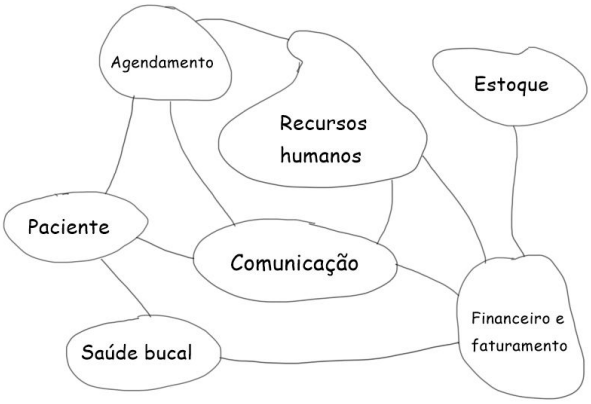


Figura 1. Mapa de contexto exibindo principais contextos e relacionamentos parciais entre eles.

Embora esteja além do escopo do presente trabalho, a "Comunicação" revelou-se fundamental tendo em vista o apoio que ela ofereceu aos demais contextos além, naturalmente, do principal contexto, Saúde bucal, responsável pelas informações clínicas em saúde bucal.

3. Resultados

Os requisitos de usuário são apresentados abaixo conforme os contextos nos quais foram agrupados (Figura 1).

A1 Agendamento

- A1.1 Gerenciar agendamento de consultas.
- A1.2 Manter agenda telefônica e de contatos (whatsapp, skype e outros).
- A1.3 Alertar retorno (via e-mail enviado automaticamente para o paciente).
- A1.4 Controlar retornos (por exemplo, quais os retornos esperados para um dia).
- A1.5 Confirmar consulta.
- A1.6 Gerenciar disponibilidade de profissionais de saúde para agendamento.
- A1.7 Gerenciar remanejamento de agendamentos.
- A1.8 Produzir relatórios de agendamentos.
- A1.9 Facilitar acesso às agendas por dia, semana ou mês.
- A1.10 Organizar fila de atendimento.

F2 Financeiro e faturamento

- F2.1 Controlar o caixa interno (receitas e despesas).
- F2.2 Vincular as contas bancárias.
- F2.3 Configurar boleto bancário.
- F2.4 Emitir relatórios de contratos.
- F2.5 Produzir relatórios administrativos.
- F2.6 Produzir métricas e indicadores em Dashboard.
- F2.7 Fazer orçamentos de procedimentos para pacientes.

- F2.8 Controlar as multas.
- F2.9 Emitir NFe.
- F2.10 Possuir integração com cielo e pagseguro.
- F2.11 Emitir boletos e recibos.
- F2.12 Cadastrar convênios (operadora de plano).
- F2.13 Vincular os pacientes aos convênios.
- F2.14 Gerenciar glosa de procedimentos.
- F2.15 Controlar o orçamento estipulado para cada convênio.
- F2.16 Controlar o faturamento de múltiplos convênios.
- F2.17 Seguir o padrão de faturamento TISS.

P3 Paciente

- P3.1 Manter cadastro de paciente vinculado ao Cartão Nacional de Saúde.
- P3.2 Controlar e registrar os atendimentos.
- P3.3 Gerenciar espera de paciente.
- P3.4 Realizar a importação de paciente.
- P3.5 Registrar o plano de indicação.
- P3.6 Gerenciar pendências do paciente.

S4 Saúde bucal

- S4.1 Manter conformidade com CID-10 e CID-O.
- S4.2 Emitir alertas.
- S4.3 Possuir odontograma gráfico.
- S4.4 Apresentar o histórico da situação bucal.
- S4.5 Registrar a prescrição.
- S4.6 Emitir atestado.
- S4.7 Registrar o planejamento.
- S4.8 Produzir anamnese clínica.
- S4.9 Anexar exames (incluindo imagens).
- S4.10 Solicitar exames.
- S4.11 Solicitar encaminhamentos.
- S4.12 Produzir relatório de procedimentos.
- S4.13 Apresentar procedimentos em aberto, realizados e cancelados.
- S4.14 Registrar a prescrição de medicamentos.
- S4.15 Registrar fórmulas.
- S4.16 Solicitar próteses.

E5 Estoque

- E5.1 Produzir inventário de estoque.
- E5.2 Manter cadastro de produtos (inclusive medicamentos a partir de bases como ANVISA).
- E5.3 Manter categorias de produtos.
- E5.4 Registrar informações de fornecedores.
- E5.5 Produzir ordem de compra.
- E5.6 Controlar compras.

- E5.7 Requisitar produtos.
- E5.8 Produzir relatórios quantitativos dos produtos.

R6 Recursos humanos

- R6.1 Cadastrar corpo clínico (inclusive por consultório).
- R6.2 Produzir estatística de produtividade.
- R6.3 Gerenciar o planejamento de férias, licenças e demais benefícios previstos para os profissionais de saúde e demais colaboradores da unidade.
- R6.4 Registrar e controlar a folha de ponto dos funcionários.
- R6.5 Produzir relatório de "todos os dentistas".
- R6.6 Produzir relatório de remunerações.

C7 Comunicação

- C7.1 Enviar e receber SMS, e-mail e notificações, e produzir relatório de enviados e recebidos.
- C7.2 Enviar mala direta.
- C7.3 Realizar pesquisa de satisfação. Dada uma pesquisa cadastrada, para determinada data, no período em questão o sistema disponibiliza página na web onde respostas poderão ser coletadas e, na sequência, os resultados compilados em um arquivo.
- C7.4 Oferecer chat para os colaboradores da clínica possibilitando a comunicação Interna.

4. Considerações finais

Em um cenário de ampliação da e-Saúde no Brasil, torna-se valioso qualquer instrumento que oriente gestores tanto no desenvolvimento quanto na aquisição de SISs. Um destes instrumentos é a definição de requisitos "essenciais" ou "indispensáveis" a serem satisfeitos por um SIS que é objeto de desenvolvimento ou aquisição. No contexto da saúde bucal foi realizado um levantamento bibliográfico que não identificou trabalhos semelhantes àquele aqui apresentado, seguido de uma análise de SISs no escopo da saúde bucal. Esse esforço foi amparado por uma ambientação com o domínio da saúde bucal. O conjunto de requisitos de usuário produzido por tais atividades foi então filtrado e classificado, o que resultou na identificação de sete contextos e seus requisitos de usuário correspondentes. Todo o esforço foi orientado para identificar um conjunto mínimo, essencial, indispensável de requisitos para servir de referência para novos desenvolvimentos ou aquisição de SIS em saúde bucal no escopo da atenção básica. Uma proposta de tal conjunto agora existe.

O principal trabalho futuro é a derivação de uma especificação de requisitos de software a partir dos requisitos de usuário identificados, o que é particularmente útil à validação dos requisitos obtidos, que pode ser estendida como uma revisão sistemática. Por fim, a existência de um conjunto de referência tem valor significativo por viabilizar a comparação entre produtos, o que é relevante em um processo de aquisição de software.

5. Referencias

BRASIL. Ministério da Saúde. Política Nacional de Informação e Informática em Saúde. Disponível em http://bvsms.saude.gov.br/bvs/publicacoes/politica_nacional_infor_informatica_saude_2016.pdf. (2016).

BRASIL. Ministério da Saúde. Estratégia e-Saúde para o Brasil. Disponível em http://portalsaude.saude.gov.br/images/pdf/2017/julho/12/Estrategia-e-saude-para-o-Brasil.pdf. (2017).

Castro, M. L.; et al. Normas para a Prescrição de Medicamentos em Odontologia. Revista Periodontia, vol 19, número 3, (2009).

Evans, E. Domain-Driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley, (2003).

SOFTEX. Melhoria de Processo de Software Brasileiro - Guia de Aquisição. Disponível em http://www.softex.br/wp-content/uploads/2013/07/MPS.BR_Guia_de_Aquisicao_2013..pdf. (2013).

Sommerville, I.; Engenharia de Software. 9. ed. — São Paulo: Pearson Prentice Hall, págs 75-76 (2011).

ISO 25010. ISO/IEC 25010. Systems and software engineering -- Systems and software Quality Requirements and Evaluation (SQuaRE) -- Systems and software quality models. (2011).

SBIS. Manual de Certificação da SBIS CFM v4.2 2016. Sociedade Brasileira de Informática em Saúde (SBIS). 2016. Disponível em http://www.sbis.org.br/certificacao/Manual_Certificacao_SBIS-CFM_2016_v4-2.pdf

Song, M.; et al. How information systems should support the information needs of general dentists in clinical settings: suggestions from a qualitative study. BMC Medical Informatics and Decision Making, vol. 10(7), (2010).
Disponível em http://www.biomedcentral.com/1472-6947/10/7.

Wiegers, K. E. and Beatty, J. Software Requirements. Microsoft Press, (2013).

