

Visualização e Interatividade com modelos 3D do corpo humano em WebGL

Aluna: Ana Paula Lourenço Ramos

Orientador: Jauvane C. de Oliveira

Resumo

Com o surgimento de novas tecnologias também surgem novas necessidades. Algumas necessidades podem ser supridas através de multiplataformas web, que proporcionam flexibilidade de desenvolvimento e capacidade de interação para o usuário.

O objetivo do projeto proposto e em andamento é uma apresentação do atlas do corpo humano, com capacidade de interação e boa representação da anatomia humana em 3D.

Para ser implementado foi utilizado o WebGL, tecnologia emergente para criação de gráficos 3D em navegadores web. Para melhor desempenho, foi acrescentado a biblioteca Three.js, construída em javascript e que permite além de controle para dispositivos de entrada e saída boa visualização dos modelos importados para o navegador.

Modelo de estudos de caso sobre a Aplicação do Modelo Atmosférico

Arthur B. Oliveira (IC), Carla Osthoff(Orientadora)

arthurbo@lncc.br – osthoff@lncc.br

Resumo

Este trabalho apresenta o desempenho paralelo de um modelo de simulação numérica global, Ocean-Land-Atmosphere Model (OLAM), em uma arquitetura híbrida multicore ambiente / GPU, sob os modelos de programação a seguir: (1) uma implementação paralela MPI executando um processo por núcleo em o nó multi-core, (2) uma implementação OpenMP paralelo executando um thread por núcleo no nó multi-core, (3) um híbrido implementação OpenMP / Cuda corredor paralelo threads OpenMP sobre os núcleos e threads Cuda na GPU. Mostramos que à medida que aumentamos o número de núcleos, a implementação paralela MPI processos em execução em cada núcleo de executa-los mais rápido outras implementações.

Introdução

Modelos numéricos têm sido largamente usados nos últimos anos a fim de compreender e prever os fenômenos climáticos e do tempo. De um modo geral, duas abordagens têm sido implementadas para o desenvolvimento de modelos numéricos: global e regional. Modelos globais têm resolução espacial de cerca de 2-5 graus de latitude e, portanto, não podem representar muito bem a escala de fenômenos climáticos regionais. Para representar todos os processos meteorológicos é necessário uma maior resolução. Consequentemente, isso requer uma grande capacidade de processamento. Os modelos regionais permitem a simulação com uma resolução maior mas são usadas para regiões específicas, mas não podem simular os fenômenos de larga escala. Recentemente, foi desenvolvido na Universidade de Duke um novo modelo que representa uma nova geração de modelos meteorológicos. A principal característica deste modelo, chamado de Ocean Land Atmosphere Model (OLAM), é a capacidade de representar fenômenos climáticos em escala global e escala regional [Walko]. Devido às grandes demandas computacionais e limitações de tempo de execução, OLAM, assim como os outros modelos, depende de processamento paralelo.. Por outro lado, durante a próxima década o grau de paralelismo no chip irá aumentar significativamente e processadores deverão conter dezenas ou mesmo centenas de núcleos. Atualmente as placas de processamento gráfico, GPU, estão sendo largamente utilizadas na área de computação científica. Este trabalho apresenta o desempenho paralelo para a aplicação do modelo de simulação atmosférica em um sistema híbrido com processadores multicore e dispositivos em GPU para modelos de programação distintas.

Configuração Típica de uma Simulação de Tempo

Realizamos o nosso estudo de caso com uma típica simulação de uma previsão de tempo global, com 40 km resolução horizontal, que gera uma subdivisão na grade icosaedro do OLAM de 25 x 25. A camada atmosférica (dimensão z) foi dividida em 28 camadas. Simulamos 24 horas de integração das equações da dinâmica atmosférica, sem qualquer cálculo adicional física (tais como umidade e processos radioativos) porque temos interesse apenas no impacto sobre o custo de execuções de dinâmica de fluidos. Cada passo de tempo de integração de 60 segundos simula o do tempo real. Esta simulação típica requer 10GB de leitura de arquivos de entrada e condições iniciais. Os arquivos de entrada do Olam não são particionados para processamento paralelo. Arquivos de entrada típicos são: condições globais iniciais de uma determinada data e hora e mapas globais que descrevem a topografia, tipo de solo, coberto de gelo áreas, Olson Global de Ecossistemas (OGE) conjunto de dados de vegetação, profundidade do solo interagindo com a zona de raiz temperatura da superfície, do mar e Índice de Vegetação por Diferença Normalizada (NDVI). Depois de ler o arquivo de entrada, o OLAM simula um número de passos_de_tempo, que evolui as condições atmosféricas no tempo em iterações discretas. Após cada iteração, os processos trocam mensagens com seus vizinhos para manter o estado atmosférico consistente. Isto é feito de uma maneira assíncrona para esconder o custo de transmissão da mensagem. Depois de executar uma série de passos_de_tempo, as variáveis que representam a atmosfera são gravadas em um arquivo de histórico. Durante esta fase, cada processo atualiza o estado atmosférico e fecha o arquivo de histórico. Estes arquivos de histórico são considerados de tamanho pequeno para os padrões de aplicações científicas. Dividimos algoritmo OLAM em três partes principais: a inicialização, o cálculo do tempo atmosférico e o estado da saída.

Avaliação de Desempenho

As medidas de desempenho foram feitas em um sistema multi-core/manycore, denotado prjCUDA, composto de dois Intel Core i7, com 12 MB de cache L3 e 24 GB de memória RAM, GTX580 e Tesla C2070. O software empregados foram MPICH versão 2-1.2.p1, VTune Performance Analyzer versão 9.1, CUDA Toolkit versão 4.0 e IGP versão FORTRAN 11,2 compilador. Como estamos executando em um sistema único nó, OLAM entrada e saída vezes execução de fase são os mesmos para todas as implementações, portanto, não estão usando-os para nossa análise. Neste experimentos alteramos o código de um loop de uma das rotinas “hotspot” do código do OLAM, com as diretivas do OpenAcc para possibilitar a execução do mesmo na placa GPU, e comparamos o desempenho do novo código com o código antigo, que dividia o domínio de dados entre os 12 processadores do desktop através do MPI.

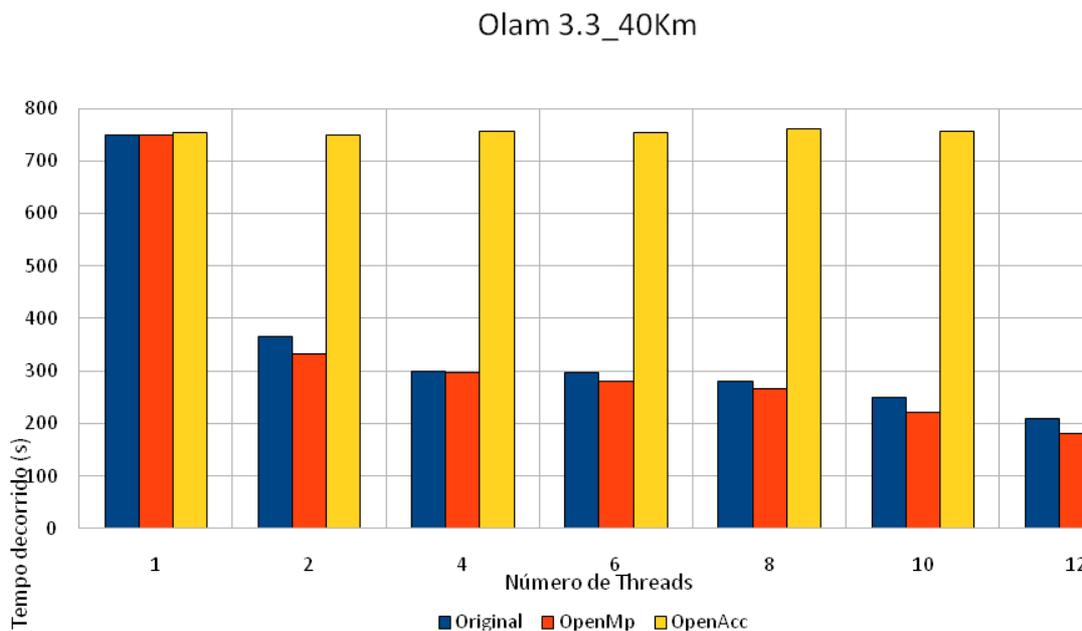
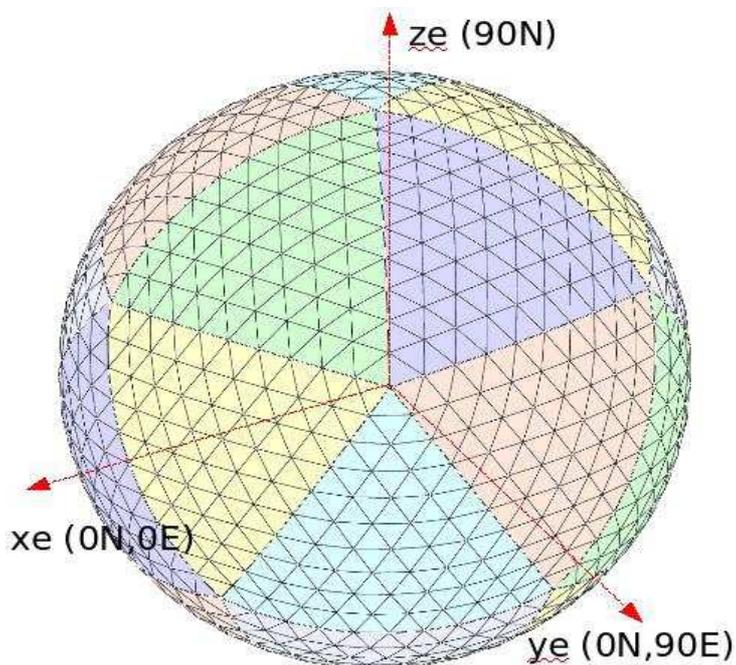


Figura 1 Esquerda: OLAM subdividida malha icosaédrica e sistema de coordenadas cartesianas



com origem no centro da Terra [Walko]. Direita: OLAM programação tempo de execução modelos como aumenta o número de núcleos em comparação com

processador multicore speedup ideal.

Figura 2 Direita: Mostra os Resultados da média de 3 rodadas, comparando o Olam original, com duas variações na Scalar_Transport. A primeira variação usou OpenMp (vermelho) que apresentou ligeiro aumento de desempenho. A segunda variação usou OpenAcc (amarelo) que não variou o desempenho, é possível afirmar que os loops paralelizados em OpenAcc não possuíam complexidade suficiente para mostrar bons resultados.

Conclusão

Nos experimentos, observamos que, à medida que aumentam o número de núcleos, a implementação MPI executa melhor do que implementações ApenACC estamos usando MPI 2 compartilhada de memória padrão para comunicação entre nós, portanto, fornece uma comunicação de alto desempenho para sistemas multicore. Observamos que o gargalo implementação MPI é contenção de memória e que o gargalo da implementação com o Acelerador OpenMP é devido ao paralelismo rotina baixa. Além disso, observa-se que tanto o MPI e OpenMP implementações de aumento de velocidade são limitadas ao número de núcleos a partir de cada nó de vários núcleos, e que o híbrido MPI / OpenMP / CUDA desempenho implementações é limitada a carga de trabalho OLAM resolução. Para o trabalho futuro, planeja executar OLAM implementação MPI / Cuda de maior carga de trabalho e de 40km para configurações Olam híbridos carga de trabalho.

Referências

OSTHOFF, C., OLIVEIRA, A., SOUTO, R.P., J., GRUNMANN, P., SILVA DIAS, P. L., BOITO, F., KASSICK, R., PILLA, L., SCHEPKE, C., MAILLARD, N., NAVAUX, P., PANETTA, J., LOPES, P.P. AND WALKO, R. I/O PERFORMANCE EVALUATION ON MULTICORE CLUSTERS WITH ATMOSPHERIC MODEL ENVIRONMENT. ATMOSPHERIC MODEL APPLICATIONS, INTECH OPEN SCIENCE BOOK, ISBN 978-953-51-0488, 2012.

WALKO, R.L. AND AVISSAR, R. OLAM: OCEAN-LAND-ATMOSPHERE MODEL - MODEL INPUT PARAMETERS - VERSION 3.0. TECH. REP., DUKE UNIVERSITY, NOVEMBER, 2008.

Modelagem de proteínas da cápsula de *Klebsiella pneumoniae*

AD Rossi¹, ATR Vasconcelos¹, CE Thompson¹

¹ Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Rio de Janeiro

O impacto de bactérias patogênicas no sistema público de saúde é enorme, causando milhares de mortes. Cepas e espécies diferentes são responsáveis por surtos hospitalares, dentre elas, *Klebsiella pneumoniae*, uma bactéria gram-negativa, que pertence à família Enterobacteriaceae. Este microrganismo é multirresistente à antibióticos e é especialmente prejudicial a pacientes imunodebilitados. A *K. pneumoniae* possui uma matriz polissacarídica que confere aderência às células do hospedeiro, prevenindo a dessecação. A composição da cápsula parece ser importante para o fator de virulência, permitindo a classificação e identificação de novos sorotipos. Dessa maneira, os estudos da evolução das proteínas estruturais de cápsula fornecem informações fundamentais sobre os fenótipos multirresistentes às drogas.

Inicialmente, uma análise extensa de todas *orfs* resultantes do pirossequenciamento de uma cepa de *K. pneumoniae* isolada em um hospital universitário em Londrina, Brasil, foi conduzida através da comparação de cada sequência com bancos de dados biológicos. Todas as sequências de proteínas utilizadas no processo de modelagem, incluindo os possíveis moldes e proteínas obtidas do pirossequenciamento foram obtidas do Centro Nacional de Informação Biológica (NCBI). Os alinhamentos das sequências de aminoácidos foram obtidos com o programa ClustalW utilizando a matriz BLOSUM para pontuação e ajustados manualmente com GeneDoc para otimização do alinhamento. A partir do melhor alinhamento obtido, o programa MODELLER foi utilizado para gerar modelos de proteínas de acordo com a metodologia de modelagem de proteínas através de comparação por homologia. Finalmente, o melhor modelo para cada proteína foi selecionado com base nos resultados obtidos pelos programas PROCHECK e VERIFY-3D, os quais executam análises estereoquímicas e avaliam o ambiente encontrado nas estruturas obtidas por cristalografia que serviram de molde, respectivamente. O ciclo de realinhamento, modelagem e validação da estrutura foi repetido até não ser possível perceber melhora nos resultados. O RMSD (“Root-Mean-Square Deviation”) entre o molde e cada modelo foi calculado, todas as figuras foram desenhadas e suas superfícies moleculares foram obtidas utilizando o programa Swiss PDB Viewer.

No total, 12 proteínas foram modeladas através de modelagem comparativa por homologia. Os modelos resultantes possuem uma boa qualidade estereoquímica, com exceção da KP03136. Três proteínas da cápsula da *K. pneumoniae* foram analisadas completamente (KP31534, KP03796 e KP03793). As

demais proteínas encontram-se em fase de validação estereoquímica através do software PROCHECK. Certas proteínas de cápsula quando inibidas, em outras cepas, causam a perda de virulência, sugerindo que possivelmente isto possa ocorrer na cepa estudada. Adicionalmente, análises filogenéticas estão sendo conduzidas a fim de auxiliar no entendimento da evolução molecular e seu impacto na estrutura das proteínas da cápsula. O próximo passo do projeto será o estudo de docking das proteínas envolvidas com a virulência.

Estudo de um modelo híbrido para a dinâmica entre presas e predadores

Flávio Henrique dos Santos, Cezar Welter,

Maurício Kritz, Jaqueline Maria da Silva.

Palavras-chave: *Populações, Lotka-Volterra, Verhulst.*

Quando estudamos o crescimento de populações é possível adotar modelos matemático-computacionais como ferramenta de estudo [1,2]. Neste trabalho, propomos o estudo gradativo de cada fator que interfere na dinâmica de interação entre duas populações. Analisamos o modelo de Verhulst e o sistema de equações de Lotka-Volterra para crescimento populacional e interação entre espécies. Assim, apresentamos um modelo populacional híbrido que aborda a dinâmica de duas populações concorrentes.

O modelo de Verhulst supõe que uma população deverá crescer até um limite máximo, isto é, que a população tende a estabilizar [3]. O modelo se apresenta na forma:

$$\frac{dx}{dt} = rx\left(1 - \frac{x}{L}\right)$$

Onde r é taxa de crescimento e L a capacidade de suporte.

O modelo de Lotka-Volterra é um modelo de competição entre duas populações, onde uma é a presa e outra um predador [3], tal que:

$$\begin{cases} \frac{dx}{dt} = ax - \alpha xy \\ \frac{dy}{dt} = -by + \beta xy \end{cases}$$

O modelo híbrido propõe a utilização do modelo de competição, adicionando a este um fator limitante, como no modelo de Verhulst. Assim:

$$\begin{cases} \frac{dx}{dt} = ax\left(1 - \frac{x}{L_1}\right) - \alpha xy \\ \frac{dy}{dt} = -by + \beta xy\left(1 - \frac{y}{L_2}\right) \end{cases}$$

Onde: a , b , α e β são parâmetros, x o número de presas, y o número de predadores, L_1 o limite de crescimento da presa e L_2 o limite de crescimento do predador.

Resultados obtidos por simulações numéricas do modelo híbrido nos mostraram que o diagrama de fases converge para um ponto de equilíbrio. A dinâmica das populações ocorre de forma periódica e tende a estabilizar, como se observa na natureza.

Referências:

- [1] J.M. Silva, "*Modelos para a Dinâmica de Vegetação em Áreas Alagáveis Amazônicas*", Tese de Doutorado, LNCC, 2011.
- [2] R.C. Bassanezi, "*Ensino-aprendizagem com modelagem matemática*", Editora Contexto, São Paulo, 2006.
- [3] RONEY R. N. "*Dinâmica de populações: um breve histórico*". In: BIENAL DA SBM - IME/UFG, 3., 2006. Goiânia: Anais eletrônicos. Goiânia, 2006.

