

Implementando Algoritmos Genéticos Paralelos

Alexandra de Aguiar, Fabiano Kist, Josué de Freitas,
Tatiana dos Santos, Cristiano Both, Márcio Kreutz,
Rafael dos Santos

Universidade de Santa Cruz do Sul - UNISC

{aaguiar, fabianokist, josuefreitas}@mx2.unisc.br, tatiana@inf.ufrgs.br, {cboth, kreutz, rsantos}@unisc.br

1 Introdução

A área de processamento de alto desempenho vem se tornando, ao longo dos anos, cada vez mais necessária para a obtenção de soluções para grandes problemas científicos. É comum que os computadores pessoais não tenham capacidade suficiente de processamento dentro de limites de tempo aceitáveis.

Dentro desse contexto, a utilização de agregados de computadores (*Clusters*) é cada vez mais atrativa, pois torna possível a obtenção de processamento distribuído e/ou paralelo a um custo relativamente baixo. Para sua utilização é necessária a paralelização da aplicação desejada. Uma alternativa é a extração de paralelismo realizada manualmente, por parte do desenvolvedor da aplicação. Apesar de se tratar de uma abordagem mais difícil, essa é usualmente a maneira mais eficiente para a obtenção de um bom desempenho [Stallings 2003].

Desse modo, o objetivo do trabalho é analisar o impacto da paralelização de uma aplicação em seu desempenho. O estudo de caso faz a análise de dados espectrográficos de fármacos através da técnica de Algoritmos Genéticos (AG) [Holland 1975] que foi paralelizado segundo dois modelos: AGP (Algoritmo Genético Paralelo) global e AGP de grossa granularidade [Adamidis 1994].

Ambas as versões foram implementadas na linguagem JAVA, largamente utilizada e aceita tanto na área comercial quanto acadêmica. Detalhes da implementação, bem como os resultados de desempenho atingidos serão descritos ao longo desse trabalho.

Este artigo está organizado como a seguir. A Seção 2 apresenta informações sobre o estudo de caso. A Seção 3, descreve as duas formas de paralelização utilizadas. Na Seção 4 é exibida a metodologia do trabalho, enquanto que a Seção 5 mostra os resultados atingidos. Finalmente, a Seção 6 apresenta as conclusões do trabalho.

2 Estudo de caso: AGs Aplicados à Farmacologia

Dentre os insumos utilizados na indústria farmacêutica, encontram-se os anti-hipertensivos que são fármacos capazes de reduzir a pressão ao atuarem na resistência periférica e/ou no débito cardíaco. Para que seu conteúdo seja caracterizado é necessário dosá-lo através de técnicas espectrográficas que resultam em uma grande quantidade de variáveis [Ferrão 2000]. Na busca do subconjunto de variáveis que melhor se ajuste ao

fármaco foi comprovado em Lucasius et.al [Beckers 1994] que AGs apresentaram melhor desempenho em relação a outras técnicas utilizadas anteriormente.

Inicialmente descrito em uma pesquisa desenvolvida por Holland em [Holland 1975], o AG baseia-se na teoria evolucionista de Darwin. Seu objetivo é melhorar uma população de indivíduos através de procedimentos como análise, seleção, cruzamento e mutação.

Uma alternativa para o aumento de desempenho de AGs é a aplicação de técnicas de paralelização. Existem três tipos básicos de Algoritmos Genéticos Paralelos (AGP): global, grossa granularidade e fina granularidade [Adamidis 1994]. Para o presente trabalho foram implementados os modelos AGP global e AGP de grossa granularidade.

3 Paralelização do Algoritmo Genético

Dois modelos de paralelização foram implementados: AGP global e o AGP de grossa granularidade. A seguir, detalhes da metodologia de implementação utilizada.

