

Universidade Federal de Uberlândia  
Faculdade de Computação

# Anais da III FACOM Techweek

30 de maio a 03 de junho de 2016  
ISSN: 2447-0406



Uberlândia  
2016

Copyright© 2016 da Faculdade de Computação (UFU)  
Todos os direitos reservados

# Sumário

<b>I Trabalhos de Graduação</b>	<b>7</b>
<b>Ctrl-car: um robusto sistema “open source” para controle de entrada e saída de veículos</b> . . . . .	<b>9</b>
<i>Marcus Evandro Teixeira Souza Junior, Rodrigo Almeida Boghossian, Shiguelo Nomura</i>	
<b>Estratégias bioinspiradas aplicadas à robótica na vigilância</b> . . . . .	<b>13</b>
<i>Claudiney R. Tinoco, Danielli A. Lima, Gina M. Barbosa</i>	
<b>Estudo e Aplicação de Técnicas de Registro de Imagens para Reconstrução 3D da Rede de Canais Vasculares do Tecido Ósseo</b> . . . . .	<b>17</b>
<i>Thiago Togo Nunes, Bruno Augusto Nassif Travençolo</i>	
<b>Estudo sobre a adoção de PDTI e CGTI pelas Universidades Federais</b> . .	<b>21</b>
<i>Gustavo Caetano Borges, Romário Vieira, Rodrigo Sanches Miani, João Batista Simão</i>	
<b>Mapeamento de ambiente para correção de trajetória de robôs</b> . . . . .	<b>25</b>
<i>Igor Gomes Arantes, Mauricio Cunha Escarpinati</i>	
<b>Agrupamento de Dados via Combinação de Partições Utilizando Amostragem de Dados</b> . . . . .	<b>29</b>
<i>Paulo Henrique Maia Soares, Daniel Duarte Abdala</i>	
<b>Caracterização de Nanofibras Geradas por Bactérias via Métodos de Processamento de Imagens</b> . . . . .	<b>33</b>
<i>Mariana Luiza Volpi Fiori, Daniel Duarte Abdala</i>	
<b>Cuca Chess: Um jogo de xadrez usando BCI</b> . . . . .	<b>37</b>
<i>Gustavo Vieira Soares, João Henrique de Souza Pereira</i>	
<b>Estudo dos parâmetros do Algoritmo Evolutivo aplicado ao Problema de Predição de Estruturas de Proteínas usando modelo HP-2D</b> . . . . .	<b>41</b>
<i>Júlia Manfrin Dias, Christiane R. S. Brasil</i>	
<b>Fatores que influenciam o recebimento de spam</b> . . . . .	<b>45</b>
<i>Danielle Silva Oliveira, Rodrigo Sanches Miani</i>	
<b>Implementação de algoritmos de agrupamento de dados via combinação de partições em hardware</b> . . . . .	<b>49</b>
<i>João Paulo Martins Arantes, Daniel Duarte Abdala</i>	

<b>Um Método para Avaliação da Variabilidade de Conjuntos de Partições</b> . . . . .	<b>53</b>
<i>Nayara G. Ribeiro, Daniel D. Abdala</i>	
<b>II Trabalhos de Pós-Graduação</b> . . . . .	<b>57</b>
<b>Abordagem de Recomendação Automática de Objetos de Aprendizagem Utilizando Tecnologias de Web Semântica Aplicadas ao Sistema Educacional Classroom eXperience</b> . . . . .	<b>59</b>
<i>Vitor César de Carvalho, Fabiano Azevedo Dorça</i>	
<b>Ambiente web para citomorfometria usando imagens 3D</b> . . . . .	<b>61</b>
<i>Eduardo Henrique Silva, Bruno Augusto Nassif Travençolo, Jefferson Rodrigo de Souza</i>	
<b>Aplicação de Redes Complexas no Estudo de Interação Proteína-Proteína</b> . . . . .	<b>67</b>
<i>Getúlio de Moraes Pereira, Anderson Rodrigues dos Santos</i>	
<b>Um Método de Acesso Métrico Eficiente para Suporte às Consultas por Similaridade em Fluxo Contínuo de Dados</b> . . . . .	<b>73</b>
<i>Régis Michel dos Santos Sousa, Humberto Luiz Razente</i>	
<b>Método de Estimção de Chuva Acumulada Utilizando Técnicas de Mineração de Dados e Processamento Digital de Sinais em Dados Coletados por Smartphones</b> . . . . .	<b>75</b>
<i>Luiz Fernando Afra Brito, Marcelo Keese Albertini</i>	
<b>Rastreamento e Evolução de Restrições em Agrupamentos Semi-Supervisionados</b> . . . . .	<b>77</b>
<i>Guilherme Alves, Maria Camila N. Barioni</i>	
<b>Avaliando Cookbooks Construídos a partir de Engenharia Reversa: Um Experimento Controlado</b> . . . . .	<b>79</b>
<i>Raquel Fialho de Queiroz Lafetá, Marcelo de Almeida Maia</i>	
<b>Avaliando Projeto de Software por Priorização de Classes-Chave</b> . . . . .	<b>81</b>
<i>Liliane do Nascimento Vale, Marcelo de Almeida Maia</i>	
<b>Classificação de Imagens Histológicas de Lesões Proliferativas Intraductais de Mama</b> . . . . .	<b>83</b>
<i>Walter Alexandre A. de Oliveira, Marcelo Zanchetta do Nascimento</i>	
<b>Métodos computacionais para caracterização e análise do desenvolvimento de Embriões de <i>Drosophila melanogaster</i></b> . . . . .	<b>85</b>
<i>Daniela Justiniano de Sousa, Bruno Augusto Nassif Travençolo</i>	
<b>Recomendação de Conhecimento da Multidão para Auxílio na Correção de Bugs Independentes de Domínio de Aplicação</b> . . . . .	<b>87</b>
<i>Eduardo Cunha Campos, Marcelo de Almeida Maia</i>	

**Sistemas Tutores Inteligentes Aplicados a Ambientes Educacionais Ubíquos:  
Proposta, Implementação e Avaliação . . . . . 89**  
*Hiran Nonato Macedo Ferreira, Renan Gonçalves Cattelan*



**Parte I**

**Trabalhos de Graduação**



# Ctrl-car: um robusto sistema “open source” para controle de entrada e saída de veículos

**Autor: Marcus Evandro Teixeira Souza Junior<sup>1</sup>**

**Rodrigo Almeida Boghossian<sup>1</sup>**

**Shiguelo Nomura<sup>2</sup>**

Faculdade de Engenharia Elétrica<sup>1</sup>

Faculdade de Computação<sup>2</sup>

Universidade Federal de Uberlândia – Uberlândia, MG – Brasil

[nunioll@hotmail.com](mailto:nunioll@hotmail.com), [digo\\_1\\_almeida@hotmail.com](mailto:digo_1_almeida@hotmail.com)  
[shigueonomura@gmail.com](mailto:shigueonomura@gmail.com)

Nível: Iniciação Científica/Trabalho de Conclusão de Curso

Bolsa de Estudo: Nenhuma

**Resumo.** *O sistema Ctrl-car consiste em um software robusto, simples e de baixo custo de entrada e saída de veículos. Seu diferencial em relação outros sistemas similares no mercado é o fato de ter uma proposta social mais abrangente. Seu custo de instalação é mínimo, com uso de poucos recursos humanos, o que faz dele excelente tanto para pequenos quanto para grandes empresários. Além disso, o modo de cobrança é justo, pois é feito pelo preço por minuto que o cliente usou o estacionamento. Também se destaca a segurança para o cliente que, quando dá entrada, tem vários dados pessoais registrados para o controle da saída, evitando roubos de veículos. Todo o trabalho foi baseado em ferramentas de desenvolvimento de código aberto tais como Apache, CGI e Emacs de uso profissional. Os resultados obtidos em testes de funcionamento do sistema mostraram que os objetivos propostos foram plenamente alcançados. Inclusive há uma grande motivação para aprimorar o sistema por estar despertando interesse de diversas instituições públicas e privadas em explorar ou diversificar o uso do sistema.*

**Palavras-Chave:** *Código aberto, controle, estacionamento, Internet, segurança, software.*

## 1. Introdução

Segundo a Associação Nacional dos Fabricantes de Veículos Automotores (Anfavea) somente nos últimos 12 meses (de maio de 2015 a abril de 2016) foram produzidos 2,12 milhões de veículos e licenciados 2,24 milhões. Juntamente com o aumento da frota de veículos, tem aumentado a procura por estacionamentos comerciais que ofereçam segurança e preço justo.

A maioria dos estacionamentos comerciais já existentes utilizam basicamente dois tipos de sistemas. O primeiro onde não possuem nenhum tipo de tecnologia eficiente e moderna, utilizando papel e caneta para registrar a entrada e saída de veículos e o segundo no qual os estabelecimentos possuem sistemas de médio e alto custo, restritos a certos sistemas operacionais (Windows principalmente), que utilizam equipamentos caros e dependem totalmente de software e hardware (trazendo risco quanto à perda de dados e à falta de segurança).

Logo, a ideia do projeto foi desenvolver um sistema de controle de entrada e saída em linguagem C script CGI, utilizando basicamente 3 ferramentas “open source”: o servidor Apache, o ambiente Emacs e a interface de comunicação CGI.

É importante ser ressaltado a inexistência de projetos similares ao proposto, aumentando a originalidade da proposta. Talvez seja esta originalidade a responsável pela procura de diferentes empresas e instituições pelo projeto. Uma destas instituições interessadas foi a Polícia Militar do estado do Rio de Janeiro que, devido ao corte de verbas, estava em busca de um sistema para controle de entrada e saída de seus veículos que fosse eficaz, de baixo custo e que pudesse substituir o uso de papéis e planilhas. Vale lembrar que está em andamento o processo de registro do software em questão junto ao INPI (Instituto Nacional da Propriedade Industrial).

## **2. Contribuição do Trabalho**

### **2.1. Contribuição de cunho social e técnico**

Conforme foi relatado anteriormente, houve um crescimento excessivo da frota de veículos nas ruas. Este fato aliado à crise dos últimos anos gerou uma procura por estacionamento mais baratos. Do ponto de vista econômico, o Ctrl-car é de extrema relevância por proporcionar baixíssimo custo de uso.

Outro fato importante, mas de cunho mais técnico, é o software ter sido todo desenvolvido em ferramentas simples, gratuitas e de código aberto de programação, como Apache, CGI e Emacs. Isto permite adaptações para usos variados do sistema. Por esta mesma razão, o próprio programa é adaptável a várias plataformas, podendo ser utilizado em diversos sistemas operacionais. Ou seja, basta um dispositivo com acesso à Internet através de navegadores e um dispositivo com servidor Apache para instalação. Um estacionamento comercial de pequeno porte pode todo ser controlado através de um simples notebook, já um grande estacionamento pode ter um sistema complexo de computadores ligados à rede com um gerenciamento centralizado. Ou ainda, é possível fazer o controle sobre vários estacionamentos através de uma só central.

### **2.1. Benefícios da implementação do sistema**

O sistema proporciona benefícios de vários tipos. O primeiro se dá na operação que deixa de lado a impressão de “tickets” ou qualquer outro meio físico entregue ao cliente para registro de entrada e saída, visto que tudo é feito digitalmente, contribuindo até para a preservação do meio ambiente.

O segundo benefício é o da segurança. O registro do carro que entra no estacionamento usa parâmetros considerados essenciais para um controle eficaz. O principal dado de entrada é a placa do veículo, única para cada um e que foi usado como chave primária. Outro dado de entrada é o tipo do veículo, que permite fácil reconhecimento pelo funcionário. Também são dados de entrada o nome do dono do veículo, assim como um telefone de contato. Este controle é um diferencial em termos de segurança, pois após o cadastro na entrada, a verificação e reconhecimento do cliente por um funcionário na saída, evita roubos e acaba com o problema de perdas de “tickets” também.

Por fim, o último benefício é para o cliente que paga de forma justa o uso do estacionamento. O cálculo é feito por minuto que o cliente deixou o veículo estacionado e não por hora, como acontece nos vários estabelecimentos atualmente.

O Ctrl-car em seu atual estado já serviria como um produto útil, principalmente para seu grande foco que são os estacionamentos de pequeno porte, substituindo-se, por exemplo, o uso de papel e caneta para controle de entrada e saída.

### 3. Proposta do trabalho

A proposta consiste basicamente no uso do HTML para estabelecer uma comunicação entre endereços de IP e o script CGI, escrito em linguagem C e armazenado no Servidor. O usuário em um computador identificado como Cliente entra com os dados através do formulário da página em HTML que envia as informações para o script CGI na máquina do Servidor. O script CGI tem como função fazer toda a parte de processamento das informações e, para isto, são criadas estruturas para gravar os dados. A proposta é proporcionar velocidades rápidas de processamento e evitar falha de sistema ou travamento. Na seção a seguir é apresentado o funcionamento do Ctrl-car.

### 4. Análise de resultados

O Ctrl-car deve ser instalado em um Servidor para gerenciar o movimento do estabelecimento. Também o Apache deve estar instalado com os scripts CGIs e o HTML organizados em pastas. Especificamente os CGIs “entrada.cgi” e “saída.cgi” devem estar na pasta cgi-bin e o arquivo em HTML “estacionamento.html” na pasta htdocs.

Para realizar os testes e obter resultados da execução do Ctrl-car, foram usados dois PCs diferentes. Um funcionando como Servidor e o outro como Cliente, para registro do veículo. Na Figura 1, pode-se observar que os IPs são diferentes. Enquanto o IP do Servidor é 192.168.0.91 (Figura 1(A)), o do Cliente é 192.168.0.61 (Figura 1(B)). Inserindo-se o IP do Servidor na barra de HTTP (<http://xxx.xxx.x.xxx/estacionamento.html>), o Cliente poderá executar o HTML, acessando-se a página inicial do Ctrl-car.

Descrição	Qualcomm Atheros AR9285 Wireless	Descrição	Atheros AR5B97 Wireless Network Ac
Endereço Físico	E0-06-E6-FF-9E-39	Endereço Físico	94-39-E5-F4-DB-EF
DHCP Ativado	Sim	DHCP Ativado	Sim
Endereço IPv4	192.168.0.91	Endereço IPv4	192.168.0.61
Máscara de Sub-rede IP	255.255.255.0	Máscara de Sub-rede IP	255.255.255.0

(A) IP da máquina Servidor

(B) IP da máquina Cliente

Figura 1 – IPs das máquinas utilizadas na execução do Ctrl-car

Na Figura 2(A) se apresenta a tela de entrada de dados do motorista e do veículo através da máquina Cliente. A entrada de dados é finalizada apertando-se o botão “Dar entrada”. Com relação à operação de saída do veículo, insere-se somente o código da placa do veículo que sai do estacionamento e finaliza-se clicando o botão “Dar saída”, como mostra a Figura 2(B). No final da operação, uma tela confirma a saída, mostrando os horários de entrada e saída, os dados do motorista e o preço a ser pago.

Quando se realizam os registros de entrada e saída, automaticamente é criado no Servidor o arquivo chamado “entrada” no formato texto com os dados de entrada e outro arquivo chamado “saída” também no formato texto com os dados de saída.

## Cadastro do Veículo

Nome do Cliente: (80 caracteres max.)  
Marcus Teixeira  
Placa: (Ex: AAA-1234)  
XYZ-1212  
Veículo:  
Gol  
Telefone: (Ex: 12345678)  
88889999  
Dar Entrada

(A) Registro de entrada do veículo na máquina cliente

## Saída do Veículo

Placa:  
XYZ-1212  
Dar saída

Pesquisar na Web e no Win

(B) Registro de saída do veículo na máquina cliente

Figura 2 – Telas de registros de entrada e saída do veículo

A Figura 3(A) mostra o conteúdo do arquivo “entrada” e a Figura 3(B) o do arquivo “saída” para o teste em curso.

```
entrada - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
Rodrigo ARE=Apache/2.2.22 (win64) SERVER_NAME=loc
Marcus Teixeira /entrada.cgi REMOTE_PORT=56308 GA
```

(A) Conteúdo do arquivo de entrada

```
saida - Bloco de notas
Arquivo Editar Formatar Exibir Ajuda
Rodrigo AAA-2233 Sandero 13:02 13:10 R$0.40
Marcus Teixeira XYZ-1212 Gol 13:52 14:22 R$1.50
```

(B) Conteúdo do arquivo de saída

Figura 3 – Arquivos de registros de entrada e saída no Servidor

## 5. Conclusão

A implementação e execução do sistema atingiram os objetivos previstos no projeto.

Testes exaustivos foram realizados e o software teve desempenho conforme o previsto. Através da utilização do Ctrl-car, experimentos demonstraram as características de alta eficiência com baixo custo de desenvolvimento e manutenção do sistema que o diferencia dos sistemas existentes no mercado. Já sua viabilidade pôde ser constatada através da sua facilidade de uso, compatibilidade com diversos sistemas operacionais e oferecimento de segurança aos clientes do estacionamento.

A utilização das ferramentas *open source* (Apache, CGI, Emacs) mostrou ser possível o desenvolvimento de um sistema de controle de estacionamento de baixo custo e de alta eficiência. O Ctrl-car permitiu a comunicação entre duas máquinas (Servidor e Cliente) com IPs diferentes conectadas na Internet para uso profissional sem depender de software de alto custo e de código não aberto.

## Referências

Tabela da Anfavea de 2016 (2016). Acessado em 11 de maio de 2016, em: <http://www.anfavea.com.br/tabelasnov.html>

The Apache Software Foundation (2016). Acessado em 11 de maio de 2016, em: [www.apache.org](http://www.apache.org).

Common Gateway Interface (CGI) (2016). Acessado em 11 de maio de 2016, em: <http://www.edb.utexas.edu/minliu/multimedia/PDFfolder/CommonGatewayInterface%28CGI%29.pdf>.

A linguagem HTML (2016). Acessado em 11 de maio de 2016, em: <http://www.ufpa.br/dicas/htm/htm-intr.htm>.

V. V. Mizrahi Treinamento em Linguagem C, Pearson Prentice Hall, 2a Edição, 2008.

# Estratégias bioinspiradas aplicadas à robótica na vigilância

Claudiney R. Tinoco<sup>1</sup>, Danielli A. Lima<sup>1</sup>, Gina M. Barbosa<sup>1</sup>

<sup>1</sup>Faculdade de Computação – Universidade Federal de Uberlândia (UFU)  
Caixa Postal 593 – 38408-100 – Minas Gerais – MG – Brasil

claudineyrt@ufu.br, danielli@ufu.br, gina@ufu.br

**Resumo.** *Nesse artigo uma estratégia de coordenação distribuída, projetada com a modelagem de Autômatos Celulares e feromônio invertido, é descrita. Os robôs se comunicam indiretamente através do feromônio repulsivo, depositado e evaporado a cada passo de tempo. Todas as decisões de movimento são estocásticas, dando uma característica não determinística ao modelo. Através da aplicação do modelo proposto, resultados da simulação são apresentados.*

**Abstract.** *In this article a distributed coordination strategy, designed with the cellular automata modeling and repulsive pheromone is described. The robots communicate indirectly through the repulsive pheromone, deposited and evaporated at each time step. All next movement decisions are stochastic giving a non-deterministic characteristic to the model. Through the application of the proposed model, simulation results are presented.*

## 1. Introdução

Os Sistemas de Multiagentes focam-se na autonomia para diferentes tarefas. Cada agente deve ser capaz de alcançar autonomamente seus objetivos, se adaptando a mudanças no ambiente. Uma inteligência global é desejada, na qual pode ser gerada pelo comportamento individual de cada agente e nas interações entre os membros do time [1]. Sabe-se que um comportamento distribuído e coordenado pode emergir de interações individuais em sistemas complexos [2]. Nesse contexto, existem vários trabalhos que estão investigando a aplicação de modelos de Autômatos Celulares (CA) na robótica de enxames [3], [4], outros o uso do Sistema Invertido de Formigas [5]. No presente trabalho, um modelo de navegação, denominado *Inverted Ant Cellular Automata* (IACA), para a robótica de enxames aplicada na tarefa de vigilância, foi descrito. Através dos resultados foi possível observar que o comportamento autônomo do enxame permitiu a execução da tarefa de forma eficiente e o feromônio invertido proporcionou um caminho com um número menor de colisões.

## 2. Descrição do Modelo

Inicialmente, a representação bidimensional do ambiente é dividida em células quadradas, e duas representações bidimensionais do ambiente são criadas, resultando em uma estrutura de duas camadas. A primeira camada é a grade de feromônio, onde é mostrada a quantidade de feromônio depositada, Figura 1 (a). A segunda camada é a camada física, onde está representado a posição atual de cada agente robótico, paredes e obstáculos, Figura 1 (b). O modelo IACA é dividido em dois níveis: um é relacionado ao comportamento individual do robô e o outro é relacionado com o comportamento global do time.

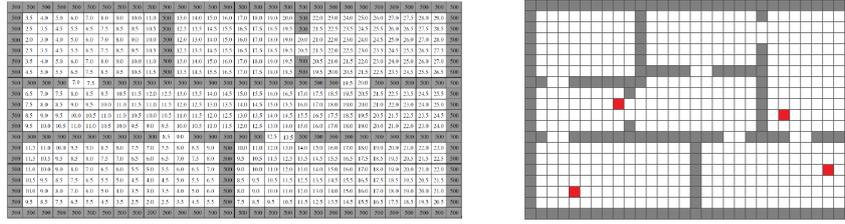


Figura 1. Grades 20×30. (a) Grade de feromônio. (b) Ambiente com 7 salas.

## 2.1. Comportamento Individual

O modelo de comportamento de cada robô pode ser descrito através de uma máquina de estado finito (FSM) com quatro estados, como mostrado na Figura 2.

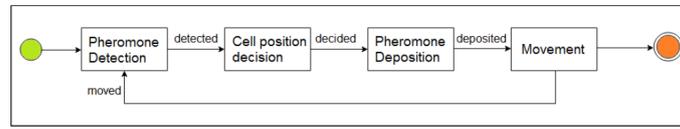


Figura 2. Comportamento individual em uma FSM adaptada de [5].

O estado de **detecção de feromônio** representa o processo de leitura do feromônio no ambiente. A leitura compreende os valores da vizinhança de Moore, definida como  $m = (2r_v + 1)^2$  no raio de visão  $r_v$  do robô e a quantidade atual de feromônio em cada célula  $x_{ij}$  é relativa aos depósitos de feromônio realizados anteriormente pelos robôs, seguidos pelo processo de evaporação. O **processo de decisão** é representado pela escolha da posição da próxima célula, a qual irá indicar o próximo movimento do robô. A quantidade de feromônio em uma determinada célula  $x_{ij}$  irá definir a probabilidade  $P(x_{ij})$  dessa célula ser escolhida no próximo passo de tempo  $t$ , de tal forma que quanto maior a quantidade de feromônio, menor é a probabilidade de uma célula ser escolhida. A quantidade de feromônio que é depositada em todas as células da vizinhança é dada por  $\rho_{max}^t = \sum_{k=1}^m x_{ij}^t$ .

$$P(x_{ij})^{t+1} = \frac{\rho_{max}^t - x_{ij}^t}{\sum_{k=1}^m \rho_{max}^t - x_{ij}^t} \quad (1)$$

Depois da escolha da próxima posição, cada robô deve depositar feromônio na célula atual e nas células correspondentes à sua vizinhança, para sinalizar a sua presença para os outros membros do enxame. O estado **processo de deposição** representa a quantidade de feromônio que cada robô é responsável por depositar na posição do robô e na vizinhança  $m$  correspondente de célula central  $x_{(i+a)(j+b)}$ , onde  $-r_v < a < 2 \cdot r_v$ . A Equação 2 representa a quantidade de feromônio depositada em cada célula a cada passo de tempo  $t$ . A constante  $\delta$  e  $\sigma$  representam, respectivamente, a taxa de feromônio e a taxa de dispersão. A atualização de feromônio é computada na Equação 3.

$$\rho_{ij}^{t+1} = \delta \cdot e^{\frac{x_{ij} - x_{(i+a)(j+b)}}{\sigma \cdot r_v^2}} \quad (2) \quad x_{ij}^{t+1} = x_{ij}^t + (\tau_{max} - x_{ij}^t) \cdot \rho_{ij}^{t+1} \quad (3)$$

O estado de **movimento** é o passo final que representa a transição do robô da célula  $x_{ij}$  para outra célula  $x_{(i+a)(j+b)}$  na vizinhança do robô. Essa ação será completada pelo controle individual do robô, o qual é responsável por decidir como controlar os componentes do robô para executar o movimento desejado.

## 2.2. Comportamento Global

O comportamento global compreende-se em dois processos: o primeiro é de evaporação e o segundo representa a interação entre os robôs e o ambiente. O **processo de evaporação** é representado pela Equação 4. Inicialmente, cada célula  $x_{ij}$  recebe no primeiro passo de tempo  $t = 0$  o valor 0 de feromônio ( $x_{ij} = 0$ ). Cada célula que é visitada por um robô é incrementada através de uma constante  $\delta$  (Equação 3) e cada célula que possui valores diferentes de zero é decrementada através de uma constante  $\beta$  (Equação 4). A utilização do feromônio invertido quase permite o tracejamento de uma rota livre de conflitos, mas casos específicos são resolvidos por um processo de **decisão de conflitos**. Através dessas interações entre os robôs e o ambiente, e entre os próprios robôs (feromônio invertido), surge um comportamento complexo que resolve a tarefa de vigilância.

$$x_{ij}^{t+1} = x_{ij}^t - (\beta \cdot x_{ij}^t) \quad (4)$$

## 3. Experimentos

Os experimentos apresentados nesta seção têm como objetivo: (i) avaliar a performance do time na cobertura do ambiente; (ii) avaliar a performance na tarefa de vigilância. As simulações utilizam os seguintes parâmetros: ambientes de tamanho  $(20 \times 30)$  células,  $T = 1000$  passos,  $\sigma = 0.43$ ,  $r_v = 1$ ,  $\delta = 0.7$ ,  $\tau_{max} = 50$  e  $N = 3$  robôs.

A primeira análise foi conduzida para verificar a cobertura do ambiente. O número de vezes que cada célula  $x_{ij}$  foi visitada durante os 1000 passos de tempo foi calculado em cada simulação, onde no gráfico azul escuro representa células menos visitadas, enquanto vermelho escuro representa aquelas que foram mais visitadas. A Figura 3 mostra que os robôs realizam uma boa cobertura do ambiente para quase todos valores de  $\beta$ . A simulação com  $\beta = 0.2$  retornou a pior performance, devido ao fato de uma sala estar sobrecarregada de visitas. Nas simulações com  $\beta = 0.05$  e  $\beta = 0.1$ , existem muitos pontos azuis dentro de algumas salas específicas, porém no geral estão homogêneas. Portanto, a taxa de evaporação  $\beta = 0.05$  e  $\beta = 0.1$  conduz o time para o melhor comportamento.

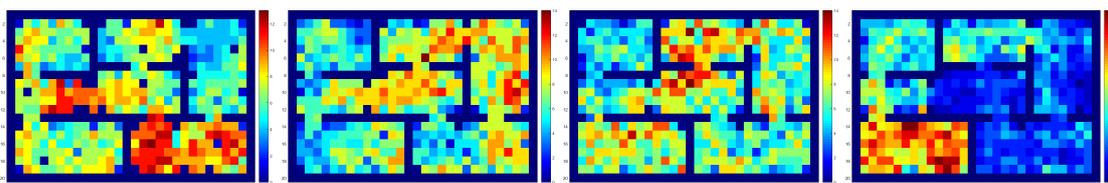


Figura 3. Mapas de passos:  $\beta = \{(a) 0.01, (b) 0.05, (c) 0.1, e (d) 0.2\}$

O segundo experimento foi realizado para analisar da performance do time, registrando o número de visitas que cada sala recebeu, Figura 4. Os gráficos foram elaborados registrando a posição atual de cada robô em cada passo de tempo. O eixo  $y$  representa as salas e o eixo  $x$  representa as iterações. Cada linha vertical vermelha indica que um ciclo de vigilância foi concluído, isto é, a iteração que o time de robôs controlado pelo algoritmo IACA foi capaz de visitar todo o ambiente. Os gráficos mostram que o sistema IACA com  $\beta = 0.001$  é capaz de concluir o primeiro ciclo da tarefa de vigilância na iteração 219, já com  $\beta = 0.01$  e  $\beta = 0.1$ , de forma mais eficiente, concluiu o primeiro ciclo nas iterações 145 e 152, respectivamente. Além disso, o sistema IACA com  $\beta = 0.001$  conclui 5 ciclos da tarefa de vigilância, utilizando  $\beta = 0.01$  e  $\beta = 0.1$ , o

número de ciclos concluídos é aumentado para 7 e 6, respectivamente. Considerando-se 3 robôs e os  $\beta$  valores avaliados, a melhor performance foi alcançada utilizando  $\beta = 0.01$ .

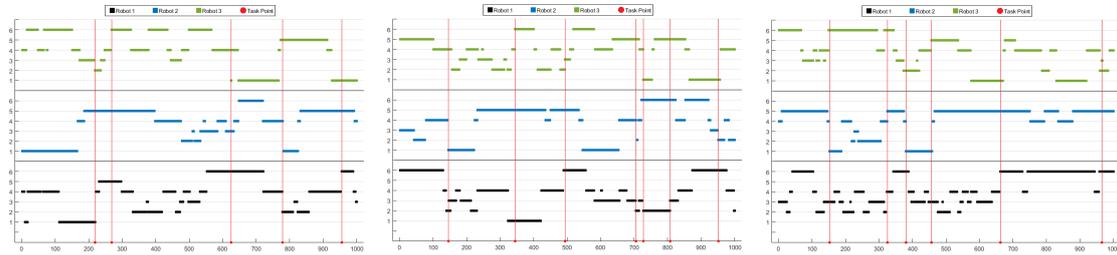


Figura 4. Pontos de tarefa: (a)  $\beta = 0.001$ . (b)  $\beta = 0.01$ . (c)  $\beta = 0.1$ .

#### 4. Conclusão

Este trabalho investigou um novo modelo de controle para robótica de enxame, o IACA. O modelo é dedicado à tarefa vigilância, o qual é muito relevante para a robótica coletiva [6], (i) é incluído em amplas classes de problemas de exploração integrada, navegação e cobertura de ambiente; (ii) a vigilância é o problema amplamente investigado nos estudos de cooperação de robô para robô; (iii) a robótica possui muitas aplicações no mundo real, busca e resgate, limpeza e exploração planetária. Com as simulações foi possível avaliar o modelo IACA em termos de performance de cobertura, análise de exploração e a melhor constante de evaporação. Um futuro trabalho é a implementação desse modelo utilizando-se plataformas de arquitetura robótica do mundo real.