Derivando Requisitos de Exigências para Certificação de Sistemas de Informação em Saúde

Leonardo Jacarandá Neves[1], Renata Dutra Braga (renata@inf.ufg.br), Rejane Faria Ribeiro-Rotta (rejanefrr@gmail.com), Fábio Nogueira de Lucena (fabio@inf.ufg.br)

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Caixa Postal 131 – 74001-970 – Goiás – GO – Brazil

leonardojacaranda@hotmail.com[1]

Abstract. *The first step in the development of any quality system depends on a well-defined and well-defined Requirements Engineering process, serving as the backbone for the rest of the system building process. A Health Information System (SIS) is a system considered complex due to the great need of the requirements used in its construction and the concern to guarantee the security of the information that is transported there. The Sociedade Brasileira de Informática em Saúde (SBIS) and the Conselho Federal de Medicina (CFM) are responsible for providing a certification process that will guarantee the quality standard for SISs. How much a SIS to be developed must meet the requirements for certification defined by SBIS/CFM, whether due to a business need or commercial orientation, what are the software requirements sufficient to satisfy such need? The present work presents the method and results obtained from an effort to identify a set of requirements that, if satisfied, ensures the certification of SBIS / CFM and the quality of a SIS.*

Resumo. *O primeiro passo para o desenvolvimento de qualquer sistema com qualidade, depende do um processo de Engenharia de Requisitos bem estruturado e bem definido, servindo como o pilar base para todo o restante do processo de construção de um sistema. Um Sistema de Informação em Saúde (SIS) é um sistema considerado complexo pela grande necessidade de requisitos utilizados em sua construção e a preocupação em garantir a segurança das informações que ali são trafegadas. A Sociedade Brasileira de Informática em Saúde (SBIS) e o Conselho Federal de Medicina (CFM) são responsáveis disponibilizar um processo de certificação que garantirá o padrão de qualidade para os SISs. Quanto um SIS a ser desenvolvido deve satisfazer as exigências para certificação definida pela SBIS/CFM, seja por força de uma necessidade de negócio ou orientação comercial, quais são os requisitos de software suficientes para satisfazer tal necessidade? O presente trabalho apresenta o método e os resultados obtidos de um esforço que teve como propósito identificar um conjunto de requisitos que, se satisfeitos, assegura a certificação da SBIS/CFM e a qualidade de um SIS.*

1. Introdução

Em um Sistema de Informação em Saúde (SIS) é imprescindível cuidados acerca da segurança e estrutura das informações, além das funcionalidades oferecidas pelo sistema. No Brasil, a Sociedade Brasileira de Informática em Saúde (SBIS), em parceria com o Conselho Federal de Medicina (CFM), oferece uma certificação de qualidade para um SIS. A certificação estabelece um conjunto de exigências a serem satisfeitas para que um SIS avaliado obtenha a aprovação no processo de auditoria da SBIS. Atualmente cerca de 50 sistemas estão certificados[1]. Para um SIS aprovado em um processo de certificação da SBIS/CFM, há uma garantia que ele foi se qualificou em todas as etapas do processo, demonstrando que o mesmo possui qualidade em suas funcionalidades e principalmente na segurança das informações, que é o foco principal no objetivo do processo de certificação da SBIS/CFM.

As exigências para a certificação estão registradas em dois manuais, o Manual de Certificação da SBIS (MCS) [SBIS 2016a], e o Manual Operacional SBIS (MOS) [SBIS 2016b], que detalha os procedimentos executados por auditores para verificar se as exigências são atendidas ou não. Naturalmente, a auditoria ocorre sobre sistemas existentes, possivelmente já em operação em estabelecimentos de saúde. Ou seja, a certificação exige acesso a um sistema em pleno funcionamento, e as exigências não se aplicam a especificações (modelos) de um sistema que ainda não foi produzido.

Antes que um SIS possa ser construído, contudo, os requisitos correspondentes precisam ser especificados, e tal especificação pode ser produzida à luz das exigências que o futuro sistema deverá atender. Nesse caso, dentre outros interessados (*stakeholders*) estaria a certificação da SBIS/CFM, que contribuiria com exigências a serem atendidas pelo sistema quando concluído. É razoável observar que ignorar as exigências da certificação da SBIS/CFM durante o desenvolvimento dos requisitos de um SIS pode comprometer a obtenção desta certificação.

Convém destacar que, um dos pontos de sucesso para um SIS é a definição de uma especificação de requisitos de qualidade. De fato, o processo de Engenharia de Requisitos pode ser considerado um dos mais importantes de todo o desenvolvimento de um software, pois é de amplo conhecimento que o defeito introduzido e não detectado nessa fase tem impacto significativo nos custos para corrigí-lo, em fase posterior [Sommerville 2011]. No presente trabalho, a falha na obtenção da certificação da SBIS/CFM é tratada como uma falha do desenvolvimento dos requisitos do sistema em questão. Ou seja, uma falha introduzida por não considerar corretamente as necessidades do interessado SBIS/CFM.

Defeitos não são inseridos exclusivamente na fase de desenvolvimento dos requisitos de um software, contudo, para efeito prático e de interesse do presente texto, não são de interesse os defeitos introduzidos por fases posteriores como o projeto (*design*) e a implementação, por exemplo.

A Engenharia de Requisitos inclui o desenvolvimento e o gerenciamento de requisitos [Wiegers 2013]. O desenvolvimento, independente do ciclo de vida definido para o sistema em questão, inclui as atividades de elicitação, análise, especificação e

[1] http://www.sbis.org.br/lista-de-sistemas-certificados

validação. A falha na certificação da SBIS/CFM, portanto, significa que uma ou mais destas atividades introduziram e/ou não detectaram um defeito ou omissão na especificação de requisitos do SIS em questão. O objetivo dessa interpretação é evitar uma adequação eventual e necessária, além de onerosa, no SIS para a obtenção da certificação da SBIS.

Tal interpretação norteou um processo extenso de síntese de requisitos de software induzidos pela necessidade de negócio de atender a certificação da SBIS/CFM, documentada na Seção 2, assim como outras certificações de SISs. A Seção 3 apresenta o método utilizado e a Seção 4 os resultados obtidos, que consideramos relevantes para o desenvolvimento de SISs, inclusive as dificuldades encontradas com os manuais da SBIS/CFM. Finalmente a Seção 5 fornece as conclusões e perspectivas de trabalhos futuros.

2. Certificação de SISs e da SBIS/CFM

No mundo existem várias certificações que definem a qualidade de um SISs como, por exemplo, a *Certified Associate in Healthcare Information & Management Systems* (CAHIMS) e a ONC *Health* IT *Certification Program*, ambas visam oferecer orientação sobre a qualidade dos SISs certificados nos EUA. No Brasil há uma única certificação que é resultado de esforço conjunto da SBIS com o CFM.

O ato de certificação é constituído de vários processos que são subdivididos em quatro etapas, são elas: preparação, inscrição e formalização, auditoria e conclusão. Um SIS que deseja ser certificado deverá obter êxito em todas as etapas, o que garantirá que o mesmo esteja apto à receber o selo de certificação.

A certificação é amparada por dois manuais logo MCS e MOS. O MCS é dividido em vários tópicos, onde esses tópicos informam toda referência teórica em que ele foi embasado, informa o conjunto de normas e condições para certificação, informa como que um SIS deve fazer para adaptar e adequar todos os processos impostos no ato de certificação, e por último, define todos os Requisitos de Conformidade que deverão ser atendidos para determinada categoria do SIS. Já o MOS apresenta o detalhamento de todas as rotinas de teste (“scripts”) que devem ser executadas durante a auditoria de um SIS para a verificação da conformidade aos requisitos estabelecidos no MCS.

O MCS divide a categoria de um SIS em dois tipos, assistencial e básica, que são relacionadas ao contexto real que o SIS atuará, e também há uma categorização relacionado ao ponto de vista de segurança da informação, definidos como categoria Nível de Garantia de Segurança 1 (NGS1) e Nível de Garantia de Segurança 2 (NGS2). Para cada combinação dessas quatro categorizações existirão conjuntos de Requisitos de Conformidades distintos, que deverão ser implementados para que haja sucesso no processo de auditoria.

Os Requisitos de Conformidade são agrupados em: Requisitos de NGS1, Requisitos de NGS2, Requisitos de Estrutura e Conteúdo, Requisitos de Funcionalidades e Requisitos para Gerenciamento Eletrônico de Documentos.

Os grupos dos requisitos são divididos em subseções que geram o total de até 50 subseções, onde cada uma tem como objetivo destrinchar mais ainda informações

específicas. Para cada requisito há um tipo de definição relacionado a sua obrigatoriedade, classificando em: requisitos mandatórios, onde esses devem ser obrigatoriamente atendidos pelo SIS, requisitos recomendados, onde são considerados importante e que possui uma grande probabilidade de se tornarem mandatórios nas versões futuras do manual, porém, não são obrigatórios, e os requisitos que não se aplicam, ou seja, requisitos não aplicável à situação apresentada.

Apesar dos manuais da certificação da SBIS/CFM fazerem referência a Requisitos de Conformidade, aqui estes requisitos são denominados de exigências que, por sua vez, dão origem a requisitos, requisitos de software, definidos pelo presente trabalho.