Variabilidade da precipitação na Região Serrana do Rio de Janeiro Estimada pelo Canal Infravermelho em Satélites Goestacionários

Aluna: Gisele Burger Vieira

Orientador: Pedro Leite Da Silva Dias

Co-Orientador: Fabiano Gomes

O painel intergovernamental de mudanças climáticas – IPCC(2007) indica claramente que surgiram evidências claras de um aumento significativo na intensidade e frequência de fenômenos climáticas extremos. Fenômenos de grande poder de destruição, que estão associados a enormes prejuízos econômicos e perdas humanas que vêm ocorrendo em diversas regiões do globo terrestre. Entretanto, o IPCC(2007) indica que no caso da precipitação, as evidências de alterações atribuídas ao aquecimento global ainda são tênues pois a variabilidade climática natural, associada a ciclos interanuais(ENSO – EL Niño/Oscilação sul) e décadas/multidecadais(p.ex., a “Pacific Decadal Oscillation”), ou ciclos solares é dominante, pelo menos até agora. Portanto faz-se necessário explorar com maior profundidade as fontes de informação que permitem a estimativa da precipitação, principalmente em regiões onde a variabilidade espacial é muito grande como, por exemplo, a região serrana do Rio de Janeiro entretanto vimos que a região serrana do Rio de Janeiro foi particularmente afetada por chuvas intensas no dia 12 de janeiro de 2011. As cidades mais afetadas foram Teresópolis, Nova Friburgo, Petrópolis, Sumidouro e São José do Vale do Rio Preto. Foram contabilizadas cerca de 900 mortes e mais de 300 desaparecidos e cerca de 35 mil desalojados em consequência dos desastres naturais. A tragédia foi considerada como o maior desastre climático da história do país, superando os 463 mortos do temporal que atingiu Caraguatatuba em 1997 com isso o principal objetivo desse projeto é obter estimativas da probabilidade de ocorrência de precipitação em diferentes percentil. Em particular, deverão ser

definidos os limiares associados aos extremos de precipitação (15%, 10%, 5%) e avaliada a variabilidade espacial desses limiares na amostra de cerca de 10 anos de dados. Numa região com topografia tão variável como na região serrana do Rio de Janeiro espera-se que a variabilidade dos limiares associados aos percentis seja muito significativa. Tais resultados serão comparados com outras estimativas de precipitação (satélite TRMM, com radar ativo à bordo).

Antes de se analisar as funcionalidades dos dados se viu a necessidade de entender de maneira mais específica os princípios básicos do funcionamento do estimador. Podendo assim estudar o como serão feitas as estatísticas desses dados. Com isso podemos ter um benévolo conhecimento em variabilidade da precipitação na região serrana do Rio de Janeiro, entretanto a pesquisa tem como foco indicar claramente que surgiram evidências de um aumento significativo na intensidade e frequência de fenômenos climáticos extremos. Fenômenos estes de grande poder de distribuição, que estão associados a enormes prejuízos econômicos e perdas humanas, e têm ocorrido em diversas regiões do globo terrestre. Sendo assim iniciou-se o estudo de ferramentas para criação de dados utilizando programas específicos: o GRADS e o MATLAB.

O GRADS é uma ferramenta computacional interativa para a análise e exibição de dados de ciência da Terra que é usado no globo todo e livremente disponível na internet. O software executa um modelo de dados de 4-Dimensões, onde as dimensões são geralmente latitude, longitude, nível e tempo. Cada conjunto de dados está localizado dentro deste espaço de 4-Dimensões através do uso de um arquivo descritor de dados. Dados em grade e de estação podem ser descritos e estar espaçados não linearmente. As grades gaussianas e de modelos de oceano de resolução variável são suportadas diretamente. O formato interno dos dados no arquivo pode ser binário, GRIB, BUFR, NetCDF, ou HDF-SDS.

O MATLAB (MATRIX LABORATORY) é um software destinado a fazer cálculos com matrizes. Os comandos do MATLAB são muito próximos da forma como escrevemos expressões algébricas, tornando mais simples o seu uso. Podem ser incorporados às rotinas pré-definidas, pacotes para cálculos específicos. Escrevemos um pacotes

chamado gaal com funções que são direcionadas para um curso de Geometria Analítica e Álgebra Linear.

Tendo em mãos os dados criados pelo GRADS, pode-se fazer as estatísticas dos mesmos através do MATLAB. Construindo assim um passo a passo de manipulação de dados e Após o termino desta primeira etapa será realizada uma análise estatística da distribuição da probabilidade de precipitação por pixel. Em seguida serão gerados mapas de probabilidade de ocorrência de precipitação estimada pelo canal infravermelho do satélite.

Foram fornecidos pelo Laboratório MASTER do IAG/USP dados de satélite Geoestacionários, tais dados se encontram em pontos de grade em diferentes resoluções espaciais por um período de aproximadamente dez anos, após serem feitos os downloads de todos os dados em formato .bin, foram analisados no GRADS, utilizando as coordenadas de latitude e longitude da região serrana localizada no estado do Rio de Janeiro. Onde a variabilidade espacial é muito grande. Até o momento somente os dados no período de dez anos que foram coletados estão sendo analisados, restando apenas as avaliações estatísticas da distribuição da probabilidade de precipitação por pixel, onde será utilizado intervalos de 5% associados aos percentis. Tais avaliações são necessárias e úteis, pois serão utilizadas para gerar mapas de probabilidade de ocorrência de precipitação estimada pelo canal infravermelho do satélite.

O projeto ainda encontra-se em andamento, havendo ainda a necessidade de rever alguns dados como também novos downloads e avaliações futuras.

Análise filogenética de proteínas da cápsula de *Klebsiella pneumoniae*

RNB da Silva¹, ATR Vasconcelos¹, CE Thompson¹

¹ Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Rio de Janeiro

Os aspectos evolutivos de relações patógeno-hospedeiro intrigam biólogos há tempos. O sistema imunológico humano evoluiu para um sistema de alta complexidade, de proteção contra agentes infecciosos, enquanto patógenos desenvolveram um grande arsenal químico que confere versatilidade genética, permitindo sua sobrevivência.

Klebsiella pneumoniae provou ser um microorganismo de grande impacto para a saúde humana, acometendo pacientes imunologicamente debilitados. Sua cápsula é considerada um fator de virulência, portanto é de grande importância o entendimento das suas proteínas constituintes, bem como dos mecanismos envolvidos na sua evolução. Tais estudos podem auxiliar no desenvolvimento de novas drogas e na identificação de novas cepas multirresistentes.

As sequências, destacadas como sequências alvo, dos genes *uge-1*, *rmlD* e *gnd* foram obtidas por intermédio do pirosequenciamento de *Klebsiella pneumoniae* em um hospital universitário de Londrina, Brasil. As três sequências de aminoácidos correspondentes aos genes supracitados foram primeiramente analisadas em razão de sua importância funcional em *Klebsiella*. As proteínas codificadas pelo pirosequenciamento foram analisadas e comparadas com banco de dados biológicos. Para os alinhamentos múltiplos das sequências alvo e das sequências obtidas por intermédio da comparação, utilizou-se o software COBALT (*Constraint-based multiple alignment tool*). Ciclos de realinhamentos foram conduzidos para eliminar sequências com um número excessivo de *gaps*. O software GeneDOC 2.6 foi empregado para realizar uma inspeção visual e rearranjo das sequências, quando necessário.

A classificação taxonômica foi baseada no banco de dados do NCBI (*National Center for Biotechnology and Information*). As análises filogenéticas foram realizadas

pelo software MEGA (Molecular Evolutionary Genetics Analysis, versão 5.05) aplicando o método *neighbor-joining* (NJ), que baseia-se no critério evolutivo mínimo, ou seja, a topologia que apresenta o menor comprimento de ramos.

No método de *neighbor-joining*, os modelos de substituição *p-distance* e Poisson, foram utilizados para a comparação das sequências de aminoácidos. As opções de deleção completa e pareada de *gaps* foram implementadas, realizando um ciclo de 1.500 repetições, pelo teste filogenético de *bootstrap*.

Não houve diferença significativa entre as árvores filogenéticas geradas pelos diferentes algoritmos utilizados. O gene *uge-1* codifica a proteína difosfato uridina galacturonato-4-epimerase, que têm por função a transformação da UDP-D-glucuronato em UDP-D-galacturonato. O melhor resultado atendendo aos parâmetros do algoritmo é referente a uma *K. pneumoniae* subespécie *rhinoscleromatis* ATCC 13884. As análises filogenéticas evidenciaram a formação de um *cluster* bem suportado pelo teste de *bootstrap*. Ademais, sequências da família *Enterobacteriaceae* também formaram um *cluster*. O gene *rmlD* codifica a proteína dTDP-4-dehidroramnose redutase e forma um *cluster* com outras espécies e cepas de *Klebsiella*. O gene *gnd* correspondente à proteína 6-fosfogluconato dehidrogenase possui como melhor resultado uma *K. pneumoniae* cepa VGH484. Na árvore filogenética apresenta-se próximo à *Escherichia coli* MS 146-1.

Análises evolutivas envolvendo as demais proteínas da cápsula de *K. pneumoniae* serão conduzidas. Além disso, pretendemos realizar um estudo filogenômico e, a partir deste, avaliar a ocorrência de eventos de transferência horizontal entre espécies de *Klebsiella*. A presença de seleção positiva também será avaliada e, posteriormente, correlacionaremos os resultados dessas análises com as estruturas tridimensionais das proteínas da cápsula, que encontram-se em fase final de validação estereoquímica.

Ambiente de Replicação para o Armazenamento Seguro dos Dados do Portal ComCiDis e Aplicações

Bolsista: Victor Dias de Oliveira

Orientador: Bruno Schulze

1 . Introdução

Atualmente é evidente o forte crescimento dos sistemas computacionais, com perspectivas promissoras para um futuro próximo. Nesse sentido, é imprescindível priorizar garantias de funcionamento ininterrupto de forma segura dos sistemas de informações dos quais certamente eles dependem.

Os servidores de bancos de dados, ao integrarem os sistemas computacionais, não podem apresentar falhas ou indisponibilidade dos dados, o que implicaria diretamente em perda de informações importantes para todos os sistemas que são dependentes deles, logo, concluímos que isto poderia ser evitado com a utilização de ambientes distribuídos visando à alta disponibilidade de informações.

Diante de tais assertivas, o Plano de Trabalho teve por finalidade dar continuidade ao trabalho de Robustez no armazenamento de dados utilizando replicação, este já iniciado no plano de trabalho anterior, tornando-o mais robusto para melhor utilização nos projetos e aplicações futuras do laboratório ComCiDis, sempre buscando a utilização de ambientes virtualizados.

A escolha por ambientes virtualizados possibilita ao usuário o aporte de confiabilidade, mobilidade e escalabilidade em suas aplicações, fazendo com que o gerenciamento, a migração de aplicações e sistemas operacionais, possa ser utilizado de forma menos complexa, inclusive, com redução de custos materiais quando se fala em hardware.

2. Motivação e Proposta

O plano de trabalho a seguir teve por objetivo dar continuidade ao trabalho anterior “Robustez no armazenamento de dados utilizando replicação” e fazer a validação do projeto utilizando os dados gerados pelo Portal do laboratório ComCiDis e pelas aplicações desenvolvidas pelo mesmo.

Inicialmente, consistira em dar continuidade ao trabalho desenvolvido, aperfeiçoando a ferramenta de geração de backup e restauração de banco de dados, utilizando no ambiente as replicação parcial ou total. Na segunda etapa, integrar a ferramenta desenvolvida de forma facilitar o gerenciamento dos dados gerados pelos projetos do ComCiDis e tornar o armazenamento destes dados mais seguro e tolerante a falhas. Pretende-se com isso aumentar a confiabilidade dos dados e a sua disponibilidade.

Logo abaixo são listadas e enumeradas as tarefas propostas e objetivos traçados:

1. Dar continuidade ao desenvolvimento do sistema de replicação de dados criado, aperfeiçoamento de suas funcionalidades e tornando-o mais compatível com o ambiente do ComCiDis;
2. Identificar e analisar novos riscos de armazenamento dos dados;
3. Integrar a ferramenta desenvolvida aos demais projetos desenvolvidos no laboratório CimCiDis (Portal ComCiDis e ICAL);

3. Metodologia

Nesse tópico é feito um estudo e levantamento dos principais pontos para entendimento dos conceitos realizados nesse trabalho realizado no período da bolsa de iniciação científica. Nessa parte são abordadas questões e análise das possíveis soluções, além dos benefícios proporcionados por cada uma delas.

3.1 Replicação Síncrona e Replicação Assíncrona

Replicação de dados se refere ao armazenamento de dados e a estratégia de backups entre computadores em locais distintos. A replicação sob uma rede de computadores pode ser feita inteiramente independente do armazenamento central de dados. A replicação é mais popularmente vista como o processo de compartilhar e distribuir informações entre bancos de dados relacionais, tornando possível a redundância e balanceamento dos dados.

Existem duas maneiras para replicar dados, são elas: a síncrona e a assíncrona. A replicação síncrona é geralmente do tipo mestre para mestre, por outro lado a assíncrona é realizada do mestre entre seus escravos. Um servidor mestre é aquele que pode realizar operações de leitura e escrita, enquanto que um servidor escravo apenas realiza operações de leitura.

Na replicação síncrona, os dados presentes são sempre os mesmo em todos os servidores. Nela as operações de escrita são distribuídas ao mesmo tempo para todos os nós que fazem parte da replicação.

A replicação assíncrona geralmente é formada por um nó mestre, encarregado de distribuir as operações de escrita para um ou mais nós escravos. A troca de mensagem nesse método é realizada da seguinte forma: todas as operações de escrita são enviadas ao nó mestre, que é encarregado de processar e enviar periodicamente as modificações realizadas na base de dados.

3.2 PgPool-II

PgPool-II é um *middleware* que atua como um servidor de banco de dados, de forma transparente para o cliente e o próprio servidor PostgreSQL. Essa ferramenta tem como características a replicação síncrona, um pool de conexões, e balanceador de carga, que será detalhado mais abaixo suas características. Outras vantagens do PgPool- II são: a capacidade de realizar detecção de falhas, adicionar outros servidores sem a necessidade de reiniciar o serviço e uma interface web conhecidos por Pgpool Admin.

A ferramenta web PgpoolAdmin auxilia na administração das funcionalidades do Pgpool-II. Instalada no próprio servidor do Pgpool-II ela facilita tarefas de monitoramento, administração do processo e alterações na configuração.

Por possuir essas características, o PgPool-II foi a ferramenta que mais se encaixou nas características do projeto. Essa ferramenta além de replicar utilizando o modo de replicação síncrona, tem várias outras características que a tornaram a ferramenta ideal para ser utilizada, uma delas, é a possibilidade de detecção de falhas autônoma, separando um nó defeituoso dos demais não interrompendo o funcionamento do sistema.

Outras características que tornaram também o PgPool-II a melhor solução para o presente projeto:

- *Pool de Conexões:* Essa característica possibilita a economia de conexões, reutilizando uma conexão com mesmas propriedades de uma já existente, reduzindo assim a sobrecarga de conexões.
- *Conexões Excedentes:* Após exceder o limite de conexões possíveis no PostgreSQL, essas conexões são rejeitadas. O PgPool-II captura essas conexões rejeitadas pelo servidor e as coloca em uma fila, para mais tarde serem utilizadas.

3.3 ICAL - Intelligent Collector Algorithm

O ICAL (Intelligent Collector Algorithm), é um agente coletor *opensource* para o monitoramento de ambientes reais e virtuais de alto desempenho, desenvolvimento dentro do nosso laboratório ComCiDis. O destaque alcançado por sistemas operacionais *Unix-like* na composição de infraestrutura e suporte a aplicações de HPC, orientaram a sua arquitetura e desenvolvimento. Possui uma arquitetura em camadas, que lhe proporciona escalabilidade e manutenibilidade.

A principal característica do ICAL é executar as tarefas de monitoramento de recursos, removendo a responsabilidade dos nós do *cluster* de processar e de armazenar as informações coletadas, enviando estas informações para servidores dedicados fora do ambiente de alto processamento, com o objetivo de reduzir a carga de processamento gerada pelo monitoramento.

A alta capacidade de configuração dos agentes do ICAL é o elemento chave por proporcionar um maior nível de detalhamento das informações. Permite que cada recurso seja monitorado independentemente. Possibilita aumentar o detalhamento das informações caso um recurso se apresente instável, ou reduzir caso apresente um comportamento considerado esperado.

O sistema apresentado é específico para ambientes de alto desempenho, sendo capaz de monitorar informações relevantes, possibilitando melhor a tomadas de decisões relativas à alocação e escalonamento de recursos, principalmente voltadas para ambientes virtuais, por meio do uso de *scripts*.

4. Funcionalidades e Arquitetura da Ferramenta Desenvolvida

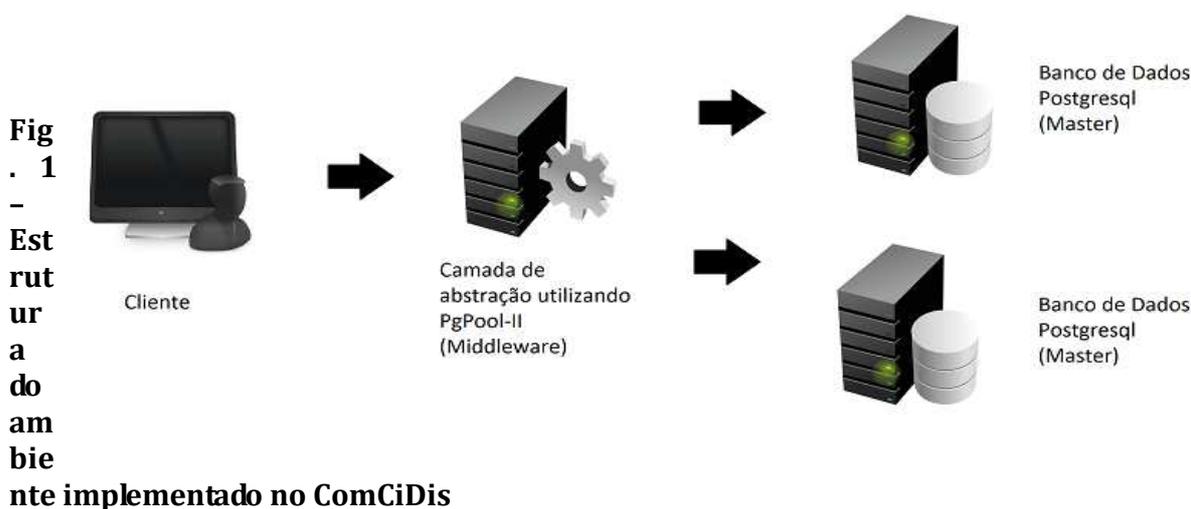
Esta seção descreve em maiores detalhes a arquitetura, o funcionamento e benefícios do sistema de robusto de armazenamento de dados e ferramenta de gerenciamento de backup, replicação e as máquinas que fazem parte do *cluster*.