#### Referências

- [1] Mustafa Servet Kıran, Mesut Gündüz, and Ömer Kaan Baykan. A novel hybrid algorithm based on particle swarm and ant colony optimization for finding the global minimum. *Applied Mathematics and Computation*, 219(4):1515–1521, 2012.
- [2] Melanie Mitchell, James P Crutchfield, Rajarshi Das, et al. Evolving cellular automata with genetic algorithms: A review of recent work. In *Proceedings of the First International Conference on Evolutionary Computation and Its Applications (EvCA'96)*. Moscow, 1996.
- [3] Giordano BS Ferreira, Patricia A Vargas, and Gina MB Oliveira. An improved cellular automata-based model for robot path-planning. In *Advances in Autonomous Robotics Systems*, pages 25–36. Springer, 2014.
- [4] C Behring, M Bracho, M Castro, and JA Moreno. An algorithm for robot path planning with cellular automata. In *Theory and practical issues on cellular automata*, pages 11–19. Springer, 2001.
- [5] Rodrigo Calvo, Janderdeson Rodrigo de Oliveira, Mauricio Figueiredo, and Roseli Ap Francelin Romero. Bio-inspired coordination of multiple robots systems and stigmergy mechanisms to cooperative exploration and surveillance tasks. In *Cybernetics and Intelligent Systems (CIS), 2011 IEEE 5th International Conference on*, pages 223–228. IEEE, 2011.
- [6] Evandro Luis S Falleiros, Rodrigo Calvo, and Renato P Ishii. Pheroslam: A collaborative and bioinspired multi-agent system based on monocular vision. In *Computational Science and Its Applications—ICCSA 2015*, pages 71–85. Springer, 2015.

# Estudo e Aplicação de Técnicas de Registro de Imagens para Reconstrução 3D da Rede de Canais Vasculares do Tecido Ósseo

**Autor: Thiago Togo Nunes**

**Orientador: Bruno Augusto Nassif Travençolo**

Programa de Iniciação Científica

Sistemas de Informação

Universidade Federal de Uberlândia – Uberlândia, MG – Brasil

[thiagotogo@gmail.com](mailto:thiagotogo@gmail.com), [travencolo@gmail.com](mailto:travencolo@gmail.com)

Nível: Iniciação Científica

Bolsa de Estudo: CNPq

**Resumo.** *Recentemente a Faculdade de Odontologia da Universidade Federal de Uberlândia adquiriu um scanner de alta resolução, que permite capturar simultaneamente imagens de diversas lâminas histológicas em alta resolução. Com isso, o número de imagens capturadas é bem grande e muito rápido quando comparado com o processo de captura em microscópios tradicionais. Por outro lado, um importante gargalo precisa ser vencido. As imagens obtidas por microtomia não preservam o alinhamento entre fatias consecutivas. A forma até o momento para fazer esse alinhamento é manualmente, com o auxílio de softwares de edição de imagens. Entretanto, essa é uma tarefa difícil, pois as deformações encontradas não são apenas deslocamento e rotações entre fatias (conhecidas como transformações rígidas), mas também encontra-se deformações no contorno e dobras, que dificultam o alinhamento. O processo de alinhamento das fatias é fundamental para as análises posteriores, a saber a reconstrução 3D do volume do osso, para posterior esqueletização e caracterização da rede vascular por meio das propriedades de redes complexas.*

**Palavras-Chave:** *Alinhamento de Imagens; Registro de Imagens*

## 1. Introdução, Motivação e Fundamentação Teórica

Na biologia, muitas descobertas científicas são realizadas por meio da observação de imagens utilizando microscópios. Tradicionalmente a análise é feita a partir de imagens bidimensionais (2D), mas o recente avanço tecnológico nas últimas décadas abriu a possibilidade de capturar e analisar, com auxílio de um computador, imagens em três dimensões (3D).

Uma das formas de obtenção de imagens 3D é ilustrada na Figura 1. Esse processo é feito utilizando um aparelho chamado micrótomo, que permite separar objetos em cortes finos. Em seguida, esses cortes são fixados em lâminas de vidro, e é importante que essas lâminas mantenham a ordem em que as fatias foram cortadas, para

que a reconstrução tenha coerência. As lâminas passam por um processo de coloração utilizando marcadores químicos e são digitalizadas, passando assim as mesmas para o computador.

Entretanto, as imagens digitalizadas não estão alinhadas entre si, isto é, durante todo processo de captura de imagens 3D é impossível garantir que uma fatia da amostra em uma lâmina seja colocada exatamente na mesma posição que a fatia da amostra em uma lâmina seja colocada exatamente na mesma posição que a fatia cortada anteriormente. Para solucionar esse problema, pode-se utilizar técnicas computacionais conhecidas como registro de imagens, em que são feitas transformações nas imagens para que elas se alinhem. Após o registro (alinhamento) de todas as imagens é que temos a imagem em 3D.

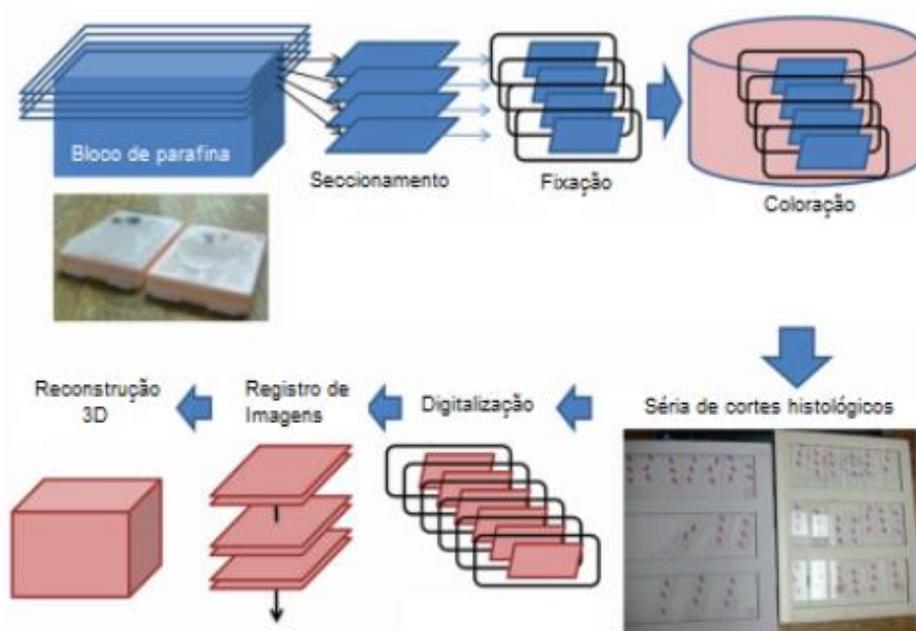


Figura 1 – Método de obtenção das imagens 3D, com destaque a parte de registro de imagens. Figura obtida de <<http://www.nature.com/articles/srep06050>>

Registro de imagens é o processo de transformação de diferentes conjuntos de dados em um sistema de coordenadas. Quando se trata de dados biológicos o registro de imagens se torna um desafio, porque problemas como deformações nas imagens são comuns. Esses problemas podem ser causados no processo de preparação das imagens, envolvendo variações de manchas, posição das imagens, falta de dados, fazendo com que os métodos atuais de registro de imagens não obtenham sua máxima performance.

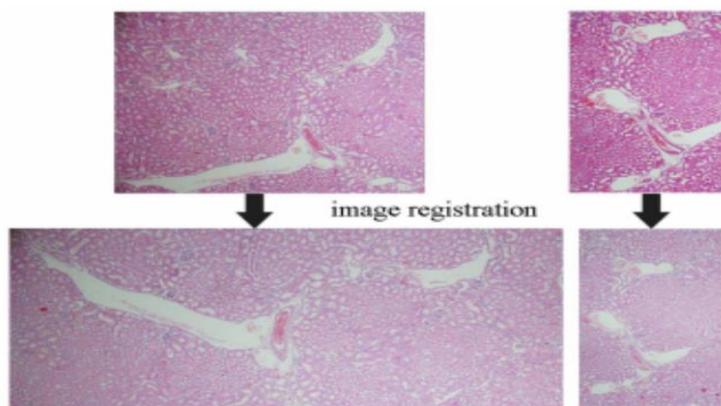


Figura 2 – Exemplo de dificuldade do registro de imagem. No conjunto de imagens à esquerda existem deformações morfológicas entre as imagens. Já no conjunto da direita existe uma forte variação na coloração das imagens. Figura obtida de <<http://www.nature.com/articles/srep06050>>

Em uma recente publicação, (WANG, KA, & CHEN, 2014), apresentam uma técnica robusta para o problema de registro de imagens. Os autores compararam a técnica proposta com seis outros métodos que são o estado da arte na área de registro *TrakEM2*, *SURF + affine transformation*, *UnwarpJ*, *bUnwarpJ* e *BrainAligner*. Na proposta dos autores foi obtida uma acurácia média de 91.49%, enquanto que as outras propostas a acurácia não passou de 70%.

Além da proposta de um novo método, Ching-Wei Wang, Shuk-Man Ka & Ann Chen, disponibilizaram um software para a realização do registro, na forma de um plugin para o software ImageJ. Entretanto, para a execução do alinhamento das imagens é necessário abrir uma imagem de cada vez no software, dificultando a automatização do processo quando o número de imagens é grande.

## 2. Contribuição do Trabalho

Este trabalho tem como objetivo desenvolver formas de automatizar o processo de registro de imagens obtidas pelo processo de microtomia. Como objetivos específicos pode-se listar:

1. Registro das imagens coletadas pela Faculdade de Odontologia
2. Estudo das técnicas de Processamento Digital de Imagens utilizadas na técnica de registro propostas por (Ching-Wei Wang, Shuk-Man Ka & Ann Chen)
3. Estudo e alteração do código disponibilizado por (Ching-Wei Wang, Shuk-Man Ka & Ann Chen) para possibilitar o alinhamento de várias imagens simultaneamente.

### 3. Análise de Resultados

Até o momento, foi estudado durante a pesquisa os conceitos principais de registro de imagens, métodos de alinhamento de imagens como por exemplo os métodos SIFT (LOWE, D. G., 2004) e SURF (BAY, H. et al., 2008).

As principais etapas envolvidas nesses algoritmos estão sendo estudadas, como por exemplo *Scale-space*, *Extrema Detection*, Localização dos pontos de interesse, Definição da orientação e descrição de um ponto de interesse. Está sendo mantido o foco na metodologia desses processos no início do projeto para que se entenda implementação do *plugin* disponibilizado por Ching-Wei Wang, Shuk-Man Ka & Ann Chen.

### 4. Trabalhos Relacionados

Serão usadas as seguintes técnicas de Processamento de Imagens: Normalização dos dados, técnicas tradicionais de registro, técnica SIFT (LOWE, D. G., 2004) e técnica SURF (BAY, H. et al., 2008).

### Referências

BAY, H. et al. Speeded-up robust features (surf). *Comput. Vis. ImageUnderst.*, Elsevier Science Inc., New York, NY, USA, v. 110, n. 3, p. 346–359, jun. 2008. ISSN 1077-3142.

LOWE, D. G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, Kluwer Academic Publishers, Hingham, MA, USA, v. 60, n. 2, p. 91–110, nov. 2004. ISSN 0920-5691.

PENG, H. et al. Brainaligner: 3d registration atlases of drosophila brains. *Nat methods*, v. 8, n. 6, p. 493–500, Jun 2011.

WANG, C.-W.; KA, S.-M.; CHEN, A. Robust image registration of biological microscopic images. *Sci Rep*, v. 4, p. 6050, 2014. Disponível em: <<http://dx.doi.org/10.1038/srep06050>>.

# **Estudo sobre a adoção de PDTI e CGTI pelas Universidades Federais**

**Autor: Gustavo Caetano Borges; Romário Vieira**

**Orientador: Rodrigo Sanches Miani; João Batista Simão**

Graduação em Sistemas de Informação

Universidade Federal de Uberlândia – Monte Carmelo, MG – Brasil

[gustavoborges@ufu.br](mailto:gustavoborges@ufu.br), [romario.sistemas@gmail.com](mailto:romario.sistemas@gmail.com), [miani@ufu.br](mailto:miani@ufu.br),  
[jbatistasimao@gmail.com](mailto:jbatistasimao@gmail.com)

Nível: Iniciação Científica

Bolsa de Estudo: CNPq

**Resumo.** A governança de TIC (Tecnologias da Informação e Comunicação) é essencial para empresas de todos os tipos e segmentos, visto que a mesma auxilia no modo como a TIC é aplicada a empresa. O presente trabalho busca compreender como instituições públicas, em particular as universidades federais, gerenciam seus planos diretores de TIC (PDTI). Com o auxílio de pesquisas usando o portal Google, foi possível traçar um mapa sobre a adoção e extensão de PDTI nas Universidades Federais brasileiras.

**Palavras-Chave:** PDTI; PETI; Comitê Gestor de TI; Universidades Federais.

## **1. Introdução, Motivação e Fundamentação Teórica**

A Governança de tecnologia da informação e comunicação (GTIC) é o sistema pelo qual as empresas e demais organizações são dirigidas, monitoradas e incentivadas, envolvendo os relacionamentos entre sócios, conselho de administração, diretoria, órgãos de fiscalização e controle e demais partes interessadas. A GTIC também busca o direcionamento das tecnologias da informação e comunicação (TIC) para atender ao negócio e o monitoramento e verificar a conformidade com o direcionamento tomado pela administração da organização Fernandes e Abreu (2012).

Segundo Lunardi (2014), ainda que um bom número de empresas venha percebendo as tecnologias da informação e comunicação (TIC) como um de seus principais ativos, as decisões sobre sua adoção, implantação e gerenciamento continuam sendo bastante complexas, o que tem feito com que muito dinheiro seja desperdiçado em aquisições tecnológicas mal realizadas.

Com objetivo de orientar e melhorar a gestão de sua robusta e complexa estrutura tecnológica, foi criada pelo governo federal brasileiro por meio de decreto nº1.048 de 21 de janeiro de 1994 e posteriormente atualizado pelo decreto nº 7.579 de 11 de outubro de 2011, o Sistema de Administração dos Recursos de Tecnologia da Informação (SISP). Além do SISP, o governo federal brasileiro instituiu a IN04/2014, que dispõe sobre o processo de contratação de soluções de TI pelos órgãos integrantes SISP. Juntamente com a IN04/2014 foram liberados diversos manuais para auxiliar a

implantação do Comitê Gestor de TI (CGTI). Em outras palavras, o SISP, a IN04/2014 e o CGTI são tentativas do governo federal brasileiro em incentivar a aplicação de uma governança de TI mais eficaz dentro de suas unidades.

Apesar da complexidade do cenário acadêmico, trabalhos recentes de Peterson (2004) e Batista (2006), mostram que algumas universidades federais brasileiras vislumbram a adoção da GTIC para melhor utilização de recursos fazendo uso principalmente de um Plano Diretor de TI (PDTI), que consiste em um documento que possibilita nortear e acompanhar a atuação do órgão da área de TI, definindo estratégias, bem como, o planejamento de ações para implantá-las. Tomando como parâmetro o documento de PDTI de uma organização, é possível identificar as estratégias de GTIC empregadas por ela. Características presentes em um PDTI como: vigência do PDTI, redação do PDTI usando como base o Plano estratégico de TI (PETI) (BARROS, 2014), quantidade de PDTI disponíveis, existência de um CGTI no escopo e existência de um portal do CGTI formam um conjunto de indicadores de qualidade para avaliar a adoção de GTIC nas organizações.

O presente trabalho, por meio de pesquisa em um buscador Web, se propôs a apresentar a situação das universidades federais brasileiras perante a utilização de práticas e como é a padronização entre as mesmas de acordo com o proposto pelo governo federal através de IN e decretos. O levantamento consta com informações ligadas ao PDTI e CGTI.

## **2. Contribuição do Trabalho**

Este trabalho contribui com a análise da atual situação da GTIC nas universidades federais, podendo assim atribuir influência da implantação de GTIC sobre aspectos a segurança da informação da instituição ou a qualidade ou reputação da instituição.

## **3. Análise de Resultados**

O presente trabalho se utiliza de certos pontos na governança de TIC das universidades federais brasileiras, procurando padrões entre os PDTI, como: período de vigência, quantidade de versões, documento modelo (PETI) tal documento apresenta características ou informações relevantes ao desenvolvimento do documento final, e também no CGTI, como existência e site próprio sobre informações relacionadas ao mesmo. De acordo com Barros (2014), o PDTI de uma universidade federal deve ter como um dos documentos de referência o PETI, mostrando assim a necessidade deste documento.

Com o intuito de identificar os documentos ligados a GTIC da universidade, foram feitas buscas no Google com os termos “PDTI”, “CGTI”, “PETI”, e “Nome Universidade”. Nesse trabalho foram consideradas as 63 Universidades Federais brasileiras. Os Institutos Federais de Educação não fizeram parte dessa pesquisa. A Tabela 1 apresenta os resultados das buscas. Pode-se notar que das 63 universidades federais, 41 (65%) possuem PDTI, 39 (62%) possuem CGTI, 12 (19%) possuem portal do CGTI, 6 (9,6%) possuem o PETI, 34 (54%) possuem PDTI e CGTI.

**Tabela 1: Informações sobre as Universidades**

Universidades	Período de Vigência (Anos)	PDTI	Qtd. Documentos	PETI	CGTI	Possui site CGTI?
FURG	2	Sim	2	Sim	Não	Não
UFABC	2	Sim	1	Não	Sim	Não
UFAC	2	Sim	1	Não	Sim	Sim
UFAL	3	Sim	1	Não	Sim	Não
UFAM	0	Não	0	Não	Não	Não
UFBA	3	Sim	1	Não	Sim	Não
UFC	5	Sim	1	Não	Não	Não
UFCA	2	Sim	1	Não	Sim	Sim
UFCG	0	Não	0	Não	Não	Não
UFCSPA	4	Sim	1	Não	Sim	Não
UFERSA	5	Sim	2	Não	Sim	Sim
UFES	6	Sim	1	Não	Não	Não
UFF	2	Sim	2	Não	Sim	Sim
UFFS	2	Sim	1	Não	Sim	Sim
UFG	2	Sim	2	Não	Sim	Não
UFGD	4	Sim	1	Não	Sim	Não
UFJF	3	Sim	1	Não	Não	Não
UFLA	2	Sim	1	Não	Sim	Não
UFMA	3	Sim	1	Não	Sim	Não
UFMG	5	Sim	1	Não	Sim	Não
UFMS	4	Não	1	Não	Sim	Não
UFMT	0	Não	0	Não	Não	Não
UFOB	0	Não	0	Não	Não	Não
UFOP	0	Não	0	Não	Não	Não
UFOPA	0	Não	0	Não	Sim	Não
UFPA	0	Não	0	Não	Não	Não
UFPB	0	Não	0	Não	Sim	Sim
UFPE	2	Sim	2	Não	Sim	Não
UFPEL	2	Sim	2	Não	Sim	Sim
UFPI	4	Sim	1	Não	Sim	Não
UFPR	4	Sim	1	Não	Sim	Não
UFRA	0	Não	0	Não	Não	Não
UFRB	0	Não	0	Não	Sim	Não
UFRGS	5	Sim	1	Não	Sim	Não
UFRJ	0	Não	0	Não	Não	Não
UFRN	2	Sim	1	Não	Sim	Não
UFRPE	1	Sim	2	Não	Sim	Não
UFRR	0	Não	0	Não	Não	Não
UFRRJ	0	Não	1	Não	Não	Não
UFS	2	Sim	1	Sim	Sim	Não
UFSB	0	Não	0	Não	Não	Não
UFSC	0	Não	0	Não	Não	Não
UFSCAR	2	Sim	2	Não	Não	Não
UFSJ	0	Não	0	Não	Não	Não
UFSM	2	Sim	1	Não	Sim	Não
UFT	1	Sim	2	Sim	Sim	Não
UFTM	0	Não	0	Não	Sim	Não
UFU	2	Sim	3	Não	Sim	Sim
UFV	4	Sim	1	Não	Não	Não
UFVJM	0	Não	0	Não	Não	Não
UNB	4	Sim	2	Não	Sim	Não
UNIFAL	0	Não	0	Não	Sim	Não
UNIFAP	5	Sim	2	Sim	Sim	Não
UNIFEI	3	Sim	2	Não	Sim	Sim
UNIFESP	2	Sim	1	Sim	Sim	Não
UNIFESSPA	0	Não	0	Não	Não	Não
UNILA	0	Não	0	Não	Não	Não
UNILAB	2	Sim	1	Não	Sim	Não
UNIPAMPA	1	Sim	1	Não	Não	Não
UNIR	2	Sim	3	Não	Sim	Sim
UNIRIO	2	Sim	1	Não	Sim	Sim
UNIVASF	0	Não	0	Não	Não	Não
UTFPR	4	Sim	2	Sim	Sim	Sim

Pode-se notar que 19 PDTIs possuem vigência de 2 anos. De acordo com o Sistema de Administração dos Recursos de Tecnologia da Informação (SISP), esse é o período mínimo que deve ser estabelecido por uma organização, além das revisões frequentes. Além disso, a variação do período de vigência (entre 1 a 6 anos) mostra que não existe um padrão estabelecido nesse quesito, o que dificulta a troca de experiência entre as universidades e o acompanhamento do governo federal. Também é possível notar que

27 universidades federais possuem apenas uma versão de PDTI, enquanto apenas duas possuem 3 PDTIs (UNIR e UFU), podendo estas serem consideradas mais adaptadas aos padrões estabelecidos pelo SISP.

Outro resultado relevante diz respeito a transparência da divulgação das informações. Mais de 80% das universidades não possuem uma página específica para divulgação do PDTI e das ações relacionadas a GTIC. Levando em consideração somente a presença do documento PDTI e do CGTI, têm-se que mais de 70% (46) das instituições estudadas estão caminhando para implantar as práticas impostas pelo governo federal. Trabalhos futuros incluem estudos sobre a avaliação da qualidade dos PDTIs e do impacto de um plano diretor em outras esferas da universidade, como por exemplo a excelência acadêmica, incidentes de segurança e gestão dos recursos.

#### **4.Trabalhos Relacionados**

Albuquerque et al. (2011) avaliam por meio de um *survey* a situação da governança de TIC em uma instituição de pesquisa pública, a Fundação Oswaldo Cruz-FIOCRUZ, na qual, o autor compara a diferença da governança de TIC nas diferentes sedes da instituição. Os resultados encontrados mostram que a governança de TIC quando é feita de forma descentralizada e pode apresentar diferenças entre visão/funcionamento do gestor de uma sede para outra. Fernandes et al. (2012) apresentam, por meio de acompanhamento da UFSM, como aconteceu a maturidade da instituição perante a GTIC, demonstrando quais eram seus pontos fracos como Planejamento e Monitoramento, assim como seu crescimento depois da aplicação de um PDTI e um CGTI sólidos.

#### **Referências**

- ALBUQUERQUE JUNIOR, A. E. de ; Machado, K. C. B.; SANTOS, Ernani M. (2011). “Estudo sobre a Percepção da Importância de Indicadores de Governança de TI em uma Instituição de Pesquisa.” In: *XXXI Encontro Nacional de Engenharia de Produção - ENEGEP 2011*, Belo Horizonte.
- BARROS, Fábio Gomes. *Elaboração do Plano Diretor de Tecnologia da Informação (PDTI)*. 2014.
- BATISTA, Fábio Ferreira. *O desafio da gestão do conhecimento nas áreas de administração e planejamento das Instituições Federais de Ensino Superior (Ifes)*. 2006.
- FERNANDES, Aguinado Aragão; ABREU Vladimir Ferraz de. *Implantando a Governança de TI: da Estratégia à Gestão dos Processos e Serviços*. 3. Ed. São Paulo: Brasport, 2012.
- LUNARDI, G.; DOLCI, J; BECKER Et Al. Análise dos mecanismos de governança de TI mais difundidos entre as empresas brasileiras. *Revista ALCANCE Eletrônica*, v.21, n. 01, p. 46-76, Jan./Mar., 2014;
- PETERSON, R. Integration strategies and tactics for information technology governance. In: *Strategies for information technology governance*, Hershey: Idea group publishing, 2004.

# Mapeamento de ambiente para correção de trajetória de robôs.

**Autor: Igor Gomes Arantes**

**Orientador: Mauricio Cunha Escarpinati**

Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal de Uberlândia – Uberlândia, MG – Brasil

igorgarantes@gmail.com, escarpinati@gmail.com

Iniciação Científica

Bolsa de Estudo: Não há

**Resumo.** *Na área da robótica, muitas vezes, existe uma necessidade do robô se localizar no ambiente em que está. Uma das formas de se fazer isso é através de técnicas baseadas em odometria, que calcula a posição atual do robô através de medições dos deslocamentos incrementais de suas rodas. O problema desta técnica é que, inevitavelmente, existe um acúmulo de erros na estimativa da posição do robô com o passar do tempo. Mas, apesar do erro acumulativo, tem baixo custo computacional e por isso é amplamente utilizada. Em contrapartida existem outras técnicas, mais precisas, que também conseguem estimar a posição atual do robô. Um exemplo eficaz é por meio de técnicas de visão computacional. Com o uso de uma câmera externa, é possível capturar imagens do ambiente e a partir delas fazer um mapeamento do cenário, e assim, estimar de maneira precisa onde estão os robôs. Apesar dessa maior precisão o custo computacional desta técnica é alto, e isso inviabiliza o seu uso. Uma possível solução neste contexto é a utilização de um sistema híbrido, que utiliza técnicas baseadas em odometria em conjunto com técnicas de visão computacional, de forma que, os erros acumulativos gerados pela primeira fossem corrigidos pela segunda, conferindo assim uma boa precisão e um bom desempenho ao sistema.*

**Palavras-chave:** *Robótica, Visão computacional, Localização*

## 1. Introdução, Motivação e Fundamentação Teórica

A odometria é um método muito utilizado para estimar a posição de um robô em movimento (Carvalho Filho, José Gilmar Nunes de, et al (2009)). Essa técnica consiste em analisar medições dos deslocamentos incrementais de suas rodas, e assim identificar a posição atual do robô em relação à sua posição original. Trata-se de uma técnica de baixo custo computacional, com boa precisão a curto prazo, mas fica menos confiável à medida que a distância percorrida cresce. Este problema ocorre devido a um inevitável acúmulo de erros inerentes a essa técnica.

Uma das maneiras de corrigir esse erro é a utilização de um sistema híbrido, ou seja, fazer a utilização desta técnica em conjunto com outra na intenção de aumentar a precisão de suas estimativas. Uma das soluções viáveis de ser utilizada é o emprego de técnicas de visão computacional, onde a partir da captura de imagens, se analisa o ambiente a fim de identificar as posições dos robôs de maneira precisa. O problema desta técnica é que o processamento das imagens tem um custo alto, o que torna

interessante seu uso em conjunto com a odometria, onde uma técnica identifica de maneira rápida a posição do robô e seu erro acumulativo é corrigido periodicamente pela outra, dando ao sistema uma boa precisão e um bom desempenho.

Com a intenção de contribuir com o desenvolvimento de métodos de correção de trajetórias de robôs este trabalho tem o objetivo de, através de técnicas de visão computacional, capturar, manipular e analisar imagens do ambiente proposto e a partir de dessas calcular de maneira precisa e robusta as posições dos robôs envolvidos na cena.

## **2. Objetivo**

### **Objetivos Gerais:**

O objetivo deste trabalho é estudar e implementar um método de correção de trajetórias de robôs através de técnicas de visão computacional.

### **Objetivos Específicos:**

Este trabalho tem como objetivo criar um programa que, através de imagens adquiridas do ambiente proposto, utiliza técnicas de visão computacional em conjunto com técnicas baseadas em odometria para calcular de forma precisa e otimizada a posição de um robô na cena.

## **3. Metodologia**

### **Aquisição das imagens do ambiente:**

Uma parte importante do trabalho é a captura das imagens que servirão de entrada para os algoritmos de análise de imagens. É interessante adquirir imagens de diferentes ângulos e tamanhos a fim de comprovar a robustez e a eficácia do projeto.

### **Implementação de métodos de pré-processamento:**

A primeira parte da implementação é o pré-processamento das imagens, que através de técnicas de processamento digital de imagens será identificado a área na cena onde os robôs e os obstáculos estão posicionados e serão feitos ajustes, preparando-as para as próximas etapas. Este primeiro momento é uma das partes mais importantes do trabalho, pois todas as próximas serão dependentes desta.

## **Implementação de métodos de identificação de objetos**

Tendo em mãos a imagem pré-processada partiremos para um trabalho de identificação dos objetos que estão na cena. Para isso, uma das alternativas, é a utilização de redes neurais artificiais (que é uma técnica que utiliza o aprendizado de máquina) para distinguir os padrões que compõem os robôs e os demais objetos, reconhecendo-os na imagem.

## **Implementação de métodos de identificação da rotação dos robôs**

Identificados os objetos e suas posições poderemos partir para a identificação da angulação da rotação de cada robô. Uma alternativa para alcançar esse objetivo é a utilização da técnica de visão computacional de fluxo óptico (que é uma técnica que descreve os deslocamentos ocorridos entre dois quadros consecutivos de um vídeo) para identificar, através do movimento dos robôs, para onde estão indo e seus ângulos de rotação.

## **Correção da trajetória dos robôs**

Em um último momento teremos um mapeamento da cena, identificando as posições dos robôs, assim como seus ângulos de rotação em relação ao ambiente, as posições dos objetos e a diferenciação de cada robô, caso haja mais do que um. Com essas informações poderemos fazer a comunicação com os robôs corrigindo, periodicamente, as estimativas das posições que a odometria nos fornece, que é o objetivo final do trabalho.

## **4. Resultados parciais**

O projeto se encontra em fase de implementação. Até o momento o trabalho tem sido focado na segmentação das imagens e na escolha de melhores métodos se trabalhar com as imagens dadas. Ao fim do projeto, espera-se como resultado uma implementação eficiente de um sistema híbrido de mapeamento de ambiente a fim de auxiliar robôs que usam método de localização a terem uma maior precisão no contexto geral de seus experimentos.

## **5. Conclusões**

Tendo em vista o que foi feito até o momento, o desenvolvimento deste trabalho pretende contribuir com resultados mais precisos em pesquisas que se utilizam de experimentos com robôs em ambiente real que tem a necessidade de se localizar sem a utilização de sensores próprios.

## Referências

- Borenstein, J., Everett, H. R. and Feng, L. (1996). “where am i?” sensors and methods for mobile robot positioning, University of Michigan.
- Duda, R. and Hart, P. (1972). Use of the hough transformation to detect lines and curves in pictures, *Communications of the ACM* 15(1): 11–15.
- Gonzalez, R. C. and Woods, R. E. (2000). *Processamento de Imagens Digitais*, Editora Edgar Bluc her.
- Jain, A. K. (1989). *Fundamentals of Digital Image Processing*, Prentice Hall.
- Machler, P. (1998). *Robot Positioning by Supervised and Unsupervised Odometry Correction*, PhD thesis, Ecole Polytechnique Fédérale de Lausanne.
- Valgas, J. H. (2002). *Uma metodologia de correção dinâmica de erros de odometria em robôs móveis*, Master’s thesis, Universidade Federal de Minas Gerais.
- Carvalho Filho, José Gilmar Nunes de, et al. "Estimação de posição e orientação para robôs móveis." (2009).

