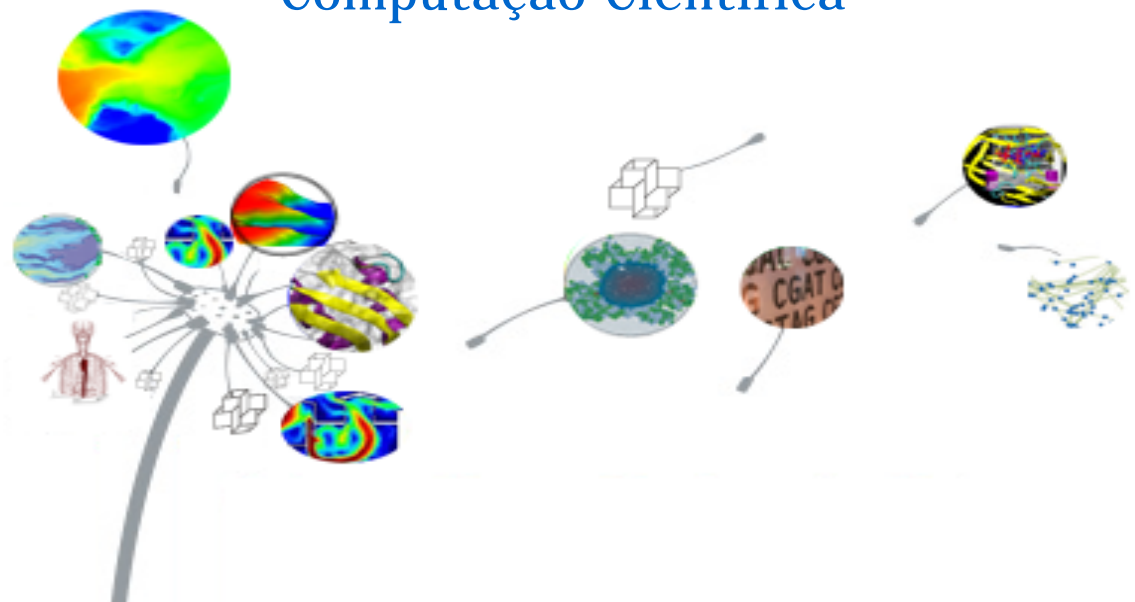


# Laboratório Nacional de Computação Científica

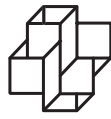


## XI Jornada de iniciação científica e tecnológica

### Livro de Resumos

PIBIC / PIBITI

05 de Outubro de 2016  
Av. Getúlio Vargas 333  
Petrópolis, RJ.



Laboratório  
Nacional de  
Computação  
Científica

XI Jornada de iniciação  
científica e tecnológica

# Livro de Resumos

Local de realização  
Hall do Auditório do  
Laboratório Nacional de Computação Científica

Endereço  
Av. Getúlio Vargas 333 - CEP 25651-075  
Tel.: (24) 2233-6150  
[www.lncc.br](http://www.lncc.br)

Quitandinha-Petrópolis, RJ  
05 de outubro de 2016

# **Jornada de Iniciação Científica e Tecnológica do LNCC**

Petrópolis, 05 de outubro de 2016.

## **Laboratório Nacional de Computação Científica – LNCC**

Diretor  
Augusto Cesar Gadelha

Coordenação de Administração - CAD  
Anmily Paula dos Santos Martins

Coordenação de Ciência da Computação - CCC  
Jauvane Cavalcante de Oliveira

Coordenação de Matemática Aplicada - CMA  
Frédéric Gerard Christian Valentin

Coordenação de Mecânica Computacional - CMC  
Márcio Arab Murad

Coordenação de Sistemas e Controle - CSC  
Carlos Emanuel de Souza

Coordenação de Sistemas e Redes – CSR  
Wagner Vieira Léo

Programa Institucional de Bolsas de Iniciação Científica &  
Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação  
Regina Célia Cerqueira de Almeida

## **Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq**

Presidente  
**Glaucius Oliva**

Coordenadora Geral do PIBIC/PIBITI  
**Lucimar Batista de Almeida**

# **Jornada de Iniciação Científica e Tecnológica do LNCC**

## **Comissão Interna do PIBIC/PIBITI-LNCC**

Regina Célia Cerqueira de Almeida  
Eduardo Lucio Mendes Garcia  
Helio José Correia Barbosa  
Jack Baczynski

## **Avaliadores Externos**

José Luis Drummond Alves - COPPE/UFRJ  
Priscila Vanessa Zabala Capriles Goliat - UFJF  
Regina Maria Maciel Braga Villela - UFJF

# Apresentação

O LNCC realiza este ano a XI Edição da Jornada de Iniciação Científica e Tecnológica, que é um fórum de divulgação das pesquisas desenvolvidas no contexto dos Programas Institucionais de Bolsas de Iniciação Científica (PIBIC) e de Bolsas de Iniciação Tecnológica (PIBITI) fomentados pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). No período de agosto de 2015 a julho de 2016, o PIBIC e PIBITI congregaram alunos de várias instituições de ensino e de diversas áreas do conhecimento. Este volume apresenta os resumos dos trabalhos desenvolvidos pelos bolsistas no período. Durante a Jornada, os trabalhos são apresentados pelos bolsistas em sessão de pôsteres e avaliados por um comitê científico externo e multidisciplinar.

Nesta XI Edição da Jornada, o Comitê Externo de Avaliação do PIBIC/PIBITI tem a seguinte composição:

Prof. José Luis Drummond Alves - COPPE/UFRJ

Prof<sup>a</sup>. Regina Maria Maciel Braga Villela - UFJF

Prof<sup>a</sup>. Priscila Vanessa Zabala Capriles Goliat - UFJF

Destacamos o papel relevante do PIBIC/PIBITI do LNCC no desenvolvimento das pesquisas no LNCC e, principalmente, na formação complementar dos bolsistas, promovendo o aprimoramento do conhecimento, espírito criativo, reflexão crítica e ética. Estas características têm contribuído para suas inserções no mercado de trabalho e em programas de pós-graduação, como o PPG em Modelagem Computacional do LNCC. Este é o resultado do esforço e dedicação de todos os participantes.

# Agradecimentos

Agradecemos ao CNPq pelas bolsas concedidas, à Direção do LNCC pelo apoio e à Comissão Interna do PIBIC e PIBITI no LNCC. Em particular, agradecemos ao Prof. Renato Portugal, que coordenou estes Programas no LNCC durante tantos anos, até junho de 2016.

Agradecemos a disponibilidade e contribuição dos membros do Comitê Externo de Avaliação.

O sucesso desta Jornada, e do Programa como um todo, é o resultado da dedicação e do esforço de toda a comunidade do LNCC. Expressamos em particular nosso reconhecimento ao apoio concedido pela secretaria do PPG-LNCC e, em particular, à Sra. Roberta Machado.

Regina Célia Cerqueira de Almeida  
*Coordenadora do PIBIC/PIBITI - LNCC*

# Índice

## **Bolsistas PIBIC com 1 ano ou mais**

- Ambiente para o apoio à execução de aplicações em *containers* em ambiente distribuído.....1  
Bolsista: Allan Matheus M. dos Santos  
Orientador: Bruno Schulze
- Modelagem e análise da malha aérea doméstica brasileira com grafos multiaspectos (MAGs)... .....2  
Bolsista: Bernardo Botelho  
Orientador: Artur Ziviani
- Simulação numérica e computacional de tráfego viário.....3  
Bolsista: Frederico Nascimento  
Orientador: Elson Toledo  
Coorientador: Regina C.P. Leal-Toledo
- Descoberta de objetos escondidos em Big Data.....5  
Bolsista: João Guilherme Rittmeyer  
Orientador: Fábio Porto
- Análise comparativa baseada em redes complexas dos efeitos da crise econômica na malha aérea doméstica brasileira.....7  
Bolsista: João Victor Marinho Bechara  
Orientador: Artur Ziviani
- Métodos numéricos aplicados à simulação de reservatórios petrolíferos.....8  
Bolsista: Lucas Carvalho de Sousa  
Orientador: Sandra Mara Cardoso Malta  
Coorientador: Cristiane O. Faria
- Controle ótimo em elasticidade.....10  
Bolsista: Monique Ribeiro da Costa  
Orientador: Jaime Rivera
- Particionamento de dados da astronomia utilizando os sistemas QEF e Hadoop.....11  
Bolsista: Rodrigo Botelho  
Orientador: Fabio Porto
- Investigação de polimorfismos de nucleotídeos únicos induzidos pela ação da Polimixina B e efeitos abióticos em *Klebsiella pneumoniae*.....14  
Bolsista: Thiago Cardoso Pereira Carneiro  
Orientador: Marisa Fabiana Nicolás  
Coorientador: Guadalupe del Rosario Quispe Saji
- Técnicas de otimização sem uso de derivadas .....16  
Bolsista: Viviane de Jesus Galvão  
Orientador: Helio José Correa Barbosa

### **Bolsistas PIBIC com menos de 1 ano**

Construção dos bancos de ligantes e alvos moleculares para o portal de triagem virtual DockThor-VS.....	17
Bolsista: Aldo Patrick Corrêa Orientador: Laurent Emmanuel Dardenne Coorientador: Isabella Alvim Guedes	
Atlas 3D de anatomia.....	19
Bolsista: Anna Caroline Wayand Martins Orientador: Jauvane Cavalcanti	
Modelagem geométrica e otimização de objeto 3D para ambientes virtuais colaborativos.....	21
Bolsista: Arthur Schulze Bittar Orientador: Jauvane C. de Oliveira	
Avaliação de sistemas de computação distribuída de alto desempenho: um estudo sobre estratégias para monitoramento de desempenho.....	23
Bolsista: Gabrieli Dutra Silva Orientador: Bruno Schulze	
Otimização de estratégias para futebol em simulação 2D através de técnicas de aprendizado de máquina.....	26
Bolsista: Lucas Borsatto Simão Orientador: André Salles Barreto	
<i>Bamba analytics</i> : uma estrutura para exploração de dados brasileiro para a biodiversidade marinha.....	35
Bolsista: Matheus Machado da Rosa Albuquerque Orientador: Luiz Manoel Rocha Gadelha Júnior Coorientador: Kary Ann del Carmen Ocaña Gautherot	
Aplicações e técnicas de aprendizagem de máquinas à análise de dados de domínio e de tempo de execução de workflows de bioinformática.....	37
Bolsista: Matheus Tonelli de Souza Orientador: Luiz Manoel Rocha Gadelha Júnior Coorientador: Kary Ann del Carmen Ocaña Gautherot	
Implementação da estrutura de dados que armazena a solução da equação de convecção-difusão no método HOPMOC.....	39
Bolsista: Thiago Daniel Quimas Simões Teixeira Orientador: Carla Osthoff Coorientador: Frederico Luís Cabral	

### **Ex-bolsistas PIBIC**

Modelagem molecular de inibidores potentes e seletivos para a enzima acetilcolinesterase.....	40
Bolsista: Ana Luiza Martins Karl Orientador: Laurent Emmanuel Dardenne Coorientador: Isabella Alvim Guedes	



Estudo de estratégias de vacinação de um modelo epidemiológico.....42  
Bolsista: Dérek Prates  
Orientador: Mauricio Kritz  
Coorientador: Jaqueline Maria da Silva

Tomografia por impedância elétrica.....43  
Bolsista: Douglas Toldo da Silva  
Orientador: Antonio André Novotny  
Coorientador: Andrey Dione Ferreira

Paralelismo e tolerância a falhas no ambiente SPiNMe.....44  
Bolsista: Emiliano Medeiros de Oliveira Neto  
Orientador: Antônio Tadeu Azevedo Gomes

Modelagem numérica de análise estática em estruturas flexíveis.....45  
Bolsista: Lucas M. R. Silva  
Orientador: Sandra M. C. Malta  
Coorientador: Cristiane O. Faria

MHOLline 2.0 – A evolução de um *workflow* científico para problemas em bioinformática e biologia estrutural.....47  
Bolsista: Victor Crisóstomo Cruz Reis  
Orientador: Laurent Dardenne  
Coorientadores: Priscila V. Z. Capriles Goliatt e Paulo M. Bisch

#### **Bolsistas PIBITI**

Otimização computacional do esquema de radiação na atmosfera em modelo numérico de previsão de tempo e clima.....49  
Bolsista: Matheus Silva de Melo  
Orientador: Roberto Souto

Análise de preditores de domínios proteicos.....50  
Bolsista: Matheus Miloski  
Orientador: Fábio Lima Custódio

Desenvolvimento de uma versão paralela híbrida para contabilização da frequência de repetição de k-mers.....51  
Bolsista: Micaella Coelho  
Orientador: Carla Osthof  
Coorientador: Fabricio Vilasbôas



## **Ambiente para o apoio à execução de aplicações em *containers* em ambiente distribuído**

Allan Matheus M. dos Santos, Bruno Schulze

Diversas aplicações são necessárias para o desenvolvimento de pesquisas apoiadas em computação científica. Uma mesma aplicação pode ser utilizada em diversas pesquisas, possivelmente em diferentes linhas de pesquisa. Entretanto, algumas aplicações podem não ser facilmente instaladas em um mesmo ambiente, por exemplo, por haver conflitos entre dependências. A criação de *appliances* de aplicações que ofereçam tudo o que é necessário para executar uma aplicação facilita a gestão do ambiente de execução e reduz o trabalho necessário para a realização de novas pesquisas.

Com o objetivo de utilizar a virtualização em nível de SO (Sistema Operacional) para apoiar a criação de *appliances* virtuais e a execução destes em sistemas distribuídos, este trabalho apresenta uma solução, composta pelos módulos Rufus-core e Rufus-orchestrator, que utilizam o LXC (*Linux Containers*) como base para a criação de uma nuvem privada que atenda as necessidades de pesquisadores, criando dinamicamente ambientes isolados dedicados a execução de aplicações científicas.

A abordagem desenvolvida neste trabalho é fruto de uma análise das infraestruturas de nuvem computacional existentes. Essa abordagem é usada no desenvolvimento de uma ferramenta de controle de *containers* para o portal VirtualIS (*Virtual Infrastructure for Science*) com objetivo de ampliar o controle da infraestrutura computacional de nuvem em apoio às aplicações científicas hospedadas pelo ComCiDis e pelo INCT-MACC.