4.1 Arquitetura do Sistema

O OpenBox, ferramenta criada neste trabalho, torna a robustez no armazenamento de dados consistente. Ela utiliza uma camada de abstração desconhecida tanto para o usuário quanto para o *cluster* de banco de dados, camada essa chamada de *middleware*. Para obter esse resultado foi utilizado o PgPool-II, mais detalhes e funcionalidade estão descritas na parte sobre metodologias utilizadas.

Inicialmente, para colocar o PgPool-II, ferramenta escolhida nesse projeto, em funcionamento, foi necessária fazer com que as máquinas que interagem nesse sistema se comunicassem visando permitir a conexão com outro computador na rede,

de forma a executar comandos de uma máquina remota. Na figura abaixo [Fig.1] é demonstrado como está organizado o ambiente distribuído de banco de dados utilizando a ferramenta PgPool-II.



Após as configurações necessárias para o funcionamento, a ferramenta foi inicializada operando satisfatoriamente. Outra configuração necessária é a utilização de troca de chaves entre as máquinas permitindo comunicação segura sem a necessidade de digitação de senhas, apenas fazendo uso da troca segura de chaves privadas e chaves públicas entre as máquinas.

Para o teste de replicação, um *dump*, arquivo contendo todas as informações e estruturas do banco de dados gtvcg, foi rodado para saber se as estruturas e informações estavam sendo replicadas para os demais nós, terminando com êxito.

Para o teste de simulação de um nó defeituoso, o PgPool-II tenta se comunicar com uma determinada máquina, porém se não obter comunicação, aquele nó defeituoso é retirado do seu sistema e prossegue o processo normalmente. Quando aquele nó retorna, a ferramenta chama um *script* que atualiza a sua base com os demais nós da rede, reinserindo o nó.

4.2 OpenBox – Backup automatizados, restauração e gerenciamento do cluster

O OpenBox é uma aplicação de backup, de restauração e de compactação, que no plano de trabalho anterior foram implementadas as seguintes funcionalidades: dados mais antigos são retirados do banco de dados, mantendo assim sempre os dados mais recentes, os dados retirados devem ser guardados em um formato de arquivo XML, onde mais tarde são compactado utilizando um algoritmo de compactação Bzip2,

permitindo a restauração destes dados sempre que necessário.

O presente plano de trabalho, deu continuidade ao desenvolvimento desta ferramenta, tornando-a mais eficiente, e integrando-a as demais aplicações e sistema existentes e desenvolvidas no laboratório ComCiDis. O trabalho desenvolvido focou dois aspectos da ferramenta: o núcleo do sistema(*core*) e a interface com o usuário.

O núcleo do OpenBox, é encarregado de descobrir todas as características do banco de dados configurada pelo usuário, ela é capaz de fazer uma varredura descobrindo as tabelas existentes além da estrutura de cada tabela. Com isso, é retirado do usuário a responsabilidade de informar à aplicação como é a estrutura e nomes das tabelas que deseja salvar as informações como também restaurar. Existe ainda a possibilidade de salvar os dados ou de remove-los do banco de dados assim que concluído o backup. Os arquivos de backups são gerados em um formato XML [Fig. 2], com os dados pedidos pelo o usuário. Esses arquivo *xml*, é mais tarde compactado utilizando o algoritmo BZIP2, diminuindo assim significativamente o espaço gasto em disco.

```
<?xml version="1.0" encoding="UTF-8"?>
<gtvcg>
  <comunidade>
    <id_comunidade>1</id_comunidade>
    <nome_comunidade>Funcionou!</nome_comunidade>
  </comunidade>
  <comunidade>
    <id_comunidade>2</id_comunidade>
    <nome_comunidade>Comunidade Testing!</nome_comunidade>
  </comunidade>
```

Fig. 2 – Arquivo gerado pelo OpenBox utilizando a linguagem de marcação XML

O código-fonte do *core* do sistema foi reescrito, permitindo que novas funcionalidades fossem a ele integradas tais como a possibilidade de realizar backup e restauração de dados com outras SGBDS existente.

Esta reestruturação do código permitiu que OpenBox ampliasse a sua possibilidade de uso permitindo que fosse capaz de trabalhar tanto com uma base de dados PostgreSQL, quanto uma MySQL. Isto faz com que ele compatibilize e sirva de intercâmbio para a migração de dados e informações entre duas bases de dados diferentes, tornando o OpenBox uma ponte para troca de informações entre elas.

Com a integração da ferramenta com o ICAL, permitiu o gerenciamento remoto e distribuído dos recursos de cada máquina que fazem parte do cluster de banco de dados. Através da interface web foi possível o gerenciamento dos recursos consumidos por cada máquina que interagem neste sistema, possibilitando o conhecimento da quantidade de memória *ram* utilizada, a carga sobre o processador, entre outros recursos referente as atividades naquelas máquinas.

Abaixo, são descritas funcionalidades do OpenBox:

- Login com senha criptografada com algoritmo de hash SHA-1[5];
- Configuração do ambiente via interface web;
- Busca automática por tabelas e estruturas do banco de dados (Recurso otimizado);
- Fácil manipulação na operação de backup e compactação de dados (Recurso otimizado);
- Gerenciador de arquivos (backups);
- Um pool de conexões, possibilitando gerenciamento com diversos banco de dados ao mesmo tempo(Otimizado);
- Restauração das informações contidas no arquivo XML, sendo possível restaurá-lo tanto para o banco de dados PostgresQL quanto para o MySQL, função essa graças a criação de dialetos (Recurso otimizado e reestruturação do código-fonte);
- Gerenciador de tarefas;
- Gerenciamento das máquinas do PgPool-II utilizando ICAL para coleta dos recursos;
- Agendamento de tarefas;

5. Atividades Complementares

Nesse período de iniciação científica, foram realizados como atividades complementares diversos experimentos com o objetivos de simular um ambiente de acesso a dados totalmente virtualizado para verificar o comportamento de um servidor de banco de dados durante um aumento gradativo da quantidade de clientes de acesso simultâneos, verificando a necessidade de expandir a infraestrutura, acrescentando outros servidores à medida que o desempenho se mostrasse degradado.

Nos experimentos foi utilizado o DBT5, ferramenta que faz uso de um padrão de teste de *benchmark* para banco de dados o TCP-E, que simula um processo de transações online de uma corretora, aonde referencia um base de dados central que executa as transações relacionadas com as contas dos clientes da empresa. O esquema, e a base de dados, e as transações foram implementadas através de regras dos modernos sistemas OLTP (On-Line Transaction Processing).

O processo foi realizado utilizando três máquinas virtuais aonde cada uma delas disparava uma quantidade específica de clientes concorrentes, sendo utilizada uma sequência de 2, 4, 8, 16, 32, 64, 80, 96 e 100 clientes simultâneos para cada domínio.

Os testes foram realizados utilizando dois tipos de virtualizadores diferentes, o VirtualBox da Oracle e o KVM, infraestrutura de virtualização integrada ao linux, concluímos que a utilização do DBT5 permite verificar o comportamento do banco de dados em um ambiente concorrente e distribuído, possibilitando retirar informações para definir o melhor momento para aperfeiçoar o sistema. Na parte da virtualização

dos sistemas, percebemos que o VirtualBox trata melhor as operações de *I/O* (Leitura e Escrita do Sistema Operacional) tendo assim um ganho superior e melhor eficiência comparado ao KVM.

6. Conclusão

Conforme apresentado neste relatório, as metas e objetivos do plano de trabalho foram cumpridos dentro do prazo, ainda foram acrescentadas funcionalidades extras na proposta inicial.

Como contribuições geradas à partir do plano de trabalho, merecem destaque:

- Melhoria no gerenciamento dos dados gerados pelos projetos desenvolvidos no laboratório do ComCiDis;
- Aumento da segurança e da disponibilidade dos dados através de uso e implementação de ferramentas de replicação de dados e balanceamento de carga;
 - Redução do espaço em disco necessário para armazenar os dados dos projetos devido a incorporação de um módulo capaz de realizar backups compactados;
 - Integração da ferramenta com demais projetos desenvolvidos no ComCiDis;
 - Integração com o ICAL (Intelligent Collector Algorithm);
 - As modificações no código-fonte do projeto, possibilitando a integração de novas funcionalidade como o agendamento de tarefas e gerenciamento dos recursos das máquinas através da interface web.

MODELAGEM E SIMULAÇÃO DE PROBLEMAS DE CRESCIMENTO TUMORAL

Anna Claudia Mello de Resende¹ Regina Célia Cerqueira de Almeida²

Palavras Chave: Crescimento Tumoral, Quimioterapia, Modelagem Físico-Matemática, Equações Diferenciais, Métodos Numéricos.

Resumo. O câncer é hoje considerado um sério problema de saúde pública mundial e apesar do bem sucedido combate à doença, ainda permanecem em aberto questões relativas ao bom desempenho de suas modalidades de tratamento. O objetivo inicial do projeto concentra-se em estudar a modelagem físico-matemática associada a modelos de crescimento tumoral e administração de quimioterapia disponíveis na literatura bem como métodos numéricos para simulação computacional dos mesmos. Até o momento foram estudados modelos espacialmente homogêneos descritos por equações diferenciais ordinárias e métodos matemático-numéricos associados à resolução desta classe de equações sendo que a análise inicial feita baseia-se na investigação quantitativa do modelo matemático de quimioterapia anti-neoplásica sob dinâmica angiogênica.

Bibliografia

1. Modelagem Matemática em Câncer: Dinâmica Angiogênica e Quimioterapia Anti-Neoplásica, Diego Samuel Rodrigues, Dissertação de Mestrado, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Botucatu, São Paulo, 2011.

¹ Universidade Federal de Juiz de Fora (anna.claudia@ice.ufjf.br)

² Laboratório Nacional de Computação Científica - LNCC/MCT (rcca@lncc.br)

2. *Elementary Numerical Analysis*, K. Atkinson and W. Han, John Wiley & Sons Inc., 2004.

Análise de Desempenho e Emprego de GPU

Bolsista: André Muniz Yokoyama

Orientador: Prof. Bruno Schulze

Resumo

Com o intuito de verificar a capacidade real de processamento do “cluster” de GPUs foram desenvolvidos e aplicados testes de “benchmark” no “cluster”. Com a análise dos resultados dos testes obteve-se dados referentes a banda de memória, banda de comunicação entre nós e a capacidade de operações sobre ponto flutuante (“flops”). A análise dos resultados dos testes também possibilitou a detecção e eventual correção de problemas no “cluster”.

1 - Introdução

Durante o período de janeiro de 2012 à junho de 2012 tendo como principal objetivo avaliar as capacidades de processamento do cluster de GPUs, foram desenvolvidos e aplicados testes de benchmark com enfoque na velocidade de acesso a memória (“Memory Bandwidth”), banda de comunicação entre os nós (ethernet/MPI) e na capacidade de realizar operações sobre ponto flutuante (com precisão simples e dupla) por segundo (FLOPS).

Os testes de memória tiveram enfoque na transferência de dados entre a memória RAM do host e a memória da GPU em 4 modalidades, “Host to Device” com memória paginável (“pageable memory”), “Host to Device” com memória page-locked, “Device to Host” com memória paginável (“pageable memory”) e “Device to Host” com memória “page-locked”.

2- Testes

Os testes de banda de comunicação tiveram por objetivo a capacidade de transferência de dados da arquitetura atual que utiliza o protocolo de comunicação MPI sobre uma interface de comunicação “ethernet”, e futuramente uma comparação com uma arquitetura “Infiniband”.

Os teste de processamento se basearam na capacidade de realizar operações aritméticas sobre ponto flutuante de precisão simples (“float”) e de precisão dupla (“double”), nos testes baseou-se na operação de “Fused Multiply Add” (FMA), e memória compartilhada nas GPUs.

3 - Resultados

Durante a realização dos testes foram detectados problemas com dois dispositivos (GPUs) quanto a banda de memória, os dispositivos com problema apresentaram performance muito inferiores a média dos demais sendo mais acentuado nas operações de memória "page-locked", onde foram constatadas reduções de 50% a 90% em relação à média.

Nas figuras 1 e 2 que se referem aos testes de banda de memória com memória paginável nos dispositivos 0 (figura 1) e 1 (figura 2) percebe-se que apesar de uma maior flutuação nos resultados dos dispositivos 1 de cada nó a média geral foi superior a 4500 MBytes/s com pico de pouco mais de 5000 MBytes/s. No entanto observa-se que a média do dispositivo 0 do nó "gpu009" apresentou média de aproximadamente 2700 MBytes/s e o dispositivo 1 do nó "gpu002" de apenas pouco mais de 350 MBytes/s.

Já nas figuras 3 e 4 que se referem aos testes de banda de memória com memória "page-locked", os testes apresentaram uma flutuação mínima não perceptível na resolução atual dos gráficos, tendo uma média superior a 5600 MBytes/s com exceção do dispositivo 0 do nó "gpu009" com média de pouco mais de 2900 MBytes/s e do dispositivo 1 do nó "gpu002" com somente pouco mais de 380 Mbytes/s.

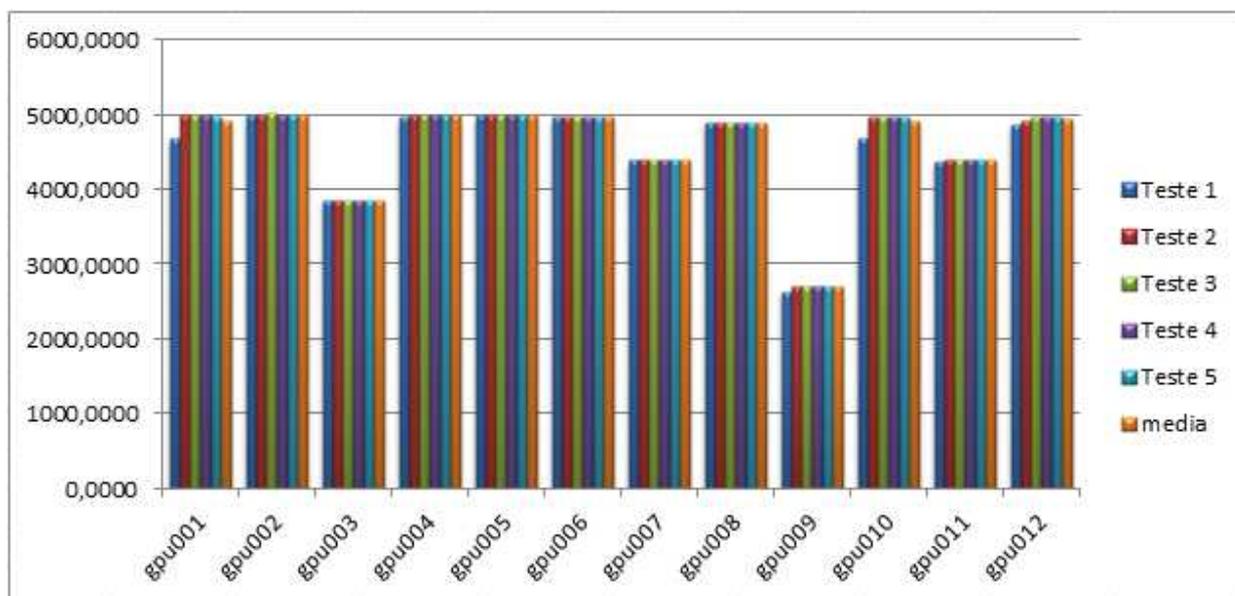


Figura 1 - Banda de memória "host to device" com memória paginável do device 0 (MBytes)

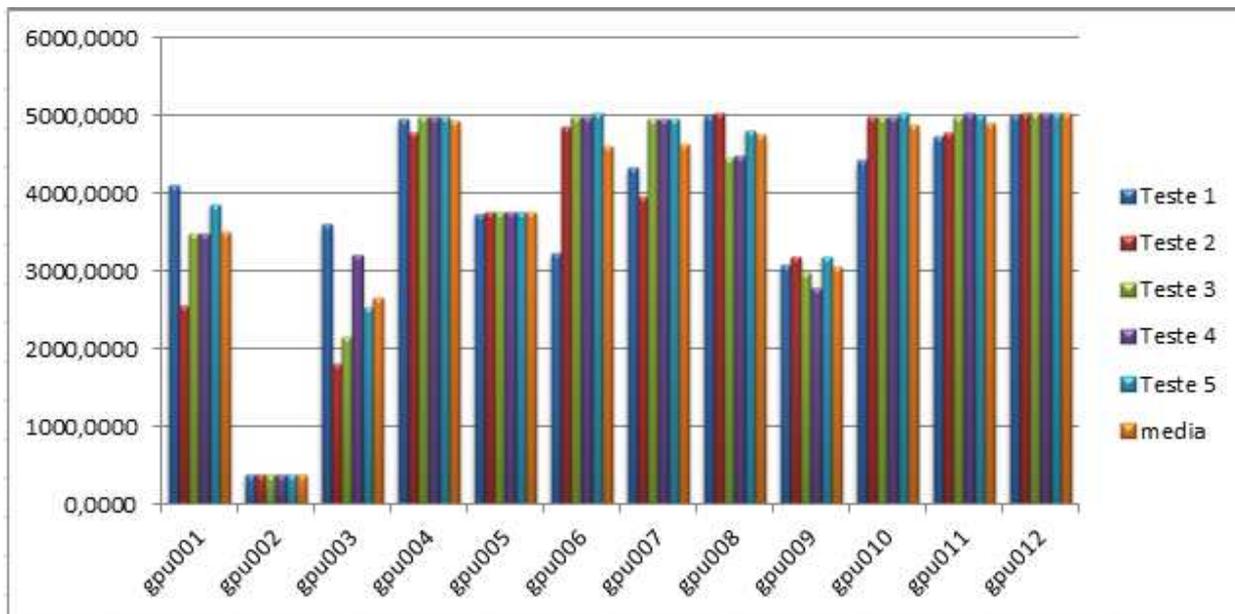


Figura 2 - Banda de memória "host to device" com memória paginavel do device 1 (Mbytes)