# Agrupamento de Dados via Combinação de Partições Utilizando Amostragem de Dados

**Autor: Paulo Henrique Maia Soares<sup>1</sup>,  
Orientador: Daniel Duarte Abdala<sup>1</sup>**

<sup>1</sup>Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

phmaias@gmail.com, abdala@ufu.br

Nível: Trabalho de Conclusão de Curso

**Resumo.** *Este trabalho aborda o problema de agrupamento de dados via combinação de partições utilizando como dados de entrada uma amostra das partições disponíveis. Tal abordagem diminui a complexidade computacional do problema habilitando sua utilização em bases de dados maiores que as atualmente elegíveis. Uma revisão extensiva dos métodos de amostragem adotados em agrupamento de dados é apresentada e os resultados experimentais acerca do impacto da amostragem no resultado final consensual são apresentados e discutidos.*

**Palavras-Chave.** *Agrupamento de Dados, Amostragem de dados*

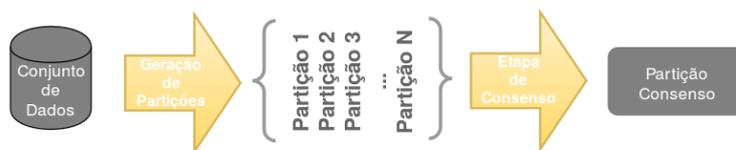
## 1. Introdução, Motivação e Fundamentação Teórica

Agrupamento de dados via combinação de partições (ADCP) tem se apresentado como uma alternativa viável para abordar problemas de agrupamento de dados onde pouco ou nada se sabe acerca da distribuição estatística dos dados no espaço característico [Abdala 2010]. Esta abordagem é interessante, pois consegue ser mais preciso que os algoritmos de agrupamento individuais, chegando a resultados inalcançáveis por um algoritmo de agrupamento tradicional [VEGA-PONS 2010].

O processo de agrupamento de dados, muitas vezes é o primeiro passo da análise de dados [Nguyen 2007]. Existem diversos métodos de agrupamento já desenvolvidos, tais como métodos hierárquicos, métodos baseados em densidade, em grid, em particionamento, dentre outros. A maior parte dos métodos de agrupamento concentra-se em encontrar grupos próximos do ideal, baseado em algum critério de agrupamento específico [Nguyen 2007]. Já o método ADCP visa combinar vários resultados de diferentes algoritmos de agrupamento em uma solução de consenso, pois dada uma base de dados onde não se conhece a priori a distribuição de dados no espaço característico, o algoritmo mais adequado a ser utilizado é igualmente desconhecido. ADCP proporciona uma melhor qualidade e robustez dos resultados em tais casos em comparação com a utilização de um algoritmo de agrupamento qualquer.

Muitos autores apresentam diferentes razões para usar técnicas de ADCP, tais como, melhorar o resultado do agrupamento final, diminuir a sensibilidade a ruídos, outliers ou variações de amostra [Nguyen 2007]. Em geral, tem-se como senso comum que a opinião de consenso de um grupo, é mais confiável do que a opinião de um único indivíduo. O processo de ADCP é apresentado na Figura 1.

Ele recebe um conjunto de dados como entrada. Em seguida na etapa de geração de partições,  $M$  partições são produzidas utilizando um ou vários algoritmos de agrupamento e



**Figura 1. Agrupamento de Dados via Combinação de Partições**

reunidas em um único conjunto. Esta etapa é fundamental uma vez que o resultado alcançado pela função de consenso depende das informações dos conjuntos.

Uma vez que o conjunto de partições está disponível, a etapa de consenso ocorre. Seu objetivo é combinar todas as partições no conjunto de agrupamento em um resultado consensual final também chamado de partição consenso. Uma revisão sistemática dos métodos consensuais para agrupamento de dados pode ser consultada em [VEGA-PONS 2010] e em [Abdala 2010].

ADCP apresenta diversas características desejáveis ao processo de agrupamento, no entanto possui uma considerável desvantagem. A complexidade computacional do processo é ditada pelo algoritmo mais lento utilizado para construir o grupo de partições. Adicionalmente, temos que considerar o número de partições na equação de complexidade e por fim a complexidade associada ao algoritmo de consenso utilizado para combinar todas as partições em uma partição consensual final. O passo inicial e ainda crítico em qualquer método de ADCP refere-se ao processo de agrupamento de dados [Ghaemi 2009]. O problema da complexidade pode ser minimizado se introduzirmos estratégias de amostragem. Pode-se utilizar amostragem frequentemente tanto para investigação preliminar dos dados quanto para a análise final dos dados, pois tem-se como princípio que uma amostra deve ser representativa da população. Uma revisão extensiva sobre amostragem aplicada ao contexto de agrupamento pode ser encontrada em [GU 2000].

Este trabalho visa avaliar a viabilidade de aplicar amostragem de dados no contexto de agrupamento de dados via combinação de partições. Considerando a Figura 1, antes da etapa consensual, cada partição é amostrada e apenas a amostra é passada para a etapa de consenso. Três métodos distintos de amostragem foram estudados. O primeiro método investigado foi amostragem aleatória sem repetições simples que conceitualmente é a mais simples das técnicas de amostragem probabilística. Neste tipo de amostragem, cada indivíduo é escolhido aleatoriamente sem repetições, de tal modo que cada indivíduo tem a mesma probabilidade de ser escolhido em qualquer fase durante o processo de amostragem, e cada subconjunto de indivíduos  $k$  tem a mesma probabilidade de ser escolhidos para a amostra [Olken 1993]. O segundo método é denominado amostragem sistemática. Neste método de amostragem cada  $k$ -ésimo elemento da população é amostrado, começando com um valor escolhido aleatoriamente em um limite de 1 a  $K$ . Este elemento  $k$ -ésimo é conhecido como índice de amostragem e é calculado dividindo o tamanho da amostra pelo tamanho da população. Em seguida o índice de amostragem é utilizado para definir os saltos utilizados para percorrer a população e selecionar os elementos da amostra. Por fim, investigou-se o método de amostragem baseado em densidade [Palmer 2000]. Este método preserva a estrutura de grupos existente em uma população, garantindo assim que grupos compostos por poucas observações estejam presentes na amostra. Considere uma população composta por 1000 indivíduos, subdividida em 3 grupos, o primeiro composto por 800, o segundo por 180 e o terceiro por apenas 20 indivíduos. Em um processo de amostragem aleatório há uma grande possibilidade de que nenhum representante do terceiro grupo seja escolhido. O método de amostragem baseado em densidade garante que pelo menos

um indivíduo deste grupo seja escolhido. Se amostrarmos 10% desta população este método garantirá que 80 indivíduos do primeiro, 18 do segundo e 2 do terceiro grupo sejam escolhidos.

## 2. Contribuição do Trabalho

A principal contribuição deste trabalho refere-se a considerar processos de amostragem no contexto de agrupamento de dados via combinação de partições. As contribuições menores referem-se a avaliar diferentes métodos de amostragem e seu impacto e estudar taxas de amostragem adequadas em relação ao tamanho do conjunto de dados.

## 3. Análise de Resultados

Para avaliar o impacto do processo de amostragem de dados no contexto de agrupamento de dados via combinação de partições foi dividido um teste que visava verificar o impacto da amostragem na qualidade da partição consensual obtida ao final do processo. Para este fim, selecionou-se quatro bases de dados sintéticas, disponíveis em [Abdala 2010] e onze bases de dados publicamente disponíveis [Frank and Asuncion 2010] e largamente utilizadas pela comunidade de reconhecimento de padrões para fins de comparação cruzada de resultados.

As quinze bases de dados foram processadas utilizando o algoritmo *EAC\_AL* [Fred 2005]. Todos os dados disponíveis foram utilizados e por fim, utilizou-se a métrica de similaridade *vi* (variation of information) [Meila 2003] como medida de comparação. Valores próximos de zero significam bons resultados. O processo foi repetido outras quatro vezes utilizando 50, 30, 20 e 10 % dos dados respectivamente. Como citado anteriormente, três algoritmos de amostragem foram utilizados, Sistematic Sampling, Random Sampling e Density Biased Sampling (DBS).

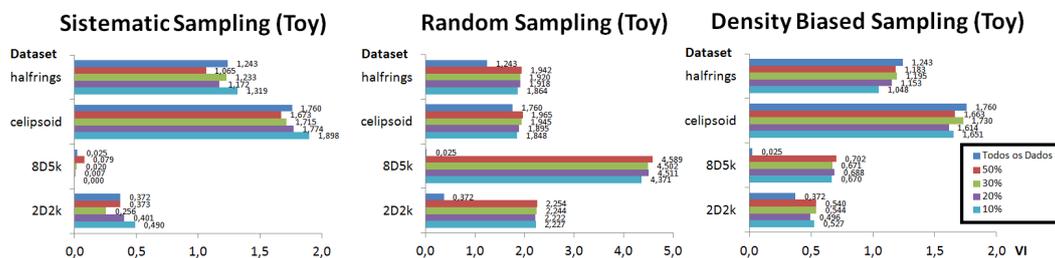
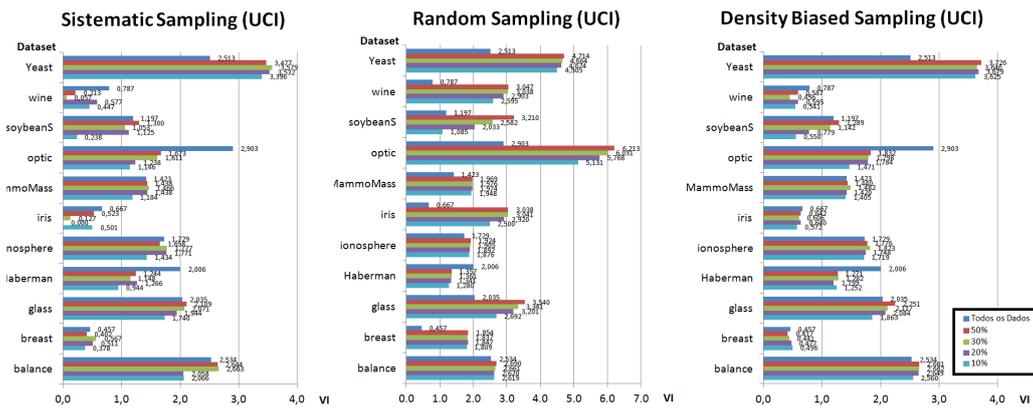


Figura 2. Resultado comparativo para as bases de dados de testes utilizando a métrica *vi* (variation of information) do algoritmo *EAC\_AL* para todo o ensemble, 50, 30, 20 e 10 % dos dados.

Como pode-se observar na Figura 2 os resultados obtidos utilizando todos os dados e as amostras de diferentes tamanhos obtiveram resultados similares. Pequenas variações podem ser observadas por exemplo nas bases "half\_hings" e "2D2K" utilizando o processo de amostragem random sampling. Índices *vi* maiores gerados pelos métodos de amostragem em comparação com a utilização de toda a base de dados indicam que houve uma perda considerável de qualidade no resultado final. O mesmo foi observado para a base "2D2K" usando o algoritmo DBS para amostragem. Ainda considerando o algoritmo DBS observa-se que para as demais bases de dados as amostras produziram resultados melhores que se considerando todos os dados disponíveis. Este é um fato surpreendente pois implica que a utilização parcial dos dados pode melhorar o resultado final. Nestes casos mais objetivos corretamente classificados foram selecionados em comparação a seleção dos erroneamente classificados, levando a um índice *vi* menor.

Para as bases UCI a maioria das bases O resultado obtido usando todos os dados é comparável ao obtido pelas amostras. Há casos em que as amostras produzem resultados melhores



**Figura 3. Resultado comparativo para as bases de dados da UCI utilizando a métrica vi (variation of information) do algoritmo *EAC\_AL* para todo o ensemble, 50, 30, 20 e 10 % dos dados.**

em comparação a utilização da base completa e em alguns poucos casos as amostras produzem resultados piores. Em geral, amostras maiores produzem resultados levemente melhores que amostras menores. O que nos leva a inferir que o tamanho da amostra, dentre os quatro valores escolhidos tem pouca relevância. Dentre os três métodos de amostragem, Random sampling foi o que apresentou o pior desempenho. Como a amostragem é randomizada, duas execuções consecutivas geram resultados distintos. Por este motivo os resultados reportados são a média de 10 amostragens. Em algumas os resultados são comparativos aos obtidos pelo DBS, o melhor algoritmo avaliado. No entanto na média, os resultados não são confiáveis, e consequentemente, random sampling não é uma boa opção. O algoritmo DBS no entanto se mostrou consistente e uma opção viável para lidar com o problema de complexidade em agrupamento de dados via combinação de partições. Como trabalhos futuros esperamos avaliar o impacto da aplicação do processo de amostragem diretamente na base de dados e não nas partições, minimizando assim também o esforço no processo de geração do conjunto de partições.

## Referências

- Abdala, D. D. (2010). Ensemble and constrained clustering with applications. (1):45–70.
- Frank, A. and Asuncion, A. (2010). Uci machine learning repository.
- Fred, A. L. (2005). Combining multiple clusterings using evidence accumulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(6):835–850.
- Ghaemi, R. (2009). A survey: cluster ensemble techniques. *Proc. of World Academy of Science, Engineering and Technology*, 38:644–657.
- GU, B. (2000). Sampling and its application in data mining: A survey. 39(1):1–27.
- Meila, M. (2003). Comparing clusterings by the variation of information. *Learning Theory and Kernel Machines*, pages 173–187.
- Nguyen, N. (2007). Consensus clusterings. *Seventh IEEE International Conference on Data Mining*, pages 1–10.
- Olken, F. (1993). Random sampling from databases. *Department of Computer Science, University of California Berkeley*.
- Palmer, C. R. (2000). Density biased sampling: An improved method for data mining and clustering. (1):1–27.
- VEGA-PONS, S. (2010). A survey of clustering ensemble algorithms. (1):1–36.

# Caracterização de Nanofibras Geradas por Bactérias via Métodos de Processamento de Imagens

**Autor: Mariana Luiza Volpi Fiori<sup>1</sup>,  
Orientador: Daniel Duarte Abdala<sup>1</sup>**

<sup>1</sup>Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

marianavolpifiori@gmail.com, abdala@ufu.br

Nível: Trabalho de Conclusão de Curso

**Resumo.** *Este projeto de pesquisa prevê o desenvolvimento de métodos computacionais para a análise e mensuração de características de nanofibras de celulose a partir de imagens de microscopia eletrônica por varredura. São realizadas as etapas de normalização de imagens, segmentação de nanofibras, resolução de problemas de cruzamento e aproximação tangencial e oclusão de partes de fibras. Uma vez isoladas, as nanofibras serão mensuradas para caracterização de seu comprimento mínimo, médio e máximo, área do perímetro e curvatura e torção das fibras. É esperado que estas medidas quantitativas sirvam como importante ferramenta para prospecção da adequação destas nanofibras e para o planejamento de futuros desenvolvimentos de novos materiais.*

**Palavras-Chave.** *nanofibra, microscopia eletrônica por varredura, processamento de imagens.*

## Introdução, Motivação e Fundamentação Teórica

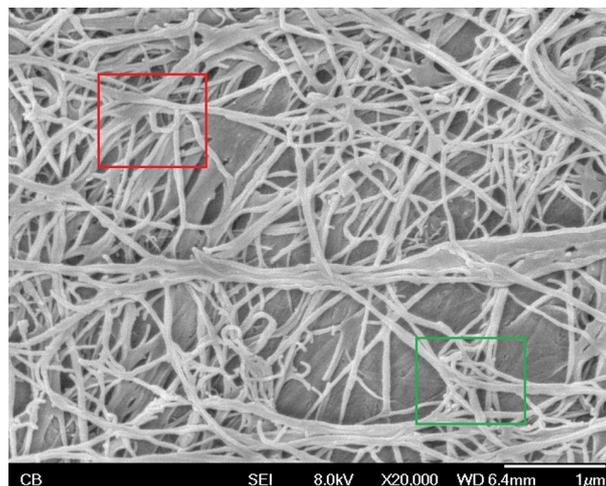
Recentemente, nanofibras têm atraído considerável atenção devido a uma ampla potencial gama de aplicações, tais como, polímeros biodegradáveis [Muller et al. 2013], termoplásticos [Schmidt et al. 2013] e próteses cardiológicas [Oliveira et al. 2013]. A caracterização física das nanofibras é de grande interesse visto que esta permite a prospecção de quais aplicações práticas podem ser mais bem atendidas por nanofibras apresentando um dado conjunto de características.

As fibras de algodão são de particular interesse devido a representarem uma considerável parte do mercado têxtil mundial. O padrão comercial para mensuração de sua qualidade refere-se ao teste Shirley SMT (Microscópio de Corrente de Tunelamento) [Huang and Xu 2002]. Este teste baseia-se na análise de seções transversais de preparados de fibra de algodão que são imaginografados via microscopia de corrente de tunelamento. Após a coleta da imagem, estas são analisadas para mensuração de características como o indicador de permeabilidade ao ar, a densidade linear e a maturidade das fibras medidas que indicam a qualidade e indiretamente as características físicas das fibras de algodão influenciando diretamente a fiação. O processo de preparação das fibras para captura de seções transversais é laborioso e caro, fato que desencadeou esforços no sentido de caracterizar fibras de algodão a partir de vistas longitudinais. Estas são mais fáceis de preparar e menos caras. Dentre os trabalhos de referência na área, podemos citar [Adel et al. 2011] e [Frydrych et al. 2010], em que ambos tem como objetivo mensurar a maturidade e finura das fibras.

Não há métodos disponíveis para caracterização de nanofibras quaisquer tal como os para fibras de algodão. A utilização direta dos métodos disponíveis não se aplica devido ao fato de que as características das nanofibras podem variar consideravelmente. Por exemplo, o tamanho médio das nanofibras é de interesse, um dado não analisado em fibras de algodão, pois seu comprimento médio é conhecido. Um trabalho recente foi desenvolvido devido a crescente utilização de "andaimos" de nanofibras para serem utilizados em engenharia de tecidos e a falta de uma solução disponível e gratuita para realização de uma análise rápida e automatizada do diâmetro das nanofibras [Hotaling et al. 2015]. Esse trabalho consiste em criar um plugin que analisa as nanofibras obtidas por microscopia eletrônica de varredura em 60 segundos.

Outras características de interesse são o perímetro e curvatura das fibras. Tais características podem ser extraídas via algoritmos de processamento de imagens. No entanto, vários desafios se apresentam. Por exemplo, as fibras podem se sobrepor causando dificuldades em decidir computacionalmente a continuidade de fibras. Dado o fato de que as imagens longitudinais não fornecem informação tridimensional acerca das fibras, o perímetro deve ser decidido de maneira indireta.

Os desafios encontrados durante a etapa de processamento de imagens podem ser vistos na Figura 1.



**Figura 1. Nanofibras via imagem de microscopia eletrônica de varredura. Na figura são destacadas duas áreas de desafios que serão encontrados durante a fase de Segmentação Tracking, são eles: em vermelho, tem o desafio das nanofibras se sobrepondo e a dificuldade de diferenciar as nanofibras do fundo com as nanofibras a frente. Em verde, tem dois desafios que dificultam a decisão da continuidade das nanofibras. Tais como, as nanofibras se tocando tangencialmente também chamado de "fiber kissing", onde as nanofibras se tocam mas não se cruzam, e também tem o caso em que as nanofibras se cruzam e há indecisão acerca de como as curvas continuam.**

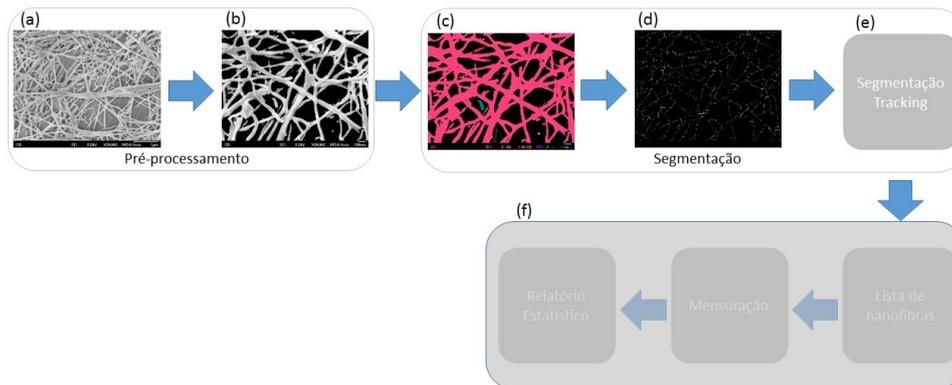
Neste trabalho é proposto a pesquisa e desenvolvimento um conjunto de algoritmos de processamento de imagens para processamento e extração de características assim como métricas para mensuração de características de interesse em nanofibras.

## **Contribuição do Trabalho**

A mensuração das características relevantes de nanofibras pode vim a ser útil na determinação do conjunto de aplicações a que estas se prestam. A disponibilidade de dados quantitativos como os supracitados pode auxiliar na determinação da necessidade e/ou viabilidade do direcionamento de pesquisa de novas formas de nanofibras que apresentem características não presentes nas nanofibras disponíveis.

## Análise de Resultados

O estado atual do trabalho consiste na etapa de segmentação da nanofibra. Como pode-se observar na Figura 1, é muito comum que nanofibras se sobreponham dificultando o processo de individualização das fibras onde partes oclusas terão que ser interpoladas. Outro problema recorrente refere-se a fibras que se cruzam e que se tocam tangencialmente fato que pode induzir a separações errôneas de fibras. A fase sendo realizada nesta etapa, é o desenvolvimento de um algoritmo capaz de decidir por qual caminho uma fibra irá seguir caso encontre alguns dos problemas supracitados.



**Figura 2. Digrama de fluxo das etapas do trabalho. Onde, (a) apresenta a imagem original obtida por microscopia eletrônica de varredura, (b) apresenta o resultado final da etapa de pré-processamento, depois de aplicado a equalização, o Fitro Anistrópico e Thresholding, (c) apresenta as componentes conexas encontradas a partir do resultado final da etapa de pré-processamento, é visto que componente conexa rosa representa a malha das nanofibras de interesse do trabalho, (d) apresenta a esqueletonização das nanofibras, (e) apresenta a fase de segmentação tracking, essa fase usa a esqueletonização para percorrer as nanofibras, (f) apresenta as etapas que ainda não foram realizadas no trabalho, são elas: Lista de Nanofibras, Mensuração e Relatório Estatístico.**

Na Figura 2 temos um diagrama representando o fluxo das etapas presentes no trabalho. Cada etapa e seus resultados serão explicados a seguir:

### 1. Pré-Procesamento:

Com a imagem original em mãos, Figura 2, em (a), é obtido o seu histograma e em cima dele foi aplicado a equalização, obtendo um melhor destacamento das nanofibras. Assim que é finalizada a fase de equalização, é aplicado um filtro escolhido, Filtro Anisotrópico, obtendo uma imagem com uma melhor visualização as nanofibras, sem ruídos e mais claras. E para encerrar a etapa de pré-processamento, é aplicado o Thresholding com liLista de Nanofibras, miar de 127, resultando em uma imagem contendo apenas as nanofibras de interesse e "limpando"o fundo para retira as nanofibras que não serão trabalhadas. É possível ver o resultado desta etapa na Figura 2, em (b).

### 2. Segmentação de nanofibras:

Finalizada a etapa de pré-processamento é iniciada a etapa de segmentação das nanofibras. Como podemos ver na Figura 2, esta etapa é dividida em 3 fases importantes. São elas, rotulação de componentes conexas (c), esqueletonização (d) e segmentação tracking (e).

Na imagem final resultante da etapa de pré-processamento, mesmo depois de ter aplicado o Thresholding para "limpar"o fundo, existe a possibilidade de ainda haver algumas nanofibras que não são de interesse para esse trabalho ou resto de uma que não

foi apagada completamente. Por esse motivo, é aplicado o algoritmo de Rotulação de Componente Conexas. Esse algoritmo é baseado em rotular cada componente conexa (objeto) presente na imagem e o nosso objeto (nanofibras) de estudo, vai ser a maior componente conexa. Em (c), é observado que esse objeto é representado pela cor rosa. Agora com uma imagem contendo apenas as nanofibras de interesse, é aplicado o algoritmo de esqueletonização. Esse algoritmo é baseado em traçar um traço (esqueleto) no meio de cada nanofibra. Em (d), é observado uma imagem com os esqueletos das nanofibras.

Aplicado os algoritmos de rotulação de nanofibras e esqueletonização, está na hora de iniciar a fase de Segmentação Tracking, fase a qual o trabalho se encontra. Essa fase inicia com a geração de uma semente aleatória, a qual será algum ponto (pixel) aleatório de uma nanofibra. Sabendo essa semente, é possível da início ao tracking, isto é, a partir da semente pode-se percorrer uma nanofibra. Antes de começar a percorrer é importante identificar se a semente está no meio ou final/início de uma nanofibra, caso esteja no meio, será necessário no final concatenar as duas partes para virar uma. Durante o tracking vai aparecer vários desafios, os quais são destacados na Figura 1. Esses desafios colocará o algoritmo de Segmentação de Tracking a decidir qual caminho continuará a percorrer a nanofibra. O algoritmo irá percorrer a nanofibra até encontrar o seu final.

### 3. Etapas futuras:

Na Figura 2, em (f), pode-se observar que existem três etapas que estão dentro de um retângulo sombreado, isto significa que elas ainda não foram desenvolvidas. No final da Segmentação Tracking, as fibras percorridas serão armazenadas em uma Lista de Nanofibras. Nessa lista, terá as nanofibras isoladas, podendo assim escolher cada uma e realizar a Mensuração para caracterização de seu comprimento mínimo, médio e máximo, área do perímetro e curvatura e torção. Obtida a Mensuração das nanofibras que se encontram na lista, é possível levantar dados estatísticos, comparando resultados e aproximando os mesmos. Finalizado o levantamento dos dados, é possível ter um Relatório Estatístico.

## Referências

- Adel, G., Faten, F., and Radhia, A. (2011). Assessing cotton fiber maturity and fineness by image analysis. *Journal of Engineered Fibers and Fabrics*, 6(2):50–60.
- Frydrych, I., Raczyńska, M., and Cekus, Z. (2010). Measurement of cotton fineness and maturity by different methods. *Fibres & Textiles in Eastern Europe*, 18(6):54–59.
- Hotaling, N. A., Bharti, K., Kriel, H., and Jr., C. G. S. (2015). Diameterj: A validated open source nanofiber diameter measurement tool. *Biomaterials*, 61.
- Huang, Y. and Xu, B. (2002). Image analysis for cotton fibers, part i: Longitudinal measurements. *Textile Research Journal*, 72(8):713–720.
- Muller, D., Rambo, C. R., and Porto, L. M. (2013). Structure and properties of polypyrrole/bacterial cellulose nanocomposites. *Carbohydrate Polymers*, pages 655–662.
- Oliveira, V. A., Rambo, C. R., and Porto, L. M. (2013). Produção e degradação in vitro de estruturas tubulares de celulose bacteriana. *Polímeros Ciência e Tecnologia*, 23(4):559–564.
- Schmidt, V. C. R., Berti, F., Porto, L. M., and Laurindo, J. B. (2013). Production of starch acetate films with addition of bacterial cellulose nanofibers. *Chemical Engineering Transactions*, 32:2251–2256.

# Cuca Chess: Um jogo de xadrez usando BCI

**Autor: Gustavo Vieira Soares<sup>1</sup>,  
Orientador: João Henrique de Souza Pereira<sup>1</sup>**

<sup>1</sup> Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

gustavovs@si.ufu.br, joaohs@ufu.br

Nível: Trabalho de Conclusão de Curso

**Resumo.** *O objetivo deste trabalho é implementar um jogo de xadrez virtual que pode ser controlado sem o uso de movimentos musculares, através de uma interface cérebro máquina (Brain Computer Interface - BCI). Essa aplicação é voltada principalmente para pessoas com deficiências motoras, como exemplo a esclerose lateral amiotrófica ou síndrome do encarceramento, embora pode ser usada por qualquer pessoa que não possui danos nas faculdades mentais. O jogo está sendo implementado na plataforma Unity, e a lógica é controlada por scripts C#. Será implementada também uma biblioteca em C, cujo objetivo é prover procedimentos de filtragem e classificação dos sinais obtidos pelo dispositivo. Serão usados potenciais evocados P300, que ocorrem quando o indivíduo é exposto a estímulos visuais de alta probabilidade enquanto espera a ocorrência de um estímulo em especial. Neste trabalho, linhas ou colunas serão destacadas aleatoriamente no tabuleiro de xadrez, e o estímulo relevante se caracteriza pelo destaque da linha ou coluna onde a peça desejada se localiza. Depois, o processo se repete para a seleção da casa de destino. O procedimento de movimentação de uma peça será ativado uma vez por lance através de um potencial espontâneo, o qual precisa apenas da intenção do usuário para ocorrer, mas em contrapartida, necessita de rotinas de treino para ser usado voluntariamente. O jogo na plataforma Unity já está operando, de forma que é possível selecionar uma peça e movimentá-la para qualquer casa do tabuleiro, mesmo que o movimento não respeite as regras do xadrez. A próxima etapa do trabalho é a aquisição, processamento e classificação de sinais e implementação das regras do xadrez.*

**Palavras-Chave.** *BCI, P300, potenciais evocados, jogos*

## 1. Introdução, Motivação e Fundamentação Teórica

Interface cérebro máquina (Brain Computer Interface - BCI) é uma interface de comunicação entre o cérebro e um dispositivo eletrônico que dispensa movimentos musculares. Através de um dispositivo munido de eletrodos posicionados em regiões específicas do escalpe do indivíduo é possível captar parte da atividade elétrica do cérebro.

As ondas elétricas captadas possuem diferentes padrões, por exemplo, frequência ou amplitude, que estão relacionados à atividades cerebrais específicas, o que permite a interface com computadores. Para que a comunicação cérebro máquina seja bem sucedida, é preciso captar o sinal, filtrá-lo, classificá-lo, e enviá-lo para a aplicação.

Aplicações podem ser baseadas em diferentes potenciais emitidos pelo cérebro: espontâneos, que são gerados a partir da intenção do indivíduo e potenciais evocados, que são gerados involuntariamente frente a um estímulo externo. Além disso, algumas emoções como raiva, interesse, ou estado de meditação podem ser reconhecidos.

Os potenciais espontâneos são gerados intencionalmente pelo indivíduo, o que quer dizer que ocorrem quando ele faz ou imagina algum movimento voluntário. Os potenciais espontâneos são relacionados principalmente com ritmos motor e sensório motor, que sofrem alteração enquanto o indivíduo realiza o movimento ou imagina este movimento. Os ritmos motor e sensório motor estão relacionados a ações motoras do indivíduo, como mover algum membro, a intenção ou imaginação do movimento. Estes ritmos ocorrem no córtex sensório motor e são caracterizados por um decréscimo nas frequências *alpha* e *beta* e acréscimo na frequência *gamma* [Yuan and He 2014]. Estudos mostram que é possível uma pessoa controlar a amplitude do ritmo sensório motor através de intenção ou imaginação de movimentos de objetos reais ou virtuais [Wolpaw and McFarland 2004] [Royer et al. 2010].

Potenciais evocados surgem frente à um determinado estímulo externo, que pode ser visual, sonoro ou somato sensorial. Potenciais evocados são caracterizados por picos de amplitude observados na região responsável por processar determinado estímulo (córtex visual em caso de estímulos visuais) [Luck and Kappenman 2011]. Alguns potenciais evocados possuem a vantagem do uso sem a necessidade de treinamento, ou seja, *plug and play* para qualquer tipo de usuário [Godói 2010]. O P300 está entre os principais potenciais evocados usados na comunicação cérebro máquina [Schomer and Da Silva 2012].