Nesse sentido, o Rufus-core permite gerenciar *containers* remotamente. Além disso, é um sistema que auxilia o usuário na criação e na execução de *appliances* virtuais de aplicações. Enquanto o Rufus-orchestrator gerencia remotamente um grupo de instâncias de Rufus-core e escalona a execução de aplicações nestas instâncias. O Rufus-orchestrator também é um *framework* para que novos algoritmos de escalonamento sejam implementados.

A principal contribuição deste trabalho é facilitar o uso de *containers* através de ferramentas adaptáveis, permitindo que diferentes estratégias de granularidade e escalonamento sejam testadas. Para isso, os módulos principais da solução apresentada neste trabalho, Rufus-core e Rufus-orchestrator, podem ser utilizadas através de *scripts* em Python criados pelo usuário, que também pode acessar as funcionalidades dos módulos através de uma interface REST. Dessa forma, o usuário pode estender as funcionalidades do sistema para obter um maior controle sobre a execução das aplicações.



## **MODELAGEM E ANÁLISE DA MALHA AÉREA DOMÉSTICA BRASILEIRA COM GRAFOS MULTIASPECTOS (MAGs)**

Bernardo Botelho Antunes da Costa, Artur Ziviani (LNCC)

**Colaboradores:** João Victor Marinho Bechara, Klaus Wehmuth

**Resumo.** Grafos MultiAspectos (MAGs) são uma generalização de grafo capaz de representar redes complexas dinâmicas com múltiplas camadas, variantes no tempo, ou com ambas as características. Um conjunto de vértices, camadas, instantes de tempo ou qualquer outra característica independente presente na rede complexa pode ser considerado como um aspecto do MAG. Este trabalho apresenta a modelagem e análise de uma rede complexa dinâmica que representa a malha aérea doméstica brasileira através de um MAG. A estrutura resultante dessa modelagem é capaz de representar com naturalidade diferentes aspectos (aeroporto, companhia aérea e tempo) da malha aérea doméstica brasileira. Isso facilita diversas análises sob diferentes perspectivas bem como permite expressar de forma simples e compacta as diversas características dessa rede complexa com base em um único objeto matemático. Para a construção desse MAG foi utilizada uma base de dados disponível no sítio eletrônico da Agência Nacional de Aviação Civil (ANAC) que refere-se à estrutura da malha aérea brasileira em junho de 2015. Assim, é possível criar um único MAG representando a malha aérea doméstica brasileira, sendo esse MAG um objeto multicamada e variante no tempo. A análise realizada utiliza como métrica principal o *k*-core, que calcula o sub-grafo conexo maximal do grafo dado, onde os vértices têm grau maior ou igual a *k*. A utilização dessa métrica fornece resultados importantes aplicados ao problema, tais como a descoberta dos *hubs* de atuação das empresas aéreas, em quais regiões elas se concentram e quais os núcleos operacionais mais importantes.



## SIMULAÇÃO NUMÉRICA E COMPUTACIONAL DE TRÁFEGO VIÁRIO

Frederico Q. Nascimento<sup>1</sup>, Guilherme V.P. Magalhães<sup>1</sup>, Elson M. Toledo<sup>2</sup>, Regina C.P. Leal-Toledo<sup>3</sup>

<sup>1</sup>Instituto de Computação/UFF (bolsistas)

<sup>2</sup>LNCC (orientador), <sup>3</sup>Instituto de Computação/UFF (co-orientador)

### RESUMO

Nos dias de hoje a qualidade de vida dos cidadãos tem sido diretamente afetada com o aumento do fluxo de veículos tanto em rodovias quanto no meio urbano, dificultando sua mobilidade, aumentando o tempo de percurso entre suas diversas atividades, gerando *stress* e contribuindo para o aumento tanto da poluição sonora quanto ambiental. O entendimento da dinâmica do tráfego viário é de fundamental importância para que políticas de planejamento e gestão sejam realizadas de forma eficiente. Nesse contexto, modelos microscópicos de tráfego viário baseados em Autômatos Celulares (AC) têm tido destaque a partir dos anos 90 quando foi proposto o modelo probabilístico conhecido como modelo NaSch (Nagel, Schreckenberg) [1]. Apesar deste modelo conseguir representar as principais características do tráfego, como a mudança de fase do fluxo livre para o congestionado, ele não consegue representar bem a região de metaestabilidade e conduz, de forma geral, a um valor de fluxo máximo conservador quando utiliza valores de probabilidade sugeridos na literatura para representar a incerteza no comportamento do motorista [2]. A maioria dos modelos sugeridos posteriormente ou o cita, ou propõe alguma modificação com o objetivo de melhor reproduzir a dinâmica do tráfego viário. Desta forma, novos modelos têm sido propostos tanto para melhorar a relação fluxo-densidade quanto para reproduzir a metaestabilidade, ou para modelar outros comportamentos ou topologias específicos do tráfego viário.

Nos últimos anos alguns modelos têm surgido objetivando representar o comportamento dos motoristas para avaliar o efeito que esses diferentes perfis podem ter no fluxo do tráfego [3-5] e é nesse contexto que esse trabalho está inserido.

Todos os modelos de AC partem do pressuposto que o movimento de cada veículo é regulado pela distância ao veículo que está à sua frente e pela movimentação deste (ou dos dois seguintes), considerando uma Função de Densidade de Probabilidade uniforme para descrever a incerteza do comportamento do motorista e o mesmo conjunto de suposições é aplicado a todos os veículos da via. Recentemente foi proposto por esse grupo de pesquisa um modelo de AC [6, 7] para o tráfego viário que para a descrição do movimento do veículo modela, inicialmente, a percepção do motorista quanto ao movimento dos veículos à sua frente e, posteriormente, ajusta a velocidade deste tendo em vista esta percepção, considerando a incerteza do comportamento do motorista nessas duas etapas. Além disso, utiliza Funções de Densidade de Probabilidade (FDP) não uniformes, a FDP Beta, utilizando a Técnica de Rejeição do Método de Monte



Carlo, para simular a incerteza no comportamento do motorista. Esse tipo de abordagem tem permitido modelar de forma mais adequadamente diversos perfis de direção com qualidade e precisão.

Este trabalho apresenta estudos realizados para calibrar os parâmetros da FDP Beta para que essa possa representar com a qualidade desejada diferentes perfis de motoristas na direção e avalia possibilidades de inclusão de novas funcionalidades para representar novas topologias de vias.

## REFERÊNCIAS

- [1] NAGEL, K.; SCHRECKENBERG, M. A cellular automata model for freeway traffic. *Journal de Physique I*, v. 2, 2221–2229, 1992.
- [2] W. KNOSPE, W. et al. Towards a realistic microscopic description of highway traffic., *Journal of Physics A: Mathematical and General*, v. 33, n. 48, 477-484, 2000.
- [3] M. SAIFUZZAMAN AND Z. ZHENG. Incorporating human-factors in car-following models: A review of recent developments and research needs. *Transportation Research Part C*, 48, 379-403, 2014.
- [4] M. E. LARRAGA AND L. ALVAREZ-ICAZA. Cellular automaton model for traffic flow based on safe driving policies and human reactions. *Physica A: Statistical Mechanics and its Applications*, 389:5425-5438, 2013.
- [5] E.B. LIMA. Modelos microscópicos para simulação do tráfego baseados em autômatos celulares. *Dissertação de Mestrado, Pós-graduação em Computação/UFF*, 2007.
- [6] M. ZAMITH. Um Modelo de Autômato Celular Aplicado ao Tráfego Viário com Múltiplos Perfis de Condutores. *Tese de Doutorado, Pós-graduação em Computação/UFF*, 2013.
- [7] M. ZAMITH,; R. C. P. LEAL-TOLEDO; E. CLUA, E.M. TOLEDO, G.V.P. MAGALHÃES, A New Stochastic Cellular Automata Model for Traffic Flow Simulation with Driver's Behavior Prediction, *Journal of Computational Science*, v. 9, p. 51-56, 2015.



## Descoberta de objetos escondidos em Big Data

João Guilherme Rittmeyer, Fábio Porto

*Big Data* é uma área discutida, explorada pela indústria e academia pelas suas características em relação ao desempenho no processamento de grande volumes de dados. Este trabalho propõe a implementação de um *workflow* científico através do QEF[4] para integrar os *frameworks* Apache Spark<sup>1</sup>[3] e o Apache Hadoop<sup>2</sup>[2] com a solução encontrada em Khatibi et al[1].

Neste trabalho, estamos interessados em desenvolver técnicas para detecção de padrões geométricos em grandes volumes de dados. Consideramos padrões tais como: constelações e efeitos de lentes gravitacionais, em catálogos da astronomia, e domos de sal e falhas, em dados sísmicos. Os padrões de interesse são expressos através de consultas exemplos, definidas segundo um conjunto dos elementos do dataset que formem uma geometria com o padrão desejado.

A partir da consulta exemplo, procuram-se subconjunto que se assemelhem ao padrão geométrico fornecido.

Na astronomia a fim de encontrar padrões geométricos a partir de um grande volume de dados é preciso fazer o uso de uma ferramenta eficiente. Para isso, utiliza-se uma implementação paralela encontrada em Khatibi et al [1] desenvolvida sob os *frameworks Spark e Hadoop*

A aplicação realiza a leitura do catálogo contendo os objetos astronômicos e cria uma árvore de indexação multidimensional em memória, em *Java*, chamada *PhTree*<sup>3</sup>. A estrutura de indexação é utilizada para restringir o acesso aos possíveis candidatos à formação de conjuntos com o padrão definido, dentro de um raio de distância definido pela consulta. Os elementos candidatos são armazenados em um arquivo no sistema de arquivos distribuído HDFS e processados pela solução implementada por Khatibi et al [1].

A implementação da descoberta de objetos em grandes volumes de dados em Khatibi et al com a utilização do *Spark e Hadoop* integrada com o sistema QEF, permite processar conjuntos de dados bastante volumosos de forma rápida, eficiente e simples.

### Bibliografia

- [1] Khatibi, A. *Efficiently Computing Geometric Composition Patterns In Big Data*. Petrópolis, RJ, 2016.
- [2] White, T. *Hadoop: The Definitive Guide*. O'Reilly Media, Inc., 4st edition, 2015.
- [3] Karau, H., Konwinski, A., Wendell, P and Zaharia, M. *Learning Spark: Lightning-Fast Data Analysis*". O'Reilly Media, Inc, 2015.

---

<sup>1</sup> <http://spark.apache.org/>

<sup>2</sup> <http://hadoop.apache.org/>

<sup>3</sup> <https://www.phmtree.org>



- [4] Porto, F., Tajmouati, O., Silva, V. F. V. D., Schulze, B., and Ayres, F. V. M. *Qef - supporting complex query applications*. In *CCGRID '07: Proceedings of the Seventh IEEE International Symposium on Cluster Computing and the Grid*, pages 846–851, Washington, DC, USA. IEEE Computer Society, 2007.
- [5] Zschke, T. *The PH-Tree revisited*, 2016.



## **ANÁLISE COMPARATIVA BASEADA EM REDES COMPLEXAS DOS EFEITOS DA CRISE ECONÔMICA NA MALHA AÉREA DOMÉSTICA BRASILEIRA**

João Victor Marinho Bechara, Artur Ziviani

**Colaboradores:** Bernardo Botelho Antunes da Costa e Klaus Wehmuth

**Resumo.** A grave crise financeira enfrentada pelo país impacta diretamente diversos setores da economia, levando estes a se ajustarem face à nova realidade.

Nesse contexto, o setor de aviação civil não é exceção e houve por parte das companhias aéreas um ajuste na sua disponibilidade de voos, tendo por consequência uma recente modificação na malha aérea doméstica brasileira. Este trabalho modela e analisa a atual malha aérea doméstica brasileira, também comparando sua estrutura com dados anteriores da mesma malha em um intervalo pouco inferior a um ano. A malha aérea doméstica brasileira foi modelada através de Grafos MultiAspectos (MAGs), uma generalização de grafo capaz de representar redes complexas dinâmicas com múltiplas camadas, variantes no tempo, ou com ambas as características. A estrutura resultante dessa modelagem é capaz de representar com naturalidade diferentes aspectos (aeroporto, companhia aérea, tempo e versão da base de dados) da malha aérea doméstica brasileira. Isso facilita diversas análises sob diferentes perspectivas bem como permite expressar de forma simples e compacta as diversas características dessa rede complexa com base em um único objeto matemático. Para a construção desse MAG foram utilizadas duas bases de dados disponíveis no sítio eletrônico da Agência Nacional de Aviação Civil (ANAC) que referem-se à estrutura da malha aérea brasileira em junho de 2015 e em maio de 2016. Assim, é possível criar um único MAG agregando as duas bases de dados, sendo esse MAG um objeto multiescala, multicamada e variante no tempo. Com os resultados obtidos a partir da análise desse objeto é possível identificar as mudanças de estratégia das companhias aéreas durante esse período, como a variação do número de voos. A análise realizada utiliza como métrica principal o k-core, que calcula o sub-grafo conexo maximal do grafo dado, onde os vértices têm grau maior ou igual a k. Analisando os dados dos dois períodos considerados é possível avaliar a evolução da malha aérea como um todo em relação ao tempo, possibilitando o estudo do impacto das mudanças ocorridas entre os dois períodos e as diferentes estratégias adotadas por cada empresa em resposta a essas mudanças.





## Métodos Numéricos Aplicados à Simulação de Reservatórios Petrolíferos

Lucas Carvalho de Sousa<sup>1</sup>, Sandra M. C. Malta<sup>2</sup> Cristiane O. Faria<sup>1</sup>,

<sup>1</sup>Universidade do Estado do Rio de Janeiro, Departamento de Análise Matemática – IME

<sup>2</sup>Laboratório Nacional de Computação Científica

### RESUMO

Na área de exploração petrolífera em poços é comum utilizar-se de simulações computacionais para se tentar replicar fielmente o comportamento dos escoamentos dos fluidos presentes e tomar decisões sobre as estratégias de produção. A modelagem matemática, em geral, é dada por sistemas de equações diferenciais parciais complexos, obtidos por leis físicas bem estabelecidas e leis empíricas constitutivas bastante estudadas. No entanto, suas soluções analíticas somente são obtidas quando simplificações no modelo são impostas, e não nos são úteis em estudos de problemas mais realísticos. Por isso, a busca por soluções numéricas fisicamente satisfatórias tem sido uma opção frequente na solução desses problemas.