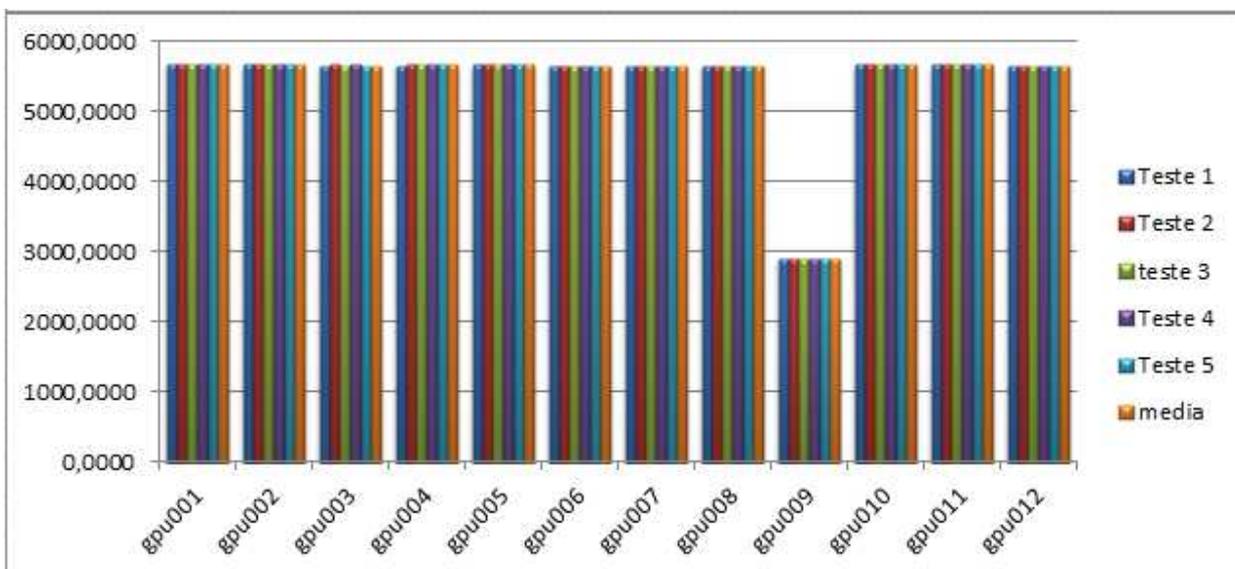


Figura 3 - Banda de memória "host to device" com memória page-locked do

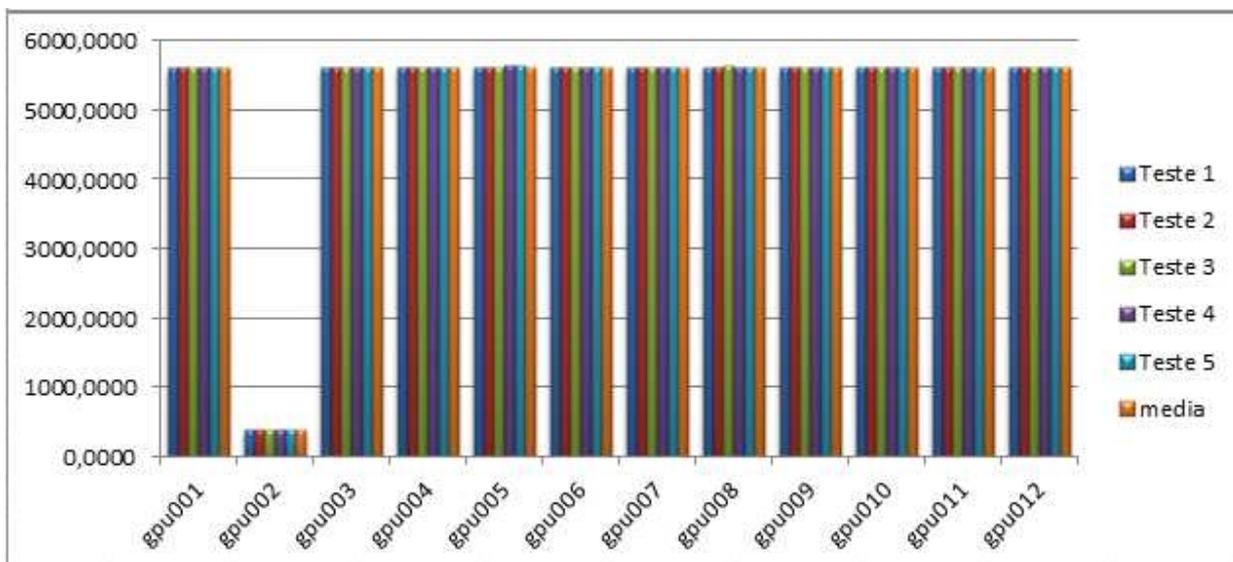
device 0 (MBytes)

Figura 4- Banda de memória "host to device" com memória page-locked do device 1 (Mbytes)

Conclusão

Com a realização dos testes foi possível obter os "benchmarks" do "cluster" avaliado, assim como a detecção e correção de problemas no "cluster", contribuindo para melhoria da arquitetura atual bem como a criação de uma base de dados para futuras comparações com diferentes arquiteturas e eventuais sugestões para melhorias do "cluster".

Assistente Virtual para Monitoramento e Auxílio em Atividades Fisioterápicas

Nome: Arthur Duarte da Silva Xavier

Orientador: Jauvane Cavalcante

Introdução

Ambientes Virtuais (AVs) permitem que usuários interajam com o sistema através da simulação de um mundo sintético controlado por computador.

Os AVs têm sido, historicamente, aplicados a diversas áreas do conhecimento, como treinamento em diversas áreas, procedimento médicos entre outros.

Alguns dispositivos de interação provenientes de jogos tem se mostrado bastante interessantes no contexto de interação em ambientes virtuais. O Microsoft Kinect que é um equipamento de reconhecimento e rastreamento do corpo humano é um deles.

O objetivo do projeto é desenvolver um sistema que auxilie profissionais na área de fisioterapia no monitoramento e análise dos pacientes que realizaram exercícios pré-formulados pelo profissional responsável.

A função do sistema é a de armazenar a execução de um exercício fisioterápico, considerando a posição, velocidade e aceleração das articulações. Uma vez gravado, o exercício realizado pelo paciente é analisado em relação ao exercício executado pelo profissional. Dessa análise são gerados relatórios, estes auxiliam o profissional a medir o desempenho do paciente na execução exercício.

Além disso, o sistema conta com um assistente virtual que guiará o paciente na execução do exercício.

Aplicação de Dispositivos de Interação para Simulação de Movimentos em Ambientes Virtuais

Orientador: Jauvane C. de Oliveira

Bolsista: Carina da Cruz Teixeira

Resumo:

O objetivo deste trabalho é desenvolver um sistema que simule a caminhada de um avatar em um ambiente virtual, seguindo os movimentos do usuário sobre o dispositivo de interação Wii Balance Board.

Uso de estimativas de precipitação por radar a bordo de satélite na região Serrana do Rio de Janeiro.

Aluna: Celso de Paiva Nogueira

Orientador: Pedro Leite da Silva Dias

Introdução

Após estudos do IPCC (Painel Intergovernamental de Mudanças Climáticas) surgiram claras evidências do aumento significativo da intensidade e frequência de fenômenos climáticos extremos em todo o globo terrestre trazendo enormes prejuízos econômicos e perdas humanas. No caso da precipitação, essas evidências são tênues devido à variabilidade climática natural associada a ciclos interanuais e decadais/multidecadais fazendo-se necessário explorar com maior profundidade as fontes de informações que permitem a estimativa da precipitação, principalmente em regiões onde a variabilidade espacial é muito grande como, por exemplo, a região serrana do Rio de Janeiro que foi particularmente afetada por chuvas intensas no dia 12 de Janeiro de 2011. A tragédia foi considerada como o maior desastre climático da história do país.

Teoria

Antes de se analisar a funcionalidade dos dados se viu a necessidade de entender de maneira mais específica os princípios básicos do funcionamento do estimador. Podendo assim estudar como serão feitas as estatísticas desses dados.

Tendo tais conhecimentos iniciou-se o estudo de ferramentas para criação de dados utilizando programas específicos: o GRADS e o MATLAB.

O **GrADS** é uma ferramenta computacional interativa para a análise e exibição de dados de ciência da Terra que é usado no globo todo e livremente disponível na internet. O software executa um modelo de dados de 4-Dimensões, onde as dimensões são geralmente latitude, longitude, nível e tempo.

O **Matlab (MATRIX LABORATORY)** é um software destinado a fazer cálculos com matrizes. Os comandos do MATLAB são muito próximos da forma como escrevemos expressões algébricas, tornando mais simples o seu uso.

Após serem organizados os dados e lidos, as informações relevantes ao estudo foram extraídas através do GRADS, podendo-se fazer as estatísticas dos mesmos através da leitura dos arquivos pelo MATLAB. Sendo assim elaborado um passo a passo para a correta manipulação dos dados.

Resultados Esperados

Pretende-se obter estimativas da probabilidade de ocorrência de precipitação em diferentes percentis. Em particular, deverão ser definidos os limiares associados aos extremos de precipitação (15, 10 e 5% superiores) e avaliada a variabilidade espacial desses limiares na amostra de cerca de 10 anos de dados. Devido a topografia tão variável da região serrana espera-se que a variabilidade dos limiares associados aos percentis seja muito significativo.

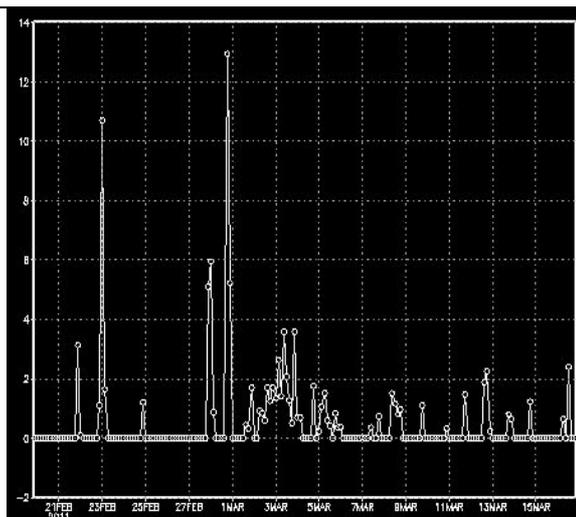
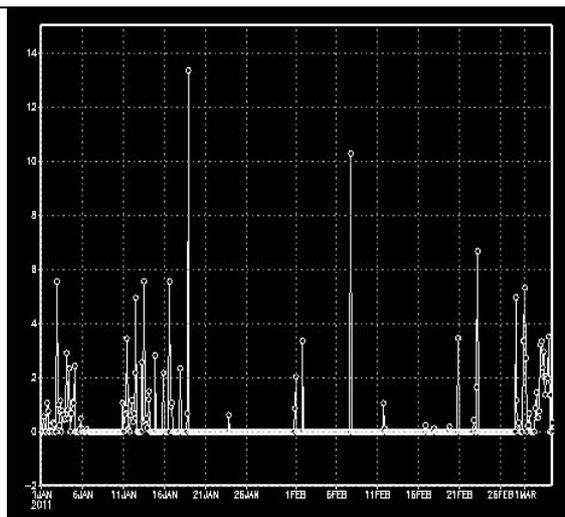
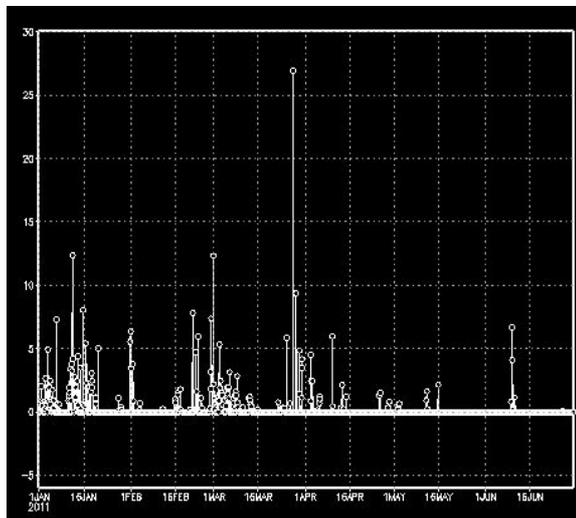
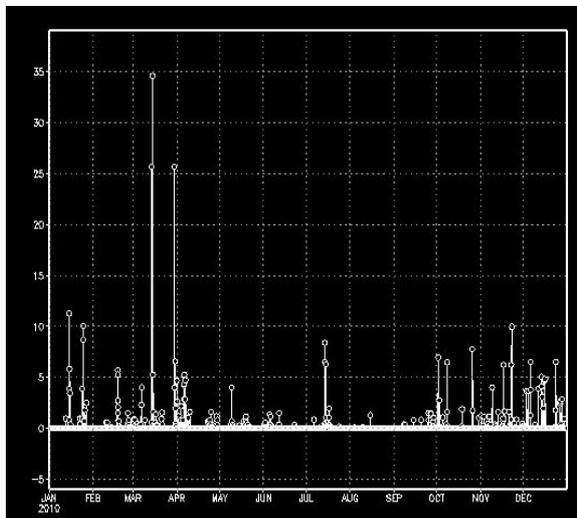
Os resultados desse trabalho serão comparados com outros projetos de estimativas de precipitação.

Conclusão

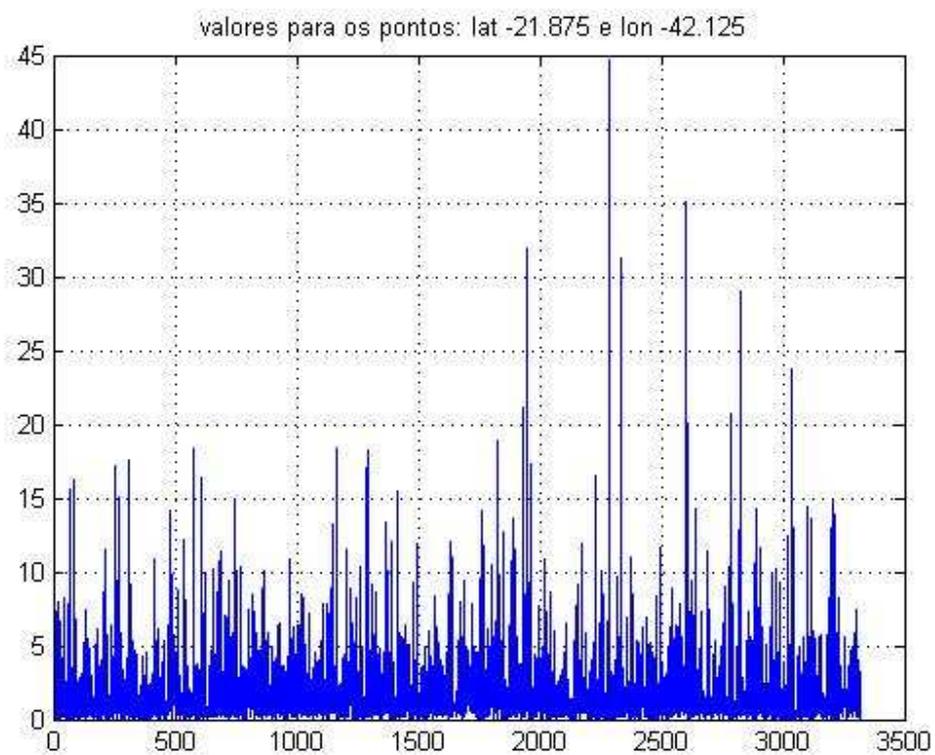
Espera-se através da análise das informações obtidas, poder realizar um estudo que possa ajudar a prever e evitar desastres como o corrido na região serrana do Rio de Janeiro, tendo como base a ocorrência dos extremos de precipitação e as suas variabilidades espaciais.

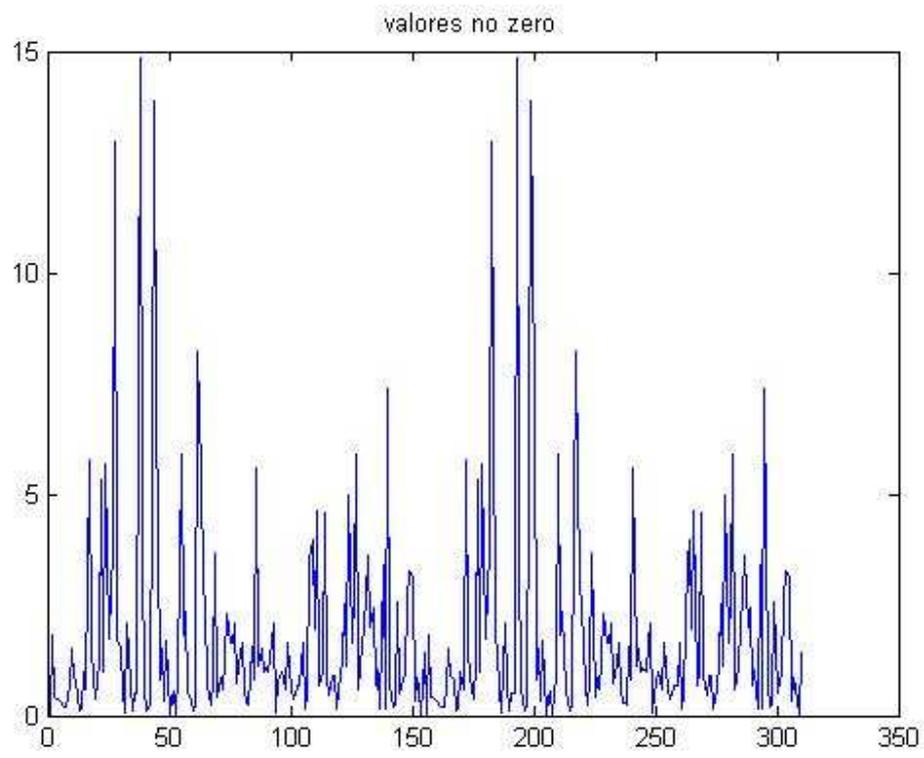
IMAGENS

GRADS



MATLAB





Análise Computacional de Fatores e sua Relação com Evasão Escolar

Bolsista: Dejamile Rocha da Silva

Orientador: Dr. José Karam Filho

Co-orientador: Dr. Fabiano Saldanha Gomes de Oliveira

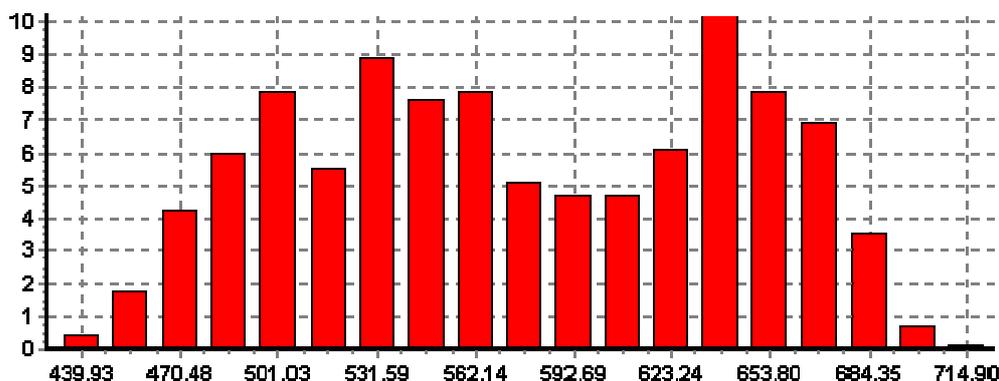
Relatório

O presente relatório tem como objetivo o desenvolvimento de metodologias quantitativas que permitam compreender de uma maneira inovadora a fenomenologia social que atua nos processos educacionais, utilizando Métodos Numéricos e Princípios Variacionais. Implementando métodos inteligentes com a calibragem dos parâmetros dos modelos e rotinas matemáticas, atribuindo dados reais, geração e análise dos resultados.