Os sinais cerebrais podem ser captados através de técnicas invasivas ou não. Técnicas invasivas são aquelas que demandam intervenção cirúrgica para serem implantadas, ou seja, os eletrodos são posicionados diretamente no cérebro do indivíduo. Por outro lado, existem técnicas não invasivas que detectam os potenciais elétricos no escalpe do usuário. As formas mais convenientes de se obter sinais neurofisiológicos para BCI são o eletroencefalograma (EEG), não invasivo, e eletrocorticografia (ECoG), invasivo.

Neste trabalho, será usado o dispositivo não invasivo Emotiv EPOC para obtenção dos sinais. No jogo de xadrez, será implementado um *speller* para ativar o potencial P300, que consiste no destaque aleatório de linhas e colunas em uma matriz, até que a linha ou coluna pretendida seja destacada e o potencial gerado após 300 milissegundos. O *speller* implementado consiste no destaque em vermelho das linhas ou colunas no tabuleiro. Desta forma, o usuário precisa selecionar uma linha e uma coluna para obter a localização atual da peça, e posteriormente outra linha e outra coluna para definir a casa destino da peça selecionada. Ao completar o movimento, o *speller* para de ocorrer até que o indivíduo gere um potencial espontâneo para ativá-lo novamente. O jogo de xadrez será implementado na plataforma Unity e controlado por scripts C#. Será implementada também uma biblioteca em C que proverá funções de filtragem e classificação de potenciais evocados P300, que poderá ser utilizada em trabalhos futuros e outros autores.

## 2. Contribuição do Trabalho

O fabricante do dispositivo usado não oferece bibliotecas oficiais para obtenção, filtragem e classificação de potenciais evocados, desta forma, este trabalho propõe uma alternativa para o uso destes potenciais.

No estado atual, foi implementado um protótipo do jogo de xadrez controlado por mouse e botão de espaço que estão substituindo provisoriamente os sinais evocados e espontâneos, respectivamente. A etapa em desenvolvimento envolve a pesquisa de formas de obtenção do sinal bruto, ou seja, sem filtros e sem classificação. A próxima etapa é o estudo de técnicas de filtragem e classificação de sinais. É preciso investigar diferentes tipos de filtros, métodos de eliminação de ruídos (sinais provenientes de outras atividades do corpo que não estão relacionadas com a aplicações), e também classificação dos sinais filtrados. Nesta aplicação é preciso

classificar o sinal por amplitude. São necessárias duas classes: ausência de estímulo (amplitude baixa) e presença de estímulo (amplitude alta). Depois de definidas as técnicas para processar o sinal obtido, será criada a biblioteca na linguagem C, para que os sinais processados possam ser convertidos em comandos para o protótipo. Por último, a implementação da versão final do jogo de xadrez, com as regras bem definidas e menu para o usuário.

### 3. Análise de Resultados

No protótipo implementado, as linhas ou colunas do tabuleiro piscam uma vez a cada meio segundo. Levando em conta que o P300 leva de 300 a 600 milissegundos para ocorrer [Schomer and Da Silva 2012] e que o destaque das linhas com tempo maior que meio segundo pode tornar o jogo pouco dinâmico, este valor será mantido. Após a finalização das etapas de processamento e classificação do sinal e integração com o jogo, serão feitos testes com pessoas com alguma deficiência motora e também pessoas sem deficiência com o intuito de encontrar o melhor *delay* entra um destaque e outro, sem comprometer o funcionamento do *speller*.

### 4. Trabalhos Relacionados

Existem diversas abordagens para desenvolvimento de jogos utilizando BCI; pode-se explorar diferentes formas de captação de sinais, e diferentes padrões a serem explorados (espontâneos e evocados). É possível também programar uma aplicações que possam utilizar diferentes tipos de potenciais para desempenhar tarefas diferentes. Abaixo alguns exemplos de aplicações usando BCI:

- Brain Chess Em 2011, foi desenvolvido um jogo de xadrez usando BCI que combina os potenciais corticais lentos (*Slow Cortical Potential - SCP*) e P300 para completar um movimento. De acordo com os autores, os padrões P300 captados em um treino específico são diferentes para cada peça que se deseja mover. Após o treino, basta o jogador pensar na peça e focar sua visão nela, que o BCI reconhecerá a intenção do usuário. Para o movimento das peças, aplicaram SCP para aquelas cujo movimento possível consiste em poucas casas, e , para peças com movimentos de muitas casas, como a rainha, um P300 *speller* de uma matriz ordenada de letras de A até H e números de 1 até 8, que são os caracteres necessários para descrever a posição da casa no tabuleiro [Maruthappan et al. 2011].
- Helicóptero Virtual: Em 2010, foi implementado um ambiente virtual onde um helicóptero voava livremente em uma área, e através de potenciais SCP e sessões de treinamento com os usuários, foi possível controlar este helicóptero via BCI, com um sistema de 4 classes: uma para cada direção. O helicóptero voava em uma velocidade constante pelo local [Royer et al. 2010].
- Jogo da memória: Em 2012, foi implementado um jogo da memória utilizando um P300 *speller* [Angeloni et al. 2012], com intuito principal de tornar o processo de treinamento mais divertido. As cartas eram posicionadas em uma matriz, viradas para baixo; o usuário escolhia duas e, caso elas fossem diferentes, eram viradas para baixo novamente.

Outros exemplos de aplicações podem ser encontrados em [Kaplan et al. 2013].

### Referências

Angeloni, C., Salter, D., Corbit, V., Lorence, T., Yu, Y.-C., and Gabel, L. (2012). P300-based brain-computer interface memory game to improve motivation and performance. In *Bioengineering Conference (NEBEC), 2012 38th Annual Northeast*, pages 35–36. IEEE.

- Godói, A. C. B. d. (2010). *Detecção de potenciais evocados P300 para ativação de uma interface cérebro-máquina*. PhD thesis, Universidade de São Paulo.
- Kaplan, A., Shishkin, S., Ganin, I., Basyul, I., and Zhigalov, A. (2013). Adapting the p300-based brain-computer interface for gaming: a review. *Computational Intelligence and AI in Games, IEEE Transactions on*, 5(2):141–149.
- Luck, S. J. and Kappenman, E. S. (2011). *The Oxford handbook of event-related potential components*. Oxford university press.
- Maruthappan, N., Iyengar, N., and Patel, P. S. (2011). Brain chess - playing chess using brain computer interface. *Int. Proc. of Computer Science and Information Technology*, 20:183–191.
- Royer, A. S., Doud, A. J., Rose, M. L., and He, B. (2010). Eeg control of a virtual helicopter in 3-dimensional space using intelligent control strategies. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 18(6):581–589.
- Schomer, D. L. and Da Silva, F. L. (2012). *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins.
- Wolpaw, J. R. and McFarland, D. J. (2004). Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17849–17854.
- Yuan, H. and He, B. (2014). Brain-computer interfaces using sensorimotor rhythms: current state and future perspectives. *Biomedical Engineering, IEEE Transactions on*, 61(5):1425–1435.

# Estudo dos parâmetros do Algoritmo Evolutivo aplicado ao Problema de Predição de Estruturas de Proteínas usando modelo HP-2D.

**Autor: Júlia Manfrin Dias<sup>1</sup>**  
**Orientadora: Christiane R. S. Brasil<sup>1</sup>**

<sup>1</sup>Faculdade de Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

[juliamanfrindias@gmail.com](mailto:juliamanfrindias@gmail.com), [christiane.ufu@gmail.com](mailto:christiane.ufu@gmail.com)

Nível: Trabalho de Conclusão de Curso

**Resumo.** *Dentre os diversos métodos de otimização computacional, destaca-se o Algoritmo Evolutivo. O objetivo principal deste trabalho é aplicar este algoritmo para o Problema de Predição de Estrutura de Proteínas, a fim de estudar e analisar o desempenho deste método com diferentes parâmetros. Para representar computacionalmente o PSP foi utilizado o modelo HP em malha 2D.*

**Palavras-Chave:** *Algoritmo Evolutivo, PSP, Estrutura, modelo HP.*

## 1. Introdução, Motivação e Fundamentação Teórica

### 1.1 Introdução

A otimização computacional é o processo que possibilita o melhor rendimento para execução de uma tarefa, que pode ser representada por um método que seja mais eficiente e apresente soluções ótimas ou aproximações para um dado problema de grande complexidade. Tais problemas são classificados como problemas NP (não polinomiais) (HAUPT; HAUPT, 2004).

Um exemplo do mundo real de problemas NP é o problema de predição de estruturas de proteínas (do inglês, *Protein Structure Problem – PSP*). As proteínas são moléculas biológicas fundamentais para a vida, sendo imprescindível a obtenção das estruturas tridimensionais destas moléculas, uma vez que as estruturas das mesmas estão diretamente relacionadas às suas características funcionais.

### 1.2 Motivação

Conhecendo as funções das proteínas por meio de suas estruturas tridimensionais, pode-se alcançar novas drogas ou vacinas para prevenção de doenças, até hoje consideradas incuráveis. Neste contexto, métodos computacionais de otimização vêm tratando do problema PSP, pois os métodos convencionais (RNM e cristalografia) não são eficientes em termo de tempo e custo.

Uma técnica de otimização bem conhecida e que aborda este problema é o Algoritmo Evolutivo (AE), desenvolvido por (HOLLAND, 1975) e popularizado por (GOLDBERG, 1989).

### 1.3. Objetivo

O objetivo deste trabalho é aplicar o AE ao problema PSP a fim de analisar o desempenho do método com diferentes parâmetros. Neste trabalho a representação dos aminoácidos das proteínas foi feita pelo modelo simplificado Hidrofóbico-Polar.

### 1.4. Fundamentação Teórica

#### 1.4.1 – Algoritmos Evolutivos

Os Algoritmos Evolutivos (AE) são uma das técnicas de otimização de problemas inspirados em teorias evolucionistas de Darwin (HOLLAND, 1975; GOLDBERG, 1989). O AE inicia com uma população aleatória, composta por indivíduos. Cada indivíduo representa uma possível solução para problema. Estes indivíduos são avaliados e recebem um valor (*fitness*) que indica o quão boa é esta solução para o problema. A cada geração do AE a população executa operações de reprodução (*crossover* e mutação). Os indivíduos são avaliados, e aqueles que não apresentam uma boa avaliação são substituídos pelos filhos. Este processo é repetido até que a resposta ótima seja encontrada ou um critério de parada for atingido.

#### 1.4.2 – Problema de Predição de Estrutura de Proteínas

Existem quatro tipos de estruturas de proteínas: primária, secundária, terciária e quaternária (GABRIEL, 2010; BRASIL, 2012). Resumidamente, é a estrutura terciária que apresenta o formato tridimensional da proteína na natureza, e está relacionada às funções da mesma, sendo o alvo do problema da predição. Essa estrutura é definida pelo enovelamento da sua cadeia, sendo direcionado por diversos critérios, destacando-se neste trabalho a hidrofobicidade. Devido à hidrofobicidade, os aminoácidos hidrofóbicos da proteína não interagem com o meio voltando-se para o centro da estrutura, enquanto que as partes polares da cadeia tem contato com o meio.

#### 1.4.3 – Modelo Hidrofóbico Polar

O modelo simplificado Hidrofóbico-Polar (HP) (LAU e DILL, 1989) é um modelo de representação de proteínas baseado em redes (*lattice model*) que pode ser 2D ou 3D. Este modelo classifica os aminoácidos em dois grupos: hidrofóbicos (H) e polares (P). Neste modelo, o cálculo da energia livre é feita do seguinte modo (GABRIEL, 2010): cada interação química entre os aminoácidos hidrofóbicos (H) que são vizinhos não conectados somam -1 à energia, sendo inversamente proporcional a quantidade de interações H-H. Portanto, esta função objetivo busca a menor energia livre possível. Quanto menor a energia, melhor a conformação da estrutura encontrada.

## 2. Contribuição do Trabalho

Um detalhe na implementação deste AE está no tratamento de colisão das soluções geradas, que ocorre da seguinte maneira: como dois aminoácidos não podem ficar na mesma posição da malha, o algoritmo de tratamento de colisão, ao detectar a ocorrência deste evento, imediatamente verifica outra posição possível para colocar o aminoácido que seria sobreposto. Poderá haver momentos em que apenas a verificação de uma nova posição não seja suficiente e, sendo assim, este tratamento usa uma pilha para retroceder aos aminoácidos anteriores ajustando novas posições.

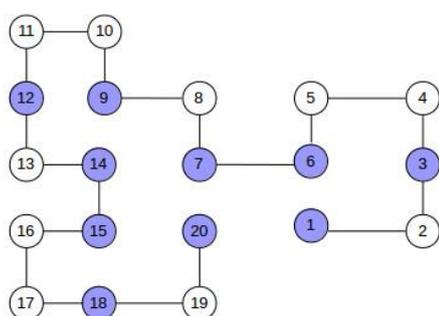
### 3. Análise de Resultados

Os testes foram realizados com base no artigo de (HUANG; YANG; HE, 2010), a sequência dos aminoácidos foram extraídos deste trabalho com o objetivo de verificar a eficiência e evolução do algoritmo.

Teste	E	Tam. Pop.	Gerações	Crossover	Qtde cross.	Mutação	Qtde Mut.
1	-9	200	100	100%	1	30%	1
2	-7	200	100	100%	2	30%	1
3	-9	100	100	100%	2	30%	1
4	-8	100	100	100%	1	30%	1

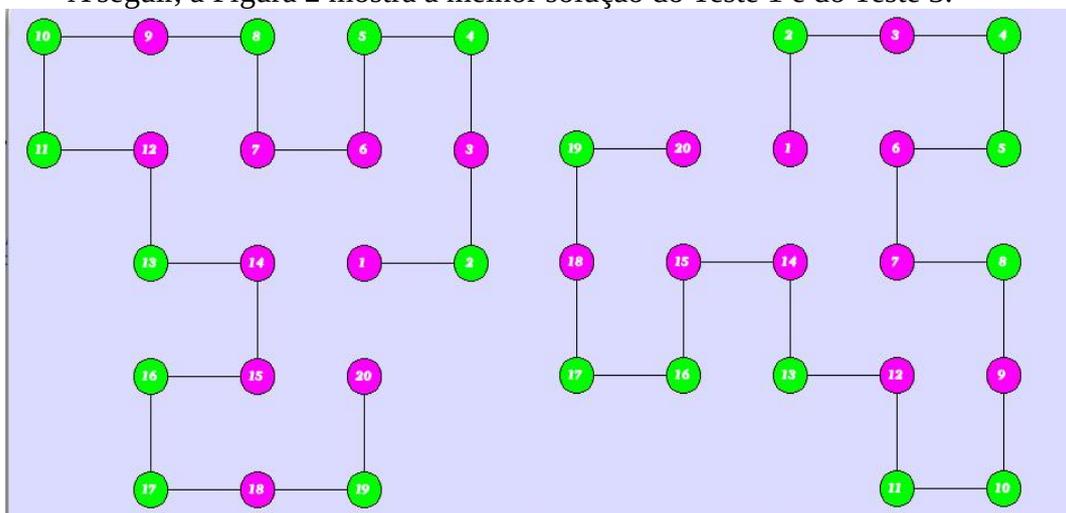
**Tabela 1:** E é a energia obtida, Tam. Pop. é o tamanho da população, Gerações é o número de iterações do AE, Crossover mostra a taxa de probabilidade do mesmo ocorrer a cada geração, Qtde cross é a quantidade de tipos de crossover usados, Mutação é a taxa de probabilidade da mesma ocorrer a cada geração, Qtde Mut é a quantidade de tipos de mutações usados.

A Figura 1 mostra a estrutura da proteína teste (HUANG; YANG; HE, 2010). Sua sequência é HPHPPHHPHPPHPPHPPH e sua menor energia é -9.



**Figura 1.** Estrutura ótima da proteína usada para os testes (HUANG; YANG; HE, 2010). Os círculos em azul representam os aminoácidos hidrofóbicos (H) e os brancos são os aminoácidos polares (P). Os números dentro dos círculos mostram a ordem dos aminoácidos da sequência.

A seguir, a Figura 2 mostra a melhor solução do Teste 1 e do Teste 3.



**Figura 2.** Melhor solução do Teste 1 (esquerda) e do Teste 3 (direita) com AE. Os pontos em rosa representam os aminoácidos hidrofóbicos (H) e os pontos em verde são os aminoácidos polares (P). As energias mínimas de conformações em ambos foram -9.

Pelos experimentos realizados, pode-se observar que tamanho de população 100 e geração 100 foi suficiente para alcançar o valor de energia mínima de -9, a mesma do artigo de (HUANG; YANG; HE, 2010), usando dois tipos de *crossover* (de um ponto e dois pontos). Em contrapartida, com a população pequena (tamanho 100) com apenas um tipo de *crossover*, não foi gerado variabilidade de indivíduos suficiente para alcançar energia -9. Com o tamanho da população 200, a energia gerou resultado pior,

pois a diversidade aumentou mais do que o necessário, prejudicando a convergência do método. A energia -9 é obtida com tamanho 200 no AE com somente um tipo de *crossover* (de um ponto). Deste modo, obteve-se a energia mínima -9 com: a população maior (200) e um tipo de *crossover* (Teste 1), ou população pequena (100), e dois tipos de *crossover* (Teste 3). Visualmente, o melhor resultado obtido foi com o Teste 1, e pela eficiência computacional o melhor desempenho foi o Teste 3.

#### 4. Trabalhos Relacionados

Um dos primeiros trabalhos no Brasil que abordaram o problema PSP foi o de (CUSTODIO; BARBOSA; DARDENNE, 2004) que realizou um estudo sobre a busca de estruturas de proteínas, usando modelo HP-3D em AE. O trabalho de (CUTELLO; NARZISI; NICOSIA, 2006), uma das principais referências da área, usa AE com o modelo *full atom*, que diferente do modelo HP, considera todos os átomos da molécula e os ângulos de torção referentes a eles. Em sua tese de mestrado, (GABRIEL, 2010) trata o problema PSP com AE simples e multiobjetivo usando o modelo HP. Pode-se citar também a tese de doutorado de (BRASIL, 2012), que utiliza o modelo *full atom* com AE multiobjetivo, sendo uma abordagem mais realista e, por isso, mais complexa. O artigo de (HUANG; YANG; HE, 2010) apresenta alguns métodos evolutivos para solucionar o PSP, sendo o artigo de comparação para esta pesquisa.

#### Referências

BRASIL, C. R. S. Algoritmo evolutivo de muitos objetivos para predição *ab initio* de estrutura de proteínas. Tese (Doutorado) — Universidade de São Paulo, 2012.

CUSTODIO, F. L.; BARBOSA, H. J. C.; DARDENNE, L. E. *Investigation of the three-dimensional lattice HP protein folding model using a genetic algorithm. Genetics and Molecular Biology*, scielo, v. 27, p. 611 – 615, 00 2004. ISSN 1415-4757.

CUTELLO, V.; NARZISI, G.; NICOSIA, G. *A multi-objective evolutionary approach to the protein structure prediction problem. Journal of The Royal Society Interface, The Royal Society*, v. 3, n. 6, p. 139–151, fev. 2006. ISSN 1742-5662.

GABRIEL, P. H. R. Algoritmos evolutivos e modelos simplificados de proteínas para predição de estruturas terciárias. Dissertação (Mestrado) — Universidade de São Paulo, 2010.

GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning. New York, NY, USA: Addison-Wesley Longman Publishing Co., Inc.*, 1989.

HAUPT, R. L.; HAUPT, S. E. *Practical genetic algorithms. 2. ed. New Jersey, USA: John Wiley and Sons, Inc.*, 2004.

HOLLAND, J. H. *Adaptation in Natural and Artificial Systems. [S.l.]: The University of Michigan Press*, 1975.

HUANG, C.; YANG, X. HE, Z. *Protein folding simulations of 2D HP model by the genetic algorithm based on optimal secondary structures. Computing Biology and Chemistry*, v. 34, n. 3, p. 137-142, jun. 2010.

LAU, K. F.; DILL, K. *A lattice statistical mechanics model of the conformational and sequence spaces of proteins. Macromolecules*, v. 22, n. 10, p. 3986-3997, 1989.

# Fatores que influenciam o recebimento de spam

**Autor: Danielle Silva Oliveira**

**Orientador: Rodrigo Sanches Miani**

Universidade Federal de Uberlândia – Monte Carmelo, MG – Brasil

[danielle@si.ufu.br](mailto:danielle@si.ufu.br), [miani@ufu.br](mailto:miani@ufu.br)

Nível: Trabalho de Conclusão de Curso

**Resumo.** *A popularização do correio eletrônico teve como consequência o envio de mensagens eletrônicas não solicitadas, os spams. Este trabalho se propõe a investigar fatores que influenciam o recebimento de spam a partir da realização de um experimento utilizando contas de e-mail reais. O objetivo do experimento é verificar de que forma o comportamento de um usuário na Web caracterizado pelo: i) uso de redes sociais, ii) realização de compras eletrônicas, iii) participação em fóruns e listas de discussão e iv) utilização de ferramentas de compartilhamento de arquivos digitais; influencia no recebimento de spam. Resultados preliminares indicam que contas associadas a redes sociais (Facebook) e as contas do provedor GMX possuem uma tendência em receber um número maior de spam do que os outros grupos estudados.*

**Palavras-Chave:** Spam, Grupos de Exposição.

## 1. Introdução, Motivação e Fundamentação Teórica

A crescente utilização da Internet e a consequente popularização do correio eletrônico, causaram um grande impacto na vida das pessoas. Um problema diretamente ligado a popularização do correio eletrônico envolve o envio de mensagens eletrônicas comerciais não solicitadas, também conhecido como spam. O custo de envio de e-mails é baixo, principalmente se comparado ao da correspondência convencional (CERF, 2005). Tal fato serve de incentivo para os *spammers*: pessoas que enviam spam.

Entre os anos de 2002 e 2010 a média de spam enviados por dia passou de 2,4 bilhões para 300 bilhões (ALMEIDA, 2010). Contudo, esse cenário está passando por modificações. Dados estatísticos de notificações de spam reportados ao Centro de Estudos, Resposta e Tratamento de Segurança no Brasil (CERT.br) apontam uma queda do volume de spam no Brasil. Essas notificações são voluntárias e representam apenas parte da quantidade de spam entre os anos de 2003 e 2014. O Relatório Anual de Segurança Cisco 2014 também relatou uma queda no volume de spam em nível global no ano de 2013. Apesar de tal fato, o relatório da Cisco 2014 afirma que spam classificados como maliciosos continuam representando uma séria ameaça aos negócios de Tecnologia de Informação e Comunicação (TIC).

Hann et al. (2006), sugeriram um estudo para confirmar se o envio de spam segue algum tipo de orientação ou se são distribuídos aleatoriamente. Alguns fatores podem determinar a taxa de envio de spam, por exemplo, provedores de email e interesses declarados em determinados produtos ou serviços. No caso de seguirem algum tipo de orientação, os autores gostariam de saber o que determina o recebimento

de spam. O segundo objetivo foi investigar a influência de outros fatores, como idade e localização, a distribuição de spam.

Para tal experimento foram criadas 288 contas de email para pessoas fictícias com características diferentes, nos provedores da *Hotmail*, *Lycos*, *Excite* e *Yahoo*. Destas, 192 foram expostas em um serviço de hospedagem de sites, chamada *Yahoo Geocities*, que foi extinta em 2009. Para o restante das contas não foi construído uma página *Web Geocities*. Algumas características consideradas na criação dos email foram: interesses declarados, como computadores e tecnologia, idade, gênero e local de origem.

Ao longo do período de 33 semanas, os spams enviados para as contas de email das pessoas fictícias foram monitorados e analisados de acordo com cada característica. As conclusões obtidas ao final do experimento foram: 1) spams não são aleatórios, mas orientados para segmentos de consumo que são relativamente mais propensos a fazer compras on-line, aqueles que declaram interesse em produtos específicos ou serviços, adultos, e os residentes dos EUA; 2) a descoberta mais surpreendente foi a de que, a mais importante influência sobre a taxa de spam foi a identidade do prestador de serviços de e-mail. Especificamente, contas do provedor *Hotmail* receberam significativamente mais spam.

## **2. Contribuição do Trabalho**

Com base no trabalho anterior, o objetivo geral deste estudo é replicar o experimento conduzido por Hann et al. (2006), usando uma base de dados atualizada e que reflita o comportamento de um usuário da Web com diferentes características, como frequentar redes sociais, fazer compras on-line e utilizar ferramentas Web para compartilhamento e hospedagem de arquivos. Esses grupos serão chamados de Grupos de Exposição para uma determinada conta de email. Ao final será feito um confronto entre os resultados obtidos pelos autores e os que serão obtidos neste estudo, afim de atualizar a pesquisa desenvolvida anteriormente e fornecer novas informações sobre o cenário atual do envio de spam.

Como objetivos específicos pode-se listar: 1) criar contas de email em três provedores distintos grátis; 2) associar as contas de email em um ou mais Grupos de Exposição, que serão classificados em Compras, Rede Social, Fórum de Discussão e Hospedagem de Arquivos e que representarão o comportamento de um usuário na web; 3) criar estratégias para investigar a relação entre cada grupo de Exposição e a distribuição de spam; 4) analisar os Grupos de Exposição afim de concluir quais deles receberam mais spam.

Ao fim deste trabalho, espera-se verificar quais comportamentos (Compras, Rede Social, Hospedagem de Arquivos e Fórum de Discussão) são indicadores de spam, assim como verificar como cada um dos provedores de email estudados tratam as mensagens eletrônicas não solicitadas. Também espera-se que o banco de dados de contas de email seja compartilhado e utilizado pela comunidade acadêmica.

## **3. Análise de Resultados**

O procedimento experimental foi desenvolvido com base nas seguintes hipóteses:

Hipótese 1: Verificar a incidência de spams considerando as características idade, sexo e nacionalidade: 1a) taxas de spam seriam maiores para contas de e-mails associadas a indivíduos com 18 e 35 anos, em relação aos indivíduos com 60. Na pesquisa desenvolvida por Statista (2014), pessoas com idade entre 15 e 34 anos são os que mais acessam a Internet. Sendo assim, podemos inferir que essa faixa etária também representa as pessoas que mais recebem spams; 1b) taxas de spam não apresentaram diferenças para contas de e-mail associados a homens em relação às mulheres; 1c) taxas de spam seriam maiores em contas de e-mail associadas aos Estados Unidos do que do Brasil, como mostrado por Spamhaus (2016). Um dos motivos é que as taxas de participação e-commerce entre americanos são maiores em relação aos brasileiros.

Hipótese 2: Verificar a incidência de spams em cada um dos grupos considerados: 2a) taxas de spam seriam maiores para as contas de e-mail associadas a todos os grupos de exposição; 2b) analisando isoladamente cada grupo, taxas de spam seriam maiores nos grupos compras e rede social. De acordo com Hann et al. (2006), as redes sociais são uma das mais influentes e importantes formas de marketing e o objetivo do spam é promover as vendas; 2c) taxas de spam seriam maiores para a combinação rede social e compras do que em fórum de discussão e hospedagem de arquivos. Redes sociais podem ser usadas para influenciar os usuários e uma das finalidades do spam é justamente promover as vendas, sendo assim, acreditamos que os *spammers* usam dessa estratégia para direcionar os consumidores a um determinado segmento de compras.

No momento estamos na fase de análise dos spams. A fase de coleta ainda não foi finalizada, mas já temos alguns resultados preliminares com base nas hipóteses citadas a cima. O envio de spam, de fato, segue orientação e alguns fatores que influenciam o recebimento de spam são: a) as contas do provedor da GMX receberam mais spams se comparado os outros provedores de serviço; b) a característica sexo foi insignificante. Não foi observado nenhuma diferença entre os e-mails indesejados recebidos por indivíduos do sexo feminino ou masculino; c) associar uma conta de email ao Facebook é um fator significativo no recebimento de spam. Tais contas receberam mais spam em relação à aquelas contas que não foram associadas à Rede Social.

#### **4.Trabalhos Relacionados**

Alves e Marques-Neto (2012) analisaram e caracterizaram os usuários destinatários de mensagens eletrônicas presentes em dois conjuntos de dados. O primeiro conjunto de dados foi gerado pelo filtro antispam de um provedor de Internet corporativo brasileiro e o segundo construído a partir de e-mails legítimos deste mesmo provedor. Com o auxílio de uma métrica que quantifica a popularidade, verificou-se a existência de um conjunto de destinatários foco de spam que recebem muito mais spams do que os demais usuários da rede. Estes usuários foram classificados como *Heavy Spam Recipients*. Quanto aos destinatários de e-mails legítimos, também foram identificados usuários foco com maior popularidade dentre os demais, os quais foram classificados como *Heavy E-mail Recipients*. Através do processamento de um algoritmo foi criado um ranking dos usuários que receberam mais spams nos 12 meses analisados. Esses usuários foram classificados, com o auxílio do administrador de redes do provedor de email em estudo, em corporativas, pessoais e inutilizadas. Dentre outras

coisas, os resultados demonstram que as contas classificadas como corporativas, geralmente associadas a meios de comunicação ou divulgação, estão presente em maior quantidade no topo da lista de *Heavy Spam Recipients*.

Já no estudo proposto por Clayton (2008), foram analisados durante 8 semanas registros coletados pelo *UK Internet Service Provider Demon Internet*. Ao longo da investigação, os pesquisadores descobriram que o caractere inicial dos endereços de correio eletrônico afeta a proporção de spam recebido. Endereços de email com nome comuns e que começam com letras comuns, são muito mais propensos a receber spam do que aquele com nomes e letras incomuns. O autor observou que para aqueles endereços de email cujo a parte à esquerda do “@” começa com a letra “A”, 50,2% de todo email recebido é spam. Clayton chamou esse grupo de “*Aardvarks*”. Os endereços de email que começam com a letra “Z”, não recebem muito email, mas sua percepção é de que 74,3% de todo email é spam. Esse grupo foi chamado de “*Zebras*”. Clayton (2008) afirmou que o mais impressionante sobre os resultados é que um endereço de email real “*Zebras*”, vai descobrir que 20% do seu email é spam, enquanto um “*Aardvark*” real irá detectar que mais de 30% do seu email é spam. Uma razão para o comportamento descrito, é a maneira com que os *spammers* criam suas listas de email. Endereços de email que começam com letras comuns, como “A”, “P” e “H” tendem a encontrar-se vítima de dicionário ou ataques *Rumpelstiltskin*, onde *spammers* ordenam uma lista alfabeticamente por destinatário.