3.1 AGP Global

Na implementação do modelo AGP global, somente a fase de análise foi distribuída, já que essa é a fase com processamento mais intenso. Esse modelo é bastante semelhante ao modelo de distribuição mestre/escravo. Na implementação desenvolvida cabe ao mestre coordenar quais escravos farão a análise de quais indivíduos. Além disso, o mestre também é o responsável por receber os resultados dos valores de avaliação de uma população. Após receber esses valores, sua função é construir novos indivíduos para a próxima iteração através de cruzamento e mutação.

3.2 AGP de Grossa Granularidade

No AGP de grossa granularidade, também conhecido como modelo da ilha, cada máquina mantém sua própria subpopulação, realizando todas as tarefas típicas de um algoritmo genético (análise, cruzamento e mutação). As diversas ilhas trabalham em conjunto trocando uma porção das suas subpopulações através de mensagens, em um processo chamado de migração.

Desse modo, é possível verificar que o número total de indivíduos numa aplicação distribuída através do modelo da ilha é igual ao número de indivíduos existente em cada uma delas multiplicada pelo número de nodos em que a aplicação é executada. O número de indivíduos em uma aplicação sequencial correspondente é, então, igual ao número total de indivíduos no modelo da ilha. Esse modelo tem como objetivo aumentar o desempenho do AG e evitar que as subpopulações fiquem pouco diversificadas, através da migração de indivíduos.

4 Metodologia de Execução

Para este trabalho foi considerada a utilização de um agregado de computadores. Nessa arquitetura, as máquinas são conectadas através de uma rede *fast-Ethernet*, cuja

vazão é na ordem de *100Mbps/s*. Foi utilizado o *Cluster* da UFRGS, cujas máquinas possuem dois processadores *Pentium III 1.1 GHz* com *1GB* de memória RAM.

Para o cálculo da média do tempo de execução foram consideradas 20 tomadas de resultado para cada configuração utilizada. O ganho de tempo é expresso através do *speed-up*. O valor do *Speed-up* foi calculado conforme demonstrado na equação 1.

$$Speed - up = \frac{T_{Sequencial}}{T_{Paralelo}} \quad (1)$$

Onde:

- $T_{Sequencial}$ = Tempo de execução do algoritmo sequencial e,
- $T_{Paralelo}$ = Tempo de execução do algoritmo paralelo.

Para cada simulação foram realizados testes com 4, 8 e 16 nodos. Já para o número de indivíduos foram utilizados valores entre 25 e 1600 indivíduos. O objetivo de ambas variações é verificar sua influência sobre o tempo final de execução.

5 Resultados

Os resultados obtidos para a implementação do AGP Global são exibidos no gráfico esquerdo da Figura 1. Em ambos os gráficos, o eixo vertical apresenta os valores de *speed-up* e o eixo horizontal denomina as diversas quantidades de indivíduos utilizadas. Cada uma das linhas apresenta o *speed-up* atingido com as diferentes quantidades de nodos.

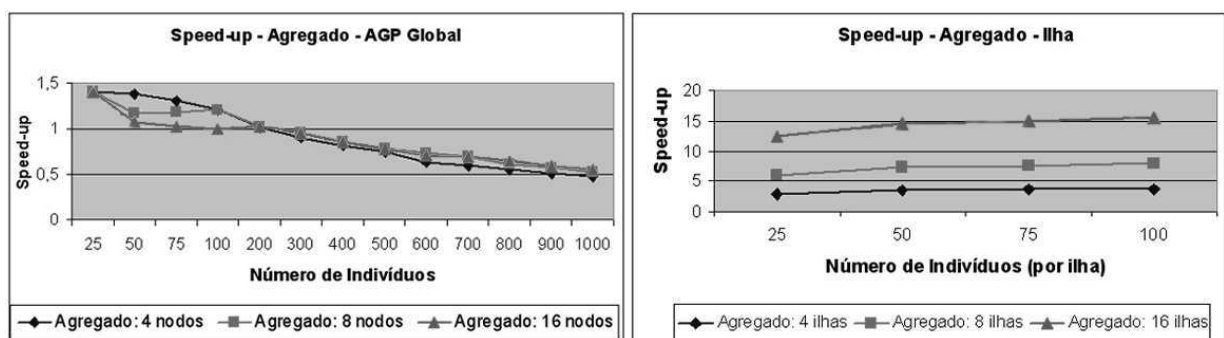


Figura 1: *Speed-up* dos Modelos AGP Global(a) e Modelo da ilha(b)