A presença de incertezas na descrição de variáveis relevantes, como por exemplo, viscosidade e permeabilidade, deve ser acrescentada nos estudo dos reservatórios para as previsões de comportamento mais próximo à realidade. Estas características são bastante relevantes e introduzem dificuldades no desenvolvimento de técnicas numéricas já utilizadas. Por isso, novas questões numéricas surgem e devem ser tratadas para a obtenção da solução numérica mais adequada.

Um escoamento monofásico incompressível estacionário em um meio saturado é modelado a partir das leis de conservação de massa e de Darcy, chamado de sistema de Darcy:

$$\frac{d(\rho \vec{v})}{dx} - \rho q = 0, \quad \text{em } \Omega,$$

$$\vec{v} = \frac{-d(Kp)}{dx} = 0, \quad \text{em } \Omega,$$

onde  $\vec{v}$  é a velocidade do fluido,  $p$  é a pressão,  $\rho$  é a densidade e  $q$  é a vazão volumétrica por unidade de volume[4]. Na lei de Darcy,  $K$  é o tensor de permeabilidade que depende da viscosidade  $\mu$ , da concentração e das propriedades de permeabilidade do meio  $k$ , e é representado por  $K = \frac{k}{\mu}$ . Os efeitos gravitacionais serão desprezados nesse trabalho.



Neste trabalho, um estudo numérico com os dois métodos mais frequentemente empregados em simuladores de poços, (Método das Diferenças Finitas[1,3] e Método dos Volumes Finitos[2,4]), avaliando a influência da permeabilidade em meios homogêneos e heterogêneos, com malhas regulares e irregulares, é apresentado para as soluções aproximadas do sistema de Darcy[1,2]. Comparações entre as duas metodologias são discutidas e ilustradas através de simulações numéricas considerando-se problemas unidimensionais.

### **REFERÊNCIAS**

- [1]-Biezuner, R. J., 2007. Métodos Numéricos para Equações Diferenciais Parciais Elípticas - Notas de Aula.
- [2]-Fortuna, A. O., 2000. Técnicas computacionais para dinâmica dos fluidos: conceitos básicos e aplicações. Edusp.
- [3]-LeVeque, R. J., 2007. Finite Difference Methods for Ordinary and Partial Differential Equations Steady-State and Time-Dependent Problems. SIAM.
- [4]-Rodrigues, J. R., 2015. Introdução à Simulação de Reservatórios Petrolíferos - Programa de Verão, LNCC 2015



### **Controle Ótimo em elasticidade**

Monique Ribeiro da Costa, Jaime E. Muñoz Rivera

Formulamos o problema de controle ótimo de uma viga no modelo de Euler Bernoulli, onde minimizamos um funcional quadrático que otimiza a deformação de acordo com uma função objetivo. Mostramos que o mínimo existe. Resolvemos o problema numericamente através do método de diferenças finitas, para encontrar os resultados numéricos usamos o MATLAB e sua plataforma gráfica para simular as deformações.



## Particionamento de dados da astronomia utilizando os sistemas QEF e Hadoop

Rodrigo Botelho, Fabio Porto

### Introdução

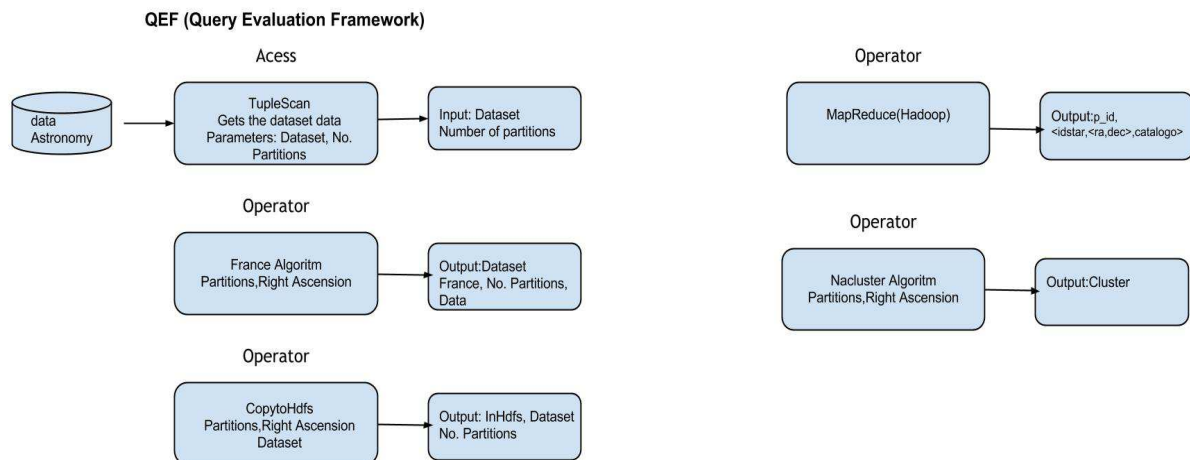
Vivemos um período de grande aumento na capacidade de armazenamento de dados eletronicamente, estamos na era dos “Zettabytes”. Contudo, esse aumento da capacidade de armazenamento não foi acompanhado pelo aumento da capacidade de leitura desses dados, acarretando um fator altamente desigual, pois em grandes volumes de dados perdemos muito tempo no processo de leitura, isso se utilizando de técnicas de processamento concentrado. Para contornarmos este problema utilizaremos técnicas e ferramentas de processamento distribuído.

### Desafio

Diferentes projetos de astronomia utilizam-se de distintos equipamentos telescópios, tais equipamentos possui configurações com relevantes diferenças como poder de visão, calibragem, tecnologia entre outras, esses telescópios mapeiam determinadas áreas do espaço obtendo seus corpos celestes em prol de seus projetos, com toda essa gama de projetos e seus catálogos relativos aos corpos celestes faremos a junção dessas diferentes áreas, contudo encontramos o problema de objetos celestes correspondente em áreas de fronteira com outros projetos.

### Objetivo

Este trabalho está focado em técnicas para particionar, processar e, finalmente, integrar os dados, tornando seu resultado final integro. Ocorre que estratégias de particionamento de dados tipicamente agem de duas formas: (i) divisão física, em blocos de dados de um certo tamanho, (ii) particionamento de valores em uma chave de identificação. Em vários cenários, no entanto, o critério de particionamento se apresenta de forma mais complexa.. O particionamento por si, pode ser visto como um processo em duas etapas: a definição do critério de particionamento, e identificação dos dados com as respectivas partições; e a alocação das partições sobre o sistema distribuído. Desta forma, este trabalho apresenta um dataflow para particionamento e alocação de dados em um sistema distribuído.



## Implementação

Framework QEF(Query Evaluation Framework), atua de maneira à gerenciar a execução do dataflow. sobre a implementação deste dataflow, iniciamos com as configurações do Framework QEF, onde alimentamos as informações de controle dentro do Query Execution Plan (QEP, arquivo xml, onde estão informações do plano de execução). no QEF, um “request” é um grafo de operações algébricas definidas pelo usuário, comunicando uns com os outros e com o objetivo de produzir um resultado. Estas operações estão representadas dentro do plano de execução. Posteriormente, implementamos nossas classes em java com tarefas específicas dentro da aplicação, estas classes serão nossa base dentro do plano de execução e irão acionar todas as tarefas de processamento dentro da aplicação.

## Aplicação

Faremos uso de uma gama de aplicações, a nível de hardware utilizaremos um cluster Petrus com um nó master e 6 nós slaves alocado de maneira distribuída, o sistema operacional em questão é Centos OS. Nas aplicações utilizaremos o Qef para encapsulamento de todo workflow. Para a fragmentação das partições faremos uso da aplicação FRANCE(fragmentador de catálogos espaciais) um algoritmo iterativo para particionar dados em espaços unidimensionais. Faz parte também de nossa aplicação o NAcluster(N-way Astronomical ) algoritmo de clusterização com o objetivo de dividir em clusters os grupos de objetos e apresentar N catálogos de modo que em cada agrupamento haja apenas objetos de diferentes catálogos e representando o mesmo objeto real. Por último com seu nível de relevância destacado dentro do trabalho a framework hadoop é seu sistema de arquivo distribuído HDFS(hadoop file system).



## Referências

WHITE, T. Hadoop: The Definitive Guide, 4th Edition, storage and analysis at Internet scale, O'Reilly Media, 2015. VENNEN, J.; SIDDALINGAIAH M.; WADKAR, S. Pro Apache

Hadoop, 2nd Edition, Apress; 2014. MURTHY, A. C.; EADLINE, D.; MARKHAM J.; NIEMIEC J.; KUMAR, V.V.

Apache Hadoop YARN, Moving beyond MapReduce and Batch Processing with Apache Hadoop 2, Addison Wesley, 2015

GASPAR, D.; PORTO F. FRANCE: Fragmentador de Catalogos Espaciais, HOSCAR Meeting, Gramado, 2014

OFREIRE, V. ; PORTO, F. ; MACEDO, J.A. ; AKBARINIA, R.

NACluster: A Nonsupervised Clustering Algorithm for Matching Multi Catalogues, eScience (eScience), 2014 IEEE 10th International Conference, 2014



## **Investigação de polimorfismos de nucleotídeos únicos induzidos pela ação da Polimixina B e efeitos abióticos em *Klebsiella pneumoniae***

Thiago Cardoso Pereira Carneiro, Guadalupe del Rosario Quispe Saji, Marisa Fabiana Nicolás

### Resumo

SNP ou *Polimorfismo de nucleotídeo único* pode ser caracterizado como a variação da seqüência observada em uma posição individual do genoma, e isso pode implicar de forma direta e relevante na formação de uma proteína. A detecção de SNP já acontece desde os primeiros sequenciamentos de fragmentos específicos [1]. A *Klebsiella pneumoniae* é uma bactéria gram-negativa, que tem como uma de suas características a resistência a antibióticos, tendo a polimixina B (ou colistina B) como ultima saída para tratamentos quando se trata de bactérias resistentes as betalactâmicos. Embora as estatísticas de várias partes do mundo estejam indisponíveis, evidências acumuladas ainda sugerem que a resistência às cefalosporinas de espectro estendido em *E. coli* e, bem como em *K. pneumoniae* tornaram-se um problema mundial [2]. Em um estudo recente utilizando *Pseudomonas aeruginosa* foi constatado que em determinados níveis de concentração de ions de ferro ocorre uma sub inibição que acarreta na facilitação de um excesso de captação desse metal pelas células bacterianas, com isso levando a um aumento da mutagênese [3]. Os resultados desse trabalho mostram também que há uma maior taxa dessa mutação se o Fe e a colistina estiverem presentes.

Tendo em vista essas informações o presente estudo tem como proposta investigar os SNPs induzidos pelo efeito da polimixina B conjuntamente com efeitos abióticos (incluindo a concentração elevada de Fe).

Os dados de transcriptoma analisados de *Klebsiella pneumoniae subsp. pneumoniae Kp13* foram gerados através de sequenciamento de nova geração (NGS) utilizando a plataforma illumina Hiseq ([www.illumina.com](http://www.illumina.com)), nas condições induzida (polimixina B) e concentrações elevadas de Fe e pH.

A busca por SNP no transcriptoma de Kp13 foi feita através de ferramentas de bioinformatica. Os dados gerados foram trimados utilizando o programa Skewer ([sourceforge.net/projects/skewer](http://sourceforge.net/projects/skewer)) e em seguida foram mapeados contra a referência através do programa Bowtie ([bowtie-bio.sourceforge.net](http://bowtie-bio.sourceforge.net)). Após o mapeamento, os dados gerados foram convertidos e ordenados através do programa Samtools ([samtools.sourceforge.net/](http://samtools.sourceforge.net/)) que também foi utilizado para indexar o arquivo fasta referencia.

A chamada de SNP foi feita pelo programa GATK ([software.broadinstitute.org/gatk](http://software.broadinstitute.org/gatk)) e o programa Picard ([broadinstitute.github.io/picard](http://broadinstitute.github.io/picard)), o qual gerou o dicionário para as seqüências.

O arquivo em formato bed foi gerado manualmente a partir do arquivo gff referencia. A lista contendo os SNPs foi usada para a criação de diagramas de venn contendo SNPs em comum entre as condições. O programa Igv ([software.broadinstitute.org/software/igv](http://software.broadinstitute.org/software/igv)) possibilitou uma melhor visualização desses SNPs dentro do mapeamento.

Finalmente, os SNPs foram classificados segundo sua incidência em uma região intergênica ou codificante (coding sequence, CDS) e quando caindo em CDS pelo tipo de mutação (sinônima ou não sinônima). Para o caso das mutações não sinônimas foi investigada a relevância da CDS quanto seu papel na resistência bacteriana aos antimicrobianos.



**Palavras Chave:** *Bioinformática, Klebsiella pneumoniae, SNPs.*

**Referencias:**

1-Orita, M, IWAHANA, H, KANAZAWA, H, et al. **Detection of polymorphisms of human DNA by gel electrophoresis as singlestrand conformation polymorphisms.** Proceedings of the National Academy of Sciences USA, v.86, n.8, p.2766-70, 1989.

2-Giske, Christian G. et al. **Clinical and Economic Impact of Common Multidrug-Resistant Gram-Negative Bacilli** . *Antimicrobial Agents and Chemotherapy* 52.3 (2008): 813–821. *PMC*. Web. 20 Sept. 2016.

3-Rodríguez-Rojas, Alexandro et al. **Cationic Peptides Facilitate Iron-Induced Mutagenesis in Bacteria.** Ed. Ivan Matic. *PLoS Genetics* 11.10 (2015): e1005546. *PMC*. Web. 20 Sept. 2016.





## **Técnicas de Otimização sem Uso de Derivadas**

Viviane de Jesus Galvão, Helio José Corrêa Barbosa

Pode-se encontrar nas Ciências Exatas e Engenharias problemas de otimização com restrições laterais, em espaço de busca contínuo, que possuem análise de sua diferenciabilidade inviável e sua avaliação com alto custo computacional. Estas dificuldades induzem o uso de uma classe de métodos apropriada: os métodos de Otimização sem Derivadas.

