



RESUMO

PARALELIZAÇÃO DA LINGUAGEM R EM AMBIENTES COMPUTACIONAIS MULTICORE

AUTOR PRINCIPAL:

Paloma Rizzi

E-MAIL:

101082@upf.br

TRABALHO VINCULADO À BOLSA DE IC::

Pibic UPF ou outras IES

CO-AUTORES:

Paulo Ricardo Rodrigues de Souza Júnior, Rafael Machado Sampaio

ORIENTADOR:

Prof. Dr. Carlos Amaral Hölbig

ÁREA:

Ciências Exatas, da terra e engenharias

ÁREA DO CONHECIMENTO DO CNPQ:

1.03.00.00-7 - Ciência da Computação

UNIVERSIDADE:

Universidade de Passo Fundo

INTRODUÇÃO:

Modelos de simulação do crescimento e desenvolvimento de culturas têm sido usados com sucesso ao redor do mundo na agricultura para aumentar a produtividade e reduzir custos. Uma das tecnologias que estão sendo utilizadas atualmente para implementar estes modelos é a linguagem R (ADLER, 2007). Trata-se de uma suíte de softwares integrados que proporcionam facilidades na manipulação de dados, no uso de funções estatísticas e na geração de gráficos. Entretanto, com a crescente complexidade dos modelos, torna-se cada vez mais necessários rodar estes modelos em ambientes computacionais paralelos. Devido a isto, este trabalho foca o uso da linguagem R em ambientes computacionais multicore (SCHMIDBERGER, 2009) visando a paralelização de modelos de culturas. Esta paralelização está sendo realizada por meio do uso do pacote Multicore (MULTICORE, 2011), um pacote para a linguagem R que possibilita uma maneira de executar em paralelo programas implementados em R para máquinas multicore.

METODOLOGIA:

A linguagem R é um projeto open source que está disponível para a maioria das plataformas computacionais. Além de ser uma linguagem de programação também é um ambiente para computação estatística, modelação e visualização de dados (TORGO, 2009). Por estas características pretende-se integrá-lo com o núcleo dos modelos de cultura que estão implementados na linguagem Fortran. Estes modelos apresentam um constante aumento de dados tornando os problemas muito complexos e demandando um grande esforço computacional, o que eleva muito o tempo de processamento dos modelos. Para ser viável trabalhar com este grande número de dados é cada vez mais importante o uso de computação paralela. A linguagem R apresenta diversos pacotes que possibilitam a paralelização. Existem pacotes para Clusters, Grids e para máquinas multicore, sendo este último o contexto escolhido, em um primeiro momento, para a pesquisa. Estes pacotes serão instalados, analisados e comparados por esta pesquisa.

RESULTADOS E DISCUSSÕES:

Existem diversos pacotes para programação paralela em R, por esse motivo foi feito um estudo para comparar os pacotes e decidir qual a melhor opção. Para clusters os principais pacotes utilizados são Rmpi, e Rpvm, respectivamente invólucros de MPI e PVM para uso em R. O RPVM, R Parallel Virtual Machine, é projetado para permitir uma rede de Unix heterogênea e ou máquinas Windows serem usados como um único computador paralelo distribuído. O RPVM é complexo de ser utilizado por valer-se de funções de baixo nível. Em cluster PVM foi largamente utilizado, entretanto vem perdendo espaço para o MPI que esta se tornando padrão na computação paralela. RMPI, Message-Passing Interface, é um sistema padronizado e portátil de transmissão de mensagens em computação paralela, fornecendo uma interface R para funções MPI de baixo nível. Desta forma, o utilizador R não precisa conhecer os detalhes das implementações de MPI. Assim, para clusters o RMPI apresenta melhor solução que o RPVM. Para multicore os principais pacotes são Fork e Multicore. O pacote Fork utiliza basicamente os recursos do sistema UNIX para efetuar a paralelização. Possui uma utilização relativamente simples por ter poucas funções mas não apresenta suporte para funções de alto nível como `apply()`. O pacote multicore apresenta além das chamadas de funções do sistema UNIX, outras rotinas próprias. Sua utilização é mais complexa por apresentar mais funções, porém tem suporte para funções de alto nível. Para ambientes multicore, o pacote Multicore fornece melhores soluções pois reúne mais recursos que o pacote Fork. A utilização de cluster possui um custo maior que a utilização de computadores com processador multicore, além de terem uma implementação mais complexa e uma usabilidade menor. Com a popularização dos multicores, optou-se por este tipo de paralelização dos modelos de cultura, utilizando o pacote Multicore como meio para isto

CONCLUSÃO:

Os modelos de simulação vem apresentando cada vez mais dados é preciso encontrar formas de otimizar o desempenho dos mesmos. Para tanto, a paralelização mostra ser uma das alternativas, pois pode melhorar efetivamente o tempo de execução dos programas. O pacote Multicore vem ao encontro das necessidades viabilizando a programação concorrente destes

REFERÊNCIAS BIBLIOGRÁFICAS:

RIZZI, P., MATOS, E., ALVES, E., HÖLBIG, C. A. Uso da Linguagem R em Ambientes Multicore Voltado a Paralelização de Modelos de Culturas In: VIII Simpósio Brasileiro de Agroinformática, 2011, Bento Gonçalves. Florianópolis: UFSC, 2011.
SCHMIDBERGER, M.; MORGAN, M.; EDELBUETTEL, D.; YU, H.; TIERNEY, L.; MANSMANN, U. State of the Art in Parallel Computing with R. Journal of Statistical Software, vol. 31, Issue 1, p. 1-27, 2009.

Assinatura do aluno

Assinatura do orientador