3. Metodologia

A transformação das exigências da certificação da SBIS/CFM em um conjunto de requisitos de software correspondente pode ser gerada de várias formas. Young [Young 2010] propõe um método para sintetizar requisitos de software a partir de documentos contendo restrições a serem observadas pelo futuro software quando em execução. O método inclui as seguintes atividades:

1. Analisar os documentos com o intuito de produzir um conjunto de elementos que poderão ser analisados independentemente, cada um deles chamado de sentença.
2. Classificar as sentenças geradas em (1).
3. Gerar os requisitos a partir da classificação produzida (2).

O presente trabalho considerou o método proposto por Young [Young 2010] na definição de um método específico para essa transformação, subdividido em três processos: síntese, especificação e verificação. A síntese é o processo que deriva um ou mais requisitos de software a partir da análise de um exigência. A especificação é o processo de documentar os requisitos produzidos pelo processo anterior, o que inclui a identificação única, dentre outras atividades. A verificação é responsável por identificar eventuais defeitos introduzidos nos processos anteriores e, se for o caso, providenciar os ajustes correspondentes. A verificação inclui a formalidade, a qualidade, a corretude e a funcionalidade do requisito gerado.

O método repete os três processos para cada exigência ou conjunto de exigências até que não existam mais exigências a serem contempladas.

Convém ressaltar que a especificação ainda classifica cada requisito gerado conforme a ISO/IEC 25010:2015 [ISO 2015], cujo principal objetivo é orientar o desenvolvimento de produtos de software através da especificação de requisitos de qualidade e as características da qualidade de avaliação. A qualidade de um sistema é o grau em que o sistema satisfaz as necessidades declaradas e implícitas de suas diversas partes interessadas e, portanto, fornece valor. As necessidades dos *stakeholders* são exatamente o que é representado no modelo, que categoriza a qualidade do produto em características. Naturalmente, boa parte dos requisitos são classificados como requisitos

de segurança, tendo em vista que este aspecto é um dos principais componentes da certificação da SBIS/CFM.

4. Resultados

Está além do espaço disponível citar os 527 requisitos gerados a partir das exigências da certificação da SBIS/CFM. De fato, na planilha encontra-se os requisitos produzidos e uma matriz de rastreabilidade que liga os requisitos gerados às exigências das quais foram sintetizados. Na planilha existe um total de 4 páginas: mapeamento, requisitos, dicionário de dados e o glossário. Na página de mapeamento descreve quais requisitos que foram criados que atendem diretamente às exigências do MCS. Na página de requisitos é onde há a descrição de todos os requisitos gerados, o tipo correspondente de cada um conforme a ISO/IEC 25010:2015 [ISO 2015], o código único de cada requisito, e as observações consideradas relevantes para determinado requisito. Na página de dicionário de dados há dados que são utilizados nas descrições dos requisitos que auxiliaram a especificação dos mesmos. Na página de glossário descreve algumas definições que surge no decorrer das especificações.

No MCS há um total de 238 exigências que resultaram nos 521 requisitos obtidos a partir dos processos de: síntese, especificação e verificação. O fato da quantidade de requisitos gerados ser relativamente maior do que a quantidade de exigências é considerado bastante corrente, pois as exigências não são subdivididas em sub-requisitos, o que não é muito adequado, podendo gerar uma ambiguidade nas informações do requisito e um problema de verificação, impossibilitando de deduzir qual funcionalidade de o SIS ele está atendendo. No trabalho desenvolvido, os requisitos possuem uma hierarquia, o que garante que requisitos que subdivididos possuem uma relação direta no papel de suas funcionalidades. A Tabela 1 abaixo mostra como a exigência *NGS1.02.05* do MCS está especificada. A primeira coluna corresponde ao número de identificação da exigência, a segunda coluna o nome correspondente e a terceira coluna a descrição do que a exigência pretende especificar.

Tabela 1 - Modelo MCS [Manual de Certificação da SBIS 2016].

ID	Título	Requisito
NGS1.02.05	Controle de tentativas de login	a) O S-RES deve possuir, em todos os processos de autenticação de usuário, independentemente do método utilizado, mecanismos para bloquear a conta deste usuário no S-RES (ou seja, bloquear seu acesso ao S-RES), após um número máximo configurável de tentativas consecutivas de login com autenticação inválida, que não exceda a 10 tentativas. b) Após o bloqueio de conta de um usuário, o sistema só deve permitir login deste após o desbloqueio de sua conta de usuário.

A partir da metodologia utilizada no trabalho e exigência *NGS1.02.05* passou a ser representado por quatro requisitos, como é ilustrado na Tabela 2.

Tabela 2 - Modelo obtido após o processo de síntese.

Código do requisito	Descrição
[R.09]	O sistema deverá possuir um controle de tentativas de login sem êxito.
[R.09.01]	O sistema bloqueará um usuário caso ocorra 10 tentativas sem êxito para autenticação.
[R.09.02]	O sistema não deve permitir o login pelo usuário cujo a conta foi bloqueada.
[R.09.03]	O usuário poderá redefinir sua senha, a partir de sua senha antiga.

O importante destaque é pela quantidade de requisitos de segurança que foi gerado. Como o principal foco da certificação é garantir a qualidade e a segurança das informações que são trafegadas e armazenadas no ambiente do SIS, é previsível que os requisitos de segurança sejam uma grande parte do conjunto total que foi gerado. Os requisitos funcionais são os que possuem uma maior quantidade, pela grande necessidade de funcionalidades de um SISs. Podemos observar na Figura 1, uma relação entre a quantidade de requisitos para cada categorização.

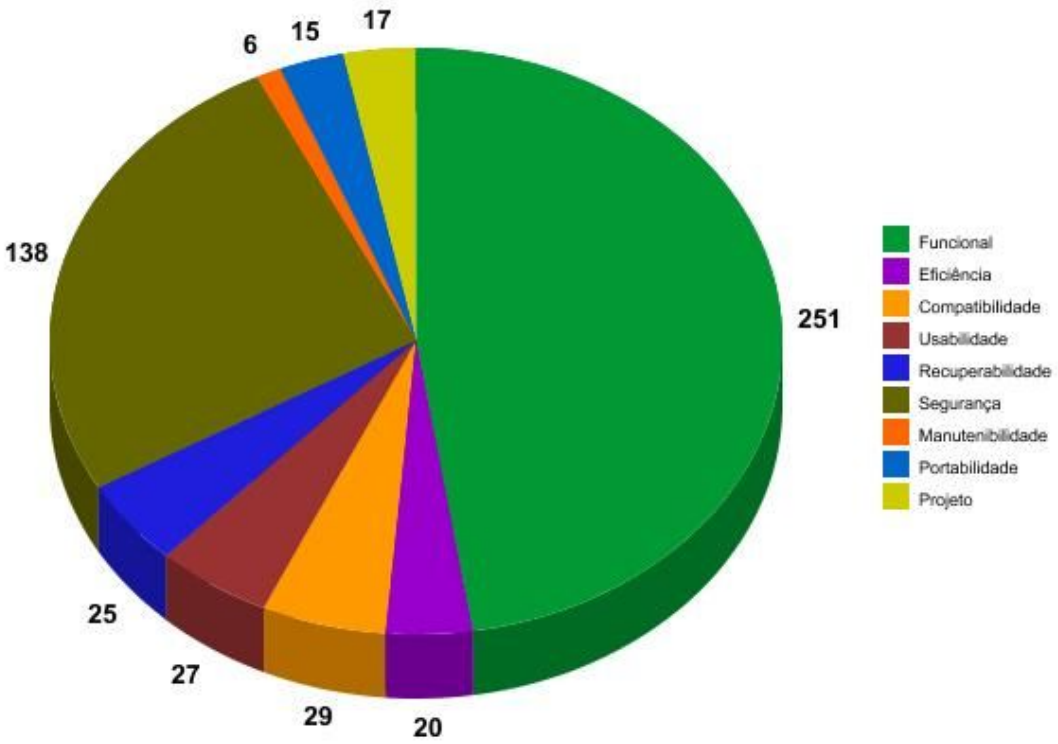


Figura 1 - Categorização conforme a ISO/IEC 25010:2015

Todo o processo do trabalho garante que os requisitos gerados atendem todas as exigências mencionadas conforme é exigido pela SBIS. Em consequência, para o escopo em questão, os processos pertinentes à engenharia de requisitos serão realizados de forma mais rápida e com menos riscos. Convém observar que essa especificação é parcial no sentido em que inclui apenas os requisitos derivados dos manuais que documentam as exigências de certificação. O resultado é bastante satisfatório pois um trabalho árduo e longo pode ser evitado caso um SISs deseja embasar e utilizar a planilha pronta como o primeiro passo da sua construção.

5. Considerações finais

A identificação de requisitos de software a partir das exigências da certificação da SBIS/CFM demandou a adequação de um método específico para essa finalidade. O método inicialmente empregado foi ajustado para acomodar as lições aprendidas ao longo do processo. A execução do método proposto revelou algumas exigências que, mesmo com o apoio do MOS, permaneceram obscuras e, consequentemente, precisam ser investigados de forma ainda mais criteriosa em ação posterior. De fato, uma ampla verificação do resultado é o principal trabalho futuro esperado.