O trabalho aqui apresentado propõe a combinação de métodos estocásticos e determinísticos para resolver os tipos de problemas citados. Como métodos estocásticos foram estudadas as metaheurísticas de Otimização por Enxame de Partículas e Evolução Diferencial. A Otimização por Enxame de Partículas é um método que foi inspirado no comportamento social de bandos de pássaros na procura de alimentos, simulando os mecanismos de inteligência coletiva. A Estratégia Evolutiva é um método baseado no modelo biológico de processos de adaptação e evolução de espécies, simples e eficiente, voltado para resolver problemas definidos em domínio contínuo, guiado por mutação de vetor de diferenças. A estratégia determinística adotada aqui é uma Busca Padrão, um método que não faz uso das derivadas da função objetivo e que opera movimentos exploratórios através de dois passos: o passo de sondagem do espaço e o passo de avaliação, que é executado se o passo de sondagem falhar. É proposto também modificar o passo de avaliação da Busca Padrão adotada aqui com o intuito de realizar menos cálculos da função objetivo, fazendo a memorização da última coordenada na qual se obteve sucesso e, na sua próxima execução, a busca começa pela coordenada posterior.

A proposta da combinação dessas duas classes de métodos induz a junção das características positivas de tais, com o intuito de usar menos cálculos da função objetivo e procurando obter os melhores resultados possíveis.

Experimentos computacionais são feitos para a resolução de 100 problemas-testes disponíveis na literatura em linguagem AMPL e os resultados encontrados são utilizados para estudar, a partir de análise por Perfis de Desempenho, o impacto no desempenho pela incorporação da Busca Padrão nas metaheurísticas consideradas aqui.



## **Construção dos Bancos de Ligantes e Alvos Moleculares para o Portal de Triagem Virtual DockThor-VS**

Aldo Patrick A. Corrêa, Laurent Emmanuel Dardenne e Isabella Alvim Guedes.

### **Introdução**

A redução de custos e tempo no desenvolvimento de novos fármacos mostra-se indispensável em todos os aspectos. Nesta busca por novos candidatos a fármacos, as técnicas de triagem virtual em larga escala mostram-se indispensáveis, possibilitando a análise computacional de um grande número de moléculas e a seleção das mais promissoras<sup>1</sup>. O programa DockThor, desenvolvido pelo Grupo de Modelagem Molecular em Sistemas Biológicos (GMMSB),<sup>2-4</sup> tem obtido excelentes resultados em estudos de *docking molecular*. O sucesso neste desenvolvimento viabiliza o uso do programa DockThor em estudos de triagem virtual em larga escala, principalmente no portal de triagem virtual DockThor-VS. O presente projeto de Iniciação Científica tem como principal objetivo construir, preparar e otimizar as estruturas bidimensionais e tridimensionais de pequenas moléculas que compõem os bancos de ligantes dos diversos grupos de pesquisa em colaboração com o desenvolvimento do portal DockThor-VS. A construção de bancos de ligantes envolve o desenho, otimização e parametrização das suas estruturas 2D e 3D, sendo uma etapa fundamental no desenvolvimento do portal de triagem virtual em larga escala DockThor-VS.

### **Metodologia**

Este trabalho envolveu a construção de 4 bancos de ligantes, totalizando 114 moléculas sintetizadas por grupos colaboradores em instituições de pesquisa como a Fundação Oswaldo Cruz para o tratamento da Doença de Chagas<sup>5-8</sup>. Os desenhos da estrutura 2D dos compostos foram realizados através do programa MarvinSketch (<https://www.chemaxon.com/products/marvin/marvinsketch/>). Isômeros, estados de protonação e tautômeros dos ligantes foram preditos com a ferramenta Ligprep/Epik da suíte de aplicativos Maestro<sup>9</sup>, utilizando diferentes faixas de pH e a presença de metais. Ademais, todos os compostos, que anteriormente não possuíam estrutura tridimensional disponível, tiveram sua estrutura 3D otimizada por meio da minimização de energia no vácuo utilizando o campo de força OPLS2005.

### **Resultados**

A preparação dos quatro bancos de ligantes gerou um total 918 estruturas tridimensionais curadas, adequadas para estudos de triagem virtual considerando diversas condições experimentais e diferentes alvos moleculares. As estruturas serão disponibilizadas em formatos amplamente utilizados, como .mol2 e .sdf (estrutura tridimensional), .SMILE (estrutura unidimensional) e como ilustração. Os arquivos serão disponibilizados no portal de triagem virtual em larga escala DockThor-VS.



## Conclusões

A construção e preparação de pequenas moléculas, gerando arquivos de entrada curados para diversas condições experimentais diferentes, mostram-se extremamente importantes para experimentos de triagem virtual influenciando diretamente nos resultados obtidos. A próxima etapa deste projeto será a preparação de diversas moléculas receptoras, consideradas como potenciais alvos terapêuticos de interesse da comunidade científica.

## Referências:

Grey, J. L. & Thompson, D. H. Challenges and opportunities for new protein crystallization strategies in structure-based drug design. *Expert Opin. Drug Discov.* 5, 1039–1045 (2010).

Magalhães, C. S. de. Algoritmos genéticos para o problema de Docking proteína-ligante. (2006).

Magalhães, C. S., Barbosa, H. J. C. & Dardenne, L. E. in *Genetic and Evolutionary Computation – GECCO 2004* (ed. Deb, K.) 3102, 368–379 (Springer Berlin Heidelberg, 2004).

Magalhães, C. S., Almeida, D. M., Barbosa, H. J. C. & Dardenne, L. E. A dynamic niching genetic algorithm strategy for docking highly flexible ligands. *Inf. Sci.* 289, 206–224 (2014).

Jardim, Guilherme A. M. On the investigation of hybrid quinones: synthesis, electrochemical studies and evaluation of trypanocidal activity. (2015).

Carvalho, Samir A. Studies toward the structural optimization of new brazilonerelated trypanocidal 1,3,4-thiadiazole-2-arylhydrazone derivatives. (2008).

Silva, Ramon B. da. Synthesis and Trypanocidal Activity of Novel 2,4,5-Triaryl-N-Hydroxylimidazole Derivatives. (2013).

Emilay B. T. Diogo. Synthesis and anti-Trypanosoma cruzi activity of naphthoquinone-containing triazoles: Electrochemical studies on the effects of the quinoidal moiety. (2013).

Shelley, J. C. *et al.* Epik: a software program for pK( a ) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 21, 681–691 (2007).



## **ATLAS 3D DE ANATOMIA**

Anna Caroline Wayand Martins, Jauvane Cavalcanti

### **RESUMO**

O estudo da Anatomia do corpo humano (suas estruturas e sistemas) consiste na base para o entendimento do seu comportamento fisiológico e para a compreensão das doenças, seus mecanismos fisiopatológicos, diagnóstico e tratamento. Torna-se, portanto, indispensável no currículo dos cursos de das Ciências da Saúde, entre elas a Medicina, na qual está presente de forma geral na grade curricular dos dois primeiros anos do curso.

Os acadêmicos das áreas da Saúde têm acesso ao estudo da Anatomia Humana através de livros, que proporcionam embasamento teórico, e através de cadáveres e peças anatômicas, para o embasamento prático. No entanto, para o estudo prático, é necessário que haja peças no ambiente acadêmico, que as mesmas estejam disponíveis, e sejam dissecadas de forma a apresentarem as estruturas nas quais o aluno necessita. Quando tais fatores não são possíveis, dificultam o estudo prático da disciplina.

É possível, no entanto, substituir um cadáver por um sistema de realidade virtual que apresente as estruturas anatômicas em 3D, em um atlas de Anatomia Interativo em 3D, através de Ambientes Virtuais Colaborativos (AVCs), o que exige a agregação de uma grande quantidade de informação a ser utilizada como explicação das partes anatômicas, com base em livros publicados sobre o conteúdo para que o mesmo tenha relevância médica e para as demais áreas da saúde, tornando-o um meio não de total substituição do estudo em cadáveres, mas uma via alternativa e funcional para o estudo da Disciplina.

Primeiramente, dividiu-se o conteúdo anatômico em 2 abordagens: uma com visão geral e outra com visão aprofundada. O conteúdo geral propicia informações sobre os sistemas do organismo humano, divididos em esquelético, muscular, cardiovascular, nervoso, digestório, respiratório, urogenital feminino e masculino, endócrino e sensorial, com conteúdo abrangente, no entanto não aprofundado. Já o conteúdo com visão aprofundada detalha as estruturas apresentadas na visão geral e apresenta aspectos clínicos relevantes sobre tais estruturas, de acordo com o curso da área da saúde.

A partir do recolhimento das informações, parte-se para o desenvolvimento de um modelo instrucional a ser utilizado no Ambiente Virtual, além de caracteres técnicos com inserção das informações coletadas e dos processos de interação desenhados no modelo instrucional, enviado periodicamente ao Laboratório Nacional de Computação Científica, para que seja realizada sua compilação, formatação e inclusão no sistema de navegação e ambiente em 3D.



Com a finalização do modelo anatômico virtual, objetiva-se a realização de teste do sistema com membros da comunidade médica e demais áreas da saúde, para que o mesmo seja relevante ao aprendizado da Anatomia.

## **REFERÊNCIAS**

SOBOTTA, J. Atlas de anatomia humana. 23ª ed. Rio de Janeiro: Guanabara Koogan, 2012. 3v.

GRAY, H. Anatomia. Rio de Janeiro: Guanabara, 1988.

MOORE, K. L; DALLEY, A. F. Anatomia orientada para a clínica. 6ª ed. Rio de Janeiro: Guanabara, 2013.

NETTER, F. H. Atlas of human anatomy. 6th ed. Philadelphia: Elsevier Inc, 2014.

YOKOCHI, C; ROHEN, J. W. Anatomia humana atlas fotográfico de anatomia sistêmica e regional. 7ª ed. São Paulo: Manole, 2010.

ELLIS, H.; LOGAN, B. M.; DIXON, A. K. Anatomia seccional humana: Atlas de secções do corpo humano, imagens por TC e RM. 3ª ed. São Paulo: Santos, 2010.

MACHADO, A. B. M. Neuroanatomia funcional. 3ª ed. São Paulo: Atheneu, 2013.



## **Modelagem Geométrica e Otimização de Objeto 3D para Ambientes Virtuais Colaborativos**

Arthur Schulze Bittar, Jauvane C. de Oliveira

### **HISTÓRICO**

Utilizo técnicas comuns de modelagem poligonal (low poly) para desenvolver assets que possam ser usado em ambientes virtuais. Aplico shaders básicos (Phone, Blinn, Lambert) e eventualmente diferentes mapas de texturas (Albedo, Normal, Height) quando necessário para simular materiais com maior detalhe. Muitos dos modelos desenvolvidos têm ou tiveram várias versões ao longo de sua elaboração, mudando de acordo com as alterações de briefing dos projetos, passando por retopologia ou decimação de malha para melhoria de desempenho nas máquinas.

### **RESUMO**

Desenvolvimento e reiteração de modelos tridimensionais de acordo com a necessidade dos projetos em curso no Laboratório ACiMA.

### **OBJETIVO**

A tarefa original do bolsista consiste na geração de modelos geométricos de objetos 3D necessários aos projetos do Laboratório ACiMA, incluindo a necessária otimização das malhas e exportação nos formatos necessários a cada sistema. Tais objetos serão posteriormente importados dentro dos vários AVCs (Ambientes Virtuais Colaborativos).

### **ATIVIDADES**

Tenho me dedicado ao desenvolvimento de modelos geométricos conforme a demanda do Laboratório ACiMA. Tal modelagem 3D inclui também a otimização de malhas com o intuito de permitir um uso mais eficiente nos sistemas desenvolvidos, assim como quaisquer alterações requisitadas pelos integrantes do laboratório, como decimação de malhas, cortes de geometria, recolocação de pivots, etc. Também venho importando os modelos desenvolvidos na plataforma Unity3D e exportando pacotes prontos para uso, diminuindo o ruído entre desenvolvimento e implementação por parte da equipe.

### **Funções exercidas além de modelagem**

Durante o processo de construção e modelagem, etapas usualmente desenvolvidas por uma equipe tem sido realizadas pelo bolsista, tais etapas incluem: busca de referências e concepts; modelagem; texturização; deformações, blendshapes ou outras animações simples; fechamento de arquivos.

### **Problemas encontrados durante o desenvolvimento das atividades**

Limitações de hardware, especificamente o Phantom, que não sustenta todo o potencial de trabalho com geometria oferecido pela plataforma Unity3D (observado no projeto de cirurgia de catarata de Letícia Fonseca). Também seria interessante saber mais do projeto de antemão para diminuir a recorrência de



remodelagem, ou receber um briefing fechado com referências e desenhos/clippings de material que deve ser desenvolvido (visual aids), visto que a maioria dos pedidos é repassado verbalmente.

#### MODELOS DESENVOLVIDOS DURANTE O PERÍODO DA BOLSA

Durante a bolsa foram desenvolvidos vários assets.

- Foi reimportado e manipulado o aeroporto e diversos de seus elementos, criada uma pista de decolagem, prédios ao entorno, hangares e torre de comando. O objetivo do projeto era simular a situação de embarque para pessoas que têm pavor de viajar de avião.
- Foi criado um porta-eletrodo e uma sala simples para uma simulação de aprendizado de métodos de solda com catodo.
- Também foi modelado um olho para simulação de cirurgia de catarata, com várias camadas e elementos de marcação de corte, lente fragmentada e uma lente IOL de reposição do cristalino.
- Por último, foi encontrado um modelo freeware de uma cidade básica, ao qual foi adicionado um prédio, um elevador panorâmico, sinalização de rua, postes, lixeiras e uma praia. O objetivo desse projeto é simular uma situação para pessoas com pavor de altura, e ainda está em andamento.