O Princípio Variacional é utilizado com o cálculo de variações as quais desenvolvem métodos para encontrar funções que minimizem ou maximizem os valores de quantidades que dependam de tais funções.

Simulação Estocástica

A simulação estocástica é um mecanismo de análise para avaliar modelos não-determinísticos. Estes modelos, também chamados de modelos estocásticos, permitem a representação de variáveis de incerteza e seus valores. Devido a estas variáveis de incerteza, toda vez que um modelo não-determinístico é avaliado, mesmo com os mesmos parâmetros fornecidos, seu comportamento pode mudar.



Métodos de simulação estocástica são procedimentos que envolvem a geração de números aleatórios (pseudo-aleatórios) com o objetivo de explorar o espaço de incerteza ou campo de possibilidades de um dado fenômeno físico ou qualquer outro tipo de variável de estudo cujo comportamento possa ser quantificado matematicamente. Os métodos de simulação fazem parte da ciência de processos estocásticos e utilizam o método de Monte-Carlo nos seus algoritmos.

A simulação de Monte Carlo, por exemplo, é um mecanismo que permite a determinação da função de distribuição de probabilidade de resultados de um modelo, baseada na distribuição de parâmetros do modelo. Isto envolve avaliar o mesmo modelo centenas ou milhares de vezes, gerando a cada vez um número aleatório para cada parâmetro descrito como uma distribuição. Cada avaliação calcula um valor diferente para cada resultado do modelo. Após diversas execuções, estes resultados podem ser organizados em histogramas de frequência, que são representações gráficas do número de vezes que cada valor foi resultante em uma simulação. Normalizado, de forma que a área sob a curva se reduza a 1, o histograma pode ser visto como a função de distribuição de probabilidade dos resultados do modelo.

Uma forma de descrever uma variável de incerteza é definir seu valor como uma função de distribuição de probabilidade, ao invés de um valor numérico determinístico. A incerteza expressa no parâmetro do modelo é propagada através das equações do modelo, afetando os resultados. Como os parâmetros do modelo, os resultados também são descritos por funções de distribuição de probabilidade, ao invés de um número fixo.

O projeto em questão foi desenvolvido utilizando o MatLab em conjunto com as Funções de Densidade de Probabilidade (Probability Density Function-PDF). O uso de

funções densidade de probabilidade está diretamente ligado à natureza dos dados a que ela se relaciona. Algumas têm boa capacidade de estimação para pequeno número de dados, outras requerem grande série de observações. Devido ao número de parâmetros de sua equação, algumas podem assumir diferentes formas, enquadrando-se em um número maior de situações, ou seja, são mais flexíveis. Desde que respeitado o aspecto da representatividade dos dados, as estimativas dos seus parâmetros, para uma determinada região, podem ser estabelecidas como de uso geral, sem prejuízo da precisão na estimação da probabilidade. Foi criado um algoritmo que engloba cinco funções de densidade de probabilidade citadas abaixo:

- Normpdf
- Binopdf
- Exppdf
- Lognpdf
- Geopdf

REFERÊNCIAS:

- [1]. Revista Rio Pesquisa Ano III, Nº. 9, publicada pela FAPERJ em Dezembro de 2009.
- [2] Jordan, M.I.; "Learning in Graphical Models". MIT Press 1998.
- [3] Pearl, J.; "Causality". Cambridge Press 2000.
- [4] McLachlan, G.J.; Krishnan T.; "The EM Algorithm and Extensions", Wiley Series in Probability and Statistics 1997.
- [5] Lauritzen, S.; "Graphical Models". Oxford 1996.
- [6] Oliveira, F. & Veiga, A.; "A EM Algorithm to Multiple Imputation in Censored Multivariate Data". Artigo aceito no ISF 2002.

[7] Shumway, R.H.; Stoffer, D.S.; “An Approach to the Time Series Smoothing and Forecasting using the EM Algorithm”, *Journal of Time Series Analysis* Vol. 3 No. 4 1982.

[8] Dempster, A. P.; Laird, N. M.; Rubin, D. B.; “Maximum Likelihood from Incomplete Data via The Em Algorithm”, *Journal of the Royal Statistical Society B* 39 1977.

DESENVOLVIMENTO DE ROTINAS DE ANÁLISE INTELIGENTE DE DADOS MULTIVARIADOS PARA MINERAÇÃO DE DADOS

Aluno: Elias Junior Cescon

Orientador: Dr. Paulo Cesar Marques Vieira

Co-orientador: Dr. Fabiano Gomes

RESUMO

Este projeto prevê o desenvolvimento de rotinas da Análise de Dados Multivariados (ADM), para extração de interdependência entre as variáveis que compõem a matriz de dados de um problema multivariado de dimensão grande. Os dados usados neste estudo possuem a ordem de grandeza de dezenas de variáveis independentes, ou matriz de grandes dimensões.

Para que este procedimento de mineração de dados funcione, serão usados dados de consumo de energia elétrica, agregados aos métodos numéricos da ADM: análise de componentes principais, análise de fatores, análise de agrupamento e análise de variância.

Ao longo do trabalho foi necessário desenvolver os conhecimentos em programação, estatística, métodos numéricos e ADM.

Foi utilizado o ambiente de programação MatLab para desenvolver algoritmos capazes de tratar eficientemente dados de origem econômica e extrair dos mesmos informações úteis para a tomada de decisão.

Estes algoritmos permitirão o trabalho com Modelos Multivariados, como alguns dos mais comuns: Mistura de Gaussianas, Análise de Fatores, Análise de Agrupamento, Sistemas Dinâmicos (lineares ou não), Modelos a espaço de estado, Simulação de Eventos em Sistemas de Informação, Algoritmos Inteligentes (aprendizado), Estruturas incompletas (fazendo uso do algoritmo EM), Algoritmos de

pesquisa, Algoritmos da Teoria de Decisão, Modelos Lineares Generalizados, Modelos Hierárquicos.

A rotina desenvolvida avaliará os dados de entrada quanto a sua variância e aplicando o modelo de regressão linear, buscar estimar um valor esperado para esta amostra. O resultado desejado é uma rotina capaz de informar qual o conjunto de dados melhor representa a variância total da amostra.

REFERÊNCIAS

- [1]. Revista Rio Pesquisa Ano III, No. 9, publicada pela FAPERJ em Dezembro de 2009.
- [2] Jordan, M.I.; "Learning in Graphical Models". MIT Press 1998.
- [3] Pearl, J.; "Causality". Cambridge Press 2000.
- [4] McLachlan, G.J.; Krishnan T.; "The EM Algorithm and Extensions", Wiley Series in Probability and Statistics 1997.
- [5] Lauritzen, S.; "Graphical Models". Oxford 1996.
- [6] Oliveira, F. & Veiga, A.; "A EM Algorithm to Multiple Imputation in Censored Multivariate Data". Artigo aceito no ISF 2002.
- [7] Shumway, R.H.; Stoffer, D.S.; "An Approach to the Time Series Smoothing and Forecasting using the EM Algorithm", Journal of Time Series Analysis Vol. 3 No. 4 1982.
- [8] Dempster, A. P.; Laird, N. M.; Rubin, D. B.; "Maximum Likelihood from Incomplete Data via The Em Algorithm", Journal of the Royal Statistical Society B 39 1977.

SISTEMA INTELIGENTE DE ANÁLISE DE DADOS

Aluno: Fabio Magest

Orientador: Dr. Paulo Cesar Marques Vieira

Co-orientador: Dr. Fabiano Gomes

1. RELATÓRIO

Meus estudos são baseados nos modelos estatísticos conhecidos como gráficos ou Redes Bayesianas, eles proporcionam o gerenciamento da incerteza em sistemas de computação, unificam a maior parte da literatura de modelos estocásticos multidimensionais. Estes modelos gráficos geram soluções para os problemas de diagnóstico, previsão e controle, em alguns tipos de redes. Os modelos gráficos do tipo orientados são as conhecidas Redes Bayesianas ou Redes de Decisão, eles são usados em aplicações de economia, modelagem de sistemas sociais e de computação, já os modelos não orientados são usados em aplicações de inteligência artificial e estatística computacional.

A ausência de dados em sistemas e informação é tratada através de estimativas de máxima verossimilhança de parâmetros e do cálculo dos intervalos de confiança destas estimativas. O algoritmo EM proposto por Dempster, Laird e Rubin, é um método que calcula a partir das observações de algumas variáveis existentes no sistema as estimativas dos valores que faltam no mesmo. Ele substitui estes os valores ausentes pelas estimativas, ou seja, ele pode ser usado, em grandes massas de dados contendo valores não observados, para a recuperação da informação ausente. A utilização do algoritmo EM com as Redes Bayesianas amplia a gama de problemas onde esta ferramenta pode ser aplicada.

Descrição de algumas das características e aplicações dos modelos gráficos:

1. Construção de Motores de Inferência probabilísticos para Sistemas de Banco de Dados.
2. Análise Top-down e Botton-up da causalidade.
3. Modelagem matemática da causalidade.
4. Independência condicional permitindo a transformação de grafos orientados em não orientados através da inserção de nós auxiliares
5. Modelos temporais.
6. Modelos Multivariados.

Alguns modelos comumente usados em redes são casos particulares de modelos gráficos. Por exemplo:

- Mistura de Gaussianas;
- Análise de Fatores;
- Análise de Agrupamento;
- Sistemas Dinâmicos (lineares ou não);
- Modelos a espaço de estado;
- Simulação de Eventos em Sistemas de Informação;
- Algoritmos Inteligentes (aprendizado);
- Estruturas incompletas (fazendo uso do algoritmo EM);
- Algoritmos de pesquisa;
- Algoritmos da Teoria de Decisão;
- Modelos Lineares Generalizados;
- Modelos Hierárquicos.

2. APLICAÇÃO

A aplicação se dá ao criar um pequeno algoritmo no sistema MatLab que permite usar os métodos descritos acima usando como parâmetros de entrada um banco de dados que contem as informações de uma companhia de energia elétrica. Neste banco temos as informações de consumo no intervalo de tempo de dois anos de 24.000 clientes. Será realizada uma análise nas informações do intervalo de um ano de consumo dos clientes. De posse dessas informações será criada uma estimativa de dados usando tais informações como base de treinamento. Após a criação dessa estimativa ela será confrontada com os dados reais existentes no banco de dados relacionadas ao segundo ano de consumo dos clientes.

3. REFERÊNCIAS:

- [1]. Revista Rio Pesquisa Ano III, Nº. 9, publicada pela FAPERJ em Dezembro de 2009.
- [2] Jordan, M.I.; “Learning in Graphical Models”. MIT Press 1998.
- [3] Pearl, J.; “Causality”. Cambridge Press 2000.
- [4] McLachlan, G.J.; Krishnan T.; “The EM Algorithm and Extensions”, Wiley Series in Probability and Statistics 1997.
- [5] Lauritzen, S.; “Graphical Models”. Oxford 1996.
- [6] Oliveira, F. & Veiga, A.; “A EM Algorithm to Multiple Imputation in Censored Multivariate Data”. Artigo aceito no ISF 2002.
- [7] Shumway, R.H.; Stoffer, D.S.; “An Approach to the Time Series Smoothing and Forecasting using the EM Algorithm”, Journal of Time Series Analysis Vol. 3 No. 4 1982.

[8] Dempster, A. P.; Laird, N. M.; Rubin, D. B.; “Maximum Likelihood from Incomplete Data via The Em Algorithm”, Journal of the Royal Statistical Society B 39 1977.

Modelagem de Incertezas em Problemas de Condução de Calor

Gabriel C. Amorim

Universidade Católica de Petrópolis

Orientadores: Regina C. Almeida, Renato S. Silva

Laboratório Nacional de Computação Científica

Resumo

As simulações numéricas de problemas de engenharia vêm acompanhadas de incertezas e erros. Estas características são muitas vezes agravadas pela precariedade ou ausência das informações necessárias para caracterizar a fenomenologia.

O tratamento de incertezas envolve três etapas essenciais:

- obtenção e tratamento das informações relevantes;
- formulação do modelo matemático que representa o fenômeno que se deseja modelar levando em consideração as informações disponíveis;
- tratamento numérico do modelo matemático.

As incertezas envolvidas em simulações numéricas são principalmente de dois tipos: aleatória e epistêmica. A incerteza aleatória ocorre devido a variações naturais como, por exemplo, a variação da condutividade térmica dos materiais devido a variação da composição química dos mesmos. Já a incerteza epistêmica ocorre devido à falta parcial ou total de conhecimento sobre o comportamento de determinado parâmetro ou condição inicial.

Na fase inicial da simulação é comum a indisponibilidade de dados como propriedades do meio, condições de contorno e iniciais. Nestes casos, quando não são conhecidas informações suficientes ou quando nenhuma informação está disponível, a análise não probabilística de incertezas se torna necessária.

Neste projeto de iniciação científica, a partir de um problema modelo simples de transferência de calor, propõe-se estudar típicas fontes de incerteza e seus impactos na simulação computacional do problema modelo. São analisadas as incertezas associadas ao armazenamento de energia solar. O armazenamento de energia solar é de fundamental importância para que usinas solares possam fornecer energia durante períodos com a ausência do sol. Existem várias tecnologias em fase de testes. Neste

trabalho serão analisadas as incertezas da tecnologia que utiliza um sólido como reservatório de calor. O sólido é o concreto, utilizado por razões econômicas, cujas propriedades físicas como calor específico e condutividade térmica variam com a composição usada na sua elaboração. Assim, este trabalho visa a modelagem numérica do processo de condução de calor e o seu armazenamento via o método de diferenças finitas, em uma seção de um bloco feito de concreto, onde as incertezas nos parâmetros calor específico e condutividade térmica serão consideradas.

Modelagem Numérica de Mecânica dos Fluidos

Bolsista: Jeozadaque Marcos da Silva

Orientador: Jiang Zhu

Introdução

O programa deverá possibilitar ao estudante: elaborar e analisar modelos matemáticos em mecânica dos fluidos e aprender métodos numéricos para resolvê-los.

Metodologia

A metodologia será baseada em estudo dirigido, resolução de exercícios, seminários e reuniões regulares com o orientador e participação como aluno ouvinte nos cursos de extensão do LNCC.

Os seguintes temas serão estudados: Cálculo Avançado, Álgebra Linear Avançada, Métodos Numéricos, Linguagens de Programação.

Cronograma

1º semestre: serão aprendidas técnicas básicas de modelagem, Cálculo Avançado e Álgebra Linear Avançada.

2º semestre: aprender técnicas de Métodos Numéricos. Estudar as linguagens de computação como o Fortran, C++ ou Matlab

Bibliografia

- T.L. Saaty & J.M. Alexander - **Thinking with Models - Mathematical Models in Physical, Biological and Social Sciences**, Pergamon Press, 1981
- Axelsson, O. **Iterative Solution Methods**. New York: Cambridge University Press, 1996
- Young, D. M. **Iterative solution of large linear systems**. New York Academic Press, 1971.
- H. Schildt. **C completo e total**, Makron Books, 1997
- S. Chapman, **Fortran 90/95 for scientists and engineers**, McGraw Hill, 1998
- G. Bartle; Donald R. Sherbert, John Wiley, **Introduction to Real Analysis** -New York, 1982
- Lima, Elon Lages **Curso de Análise** -, vol. I e II.
- Hoffman, Kenneth; Kunze, Ray. **Linear Algebra**. Prentice-Hall, 197

Sistema para Análise e Simulação de Fluidos via Lattice Boltzmann

José Guilherme Mayworm, Sicilia F. Judice e
Gilson A. Giraldi

Laboratório Nacional de Computação Científica (LNCC/MCT)

Petrópolis, RJ, Brasil

{jmayworm, sicilia, gilson}@lncc.br

Resumo. Neste trabalho é apresentado um sistema que integra técnicas de simulação de fluidos via Lattice Boltzmann com técnicas de visualização científica. O método Lattice Boltzmann baseia-se na idéia fundamental de construir modelos cinéticos simplificados que incorporam aspectos relevantes da física de processos microscópicos de forma que as propriedades macroscópicas satisfaçam as equações de fluidos. Os campos gerados na simulação necessitam da aplicação de técnicas de visualização para análise. Neste trabalho é apresentada uma breve descrição do software desenvolvido bem como dois exemplos de simulação de fluidos utilizando o método proposto.

1. Introdução

Nas últimas décadas, técnicas baseadas em modelos físicos para animação de elementos naturais, como fluidos (gás ou líquido), têm chamado a atenção dos pesquisadores em computação gráfica. A motivação reside no potencial existente nas aplicações dessas técnicas bem como na complexidade e beleza dos fenômenos naturais envolvidos. Em particular, técnicas no campo da Dinâmica de Fluidos Computacional (DFC) têm sido implementadas para animação de fluidos em aplicações que envolvem interações fluido-fluido e fluido-superfície [1], simuladores de cirurgia virtuais [2], efeitos visuais para a indústria cinematográfica [3] e jogos digitais [4].

Os métodos tradicionais de animação de fluidos em computação gráfica utilizam abordagens baseadas em malhas 2D/3D, motivadas pelos métodos eulerianos de Elementos Finitos (EF) e Diferenças Finitas (DF) em conjunto com as equações de Navier-

Stokes para fluidos [5]. Ao final do pipeline de animação, técnicas de rendering precisam ser aplicadas para garantir o nível de realismo desejado nas imagens geradas.