Tendo em vista que a matriz de rastreabilidade produzida estabelece um mapeamento entre as exigências da certificação e os requisitos identificados, o que agiliza a verificação do resultado e, possivelmente, a aceitação integral do requisito produzido para a exigência em questão, além do resultado produzido já incluir um esforço considerável de desenvolvimento de requisitos, torna-se evidente que os requisitos definidos já permitem alimentar o desenvolvimento de um novo SIS reduzindo custos e riscos.

6. Referências

Wiegers, K. E. and Beatty, J. Software Requirements. Microsoft Press, 2013.

Jessica D. Young and Annie I. Antón. A Method for Identifying Software Requirements Based on Policy Commitments, 2010.

SBIS. Manual Operacional de Ensaios e Análises para Certificação de S-RES v2.1. Sociedade Brasileira de Informática em Saúde. Disponível em https://goo.gl/clFo9B, 2016, último acesso: 29/04/2017.

SBIS. Manual de Certificação da SBIS CFM v4.2 2016. Sociedade Brasileira de Informática em Saúde (SBIS). 2016. Disponível em https://goo.gl/Gxpm7I, último acesso: 29/04/2017.

SBIS. Sociedade Brasileira de Informática em Saúde. Portal acessado em 18/04/2017: http://www.sbis.org.br. 2017.

SOMMERVILLE, I. Software Engineering. 9th edition. Chapter 4 - Requirements engineering p.110-111, 2011.

ISO/IEC 25000 - System and Software Quality Requirements and Evaluation (SQuaRE) series of standards, 2015.

Matriz de rastreabilidade e Requisitos de Software gerados pelo estudo de caso, https://docs.google.com/spreadsheets/d/1M1XA9lI0Bb0wiHLoF4LP-EDvXa8V6A552LQdkZpdd-4/edit#gid=0.

BPM das Coisas
Uma abordagem prática

Walmir Nascimento B. de Melo[1], **Gilmar F. Arantes**[1]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

{walmirmelo, gilmar}@inf.ufg.br

Abstract. *The Internet of Things (IoT) has become a subject quite discussed in business and beyond. It is a concept with the potential to change the way society lives and works. IoT's challenge is to securely and efficiently capture vast amounts of data for analysis and information that can enhance business and many other processes. By integrating other concepts such as BPM (Business Process Management), it is possible to deepen the understanding of IoT's value generation in organizations. This article aims to address the theoretical and practical aspects of BPM integration with the Internet of Things and through a case study, where the benefits and challenges will be presented the joint use of these two concepts.*

Resumo. *A internet das Coisas (IoT) tornou-se um tópico bastante discutido nos negócios e fora deles. É um conceito com potencial de mudar a maneira como a sociedade vive e trabalha. O desafio da IoT é capturar, de forma segura e eficiente, enormes quantidades de dados para análises e informações que possam aprimorar os negócios e muitos outros processos. Com a integração de outros conceitos como o BPM (Business Process Management), é possível aprofundar o entendimento da geração de valor da IoT nas organizações. Este artigo tem como objetivo abordar aspectos teóricos e práticos da integração do BPM com Internet das Coisas e através de um estudo de caso, onde serão apresentados os benefícios e desafios da utilização em conjunto destes dois conceitos.*

1. Introdução

A internet das coisas vem ameaçando trazer mudanças drásticas para os negócios já há algum tempo [Soto 2015]. A IoT (*Internet of Things*) é uma grande rede de objetos conectados que são capazes de coletar e trocar dados entre si através de sensores embutidos [Meola 2016c]. Em resumo, IoT é a comunicação que ocorre entre esses objetos conectados a outros dispositivos, pessoas e sistemas. Porém, sem um controle adequado, todos esses dispositivos inteligentes serão apenas uma grande massa de dados fluindo sem direção. Segundo [Hat 2017], os petabytes gerados pela IoT serão apenas um peso a mais, a não ser que se tornem acionáveis. É preciso saber detectar as possibilidades de transformar dados isolados em informações de valor para o negócio, e com isso impactar positivamente seus processos internos e a experiência dos clientes.

Nesse contexto, para identificar onde e quando implementar as soluções de IoT em um ambiente corporativo, é necessário entender quais processos dentro da empresa carecem desse tipo de tecnologia. Tendo um mapa mais completo de onde efetivamente a

IoT pode ser aplicada através de metodologias voltadas para gerenciamento de processos de negócios, como o BPM.

Para atingir o objetivo proposto, o restante deste artigo está estruturado da seguinte forma: Na Seção 2 é apresentada a fundamentação teórica em que aborda os conceitos IoT e BPM-IoT; A Seção 3 é descreve um estudo de caso que aborda a aplicação prática destes conceitos; Na Seção 4 são apresentadas algumas discussões acerca do estudo de caso e finalmente na Seção 5 são apresentadas as considerações finais.

2. Fundamentação Teórica

2.1. IoT

A internet das coisas é composta por dispositivos conectados. Combinando esses dispositivos conectados com sistemas automatizados, é possível obter informações, analisá-las e criar ações para ajudar alguém com determinada tarefa, ou aprender de algum processo [Burgess 2017]. Isso significa que, a comunicação desses objetos com outros dispositivos e sistemas através da rede é o que os tornam diferentes de outros dispositivos desconectados.

"De acordo com Gartner até 2020 cerca de 20 bilhões de dispositivos estarão conectados" [Garter 2017]. Nota-se que é um campo em expansão e as áreas de aplicação da IoT são vastas e variadas, como o setor de agriculura, setor de transportes, para o comércio e saúde. Dentre estas, pode-se destacar a área da saúde, que de acordo com [Meola 2016b] será umas das indústrias com maior crescimento dos investimentos em IoT nos próximos anos. Outro setor que está rapidamente aderindo à IoT e o da agricultura, que tem previsão de crescimento dos investimentos em IoT em mais 50% até 2020 [Meola 2016d]. Fazendeiros poderão, por exemplo, usar *smartphones* para monitoramento remoto dos seus equipamentos e obter informações em tempo real de tempo e umidade, com o intuito de gerar resultados estatísticos.

Por outro lado, mesmo com todas suas aplicações e vantagens, a IoT ainda tem muitos desafios até consolidar-se no cotidiano das pessoas. Um deles, é a segurança [Meola 2016a], que é muito debatida nos dias atuais e tornou-se primordial nas organizações. Além disso, com a massiva troca de dados que haverá entre esses dispositivos, a IoT esbarra em outro desafio: o de possuir uma infraestrutura eficaz para possibilitar essa comunição sem gargalos. De acordo com [James Manyika 2015], estima-se que até 2025, os dispositivos de IoT acrescentarão entre 4 e 11 trilhões de dólares à economia mundial a cada ano. Dada essa perspectiva, o desafio é adequar as infraestruturas de rede para que se ajustem às novas demandas.

De forma análoga, mesmo que o Brasil já tenha dado um passo para adoção da IoT, com o plano nacional de IoT, em que reúne iniciativas para implementar a tecnologia no país, o ritmo de crescimento deverá ser lento. Segundo [Bucco 2017], o Brasil irá enfrentar algumas barreiras como a tributação e o preço, porém o volume de conexões mostra o grande potencial econômico. Além disso,"A agricultura é a nossa grande oportunidade. Imagine que em 2015 havia 45 mil colheitadeiras em uso, e menos de 10% eram conectadas", ressaltou.

2.2. BPM-IoT

O BPM foca em conectar pessoas, processos e sistemas de informação através da melhoria dos processos chave de uma organização. Dessa forma, o BPM pode prover valor aos projetos de IoT e fazer com que esses dispositivos inteligentes possam ser mais integrados aos processos de negócio. Assim sendo, o termo *"BPM of Things"*[Ozil 2015] já vem ganhando bastante credibilidade internacional e se tornando um campo promissor para o crescimento econômico das organizações.

A IoT ganha espaço, enquanto as organizações começam a avaliar o valor combinado da IoT com o BPM e se essa integração será crucial para a tomada de decisões. Os benefícios que o BPM trouxe às organizações possibilitou atingir maiores padrões de flexibilidade, eficiência e agilidade, através da modelagem de seus processos. De maneira similar, o impacto da IoT está transformando a indústria e toda a sua estrutura com o termo indústria 4.0 [Bernard 2016], em que sistemas monitoram todo o processo de fábrica e tomam decisões decentralizadas, comunicando e cooperando entre os dispositivos através da *web*.

No entanto, os dispositivos IoT trocam muitos dados entre si, e toda essa informação gerada será apenas um peso na infraestrutura, sem beneficio algum que possa gerar uma vantagem competitiva. É preciso tornar toda essa massiva troca de dados acessível, de forma que possa ser gerenciada em tempo real. Com o BPM, esses dados podem ser ligados a regras de negócio e processos, para que gerem valor. Já que na IoT, os objetos inteligentes mandam informação de um dispositivo conectado através de sensores para ativar uma resposta. Essa resposta, pode ser orquestrada por um processo completo, invocando outras respostas subsequentes como, por exemplo, na Figura 1, em que há um dispositivo que verifica a quantidade de lixo, e se caso a lixeira esteja cheia, envia tal informação para a central então recolhe-lo.