## **Avaliação de Sistemas de Computação Distribuída de Alto Desempenho: Um Estudo sobre Estratégias para Monitoramento de Desempenho**

Gabrieli Dutra Silva, Bruno Schulze

### 1. Introdução

A Computação Científica Distribuída de Alto Desempenho (HPDC) é um segmento da ciência da computação que tem como objetivo a melhoria do desempenho de aplicações distribuídas e paralelas, utilizando complexas infraestruturas computacionais. Com ela é possível o processamento de grandes volumes de dados experimentais criados por uma nova geração de instrumentos e aplicações científicas, as quais exigem grande largura de banda, redes de baixa latência e recursos computacionais de alto desempenho. Esse tipo de infraestrutura tem se tornado crucial para a pesquisa científica em muitos domínios de investigação. Um desses domínios é a indústria de energia, cujas aplicações envolvem intensivas simulações numéricas, fundamentais na exploração de novas fontes de energia, tais como, energia eólica e petróleo e gás. Porém, cada domínio de investigação tem aplicações com diferentes requisitos computacionais, os quais dependem da definição adequada dos sistemas de HPDC envolvidos para se obter a eficácia e eficiência na resolução dos seus problemas. Assim, o tomador de decisão (o pesquisador) se depara com decisões complexas e que na maioria das vezes não fazem parte da sua especialidade, pois avaliar o desempenho das aplicações nessas infraestruturas não é tarefa trivial.

Uma nova abordagem proposta por [Ferro:2015], busca tratar esse tipo de problema avaliando o desempenho de uma infraestrutura de HPDC por meio de uma metodologia, voltada para as características das aplicações científicas. Essas aplicações possuem características bastante distintas em termos dos requisitos computacionais exigidos em sua execução. Resultados experimentais mostram que as diferentes classes de aplicações, quando executadas em diferentes arquiteturas de HPDC e combinadas a diversos Elementos Essenciais de Análise (EEA), os quais correspondem aos parâmetros relevantes a serem avaliados na metodologia proposta, podem levar a resultados de desempenho completamente distintos. Essas análises de desempenho vêm sendo investigadas e os resultados demonstram a sua importância, bem como a de uma metodologia formal que auxilie a avaliar a eficácia e eficiência operacional das arquiteturas de HPDC voltadas à execução de aplicações científicas.

Com base no que foi exposto, o objetivo do plano de trabalho é investigar diferentes abordagens de monitoramento, coleta e análise de desempenho em ambientes paralelos e distribuídos de alto desempenho e que viabilizem a coleta e análise dos EEA. Assim, é possível definir com segurança os EEA, voltados a cada aplicação científica, e definir a maneira mais eficiente para coleta dos seus valores operacionais, pois essas definições são fundamentais para a conclusão da metodologia proposta.

### 2. Atividades Realizadas

De acordo com o objetivo proposto, nesses quatro meses iniciais foi realizado um estudo sobre a importância de analisar o desempenho de aplicações paralelas, além de assimilar como essa análise pode ser feita comparando duas abordagens de monitoramento, a *online* e a *offline*. Também buscou-se compreender alguns conceitos existentes em ambientes HPC, importantes para a análise de desempenho de aplicações. As próximas subseções abordam cada um dos estudos aqui mencionados.





## 2.1. Importância da análise de desempenho de programas paralelos

A análise de desempenho de aplicações é um importante fator contribuinte para a metodologia proposta, mencionada na Seção 1, pois possibilita definir com segurança os EEA e a maneira mais eficiente para coleta dos seus valores operacionais. Além disso, com as análises e medidas realizadas, é possível definir os requisitos de software, hardware e infraestrutura que viabilizem a exaescala da computação, contribuindo com o projeto *High Performance Computing for Energy (HPC4E)*<sup>4</sup> no qual este trabalho se insere.

O projeto *HPC4E* visa o desenvolvimento de simulações altamente escaláveis que viabilizem a geração de energia de maneira mais eficiente. Entre seus objetivos técnicos está a análise de desempenho de aplicações científicas para identificar possíveis gargalos na sua execução, bem como a eficiência no consumo de energia elétrica dos códigos e quais seus principais requisitos em diferentes arquiteturas computacionais. Portanto, para alcançar os objetivos do projeto, monitorar aplicações e compreender seus requisitos computacionais são tarefas indispensáveis.

Outro fator de grande importância é a avaliação continuada dos ambientes computacionais de HPC, verificando seus níveis de desempenho, possíveis falhas e necessidades de melhorias. Com isso os custos com manutenção poderiam ser previstos e não haveria tanto impacto na eficiência dos ambientes de HPC. Porém, realizar todas essas avaliações não é tarefa trivial, pois cada um desses aspectos envolvem um conjunto de parâmetros diferentes, obtidos por meio de um grande conjunto experimental. Por exemplo, na avaliação das aplicações individuais, para caracterizar os seus requisitos computacionais é necessário medir os percentuais e tempos de utilização de recursos (CPU, E/S, Memória, etc.), necessários para a sua execução, além de medidas do sistema que permitam a sua avaliação continuada e também o consumo de energia elétrica durante a execução das aplicações.

## 2.2. Conceitos Fundamentais: Efeito de sonda e intrusão

Todo processo de análise de comportamento de uma aplicação passa por duas fases primárias, a coleta de dados e a análise dos mesmos. Na fase de coleta de dados são registradas informações consideradas pertinentes para se entender o comportamento da aplicação. Existem várias técnicas diferentes, que serão citadas no próximo resumo de atividades, que podem ser utilizadas para realizar o registro dessas informações comportamentais.

O tempo gasto no registro do comportamento é chamado de efeito de sonda, ou seja, é a intrusão causada pela sonda de monitoramento para registrar o comportamento [Schnorr:2014]. Este efeito sonda pode ser tanto a nível de inserir instrumentação na aplicação, quanto ao fato do software utilizado para o monitoramento consumir recursos computacionais juntamente com aplicação. Estas duas formas de sonda podem, respectivamente, modificar o comportamento da aplicação ou impactar nos valores de utilização de recursos coletados.

Portanto, procura-se sempre que o efeito de sonda seja o menor possível, pois é difícil anular completamente a intrusão causada pela coleta. O objetivo é sempre coletar os dados de desempenho com a aplicação o mais próximo possível de seu comportamento natural.

---

<sup>4</sup> <https://hpc4e.eu/>



### 2.3. Abordagens de monitoramento

Levando em consideração as duas fases primárias para monitorar aplicações, existem duas formas de combiná-las: na abordagem *online* tanto a coleta quanto o procedimento de análise são realizados simultaneamente, e ferramentas de monitoramento, são executadas em paralelo com a aplicação para observá-la em tempo real. A abordagem *offline* realiza as duas fases separadamente e a análise ocorre sempre após o fim da execução da aplicação. Neste tipo de análise para monitorar o comportamento da aplicação, normalmente é necessário modificar seu código fonte inserindo diretivas (instrumentação).

Porém, ambas as abordagens possuem vantagens e limitações quanto a sua utilização. Entre as vantagens da análise *online* estão a ausência do custo de manutenção de dados e a interatividade na análise dos mesmos. E entre suas limitações estão o custo da transferência dos dados e o fato de os recursos computacionais não estarem dedicados apenas à aplicação, uma vez que o sistema de monitoramento também faz uso de tais recursos. Além disso, as ferramentas de monitoramento *online*, na maioria das vezes, tem alta complexidade, o que dificulta a instalação e o uso, ou exigem grande custo computacional para experimentos que muitas vezes deveriam ser rápidos e práticos de serem executados. Já na abordagem *offline* as vantagens e limitações são opostas àquelas da análise *online*, mas pode-se considerar como principal fator limitante a dificuldade de inserir as diretivas nos locais adequados dos códigos das aplicações, pois as instrumentações devem estar em locais estratégicos para não alterar seu funcionamento.

### 3. Atividades futuras

Para os próximos meses, ferramentas que fazem uso da abordagem *online* serão estudadas, como por exemplo Nagios, Paraver e Extrae. O objetivo é compreender qual técnica de coleta de dados essas ferramentas utilizam e que tipo de parâmetros podem ser coletados. Para a abordagem *offline*, serão realizados estudos com a inserção manual da instrumentação em aplicações e a ferramenta PAPI também será analisada, a fim de verificar a possibilidade de fazer a instrumentação com o auxílio desta ferramenta.

### Referências

- Ferro, M. (2015, Maio). *Avaliação de Sistemas de Computação Massivamente Paralela e Distribuída: Uma metodologia voltada aos requisitos das aplicações científicas*. Tese de doutorado. Laboratório Nacional de Computação Científica, Petrópolis/ RJ.
- Schnorr, L. M. (2014). Análise de desempenho de programas paralelos. Sociedade Brasileira de Computação (Ed.), *XIV Escola Regional de Alto Desempenho do estado do Rio Grande do Sul*, Alegrete, RS, Brazil, página. 57- 82.



## **Otimização de Estratégias para Futebol em Simulação 2D Através de Técnicas de Aprendizado de Máquina**

Lucas Borsatto Simão, André Salles Barreto

### **Introdução**

Uma das áreas da computação que podem ser destacadas por mais ter crescido em importância nos últimos anos é o aprendizado de máquina. Isso ocorre por conta de o número de aplicações de métodos dessa área de pesquisa serem variados, abrangendo problemas que vão desde reconhecimento e descrição de imagens até definição de uma sequência ações que um robô.

Nos últimos anos a área de aprendizado tem apresentado avanços consideráveis, como o desafio de Go ganho pelo algoritmo da DeepMind, denominado AlphaGo [1], que há anos vinha sendo tratado como o maior desafio para muitos pesquisadores. Outro destaque que foi apresentado foi a aprendizagem de jogos de Atari somente através de pixels [2], tratando como entrada para o algoritmo somente a tela de jogo e os pontos que o agente adquiria ao longo de uma partida.

Um problema bastante também bastante explorado hoje é o de simulação de futebol em 2D, categoria gerenciada pela RoboCup, que tem como um dos principais objetivos servir como ferramenta de desenvolvimento para algoritmos de aprendizado de máquina. Por exemplo, em [3] Peter Stone e Richard Sutton descrevem um algoritmo conhecido como Sarsa( $\lambda$ ) para determinação de quais as melhores ações para manter a bola longe dos agentes adversários. Já em [4] os autores se utilizam do algoritmo MAXQ, uma espécie de árvore hierárquica, com busca através de métodos heurísticos para descobrir qual a melhor ação a ser tomada durante um jogo.

A categoria se apresenta como possibilidade de aplicação para tais algoritmos por que possui muitas características interessantes e que a fazem se assemelhar a um sistema real, o que é essencial para testes de efetividade de um algoritmo.

Assim como os trabalhos mencionados, este irá abordar técnicas de aprendizagem que tem como finalidade a escolha de ações ótimas para agentes envolvidos em uma partida simulada de futebol desta categoria da RoboCup. O objetivo é verificar qual a melhor técnica que se mostra mais apta a se adaptar para vencer uma partida contra um oponente.



## Descrição do problema

A categoria *RoboCup Soccer Simulation 2D* simula um ambiente que se assemelha com um jogo de futebol real. Cada um dos agentes representados em campo é um programa independente executado por um servidor, que os gerencia e rege as situações de jogo. Através da figura 1 podemos ver a representação de uma partida, em que cada um dos círculos representa um agente.



Figura 1 - Representação de uma partida da *RoboCup Soccer Simulation 2D*

Os agentes possuem características inerentes a um jogador, como estamina, esforço e recuperação [5]. Os jogadores podem ter comunicação entre si, mas ela é limitada a direção para a qual o jogador que manda a mensagem esta virado e a uma distância que é determinada aleatoriamente pelo servidor.

Os agentes ainda possuem outras características que condizem com o que esta presente em um jogador de uma partida convencional, como direção e força no toque de bola, direção para a qual andar ou correr e a função de pegar a bola.

Além disso há a representação dos estados do jogo em cada instante de tempo disponibilizado para cada agente. Pela definição de processos de decisão de Markov dada em [6], pode-se analisar que o *RoboCup Simulation 2D* trata-se de um Processo de Decisão de Markov Parcialmente Observável (POMDP, do inglês: *Partial Obervable Markov Decision Process*). Isso por que é dado aos agente apenas uma parte do estado atual da partida, omitindo, por exemplo, a posição da bola caso esta não esteja na linha de visão do jogador.



A partida também é realizada através de processo estocástico. Isso por que o servidor pode alterar de maneira aleatória variáveis ligadas tanto ao estado de jogo quanto á propriedades do próprio agente. Por exemplo, quando um jogador calcula através do comportamento programado para ele que ele precisa chutar a bola com uma força arbitrária  $F$  e na direção  $\vec{v}$ , o servidor pode incluir pequenas perturbações de forma que a execução da ação não irá ser determinística.

### Métodos Analisados

Nesta seção serão mostradas as análises feitas de alguns algoritmos de aprendizado por reforço para aplicação no futebol simulado. Isto foi feito através de implementações em problemas de natureza mais simples para que as vantagens e desvantagens de cada algoritmo pudesse ser entendida.

O objetivo da análise é definir as características de cada uma das técnicas e possíveis alterações para melhora de performance que podem ser feitas. Isso permite uma compreensão mais profunda para futura implementação da técnica escolhida como algoritmo de aprendizado para um time da categoria.

### Neural Evolution of Augmenting Topologies

Técnica inicialmente proposta por Stanley em [7], esta técnica se baseia na criação de redes neurais que tenham suas topologias evoluídas através de gerações e premiação que cada um dos genes da população recebe. As redes neurais usam como operação de mutação a adição de nós e ligações entre as camadas, e também podem realizar a operação de *crossover* entre si. Um exemplo desse tipo de rede pode ser visto na figura 2, onde a camada de entrada da rede neural possui quatro entradas e a camada de saída possui somente um nó.

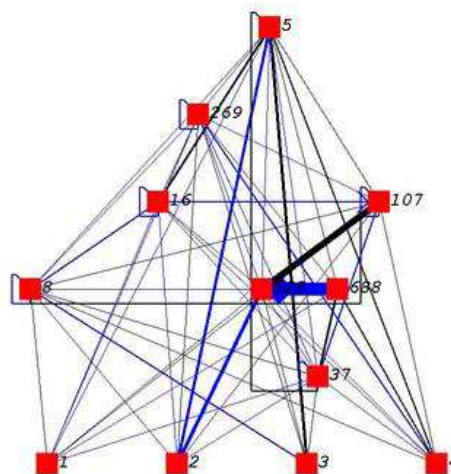


Figura 2 - Exemplo de rede feita pelo algoritmo NEAT



Para aprofundamento no estudo da técnica, foi criado um ambiente de simulação onde um carro autônomo possui seis sensores, cada um detectando a distância para o obstáculo, que neste caso são as laterais da pista. O cálculo de premiação do algoritmo se dá segundo o avanço do carro na direção da linha de chegada, e o valor de punição é dado segundo os sinais captados pelos sensores.