## Referências

- ALMEIDA, T. A. DE. SPAM: *do Surgimento à Extinção*. [s.l: s.n.], 2010.
- ALVES, T.; MARQUES-NETO, H. T. Caracterização e análise do comportamento de destinatários de SPAMs. p. 101–114, 2011.
- CISCO. *Cisco 2014 annual security report*. p. 1–81, 2014.
- CERF, V. G. Spam, spim, and spit. *Communications of the ACM*, v. 48, n. 4, p. 39, 2005.
- CERT.BR. *Estatísticas de Notificações de Spam Reportadas ao CERT.br*. 2015. Disponível em: <<http://www.cert.br/stats/spam/>>.
- CLAYTON, R. Do Zebras get more Spam than Aardvarks? *Proceedings of the Fifth Conference on Email and Anti-Spam*, 2008.
- HANN, I; HUI, K; LAI, Y; LEE, S. Y. T; PNG, I. P. L. Who Gets Spammed? *Communications of the ACM*, v. 49, n. 10, p. 83–87, 2006.
- STATISTA. *Distribution of internet users worldwide as of November 2014, by age group*. 2014. Disponível em: <<http://www.statista.com/statistics/272365/age-distribution-of-internet-users-worldwide/>>.
- SPAMHAUS. *The World's Worst Spam Enabling Countries*. 2016. Disponível em: <<https://www.spamhaus.org/statistics/countries/>>.

# Implementação de algoritmos de agrupamento de dados via combinação de partições em hardware

**Autor:** João Paulo Martins Arantes<sup>1</sup>,  
**Orientador:** Daniel Duarte Abdala<sup>1</sup>

<sup>1</sup>Bacharelado em Sistemas de Informação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

j.p\_martins@live.com, danielabdala@gmail.com

Nível: Trabalho de Conclusão de Curso

**Resumo.** *A utilização de métodos de ensemble clustering aumenta a confiabilidade dos resultados de clustering, aplicabilidade que por sua vez, necessita de alta capacidade de processamento. A implementação de algoritmos para processamento de instruções específicas em hardware não é recente, a transcrição dos algoritmos de ensemble para linguagem de hardware será uma maneira de reduzir o custo de processamento e até mesmo possibilitar que os métodos de ensemble sejam aplicados facilmente em massivas bases de dados. Transcrever os algoritmos de ensemble para linguagem de hardware possibilita a realização de simulações do processamento, medindo o tempo utilizado para execução completa do método. Um estudo para identificação e viabilidade de paralelizar o processamento de funções distintas poderá reduzir drasticamente o tempo necessário para a execução completa dos métodos.*

**Palavras-Chave.** *clustering, ensemble clustering, paralelismo, hardware*

## 1. Introdução, Motivação e Fundamentação Teórica

O crescimento exponencial de informações públicas ou privadas dos diversos tipos e a massiva quantidade de dados em uma mesma base, impulsiona estudos para gerir soluções algorítmicas no processamento de dados. Técnicas de agrupamento de dados são utilizadas em diversas frentes tecnológicas, por exemplo, na internet [Dorogovtsev and Mendes 2013, Castells 2003], em mineração de dados e conceitos de big data [Shirkhorshidi et al. 2014]. Alguns estudos apontam que a quantidade de informação em formato digital poderá atingir um tamanho total de 44 zettabytes até 2020 [EMC 2014].

Clustering ou agrupamento de dados é uma técnica para criação de grupos de objetos, onde os objetos em um grupo são bastante semelhante e o mais dessemelhante possível dos outros grupos. A grande diversidade de informações caracterizou a criação de diversos tipos de algoritmos de agrupamento de dados, dificultando a identificação de qual algoritmo terá melhor processamento em uma determinada base de dados, ou seja, qual algoritmo trará um resultado de agrupamento mais assertivo [Gan et al. 2007]. Esta dificuldade na identificação do melhor algoritmo a ser aplicado impulsionou na criação da técnica de ensemble (combinação) que poderá ser utilizada para refinar o resultado da aplicação de algoritmos de clustering, aumentando assim a confiabilidade do resultado [Abdala 2010]. O processo exemplificado na figura 1.

O custo de processamento para aplicação de algoritmos de clustering em geral é  $O(n^2)$ , a aplicação de técnicas de ensemble nos resultados de clustering torna o processo ainda mais árduo. A massiva quantidade de informações distintas demonstra que o processamento destes

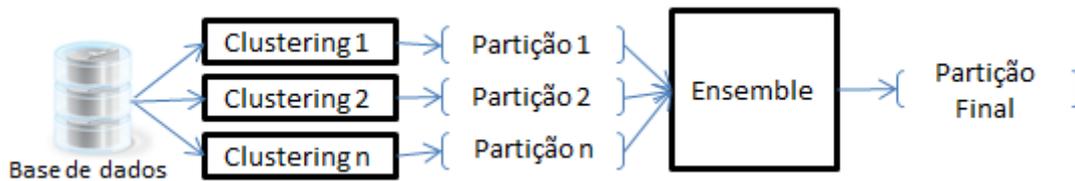


Figura 1. Processo de ensemble clustering

algoritmos em sistemas de propósito geral, por exemplo PC's [de Araujo 2003], expõem problemas com necessidade de alto tempo de processamento e custos financeiros com equipamentos de hardware relativamente alto. A alternativa para evitar estes custos e o alto tempo de processamento é a implementação destes algoritmos de ensemble em hardware, pois o processamento em sistemas de propósito geral tende a ser mais vagaroso comparado ao processamento em hardware [Hamada et al. 2009].

## 2. Contribuição do Trabalho

O processamento de métodos de ensemble clustering executados em computadores de propósito geral requer uma alta capacidade de processamento se aplicados em massivas base de dados, podendo até mesmo ser inviável o tempo necessário para aplicação destes métodos. A implementação destes algoritmos em hardware é uma forma de reduzir o tempo gasto em processamento e em alguns casos tornar possível a aplicação destes métodos em massivas base de dados. Além da implementação a aplicabilidade de paralelismo em hardware possibilita a execução de diversas funções simultaneamente, por exemplo, executar uma mesma função de cálculo de distância para diversas partições resultantes da aplicação de clustering ao mesmo tempo. [Anselmo et al. , Arias García 2011, Sotiropoulou et al. 2014]. As funções de cálculo de distâncias são utilizadas na aplicação das técnicas de ensemble [Goder and Filkov 2008].

## 3. Análise de Resultados

Os testes iniciais apontaram que a aplicabilidade dos métodos ensemble nas partições resultantes de clustering consumiu um baixo tempo de processamento. Utilizando o software Model-Sim [Altera 2015] foi possível realizar a simulação de processamento do método ensemble em hardware. A base de dados *EM\_1\_2D2k* e *EM\_1\_Yeast* possui dez partições resultantes da aplicação do método de clustering K-Means, cada partição contém o grupo em que o objeto  $N$  foi alocado [Abdala 2010].

- *EM\_1\_2D2k*: 1000 objetos;
- *EM\_1\_Yeast*: 1484 objetos.

O tempo de processamento medido pelo simulador foi inferior a um picosegundo, para o processamento correto das instruções no momento da simulação foi utilizada a função delta delay, esta função define um atraso mínimo de tempo (delay) para atualização dos sinais do circuito [Perry 2002]. Em continuidade da pesquisa será utilizado novas bases de dados com crescimento de tamanho exponencial, por exemplo, o número de objetos da base um seria 10000, base dois 100000 e assim sucessivamente até que seja possível registrar a variação do tempo gasto no processamento dos métodos ensemble.

## 4. Trabalhos Relacionados

A ideia de se implementar algoritmos em hardware não é nova, diversos estudos apontam resultados promissores na utilização deste meio de processamento, por exemplo, processamento de

sinal [Lyons 2010]. A utilização de hardware como meio de processamento para algoritmos de clustering mostrou-se um potencial meio para atingir ganhos de eficiência no processamento, no estudo "*Novel Dynamic Partial Reconfiguration Implementation of K-Means Clustering on FPGAs: Comparative Results with GPPs and GPUs*" [Hussain et al. 2012], a transcrição do algoritmo K-Means em linguagem HDL possibilitou a síntese no circuito FPGA e resultou na aceleração da análise dos dados, ultrapassando a limitação de processamento então imposta por computadores de propósito geral.

A aplicabilidade dos algoritmos de clustering em hardware demonstra ganhos em diversas áreas, segundo o estudo "*Dynamically and Partially Reconfigurable Hardware Architectures for High Performance Microarray Bioinformatics Data Analysis*" [Hussain 2012], Bioinformática e Biologia Computacional (BCB) lida com o armazenamento, organização, recuperação e análise de conjuntos de dados biológicos. Idealizado pelo problema da massiva quantidade de informações dos diversos tipos existentes em uma mesma base de dados a serem processadas, foi desenvolvido o processamento destas informações em hardware, que até então eram feitos por computadores de propósitos geral, um meio de processamento que tornou-se ineficiente devido a alta demanda de dados a serem processados.

Após uma extensiva busca não foi identificado estudos de implementação de algoritmos para aplicação de métodos ensemble em hardware.

## Referências

- Abdala, D. D. (2010). *Ensemble and Constrained Clustering with Applications*. PhD thesis, Universität Münster.
- Altera (2015). *Modelsim*. Altera.
- Anselmo, D., Martins, C., Soares, G., Itaúna, T., de Minas, P., and Gerais, I.-M. Arquitetura paralela reconfigurável em fpga para implementação de operadores elementares da aritmética intervalar.
- Arias García, J. (2011). Implementação em fpga de uma biblioteca parametrizável para inversão de matrizes baseada no algoritmo gauss-jordan, usando representação em ponto flutuante.
- Castells, M. (2003). *A Galáxia Internet: reflexões sobre a Internet, negócios ea sociedade*. Zahar.
- de Araujo, R. B. (2003). Computação ubíqua: Princípios, tecnologias e desafios. In *XXI Simpósio Brasileiro de Redes de Computadores*, volume 8, page 51.
- Dorogovtsev, S. N. and Mendes, J. F. (2013). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford University Press.
- EMC (2014). The digital universe of opportunities: Rich data and the increasing value of the internet of things. *EMC*.
- Gan, G., Ma, C., and Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics / American Statistical Association.
- Goder, A. and Filkov, V. (2008). Consensus clustering algorithms: Comparison and refinement. In *ALENEX*, volume 8, pages 109–117. SIAM.
- Hamada, T., Benkrid, K., Nitadori, K., and Taiji, M. (2009). A comparative study on asic, fpgas, gpus and general purpose processors in the  $O(n^2)$  gravitational n-body simulation. In *Adaptive Hardware and Systems, 2009. AHS 2009. NASA/ESA Conference on*, pages 447–452. IEEE.

- Hussain, H. M. (2012). Dynamically and partially reconfigurable hardware architectures for high performance microarray bioinformatics data analysis.
- Hussain, H. M., Benkrid, K., Ebrahim, A., Erdogan, A. T., and Seker, H. (2012). Novel dynamic partial reconfiguration implementation of k-means clustering on fpgas: comparative results with gpps and gpus. *International Journal of Reconfigurable Computing*, 2012:1.
- Lyons, R. G. (2010). *Understanding digital signal processing*. Pearson Education.
- Perry, D. L. (2002). *VHDL: programming by example*, volume 4. McGraw-Hill.
- Shirkhorshidi, A. S., Aghabozorgi, S., Wah, T. Y., and Herawan, T. (2014). Big data clustering: A review. In *Computational Science and Its Applications–ICCSA 2014*, pages 707–720. Springer.
- Sotiropoulou, C., Petridou, C., Nikolaidis, S., Annovi, A., Kordas, K., Gkaitatzis, S., Beretta, M., and Volpi, G. (2014). A parallel fpga implementation for real-time 2d pixel clustering for the atlas fast tracker processor. Technical report, ATL-COM-DAQ-2014-016.

# Um Método para Avaliação da Variabilidade de Conjuntos de Partições

**Autor: Nayara G. Ribeiro**

**Orientador: Daniel D. Abdala**

Graduação em Ciência da Computação  
Universidade Federal de Uberlândia – Uberlândia, MG – Brasil

[nayarag.ribeiro@gmail.com](mailto:nayarag.ribeiro@gmail.com), [abdala@facom.ufu.br](mailto:abdala@facom.ufu.br)

Nível: Trabalho de Conclusão de Curso

**Resumo.** *Este trabalho tem como objetivo produzir um algoritmo para geração automática de conjuntos de partições para servir como etapa de geração de dados em processos de agrupamento de dados via combinação de partições. O algoritmo visa gerar um conjunto de partições utilizando diferentes algoritmos de agrupamento de dados, utilização de parametrizações distintas no processo de agrupamento, com o intuito de gerar o conjunto de partições o mais variável possível. Resultados experimentais demonstram a viabilidade da proposta.*

**Palavras-Chave:** *Ensemble Clustering, variabilidade, agrupamento*

## 1. Introdução, Motivação e Fundamentação Teórica

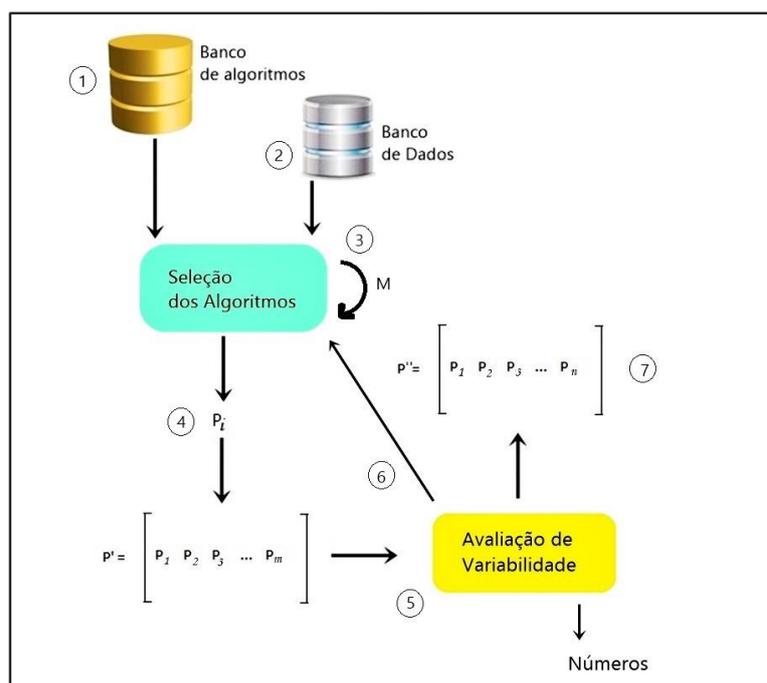
Ensemble clustering (combinação de partições) surgiu como uma opção de agrupamento de dados. Esta técnica é uma maneira de lidar com o problema da escolha do algoritmo de agrupamento em casos em que pouco ou nada se sabe sobre o conjunto de dados [Abdala 2010]. Ele também suaviza o resultado final quando partições diferentes apresentam distribuições consideravelmente distintas.

Neste contexto, este trabalho tem como objetivo desenvolver um método para análise de variabilidade intra-partições e recomendação de partições para fins de combinação de modo que a variabilidade seja maximizada ou minimizada.

Para realizar o agrupamento de dados, utilizamos algoritmos aproximados. Atualmente, há uma infinidade de algoritmos aproximados, que utilizam diferentes heurísticas para encontrar resultados sub ótimos. Estes algoritmos são desenvolvidos especialmente para conjuntos de dados específicos. Dado um novo conjunto a ser agrupado, não se sabe, a priori, qual o melhor algoritmo a ser aplicado.

A utilização da técnica de clustering determina o agrupamento intrínseco em conjuntos de dados. O processo de organização dos dados surge da semelhança entre os dados de alguma forma. Um cluster é, portanto, coleções de dados que são semelhantes entre eles, e que são diferentes dos dados pertencentes a outros clusters.

As etapas que serão seguidas para o desenvolvimento de um trabalho de seleção de partições que priorize aquelas que apresentam maior variabilidade entre si, podem ser observadas na figura 1.



**Figura 1.** Visão de alto nível das etapas do processo de ensemble clustering.

Fazem-se necessários dois bancos de dados, um de algoritmos de clustering, conforme etapa 1, e outro de dados, conforme etapa 2. Os dados são processados pela base de algoritmos, de acordo com etapa 3, e após  $M$  iterações serão geradas  $M$  partições  $P_i$ . Cada partição gerada, conforme etapa 4, será colocada em uma matriz  $P'$ . A matriz  $P'$  é avaliada na etapa 5, onde são selecionadas as partições com mais alta variabilidade. Nesta etapa, caso uma partição não tenha uma de boa variabilidade, seleciona-se um novo algoritmo para a geração de uma nova partição para substituí-la. Finalmente na etapa 7 uma matriz  $P''$  é produzida com mais alta variabilidade.

A escolha dos algoritmos na etapa 3 é dada aleatoriamente entre algoritmos de clustering dos tipos: algoritmos aglomerativos, algoritmos de densidade e algoritmos hierárquicos. Caso o algoritmo selecionado não garanta uma partição com uma boa variabilidade, o algoritmo prevê a troca do algoritmo para geração de uma nova partição. A permuta deste algoritmo pode ser por outro algoritmo do mesmo conjunto, ou pode ser trocado por um algoritmo dos outros conjuntos.

## 2. Contribuição do Trabalho

O desenvolvimento de um algoritmo de seleção de partições que priorize aquelas que apresentam maior variabilidade entre si resolve o problema de bias introduzido pela pré-existência de partições similares no ensemble.

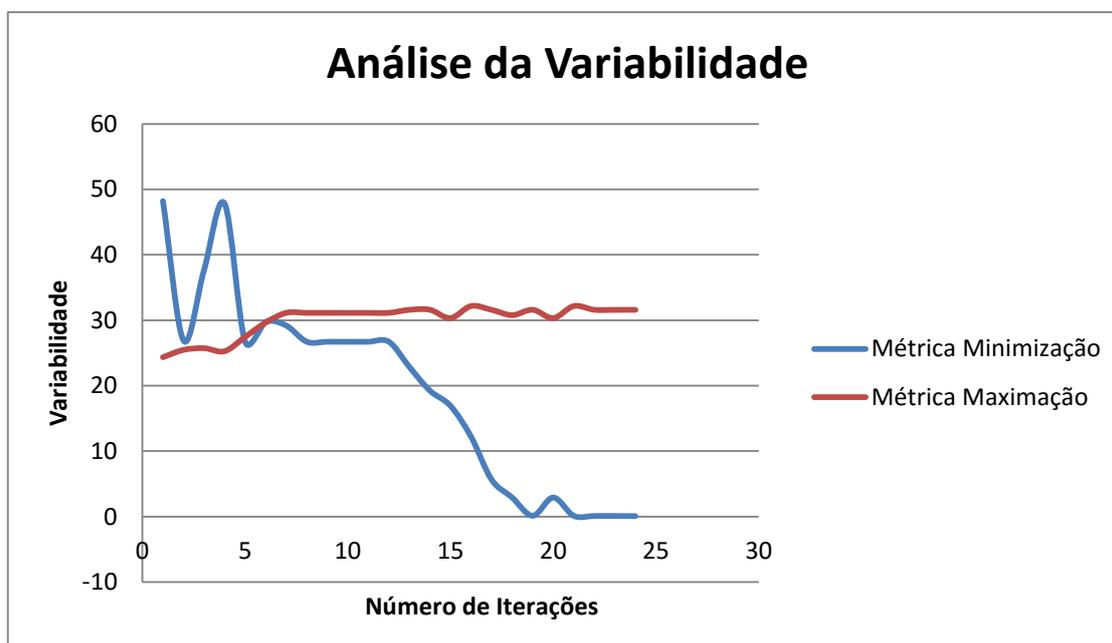
Estes métodos de agrupamento de dados via combinação de partições, encontram aplicações nos mais diversos segmentos da ciência. Métodos confiáveis de seleção de conjunto de partições diminuirão problemas de bias. Em estatística o conceito de bias está associado à diferença entre a média dos resultados e o valor verdadeiro. Ou seja, é a diferença entre o valor esperado e o valor produzido pelo estimador.

Técnicas de agrupamento de dados visam combinar partições geradas a partir de diversos algoritmos com o intuito de eliminar a necessidade de um estudo prévio acerca dos dados a serem agrupados de modo a definir qual o algoritmo de agrupamento mais adequado. Embora esta seja uma área ativa e recente em reconhecimento de padrões pouco se sabe acerca do impacto da variabilidade do ensemble no resultado final. Acredita-se que este trabalho se configurará como uma contribuição válida de modo a se entender melhor as restrições e áreas de aplicação da metodologia de agrupamento de dados via combinação de partições.

### 3. Análise de Resultados

Atualmente, as etapas de pesquisa bibliográfica e modelagem do sistema foram finalizadas. Os algoritmos que comporão a base de algoritmos já foram pré-selecionados, foram prospectados conjuntos de dados que comporão a base de dados de testes, e, também desenvolvido os algoritmos para compor a matriz de partições e a analisar a variabilidade das partições.

A etapa em desenvolvimento compõe a matriz de partições escolhendo M algoritmos aleatoriamente e executa uma métrica para análise da variabilidade entre as partições. Caso o algoritmo selecionado não garanta uma partição com uma boa variabilidade, de acordo com a métrica escolhida, o algoritmo é então trocado para geração de uma nova partição.



**Gráfico 1:** Resultado da execução de duas métricas diferentes para análise da variabilidade.

As métricas de variabilidade podem maximizar ou minimizar a variabilidade dependendo do método escolhido para avaliação. Cada métrica avalia uma partição com todas as partições pertencentes a matriz e tem como resultado números. Os resultados dessas aplicações podem ser analisados de duas maneiras como maximização e minimização. O intuito da execução de uma métrica de maximização é aumentar a

variabilidade das partições, já à execução de uma métrica de minimização é diminuir a variabilidade das partições.

Com os resultados preliminares demonstrados no gráfico 1, chegamos a conclusão preliminar que a execução do algoritmo maximiza/minimiza a variabilidade intra-partições. O próximo passo para compor os resultados finais, é avaliação do método proposto, ou seja, dado os ensembles com alta variabilidade, serão aplicados métodos de ensemble clustering para verificar se eles realmente geram resultados superiores aos obtidos por métodos ingênuos de geração de partições.

## **4. Trabalhos Relacionados**

### **4.1- A Mixture Model for Clustering**

Inúmeros algoritmos de agrupamento são capazes de produzir partições diferentes dos mesmos dados que capturam vários aspectos distintos dos dados [TOPCHY; PUNCH,2004]. O foco nesse artigo é a pesquisa em ensemble clustering, buscando uma combinação de múltiplas partições que proporcionam maior agrupamento geral dos dados fornecidos. No mesmo, relata sobre a maior dificuldade em encontrar uma partição consenso das partições de saída dos vários algoritmos de agrupamento. Outra questão difícil é a escolha do algoritmo de agrupamento para o conjunto.

### **4.2- Combining Multiple Clusterings Using Evidence Accumulation**

O objetivo do agrupamento é particionar um conjunto de objetos não rotulados em grupos homogêneos ou clusters [FRED; JAIN, 2005]. Nesse artigo é relatada a existência de centenas de algoritmos de agrupamento, e a produção de resultados distintos até para o mesmo algoritmo. A abordagem proposta é o conceito de evidência de acumulação de clustering, que mapeia as partições de dados individuais em um conjunto de cluster em uma nova medida de similaridade entre os padrões.

## **Referências**

- Jain M.N. Murty, P. F. A. Data clustering: A review ,1999.
- Abdala, D. D. Ensemble and constrained clustering with applications. 2010.
- TOPCHY, A. K. J. A.; PUNCH, W. A mixture model for clustering ensembles. 2004.
- FRED,A. L.; JAIN, A. K. Combining multiple clusterings using evidence accumulation. 2005.

**Parte II**  
**Trabalhos de Pós-Graduação**



# Abordagem de Recomendação Automática de Objetos de Aprendizagem Utilizando Tecnologias de Web Semântica Aplicadas ao Sistema Educacional Classroom eXperience

**Autor:** Vitor César de Carvalho<sup>1</sup>,  
**Orientador:** Fabiano Azevedo Dorça<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

vitorcarvalho@mestrado.ufu.br, fabianodor@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 03/2015

Mês e ano de previsão de conclusão: 03/2017

**Resumo.** *Este trabalho objetiva a construção de um modelo eficiente para recomendação automática de Objetos de Aprendizagem. Para tal, os estudantes serão modelados a partir do modelo Felder de Estilos de Aprendizagem e os Objetos de Aprendizagem devem ter seus metadados descritos no padrão IEEE LOM. O trabalho utiliza de técnicas de Web Semântica, modelando todo o processo através de Ontologias e inferindo o que é preciso via regras de inferência em SWRL. Além disso, esse modelo será aplicado ao Sistema Educacional Classroom Experience para que seja possível avaliar as diferenças de desempenho entre os alunos que puderam fazer uso do sistema de recomendação e daqueles que não tiveram habilitada essa função.*

**Palavras-Chave.** *web semântica, ontologias, informática para educação, ambientes virtuais de aprendizagem*

## 1. Introdução e Motivação

O uso de Objetos de Aprendizagem (OA) como conteúdo apresentado em cursos e aulas pela Web é uma característica amplamente adotada pelos Ambientes Virtuais de Aprendizagem. OAs são quaisquer entidades, digitais ou não, que possam ser utilizadas, reutilizadas ou referenciadas no ensino assistido por tecnologia. Este trabalho utiliza o padrão LTSC/IEEE *Learning Object Metadata* (LOM) como modelo de metadados para representação dos OAs. Assim como os OAs, os estudantes também possuem seu modelo de representação. Neste trabalho, os estudantes são modelados via Estilos de Aprendizagem (EA) de Felder-Silverman.

Este artigo apresenta uma abordagem eficiente para personalização do processo de ensino baseada no mapeamento de características de EAs em metadados de OAs através da utilização de Ontologias e outras tecnologias da Web Semântica. São consideradas as preferências dos estudantes para fazer a ligação entre o padrão de metadados LOM e os EA's. Além disso, espera-se que essa abordagem seja implementada e validada no Classroom eXperience, Ambiente Virtual de Aprendizagem ubíquo que é utilizado em algumas disciplinas da Faculdade de Computação e seus programas de pós-graduação da Universidade Federal de Uberlândia.

## 2. Fundamentação Teórica

Dentre os conceitos mais importantes a serem conhecidos para um bom entendimento do trabalho, estão a Web Semântica, os estilos de aprendizagem e objetos de aprendizagem; estes dois últimos comentados na seção 1.

A Web Semântica é a parte da Web que provê métodos e padrões que nos permitem atribuir significado as informações presentes na Web convencional. Suas funcionalidades e tecnologias estão dispostas por suas camadas, apresentadas, que são: Unicode, XML, RDF, Ontologia, Lógica, Prova e Confiança.

As funcionalidades que a Web Semântica nos provê são interessantes por auxiliar tanto na escolha dos dados recuperados, quanto na utilização da informação recuperada na Web. Dentre as suas tecnologias mais importantes, estão o *Resource Description Framework* (RDF) da camada RDF que define mais semântica ao documento e a *Web Ontology Language* (OWL) da camada de Ontologias. O uso de Ontologias no sistema, permite que sejam utilizadas regras em SWRL, que é uma linguagem com grande poder de inferência.

### 3. Trabalhos Relacionados

Diversos são os trabalhos que utilizam de Ontologias e outros recursos da Web Semântica para melhorar abordagens existentes em trabalhos na área de Informática para Educação. Gasparini et al. 2011 por exemplo, utilizou de Ontologias para melhorar alguns aspectos AdaptWeb.

Outro bom exemplo é o do trabalho de Rezende et al. 2015 que também criou uma abordagem de recomendação de Objetos de Aprendizagem baseada na utilização de Ontologias. Porém, enquanto Rezende et al. 2015 utiliza do contexto e de um modelo de estudante com informações diversas para recomendação, este trabalho se baseia na utilização de regras de inferência construídas na linguagem SWRL e baseadas num mapeamento obtido entre o padrão IEEE LOM de OA e o modelo de Felder de EA.

### 4. Estado Atual do Trabalho e Contribuição Esperada

O trabalho se encontra em desenvolvimento, de sua construção foi criada até o momento a Ontologia que servirá como base para todo o sistema. Dando subsídio para realização do mapeamento entre os Objetos de Aprendizagem e os Estilos de Aprendizagem e também para a recomendação destes OA.

Este trabalho tem como principal contribuição permitir representar o conhecimento sobre os OA em relação aos EA que atendem de forma eficiente, e então realizar inferências sobre os OA, gerando representações vetoriais que podem ser facilmente utilizadas para descobrir o quanto um OA atende às preferências de determinado estudante. Com essa representação vetorial, vários trabalhos serão facilitados, como por exemplo, aqueles que envolvem a utilização de algoritmos de Mineração de Dados, como a clusterização de Objetos de Aprendizagem.

### Referências

- Gasparini, I., Kemczinski, A., Pimenta, M. S., and de Oliveira, J. P. M. (2011). Modelo do usuário sensível ao contexto cultural em um sistema e-learning adaptativo. *Informática na educação: teoria & prática*, 14(1).
- Rezende, P. A. A., Pereira, C. K., Campos, F., David, J. M. N., and Braga, R. (2015). Persona: proposta de ontologia de contexto e perfil de alunos para recomendação de objetos de aprendizagem. *Revista Brasileira de Informática na Educação*, 23(01):70.

# Ambiente web para citomorfometria usando imagens 3D

**Autor: Eduardo Henrique Silva,**  
**Orientador: Bruno Augusto Nassif Travençolo,**  
**Coorientador: Jefferson Rodrigo de Souza**

Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

eduardo.hssi@gmail.com, travencolo@gmail.com, jrsouza@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 03/2015

Mês e ano de previsão de conclusão: 03/2017

Data de aprovação da proposta de dissertação: 19/02/2016

**Resumo.** *A instalação e configuração de um bioimaging software pode gerar empecilhos em sua utilização quando cientistas com pouco conhecimento em informática deseja utilizá-lo. A disponibilização de um software na web pode facilitar e impulsionar sua utilização na comunidade, permitindo o acesso em qualquer lugar e em qualquer hora por meio da internet. Com o ambiente web proposto pretende-se auxiliar patologistas e outros usuários na tomada de decisões, minimizando a subjetividade de suas avaliações. Para validá-ló serão analisadas amostras de carcinomas de colo de útero, espera-se que a análise dos dados obtidos a partir da citomorfometria em imagens tridimensionais reconstruídas com algoritmos de visualização volumétrica possa prever a graduação de um tumor.*

**Palavras-Chave.** *Visualização Científica, Citomorfometria, Software de Bioimaging*

## 1. Introdução

Os *bioimaging softwares* fornecem métodos computacionais para a manipulação de imagens microscópicas. A maioria desses softwares são executadas como sistemas locais. Um estudo publicado na *Nature Methods* apresentou que a usabilidade nos *bioimaging software* tem sido um desafio e pode ser um fator limitante para o seu uso, pois cientistas com conhecimento básico em computação devem ser capazes de realizarem os seus experimentos no software disponibilizado. Características como a facilidade na instalação, suporte a várias plataformas, atualização automática e integração com outros softwares podem contribuir com a usabilidade. O estudo ainda ressalta que a disponibilização do software na nuvem evita muitos problemas de instalação e possibilita sua utilização remota, além de permitir o compartilhamento dos resultados com outros cientistas, quebrando barreiras geográficas [Cardona and Tomancak 2012].

Assim, pretende-se nessa dissertação desenvolver um ambiente para análise citomorfométrica em imagens extraídas de microscópicos por meio da reconstrução 3D, permitindo a exploração das informações no navegador web. Além disso, serão feitas análises em amostras de células cancerígenas do colo do útero utilizando o ambiente proposto. Por meio da reconstrução 3D e da extração dos dados quantitativos, espera-se fornecer informações relevantes para auxiliar o patologista sobre a graduação do tumor presente nas amostras analisadas e também validar o ambiente proposto.