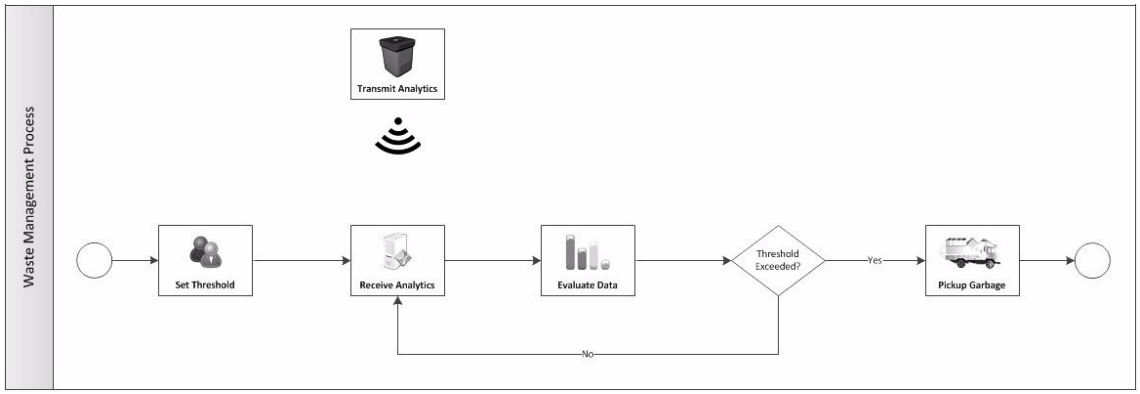


Figura 1. Processo de coleta de lixo [Javed 2015]

3. Estudo de Caso

Para investigar se existem benefícios com o uso integrado do BPM com a IoT, foi realizado, o mapeamento de alguns processos de uma indústria de fertilizantes localizada em Aparecida de Goiânia - GO. Foram mapeados 3 processos chave, e devido a extensão de cada um, optou-se por resumi-los às principais etapas relevantes para o estudo. São eles:

3.1. Compra de Matéria Prima (MP)

A indústria selecionada para o estudo possui anos de experiência com a produção e venda de fertilizantes para o agronegócio. Um de seus principais processos que impacta nos resultados do negócio é o de compra de matéria prima, conforme Figura 2.

Figura 2. Processo de Compra da Máteria Prima

Foi verificado que o principal gargalo desse processo é logo no inicio, quando é necessário o planejamento dos suprimentos e a geração de um documento com a quantidade de matéria prima. Pois, esse trabalho é feito pelo gerente de produção juntamente com o operacional, que precisam checar periodicamente a quantidade restante nos armazéns.

Diante desse cenário, a empresa está implementando balanças em todas as plataformas de matéria prima, para ter números exatos da quantidade restante e melhor previsão de compra futura. A Figura 3, ilustra o uso da transmissão de dados dos armazéns diretamente ao ERP, que por meio da quantidade previamente estabelecida pelo gestores, irá avaliar a quantidade necessária para compra.

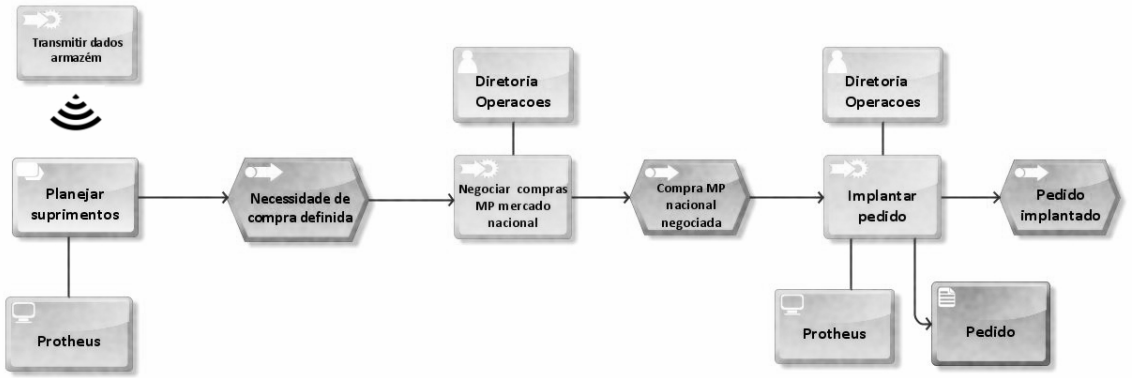


Figura 3. Processo de Compra da Máteria Prima integrado a IoT

3.2. Produção do Produto Acabado

Atualmente, o processo de produção está muito bem definido, conforme verificamos na Figura 4.

Porém, o principal problema está no desperdício e na qualidade do produto acabado. Tendo em vista que a matéria prima é armazenada em lugares fechados, fatores

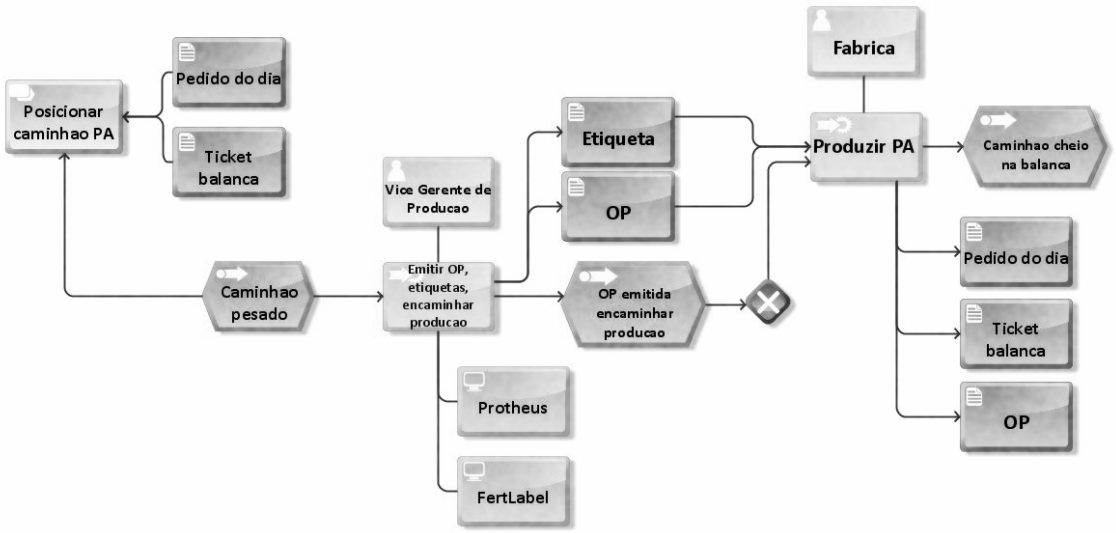


Figura 4. Processo de Produção de Produto Acabado

como a umidade do ar e temperatura afetam drasticamente o produto final. Dessa forma, a empresa tem em seu planejamento, a implantação de diversos dispositivos de temperatura e umidade conectados a rede nos seus armazéns. Com o intuito de obter melhor controle e tomar decisões em tempo real para evitar a perda de qualidade e evitar desperdícios. Na Figura 5, verifica-se o conceito de IoT inserido no processo de produção de fertilizantes. Em que o vice gerente de produção receberá as informações em tempo real das condições da matéria prima no armazém.

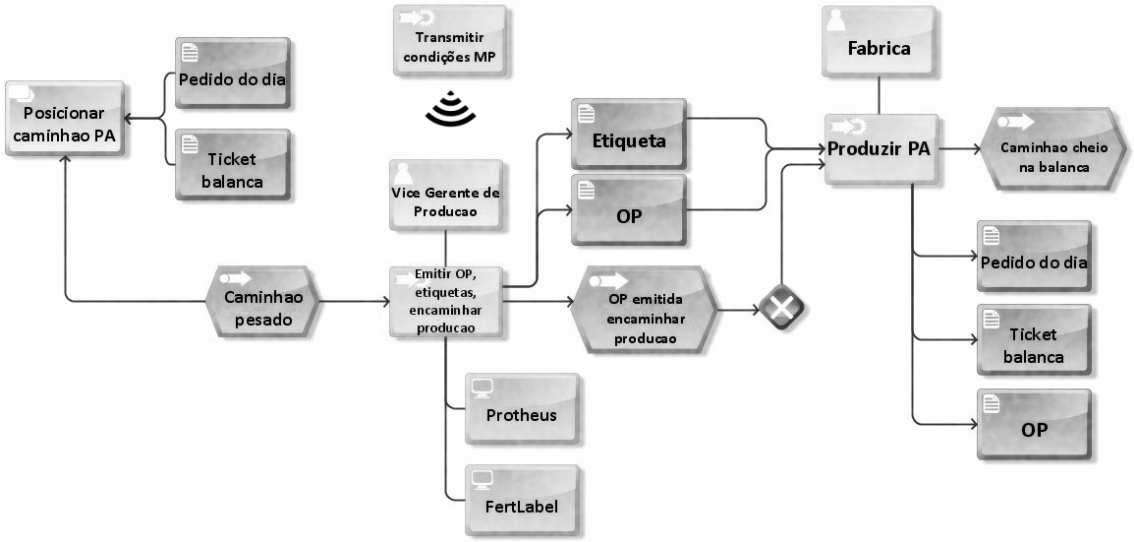


Figura 5. Processo de Produção de Produto Acabado utilizando IoT

4. Discussões

Com base nas informações obtidas, foi possível obter um entendimento de alguns processos chave. A empresa está passando por uma reestruturação de seus processos, e foi

interessante para a organização fazer um trabalho que pode vir a ser usado no futuro. Automatizar e integrar cada vez mais os seus processos de forma contínua é objetivo de muitas organizações que lutam para continuar entregando valor para seus clientes mesmo em tempos de crise.