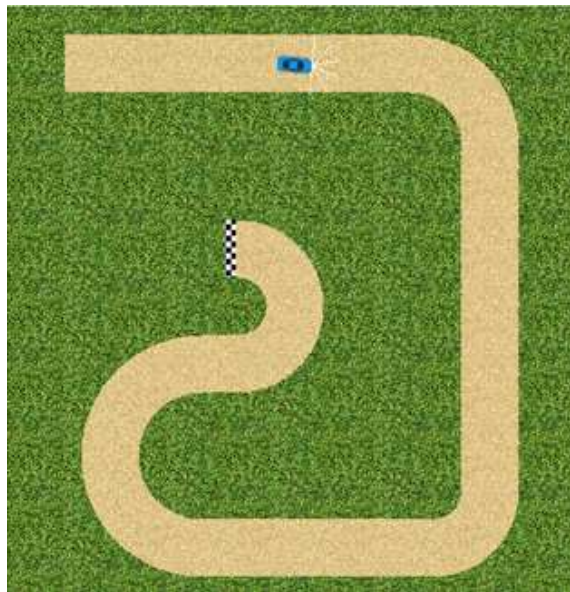


Figura 3 - Carro autônomo criado através do algoritmo NEAT

Uma grande vantagem que o NEAT apresenta é o fato de que ele apresenta resultados satisfatórios de maneira rápida, com menor custo computacional do que outros algoritmos. Para o problema descrito, para a maior parte das simulações foi necessária apenas uma geração para que ele apresentasse uma solução.

O problema com esse tipo de técnica é que, para que a evolução seja alcançada, é necessário que sejam realizados testes entre os genes da população para que se possa atribuir as premiações de cada um. Isso dificulta a implementação deste na categoria de simulação 2D de futebol, pois o processo representado pelo MDP possui função de premiação dinâmica por conta dos diferentes times e diferentes situações de jogo, o que requer aprendizado *online*. Entretanto, o caso analisado tem premiação estática e processo de evolução relativamente lento, que dura aproximadamente de um segundo, o que impede a sua aplicação *online*.



### **Value Iteration**

O método conhecido como *Value Iteration* [6] é implementado através do cálculo do valor de cada um dos estados presentes no problema em questão. Esta é uma técnica conhecida como baseada em um modelo, o que quer dizer que para seu funcionamento e convergência para os valores finais é necessário que seja apresentado um modelo de probabilidades da tomada de cada ação em cada um dos estados. Sua equação característica pode ser vista em (1).

$$v_{k+1} = \max_a \sum_{s'} p(s'|s, a) [r(s, a, s') + \gamma v_k(s')] \quad (1)$$

Onde  $v_{k+1}$  é o próximo valor do estado,  $v_k$  é o valor do estado naquele instante,  $p(s'|s, a)$  é a probabilidade em que, tomada a ação  $a$  no estado  $s$ , o agente vá para o estado  $s'$ . A constante  $\gamma$  representa o fator de desconto e  $r(s, a, s')$  é a premiação recebida pela ação tomada.

O algoritmo foi analisado através da implementação de um labirinto onde o agente deveria chegar na bola, que pode ser visto através da figura 4. O modelo usado foi que para a ação que o algoritmo selecionasse para o agente tem 80% de chance de ser escolhida, sendo os outros 20% dividido entre os demais. A premiação por chegar até a bola é de 10, enquanto as laterais do campo tem punição de -5. Os obstáculos não influenciam nos cálculos.

A dificuldade com esse método se apresenta por conta da necessidade de apresentar o modelo de probabilidades, o que é muito difícil de ser feito em um problema como o de futebol simulado. Além disso, o custo computacional aumenta exponencialmente de acordo com o número de estados do jogo, visto que o algoritmo precisa mostrar a convergência dos valores para cada um deles.

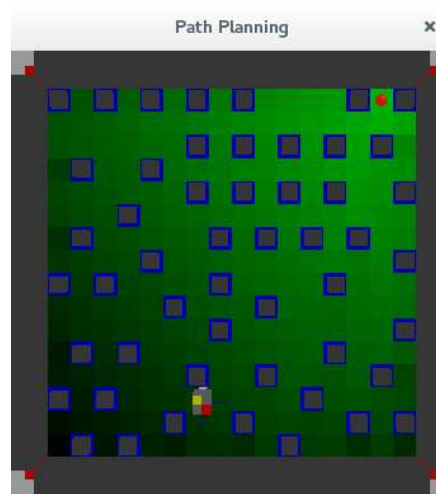


Figura 4 - Representação do

### ***Q-Learning***

Esta é uma técnica conceitualmente parecida com a de *Value Iteration*. A diferença está no fato de que se utiliza da técnica de Monte Carlo juntamente com aprendizado por diferenciação temporal [6] para realizar seus cálculos. Isso permite que ela esteja apta a calcular os estados de forma *online*, o que é uma vantagem diante dos dois primeiros algoritmos apresentados. Sua fórmula característica é mostrada em (2).

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha [R_{t+1} + \gamma \max_a Q(s_{t+1}, a_t) - Q(s_t, a_t)] \quad (2)$$

Onde  $Q(s_t, a_t)$  representa o valor da tomada de determinada ação  $a_t$  em determinado estado  $s_t$ ,  $\alpha$  é uma constante que deve ser decrescente para a garantia da conversão da equação.  $R_{t+1}$  é a premiação recebida pela tomada da ação e  $Q(s_{t+1}, a_t)$  é o valor individual de cada uma das ações no estado seguinte. Por fim,  $\gamma$  é o fator de desconto.

O problema foi implementado através de um exemplo muito conhecido, chamado *Crawler*. Consiste em um agente que possui um braço com duas articulações. O agente possui premiação positiva para toda vez que sua ação causa um avanço para ele, e caso contrário recebe uma punição de valor negativo.



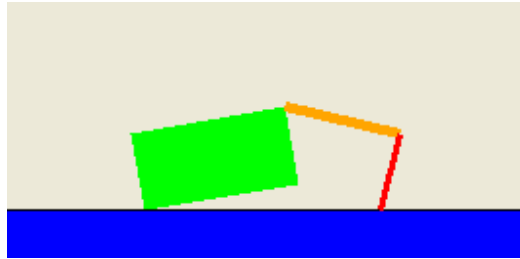


Figura 5 - Representação de um agente *Crawler*.

Ainda que apresente a possibilidade de aprendizado *online*, o algoritmo de *Q-Learning* não tem uma conversão rápida e ainda possui um custo computacional muito grande, devido a necessidade de se atravessar toda a fase de exploração até que os valores estejam aproximados dos valores de conversão final.

### Deep Q-Learning

Este algoritmo foi pesquisado por finalidade puramente científica, pois sua conversão é notoriamente lenta [1]. Foi aplicado no mesmo ambiente em que foi aplicado o algoritmo de *Value Iteration*, com a diferença de que a entrada é feita com os valores dos próprios pixels. Ele se utiliza da equação mostrada em (2) para convergência e utiliza uma técnica chamada de *experience replay*, que garante sua convergência mais abrangente e não para um ótimo local. Para manipular e fazer os cálculos é utilizada também o que se chama de rede neural convolucional. Uma demonstração do algoritmo pode ser vista através da figura 6.

Apesar de sua conversão lenta ser conhecida, ainda sim se justifica seu estudo por conta de uma tendência da área de aprendizado de máquina de utilizar algumas características usadas no algoritmo para realizar o processo de aprendizagem, como pode ser visto em [1], onde redes neurais convolucionais são utilizadas em conjunto com a técnica de *Monte Carlo Search Tree*.

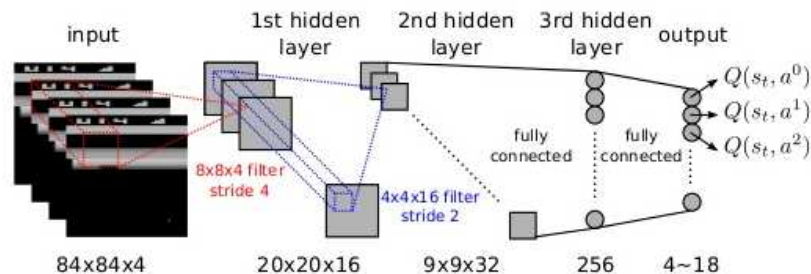


Figura 6 - Representação de jogo de Atari para processamento da DQN.



### **Considerações finais e trabalhos futuros**

Os algoritmos vistos ao longo deste trabalho são estudados amplamente pela comunidade de aprendizado de máquina, seja através de uso direto ou indireto através de conceitos dados por cada um deles. Os problemas onde foram aplicados, apesar de serem simplificados com o objetivo de facilitar a implementação, ainda servem como parâmetro de comparação e avaliação de suas performances em ambientes mais complexos, como o próprio simulador de futebol 2D.

Apesar de o trabalho de pesquisa até agora não ter encontrado em nenhum dos algoritmos aplicados características que os levem a serem implementados no *Robocup Soccer Simulation 2D*, o conhecimento adquirido será útil para que trabalhos futuros possam ser implementados.

Técnicas que envolvem a concepção de árvores hierárquicas é uma técnica que não foi implementada, mas que lida com alguns dos problemas encontrados pelos algoritmos citados, como tempo de convergência lento. O MAXQ-OP [4] é um exemplo de que, através da simplificação é possível se atingir resultados bastante aceitáveis.

É, portanto, planejado um estudo mais aprofundado nesse tipo de técnica, juntamente com a incorporação também do que é conhecido como *Transfer Learning* [8]. A técnica pode ser utilizada para que cada nó da árvore hierárquica aprenda de uma maneira mais rápida através do aproveitamento do que já foi aprendido. Isso facilita a conversão do algoritmo, pois evita o ato de recalcular situações que possam ser semelhantes.

### **Bibliografia**

- [1] D. Silver e A. Huang. *Mastering the Game of Go with Deep Neural Networks and Tree Search*, Google DeepMind. 2016.
- [2] D. Silver, V. Mnih, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller. *Playing Atari with Deep Reinforcement Learning*, DeepMind Technologies. 2013.
- [3] P. Stone e Richard S. Sutton. *Scaling Reinforcement Learning toward RoboCup Soccer*, AT&T Shannon Laboratory. 2001.
- [4] A. Bai, F. Wu e X. Chen. *Online Planning for Large MDPs with MAXQ Decomposition*, School of Computer Science Univ. of Sci. & Tech. of China. 2012.
- [5] P. Stone. *Layered Learning in Multi-Agent Systems*, School of Computer Science, Carnegie Mellon University, 1998.



- [6] R. S. Sutton e A. G. Barto. *Reinforcement Learning: An Introduction*, Cambridge, MIT Press. 1998.
- [7] K. O. Stanley e R. Miikkulainen. *Evolving Neural Networks through Augmenting Topologies*. Department of Computer Sciences, The University of Texas at Austin. 2001.
- [8] A. Barreto, R. Munos, T. Schaul e D. Silver. *Successor Features for Transfer in Reinforcement Learning*, Google DeepMind, Londres. 2016.



## **Bamba Analytics : Uma Estrutura para Exploração de Dados Brasileiro para a Biodiversidade Marinha**

Matheus Machado da Rosa Albuquerque, Luiz Manoel Rocha Gadelha Júnior e Kary Ann del Carmen Ocaña Gautherot

### **Resumo**

O presente trabalho visa o desenvolvimento de um ambiente gráfico para análise e manipulação de dados via web da Biodiversidade Marinha Brasileira, através do projeto BiotecMar (Meirelles, P. M. et al., 2015). O projeto aborda a problemática na análise e gerenciamento de dados coletados pelo projeto BiotecMar. Constatamos a dificuldade de pesquisadores de se obter dados consistentes e completos além de compatíveis com os softwares usados por eles respectivamente. As etapas de desenvolvimento da aplicação envolvem desde a concepção de um banco de dados que permita armazenar tais dados, até a implementação de rotinas que garantam a qualidade dos mesmos. Temos a proposta de desenvolver uma interface totalmente online, onde o usuário poderá com facilidade minerar os dados e gerar estatísticas com eficiência. A interface está sendo desenvolvida na linguagem estatística R. A escolha é justificada pois o R torna a implementação da interface robusta e conseqüentemente, permite a inclusão de scripts específicos em R para manipulação dos dados. Há a intenção de se desenvolver um ambiente onde o pesquisador poderá fazer o upload de seus próprios scripts e executá-los. O banco de dados foi implementado em PostgreSQL e faz utilização de alguns frameworks que são responsáveis por assumir o controle de diversos procedimentos. Através desta aplicação web, uma quantidade massiva de dados contidos em planilhas, são importados e armazenados em um banco de dados relacional. A partir desta importação e armazenamento será possível efetuar buscas pontuais, demonstrar estatísticas e obter uma visualização geográfica através de mapas. O sistema está integrado ao repositório de dados GBIF (Edwards, J. L., 2000), através de uma API via serviço RESTful onde é possível importar informações de datasets de diversas categorias de dados e seus metadados que seguem padrões como o Darwin Core (Wieczorek, John et al., 2012), EML (Fegraus, E. et al., 2005). A aplicação será instalada no servidor Buriti do LNCC. Assim permite eliminar a etapa de baixar os dados e executá-los localmente. Passo esse que exigia hardware local do usuário para que a execução dos scripts não fosse tão lenta. Por conseguinte, vamos possibilitar um grande poder de processamento aos mesmos sem a necessidade de requisitos locais. Estimamos que a aplicação tornaria a execução de pesquisas e consultas de dados biológicos, simples e eficiente.

### **Referências**

- Meirelles, P. M. et al., 2015. Biotecmar.  
Edwards, J. L., 2000. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. Science.  
Wieczorek, John et al., 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7, no. 1.