Alternativamente, métodos baseados em malha, tais como *Lattice Gas Cellular Automata* (LGCA) e *Lattice Boltzmann Method* (LBM), podem ser usados. A idéia básica por trás destes métodos consiste em obter a dinâmica macroscópica de um fluido através do comportamento microscópico de diversas partículas. O LGCA simplifica a dinâmica através de regras simples e locais para a interação das partículas. Em contrapartida, o LBM constrói um modelo cinético, uma simplificação da equação de *Boltzmann*, a qual incorpora a física microscópica essencial de modo que as quantidades macroscópicas estimadas obedecem as equações de interesse [6].

O modelo LBM foi desenvolvido como uma técnica numérica alternativa e promissora para simulação de fluidos. Este método propõe substituir os dados booleanos utilizados nos modelos LGCA por números reais entre 0 e 1, representando probabilidade de presença de partículas em determinada direção de movimento. O método LBM pode ser usado para simular modelos 2D ou 3D.

O presente trabalho apresenta um software, o qual incorpora um simulador de LBM 3D com capacidade de visualização dos campos gerados, através de técnicas simples como mapas de cor e setas orientadas, disponíveis na biblioteca OpenGL, além da implementação de interfaces gráficas usando a biblioteca GLUT, que permitem a configuração e controle das simulações.

O texto a seguir é organizado da seguinte forma. A seção 2 descreve o método Lattice Boltzmann e explica brevemente o modelo D3Q19. A seção 3 apresenta o *software* desenvolvido. Na seção 4 são apresentados os resultados e finalmente na seção 5 a conclusão.

2. O Método Lattice Boltzmann

Nos últimos anos, o método *Lattice Boltzmann* tem chamado a atenção da comunidade científica devido à sua facilidade de implementação e eficiência computacional. Especificamente na área de dinâmica de fluidos, o LBM tem sido utilizado pela facilidade de implementação de condições de fronteira e pela estabilidade numérica [7]. Este originou-se do *Lattice Gas Cellular Automata* (LGCA) o qual, apesar de suas vantagens, apresenta certas limitações pela sua natureza discreta. O LBM foi introduzido

por [8], onde os autores mostram a vantagem de se estender a dinâmica booleana dos autômatos celulares para se trabalhar diretamente com valores numéricos reais representando probabilidades de presença de partículas.

No LBM, o domínio de interesse é discretizado em uma malha e o fluido é considerado uma coleção de partículas. Tais partículas se movem em intervalos discretos de tempo, com uma velocidade apontando ao longo de uma das direções da malha. Além disso, partículas colidem uma com as outras e quantidades físicas de interesse associadas com os nós da malha são atualizadas em cada instante de tempo. O cálculo em cada nó é dependente das propriedades inerentes ao próprio nó bem como aos nós vizinhos no instante de tempo anterior [7, 6].

A dinâmica deste método é governada pela equação de *Boltzmann*:

$$f_i(\vec{x} + \Delta x \vec{c}_i, t + \Delta t) - f_i(\vec{x}, t) = \Omega_i(f), i = 1, \dots, z, \quad (1)$$

onde f_i é a função de distribuição, \vec{x} é o nó da malha, \vec{c}_i é uma das direções, Δx é o espaçamento da malha, Δt é o passo de tempo, $\Omega_i(f)$ é o termo de colisão, e z é o número de direções da malha.

Diversificação funcional da álcool desidrogenase em animais e fungos

JB Maciel¹, ATR Vasconcelos¹, CE Thompson¹

¹ Laboratório de Bioinformática, Laboratório Nacional de Computação Científica, Rio de Janeiro

A superfamília das desidrogenases e redutases de cadeia média (MDR) é formada por proteínas com cerca de 350 resíduos de aminoácidos que inclui *quinone reductases*, *leukotriene B4 dehydrogenases*, *polyol dehydrogenases*, *Zn²⁺ dependent ADHs* (álcool desidrogenases). Tais proteínas possuem várias atividades enzimáticas diferentes e podem ser encontradas em todos os reinos. Atualmente, as maiores famílias de *MDRs* são as *ADHs* e a *cinnamyl alcohol dehydrogenases (CAD)*.

A enzima álcool desidrogenase (ADH) é produzida a partir de uma família de genes de mesmo nome. Esses genes codificam enzimas que metabolizam uma variedade de substratos, incluindo, etanol, retinol, outros alcoóis alifáticos e hidroxiesteróides. O conjunto das ADHs apresenta uma grande complexidade genética e funcional. Essas enzimas variam bastante em suas propriedades farmacocinéticas. As várias formas da enzima ADH são divididas em classes distintas, de acordo com as subunidades, composição de suas isoenzimas e suas propriedades físico-químicas. Há evidência de que a álcool desidrogenase desempenha funções distintas em mamíferos, tendo sido relacionada ao metabolismo da norepinefrina, dopamina, serotonina e ácido biliar. Adicionalmente, essa enzima pode catalisar a oxidação do retinol *in vitro* e *in vivo*. Com o objetivo de futuramente avaliar o impacto da diversificação funcional na estrutura tridimensional deste grupo de proteínas, a técnica de Modelagem Molecular Comparativa por Homologia foi usada para construção de modelos tridimensionais.

O processo para a obtenção do modelo proteico 3D através da modelagem molecular comparativa por homologia envolve algumas etapas, entre elas: busca de proteínas homólogas, alinhamento das sequências, construção e otimização dos modelos e a validação dos mesmos. O primeiro passo na modelagem é a identificação de estruturas conhecidas que possam atuar como estruturas molde. A seleção destas estruturas foi realizada levando em consideração alguns aspectos como a identidade de no mínimo 25% e a disponibilidade de estruturas 3D nos bancos de dados. Tais sequências foram obtidas do *National Center of Biotechnology Information* (NCBI) no formato FASTA. A procura pelos moldes foi realizada através do software BLASTp do NCBI. Para realizar os alinhamentos múltiplos entre os possíveis moldes e a proteína a ser modelada foi utilizado o CLUSTALW. As correções manuais foram realizadas com o software GENEDOC. Após a identificação do melhor molde, foi utilizado o MODELLER para gerar 10 modelos tridimensionais, que foram visualizados através do programa *Swiss Protein Data Bank Viewer* (SPDBV). Para verificar a qualidade estereoquímica dos modelos foi utilizado o PROCHECK. Já o VERIFY-3D foi importante na avaliação do ambiente químico de cada resíduo de aminoácido no modelo.

A sequência 1CDO mostrou uma elevada identidade com a adh1DANRER. Posteriormente, foi realizado o alinhamento global entre as sequências primárias, da proteína-molde e da proteínas-alvo, que permitiu a identificação de regiões conservadas. Os modelos foram comparados entre si e submetidos à validação, passo que ainda encontra-se em andamento. Tais avaliações são necessárias e úteis, pois definem os resíduos que se encontram nas regiões energeticamente mais favoráveis e desfavoráveis, além de orientar a avaliação da qualidade para modelos teóricos e experimentais de proteínas. Através desses resultados, podemos determinar o melhor modelo. Ao todo, 15

proteínas foram modeladas com o método de modelagem comparativa por homologia e parcialmente analisadas. Alguns modelos foram selecionados para exibição.

Desenvolvimento de um Atlas do Corpo Humano utilizando WebGL.

Bolsista: Leonardo Cardia da Cruz

Orientador: Jauvane C. de Oliveira

RESUMO:

O intuito do trabalho é desenvolver um atlas do corpo humano totalmente em 3D sem a necessidade de plugins para visualização no navegador utilizando a API multi-plataforma WebGL. Desta forma, será possível observar cada parte do corpo humano com descrições e detalhes, sem que o usuário procure por livros de anatomia.

Otimização por Enxame de Partículas (PSO)

Luís Henrique Carvalho do Nascimento, Eduardo Krempser, Helio J. C. Barbosa

{luishcn, krempser, hjcbm}@lncc.br

O método de otimização conhecido como Otimização por Enxame de Partículas, aqui abreviado por PSO, do inglês Particle Swarm Optimization, criado por Kennedy e Eberhart, cria uma população de partículas que se movem em torno do espaço de busca. Tais partículas se movem com base no conhecimento adquirido pelo enxame (a melhor partícula do enxame), e por experiências próprias.

O PSO foi inspirado no comportamento de um bando de pássaros na busca de alimentos, onde os pássaros correspondem as partículas e os alimentos ao ponto ótimo.

Observando o bando notou-se que as partículas sofrem influências

pelas experiências individuais acumuladas

e também pelo resultado acumulado pelo grupo. Ou seja,

a partícula armazena o lugar onde ela obteve o seu melhor resultado

e também o melhor lugar encontrado por todo o enxame. Com essas experiências

a partícula calcula sua nova velocidade e assim conhece sua nova posição.

A grande vantagem de usar um método de otimização como o PSO é que ele não necessita de que o problema a ser otimizado se apresente de forma explícita, além de não utilizar o gradiente do problema, o que se torna bastante útil quando não se tem sua formulação ou até mesmo impossível de ser derivado. Fazendo assim, com que o PSO possa ser empregado a vários problemas de otimização.

Porém o uso de parâmetros influencia no comportamento do algoritmo, sendo assim, esse trabalho visa a avaliação de um conjunto adequado de parâmetros para a sua aplicação.

Dockthor: Comparative Analysis of Receptor-Ligand Docking Performance

Vizani, L. A.¹; Guedes, I. A.¹; Marinho, D.¹, De Magalhães, C.S.² and Dardenne, L. E.¹

Grupo de Modelagem Molecular de Sistemas Biológicos, LNCC, RJ, Brazil

¹ **Laboratório Nacional de Computação Científica - LNCC/MCTI**

² **Universidade Federal Rural do Rio de Janeiro - UFRRJ**

The protein-ligand docking methodology is an important tool for structure-based rational drug design studies. Dockthor is a grid based docking program, developed by our research group, and was implemented to deal with highly flexible ligands using a multiple-solution steady state genetic algorithm and a scoring function based on the MMFF94 force field. In this work, the Dockthor performance was evaluated through a comparative analysis of redocking experiments with Glide, GOLD and Autodock Vina docking programs. This study was carried out using a high quality test set of 115 protein-ligand complexes. The overall test set covers a wide range of different protein families (including relevant drug or agrochemicals targets like kinases, phosphodiesterases, nuclear receptors) and ligands with distinct chemical properties. The test set consists in 5 complexes from HIV-Protease with a high degree of freedom plus 35 complexes from a *in house* test set plus 75 of the 85 selected by Astex Therapeutics, Ltd

for a study of validation of docking programs. For each complex, the programs that consists in a deterministic method (Glide and Vina) were executed one time using their best protocol, and those which consists in stochastic method (Dockthor and Gold) were executed several times (30 runs) using their best docking protocol. In the case of Gold, each function (GoldScore, ChemScore and ChemPLP) were executed 30 times. The best ligand energy solution obtained from each program was compared with the experimental conformation. A RMSD value $\leq 2.5 \text{ \AA}$ from the experimental structure was considered as a success criterion. Dockthor obtained a success rate of 89,6%, showing a very good performance when compared to the programs Glide (86,1%), GOLD (ChemScore - 88,7%, GoldScore - 87,8%, ChemPLP - 86,1%) and Autodock Vina (75,6%). The program Dockthor obtained an excellent performance for highly flexible ligands (HIV-Protease) in comparison with the others programs. We conclude with the results obtained that Dockthor has a great potential to be widely used in real receptor-ligand studies.

**Keywords: Docking, Molecular Modeling, Dockthor.
Supported by: FAPERJ, CNPq and CAPES**

SISTEMA INTELIGENTE DE GESTÃO FAZENDÁRIA IMPLEMENTAÇÃO DE UM MÓDULO DE TOMADA DE DECISÃO USANDO REDES BAYESIANAS

Bolsista: Maria Luiza Mondelli

Orientador: Dr. Paulo Cesar Marques Vieira

Co-orientador: Dr. Fabiano Gomes

Introdução

O presente trabalho tem como objetivo o desenvolvimento de metodologias que associem estruturas de dados, algoritmos de inteligência artificial e análises de dados, possibilitando a construção de módulos capazes de analisar situações complexas e tomar decisões de forma autônoma. Como este trabalho envolve o desenvolvimento de algoritmos de prospecção de dados, que consiste no processo de explorar grandes quantidades de [dados](#) à procura de padrões consistentes, a fim de detectar relacionamentos entre [variáveis](#), o ambiente de desenvolvimento que será utilizado é o MatLab. Nele será criado um protótipo de toolbox para análise e tomada de decisão baseado em redes bayesianas, modelos estatísticos conhecidos como gráficos, que proporcionam formalismo para o gerenciamento da incerteza em sistemas de computação, unificando a maior parte da literatura de modelos estocásticos multidimensionais. Posteriormente as rotinas criadas serão convertidas em programas executáveis.

Redes Bayesianas

Os modelos gráficos são representados por grafos nos quais os nós são variáveis aleatórias e os arcos representam a dependência entre estas variáveis. Dessa forma, a

ausência de arcos está ligada ao conceito de independência das hipóteses. Os modelos gráficos do tipo orientados são as conhecidas Redes Bayesianas ou Redes de Decisão e combinam a teoria de grafos com a estatística Bayesiana e também a distribuição de probabilidades, a fim de representar uma situação, suas variáveis e estados e a partir disso ser capaz de realizar inferências. A construção da estrutura do grafo é guiada pelo conceito de causalidade.

Uma rede Bayesiana é aquela em que as regras de inferência do modelo se baseiam na regra de Bayes, que representa a relação entre a probabilidade condicional, ou seja, a probabilidade de um evento dado a ocorrência de outro evento, e a sua inversa. Baseado nesta regra, a determinação da distribuição a posteriori da estrutura nos informa qual a possibilidade de um evento ocorrer ou não para cada combinação de eventos anteriormente ocorridos. Seu modelo não precisa estar comprometido com a estatística Bayesiana. A rede consiste em:

- Conjunto de variáveis e conjunto de arcos ligando essas variáveis;
- Cada variável possui um conjunto finito de estados mutuamente exclusivos;
- As variáveis e arcos formam um grafo acíclico direcionado (GAD). Os arcos não podem ligar a saída de um nó a sua própria entrada.
- Para cada variável A com pais B_1, \dots, B_n , existe um tabela $P(A | B_1, \dots, B_n)$.

O propósito da construção de uma Rede Bayesiana é oferecer estimativas de certezas para eventos não observáveis e a primeira coisa a ser feita na construção desses modelos é a identificação dos eventos. É necessário que as propriedades da rede mantenham o domínio modelado, possibilitando então uma estrutura que modela um processo de decisão, através da escolha dos caminhos no grafo que dependem do resultado dos testes.

Tratamento de Dados e Modelo

Como parte do estudo e início da construção e implementação de um modelo de rede bayesiana, foi obtida uma base de dados com aproximadamente 23 mil registros contendo

informações da medição de contas de luz realizadas durante 2 anos e 3 meses, com dados como os códigos do cliente e localidade, número do lote, tipo de ligação, a data da baixa da medição, o tipo de seleção e se foi detectado fraude ou não. Para que fosse possível o estudo desses dados, todos os que estavam em forma de texto foram substituídos por números e no caso de ausência de dados, pelo fato de não ter sido feita medição, os espaços vazios foram substituídos por números negativos. Feito isso, a base de dados foi ordenada pelo código do cliente e foi dividida em três arquivos: o primeiro com todos os registros, o segundo com 80% dos registros, totalizando aproximadamente 18 mil registros, e o terceiro com 20%, totalizando pouco mais de 4 mil registros.

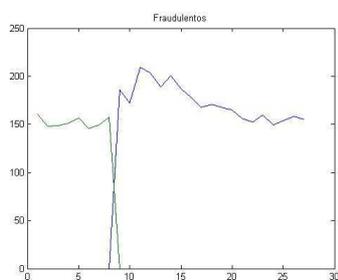


Figura 1. Detecção de fraudes

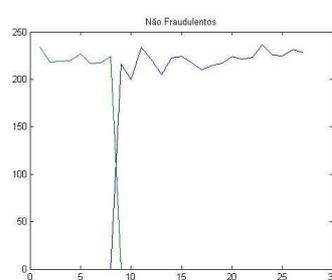


Figura 2. Não detecção de fraudes

Com o arquivo que contém todos os registros, foram feitas duas contagens: a primeira de medições onde foram detectadas fraudes e a segunda onde não foram detectadas e foram gerados dois gráficos (Figuras 1 e 2). As quedas indicam a ausência de medições em determinado período. A partir disso, será analisado o comportamento das medições dos outros arquivos e então será construído e implementado um modelo de Rede Bayesiana.

Referências

[1] <http://www.cos.ufrj.br/~ines/courses/cos740/leila/cos740/aprBayesianas.pdf>

[2]<http://www.das.ufsc.br/~componog/Disciplinas/DAS-5341/Slides/LBn2.pdf>

[3]Jensen, V. Finn. 'Bayesian Networks and Decisions Graphs'. Springer-Verlag, 2011

Aproximação Numérica do Escoamento Bifásico em Meios Porosos

Bolsista: Michel Borges

Orientador: Marcio Borges

1 INTRODUÇÃO

As novas descobertas de reservatórios de petróleo na costa brasileira, conhecidos como "camada pré-sal", e localizados a grandes profundidades, representam um novo potencial petrolífero para o país e novos desafios científicos e tecnológicos visando sua exploração. Neste sentido, dentre os objetivos estratégicos previstos no Plano Diretor do LNCC (2011-2015) está a linha de pesquisa aplicada na área de Petróleo, Água e Gás, que tem por objetivo consolidar a atuação do LNCC como um centro de referência nas áreas de modelagem numérica de processos de prospecção de óleo, gás natural e águas subterrâneas, bem como transporte de poluentes em solos e armazenamento de em reservatórios de petróleo e aquíferos, desenvolvendo pesquisa aplicada e formação qualificada de recursos humanos e exibindo potencial de interação com o setor produtivo. É neste contexto que o presente projeto se enquadra.

2 METODOLOGIA

O escoamento bifásico em um reservatório rígido heterogêneo é modelado por um sistema de equações diferenciais parciais, basicamente composto por um subsistema para a determinação do campo de velocidades e uma equação hiperbólica não-linear para o transporte das fases que escoam (equação da saturação) (LeVeque, 1990; Chen et al., 2006). O desenvolvimento de métodos numéricos localmente conservativos desempenha papel fundamental para a simulação acurada da prospecção secundária em reservatórios de petróleo e a injeção de CO_2 em aquíferos, visando o seu sequestro geológico. Especificamente, propomos neste projeto a aplicação de uma família de métodos de diferenças finitas, não-oscilatórios, de alta ordem, baseados em esquemas centrais, para a equação hiperbólica não-linear que governa a saturação das fases.