Com a modelagem dos processos vistos na Seção 3, verificou-se os principais gargalos de dois processos chave na organização e em colaboração com visão futura da empresa. Então, foi modelado um cenário hipotético com a integração da IoT nesse ambiente. Nesse cenário criado, foi constatado que de fato, seus processos foram melhorados com a integração da IoT. Além da redução de passos, para uma determinada tarefa, houve melhoria na integração do atual ERP da empresa. Todavia, para que esse cenário torna-se realidade a empresa esbarra em problema que será comum a muitas outras, que é o de investimentos em infraestrutura para suportar esses dispositivos conectados.

5. Considerações Finais

Apesar de toda a inovação e agilidade em integrar o BPM com IoT, um aspecto é necessário para que o BPM das Coisas possa ser tornar disruptiva em todos os aspectos de como os negócios são conduzidos, trata-se da integração com tecnologias comuns. Pois os resultados deve ser analisados e comunicados para a tomada de decisões praticamente em tempo real, para ser efetivo. Se houver qualquer atraso entre o tempo que a IoT leva para coletar os dados, e o tempo que o BPM leva para processar e submeter tais dados aos tomadores de decisões, a agilidade do BPM das Coisas torna-se menos eficiente. Muitos passos podem gerar gargalos que diminuem os efeitos da integração BPM e IoT.

Para que sejam verificados realmente os benefícios da colaboração entre BPM e IoT deve-se aprofundar este trabalho à medida em que for mapeando os outros processas da empresa. Tendo em vista as dificuldades e limitações de se trabalhar com esse tema, tendo em vista pequena quantidade de referências disponíveis, além da falta de informação do atual mercado brasileiro em relação ao uso de IoT. Recentemente as empresas brasileiras começaram a notar o potencial do BPM em seus negócios é só uma questão de tempo até que passem a notar também a IoT.

Os dispositivos conectados estão prevalacendo em vários setores, desde do consumidor até a indústria, ocorrendo interações entre máquinas, e não apenas pessoas e máquinas. Essa união de tecnologias está apenas em seu estágio inicial e esse artigo tem como intuito promover a pesquisa em campo tão amplo e forte crescimento nos próximos anos.

Referências

Bernard, M. (2016). What everyone must know about industry 4.0. Disponível em: https://www.forbes.com/sites/bernardmarr/2016/06/20/what-everyone-must-know-about-industry-4-0/#11206f79795f, acesso em 07/09/2017.

Bucco, R. (2017). Expansão da iot no brasil deve ser lenta. Disponível em: http://www.telesintese.com.br/expansao-da-iot-no-brasil-deve-ser-lenta-preve-teleco/, acesso em 15/09/2017.

Burgess, M. (2017). What is the internet of things? wired explains. Disponível em: http://www.wired.co.uk/article/internet-of-things-what-is-explained-iot, acesso em 22/09/2017.

Garter, P. (2017). Gartner says 8.4 billion connected "things"will be in use in 2017, up 31 percent from 2016. Disponível em: https://www.gartner.com/newsroom/id/3598917, acesso em 15/09/2017.

Hat, R. (2017). Bpm and iot: Transformation to the physical world. Disponível em: https://www.redhat.com/en/resources/bpm-and-iot-transformation-physical-world, acesso em 11/09/2017.

James Manyika, R. D. (2015). The internet of things: Mapping the value beyond the hype. Disponível em: http://globaltrends.thedialogue.org/publication/the-internet-of-things-mapping-the-value-beyond-the-hype/, acesso em 14/09/2017.

Javed, A. (2015). Business process improvement using internet of things (part 2). Disponível em: http://www.bpmleader.com/2015/02/11/business\-process-improvement-using-internet-of-things-part-2, acesso em 07/09/2017.

Meola, A. (2016a). How the internet of things will affect security and privacy. Disponível em: http://www.businessinsider.com/internet-of-things-security-privacy-2016-8, acesso em 14/09/2017.

Meola, A. (2016b). Internet of things in healthcare: Information technology in health. Disponível em: http://www.businessinsider.com/internet-of-things-in-healthcare-2016-8, acesso em 14/09/2017.

Meola, A. (2016c). What is the internet of things (iot)? Disponível em: http://www.businessinsider.com/what-is-the-internet-of-things-definition-2016-8, acesso em 18/09/2017.

Meola, A. (2016d). Why iot, big data and smart farming are the future of agriculture. Disponível em: http://www.businessinsider.com/internet-of-things-smart-agriculture-2016-10, acesso em 14/09/2017.

Ozil, P. (2015). Bpm of things: the next generation of the internet of things. Disponível em: http://data-informed.com/bpm-of-things-the-next-generation-of-the-internet-of-things/, acesso em 08/09/2017.

Soto, C. (2015). Bpm and iot: The signal of true disruption. Disponível em: http://software.schneider-electric.com/pdf/white-paper/wonderware-bpm-and-iot---the-signal-of-true-disruption/, acesso em 12/09/2017.

EasYii Gii, Um Gerador de código para aplicações MVC, em PHP, baseado em banco de dados relacionais

Matheus Evangelista Morais[1], Alessandro Rodrigues e Silva[2]

[1,2] Instituto Federal de Educação, Ciência e Tecnologia de Goiás – Campus Anápolis
Av. Pedro Ludovico, s/n, Residencial Reny Cury. CEP: 75131-457. Anápolis-GO

[1]thtmorais@hotmail.com, [2]alessandro.rodrigues@ifg.edu.br

Abstract. *This work details the operation of a code generator for MVC applications based on relational databases, using the Yii2 framework for PHP, which reduces the production time of custom applications generating standardized code. A set of templates that execute CRUD operations are used, considering all information in the relational schema. For the mapping of the information of the scheme we use pattern recognition algorithms.*

Resumo. *Este trabalho detalha o funcionamento de um gerador de código para aplicações MVC baseado em banco de dados relacionais, utilizando o framework Yii2 para PHP, que reduz o tempo de produção de aplicações personalizadas gerando códigos padronizados. É utilizado um conjunto de templates que executam operações CRUD considerando todas as informações do esquema relacional. Para o mapeamento das informações do esquema utiliza-se de algoritmos de reconhecimento de padrões.*

1. Introdução

Atualmente o desenvolvimento de software de qualidade em menor tempo tem sido bastante almejado. Este problema pode ser resolvido através de geradores de códigos padronizados de acordo com dados e regras estabelecidas. O uso dos geradores traz vantagens em relação a codificação manual em vista da redução significante do tempo necessário para a escrita de regras comuns, além de aumentar a qualidade do código fonte, produzindo código padronizado, portanto, facilitando a manutenção. O desenvolvimento de um gerador como um componente de software altamente sofisticado motiva os desenvolvedores de software e reduz a monotonia que surge durante a codificação manual de código semelhante ou repetitivo [HERRINGTON, 2003].

Existe vários geradores de código para PHP, como o *Grocery CRUD* [SKOUMBOURDIS, JOHN. 2017] e o *Gii* [BOGDANOV , ANDREW.; ELISEEV , DMITRY., 2016]. Ambos utilizam o padrão *Active Record* que proporciona o relacionamento entre o usuário, o banco de dados e um objeto; onde as tabelas são representadas através de classes, os registros através de instâncias e os campos da tabela por atributos do objeto. As ações com as linhas da tabela -*CRUD* (criar, ler, atualizar e excluir) – se realiza pelos métodos desta classe. Outro benefício do Active Record é o uso do *ORM* [SAFRONOV, M.; WINESETT, J., 2014] – *Object-Relational Mapping*, uma maneira eficiente de armazenar os dados em banco de dados relacionais e com ela se elimina a necessidade do mapeamento manual.

Diante das soluções existentes, que já trazem uma grande redução de código escrito, optou-se por complementar as funcionalidades do *Gii*, acrescentando-se

funcionalidades não existentes ou incompletas. O *Gii* é um componente do framework *Yii2*.

O framework *Yii2* implementa o *Active Record* pela classe *yii/db/ActiveRecord* eliminando-se a necessidade de escrever instruções *SQL*, principalmente as que estão relacionadas ao *CRUD*.

A solução, o *EasYii Gii*, adiciona novas funcionalidades como segurança baseada em papéis (RBAC), ações em funções *CRUD* como de exclusão de registros referenciados por chaves estrangeiras, validações e integridade de campos como CPF/CNPJ, datas, arquivos, além de incorporar componentes visuais mais amigáveis para dispositivos móveis.