## 2. Fundamentação Teórica

### 2.1. Visualização Científica (ViSC)

A ViSC é uma disciplina da visualização da informação a fim de aplicar técnicas para permitir a transformação de dados abstratos de observações em modelos que podem ser facilmente compreensíveis para a exploração científica [Hansen et al. 2014]. Os algoritmos de visualização volumétrica são capazes de criar uma projeção tridimensional em um plano de imagens bidimensionais (e.g. Figura 1(a)) para compreensão e exploração dos dados volumétricos, eles estão divididos em duas categorias [Elvins 1992]:

- **Renderização Volumétrica Direta (DVR):** o DVR é capaz de criar imagens de um conjunto de dados que não apresentam uma forma geométrica definida. Assim, ele realiza um mapeamento dos elementos diretamente no plano sem utilizar a geometria primitiva para criar a representação (e.g. Figura 1(b)). O algoritmo *Volume Ray-casting* é amplamente utilizado em imagens microscópicas [Elvins 1992].
- **Ajuste de superfície (SF):** o ajuste de superfície é capaz de extrair uma superfície em um volume por meio do ajuste de primitivas geométricas, como polígonos, para contornos de superfícies de valor constante (e.g. Figura 1(c)). O *Marching Cubes* é um método de ajuste de superfície amplamente implementado em imagens médica. Ele cria uma malha de triangulações que representam uma isosuperfície [Lorensen and Cline 1987].

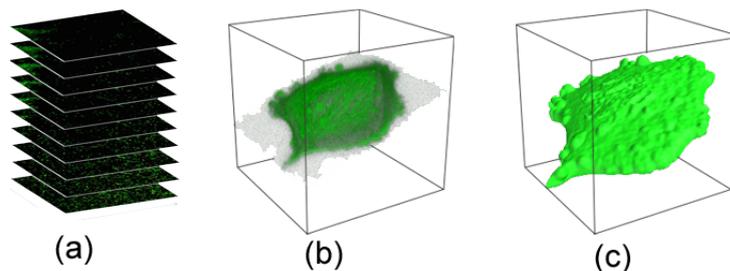


Figura 1. Reconstrução tridimensional com o software BiImageXD. (a) Imagens fatiadas. (b) Visualização de volume. (c) Isosuperfície.

O método de *Marching Cubes* possui um desempenho melhor em relação ao método de *Volume Ray-casting*, pois ele não percorre todo o conjunto de dados a cada vez que a imagem é renderizada. Porém, o método *Marching Cubes* pode apresentar isosuperfícies não tão representativas de acordo com o modelo real, devido a reconstrução com falsos pedaços negativos e positivos de superfícies [Elvins 1992].

### 2.2. Análise citomorfométrica

A morfometria é um conjunto de técnicas que permite descrever quantitativamente características estruturais por meio da extração das informações, como, área da superfície, volume, circularidade e curvatura [Weibel 1969]. Na patologia, alterações morfológicas nas estruturas celulares são importantes no auxílio dos métodos de diagnósticos. A morfometria citológica ou citomorfometria é um conjunto de técnicas de morfometria para descrever características dos componentes celulares [Waldemarin et al. 2004].

Para a análise citomorfométrica os atributos como, área da superfície, volume, esfericidade, diâmetro esférico equivalente, coeficientes de curvatura e convexidade podem serem extraídos de uma isosuperfície, permitindo quantificar os objetos de interesses.

### 3. Contribuição do Trabalho

Pretende-se contribuir com um ambiente web de visualização 3D e análise citomorfométrica em imagens microscópicas, e com resultados das análises morfométricas em amostras de carcinomas de colo de útero.

### 4. Estado atual do Trabalho

Para as análises em amostras de células cancerígenas do colo do útero será utilizado uma base de dados com 34 amostras de carcinomas fornecidas pelo Laboratório de Patologia do Hospital de Clínicas da Universidade Federal do Triângulo Mineiro, conservadas em blocos de parafina. As amostras estão em fase de digitalização utilizando microscópio confocal, uma amostra foi disponibilizada para auxiliar no desenvolvimento do ambiente.

O ambiente está sendo desenvolvido utilizando a arquitetura cliente/servidor com o VTK. O VTK é uma biblioteca *open-source* para processamento e visualização 3D utilizada em diversos segmentos, tais como, visualização médica e visualização biológica. Aplicações de visualização na web que realizam o processamento no lado do cliente são ineficientes quando grandes conjuntos de dados são renderizados [Jourdain et al. 2010]. No ambiente proposto todo processamento ocorre no servidor, e é utilizada a estratégia de carregamento progressivo, ou seja, quando um cliente solicita um objeto tridimensional o servidor envia pequenos blocos de dados, o cliente vai recebendo os blocos e apresentando para o usuário. Isso evita que a biblioteca JavaScript no lado do cliente possa sobrecarregar e não apresentar corretamente ao usuário o modelo 3D.

No ambiente proposto o maior espaço da página principal é destinada à visualização, permitindo uma maior liberdade ao usuário ao interagir com a amostra em análise. As páginas web são adaptáveis à diferentes tamanhos de tela, em que a maior proporção da tela é também destinada a visualização. A Figura 2 apresenta a página de visualização e análise.

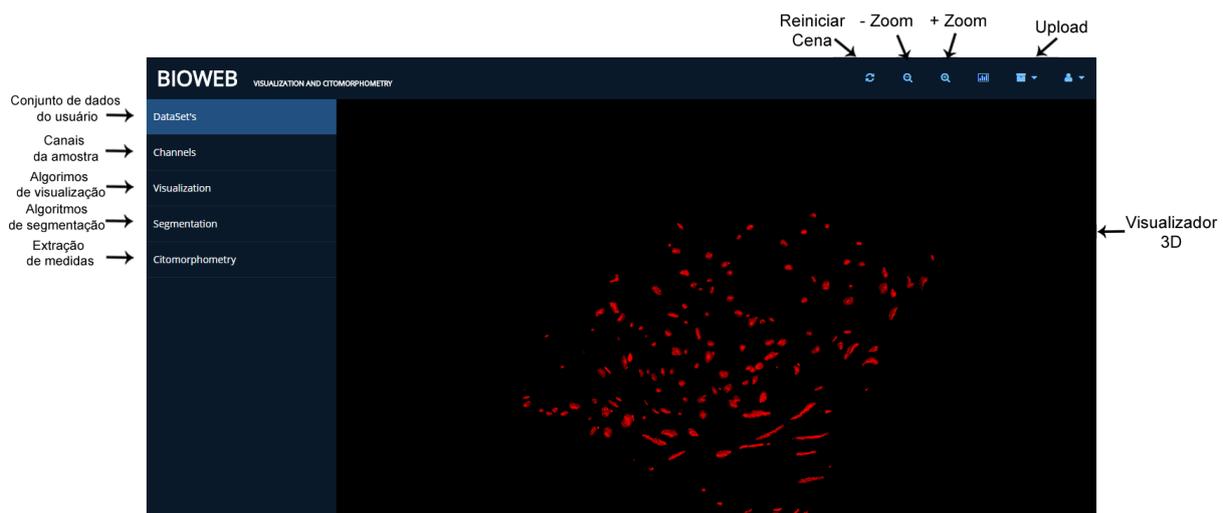


Figura 2. Visão geral do BioWeb.

Após efetuar o *upload*, as amostras ficam disponíveis para visualização ou exclusão (Figura 3(d)). Ao selecionar uma amostra para visualização é feita uma solicitação de reconstrução tridimensional ao servidor. Na reconstrução tridimensional pode-se utilizar os algoritmos *Volume Ray-casting* e *Marching Cubes* (e.g. Figura 3(a)).

Alguns microscópicos fornecem imagens com múltiplos canais, o ambiente permite a visualização de cada canal separadamente também permite usuário selecionar os canais desejados para gerar a visualização 3D (e.g. Figura 3(e)). Segmentação, filtragem e citomorfometria é realizada em cada canal separadamente.



**Figura 3. Funcionalidades para ViSC e citomorfometria. (a) Seleção e configuração dos algoritmos de ViSC. (b) Extração das medidas utilizando métodos de citomorfometria. (c) Medidas extraídas. (d) Seleção do conjunto de dados. (e) Escolha dos canais. (f) Segmentação.**

O filtro Gaussiano é utilizado para suavizar (*smoothing*) as imagens, reduzindo possíveis ruídos (e.g. Figura 3(a)). Esses ruídos podem interferir nos resultados, pois pequenas triangulações serão consideradas como objetos e suas medidas serão extraídas. Para segmentação é fornecido o método de limiarização por meio da escolha de um limiar mínimo e máximo (e.g. Figura 3(f)), pretende-se fornecer mais métodos para segmentação.

O ambiente realiza a extração de dados quantitativos utilizando a superfície obtida com o algoritmo *Marching Cubes* (e.g. Figura 3(b)). Na mensuração das características o filtro de conectividade é utilizado para separar os objetos e então cada superfície do objeto é submetido às técnicas de citomorfometria. Os resultados obtidos podem ser armazenados, exportados para um arquivo CSV e visualizados no próprio ambiente (Figura 3(c)).

As interações com a cena são realizadas por meio da barra de ferramentas superior ou por meio do mouse. Na barra de ferramentas é possível aumentar e diminuir o zoom e também reiniciar a câmera da cena em seu estado original. O visualizador 3D permite por meio do mouse visualizar a cena em diferentes perspectivas, aumentar e diminuir o zoom e selecionar um objeto.

Ao visualizar uma superfície é possível selecionar um objeto de interesse de duas formas: pelas coordenadas  $x$  e  $y$  obtidas por um clique duplo ou pela extração das medidas de citomorfometria. Após encontrar o objeto de interesse, o ambiente redireciona o usuário para uma nova página, conforme a Figura 4.

Na visualização do objeto de interesse todas as medidas são apresentadas na barra la-

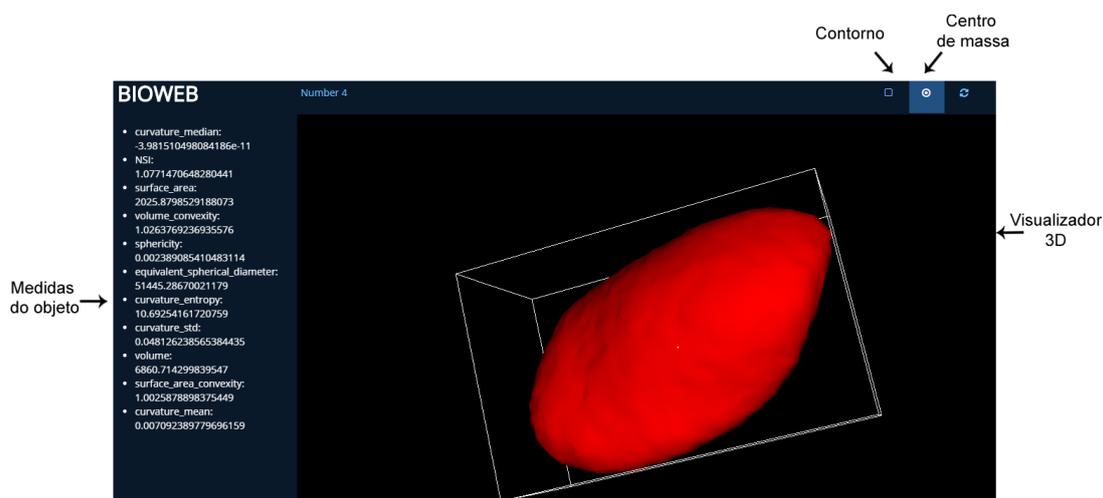


Figura 4. Visualização de um objeto de interesse.

teral, além de interações com o objeto é possível adicionar um delimitador de fronteira e um indicador do centro de massa do objeto.

## 5. Análise de Resultados

Nenhum resultado ainda pode ser inferido, pois as amostras estão em fase de digitalização. Para afirmar se os algoritmos disponibilizados pelo ambiente podem auxiliar os patologistas será necessário a comprovação dos resultados obtidos por meio da análise estatística sobre os dados quantitativos extraídos com a citomorfometria.

## 6. Trabalhos Relacionados

O ambiente proposto pode ser comparado com os trabalhos relacionados: Bisque [Kvilekval et al. 2010], OMERO [Allan et al. 2012], BioIMAX [Loyek et al. 2011] e Slicer Drop [Haehn 2013]. Esses trabalhos possuem mecanismos de visualização 3D na web com o foco em imagens microscópicas ou médicas.

A Tabela 1 apresenta uma comparação entre o estado atual do trabalho proposto com os trabalhos relacionados.

Tabela 1. Comparação do ambiente proposto com outros trabalhos.

Função	Ambiente Proposto	Bisque	Omero	BioIMAX	Slicer Drop
Tecnologia web de visualização	WebGL	WebGL	JavaScript	Adobe Flex	WebGL
Qtd. de formatos de imagens suportadas	+ de 100	+ de 100	+ de 100	3	+ de 10
Integração com outros ambientes	Não	Sim	Sim	Não	Não
Segmentação 3D	Sim	Sim	Não	Não	Sim
Separação de objetos 3D	Sim	Não	Não	Não	Não
Redução de ruídos	Sim	Não	Não	Não	Não
Renderização Volumétrica	Sim	Sim <sup>1</sup>	Não	Não	Sim <sup>1</sup>
Extração de isosuperfície	Sim	Não	Não	Não	Não
Multicanais	Sim	Sim	Não	Não	Não
Interação com o usuário	Sim	Sim	Sim	Sim	Sim
Suporte à tablets e smartphones	Sim	Sim	Sim	Não	Sim
Métodos de análise citomorfometria	14	2	0	0	0

<sup>1</sup> Apresentou erros ao realizar renderização volumétrica de grandes quantidades de dados.

O Bisque é um ambiente web para gestão, compartilhamento, análise e extração de dados quantitativos em imagens multidimensionais [Kvilekval et al. 2010]. O OMERO é uma ferramenta para gerenciamento de dados biológicos e o seu foco não é a análise e nem a visualização [Allan et al. 2012]. O BioIMAX é um ambiente web desenvolvido em Adobe Flex voltado para exploração e mineração de dados, necessitando do *plug-in* do Adobe Flash instalado no navegador. Ele permite visualizar imagens em planos bidimensionais e alguns navegadores não possuem mais suporte ao Adobe Flash [Loyek et al. 2011]. O Slicer Drop é um ambiente voltado para imagens médicas que utiliza o XTK (biblioteca JavaScript de renderização volumétrica) [Haehn 2013].

## Referências

- Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., MacDonald, D., Moore, W. J., Neves, C., Patterson, A., et al. (2012). Omero: flexible, model-driven data management for experimental biology. *Nature methods*, 9(3):245–253.
- Cardona, A. and Tomancak, P. (2012). Current challenges in open-source bioimage informatics. *Nature methods*, 9(7):661–665.
- Elvins, T. T. (1992). A survey of algorithms for volume visualization. *ACM Siggraph Computer Graphics*, 26(3):194–201.
- Haehn, D. (2013). Slice:drop: Collaborative medical imaging in the browser. In *ACM SIGGRAPH 2013 Computer Animation Festival, SIGGRAPH '13*, pages 1–1, New York, NY, USA. ACM.
- Hansen, C., Chen, M., Johnson, C., Kaufman, A., and Hagen, H. (2014). *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*. Mathematics and Visualization. Springer London.
- Jourdain, S., Ayachit, U., and Geveci, B. (2010). Paraviewweb: A web framework for 3d visualization and data processing. In *IADIS international conference on web virtual reality and three-dimensional worlds*, volume 7, page 1.
- Kvilekval, K., Fedorov, D., Obara, B., Singh, A., and Manjunath, B. (2010). Bisque: a platform for bioimage analysis and management. *Bioinformatics*, 26(4):544–552.
- Lorensen, W. E. and Cline, H. E. (1987). Marching cubes: A high resolution 3d surface construction algorithm. In *ACM siggraph computer graphics*, volume 21, pages 163–169. ACM.
- Loyek, C., Rajpoot, N. M., Khan, M., and Nattkemper, T. W. (2011). Bioimax: A web 2.0 approach for easy exploratory and collaborative access to multivariate bioimage data. *BMC bioinformatics*, 12(1):297.
- Waldemarin, K. C. A., Beletti, M. E., and Costa, L. F. (2004). Nuclear morphometry of neoplastic cells as a method for diagnosis of histiocytoma, mastocytoma and transmissible venereal tumor in dogs. *Real-Time Imaging*, 10(4):197–204.
- Weibel, E. R. (1969). Stereological principles for morphometry. *Int rev cytol*, 26:235–302.

# Aplicação de Redes Complexas no Estudo de Interação Proteína-Proteína

**Autor: Getúlio de Moraes Pereira<sup>1</sup>,  
Orientador: Prof. Dr. Anderson Rodrigues dos Santos<sup>1</sup>**

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

getulio@ufu.br, santosardr@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 08/2015

Mês e ano de previsão de conclusão: 08/2017

**Resumo.** *A identificação de proteínas alvo para a produção de fármacos utiliza frequentemente a análise PPI como um modelo computacional em Rede Complexa. A aplicação de medidas de centralidade sobre uma PPI possibilita a identificação de aglomerados proteicos e, como consequência, apontar proteínas alvo para tal finalidade. Por meio de bridging centrality e algoritmos de Inteligência Artificial, pretende-se realizar a análise de redes PPI geradas ab initio.*

**Palavras-Chave.** *Redes Complexas, Bioinformática, Interação Proteína-Proteína, Inteligência Artificial*

## 1. Introdução e Motivação

A identificação de proteínas alvo para a produção de fármacos utiliza frequentemente a análise PPI (do inglês *Protein-Protein Interaction*) [Ananthasubramanian et al. 2012, Liu et al. 2012], em que as relações de interação proteína-proteína consideram características como: (i) a proximidade de seus genes em um cromossomo; (ii) seus genes estão fundidos de modo a facilitar a produção e interação de seus produtos proteicos; (iii) evoluem conjuntamente; (iv) são expressas conjuntamente; (v) existem no mesmo local subcelular; (vi) estão associados à mesma função molecular; e (vii) participam do mesmo processo biológico [Von Mering et al. 2007, Consortium et al. 2004].

Uma das potencialidades de PPIs é identificar quais seriam as consequências de se ativar/desativar a expressão de um determinado gene em um organismo. No entanto, as técnicas *in vitro* e *in vivo* demandam grande quantidade de recursos, produzindo apenas uma fração das PPIs de uma bactéria sob estudo [Braun 2012]. Uma alternativa é a predição *in silico* de PPIs [Larsen et al. 2012] em que modelos computacionais de predição de PPIs podem ser reaproveitados entre organismos evolutivamente próximos [Ananthasubramanian et al. 2012]. Esse resultado abre a perspectiva de reutilizar o conhecimento a respeito de organismos para os quais as PPIs sejam conhecidas para outros sobre os quais ainda não exista informação experimental. Tal análise baseia-se na hipótese de que proteínas envolvidas em um mesmo processo biológico possuem algumas das características acima relatadas (i até vi) com grande similaridade. Dessa forma, quando tais proteínas são representadas graficamente em um espaço multidimensional (por meio de uma Rede Complexa), estariam propensas a formarem aglomerados mais compactos do que aquelas que não pertencem ao mesmo processo biológico. Atualmente existe

uma gama de algoritmos para identificar tais aglomerados como, por exemplo, a Clusterização Baseada na Densidade e a Clusterização Hierárquica que apresentam boa acurácia preditiva em redes de PPIs [Zhang 2009].

Os grafos produzidos como consequência da geração de PPIs de um proteoma são complexos por possuírem uma quantidade relevante de caminhamentos circulares, tornando difícil perceber onde começa e onde termina um possível processo biológico. Nos grafos construídos a partir de milhares de proteínas a quantidade de ciclos aumenta consideravelmente, e a análise de suas interações mais simples tende a tornar-se inviável, mesmo com o uso dedicado dos computadores mais velozes. Essa complexidade dificulta o uso de grafos oriundos de PPIs para o mapeamento de relações menos óbvias, aquelas que não estão nas vizinhanças imediatas das proteínas [Zhang 2009].

Para diminuir a complexidade da análise de PPIs, propõe-se a utilização do conceito de *bridging centrality* [Hwang et al. 2006] para definir um arcabouço central de vértices do PPI em estudo. Apesar da análise de aglomerados sugerir proteínas relacionadas a um mesmo processo biológico, não fica evidente, por exemplo, quais proteínas poderiam estar associadas a processos mais complexos, ou seja, que utilizariam proteínas presentes em diferentes aglomerados proteicos.

Este trabalho visa analisar a aplicação da técnica *bridging centrality* em redes PPI, além de criar um mecanismo de Inteligência Artificial para aproximar os cálculos exatos gerados pela *bridging centrality*. Conhecidos os nós de uma rede complexa com a característica de servirem de ponte entre aglomerados proteicos, é possível a identificação de aglomerados proteicos que venham a ser candidatos em potencial para a produção de medicamentos [Hwang et al. 2008]. Esta estatística considera a quantidade de arestas adjacentes aos vértices da vizinha direta de um vértice sob análise, bem como a quantidade de caminhamentos passando por esse vértice, cujos valores são utilizados para computar os coeficientes *bridging* e *betweenness*, respectivamente [Hwang et al. 2006]. Na Figura 1, os pontos pretos representam proteínas com essa característica no grafo de PPI da levedura e é possível perceber como as proteínas com os maiores valores da estatística *bridging centrality*, denominadas *bridging nodes*, estariam conectando aglomerados de proteínas.

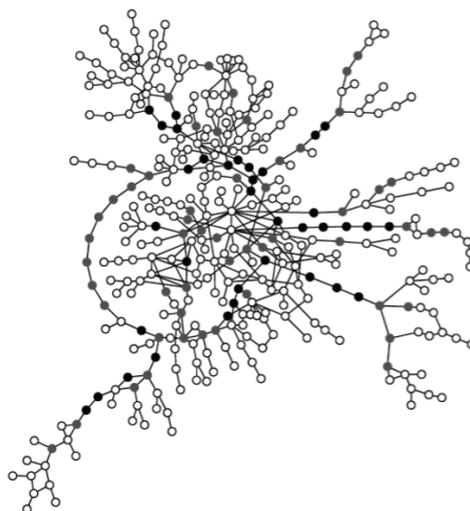


Figura 1. A rede metabólica da levedura com 359 nós e 435 arestas. Os nós com os maiores valores de *bridging centrality* são destacados em círculos pretos; os nós destacados em cinza seguem logo atrás situados entre os percentis 10 e 25. Adaptado de [Hwang et al. 2006].

## 2. Fundamentação Teórica

A literatura sobre Redes Complexas apresenta os seguintes modelos.

### 2.1. Grafos Randômicos de Erdős e Rényi

Este modelo é formado a partir de um conjunto de  $N$  vértices aos quais são adicionados  $L$  arestas aleatoriamente, segundo uma probabilidade  $p$  para cada nova aresta, evitando-se auto-conexões e conexões múltiplas, como ilustrado na Figura 2 (a). Para este modelo, a distribuição de graus segue a distribuição de Poisson (Figura 2 (b)). [Costa et al. 2007]

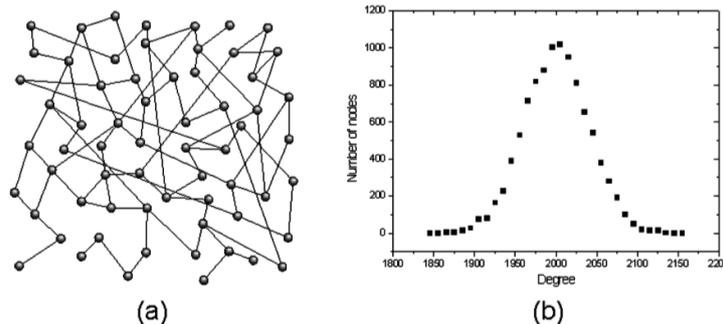


Figura 2. Grafo randômico de Erdős e Rényi: (a) exemplo de grafo randômico (b) distribuição média de grau sobre 10 grafos randômicos constituídos por 10.000 vértices, usando probabilidade  $p = 0.2$ . Fonte: [Costa et al. 2007]

### 2.2. O Modelo de Pequeno Mundo de Watts e Strogatz

Proposto por Watts e Strogatz, é o modelo de rede randômica mais popular que apresenta as características de pequeno mundo (*small world*) e uma grande quantidade de triângulos. A característica de pequeno mundo indica que a maioria dos vértices podem ser alcançadas, partindo-se de outros vértices, a poucas arestas de distância. Este conceito tem sua origem no experimento de Milgram, onde ele descreve que dois cidadãos norte-americanos tomados ao acaso estão conectados a 6 conhecidos de distância, em média. Um triângulo, em uma rede *small world*, indica que, se um nó  $i$  está ligado aos nós  $j$  e  $k$ , existe uma grande probabilidade de  $j$  e  $k$  também estarem ligados. [Costa et al. 2007]

Uma forma de se obter uma rede complexa desse modelo é, a partir de um reticulado regular (Figura 3 (a)), reconectar suas arestas, tomadas ao acaso, segundo uma probabilidade  $p$  (Figura 3 (b)).

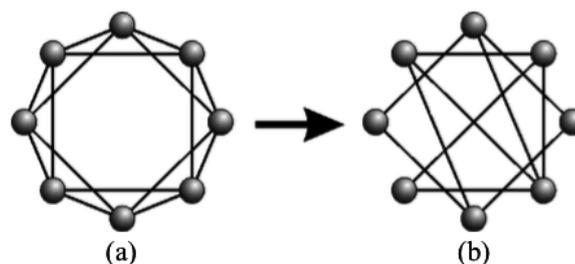
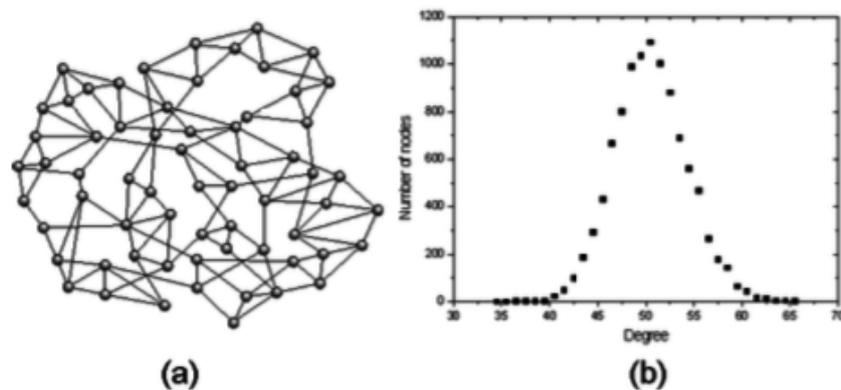


Figura 3. Obtenção de rede pequeno-mundo, de acordo com Watts and Strogatz: partindo-se de uma rede regular, reconectando-se suas arestas com probabilidade  $p$ . Se  $p \approx 0$  a rede é regular, apresentando muitos triângulos e grandes distâncias entre seus vértices. Se  $p \approx 1$ , tem-se uma rede randômica, com poucos triângulos e distâncias menores. Fonte: [Costa et al. 2007]

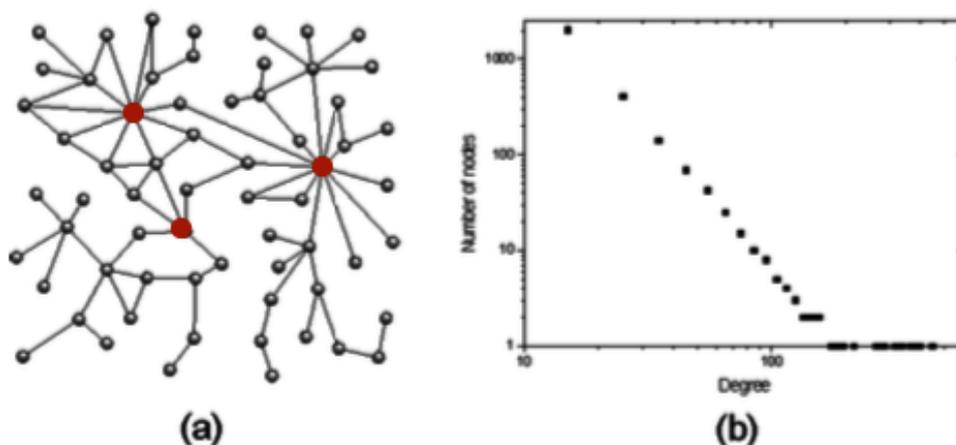
A Figura 4 (a) apresenta um exemplo de rede pequeno mundo contendo 64 vértices. A Figura 4 (b) ilustra que a distribuição de graus de 10 redes desse modelo é similar à distribuição de Poisson [Costa et al. 2007].



**Figura 4. Modelo de WS: (a) exemplo de rede WS com  $N = 64$  vértices, grau  $k = 2$  e probabilidade  $p = 0.1$  (b) distribuição média de grau para 10 redes WS, com 10000 vértices, grau  $k = 25$  e probabilidade  $p = 0.3$ . Fonte: [Costa et al. 2007]**

### 2.3. Redes Livres de Escala de Barabási e Albert

Barabási e Albert demonstraram que alguns sistemas reais apresentam a distribuição de nós de forma desigual, ou seja, alguns vértices são altamente conectados ao passo que outros possuem baixo número de ligações. De forma mais precisa, a distribuição de graus dos vértices dessas redes segue a Lei de Potência ( $P(k) \sim k^{-\lambda}$ ), como ilustra a Figura 5 (b). Outra característica marcante dessas redes é a presença de *hubs*. Um *hub* é um nó que está ligado a uma fração significativa de arestas da rede como um todo, o que é ilustrado pelos nós marcados na Figura 5 (a) [Costa et al. 2007].



**Figura 5. Rede livre de escala de Barabási and Albert. (a) exemplo de uma rede livre de escala (b) distribuição de grau médio para 10 redes livres de escala, com 10000 vértices. Adaptado de: [Costa et al. 2007]**

Várias redes biológicas seguem o modelo livre de escala, incluindo Redes PPI [Zhang 2009]. Nessas redes PPI, as proteínas identificadas como *hubs* são tidas como alvos promissores para produção de drogas. Identificar estes *hubs* é objetivo deste trabalho.

### 3. Estado atual do Trabalho

O projeto realizado por [Rodrigues 2013] teve como resultado a geração de PPI *ab initio*, ou seja, sem uso de homologia entre proteínas. A vantagem dessa técnica é identificar PPI's específicas de uma bactéria mesmo que essa PPI não esteja presente em organismos similares.

Nesse sentido, este projeto tem como objetivo dar continuidade aos trabalhos de [Rodrigues 2013], criando métodos de análise dessas PPI's utilizando *bridging centrality* além de algoritmos de Inteligência Artificial para tentar aproximar os cálculos exatos de *bridging centrality*.

A IA será utilizada em uma tentativa de mapear características de nós e arestas de uma rede complexa que possam servir como fonte de análise da *bridging centrality*. Para esse propósito pretende-se utilizar diversos algoritmos de aprendizagem como por exemplo Redes Neurais, Árvores de Decisão, Florestas Randômicas, entre outros, procurando o algoritmo que melhor se adapta ao problema em questão. Uma vez que este trabalho encontra-se em fase de planejamento, está em análise quais *features* de uma rede PPI serão objetos desse estudo.