Fegraus, E. et al., 2005. Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. Bull. of the Ecological Society of America 86.



## **Aplicações e técnicas de aprendizagem de máquinas à análise de dados de domínio e de tempo de execução de workflows de bioinformática**

Matheus Tonelli de Souza , Luiz M. R. Gadelha Jr. e Kary Ocaña

Um workflow se refere a execução de programas encadeados que consomem e produzem grande volumes de dados, o que acaba se tornando uma tarefa com alto custo computacional, conseqüentemente consome um tempo considerável. A partir deste problema buscam-se maneiras de otimizar o algoritmo. Um experimento que produz excelentes resultados é a aplicação de aprendizagem de máquinas para reconhecer padrões estruturados nos bancos de dados, gerando métodos de classificações eficientes.

Constatado o problema da grande demanda de tempo na execução do workflow, deve se encontrar maneiras de visualizar onde existe a tarefa consumidora de tempo. O método utilizado com maior frequência é a utilização de perfiladores<sup>5</sup> para localizar a tarefa do workflow que mais requer recursos, assim como predições de sua execução.

Para o desenvolvimento deste trabalho, foi utilizada a linguagem R que é amplamente utilizada para análise de dados e possui uma vasta gama de bibliotecas. Dentro do R, existe um sistema de desenvolvimento chamado Shiny, que se trata de uma aplicação Web que permite a interação do usuário com o *HPSW-Prof*<sup>6</sup>, permitindo selecionar qual workflow deseja analisar, assim como pode selecionar quais informações são de maior relevância para atender a necessidade do usuário.

Neste trabalho foi desenvolvido um perfilador *HPSW-Prof* que visa auxiliar o usuário com um conjunto de ferramentas para tratamento estatístico e manipulação de informações de proveniência que deverão ser gerados a partir da execução de algum experimento científico. Através desta ferramenta bem vantajosa, o usuário torna-se capaz de visualizar de maneira simples, o funcionamento do workflow e otimizar seu funcionamento, uma vez que se torna mais transparente o seu funcionamento durante cada etapa.

---

5 Perfiladores: Aplicação capaz de gerar um perfil sobre o workflow.  
6 Disponível em: <https://github.com/mmondelli/swift-phylo>



## Referências

Dun, N., Taura, K., and Yonezawa, A. (2010). ParaTrac: a fine-grained profiler for data-intensive workflows. In *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing – HPDC'10*, page 37, NY, USA.

Juve, G., Chervenak, A., Deelman, E., Bharathi, S., Mehta, G., & Vahi, K. (2013). Characterizing and profiling scientific workflows. *Future Generation Computer Systems*, 29(3), 682–692.

Freire, J., Koop, D., Santos, E., and Silva, C. (2008). *Provenance for Computational Tasks: A Survey*. *Computing in Science & Engineering*, 10(3):11-21.

Gadelha, L. M. R., Wilde, M., Mattoso, M., and Foster, I. (2012) MTCProv: a practical provenance query framework for many-task scientific computing. *Distributed and Parallel Databases*, 30(5-6):351-370.



## **Implementação da estrutura de dados que armazena a solução da equação de convecção-difusão no método HOPMOC**

Thiago Daniel Quimas Simões Teixeira, Carla Osthoff e Frederico Luís Cabral

O método HOPMOC foi proposto como um método para resolução de problemas de convecção e difusão com dominância convectiva para ser executado em ambientes compostos por máquinas paralelas. Ele permite uma decomposição espacial do domínio do problema, onde cada subproblema pode ser resolvido independentemente.

De forma a otimizar a execução do método em uma arquitetura Xeon da Intel, foram introduzidas diversas alterações na implementação da estrutura de dados que armazena a solução de convecção e difusão em uma dimensão espacial do HOPMOC, sendo executado em processadores Intel (arquitetura Haswell/Broadwell), que oferecem suporte à vetorização, que por sua vez, permite que se aproveite com maior eficiência os recursos da microarquitetura, principalmente na execução fora de ordem (Out Of Order execution) do pipeline.

A estratégia consiste basicamente em separar os pontos da discretização em diferenças finitas do algoritmo em duas partes, que no algoritmo HOPMOC original estão entrelaçados, que correspondem aos pontos que são computados pelos operadores explícitos e implícitos.

A vantagem principal em se desenvolver estratégias que facilitam ao compilador gerar um código com vetorização mais eficiente de forma automática, se torna mais evidente quando se compara o ganho obtido através do paralelismo multithread (OpenMP) com e sem a vetorização. Neste trabalho demonstramos que o código otimizado apresenta um ganho de 2,4 vezes em relação ao original.





## **MODELAGEM MOLECULAR DE INIBIDORES POTENTES E SELETIVOS PARA A ENZIMA ACETILCOLINESTERASE**

Ana Luiza Martins Karl, Laurent Emmanuel Dardenne e Isabella Alvim Guedes

**Introdução.** A acetilcolinesterase (AChE) é uma enzima responsável pela regulação da transmissão dos impulsos nervosos através da hidrólise rápida de moléculas de acetilcolina (ACh)<sup>[1]</sup> nas fendas sinápticas. Deficiências no funcionamento desta enzima estão associadas a uma série de quadros patológicos, dentre eles a Doença de Alzheimer (DA). Em condições normais e, de forma mais importante quando a atividade da AChE diminui, a enzima Butirilcolinesterase (BChE) participa na modulação dos níveis de ACh nos neurônios colinérgicos e, por esse motivo, também é um importante alvo farmacológico da DA<sup>[2]</sup>. O objetivo deste estudo é guiar a etapa de otimização racional de inibidores potentes e seletivos (sintetizados por pesquisadores do Instituto de Química da UFRGS) para a AChE, com atividade multialvo, através da realização de estudos de *ensemble docking* em diferentes conformações das enzimas AChE e BChE, considerando implicitamente a flexibilidade da proteína.

**Metodologia.** Atualmente existem diversas estruturas tridimensionais da AChE disponíveis no banco de estruturas *Protein Data Bank* (PDB). As estruturas PDB de código 1q83, 1zgc e 2ckm de AChE, e 4bds e 4tpk de BChE, foram selecionadas para serem utilizadas nos estudos de *ensemble docking* seguindo critérios pré-estabelecidos. Após a seleção, um alinhamento estrutural foi realizado utilizando a ferramenta *super* do programa Pymol para investigar a presença de moléculas de águas altamente conservadas e a variabilidade conformacional de resíduos de aminoácidos na cavidade de ligação da ACh. Os complexos selecionados foram preparados usando a ferramenta Prepwizard<sup>[3]</sup> definindo pH = 7 e respeitando os estados de protonação já descritos na literatura. Os ligantes estudados neste trabalho foram preparados utilizando a ferramenta LigPrep/Epik<sup>[4]</sup>. Os estudos de *redocking*, realizados para validar a metodologia e protocolo de preparação, e *ensemble docking*, com os inibidores testados nesse trabalho, foram feitos utilizando os programas de docking DockThor<sup>[5]</sup> e Glide<sup>[6]</sup>.

**Resultados.** O alinhamento estrutural mostrou diferenças significativas da AChE entre as três estruturas, especialmente o resíduo de aminoácido Trp279. A conformação deste resíduo parece estar diretamente relacionada com o modo de ligação de diferentes inibidores com o sítio PAS da enzima. Os estudos de *ensemble docking* com os inibidores sintetizados pelo grupo da UFRGS utilizando os complexos da enzima AChE (1q83, 2ckm e 1zgc) tiveram resultados satisfatórios. De forma geral, o programa de *docking* Glide conseguiu prever o modo de ligação descrito na literatura para o grupo análogo à tacrina: interações do tipo pi-stacking com os resíduos Trp84 e Phe330, e ligação hidrogênio entre o NH da tacrina e o oxigênio carbonílico da His440. Os inibidores mais potentes testados apresentaram manutenção de ligações pi-stacking com os resíduos Trp279 e Tyr70 no PAS e T-stacking com resíduos Tyr72 e Tyr341, além de ligações de hidrogênio com o resíduo Tyr121. Os resultados de *docking* nas estruturas de BChE não foram satisfatórios, uma vez que o sítio de ligação desta enzima está adaptado para interagir com ligantes pequenos, não havendo espaço suficiente para acomodar inibidores volumosos e flexíveis como os estudados neste trabalho.



**Conclusões.** A partir dos resultados obtidos nesse trabalho, é possível concluir que a seleção de três estruturas representativas da AChE possibilitou considerar parcialmente a flexibilidade dessa enzima, dado que o Trp279 apresenta diferentes orientações nas três estruturas representativas selecionadas. Além disso, a metodologia de *ensemble docking* utilizada permitiu a observação de interações importantes dos inibidores testados, que auxiliaram no entendimento da seletividade e potência das moléculas estudadas<sup>[7]</sup>. A previsão do modo de ligação desses inibidores constitui uma importante informação para futuros estudos visando a otimização desses compostos.

**Referências:**

- [1] J. L. Sussman, M. Harel, F. E. Frolow, et. al., Science vol 253, 1991 page 872 – page 879.
- [2] Nachon, F., Carletti, E., et. al., Biochem. J., Vol 453, 2013 page 393 – 399.
- [3] Schrödinger Release 2013-3: Schrödinger Suite 2013-3 Protein Preparation Wizard, Schrödinger, LLC, New York, NY, 2013.
- [4] LigPrep, version 2.9, Schrödinger, LLC, New York, NY, 2014.
- [5] C. S. Magalhães, D. M. Almeida, H. J. C. Barbosa and L. E. Dardenne, Information Sciences vol 289, 2014 page 206- page 224.
- [6] Small-Molecule Drug Discovery Suite 2013-3 Glide, version 6.1, Schrödinger, LLC, New York, NY, 2013.
- [7] Ceschi, M. A. et al., European Journal of Medicinal Chemistry, Vol 121, 2016 page 758-772.



## **Estudo de estratégias de vacinação de um modelo epidemiológico**

Prates, D.B., Kritz, M. V. e SILVA, J. M.,

### **RESUMO**

As epidemias são historicamente responsáveis por inúmeras mortes no mundo, o que as tornam um assunto de saúde pública. No estudo de epidemiologia há um grande número de modelos e técnicas matemáticas que possibilitam a simulação de fenômenos epidemiológicos. Esses modelos são baseados em equações diferenciais que possibilitam a análises e estudos das características endêmicas, facilitando a compreensão do comportamento dinâmico de epidemias em determinadas populações.

O modelo SIR, proposto por Kermack e McKendrick em 1927, foi o pioneiro na modelagem epidemiologia e propõe dividir a população em três classes: suscetíveis, infectados e removidos. Com a inserção de indivíduos infectados na população, a epidemia pode ser analisada apresentando resultados conclusivos. O modelo é constantemente modificado para simular as características intrínsecas de determinada doenças, como por exemplo, quando admite a inserção de vacinações, de efeitos de sazonalidade e de novos parâmetros e classes ao modelo [1].

Estudos detalhados sobre dois métodos de vacinação são apresentados em [3]: um método de vacinação constante e outro de vacinação em pulsos, ambos aplicados, separadamente, a apenas uma parcela da população. Neste trabalho, ambas as estratégias são alternadamente aplicadas e algumas discussões sobre os resultados obtidos são realizadas.

### **Referências**

- [1] Jardim C. L. T. F., Prates D. B., Silva J. M., Ferreira L. A. F. and Kritz M. V. 2015 Simulations of a epidemic model with parameters variation analysis for the dengue fever Journal of Physics: Conference Series 633 012008 In IOPscience.
- [2] SABETI, Mehran. Modelo Epidêmico Discreto SIR com estrutura etária e aplicação de vacinação em pulsos e constante. UFPE, Phd Thesis, 2011.
- [3] PRATES, D. B. et al. An epidemiological model with vaccination strategies. In: ICNAAM 2015. AIP Publishing, 2016. p. 480035.
- [4] PRATES, D. B. et al. Vaccination Strategies: a comparative study in an epidemic scenario. In: Journal of Physics: Conference Series. IOP Publishing, 2016. p. 012083.



## **TOMOGRAFIA POR IMPEDÂNCIA ELÉTRICA**

D. TOLEDO DA SILVA, A.A. NOVOTNY e A.D. FERREIRA

Neste trabalho apresentamos o problema do EIT e um método não iterativo para resolvê-lo. Também discute-se a possibilidade de resolver esse problema mesmo que não seja conhecida a condutividade elétrica do corpo e das anomalias procuradas. Por último, apresentamos um estudo sobre as diferentes configurações de eletrodos que são usados para ler o potencial elétrico e injetar ou drenar corrente elétrica do corpo estudado.

### Referências

- D.C. Baber and B.H. Brown. Applied potential tomography. J. Phys. E. Sci. Instrum., 17:723733, 1984.
- A. D. Ferreira and A. A. Novotny. A new non-iterative reconstruction method for the electrical impedance tomography problem. Submetido para publicação.
- A. A. Novotny and J. Sokoowski. Topological derivatives in shape optimization. Interaction of Mechanics and Mathematics. Springer, 2013.
- G. Uhlmann. Electrical impedance tomography and Calderon's problem. Inverse Problems, 25(12):123011, 2009.



## PARALELISMO E TOLERÂNCIA A FALHAS NO AMBIENTE SPiNMe

Emiliano Medeiros de Oliveira Neto, Antônio Tadeu Azevedo Gomes

### Resumo

Neste projeto busca-se a compreensão e aperfeiçoamento de funcionalidades do simulador numérico MHM (Multiscale Hybrid Mixed) desenvolvido nas linguagens Erlang e C++. O objetivo é propor e implementar melhorias de desempenho, suporte a tolerância a falhas, maior escalabilidade e flexibilidade, e melhor aproveitamento possível da infraestrutura disponível para a execução desse simulador.