Na análise dos resultados, observa-se que, apesar do aumento de desempenho em torno de 40% na versão com 50 indivíduos, o desempenho do AGP Global não foi crescente diante do aumento do número de indivíduos. Isso ocorre, pois quanto mais indivíduos são utilizados, maior é o tráfego de mensagens, interferindo de modo negativo na implementação. Além disso, o tempo gasto para analisar um indivíduo em uma só máquina mostra-se menos custoso que enviá-lo aos nodos e esperar por seu resultado.

Já o modelo da ilha, tem o potencial de reverter esses resultados. As linhas que descrevem seu *speed-up* são apresentadas no lado direito da Figura 1. Ao contrário do

AGP Global, o modelo da ilha apresenta um aumento significativo de desempenho conforme o número de nodos aumenta. Os resultados obtidos com 4, 8 e 16 nodos apresentaram desempenho superior às suas versões sequenciais correspondentes, sendo que a última conseguiu um desempenho final quinze vezes maior.

Os resultados evidenciam a fragilidade do modelo AGP Global principalmente ao se aumentar o número de indivíduos. Seu problema, além do excesso de comunicações, é a ociosidade dos nodos escravos enquanto o mestre executa suas tarefas. Assim, tornam-se claras as vantagens da utilização do modelo da ilha, principalmente quando o objetivo é aumentar o *speed-up* do AG.

6 Conclusão

Esse trabalho descreve a implementação paralela de um AG aplicado à farmacologia. O objetivo é conseguir que a paralelização ofereça ao complexo processo de separação de fármacos, uma aplicação mais eficiente e eficaz. Para tal, foram implementados dois modelos de AGPs: os modelos AGP global e de grossa granularidade (ilha).

Os resultados alcançados pelo AGP global ofereceram um pico de desempenho de 40% em relação à sua versão sequencial. No entanto, esse mesmo modelo apresentou perda de desempenho em diversos pontos. Já o modelo da ilha, apresentou resultados significativos. Foi obtido um ganho máximo de 15 vezes sobre a versão sequencial.

Como trabalho futuro, a implementação paralela de acordo com o modelo da ilha será utilizada como aplicação base na prototipação de um agregado de FPGAs. Espera-se que o resultado seja ainda superior ao atingido em *software*, já que o agregado de FPGAs implementará diretamente o algoritmo em *hardware*. A plataforma de desenvolvimento, bem como protocolos de comunicação já estão completamente funcionais e descritas em [de Freitas 2004].

Referências

- [Adamidis 1994] Adamidis, P. (1994). *Review of Parallel Genetic Algorithms Bibliography*. Aristotle University of Thessaloniki.
- [Beckers 1994] Beckers, C. L. M. (1994). *Genetic Algorithms in wavelength selection: A comparative study*. Anal. Chem. Acta.
- [de Freitas 2004] de Freitas, J. P. J. (2004). *Integração em FPGA de um microcontrolador FemtoJava com uma Pilha TCP/IP*. Universidade de Santa Cruz do Sul - Trabalho de Conclusão de Curso.
- [Ferrão 2000] Ferrão, M. F. (2000). *Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo*. PhD thesis, Universidade Estadual de Campinas, Instituto de Química.
- [Holland 1975] Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press.
- [Stallings 2003] Stallings, W. (2003). *Computer organization and architecture : designing for performance*. Prentice Hall: Upper Saddle River, 6 edition.