### Referências

- Ananthasubramanian, S., Metri, R., Khetan, A., Gupta, A., Handen, A., Chandra, N., and Ganapathiraju, M. (2012). Mycobacterium tuberculosis and clostridium difficile interactomes: demonstration of rapid development of computational system for bacterial interactome prediction. *Microbial informatics and experimentation*, 2(1):1.
- Braun, P. (2012). Interactome mapping for analysis of complex phenotypes: insights from benchmarking binary interaction assays. *Proteomics*, 12(10):1499–1518.
- Consortium, G. O. et al. (2004). The gene ontology (go) database and informatics resource. *Nucleic acids research*, 32(suppl 1):D258–D261.
- Costa, L. d. F., Rodrigues, F. A., Travieso, G., and Villas Boas, P. R. (2007). Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242.
- Figueiredo, D. R. (2011). Introdução a redes complexas. *Atualizações em Informática*, pages 303–358.
- Hwang, W., Cho, Y.-r., Zhang, A., and Ramanathan, M. (2006). Bridging centrality: identifying bridging nodes in scale-free networks. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 20–23.
- Hwang, W.-C., Zhang, A., and Ramanathan, M. (2008). Identification of information flow-modulating drug targets: a novel bridging paradigm for drug discovery. *Clinical Pharmacology & Therapeutics*, 84(5):563–572.
- Larsen, P., Hamada, Y., and Gilbert, J. (2012). Modeling microbial communities: Current, developing, and future technologies for predicting microbial community interaction. *Journal of biotechnology*, 160(1):17–24.
- Liu, Z.-P., Wang, J., Qiu, Y.-Q., Leung, R. K., Zhang, X.-S., Tsui, S. K., and Chen, L. (2012). Inferring a protein interaction map of mycobacterium tuberculosis based on sequences and interologs. *BMC bioinformatics*, 13(Suppl 7):S6.
- Rodrigues, I. C. (2013). Mapeamento de interações proteína-proteína comuns a bactérias do grupo cmnr e análise de interações. Trabalho de conclusão de curso. (graduação em sistemas de informação)., Universidade Federal de Uberlândia.

Von Mering, C., Jensen, L. J., Kuhn, M., Chaffron, S., Doerks, T., Krüger, B., Snel, B., and Bork, P. (2007). String 7—recent developments in the integration and prediction of protein interactions. *Nucleic acids research*, 35(suppl 1):D358–D362.

Zhang, A. (2009). *Protein interaction networks: computational analysis*. Cambridge University Press.

# Um Método de Acesso Métrico Eficiente para Suporte às Consultas por Similaridade em Fluxo Contínuo de Dados.

**Autor: Régis Michel dos Santos Sousa<sup>1</sup>,  
Orientador: Prof. Dr. Humberto Luiz Razente<sup>1</sup>**

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

regismaicon@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 03/2015

Mês e ano de previsão de conclusão: 03/2017

Data de aprovação da proposta de dissertação/plano de tese: 19/02/2015

**Resumo.** *A busca e recuperação por similaridade são operações fundamentais para aplicações que lidam com dados complexos. Este trabalho tem como objetivo propor um método para busca por similaridade em ambiente de fluxo contínuo de dados que permita consultas por similaridade de modo eficiente. O método será dinâmico permitindo operações de inserção na base de dados e contará com uma estrutura auxiliar que contribuirá para minimizar a sobreposição da estrutura de indexação. As estruturas serão implementadas e o projeto está em fase inicial de desenvolvimento.*

**Palavras-Chave.** *Consultas por Similaridade, Fluxo Contínuo de Dados, Estrutura de Indexação, dados complexos.*

## 1. Introdução e Motivação

A nova forma de gerar e consumir dados contribui para o surgimento de sistemas que processam em *streams*, ou seja, em fluxos contínuos de dados (FCD). Nesse ambiente é necessário que os algoritmos tenham a capacidade de aprender continuamente para representar adequadamente o conhecimento, visto que, a distribuição dos dados pode alterar ao longo do tempo [Gama 2012].

A possibilidade de conhecer mais sobre estes dados gera a necessidade da criação de estruturas e estratégias para recuperar dados de forma eficiente. As estruturas que permitem indexar e recuperar rapidamente um determinado objeto são conhecidas como Métodos de Acesso Métricos (MAMs).

Para dados complexos (dados vetoriais, de trajetórias, dados multimídia, grafos, séries temporais) é empregada a noção de similaridade quando o objetivo é a recuperação, sendo útil recuperar os documentos similares (ou menos dissimilares) aos de referência.

Os custos dos algoritmos que determinam a similaridade entre pares objetos tornam as buscas por similaridade operações de custo elevado. Estes eventos têm motivado estudos com objetivo de tornar a busca por similaridade eficiente sobre grandes conjuntos de dados.

Deste modo, identifica-se o problema pois a medida que os dados são gerados pelo fluxo, eles devem ser processados e armazenados na estrutura. O processo de indexação do elemento pode ocasionar a degradação da estrutura devido ao grau de sobreposição entre a subárvores.

Neste sentido, pretende-se, com este trabalho, propor um método para melhorar o desempenho de MAMs dinâmicos (Slim-Tree, M-Tree, entre outros), para acelerar consultas por similaridade em ambiente de FCD, tendo com fator fundamental não comprometer a dinamicidade da estrutura. A abordagem proposta é a criação de uma estrutura auxiliar, para armazenamento dos dados provenientes do fluxo até que estes possam ser indexados na estrutura com a finalidade de minimizar o grau de sobreposição tornando as estruturas mais eficientes.

## 2. Trabalhos Relacionados

Nesta seção, serão descritos brevemente alguns dos estudos existentes na literatura relacionados ao trabalho proposto. Em [Brisaboa et al. 2006] é proposto um novo método de busca baseado em pivôs dinâmicos onde a estrutura de índice pode ser armazenada com eficiência na memória secundária. No trabalho proposto por [Oliveira et al. 2015] é utilizada uma estrutura dinâmica para lidar com dados complexos online. Já no estudo realizado [Barrios et al. 2012] é apresentada uma estrutura de índice dinâmico para apoiar consultas K-NN em fluxo contínuo de dados. Por serem estudos com resultados relevantes podem contribuir para este trabalho.

## 3. Contribuição do Trabalho

O presente trabalho tem como pressuposto contribuir para o aprimoramento das técnicas utilizadas para busca por similaridade e recuperação de conteúdo em ambiente de fluxo contínuo de dados.

## 4. Estado Atual

O trabalho está no estágio de implementação das estruturas e métodos. A estrutura de indexação escolhida foi Slim-Tree devido a característica dinâmica da estrutura e pelo fato de sua construção ser incremental. A estrutura auxiliar será uma janela de eventos, que terá seu tamanho variado dependendo da base de dados utilizada.

## Referências

- Barrios, J. M., Bustos, B., and Skopal, T. (2012). Snake table: A dynamic pivot table for streams of k-nn searches. In *Similarity Search and Applications*, pages 25–39. Springer.
- Brisaboa, N. R., Farina, A., Pedreira, O., and Reyes, N. (2006). Similarity search using sparse pivots for efficient multimedia information retrieval. In *Multimedia, 2006. ISM'06. Eighth IEEE International Symposium on*, pages 881–888. IEEE.
- Gama, J. (2012). A survey on learning from data streams: current and future trends. *Progress in Artificial Intelligence*, 1(1):45–55.
- Oliveira, P. H., Traina Jr, C., and Kaster, D. S. (2015). Improving the pruning ability of dynamic metric access methods with local additional pivots and anticipation of information. In *Advances in Databases and Information Systems*, pages 18–31. Springer.

# Método de Estimação de Chuva Acumulada Utilizando Técnicas de Mineração de Dados e Processamento Digital de Sinais em Dados Coletados por Smartphones

**Autor: Luiz Fernando Afra Brito<sup>1</sup>,  
Orientador: Marcelo Keese Albertini<sup>1</sup>**

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

luiz.brito@ufu.br, albertini@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 06/2016

Mês e ano de previsão de conclusão: 06/2016

Data de aprovação da proposta de dissertação/plano de tese: XX/XX/20\*\*

**Resumo.** *Este trabalho apresenta estudos realizados para o desenvolvimento de um método de estimação de chuvas acumuladas utilizando dados de smartphones. Para isso, técnicas de processamento digital de sinais serão aplicadas seguindo uma abordagem de análise temporal dos dados coletados a fim de extrair características relevantes. Após a extração das características, técnicas de mineração de dados serão utilizadas para geração de modelos de regressão para estimar o estado atual. O trabalho ainda se encontra em estágio de coleta dos dados e não possui resultados.*

**Palavras-Chave.** *chuva, coleta de dados, estimação, mineração de dados, nowcasting, smartphone, precipitação, processamento de sinais, sistemas de alerta*

## 1. Introdução e Motivação

*Nowcasting* é o termo utilizado para descrever previsões meteorológicas que antecipam algumas horas. O estado meteorológico atual é monitorado e modelos numéricos são utilizados para extrapolar variáveis do ambiente. Devido ao curtíssimo prazo de antecipação, o *nowcasting* possibilita a obtenção de informações mais precisas e detalhadas que métodos convencionais de previsão.

Atualmente, dados coletados por radares meteorológicos terrestres são utilizados para *nowcasting*. Esses radares possuem alta taxa de amostragem e apresentam boa resolução espacial dentro do raio de atuação. Contudo, apenas grandes cidades e capitais são beneficiadas devido ao alto custo envolvido. Recentemente, estudos mostraram que a atenuação sofrida em sinais eletromagnéticos ao longo de enlaces de comunicação sem fio pode ser utilizada para para estimar precipitações atmosféricas.

Este trabalho busca estudar dados de comunicações sem fio coletados por *smartphones* para estimar o volume de chuvas acumuladas durante o período de uma hora. Primeiramente, serão utilizadas técnicas de processamento digital de sinais para análise, pré-processamento e extração de características relevantes do dados coletados. Depois, serão utilizadas técnicas de mineração de dados para mensurar a taxa de precipitação de chuva do estado meteorológico atual. Para avaliação do método proposto serão realizados experimentos comparativos utilizando leituras de estações meteorológicas automáticas do Instituto Nacional de Meteorologia (INMET).

## 2. Trabalhos Relacionados

Em trabalhos recentes a atenuação de sinais de torres de comunicação sem fio de alta frequência tem sido relacionados com fatores atmosféricos. Messer Messer (2007) mostrou que as redes de comunicação com frequência acima de 10 Ghz são severamente afetadas por precipitações atmosféricas. Em Zinevich (2010), a partir da avaliação da atenuação de sinais, os autores construíram um modelo para estimar precipitações atmosféricas com acurácia de até 90%.

## 3. Contribuição do Trabalho

Esse trabalho tem como objetivo a validação de um método para estimação do volume de chuvas acumuladas utilizando dados de *smartphones*. Para isso, 4 problemas, com suas respectivas contribuições, serão solucionados. O primeiro é a criação de coleções de dados que representem as variáveis coletadas. O segundo é aplicação de filtros que melhorem a qualidade das coleções de dados. O terceiro é a seleção de características relevantes tendo em vista a abordagem de análise temporal dos dados. O quarto é a geração de um modelo de regressão que obtenha taxa de acerto adequada em situações reais. Com a obtenção de bons resultados pretende-se melhorar a resolução e a capacidade de estimação de modelos *nowcasting*.

## 4. Metodologia

Neste trabalho três tipos de comunicações sem fio estabelecidas por *smartphones* serão estudadas: com satélites de geolocalização, com redes *wifi* e com torres de operadoras telefônicas. Adicionalmente, outras variáveis também serão estudadas a fim de verificar a viabilidade de inclusão no modelo de estimação.

A coleta será feita conforme amostragem uniforme pois, dessa maneira, é possível analisar mudanças de variáveis observadas em uma abordagem temporal [Lyons 2010, Capítulo 2]. Depois da coleta dos dados, filtros serão utilizados para remover informações desnecessárias presentes nos dados coletados. Na etapa seguinte serão extraídas as principais características tendo como abordagem o estudo de sinais digitais. Por fim, técnicas de mineração de dados para regressão serão utilizadas para estimar o volume de chuva acumulada durante uma hora.

## 5. Estado atual do Trabalho

O trabalho se encontra em estágio inicial. Até o momento, um aplicativo para *smartphones Android* foi desenvolvido a fim de coletar cerca de 60 variáveis. Também foi implementado um servidor para receber os dados coletados pelos usuários do aplicativo. Além disso, um módulo para obtenção do volume de chuvas acumuladas foi implementado. Do qual os dados são provenientes de estações meteorológicas automáticas do INMET.

Nas próximas etapas pretende-se realizar coleta, pré-processamento e extração das principais características dos dados coletados e, posteriormente, utiliza-los na criação de modelos de regressão.

## Referências

- Lyons, R. G. (2010). *Understanding digital signal processing*, chapter 2, pages 21–44. Pearson Education.
- Messer, H. (2007). Rainfall monitoring using cellular networks [in the spotlight]. *IEEE Signal Processing Magazine*, 24(3):144–142.
- Zinevich, A., Messer, H., and Alpert, P. (2010). Prediction of rainfall intensity measurement errors using commercial microwave communication links. *Atmospheric Measurement Techniques*, 3(5):1385–1402.

# Rastreamento e Evolução de Restrições em Agrupamentos Semi-Supervisionados

Autor: Guilherme Alves<sup>1</sup>,  
Orientadora: Maria Camila N. Barioni<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

{guilhermealves, camila.barioni}@ufu.br

Nível: Mestrado

Mês e ano de ingresso no programa: 04/2015

Mês e ano de previsão de conclusão: 04/2017

Data de aprovação da proposta de dissertação: 19/02/2016

**Resumo.** *A disponibilidade abundante de dados torna cada vez mais inviável a busca manual por informações relevantes. Métodos automáticos para organizar os dados propiciam o acesso em tempo hábil. Todavia, a organização dos dados muda ao longo do tempo e um pequeno subconjunto pode fornecer informação antecipada de que a estrutura da organização mudará. Neste trabalho é investigado a evolução desse subconjunto por meio do rastreamento das informações nele contido. Assim, espera-se que modificações no subconjunto possam auxiliar no entendimento das alterações na organização dos dados bem como permitir a predição das mudanças futuras.*

**Palavras-Chave.** *evolução de restrições, evolução de agrupamentos, agrupamento semi-supervisionado em fluxo contínuo de dados*

## 1. Introdução e Motivação

O advento das redes sociais e da computação ubíqua, por exemplo, são responsáveis pelo aumento vertiginoso da geração de dados. O imenso volume de informações, no entanto, é um problema, pois dificulta a busca de informações úteis pelo usuário. A análise de agrupamentos surge com o propósito de agrupar os dados objetivando resumir as informações ou auxiliar na organização dos mesmos. Em determinados cenários, quando detém-se um conhecimento de domínio ou um subconjunto de dados cuja variável resposta é conhecida, pode-se utilizar esse tipo de informação adicional para guiar a detecção. Técnicas que suportam essa informação adicional, também referida como restrições, constituem a área de pesquisa denominada de aprendizagem semi-supervisionada. Inobstante, preferências e padrões podem mudar ao longo do tempo. Então, os algoritmos em fluxo (*stream*) são utilizados de modo a considerar a característica não estacionária e contínua da informação ao longo do tempo.

Considere que uma empresa de *marketing* deseja segmentar o mercado com o objetivo de encontrar grupos de possíveis clientes. Assim, um algoritmo de detecção de agrupamentos semi-supervisionado é executado em uma determinada época. No entanto, a medida que o tempo passa, as preferências dos clientes mudam e a estrutura dos grupos se altera. Em um determinado intervalo de tempo as preferências dos clientes de um grupo específico alteram-se a ponto de um subconjunto de clientes não pertencer ao grupo original. Então, as novas restrições, provenientes das preferências dos clientes, impõem uma nova estrutura particionando o grupo

original em dois novos grupos. Observe ainda que novos clientes podem aparecer, mas o espaço de características permanece inalterado, ou seja, a representação dos clientes não muda ao longo do tempo, somente as preferências.

O cenário apresentado é uma instância do problema investigado neste trabalho. Afinal, as restrições também podem mudar ao longo do tempo, o que pode explicar as transições dos agrupamentos: surgimento, morte, divisão, alterações na forma espacial etc. Este trabalho parte da suposição que a evolução das restrições pode ter impacto nas alterações sofridas pelos agrupamentos ao longo do tempo.

## 2. Contribuição do Trabalho

A principal contribuição deste trabalho é um conjunto de técnicas que modelam e monitoram a evolução das restrições no cenário de detecção de agrupamentos semi-supervisionado em *data streams*. Assim, as técnicas a serem desenvolvidas deverão contribuir para o entendimento do comportamento das transições dos agrupamentos ao longo do tempo e também prever as futuras transições.

## 3. Estado atual do Trabalho

Os conjuntos de dados apropriados foram obtidos como primeira etapa. Atualmente, um algoritmo de detecção de agrupamentos semi-supervisionado foi adotado para ser estendido no cenário de fluxo contínuo [Lai et al. 2014]. Ademais, o algoritmo também está sendo estendido de forma a ser capaz de construir dinamicamente um modelo gráfico probabilístico.

## 4. Trabalhos Relacionados

**Evolução de agrupamentos.** O *framework* MONIC monitora e analisa as mudanças em agrupamentos com o objetivo de entender a natureza da mudança [Spiliopoulou et al. 2006]. Nesse trabalho são formalizadas as transições dos agrupamentos. No *framework* MEC um grafo bipartido é utilizado para modelar as transições dos grupos ao longo do tempo [Oliveira and Gama 2010].

**Agrupamento semi-supervisionado em fluxo.** Diversos algoritmos de detecção de agrupamentos foram estendidos para suportarem restrições e também tratarem dados em fluxo contínuo, a saber: C-DENSTREAM [Ruiz et al. 2009] e I-SSFCM [Castellano et al. 2013]. No entanto, esses trabalhos não investigam a evolução das restrições bem como as contradições que podem aparecer ao longo do tempo e o impacto na estrutura dos agrupamentos.

## Referências

- Castellano, G., Fanelli, A. M., and Torsello, M. A. (2013). Shape Annotation by Incremental Semi-supervised Fuzzy Clustering. In *10th International Workshop WILF*, pages 193–200, Cham. Springer.
- Lai, H. P., Visani, M., Boucher, A., and Ogier, J.-M. (2014). A new interactive semi-supervised clustering model for large image database indexing. *Pattern Recognition Letters*, 37(1):94–106.
- Oliveira, M. and Gama, J. (2010). Bipartite graphs for monitoring clusters transitions. In *9th International Symposium IDA*, pages 114–124, Tucson. Springer.
- Ruiz, C., Menasalvas, E., and Spiliopoulou, M. (2009). C-DenStream: Using Domain Knowledge on a Data Stream. In *12th International Conf. DS*, pages 287–301, Berlin. Springer.
- Spiliopoulou, M., Ntoutsi, I., Theodoridis, Y., and Schult, R. (2006). MONIC - Modeling and Monitoring Cluster Transitions. In *12th International Conf. KDD*, page 706, NY. ACM.

# Avaliando *Cookbooks* Construídos a partir de Engenharia Reversa: Um Experimento Controlado.

**Aluna: Raquel Fialho de Queiroz Lafeté<sup>1</sup>, Orientador: Marcelo de Almeida Maia<sup>1</sup>**

<sup>1</sup>Pós-Graduação em Ciência da Computação - Universidade Federal do Uberlândia (UFU)  
Nível: Doutorado - Início: 03/2012 - Término:03/2017 - Plano de Tese: 08/02/2014

raquel.rafialho@gmail.com, marcelo.maia@ufu.br

**Resumo.** *Nesta pesquisa buscamos verificar os benefícios de uma nova forma de documentação de frameworks, baseada no código fonte. Nossa hipótese é que o próprio código fonte do framework e instâncias existentes podem ser utilizados para criar documentação apropriada na forma de cookbook. Pretendemos verificar se o cookbook proposto é uma alternativa viável de documentação. Para este fim, será necessário comparar o desempenho de desenvolvedores utilizando o cookbook proposto e utilizando a documentação tradicional.*

**Palavras-Chave.** *Framework. Livro de receitas. Análise estática. Análise dinâmica.*

## 1. Introdução e Proposta

Frameworks são amplamente utilizados e são uma forma de reúso de software efetiva. Instanciação de frameworks é uma atividade que exige um esforço de entendimento substancial. No entanto, na maioria dos casos, a documentação adequada não existe ou não é atualizada [Heydarnoori 2009]. Diante deste problema, propomos em [Lafeté et al. 2015] uma abordagem baseada em exemplos para geração semi automática de documentos na forma de *cookbooks* composto por receitas para instanciar características do framework. Baseando-nos na afirmativa de que, muitas vezes, a melhor maneira de aprender um framework é por meio de exemplos [Johnson 1992]. Utilizamos engenharia reversa, análise dinâmica e análise estática para localizar o código fonte ligado a instanciação e obter informações sobre os elementos de código, suas interações, exemplos de uso e padrões de projeto.

## 2. Objetivo

Este trabalho tem como objetivo demonstrar que o modelo de *cookbook* derivado de código é uma alternativa viável para guiar o desenvolvedor no processo de instanciação, podendo ser tão bom quanto ou até melhor do que a documentação gerada manualmente, em termos de taxa de acerto e tempo de execução.

## 3. Resultados

Em [Lafeté et al. 2015], estudos preliminares, com o framework JHotDraw, mostraram que a abordagem proposta pode obter as informações definidas para compor *cookbook* com um bom *recall*, ou seja, suficientes para guiar a instanciação. Em estudo com 5 sujeitos foi possível perceber a utilidade das informações, a receita guiou e auxiliou os sujeitos no processo de instanciação. Estes estudos nos direcionaram e levaram a melhorias na abordagem.

#### 4. Próximas Atividades e Cronograma

Esta seção apresenta as próximas atividades do nosso trabalho e um cronograma (Figura 1). Atualmente, estamos aplicando a abordagem sobre os frameworks JMF (*Java Media Framework*) e Android ou Eclipse para obter receitas e verificar se a abordagem com melhorias apresenta bom *recall*. Além de gerar *cookbooks* que serão utilizados nos experimentos.

Pretendemos realizar experimentos controlados com sujeitos desenvolvedores. Serão duas etapas experimentais, com dois frameworks distintos. Para cada etapa, iremos:

- Selecionar entre 20 à 30 participantes com conhecimento mínimo para execução das atividades e distribuição equilibrada em dois grupos. Cada grupo hora atuará usando o *cookbooks* e hora atuará usando documentação tradicional, alternados em dois experimentos com o mesmo framework.
- Definir 2 atividades reais de instanciação para cada framework com o mesmo nível de complexidade, onde sejam necessárias modificações no código. A execução da primeira atividade não deve influenciar na execução da segunda devido à alternância dos grupos.
- Montar materiais de apoio: i) documento para explicar o *cookbook* e documentação tradicional; ii) para explicar as atividades de instanciação; iii) questionários para seleção dos participantes; iv) questionário para coleta informações qualitativas e quantitativas.
- Treinar os participantes para realização do experimento controlado.
- Preparar os ambientes para a instanciação do framework em questão.
- Conduzir os experimentos aplicando as atividades instanciação para cada grupo. O tempo de execução e resultados das atividades serão coletados a fim de evidenciar o desempenho dos grupos e compará-los. O questionário será aplicado para obter informações sobre satisfação, nível de dificuldade, utilidade das informações e usabilidade.

Os resultados serão analisados por meio de estatística descritiva com boxplots. Serão realizados testes estatísticos para avaliar a possibilidade de rejeitar a hipótese nula de que não diferença entre os participantes que usam os cookbooks e os participantes que usam a documentação tradicional. Serão calculados intervalos de confiança para avaliar o tamanho do efeito da variável.

Identificação da Etapa	Início	Término
Análise do protocolo do projeto pelo Conselho de Ética da UFU	25-03-16	30-06-16
Aplicar abordagem para o framework JMF - Gerar Cookbook	25-03-16	30-05-16
Experimento Controlado (Primeiro Estudo - Alunos - Framework JMF)	01-07-16	30-07-16
Aplicar abordagem para Eclipse ou Android - Gerar Cookbook	01-06-16	01-08-16
Experimento Controlado (Segundo Estudo – Profissionais – Framework 1)	01-08-16	30-08-16
Experimento Controlado (Segundo Estudo – Profissionais – Framework 2)	01-09-16	30-09-16
Tabulação, Análise dos dados e Redação de Artigos	30-09-16	30-12-16
Preparação para defesa da Tese	01-01-17	01-03-17

Figura 1. Cronograma resumido das próximas atividades.

#### Referências

- Heydarnoori, A. (2009). *Supporting Framework Use via Automatically Extracted Concept-Implementation Templates*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada.
- Johnson, R. E. (1992). Documenting frameworks using patterns. In *Conference Proceedings on Object-oriented Programming Systems, Languages, and Applications*, OOPSLA '92, pages 63–76, New York, NY, USA. ACM.
- Lafetá, R. F. Q., Maia, M. A., and Röthlisberger, D. (2015). Framework instantiation using cookbooks constructed with static and dynamic analysis. In *2015 IEEE 23rd International Conference on Program Comprehension*, pages 125–128.

# **Avaliando Projeto de Software por Priorização de Classes-Chave**

**Aluna: Liliane do Nascimento Vale Orientador: Marcelo de Almeida Maia**

Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal de Uberlândia – Uberlândia, MG – Brasil  
liliane.ufg@gmail.com, marcmaia@facom.ufu.br

Nível: Doutorado - Início: 03/2013 – Término: 03/2017 - Plano de Tese: 08/2014

**Resumo.** Neste trabalho, analisamos uma nova alternativa de compreender e avaliar o projeto de software a partir de um conjunto reduzido de classes relevantes arquiteturalmente (classes-chave). Inicialmente, usando técnicas de engenharia reversa sobre o código fonte e de processos de redução do volume de dados, torna-se possível obter as classes-chave. Em seguida, avaliaremos se tais classes podem ser priorizadas como um ponto de partida para entender arquitetura de software em relação às propriedades social (ownership) e estruturais. Finalmente, será necessário avaliar o desempenho dos desenvolvedores utilizando as classes-chave em tarefas de manutenção.

**Palavras-Chave:** classes-chave, análise dinâmica/estática, smells, métricas de complexidade.

## **1. Introdução e Proposta**

O custo e o esforço demandado para compreender e adaptar elementos internos de sistemas de software está relacionado à investigação artefatos tais como, código e documentação. Em sistemas bem projetados, artefatos de software estão organizados de tal maneira que eles podem ser compreendidos e mudados independentemente um do outro. Portanto, a avaliação de projeto é uma importante tarefa para melhorar a manutenção e evolução de software. Porém, em muitos casos a documentação contendo decisões de projeto é perdida, ou quando existe ela está incompleta ou desatualizada. Neste sentido, os desenvolvedores são induzidos à analisar o código fonte que é a única fonte de informação confiável para avaliar o projeto de software. A avaliação do projeto de software considerando-se todas as classes do sistema como ponto de partida é uma tarefa difícil. Algumas técnicas de reconstrução de arquiteturas foram propostas, porém a maioria destas abordagens são difíceis de usar e tem baixa precisão. Diante deste problema, nós propomos Keele, uma maneira semi-automática de encontrar classes-chave em sistemas orientados a objetos que são usadas para documentar a arquitetura de software e portanto, priorizadas para a compreensão e avaliação do projeto sob o ponto de vista de: importantes relações de dependência; serem mais propensas a bad smells; assumirem os piores valores em métricas de coesão/acoplamento; e uma alta concentração de desenvolvedores dedicados à modificação das mesmas.

## **2. Objetivo**

Este trabalho tem como objetivo mostrar que em um contexto dinâmico classes-chave extraídas das raízes de árvores de rastros de execução contém importantes propriedades. Em particular, tais classes estão localizadas em níveis mais altos da árvore de rastros, e portanto, detém de um forte controle sobre a aplicação, uma vez que todas as outras chamadas de métodos (nós da árvore) são controlados por esta raiz. Conseqüentemente, estas classes estão intrinsecamente relacionadas à qualidade do projeto de software. Para entender o impacto destas classes sobre o projeto de software, em um contexto estático nós investigamos a organização destas classes em um estilo em camadas para analisar as relações de dependências, bem como se indicadores

clássicos de modularidade (coesão/acoplamento) são diferentes quando comparados as demais classes do sistemas e se estes indicadores podem estar associados à ocorrência de *smells*.

### 3. Resultados

Em [Vale et al, 2015] um estudo preliminar com dois sistemas de código aberto mostraram que a abordagem é efetiva para recuperar classes-chave em termos de recall e precisão quando comparados com trabalhos correlatos [Zaidman et al, 2015].

### 4. Próximas Atividades e Cronograma

Esta seção apresenta as próximas atividades do nosso trabalho e um cronograma (Figura 1). Atualmente, estamos ampliando os experimentos da abordagem com novos sistemas e avaliando a capacidade das classes-chave como indicadoras de qualidade do projeto de software, usando métricas de coesão e acoplamento, detecção de bad smells e a informação do *ownership* como mencionados anteriormente. Novas atividades serão desempenhadas entre elas:

**1) Experimentos com sujeitos desenvolvedores:** nesta atividade vamos selecionar usando questionários, desenvolvedores com níveis de conhecimento distintos que serão submetidos à desempenhar atividades de reuso e manutenção de software usando as classes-chave como guia para localização mais rápida das informações. Para conduzir os experimentos forneceremos treinamento, material de apoio, configuração do ambiente, questionário para coleta informações qualitativas e quantitativas para obter dados sobre satisfação, nível de dificuldade, utilidade/reusabilidade das classes-chave e redução do esforço analítico do desenvolvedor.

**2) Alinhamento de classes-chave:** nesta atividade, nós queremos investigar se alinhando classes-chave de um mesmo domínio de aplicação é possível obter conceitos que caracterizam classes-chave como sendo classes de referência em uma arquitetura de referência. Por exemplo, considerando as classes-chave de Tomcat<sup>1</sup>, Undertow<sup>2</sup> e Jetty<sup>3</sup>, serão alinhadas semanticamente para avaliar se uma classe-chave desempenha o mesmo papel, implementa o mesmo propósito e conceito que as de outros sistemas. Nós podemos investigar também, se uma classe-chave implementa funcionalidades extras que podem ser incorporadas em outros sistemas de mesmo domínio, como por exemplo, inserir funcionalidades do Tomcat no Undertow.

Atividade	2016							2017			
	mai	jun	jul	ago	set	out	nov	dez	jan	fev	mar
Elaboração/Experimento com sujeitos desenvolvedores	X	X	X	X	X						
Alinhamento de classes-chave					X	X	X				
Análise de dados e redação de artigos						X	X	X			
Escrita e defesa da tese								X	X	X	X

Figura 1: Cronograma resumido das próximas atividades

### 5. Referências

Vale, N. L. and Maia, M. A., “Keele: Mining key architecturally relevant classes using dynamic analysis,” in Proc. ICSME. IEEE, 2015, pp. 566–570.

A. Zaidman and S. Demeyer, “Automatic identification of key classes in a software system using webmining techniques<sup>1</sup>,” J. Softw. Maint. Evol., vol. 20, no. 6, pp. 387–417, Nov. 2008.