Durante o primeiro ano deste projeto, as atividades concentraram-se nos módulos desenvolvidos em Erlang, que têm como foco o tratamento da distribuição dos processos C++ e o suporte a tolerância a falhas nesses processos. Das contribuições para aumento do desempenho na execução do simulador, destacam-se a análise e implementação realizadas para a redução dos custos computacionais, diminuindo a frequência com que novos processos C++ são criados e terminados em cada nó processador na infraestrutura. Com a substituição do antigo método de criação e terminação dos processos C++ pela troca de mensagens entre esses processos e os processos correspondentes em Erlang, observou-se uma redução significativa no tempo de execução para um cenário com processos C++ de baixo custo computacional, em ambiente de execução local. Além disso, duas funcionalidades foram incluídas no código do simulador. A primeira delas diz respeito ao suporte no simulador à resolução em paralelo de problemas transientes. A segunda funcionalidade foi a flexibilização do simulador para funcionamento em cloud. Nesta segunda funcionalidade, ainda em desenvolvimento, busca-se implementar módulos Erlang que permitam a execução dos processos C++ tanto em ambiente de cluster como de cloud sem modificação do código C++ correspondente, permitindo a evolução desacoplada e simultânea de ambos os códigos C++ e Erlang.



## **Modelagem Numérica de Análise Estática em Estruturas Flexíveis**

Lucas M. R. Silva<sup>1</sup>, Cristiane O. Faria<sup>1</sup>, Sandra M. C. Malta<sup>2</sup>

<sup>1</sup>Universidade do Estado do Rio de Janeiro, Departamento de Análise Matemática – IME,

<sup>2</sup>Laboratório Nacional de Computação Científica,

### **RESUMO**

A Mecânica dos Meios Contínuos, e mais especificamente a Teoria da Elasticidade, tem como preocupação básica o desenvolvimento de modelos matemáticos que possam representar adequadamente a situação física real de componentes industriais sujeitos a esforços mecânicos. Em análise estrutural, o objetivo pode ser a determinação do campo de deslocamentos, as deformações internas ou as tensões atuantes no sistema devido à aplicação de cargas, além de outros. Porém, a aplicação de tais teorias a casos práticos apresenta dificuldades às vezes intransponíveis. Por exemplo, na análise estrutural, a perfeita representação matemática de carregamentos, geometria, condições de contorno, comportamento dos materiais etc, em muitas situações, apresenta-se de forma complexa, havendo, assim, a necessidade de se introduzir muitas hipóteses simplificativas no problema real, para permitir alguma forma de modelagem matemática que conduza a soluções mais simples.

Por outro lado, engenheiros têm demonstrado um interesse crescente por estudos mais precisos para a análise de estruturas. Este interesse vem unido a uma necessidade cada vez maior de se estudar o comportamento de elementos estruturais complexos, o que conduz a tratamentos analíticos mais elaborados, baseados em teorias gerais, e que são, via de regra, de soluções extremamente difíceis.

Mas nem sempre é possível encontrar soluções analíticas dessas equações diferenciais, desse modo, devido a sua simplicidade, nesse projeto utilizamos o Método das Diferenças Finitas (MDF) partindo de uma abordagem pedagógica. No desenvolvimento do projeto empregamos ferramentas matemáticas utilizadas no MDF e sua aplicação prática na elasticidade e no problema de vigas.

O modelo matemático estudado neste projeto foi o de Euler-Bernoulli. Este modelo considera que o cisalhamento e a inércia de rotação são desprezados, ou seja, supõe-se que as seções transversais planas permanecem planas e perpendiculares ao eixo longitudinal da viga mesmo após a deflexão. Na teoria de vigas, utiliza-se a hipótese de pequenas deformações, por isso as variações de geometria podem ser desconsideradas.



Para a implementação desse método, consideramos três tipos comuns de vigas – biapoiada, biengastada e engastada-apoiada – sujeitos à ação de um carregamento uniformemente distribuído no espaço. Analisamos, para cada tipo de viga, a consistência das aproximações para a equação da elasticidade por Diferenças Finitas e, após a implementação computacional utilizando a linguagem C, os resultados das soluções analíticas e das soluções aproximadas foram comparadas numericamente.

## **REFERÊNCIAS**

SOARES, Alexandre A. B.; Universidade da Amazônia. *O Método das Diferenças Finitas aplicado à Teoria das Vigas*, 2010. Monografia (Graduação).

de MELLO, Eduardo Henrique V. M., Universidade Federal do Paraná, *Análise Dinâmica de Vigas de Euler-Bernoulli e de Timoshenko com o Método de Diferenças Finitas*, 2014. Monografia (Graduação).



## **MHOLline 2.0 - Workflow para problemas científicos em Bioinformática e Biologia Estrutural**

V.C.C. Reis, L.E. Dardenne, P.V.S.Z. Capriles, A.D.Rossi e P.M. Bisch

### **RESUMO**

Genomas e suas proteínas podem ser analisados de diferentes perspectivas e com diversos objetivos, sendo que os resultados são aplicáveis a várias áreas de estudo. Em pesquisas em bioinformática e biologia computacional normalmente aplicam-se múltiplas combinações de programas para se chegar ao resultado esperado. Escolher o programa correto, setar seus parâmetros, filtrar e tratar os arquivos de entrada, extrair informação dos arquivos de saída de cada programa e ainda automatizar estas tarefas pode ser muito desafiador e nada trivial.

O MHOLline é um workflow científico que proporciona um conjunto de módulos que auxiliam na análise global e em larga escala de proteínas, facilitando portanto boa parte do trabalho de seu usuário. A versão 1.0 foi lançada em 2010 através de uma colaboração entre a Universidade Federal do Rio de Janeiro (UFRJ) e o Laboratório Nacional de Computação Científica (LNCC) e se encontra no site [www.mholline.lncc.br](http://www.mholline.lncc.br) [Capriles & Guimarães et al. 2010, Guimarães & Capriles et al. 2013]. O sistema está sendo completamente reformulado e expandido na versão 2.0, que já encontra-se em Beta Test em [www.mholline2.lncc.br](http://www.mholline2.lncc.br).

Nossos objetivos para a versão 2.0 já alcançados: (i) atualização dos módulos antigos para suas versões mais novas; (ii) adição de novos softwares para fazer o MHOLline cada vez mais completo e abrangente; (iii) refatoração, otimização e padronização do código para que ele se torne mais legível e rápido; (iv) resolução de problemas de segurança; (v) criação de uma página de visualização de resultados em detalhes; (vi) Criação de um sistema de refinamento de resultados em que o usuário pode reconstruir um alinhamento individual e resubmetê-lo usando algumas restrições na construção do modelo tridimensional da proteína; (vii) desenvolvimento de uma nova interface compatível com os principais browsers do mercado e que utiliza dos conceitos de Interação Humano Computador para proporcionar uma melhor experiência de uso ao usuário.

Os softwares já adicionados são: (i) MHOLPROBITY [Chen et al. 2010] para validar as propriedades estereoquímicas dos modelos 3D gerados; (ii) SIGNALP [Petersen et al. 2011] para identificar regiões de peptídeo sinal em sequência de proteínas; (iii) TMHMM [Krogh et al. 2001] para identificação de hélices de transmembranas; e (iv) PSIPRED [McGuffin et al. 2000] para predizer a estrutura secundária de proteínas.

Até o final do ano a área administrativa do sistema estará completamente reformulada e contará com novas funcionalidades para facilitar o controle e a manutenção do sistema pelos administradores e mantenedores. Paralelamente estamos coletando o feedback de usuários da versão beta para corrigir





possíveis bugs e ou fazer pequenas melhorias. Após a finalização da interface administrativa, de possíveis erros e de melhorias, lançaremos a versão 2.0.

## **REFERÊNCIAS**

Capriles, P.V.S.Z.; Guimarães, A.C.R; Otto, T.D; Miranda, A.B.; Dardenne, L.E.; Degrave, W.M. Structural modelling and comparative analysis of homologous, analogous and specific proteins from *Trypanosoma cruzi* versus *Homo sapiens*: putative drug targets for chagas disease treatment. *BMC Genomics*, 11:610, 2010.

Chen, V.B.; Arendall III, W.B.; Headd, J.J., Keedy, D.A.; Robert M. Immormino, R.M.; Kapral, G.J.; Murray, L.W.; Richardson, J.S.; Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica D66*: 12-21, 2010.

Krogh, A.; Larsson, B.; Heijne, G.; Sonnhammer, E.L.L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567-580, 2001.

Petersen, T.N.; Brunak, S.; Heijne, G.; Nielsen, H. SignalP 4.0: discriminating signal peptides from transmembrane regions, *Nature Methods*, 8:785-786, 2011.

McGuffin, L.J.; Bryson, K; Jones, D.T. The Psipred protein structure prediction server. *Bioinformatics*, 4:401-405, 2000.



## Otimização computacional do esquema de radiação na atmosfera em modelo numérico de previsão de tempo e clima

Mateus Silva de Melo, Roberto Pinto Souto

O modelo numérico de previsão de tempo e clima BRAMS (*Brazilian developments on the Regional Atmospheric Modeling System*) é um modelo em escala regional desenvolvido no INPE/CPTEC, com base no RAMS (*Regional Atmospheric Modeling System*), onde foram introduzidas diversas alterações no código-fonte que resultaram em uma descrição mais realista dos processos atmosféricos tropicais e em uma eficiente paralelização do modelo, através da decomposição de domínio, usando MPI. Neste trabalho foi utilizado a versão 5.2 do BRAMS, no qual foi realizado um perfil de desempenho, descobrindo que a rotina “*rrtm\_driver*”, responsável por ser driver da rotina que calcula a radiação na atmosfera, é a mais custosa à aplicação. Portanto, foram implementadas diretivas *OpenMP* em um laço de repetição existente nessa rotina, apresentado razoável escalabilidade. No entanto, posteriormente foi verificada uma ineficiência em um trecho deste laço que, ao ser corrigida, resultou em código cerca de três ordens de grandeza mais rápido. Como consequência, o *driver* da radiação deixou de ser um dos gargalos de desempenho da aplicação.



## **Análise de Preditores de Domínios Proteicos**

Matheus Miloski, Fábio Lima Custódio

Um domínio proteico representa uma parte da cadeia polipeptídica que pode enovelar-se de maneira independente, formando uma estrutura compacta e estável. A identificação de domínios é uma etapa crucial para o sucesso de métodos de predição de estrutura de proteínas, principalmente em abordagens *free-modeling*. Ao fragmentar o alvo em domínios e fazer a predição destes módulos menores, o ataque à proteínas multidomínios é aprimorado. Somente após a obtenção das estruturas para cada domínio que o modelo para a proteína completa é construído. Essa estratégia vem sendo aplicada em metodologias de sucesso na área, como a do Rosetta, e resultados superiores àquelas na qual a proteína é predita sem a separação de domínios vem sendo obtidos. Uma variedade de metodologias vem sendo apresentada nos últimos anos para a predição de domínios, mas o problema ainda representa um grande desafio na área. O alinhamento múltiplo é o princípio mais aplicado nas estratégias de predição de domínios.

O presente estudo busca avaliar o desempenho de preditores de domínios proteicos. Oito preditores foram selecionados: DomCut (Suyama & Ohara, 2003), Scooby (George et al., 2005), InterProScan (Quevillon et al., 2005), Cdvist (Adebali et al., 2014), DoBo (Eickholt et al., 2011), Drop (Ebina et al., 2011), H-drop (Ebina et al., 2014) e ThreaDom (Xue et al., 2013). Como conjunto teste, foram utilizados 20 alvos obtidos das últimas três edições do CASP (*Critical Assessment of protein Structure Prediction*) que apresentavam mais de um domínio em sua estrutura tridimensional. A análise da acurácia dos preditores foi feita através da comparação dos resultados obtidos por eles com os dados referentes aos domínios reais de cada alvo, que são disponibilizados pelo CASP.

O InterProScan e o Cdvist obtiveram os melhores resultados dentre os preditores avaliados, apresentando uma maior capacidade preditiva. O InterProScan demonstrou-se ainda mais promissor devido à sua disponibilidade para uso local, enquanto o Cdvist funciona apenas através de um servidor *on-line*. Embora o InterProScan tenha se destacado dos demais preditores, seus resultados apresentaram acurácia elevada em apenas 44% dos alvos avaliados. Assim, observou-se que a identificação de domínios proteicos não é um problema trivial e continua sendo um desafio para as metodologias atuais, fazendo-se necessário o aprimoramento dos métodos e o desenvolvimento de novas estratégias. Atualmente, o InterProScan está sendo aplicado como parte da estratégia de predição de domínios do programa de predição de estruturas de proteínas GAPF.

**Palavras-chave:** Predição de Domínios Proteicos; Predição de Estrutura de Proteínas; Proteínas Multidomínios.



## **Desenvolvimento de uma versão paralela híbrida para contabilização da frequência de repetição de k-mers**

Micaella Coelho, Carla Osthoff e Fabrício Vilasbôas

O desenvolvimento tecnológico na área de bioinformática tem gerado uma grande quantidade de dados. Esta grande quantidade de dados tem demandado um maior poder computacional e, graças aos avanços das plataformas computacionais de alto desempenho, esse processamento tem sido possível em tempo hábil e com a precisão desejada. Com o advento de novas tecnologias em processamento de alto desempenho, surge um novo paradigma. Este paradigma consiste em desenvolver algoritmos que utilizem todos os recursos disponíveis em um sistema computacional.

Nesse projeto, temos como objetivo apresentar técnicas de otimização utilizando uma arquitetura manycore em GPU para a classificação de metagenomas. O algoritmo utilizado é o k-mer. O algoritmo k-mer aplicado a metagenoma consiste em contabilizar a frequência de repetição de todas as possíveis combinações de k nucleotídeos em uma sequência. Este estudo apresenta uma análise de desempenho de uma versão paralela híbrida envolvendo a biblioteca MPI e o processamento em GPU's. Mostramos que a versão híbrida possibilita o aumento do desempenho do algoritmo ao possibilitar a execução em paralelo em múltiplas GPU's. Mostramos também que a versão aumenta o desempenho em uma única GPU que apresenta funcionalidade Hyper-Q.

Os dados utilizados para os testes são dados reais obtidos da base de dados de metagenoma do *NCBI* (*National Center for Biotechnology Information*). O sequenciamento foi feito utilizando equipamentos da nova geração de sequenciadores, o *Illumina*, que geram uma grande quantidade de dados com alta fidelidade. Equipamentos como este necessitam de novas tecnologias, com capacidade de processamento de dados muito superior aos utilizados para o processamento dos dados gerados pelos sequenciadores das gerações anteriores.