2. Metodologia

Com as classes geradas com o *ORM* padrão do *Yii2* é possível identificar os tipos de dado, o nome das colunas e relações com outras tabelas. Mas um problema identificado é que os nomes dos campos dos formulários são extraídos diretamente dos nomes das colunas, gerando nomes não muitos amigáveis. Como solução foi utilizado o texto inserido como comentário em cada coluna de uma tabela do banco de dados para denominar um rótulo (label) de um campo em um formulário HTML. A Figura 1 mostra um código que extrai os comentários de cada campo de uma tabela.

```
SELECT c.column_name,pgd.description
FROM pg_catalog.pg_statio_all_tables AS st
INNER JOIN pg_catalog.pg_description pgd ON (pgd.objoid=st.relid)          INNER JOIN information_schema.columns c
ON (pgd.objsubid=c.ordinal_position
AND c.table_schema=st.schemaname AND c.table_name=st.relname)
WHERE c.table_name='$table' AND c.column_name='$column'
```

Figura 1: Consulta SQL para gerar novos nomes para campos.

Outra consulta necessária se refere a coleta das informações do esquema quanto a chaves estrangeiras para obter mais informações tais como: tabelas que referenciam uma determinada chave primária, o nome do campo de referência e principalmente qual ação dever ser tomada ao excluir uma chave. O SGBD preferencialmente a se utilizar, o *PostgreSQL*, possibilita três ações: *CASACADE*, *RESTRICT* e *NO ACTION*.

RESTRICT impede a exclusão de uma linha referenciada. *NO ACTION* significa que, se ainda houver linhas de referências ainda quando a restrição for verificada, um erro é gerado. *CASCADE* especifica que, quando uma tupla é excluída, qualquer tupla que possui uma chave para àquela excluída devem ser excluídas.

```
SELECT DISTINCT(ccu.table_name) AS tablefk, kcu.table_name
AS tableref, kcu.column_name AS columnref, ccu.column_name
AS columnfk, rc.delete_rule AS actiondelete
FROM Information_schema.table_constraints AS tc
JOIN information_schema.key_column_usage AS kcu
USING (constraint_schema, constraint_name)
JOIN information_schema.constraint_column_usage AS ccu
USING(constraint_schema, constraint_name)
JOIN information_schema.referential_constraints AS rc
USING (constraint_schema, constraint_name )
WHERE constraint_type = 'FOREIGN KEY' AND ccu.table_name='$table'
```

Figura 2: Consulta SQL para ações ao efetuar exclusões.

A Figura 2 mostra a consulta em *SQL* para obter-se estas informações, com base na saída desta consulta, gera-se um código que restringe na aplicação as exclusões caso as permissões forem do tipo *RESTRICT* ou *NO ACTION*. Sem as restrições na aplicação, caso o usuário tentar apagar alguma informação que é uma referência de alguma chave estrangeira, uma exceção será exibida na tela ao invés de uma mensagem amigável ao usuário. O resultado desta consulta é similar ao da Tabela 1:

O *template* gerado com *Gii* do *Yii* referência controles *HTML* simples, não aconselháveis para uso em dispositivos móveis, não considera várias restrições do modelo relacional, como chaves únicas, além de não assumir a presença de controle de acesso para garantir a segurança da aplicação. Entretanto, qualquer gerador de código não atende, completamente, as necessidades específicas de uma determinada equipe de desenvolvimento, o mesmo necessita ser customizado.

Um *template*[POREBSKI, BARTOSZ; PRZYSTALSKI, KAROL; NOWAK, LESZEK, 2011] permite que você desenvolva sistemas com diferentes *layouts* ou temas para a mesma funcionalidade do núcleo. E também é um exemplo de *DRY (Don't Repeat Yourself)*. Pode-se usar apenas um arquivo de modelo para algumas funcionalidades do código, isto é possível devido a separação que o *MVC* fornece. Um exemplo de *template* é mostrado na Figura 4, onde para qualquer exclusão, apenas os dados são modificados e a lógica continua sempre a mesma.

Tabela 1. Resultado de uma consulta SQL sobre restrições ao excluir.

Tablefk	tableref	columnref	Columnfk	Actiondelete
Pessoa	solicitacao	id_pessoa	Cpf	RESTRICT
Pessoa	aluno	Aluno	Cpf	RESTRICT
Pessoa	professor	cpf_pessoa	Cpf	RESTRICT

No processo de gerar o código, para todo campo da tabela do banco de dados, é verificado qual é o tipo de dado, qual o nome do campo. Com está verificação realizada, define-se qual e se precisa utilizar *plugins* de terceiros como *Select2*, *DataPicker*, *InputFile*, *MaskedInput* e dentre outros (Figura 5) (Figura 7). No processo geração de código do *EasYii Gii* pode-se informar se há relações com outras tabelas, caso este campo seja preenchido, o gerador cria *behaviors* para as relações *NxM*. No *EasYii Gii*, para gerar um código é necessário apenas informar a tabela desejada, os caminhos são definidos de forma padrão (Para o *MODEL*: *@app/models/Model*; para o *CONTROLLER*: *@app/controllers/GeradoController*; e para a *VIEW*: @app/views/*Gerado*).

3. Resultados e discussões

A seguir analisa-se dois códigos gerados pelo framework *Yii* 2.0, o primeiro utiliza-se de seu gerador padrão (Figura 3) e o segundo utiliza-se o gerador desenvolvido (Figura 4), ambos para a operação *CRUD* de exclusão.

```
public function actionDelete($id){
            $this->findModel($id)->delete();
            return $this->redirect(['index']);
}
```

Figura 3: Código padrão gerado pelo Yii 2.0 para exclusão.

O código padrão que é produzido pelo framework *Yii* 2.0 não verifica restrições de chaves estrangeiras, e se caso houver algum tipo de restrição, como o *RESTRICT,* irá retornar uma exceção ao usuário pois o SGBD impossibilitou a exclusão. Com o gerador desenvolvido:

```
public function actionDelete($id){
    if (Solicitacao::find()->where(['id_pessoa'=>$id])->all()){
        Yii::$app->session->setFlash('danger', "Pessoa está cadastrado.");
        return $this->redirect(['index']);
    }
    if (Aluno::find()->where(['aluno'=>$id])->all()){
        Yii::$app->session->setFlash('danger', "Pessoa está cadastrado.");
        return $this->redirect(['index']);
    }
    if (Professor::find()->where(['cpf_pessoa'=>$id])->all()){
        Yii::$app->session->setFlash('danger', "Pessoa está cadastrado.");
        return $this->redirect(['index']);
    }
    $this->findModel($id)->delete();
    return $this->redirect(['index']);
}
```

Figura 4: Novo código gerado para exclusão.

O código produzido pelo novo gerador faz verificações para todos os dados fornecidos pelo campo 'tableref' (Figura 2). Caso aquela tabela da base de dados não possua restrições ao excluir ou fosse do tipo *CASCADE*, não será preciso fazer nenhum tipo de verificação ao excluir e o código produzido pelo novo gerador será igual ao gerador padrão do framework.

No novo gerador acrescentou-se identificadores de padrões para gerar campos de formulários personalizados de acordo com o tipo de dado ou com o nome do campo da tabela. Como mostra a Figura 5, caso o tipo de dado for '*datetime'* irá gerar um campo com calendário para que o usuário final selecione a data facilmente evitando assim a entrada de dados inválidos – não elimina a necessidade de verificação posteriormente no servidor. A Figura 6 mostra a visualização que o usuário terá para este tipo de dado.

```
if ($column->dbType === 'datetime') {
return "'$attribute' => ['type' => TabularForm::INPUT_WIDGET,
    'widgetClass' => \\kartik\\datecontrol\\DateControl::classname(),
    'options' => [
        'type' => \\kartik\\datecontrol\\DateControl::FORMAT_DATETIME,
        'saveFormat' => 'php:Y-m-d H:i:s',
        'ajaxConversion' => true,
        'options' => [
            'pluginOptions' => [
                'placeholder' => " . $this->generateString('Choose ' . $humanize) . ",
                'autoclose' => true,
            ]
        ],
    ]
    ]";
}
```

Figura 5: Gerador de campos do tipo data.

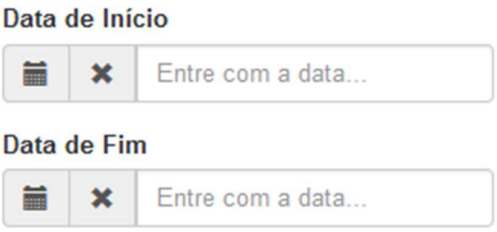


Figura 6: Campos de formulário do tipo 'datetime'.

A identificação por nome de campos se dá de forma similar à de tipo de dado, com ele a validação de campos pode-se ser gerada automaticamente como ocorre nos casos de CPJ/CNPJ, e-mail e arquivos. A Figura 7 mostra o código de validação produzido para campos CPF/CNPJ, e-mail ou arquivo:

```
if($line['column_name']=='email'){
    array_push($rule,"[['email'],'email']");
}
else if($line['column_name']=='cpf'){
    array_push($rule,"['cpf', CpfValidator::className()]");
}
else if($line['column_name']=='cnpj'){
    array_push($rule,"['cpf', CnpjValidator::className()]");
}
else if($line['column_name']=='arquivo'){
    array_push($rule,"['arquivo', 'file']");
}
```

Figura 7: Validações para CPF/CNPJ, arquivo e e-mail.