<sup>1</sup> <http://tomcat.apache.org/> <sup>2</sup> <http://undertow.io/> <sup>3</sup> <http://www.eclipse.org/jetty/>

# Classificação de Imagens Histológicas de Lesões Proliferativas Intraductais de Mama

Walter Alexandre A. de Oliveira<sup>1</sup>,  
Marcelo Zanchetta do Nascimento<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

waaao@doutorado.ufu.br, marcelo.nascimento@ufu.br

Nível: Doutorado

Mês e ano de ingresso no programa: 02/2016

Mês e ano de previsão de conclusão: 02/2020

Data de aprovação da proposta de dissertação/plano de tese: –

**Resumo.** *Um novo método para descrever imagens histológicas é apresentado. A abordagem é baseada em informações de forma e cor extraídas de uma imagem usando a curva de preenchimento de espaço de Hilbert. Deconvolução de cor foi realizada para separar a contribuição de diferentes pigmentos utilizados em imagens histológicas. O propósito do descritor é a classificação das imagens como em classes benigno/maligno.*

**Palavras-Chave.** *descritor, lesão proliferativa intraductal, câncer de mama*

## 1. Introdução e Motivação

Ferramentas que auxiliam a análise de imagens médicas são de grande valor para pesquisa e prática clínica [Nanni and Melucci 2016]. Carcinoma ductal *in situ* (DCIS) e hiperplasia ductal usual (UDH) são exemplos de condições que ocorrem em lesões intraductal proliferativas na mama [bre b, bre a]. Sabendo ser o DCIS o tipo de câncer de mama não invasivo mais comum (porém com alto risco de se tornar invasivo), é de extrema importância obter diagnósticos rápidos e precisos [bre b, Dong et al. 2014].

## 2. Fundamentação Teórica

Curva de preenchimento de espaço (SFC) é um caminho em uma grade  $n$ -dimensional que percorre todos os pontos e nunca se cruza [Yasser Ebrahim 2009]. Ela pode ser usada para determinar a ordem em que pixels são extraídos de uma imagem [Yasser Ebrahim 2009].

A combinação de hematoxilina e eosina é uma técnica de pigmentação de imagens histológicas bastante comum [Brown 2013]. A deconvolução de cor é um processo usado para separar a contribuição destes corantes em uma imagem [Ruifrok and Johnston 2001]. A técnica será utilizada para obter informações contidas em vários canais de cor da imagem.

## 3. Contribuição do Trabalho

Conforme [Dong et al. 2014], um método capaz de discernir entre classes benigno/maligno é de grande valor para a realização de diagnósticos patológicos, permitindo uma análise mais rápida, mais precisa e de menor custo.

O presente trabalho propõe um novo descritor capaz de classificar imagens histológicas de lesões intraductais proliferativas de mama como benigno (UDH) ou maligno (DCIS). O método proposto requer pouca intervenção manual (poucos parâmetros serão manipulados).

#### 4. Estado atual do Trabalho

O método proposto está ainda sendo implementado. Experimentos de classificação de imagens usando redes neurais são previstos para validação do método que será avaliado, e comparado a outros métodos, através de curva ROC e área sob a curva ROC (AUC).

#### 5. Trabalhos Relacionados

[Gertych et al. 2015] usa histogramas de intensidade de pixels dos corantes hematoxilina e eosina (H&E) capturando a diferença arquitetural entre glândulas e stromas. Eles juntam histogramas de padrões binários locais e variância local a fim de obter uma característica de textura e classificar tecidos de próstata como benigno ou maligno (câncer). Com um conjunto de dados de 210 imagens, obtiveram uma precisão média de 81,42% em seu melhor resultado.

Em [Hervé et al. 2011] combinam-se descritores estatísticos de cor e textura em imagens histológicas usando H&E. Usando *support vector machines* (SVM) como classificador, seus experimentos, executados 5000 vezes, foram avaliados com medida de precisão média. Observa-se que bons resultados são obtidos apenas quando os descritores, isoladamente, também conseguem bons resultados.

#### Referências

- Certain breast changes. [http://www.breastcancer.org/risk/factors/breast\\_changes](http://www.breastcancer.org/risk/factors/breast_changes). Accessed in 05/14/2016.
- Diagnosis of dcis. <http://www.breastcancer.org/symptoms/types/dcis/diagnosis>. Accessed in 05/14/2016.
- Brown, S. (2013). The science and application of hematoxylin and eosin staining.
- Dong, F., Irshad, H., Oh, E.-Y., Lerwill, M. F., Brachtel, E. F., Jones, N. C., Knoblauch, N. W., Montaser-Kouhsari, L., Johnson, N. B., Rao, L. K. F., Faulkner-Jones, B., Wilbu, D. C., Schnitt, S. J., and Beck, A. H. (2014). Computational pathology to discriminate benign from malignant intraductal proliferations of the breast. *PLoS ONE*, 12(9).
- Gertych, A., Ing, N., Ma, Z., Fuchs, T. J., Salman, S., Mohanty, S., Bhele, S., Velásquez-Vacca, A., Amin, M. B., and Knudsen, B. S. (2015). Machine learning approaches to analyze histological images of tissues from radical prostatectomies. *Computerized Medical Imaging and Graphics*.
- Hervé, N., Servais, A., Thervet, E., Olivo-Marin, J.-C., and Meas-Yedid, V. (2011). Statistical color texture descriptors for histological images analysis. *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 724–727.
- Nanni, L. and Melucci, M. (2016). Combination of projectors, standard texture descriptors and bag of features for classifying images. *Neurocomputing*, 173:1602—1614.
- Ruifrok, A. and Johnston, D. (2001). Quantification of histochemical staining by color deconvolution. *Analytical and Quantitative Cytology and Histology*, 23:291–299.
- Yasser Ebrahim, Maher Ahmed, W. A. S.-C. C. (2009). Shape representation and description using the hilbert curve. *Pattern Recognition Letters*, (30):348–358.

# Métodos computacionais para caracterização e análise do desenvolvimento de Embriões de *Drosophila melanogaster*

**Autor:** Daniela Justiniano de Sousa<sup>1</sup>,  
**Orientador:** Bruno Augusto Nassif Travencolo<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

justinianoufu@gmail.com, travencolo@gmail.com

Nível: Doutorado

Mês e ano de ingresso/conclusão no programa: Março-2015/Março-2019

**Resumo.** Este projeto tem como objetivo o desenvolvimento e a aplicação de métodos computacionais para o estudo da dinâmica da formação e diferenciação de tecidos embrionários em *Drosophila melanogaster*. Essa tarefa envolve a aquisição de imagens em 3D relativas à padrões de expressão gênica, às quais serão aplicadas técnicas de Processamento de Imagens e Visualização Computacional. Os resultados das análises permitirão aprofundar o conhecimento do desenvolvimento tecidual, bem como, a obtenção de novas ferramentas automáticas para o estudo da diferenciação em diversos organismos.

**Palavras-Chave.** Processamento de Imagens, *Drosophila melanogaster*.

## 1. Introdução e Motivação

A inserção da Ciência de Computação em outras áreas do conhecimento tem sido fundamental para a pesquisa científica e o aumento potencial de novas descobertas. Um exemplo importante é a aplicação de técnicas computacionais para o processamento e armazenamento de dados relativos às ciências biológicas. Um dos desafios atuais visa entender como a relação entre os vários componentes de um sistema biológico dão origem às funções realizadas por células ou organismos. Nesse sentido, a principal linha de estudo que está sendo seguida é a análise dos padrões de expressão gênica, uma vez que esse tipo de dados é a chave da diferenciação celular. Um modelo biológico notório desse cenário é a *Drosophila melanogaster*. De fato, é crescente o número de bases na Internet que disponibilizam dados de expressão gênica sobre essa espécie, sendo normalmente encontradas na forma de imagens. É possível obtê-las em diferentes resoluções, dimensões ou estado da espécie (*in vivo* ou *post mortem*). No entanto, apenas conhecê-las não é suficiente para entendimento do desenvolvimento de um organismo. É de fundamental importância caracterizar os diferentes padrões de expressão dos genes, tanto temporalmente quanto espacialmente. Considerando a complexidade e a enorme quantidade dos dados, esse cenário apresenta-se como um potencial motivador a grupos de computação, levando-os a propor novas ferramentas para armazenar e processar esse tipo informações.

## 2. Fundamentação Teórica

A *Drosophila* é um inseto díptero com aproximadamente 3 mm de comprimento, popularmente conhecido como a “mosca da fruta” ou “mosca do vinagre”. Essa espécie é ideal para estudo de interações gênicas, devido à clareza na visualização de seus padrões de expressão gênica.

Entende-se por padrão de expressão o conjunto de genes expressos por uma célula em um dado momento. Esse tipo de imagem fornece informações importantes sobre o funcionamento de uma célula. Dentro desse contexto, a *Drosophila* é um dos modelos mais apropriados para o estudo e análise da dinâmica celular. Particularmente, essa tarefa envolve a integração de áreas da Computação e Biologia, em que a análise se dá por meio de técnicas computacionais.

### **3. Contribuição do Trabalho**

A principal contribuição do projeto é o desenvolvimento de métodos computacionais para extração e análise de dados que suportará o entendimento do desenvolvimento tecidual, em específico a análise quantitativa da dinâmica celular, a partir de três contextos biológicos fundamentais. O conjunto das ferramentas propostas integrará um projeto amplo, que em parceria com a UFRJ, gerará dados para o melhor entendimento dos sistemas estudados, e terá o potencial de permitir a utilização das rotinas e modelos desenvolvidas para a análises de outros tecidos animais.

### **4. Estado atual do Trabalho**

Uma breve descrição que sintetize o desenvolvimento do projeto de doutorado é delimitado em marcos, ainda que seus contextos sejam dinâmicos e não possam estar fechados em informações precisas, a saber: (1) Disciplinas teóricas, (2) Base de Dados, (3) Atividades Correlatas, (4) Disciplinas práticas, (5) Revisão da literatura, (6) Rotinas computacionais, (7) Experimentação, (8) Documentação e (9) Defesa. As atividades estão sendo realizadas dentro do previsto e encontram-se atualmente no Marco 3.

### **5. Análise dos resultados**

O projeto descrito apresenta caráter multidisciplinar, devido a isso, a análise dos resultados obtidos ocorrerá de forma cooperativa entre os colaboradores envolvidos. A equipe envolvida tem larga experiência na combinação de estratégias experimentais, computacionais e teóricas diretamente relacionadas aos objetivos do presente projeto.

### **6. Trabalhos Relacionados**

Na literatura é possível encontrar uma diversidade de pesquisas relativas ao processamento computacional de imagens da *Drosophila melanogaster* [Surkova et al. 2008] [de Sousa et al. 2013]. No entanto, a temática desse projeto apresenta-se como um desafio. Isto porque, até o presente momento, não se encontraram propostas que sugiram um ferramental computacional que suporte compreensão dos processos envolvidos na dinâmica da formação e diferenciação de tecidos embrionários. Outro ponto que pode ser melhor explorado é a formação de bases de bioimagens, em que o desafio é criar atlas digitais de organismos ou órgãos, onde ficam integradas informações anatômicas e ontológicas, juntamente com outros tipos de dados, como os de expressão gênica.

### **Referências**

- de Sousa, D. J., Cardoso, M. A., Bisch, P. M., Lopes, F. J. P., and Travençolo, B. A. N. (2013). A segmentation method for nuclei identification from sagittal images of drosophila melanogaster embryos. In *WSCG*, pages 133–142.
- Surkova, S., Myasnikova, E., Janssens, H., Kozlov, K., Samsonova, A., Reinitz, J., and Samsonova, M. (2008). Pipeline for acquisition of quantitative data on segmentation gene expression from confocal images. *Fly (Austin)*, 2(2):58–66.

# Recomendação de Conhecimento da Multidão para Auxílio na Correção de Bugs Independentes de Domínio de Aplicação

**Autor: Eduardo Cunha Campos**<sup>1</sup>,  
**Orientador: Marcelo de Almeida Maia**<sup>1</sup>

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

eduardocunha11@gmail.com, marcelo.maia@ufu.br

Nível: Doutorado

Mês e ano de ingresso no programa: 04/2015

Mês e ano de previsão de conclusão: 04/2019

Data de aprovação do plano de tese: A ser submetido em 06/2016

**Resumo.** *O presente trabalho possui o propósito de recomendar correções para API-related-usage bugs, i.e., bugs independentes dos requisitos funcionais do software que geralmente causam uma saída inesperada e incorreta resultante de um uso incomum ou contra-intuitivo de um método da API (Application Programming Interface). Nós propomos uma abordagem para encontrar correções de API-related-usage bugs, que é baseada no casamento entre trechos de código que estão sendo depurados pelo desenvolvedor e trechos de código relacionados presentes no site de pergunta-resposta Stack Overflow (SO). Nós analisamos trechos de código reais retirados de projetos de software hospedados no site OHLOH Code Search que contém potenciais API-related-usage bugs. Além disso, foi conduzido um estudo com estes trechos de código para investigar com qual extensão a abordagem proposta provê informação apropriada a partir do SO para corrigir estes API-related-usage bugs.*

**Palavras-Chave.** *crowd debugging, crowd knowledge, recommendation systems*

## 1. Introdução e Motivação

Desenvolvedores de *software* frequentemente precisam utilizar métodos de API (*Application Programming Interface*) que eles não estão familiarizados com ou que não se lembram como usá-los adequadamente [Moreno et al. 2015]. Como resultado disto, alguns *bugs* podem ser introduzidos no código-fonte do projeto devido a um uso incorreto destes métodos por parte dos desenvolvedores. O presente trabalho visa endereçar uma classe particular de *bugs*, i.e., *bugs* relacionados a um não entendimento comum entre os desenvolvedores sobre como utilizar alguns métodos de API. Estes *bugs* possuem uma característica comum: eles ocorrem repetidamente em diferentes domínios de aplicação e são independentes dos requisitos funcionais do *software*. Esta classe de *bugs* é chamada de “*API-usage-related bugs*” pois é provável que este tipo de *bug* já tenha ocorrido em diferentes projetos e existe uma descrição do problema em algum lugar da *Web*, juntamente com a sua explicação e correção. Em outras palavras, existe uma grande chance da multidão ter identificado o *bug* e sua solução.

Apesar dos esforços da comunidade acadêmica e da natureza geral dos *API-related-usage bugs*, muitos problemas ainda não foram resolvidos com relação a esta classe de *bugs*: (i) A correção deles ainda é um processo manual e que consome tempo e esforço por parte do desenvolvedor; (ii) Não existe uma abordagem de recomendação com resultados promissores que sugere correções para este tipo de *bug*.

## 2. Fundamentação Teórica

Sistemas de Recomendação para Engenharia de *Software* representam uma solução possível para este tipo de problema. Um sistema de recomendação recupera e analisa os dados, identifica artefatos úteis e sugere os mesmos para o desenvolvedor [Robillard et al. 2010]. A ideia deste trabalho é propor um sistema de recomendação que recomenda conteúdo do SO para ajudar os desenvolvedores a corrigirem *API-related-usage bugs* em seus trechos de código.

## 3. Contribuição do Trabalho

As contribuições do presente trabalho deverão ser: (i) Construir e disponibilizar um *dataset* contendo trechos de código reais com *API-related-usage bugs* retirados de projetos Java e JavaScript hospedados no GitHub <sup>1</sup>; (ii) Propor estratégias de recomendação para trechos de código com *API-related-usage bugs* para auxiliar o desenvolvedor na correção dos mesmos. Estas estratégias utilizarão funções de pré-processamento de código inéditas que estão sendo investigadas no presente trabalho de Doutorado.

## 4. Estado atual do Trabalho

O trabalho atualmente necessita de um *dataset* maior contendo contextos de código com *API-related-usage bugs*. A nossa ideia é retirar estes contextos de código de projetos hospedados no GitHub. Além disso, seria interessante realizar uma avaliação com sujeitos humanos para investigar se os posts do SO recomendados estão sendo úteis para auxiliar os desenvolvedores durante as atividades de correção destes *bugs*.

## 5. Análise de Resultados

Os resultados preliminares são encorajadores: 66.67% dos trechos de código Java com potenciais *API-related-usage bugs* tiveram suas correções encontradas nos top-10 resultados de busca. Considerando os trechos de código em JavaScript, correções foram encontradas para 40% deles. Estes resultados indicam que as nossas técnicas de busca de código superam o Google e o SO na pesquisa por correções de *API-related-usage bugs*.

## 6. Trabalhos Relacionados

Moreno et al. [Moreno et al. 2015] propuseram uma abordagem para minerar exemplos de código concretos que ensinam como utilizar um determinado método de API.

Ponzanelli et al. [Ponzanelli et al. 2014] propuseram uma abordagem que, dado um contexto de código na IDE, automaticamente retorna discussões pertinentes a partir do SO, avalia a relevância delas, e se, um dado limiar de confiança for ultrapassado, notifica o desenvolvedor sobre a ajuda disponível. A abordagem deles foi implementada em um *plug-in* para Eclipse IDE denominado PROMPTER.

## Referências

- Moreno, L., Bavota, G., Di Penta, M., Oliveto, R., and Marcus, A. How Can I Use This Method? In *Proceedings of the ICSE '15*, pages 880–890, Piscataway, NJ, USA. IEEE Press.
- Ponzanelli, L., Bavota, G., Di Penta, M., Oliveto, R., and Lanza, M. Mining StackOverflow to Turn the IDE into a Self-confident Programming Prompter. In *Proceedings of the MSR'14*, pages 102–111, New York, NY, USA. ACM.
- Robillard, M., Walker, R., and Zimmermann, T. (2010). Recommendation Systems for Software Engineering. *IEEE Software*, 27:80–86.

---

<sup>1</sup><https://github.com/>

# Sistemas Tutores Inteligentes Aplicados a Ambientes Educacionais Ubíquos: Proposta, Implementação e Avaliação

**Autor: Hiran Nonato Macedo Ferreira<sup>1</sup>,  
Orientador: Prof. Dr. Renan Gonçalves Cattelan<sup>1</sup>**

<sup>1</sup>Programa de Pós-Graduação em Ciência da Computação  
Universidade Federal do Uberlândia (UFU)  
Uberlândia – MG – Brasil

hirannonato@ufu.br, renan@ufu.br

Nível:Doutorado

Mês e ano de ingresso no programa: 09/2014

Mês e ano de previsão de conclusão: 08/2018

Data de aprovação do projeto de tese: 19/02/2016

**Resumo.** *Na literatura é possível encontrar diversos trabalhos que tratam Ambientes Educacionais Ubíquos (AEUs) e Sistemas Tutores Inteligentes (STIs) como duas linhas de pesquisa isoladas, focando em questões voltadas a ubiquidade e personalização de conteúdo, ou buscam explorar mais sobre questões de avaliação de conhecimento e comportamento. Dessa forma, existe pouca integração entre essas duas áreas. Para suprir essa lacuna, esta proposta pretende investigar a incorporação de recursos de STIs a AEUs. A solução proposta será composta por módulos de software fracamente acoplados e modelos de dados semânticos para representar e acompanhar o processo de aprendizagem do estudante por meio de análises de navegação, interação e comportamento. Estudos e avaliações serão conduzidos buscando identificar e compreender o impacto e a efetividade da abordagem proposta.*

**Palavras-Chave.** *AEUs, STIs, Modelo de Estudante Semântico*

## 1. Introdução e Motivação

A *Computação Ubíqua* é capaz de fornecer diversos subsídios para a construção de arquiteturas de software adaptáveis voltadas para educação. No entanto, é perceptível a necessidade de ambientes inteligentes que forneçam informações a essas aplicações. Informações como o nível de conhecimento do estudante são um importante requisito que pode auxiliar o ambiente ubíquo na personalização do conteúdo educacional. Descobrir quais são as capacidades do estudante, quais suas principais afinidades, em qual disciplina/conteúdo ele possui mais domínio e como o usuário se comporta durante a utilização do ambiente ainda são questões em aberto.

Uma das principais funções de um *Sistema Tutor Inteligente (STIs)* é auxiliar na tarefa de tentar descobrir o nível de conhecimento do estudante. Os *STIs* são formados por um conjunto de sistemas de software que dão suporte a atividades de aprendizagem. Aplicando técnicas de inteligência artificial, eles procuram se adaptar ao contexto do usuário e buscam agir da forma mais semelhante possível a um tutor humano (Zhiping *et al.*, 2012).

É possível encontrar na literatura diversas pesquisas que tratam sobre proposta e construção de *STIs* para os mais diversos fins. No entanto, percebe-se pouca integração desses estudos com as pesquisas em *AEUs*. Raras são as propostas para integrar capacidade de tutoria inteligente a cenários que exploram recursos ubíquos.

## 2. Fundamentação Teórica

Uma das propostas para apoiar atividades em sala de aula consiste no uso de *Ambientes Educacionais Ubíquos* (AEUs) (Settle *et al.*, 2011) – espaços físicos equipados com recursos computacionais (sensores, lousas eletrônicas, câmeras, microfones, entre outros), tendo potencial de incorporar a visão de *Computação Ubíqua* (Weiser, 1991) à área de educação. Dentre os diversos AEUs propostos na literatura, destaca-se o Classroom eXperience (CX)<sup>1</sup> (Ferreira *et al.*, 2012; Araújo *et al.*, 2013). O CX é uma plataforma multimídia para captura de aulas em uma sala instrumentada com dispositivos computacionais ubíquos. Desde 2012, na sua primeira versão, o CX tem se mostrado muito eficiente e eficaz para auxiliar atividades de ensino em ambiente instrumentado.

Outro tópico relacionado com essa pesquisa que merece atenção são os STIs. Estes podem personalizar o conteúdo educacional, de acordo com as informações modeladas sobre o desempenho, perfil e estados afetivos do aluno. Essa capacidade de adaptação do sistema é o que o torna inteligente. Segundo Woolf (2007), os STIs buscam raciocinar sobre o seu processo de aprendizado, entender suas necessidades individuais, fornecer representações alternativas de conteúdos, possibilitar diferentes caminhos de aprendizagem e formas de iteração e, por fim, compreender como a emoção influencia no processo de aprendizagem.

## 3. Contribuição do Trabalho

Ao final deste projeto de doutoramento, almeja-se a construção de uma arquitetura que incorpore recursos de STIs a AEUs. Serão desenvolvidos modelos e módulos de software – como, modelo do aluno, modelo de tutoria, módulo de acompanhamento de interações, módulo de gamificação, módulo de avaliação, módulo de acompanhamento de rastros e perfil de navegação – para apoiar essa integração. As soluções propostas utilizarão tecnologias de Web Semântica e ontologias para representação do conhecimento e estruturação dos dados. Possivelmente, serão utilizados também recursos de mineração de dados para auxiliar no processo de identificação e organização de perfis. Pretende-se, ainda, realizar experimentos mensurando diversos indicadores para avaliar o impacto da arquitetura em um ambiente real de ensino, integrando os modelos desenvolvidos e os módulos implementados ao CX.

## 4. Estado atual do Trabalho

É importante frisar que, até o momento, algumas etapas importantes desta pesquisa já foram iniciadas. Inicialmente foi realizado um levantamento bibliográfico acerca dos principais temas propostos. Uma segunda atividade deste projeto trata de abordagens para melhorar o engajamento dos estudantes ante a utilização de ambientes computacionais de ensino. Esta abordagem já está integrada ao ambiente e testes já estão sendo realizados. Outra atividade já em andamento é a definição de um Modelo de Estudante baseado em Ontologias e Redes Bayesianas voltado para ambientes adaptativos. Os próximos passos para continuação da pesquisa versarão sobre a integração do Modelo de Estudante ao ambiente Ubíquo, bem como a definição de regras e procedimentos para avaliar rastros e permitir um acompanhamento do progresso do estudante de forma mais acurada.

## 5. Análise de Resultados

As etapas já desenvolvidas apresentam diversos resultados satisfatórios. O Módulo de gamificação está integrado e em pleno uso na plataforma CX. Resultados da integração estão disponíveis

---

<sup>1</sup><http://cx.facom.ufu.br/>

em Ferreira *et al.* (2015). Outra etapa que apresenta bons resultados diz respeito a definição do Modelo de Estudante. Este foi proposto utilizando uma abordagem híbrida, por meio de ontologias e Redes Bayesianas. Resultados iniciais dessa etapa serão submetidos para publicação em eventos da área.

## 6. Trabalhos Relacionados

Diversos trabalhos propõem abordagens para aprendizado baseado em ambientes instrumentados e equipados com dispositivos eletrônicos para auxiliar no processo de ensino. Esses ambientes fornecem subsídios para que a aprendizagem aconteça independente de contexto físico, permitindo assim um ensino mais facilitado e atrativo (Truong e Hayes, 2009; Aihua, 2010). Pesquisas têm mostrado que esses ambientes trazem resultados positivos do ponto de vista educacional e pedagógico (Settle *et al.*, 2011; Kinshuk e Graf, 2012).

Uma das características que tornam os ambientes ubíquos uma linha de pesquisa promissora é sua capacidade de integração com outras áreas (Chu e Lin, 2013; Thiprak e Kurutach, 2015). Do ponto de vista colaborativo, alguns trabalhos exploram e fornecem ferramentas para compartilhamento do material educacional, anotações, resultados de aprendizagem, entre outros recursos que podem auxiliar os estudantes no processo de aprendizagem (Hwang *et al.*, 2014; Shih *et al.*, 2012).

Uma boa estratégia para organização e classificação de conteúdo educacional é provida, muitas vezes, pelos STIs, os quais se inspiram em técnicas de inteligência artificial para tentar encontrar o melhor modelo didático a ser disponibilizado individualmente para o aluno (Kusumawardani *et al.*, 2014; Chrysafiadi e Virvou, 2015). STIs têm sido empregado para auxiliar o processo de ensino em diversos domínios. Alguns exemplos de sucesso incluem: *Andes Physics Tutor* (Gertner e VanLehn, 2000) para ensino de física e informática; *AutoTutor* (Graesser *et al.*, 2001) para ensino física newtoniana; *PACT* (Koedinger *et al.*, 1997) para ensino de álgebra.

## Referências

- Aihua, Z. (2010). Study of Ubiquitous Learning Environment Based on Ubiquitous Computing. In *Proc. of IEEE U-Media*, p. 136–138, Jinhua, CN.
- Araújo, R. D., Brant-Ribeiro, T., Cattelan, R. G., Amo, S. A., e Ferreira, H. N. (2013). Personalization of Interactive Digital Media in Ubiquitous Educational Environments. In *Proc. of IEEE SMC*, p. 3955–3960, Manchester, UK.
- Baxter, E. (1990). Comparing conventional and resource based education in chemical engineering: student perceptions of a teaching innovation. *Higher Education*, 19(3):323–340.
- Chrysafiadi, K. e Virvou, M. (2015). Fuzzy Logic for Adaptive Instruction in an E-learning Environment for Computer Programming. *IEEE Transactions on Fuzzy Systems*, 23(1):164–177.
- Chu, H.-C. e Lin, C.-W. (2013). The Development and Application of a Repertory Grid-Oriented Ubiquitous Augmented Reality Learning System. In *International Conference on Advanced Applied Informatics*, p. 207–210.
- Diniz de Carvalho, S., Flores, E., Ramos de Melo, F., Fernando Batista Loja, L., e Bueno Pereira Carneiro, M. (2014). Hybrid Intelligent Tutoring System With Didactic Transposition Of The Subjects Guided By Expert Knowledge And Self Organizing Maps Neural Network. *Latin America Transac.*, 12(8):1539–1544.
- Ferreira, H., Dias Araújo, R., De Amo, S., e Cattelan, R. (2012). Classroom Experience: A Platform for Multimedia Capture and Access in Instrumented Educational Environments. In *Proc. of SBSC 2012*, p. 59–64.

- Ferreira, H. N. M., Araújo, R. D., Souza, P. C., Júnior, S. C., Dorça, F. A., e Cattelan, R. G. (2015). Gamificação em Ambientes Educacionais Ubíquos. *Proc. of 4<sup>o</sup> CBIE*
- Gertner, A. S. e VanLehn, K. (2000). Andes: A Coached Problem Solving Environment for Physics. In *Proc. of 5th ITS*, p. 133–142, London, UK.
- Graesser, A. C., VanLehn, K., Rosé, C. P., Jordan, P. W., e Harter, D. (2001). Intelligent Tutoring Systems with Conversational Dialogue. *AI Mag.*, 22(4):39–51.
- Hwang, G.-H., Chu, H.-C., Chen, B., e Cheng, Z. S. (2014). Development and Evaluation of a Web 2.0-Based Ubiquitous Learning Platform for Schoolyard Plant Identification. *Int. J. Distance Educ. Technol.*, 12(2):83–103.
- Kinshuk e Graf, S. (2012). Ubiquitous Learning. *Encyclopedia of the Sciences of Learning*, p. 3361–3363. Springer USA.
- Klasnja-Milicevic, A., Vesin, B., Ivanovic, M., e Budimac, Z. (2011). E-Learning personalization based on hybrid recommendation strategy and learning style identification. *Computers & Education*, 56(3):885 – 899.
- Koedinger, K. R., Anderson, J. R., Hadley, W. H., e Mark, M. A. (1997). Intelligent Tutoring Goes to School in the Big City. *International Journal of Artificial Intelligence in Education*, p. 30–43.
- Kusumawardani, S. S., Prakoso, R. S., e Santosa, P. I. (2014). Using Ontology for Providing Content Recommendation Based on Learning Styles inside E-learning. *2nd AIMS*, p. 276–281.
- Lesgold, A., Lajole, S., Bunzo, M., e Eggan, G. (1988). Sherlock: A Coached Practice Environment for an Electronics Troubleshooting Job. In J. Larkin, R. Chabay, e C. Scheftic, editores, *Computer assisted instruction and ITS*. Hillsdale N.J: LEA.
- Liu, T.-Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning*, 25(6):515–527.
- Settle, A., Dettori, L., e Davidson, M. J. (2011). Does Lecture Capture Make a Difference for Students in Traditional Classrooms. In *Proc. of 16th ITiCSE*, p. 78–82, Darmstadt, DE.
- Shih, S., Kuo, B., e Liu, Y. (2012). Adaptively Ubiquitous Learning in Campus Math Path. *Educational Technology & Society*, 15(2):298–308.
- Thiprak, S. e Kurutach, W. (2015). Ubiquitous computing technologies and Context Aware Recommender Systems for Ubiquitous Learning. *12th ECTI-CON*, p. 1–6.
- Truong, K. N. e Hayes, G. R. (2009). Ubiquitous Computing for Capture and Access. *Foundations and Trends in Human-Computer Interaction*, 2(2):95–171.
- Vinchurkar, D. e Sasikumar, M. (2015). Intelligent Tutoring System for Voice Conversion in English. *15th ICALT*, p. 314–316.
- Weiser, M. (1991). The computer for the 21st Century. *Scientific American*, 265(3):94–104.
- Woolf, B. P. (2007). *Building Intelligent Interactive Tutors: Student-centered Strategies for Revolutionizing e-Learning*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Zhiping, L., Yu, S., Tianwei, X., e Yang, L. (2012). The research of classical learner models in intelligent tutoring systems. *7th ICCSE*, p. 1204–1207.