Dentro da ampla classe de métodos de diferenças finitas, esquemas centrais têm sido usados com eficácia na resolução de leis de conservação não-lineares (Nessyahu e Tadmor, 1990;

Kurganov e Tadmor, 2000). Tais esquemas consistem de uma classe alternativa de métodos simples e eficientes usualmente construídos usando o conceito de malha deslocada, evitando assim a necessidade da solução de problemas de Riemann sobre as interfaces das células para o cálculo do fluxo numérico. O precursor de tais métodos é o esquema de Lax-Friedrichs de primeira ordem, que consiste em aproximar a solução por funções constantes por partes. No entanto, obteve pouca aceitação devido à grande dissipação numérica, $O(\Delta x^2/\Delta t)$. Em 1990, Nessyahu e Tadmor (Nessyahu e Tadmor, 1990) apresentaram uma generalização de segunda ordem do método de Lax-Friedrichs. Esta generalização consiste em utilizar malha deslocada com o objetivo de evitar a resolução, explícita ou aproximada, de problemas de Riemann e utiliza uma reconstrução linear por partes da solução. O esquema construído desta forma conserva a simplicidade do esquema de Lax-Friedrichs, porém com menor difusão numérica, $O(\Delta x^4/\Delta t)$. Entretanto, este esquema quando aplicado a escoamentos multifásicos em reservatórios de petróleo heterogêneos, uma redução drástica do passo de tempo devido às restrições do tipo CFL podem levar a uma excessiva difusão numérica na solução. Para contornar este problema, Kurganov e Tadmor (2000) propuseram um novo esquema que permite uma formulação semidiscreta, contínua no tempo, que possui um termo de difusão da $O(\Delta x^3)$.

3 OBJETIVOS ESPECÍFICOS

O objetivo principal deste projeto é a modelagem computacional do escoamento de dois fluidos imiscíveis em um meio poroso. Mais especificamente, será estudada a modelagem matemática dos problemas de escoamento água-óleo e água- CO_2 . No segundo caso, incorporando as forças de empuxo. Os métodos numéricos, citados anteriormente, serão usados para aproximar as soluções das equações provenientes da modelagem matemática.

4 PLANO DE TRABALHO DO BOLSISTA

Para alcançar os objetivos propostos o bolsista terá de cumprir as seguintes etapas:

- Estudo da modelagem matemática do escoamento bifásico;
- Estudo dos métodos numéricos para aproximação da solução das equações que descrevem o escoamento bifásico em meios porosos;
- Implementação um código computacional (1D) com os métodos estudados;

- Estudo da qualidade das soluções de cada um dos métodos implementados;
- Elaboração do relatório de pesquisa.

5 CRONOGRAMA DE ATIVIDADES

As atividades do bolsista durante o período de vigência do projeto estão descritas na Tabela 1.

Tabela 1: Cronograma de atividades

| Atividade | Jun | Jul | Ago | Set | Out | Nov | Dez | Jan | Fev | Mar | Abr | Mai |
|------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Modelagem do Problema | x | x | x | | | | | | | | | |
| Estudo dos Métodos Numéricos | | | x | x | x | | | | | | | |
| Implementação dos Códigos | | | | | | x | x | x | | | | |
| Estudos das Soluções | | | | | | | | x | x | x | | |
| Preparo do Relatório | | | | | | x | x | x | x | x | x | x |

REFERÊNCIAS

- Z. Chen, G. Huan, e Y. Ma. *Computational Methods for Multiphase Flows in Porous Media*. Computational Science And Engineering. Society for Industrial and Applied Mathematics, 2006.
- A. Kurganov e E. Tadmor. New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations. *J. Comput. Phys*, 160:241-282, 2000.
- R. J. LeVeque. *Numerical Methods for Conservation Laws*. Birkhäuser, 1990.
- N. Nessyahu e E. Tadmor. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics*, 87(2):408-463, 1990.

Visualização Científica de Rendering de Fluidos

Bolsista: Pedro Henrique Vieira de Oliveira Azevedo

Orientador: Gilson Antonio Giraldi

Co-orientador: Sicilia Ferreira Judice

Introdução

A Dinâmica de Fluidos Computacional (DFC) é uma área de intensa pesquisa devido a suas aplicações tecnológicas e científicas. O comportamento de estruturas mecânicas sujeitas a fortes correntes de ar ou ao movimento das marés, o estudo de fenômenos atmosféricos tais como ciclones e tornados são alguns exemplos da importância das pesquisas nesta área.

Tradicionalmente, o comportamento do fluido é modelado por um sistema de equações diferenciais parciais sujeitos a certas condições iniciais e de contorno. Dificilmente este sistema apresenta solução analítica, o que obriga os fluidicistas a lançarem mão dos recursos da análise numérica para discretizar o problema e obter uma solução com um nível de aproximação aceitável.

Uma outra alternativa nesta área é o uso de técnicas baseadas em Autômatos Celulares como o Método de Lattice Boltzmann (LBM). Esse método busca obter a dinâmica macroscópica do fluido através do comportamento coletivo de inúmeras partículas microscópicas construindo um modelo cinético simplificado (simplificação da equação de Boltzmann) o qual incorpora a física microscópica de maneira que as propriedades macroscópicas obedeçam as equações desejadas [4].

O volume de dados gerado nas simulações é em geral constituído por campos escalares (pressão e densidade) e vetoriais (velocidade e vorticidade), podendo ser estacionários ou não-estacionários (dependentes explicitamente do tempo ou não-dependentes, respectivamente) e a sua análise depende do desenvolvimento de

técnicas específicas que possam filtrar o montante de dados realçando apenas as informações realmente relevantes. Dentre estas técnicas, as técnicas de visualização constituem um recurso de particular importância na análise dos dados [7].

O objetivo do presente trabalho consiste em pesquisar e implementar técnicas de visualização de campos escalares e vetoriais oriundos de métodos de simulação de fluidos. Para tal, será usado como informação inicial os campos de densidade e velocidade resultantes da dinâmica do Método de Lattice Boltzmann tridimensional, através do modelo D3Q19.

1. Modelos de Iluminação

Segundo [1], os modelos de iluminação podem ser agrupados como mostra a Figura 1.

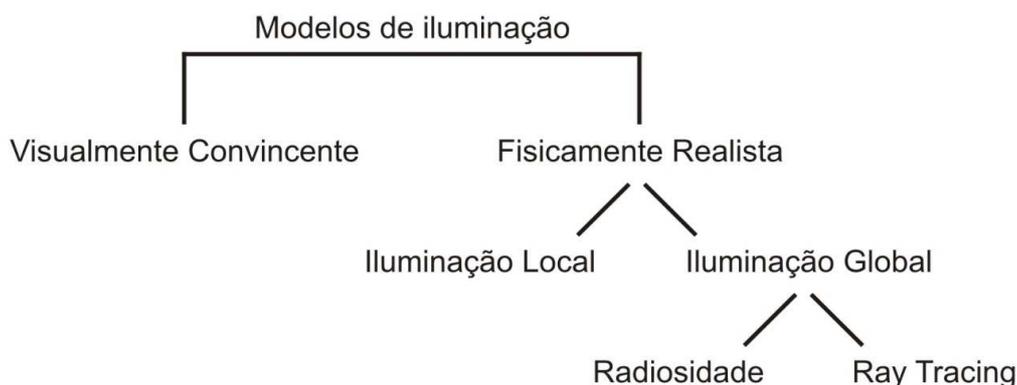


Figura 1 - Classificação dos modelos de iluminação. (Fonte: [1] p. 72)

As técnicas **visualmente convincentes** utilizam métodos heurísticos para a geração de imagens aparentemente realistas, não sendo necessário se basear nas leis da física. Em contrapartida, as técnicas **fisicamente realistas** buscam simular a

interação da luz com os objetos da cena através das leis da física. Os **modelos de iluminação local** utilizam apenas a luz emitida pelas fontes primárias e as propriedades da superfície iluminada para calcular a intensidade de luz refletida. Estes são os modelos usualmente empregados nas técnicas de visualização científica.

No caso dos **modelos de iluminação global** são consideradas também fontes de luz secundárias [6], permitindo, portanto a geração de cenas com efeitos visuais foto-realistas. Os **modelos de iluminação global** citados na Figura 1 levam em conta todo o tipo de interação luminosa que ocorre em uma cena do mundo real. Proposto por [5], o algoritmo de **Radiosidade** tem origem nos métodos fundamentais de transferência de calor. A radiosidade é definida como sendo a energia por unidade de tempo e área em cada ponto dos objetos em cena, onde a energia luminosa incidente em uma superfície é refletida com igual intensidade em todas as direções.

O **Ray Tracing** foi desenvolvido em 1968 como um algoritmo para simulação de trajetórias de projéteis balísticos e partículas nucleares [3]. A primeira implementação para computação gráfica foi desenvolvida por [2], como uma ferramenta para o cálculo de sombras. A referência [8] propôs um algoritmo *ray tracing* para o cálculo de sombras, reflexões e refrações. Por sua simplicidade, muitos softwares de *rendering* atualmente são baseados em ray-tracing.

2. Visualização Científica

As técnicas para visualização de campos escalares tais como rendering de volume e isosuperfícies, são largamente utilizadas na visualização científica. A técnica de isosuperfície é um recurso valioso na visualização de campos escalares, particularmente para o caso bidimensional onde o traçado de isosuperfícies (isolinhas), em geral, fornece boa noção do comportamento do campo. Para o caso 3D tem-se a dificuldade adicional da sobreposição visual de superfícies, o que pode exigir rotações ou mudanças de ponto de vista para melhor compreensão do resultado. Esta técnica pode ser enriquecida pela utilização conjunta de recursos para visualização de campos vetoriais ou mesmo pela utilização de recursos para mapear o comportamento de um campo escalar sobre uma isosuperfície de outro campo escalar.

As técnicas para a visualização de campos vetoriais tais como setas, linhas de corrente, superfícies de corrente e traçado de partículas, são muito convenientes em situações onde o campo não apresenta complexidade elevada. A representação mais comum para a visualização de campos vetoriais são setas, por apresentar uma grande vantagem de compactar em um único símbolo a direção, o sentido e a magnitude do campo em um ponto. No entanto, sua grande limitação está na representação de campos vetoriais 3D devido a inevitável sobreposição de setas em regiões de maior complexidade do campo e à ambiguidade das noções de direção e magnitude do campo decorrentes da projeção plana utilizada para o display na tela do computador.

Uma dificuldade básica com estas técnicas é a grande massa de dados a ser analisada. A utilização direta destas pode levar a muitos cálculos para visualizar o comportamento do fluido em regiões de pouco interesse o que exige uma fase de filtragem da massa de dados para a seleção das regiões de interesse.

Para atender as necessidades de rendering deste projeto utilizaremos os softwares ParaView (<http://www.paraview.org>) e Pov-Ray (<http://www.povray.org>), ambos seguem a filosofia de código livre com portabilidade para os sistemas operacionais Windows e Linux. O software Paraview é voltado para visualização científica, enquanto o Pov-Ray é um software para rendering foto-realista.

2.1. Software Paraview

Paraview é uma aplicação para a visualização e análise de conjuntos de dados científicos, principalmente as que são definidas de forma nativa em um espaço de duas ou três dimensões. A interface gráfica do usuário (GUI) tem uma interface aberta, flexível e intuitiva que nos permite total controle, manipulação e processamento de dados de exibição. O processamento de dados e componentes de renderização do Paraview são construídos sobre uma arquitetura paralela modular e escalável de memória distribuída em que muitos processadores operam sincronicamente em diferentes partes dos dados.

2.2. Software Pov-Ray

O PovRay cria imagens foto-realistas usando técnicas de ray-tracing. O programa lê um arquivo contendo diversas informações de cena, tais como iluminação, descrição dos objetos, dentre outros. A técnica que o programa utiliza, ray-tracing, não é um processo de renderização instantânea porém, são geradas imagens de alta qualidade onde efeitos de sombra, reflexão, refração, dentre outros, são os mais próximos da realidade.

3. Referências

[1] Adabala, N. and Manohar, S. (2002). Techniques for realistic visualization of fluids: A survey. *Comput. Graph. Forum*, 21(1):65–81.

[2] Appel, A. (1968). Some techniques for shading machine renderings of solids. In *Proceedings of the Spring Joint Computer Conference*, pages 37–45, New York, NY, USA:ACM.

- [3] Azevedo, E. and Conci, A. (2003). *Computação Gráfica: Teoria e Prática*. Elsevier Editora.
- [4] Chopard, B. and Droz, M. (1998). *Cellular Automata Modeling of Physical Systems*. Cambridge University Press.
- [5] Goral, C. M., Torrance, K. E., Greenberg, D. P., and Battaile, B. (1984). Modeling the interaction of light between diffuse surfaces. *SIGGRAPHComput. Graph.*, 18(3):213–222.
- [6] Rogers, D. F. (1998). *Procedural Elements for Computer Graphics*. Tom Casson, 2nd edition.
- [7] Rosenblum, L., Earnshaw, R., Encarnacao, J., Hagen, H., Kaufman, A., Klimenko, S., Nielson, G., Post, F., and Thalmann, D. (1994). *Scientific Visualization: Advances and Challenges*. Academic Press.
- [8] Whitted, T. (1980). An improved illumination model for shaded display. *Commun. ACM*, 23(6):343–349.

Modelagem e Monitoramento de Sistemas Ecológicos

Bolsista: Tainara Mendes de Andrade Soares

Orientador: Maurício Vieira Kritz

Introdução

Dada a atual conjuntura de desequilíbrio e destruição do meio ambiente acelerada pelo homem, cada vez mais é necessário compreender os ecossistemas não explorados para minimizar o impacto dessas mudanças. Desse modo é preciso manter esses locais preservados e para isso o monitoramento continuado seria muito útil. Visto a dificuldade de acesso a esses sistemas ecológicos e a necessidade de mantê-los intocados, seria interessante fazer isso de forma mais automática possível.

Assim, a motivação deste trabalho é discutir como o monitoramento continuado e automatizado pode ser realizado para estes sistemas ecológicos e como a modelagem matemática e computacional pode contribuir nisso. Visamos assim aprofundar os nossos conhecimentos sobre um ecossistema intocado, buscando encontrar melhores meios para identificar as mudanças ocorridas nele ao longo do tempo aprimorando sua gestão.

Objetivo

O intuito deste trabalho é discutir como a modelagem matemática e computacional pode contribuir para o monitoramento continuado e automatizado de ecossistemas intocados.

Método

Para alcançarmos este objetivo foi abordado conhecimentos sobre o método científico e discutidos com maior atenção os métodos de observação, os fenômenos biológicos, o raciocínio científico, os modelos, os eventos, os aspectos e os sistemas.

Para entender a relação entre a observação e modelagem matemática e computacional começamos com um exemplo simples, o trabalho de Hodgkin e Huxley (H-H), que se refere à propagação de pulsos elétricos em axônios. Os mecanismos iônicos que viabilizam esta propagação e que são responsáveis pela geração do potencial de ação, foram descobertos ainda na primeira metade do século XX em decorrência do trabalho de dois fisiologistas britânicos, Alan Hodgkin (1914 - 1998) e Andrew Huxley (1917 - 2012).

A partir do conhecimento obtido e do exemplo estudado a mesma sequência de aprendizado será empregada em um ecossistema simplificado. Onde será aprofundado o conhecimento dos aspectos ecológicos bem como os conceitos e métodos apreendidos. Em seguida faremos a discussão de como a modelagem matemática e computacional pode se integrar ao monitoramento continuado e automatizado deste.

Resultados alcançados

Para entendermos o modelo H-H, o raciocínio científico por trás dele e a metodologia de observação utilizada, foram aprofundados os conhecimentos sobre a célula neuronal, os axônios, a membrana neuronal, os canais iônicos, os potenciais de

ação e repouso, a geração do pulso elétrico e os elementos responsáveis por eles, nos quais foram descobertos em decorrência do trabalho de Hodgkin e Huxley.

Buscando entender a relação entre a observação e a construção do modelo fenomenológico, foram analisados o processo de observação, experimentação e elaboração do modelo matemático realizado por H-H, bem como a metodologia empregada por eles.

Por meio das seis perguntas do método científico conseguimos identificar a estrutura do modelo de H-H sobre o fenômeno da propagação de pulsos elétricos em axônios. Vimos que a partir de um experimento simples, a inserção de dois eletrodos no axônio gigante da lula, H-H fizeram um grampo de voltagem com a finalidade de medir a voltagem da corrente da membrana do axônio. Em seguida fizeram uma série de testes para medir o potencial de ação e repouso e começaram a modelar equações para explicar o padrão encontrado.

Assim, estão sendo discutidos como foram feitas as observações e como se dá a propagação de pulsos elétricos através do axônio, os fatores que o influenciam, como isso se reflete no modelo construído e dos demais conceitos apreendidos. A partir disso, a mesma sequência de aprendizado será empregada para entender como a modelagem pode auxiliar na observação de ecossistemas.

Resultado esperado

Esperamos como resultado neste trabalho que o aprendizado da aluna contemple não apenas o modelo de H-H e o ecossistema estudado, mas também as técnicas de modelagem matemática e computacional, bem como o entendimento teórico e prático do método científico. Esperamos também gerar material que possa

servir como fonte de pesquisa para os interessados na área e também contribuir com a divulgação do assunto em meio a sociedade.

Evolução Diferencial Aplicada a problemas de otimização

Bolsista: Vinicius Kreischer de Almeida

Orientador: *Helio José Corrêa Barbosa*

Co-orientador: *Eduardo Krempser*

Resumo do trabalho a ser apresentado

Problemas de otimização, os quais podem ser definidos pela maximização das características de um sistema simultaneamente a minimização das características não desejadas, são frequentemente encontrados nas mais diversas áreas, como na engenharia e na ciência. Neste contexto, diversos algoritmos foram propostos e aperfeiçoados ao longo dos anos visando à resolução de tais problemas de forma satisfatória. Dentre os algoritmos propostos encontra-se a classe dos Algoritmos Evolucionistas, os quais se baseiam em conceitos biológicos, especialmente conceitos evolucionistas.

A Evolução Diferencial, método pertencente à classe dos Algoritmos Evolucionistas, apresenta-se como uma poderosa ferramenta de otimização, além de possuir como características a simplicidade de implementação, facilidade de uso e confiabilidade. Tal método beneficia-se também do fato de não ser necessário o conhecimento explícito da função a ser otimizada, nem mesmo da necessidade do cálculo de gradientes. Os resultados obtidos recentemente em problemas reais e de interesse prático têm impulsionado cada vez mais pesquisas relacionadas a este método. Apesar dos resultados, alguns fatores ainda dificultam sua utilização de maneira mais ampla. Entre estes fatores está a correta configuração dos parâmetros necessários ao seu funcionamento e que apresentam grande influência sobre o resultado final.