Caso o usuário inserir algum dado inválido, assim que remover o cursor do mouse do campo lhe será mostrado um alerta, caso o dado esteja inválido e algumas funções de navegador estiverem desativadas, o alerta de erro será mostrado após submissão – porém os dados não são salvos – e os campos do formulário serão auto preenchidos com os dados que foram antes inseridos evitando a perca de tempo e de dados. A Figura 8 mostra como os alertas são mostrados ao usuário.

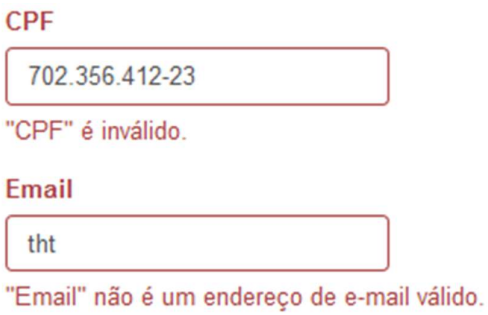


Figura 8: Dados inválidos

Para a parte de segurança, implementa-se o acesso baseado em papéis com a classe *webvimark/modules/UserManagement*. Esta classe utiliza-se do *RBAC (Role-Based Access Control – Controle de Acesso Baseado em função)* [CHANDRAMOULI et. al. 2012] onde se insere a função apropriada a usuários, grupos e aplicativos em determinado escopo, que é disponibilizado através de uma interface web.

Com este módulo é necessário apenas que os *Controllers* herdem de uma específica superclasse para restringir o acesso. Posteriormente o administrador deve criar os papéis informando quais as rotas permitidas para cada usuário ou papel. A classe também disponibiliza um menu com as funcionalidades definidas no *RBAC* para cada usuário.

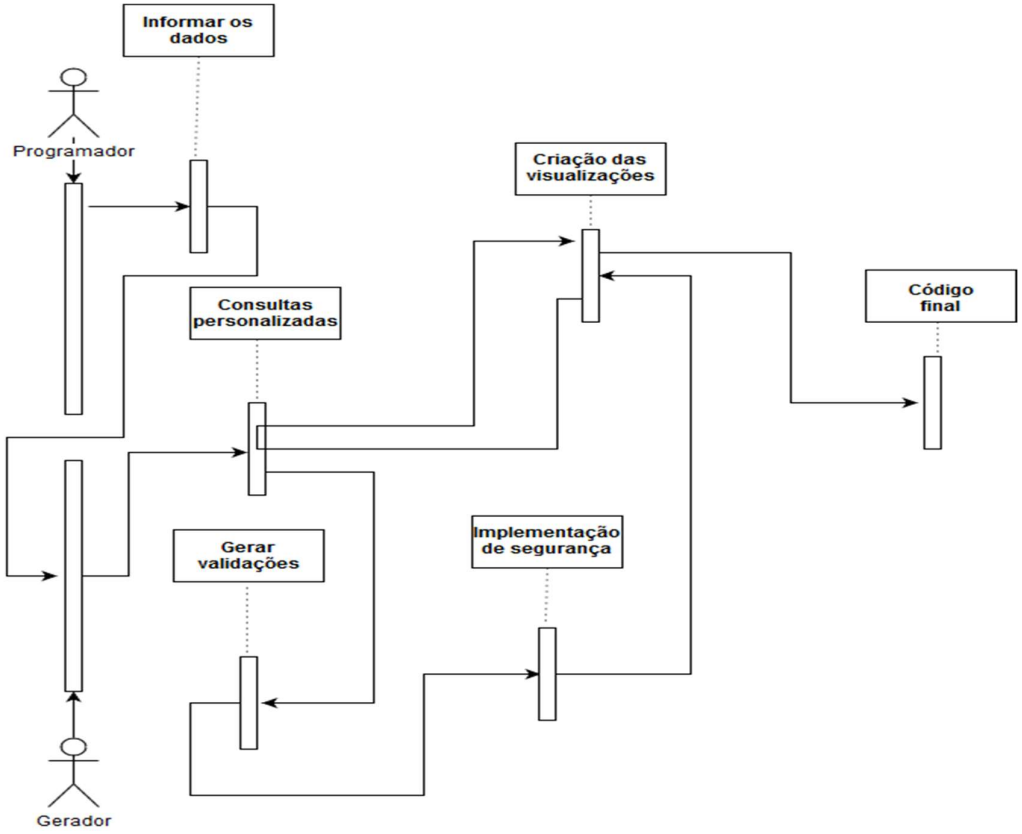


Figura 9: Diagrama simplificado da geração do código.

4. Considerações finais

O EasYii Gii consegue acelerar o processo de produção de código uma vez que muitos passos são realizados automaticamente, eliminando o tempo necessário para codificação manual além de agilizar a manutenção criando código padronizado.

Várias funcionalidades ainda estão no backlog para serem implementadas, mas não impede a ferramenta de ser utilizada. O EasYii Gii é um software livre (Apache licence - https://www.apache.org/licenses/LICENSE-2.0) disponível no endereço https://github.com/thtmorais/easyiigii

5. Referências

HERRINGTON, J. Code Generation in Action. Manning Publications Co., 2003.

FAYAD, M.; SCHMIDT, D. C.; JOHNSON, R. E. Building application frameworks : object-oriented foundations of framework design. [s.l.] Wiley, 1999.

FOWLER, M. Patterns of enterprise application architecture. [s.l.] Addison-Wesley, 2003.

SAFRONOV, M.; WINESETT, J. Web Application Development with Yii 2 and PHP. Packt Publishing, 2014.

CHANDRAMOULI, RAMASWAMY; FERRAIOLO, DAVID F.; KUHN, RICHARD D. Role-Based Administration of RBAC: Chapter 9 from Role-Based Access Control, Volume 2. Artech House Digital, 2012.

BOGDANOV , ANDREW.; ELISEEV , DMITRY. Yii2 Application Development Cookbook - Third Edition. Packt Publishing, 2016.

SKOUMBOURDIS, JOHN. (2017) "Documentation Cook Book", https://www.grocery*CRUD*.com/documentation, Outubro.

POREBSKI, BARTOSZ; PRZYSTALSKI, KAROL; NOWAK, LESZEK. Building PHP Applications with Symfony, CakePHP, and Zend Framework. John Wiley and Sons, 2011

Index of Authors

Abrahão, Diego, 119
Almeida, Dyego, 303
Alves, Rafael, 211
Arantes, Gilmar, 383
Araújo, Adailton, 279
Araújo, Júlia, 199

Batista, Mariane, 223
Borelli, Helberth, 323
Borges, Vinicius da Cunha Martins, 119
Braga, Renata, 367, 375
Brito, Leonardo, 343
Bueno, Rodolfo, 279
Bulcão Neto, Renato, 237

Cardoso, Alisson, 133
Carvalho, Marcos, 53
Carvalho, Sergio T., 159, 313, 323
Castro, Giovanna, 39
Castro, Marcelo Stehling, 79
Cavalcanti, Walison, 333
Correa, Henrique, 79
Costa, Fabio, 251
Costa, Ronaldo, 303
Cruz, Lucas Fiorini, 147, 351

David, Rodrigo Sasse, 359
Dias, Márcio, 11
Dias, Rodrigo, 39

Fernandes, Deborah, 359
Fernandes, Márcio Giovane Cunha, 359
Ferreira, Deller, 199
Ferreira, Marcus Vinícius Gonzaga, 93
Filho, Gilberto, 67
Fortes, Marcelo, 251
Franco, Ricardo, 67, 79

Gonçalves, Cristhiane, 303

Jr Oliveira, Antônio, 173
Junqueira, Natália, 313

Lemos, Rodrigo Pinto, 147, 351
Lucena, Fabio, 367, 375

Magalhães, Marcela, 147, 351
Martins, Ernane, 265
Melo, Walmir, 383
Mendonça, Emily, 199
Monção, Ana Claudia Loureiro, 223
Morais, Edison, 223
Morais, Matheus, 391
Moura, Júlia, 199

Neto, Raimundo N. de A. S., 185, 293
Neves, Leonardo, 375
Nogueira, Tiago, 185, 199, 293

Oliveira, Ana Paula de Sousa, 159
Oliveira, Leandro, 303
Oliveira,Hayanna, 211

Palhares, Pedro Henrique, 343
Povoa, Caio Jose, 343

Rezende, Sônia Regina, 185
Ribeiro, Alison, 25
Ribeiro, Rosário, 173
Ribeiro-Rotta, Rejane, 367, 375
Rocha, Flávio Geraldo Coelho, 105

Salvini, Rogerio, 39
Santana, Ygor Escatolim Castor, 333
Santos, Paulo Victor, 303
Silva, Alessandro, 391
Silva, Cláudio, 293
Silva, Gaspar, 11
Silva, Gilton, 211
Silva, Lafaiet C., 105
Silva, Leonardo, 211
Silva, Maykon Renan Pereira, 53
Silva, Murilo Barcelos, 237
Silva, Nádia, 11, 25, 333
Silveira, Beatriz, 367
Sommer, Reinoldo Oliveira, 265
Souza, Dalton Foltran, 119

Vieira, Flávio Henrique Teles, 53, 67, 79, 93, 105, 119, 133
Vilela, Jorge Tawnan Leles, 265