O presente trabalho tem como objetivo apresentar os resultados obtidos em testes nos quais se utiliza um conjunto bem definido de problemas de otimização de funções de parâmetros reais, empregando-se as estratégias mais presentes na literatura referente à Evolução Diferencial. Assim, busca-se demonstrar a influência de cada uma das estratégias sobre o comportamento de cada um dos problemas.

Otimização Estrutural Utilizando MatLab

Bolsista: Vitor Luis G. da S. David

Orientador: A.A. Novotny

Co-orientador: Marcel Duarte da Silva Xavier

Resumo

O Método dos Elementos Finitos é uma técnica utilizada para a obtenção de soluções aproximadas de equações diferenciais e é um procedimento numérico para análise de estruturas e meios contínuos.

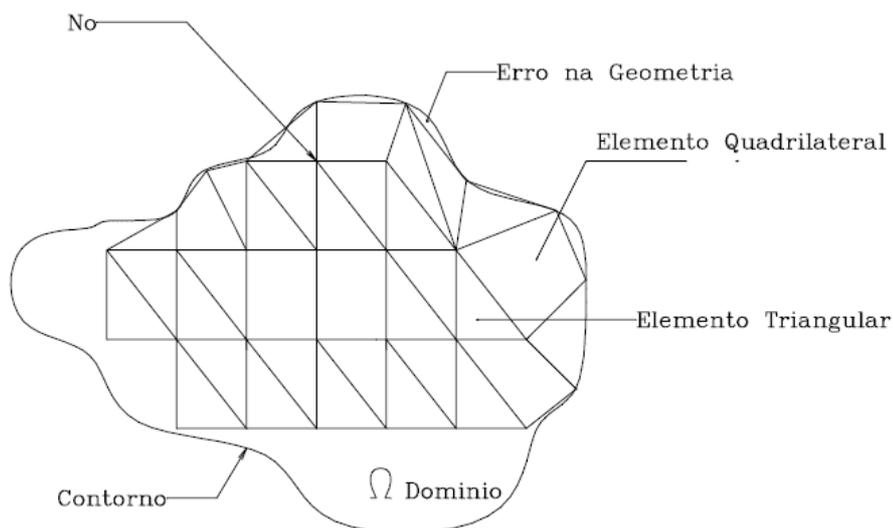
Atualmente, motivados pela forte concorrência industrial os fabricantes de máquinas e equipamentos tem buscado soluções cada vez mais otimizadas, com o intuito de reduzir custos aumentando segurança, conforto e desempenho.

Este desafio tem sido confiado às equipes de engenharia que tem se tornado cada vez mais multidisciplinares, sendo formadas não só por engenheiros das diversas modalidades, mas também por economistas, físicos, matemáticos etc.

Dentre muitas ferramentas disponíveis para o cálculo e análise de estruturas mecânicas e meios contínuos em geral, o Método dos Elementos Finitos tem se destacado como sendo uma ferramenta de uso geral, eficaz e de alto desempenho.

A utilização do Método dos Elementos Finitos é aplicado principalmente ao cálculo estrutural. Todavia, sendo o Método dos Elementos Finitos uma ferramenta geral utilizada para a solução de Equações Diferenciais, é natural que outras disciplinas também sejam abordadas, caracterizando assim a multidisciplinaridade dos problemas de mecânica.

De uma forma global o Método dos Elementos Finitos foi criado com o objetivo de se resolver os problemas de mecânica que não admitem soluções fechadas (de forma analítica). Ele é baseado em aproximações do tipo polinomial nodal em subdomínios, o que implica em processos de discretização dos domínios, que podem ter geometrias irregulares arbitrárias. Na figura abaixo pode ser visto de forma esquemática alguns elementos básicos de um modelo de elementos finitos, e sua nomenclatura.



Pode-se ver no esquema apresentado que o domínio foi discretizado em Elementos Finitos, de forma triangular ou quadrilateral. Neste caso para cada ponto n , chamado de nós, estão associadas às incógnitas do problema. Trata-se assim de uma discretização clássica de um meio contínuo.

Sumário

1 Conceitos Fundamentais; Problema Valor Fronteira;
ou Valor Limite Unidimensional Simples.

1.1 INTRODUÇÃO PRELIMINAR E NOTAÇÕES

1.2 FORTE OU CLÁSSICA FORMA DO PROBLEMA

1.3 FRACA OU VARIACIONAL FORMA DO PROBLEMA

1.4 EQUIVALÊNCIAS DE FORMAS FORTES E FRACAS; NATURAIS;
CONDIÇÕES DE CONTORNO

2 Conclusão

Referências Bibliográficas

Modelagem Computacional Aplicada a Questões Sócio-Educacionais

Bolsista: Weslei Peter de Oliveira

Orientador: Dr. José Karam Filho

Co-orientador: Dr. Fabiano Saldanha Gomes de Oliveira

Este projeto de pesquisa tem por finalidade o estudo e o desenvolvimento de ferramentas computacionais para análise tática e estratégica em educação, a fim de prever consequências sociais.

Dentre as preocupações suscitadas com a mudança de patamar de desenvolvimento alcançado no país, estão questões sociais que, acredita-se, possam estar relacionadas à questão da educação. Duas delas são a criminalidade e a pobreza. Assim pretende-se desenvolver um estudo da relação entre educação-pobreza-criminalidade, tendo como foco principal a educação para afetar de forma positiva os índices de pobreza e criminalidade.

O principal objetivo desse projeto é aplicar técnicas de modelagem e construção de gráficos estatísticos, tendo como motivação questões sócio-educacionais, podendo então generalizar as técnicas e as metodologias à problemas análogos.

Para então ter um benévolo conhecimento em técnicas de modelagem em inferência estatística, iniciou-se o estudo em ferramenta para criação de gráficos estatísticos utilizando a ferramenta OCTAVE, que utiliza uma linguagem computacional, desenvolvida para computação matemática. Os comandos do OCTAVE são muito próximos da forma em que escrevemos expressões algébricas, tornando-o mais simples para seu uso.

Como continuidade do trabalho aqui apresentado, o estudo da ferramenta será aplicado em dados reais, podendo assim apresentar e estudar melhor os dados, visto que se trata de situações reais como já citado acima.

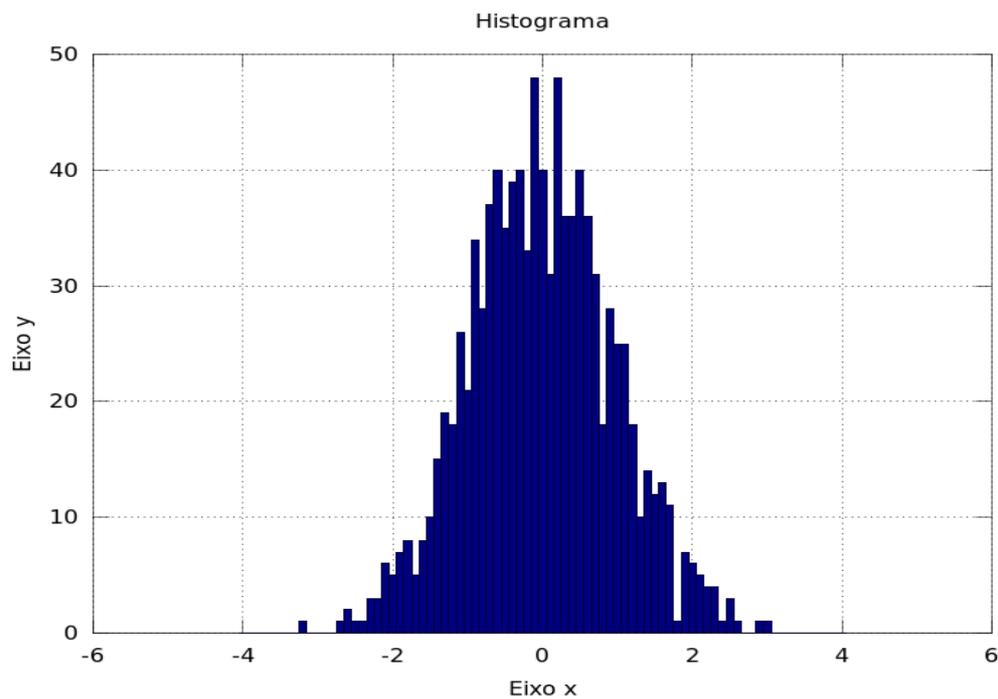


Figura 1-0 Um histograma do OCTAVE.

Referências:

[1] Estatística 4^o edição, Murray R. Spiegel, Larry J. Stephens

Dinâmica de peixes predadores

Estudo de caso: Usina de Santa Clara

Marcelo Alves Moreira
Cezar Welter
Jaqueline Maria da Silva

Instituto de Ciência, Engenharia e Tecnologia – UFJVM

Palavras chaves: populações, modelo Lotka-Volterra, Bacia do Mucuri.

Os estudos de modelos matemáticos de competição e predação de populações surgiram por volta do século XX com o matemático Vito Volterra e o biofísico Alfred Lotka. Em 1925 Volterra desenvolveu um modelo presa-predador tendo como base os estudos estatísticos do zoologista Umberto d'Ancona, que analisou o aumento da frequência de tubarões (predadores) e a diminuição de pequenos peixes (presas) durante a suspensão da pesca nos portos do Mar Adriático. Neste mesmo ano Alfred Lotka publicou o livro *Elements of Physical Biology* (1925) onde discutia a mesma ideia proposta por Volterra justificando o aumento de predadores no Mar Adriático [1, 2].

A região do Vale do Mucuri possui um indicador social e econômico muito baixo, cuja população dependente da agricultura de subsistência e da pesca que é uma das principais fontes de renda, bem como base alimentar. Propomos um estudo científico dos principais fatores que estão relacionados à dinâmica populacional dos peixes na região. Tal estudo foi realizado na Usina Hidrelétrica de Santa Clara, localizada no município de Nanuque-MG na região do Vale do Mucuri [3], fazendo uso do modelo de equações diferenciais de Lotka-Volterra.

O estudo permitiu obter uma perspectiva geral da dinâmica bem como das quantidades de peixes predadores e presas na usina de forma satisfatória. O estudo também mostrou que nem sempre o modelo descreve uma realidade biológica, pois na ausência de peixes predadores a população de peixes presas não cresce de forma

malthusiana como o modelo prevê. Quando ocorre ausência de presas, observa-se que a quantidade de predadores tende a zero, o que também não descreve realidade biológica, pois o predador pode ter outra base alimentar. Outra falha do modelo é não considerar as outras fontes de alimento das presas.

Apesar das deficiências descritas acima, o modelo de predação representou satisfatoriamente a dinâmica biológica dos peixes presas e peixes predadores na Usina Hidrelétrica de Santa Clara, exibindo através de simulações computacionais que a interação entre ambos ocorre de forma cíclica e equilibrada. Este equilíbrio apresenta uma perspectiva geral da dinâmica de predação através de oscilações nas quantidades de indivíduos de cada espécie na bacia ao longo do tempo.

Referências

- [1] Silva J.M., Modelos para a Dinâmica de Vegetação em Áreas Alagáveis Amazônica. Tese de Doutorado, LNCC. Petrópolis - RJ, 2011.
- [2] Boyce W.E.; DiPrima R.C., Equações Diferenciais Elementares e Problemas de Valores de Contorno, LTC, 2006.
- [3] Pompeu P.S.; Martinez C.B., Variações temporais na passagem de peixes pelo elevador da Usina Hidrelétrica de Santa Clara, rio Mucuri, leste brasileiro. Rev. Bras. Zool. vol.23 no.2 Curitiba Junho 2006.

Estudo da rotação de um corpo rígido para aplicação em modelagem da dinâmica de voo de uma aeronave

Bolsistas: Norberto Rodrigues Santos; Luan Márcio Leme Brito

Orientador: Leandro Baroni

Buscando descrever de forma adequada o comportamento dinâmico de uma aeronave, utiliza-se como base a segunda lei de Newton e os Ângulos de Euler. A segunda lei de Newton tem como princípio relacionar a força aplicada em um ponto com relação a sua massa e aceleração adquirida, sendo utilizada como ponto de referência para os estudos de movimentos de translação. Os Ângulos de Euler relacionam dois sistemas por meio de um conjunto de ângulos específicos entre seus eixos e são aplicados na transformação entre sistemas de coordenadas de rotação (corpo \square inercial). Com o intuito de modelar o comportamento dinâmico de atitude a partir das forças exercidas em uma aeronave utiliza-se as equações de Euler, que são obtidas a partir da segunda lei de Newton e descrevem a dinâmica de rotação da aeronave a partir dos torques aplicados sobre ela.

Considerando que as forças aplicadas sobre a aeronave fora da linha do centro de massa geram torques e ocasionam um movimento de rotação ($\mathbf{G} = \mathbf{r} \times d\mathbf{f}$, onde \mathbf{G} é o torque, \mathbf{r} o vetor posição do ponto de aplicação da força com relação a origem e \mathbf{f} é a força). Com aplicação do torque haverá uma variação no momento angular (\mathbf{h}) ou seja, $\mathbf{G} = \dot{\mathbf{h}}$, descrito no sistema inercial:

$$\frac{d}{dt}(d\mathbf{h}) = \mathbf{r} \times \dot{\mathbf{f}} dm$$

O momento angular pode ser relacionado ao momento de inércia (\mathbf{I}) e a velocidade angular $\boldsymbol{\omega}$, $\mathbf{h} = \mathbf{I}\boldsymbol{\omega}$. O momento de inércia é a resistência a mudança do estado de movimento do corpo que está diretamente relacionado a sua distribuição de massa em torno dos seus eixos de rotação e é descrita pela seguinte matriz:

$$\mathbf{I} = \begin{bmatrix} I_x & -I_{xy} & -I_{zx} \\ -I_{xy} & I_y & -I_{yz} \\ -I_{zx} & -I_{yz} & I_z \end{bmatrix}$$

onde $I_x = \int (y^2 + z^2) dm$, etc. e $I_{xy} = \int xy dm$, etc.

No caso da aeronave as forças atuantes geram torques em cada eixo. O momento angular no sistema do corpo é dado por:

$$\mathbf{G}_B = \mathbf{h}_B + \boldsymbol{\omega}_B \times \mathbf{h}_B$$

\mathbf{G}_B e \mathbf{h}_B em termos de seus componentes são $\mathbf{G}_B = [L \ M \ N]^T$, $\mathbf{h}_B = [h_x \ h_y \ h_z]^T$. Reescrevendo a equação de \mathbf{G}_B em termos de seus componentes e considerando $I_{xy} = I_{yz} = 0$ obtém-se:

$$L = I_x \dot{p} - I_{zx}(\dot{r} + pq) - (I_y - I_z)qr$$

$$M = I_y \dot{q} - I_{zx}(r^2 - p^2) - (I_z - I_x)rp$$

$$N = I_z \dot{r} - I_{zx}(\dot{p} - qr) - (I_x - I_y)pq$$

onde p , q e r são os componentes da velocidade angular $\boldsymbol{\omega}_B$.

Os valores típicos para os momentos de inércia de um jato executivo Jetstar Business Jet são: $I_x = 1577915,306 \text{ Nm}^2$, $I_y = 1805037,969 \text{ Nm}^2$, $I_z = 3234983,445 \text{ Nm}^2$ e $I_{zx} = 67236,067 \text{ Nm}^2$. Esta aeronave possui uma massa de 17327,228 kg.

Essas são as etapas iniciais do trabalho, que tem como objetivo obter o conjunto completo das equações da aeronave, realizar simulação computacional e desenvolver um sistema de controle.

Dinâmica de aglomerados populacionais

Samuel Hussin Couto
Jaqueline Maria da Silva
Cezar Welter

Instituto de Ciência, Engenharia e Tecnologia – UFVJM

Palavras-chave: Populações, Efeito Allee, Doenças contagiosas.

Este trabalho visa estudar e modelar a influência de doenças infectocontagiosas sazonais em populações, bem como as condições para a ocorrência de pandemias. Espera-se que uma alta densidade populacional seja favorável à disseminação de doenças, o que pode levar a uma redução populacional, e até mesmo à extinção local. Já uma baixa concentração populacional minimiza as relações intraespecíficas e desfavorece a proliferação de doenças. Tais relações se adequam aos conceitos do efeito Allee.

O efeito Allee (em homenagem ao ecologista Warder Clyde Allee) [1, 2, 3] é um fenômeno natural observado em algumas populações em que: (1) uma população pouco numerosa ou de baixa densidade pode apresentar um crescimento negativo, seja por dificuldades na reprodução, dificuldades para se defenderem de predadores, ou outros motivos; (2) aglomerados populacionais de elevada densidade também pode apresentar crescimento negativo, visto que seus indivíduos estarão sujeitos a limitações alimentares, sujeitos ao ataque de doenças que podem se proliferar rapidamente devido à alta densidade populacional, ou outros motivos; (3) uma população que estiver acima de um valor mínimo (crítico) e abaixo de um valor de saturação terá crescimento positivo.

Considerando um ambiente heterogêneo, temos que a distribuição de alimentos, as diferenças de temperatura e de umidade entre regiões diferentes, podem tornar um local mais atrativo do que outro para uma determinada população, propiciando a ocorrência de uma distribuição populacional não homogênea, e possibilitando a formação de aglomerados populacionais inseridos em um espaço de baixa densidade populacional, e vice versa.

O limite entre a extinção local e extinção total de uma espécie depende do alcance e velocidade da propagação da doença, do alcance e velocidade das relações entre os indivíduos de um mesmo grupo social e entre grupos distintos.

Nesta etapa do trabalho serão apresentados os processos considerados para reproduzir, computacionalmente, a formação de aglomerados populacionais sob uma matriz $N \times N$. Objetiva-se em posteriormente inserir a competição entre indivíduos saudáveis e doentes.

Referências

- [1] Berryman, A. A. Principles of population dynamics and their application. Cheltenham, UK, (1999).
- [2] Wang, M. H., e Kot, M. Speeds of invasion in a model with strong or weak allee effects. Math. Biosci. 171 (2001), 83-97.
- [3] Wang, M. H., e Kot, M., e Neubert, M. G. Integrodifference equations, allee effects and invasions. Mathematical Biology 44 (2002), 150-168.
- [4] Boyce W.E.; DiPrima R.C., Equações Diferenciais Elementares e Problemas de Valores de Contorno, LTC, 2006.