Execução Paralela de Modelos de Simulação de Culturas e Doenças

Angela Mazzonetto[1], Marcelo Trindade Rebonatto[1], Willingthon Pavan[1], Carlos Amaral Hölbig[1]

[1]Programa de Pós Graduação em Computação Aplicada (PPGCA)
Universidade de Passo Fundo (UPF)

144981@upf.br, rebonatto@upf.br, pavan@upf.br, holbig@upf.br

Resumo. *A quantidade de dados processados nos modelos de simulação de culturas e doenças e a necessidade de rodar simultaneamente o mesmo modelo para grandes áreas ocasionam um aumento no seu tempo de execução, inviabilizando, muitas vezes, a utilização dos resultados. Por isso, esta pesquisa visa estudar alguns modelos, aplicando técnicas de paralelização na sua execução, a fim de obter um melhor desempenho computacional durante sua execução.*

1. Introdução

No atual cenário científico e tecnológico muitos fatores contribuem para o aumento do volume de dados que são utilizados para a solução de aplicações nas mais variadas áreas. Entre estes dados estão os que são processados em modelos de simulação de culturas e doenças. Estes, em especial os dados observados, são posteriormente analisados e as informações deles extraídas, tem por objetivo auxiliar no processo de tomada de decisão [PAVAN 2007]. Porém, a execução destes modelos para áreas cada vez maiores, simultaneamente, podem ocasionar um aumento no tempo de execução, inviabilizando, muitas vezes, a utilização efetiva dos resultados gerados. Por este motivo, torna-se indispensável a utilização de ferramentas ou técnicas computacionais de alto desempenho, visando a otimização e a melhora do desempenho computacional da execução destes modelos [ZHAO et al. 2013].

A melhora no tempo de execução pode ser obtida por meio da paralelização da execução dos modelos de simulação desenvolvidos em linguagens como R, Fortran e Java. Este fato irá contribuir de forma significativa para a análise de séries temporais de dados observados mais amplas e variadas. Possibilitará uma quantidade maior de rodadas de execução dos modelos ou poderá, ainda, ampliar a sua área de cobertura de forma eficiente e em um tempo computacional adequado às necessidades das aplicações que desta análise dependem para ser solucionadas. Portanto, o objetivo deste trabalho é estudar o funcionamento de alguns modelos de simulação de cultura e doenças utilizados pelos grupos de pesquisa envolvidos no projeto e identificar as técnicas de paralelização adequadas para implementar a paralelização da sua execução, visando, com isto, obter resultados satisfatórios de desempenho computacional, ampliando sua área de abrangência.

2. Materiais e métodos

O trabalho que está em andamento nesta pesquisa é dividido em quatro etapas. A primeira consiste no estudo do funcionamento dos modelos de simulação de culturas e doenças

que estão sendo utilizados na Universidade de Passo Fundo e na Embrapa Trigo e que necessitam de uma melhora no seu tempo de execução, onde as limitações atuais estão impossibilitando sua maior aplicação. Estes modelos serão avaliados de forma que seja possível verificar o seu funcionamento, sua execução e como são organizados, armazenados e manipulados seus dados.

Na segunda etapa será feita uma análise das características identificadas nos modelos. Com base nestas, será definido quais técnicas, abordagens ou ferramentas computacionais paralelas serão mais adequadas para serem usadas na execução do modelo de simulação selecionado. As ferramentas que poderão ser utilizadas para a paralelização da execução dos modelos poderão ser a linguagem R, por meio de alguns de seus pacotes paralelos [SCHMIDBERGER et al. 2009, EUGSTER et al. 2011], o Message Passing Interface (MPI) e/ou o *framework* Hadoop.

Na terceira etapa será estruturado e configurado o ambiente computacional (hardware) necessário para a realização desta pesquisa. Por fim, na quarta etapa, com a execução em paralelo dos modelos selecionados e com base nos resultados obtidos nesta execução, pretende-se verificar se a qualidade numérica será mantida e, assim sendo, será avaliado o ganho de desempenho computacional que foi obtido.

3. Conclusões e trabalhos futuros

Observa-se que o processamento e análise de dados atualmente é uma atividade imprescindível para as organizações que possuem grandes bases de dados. Um exemplo são os dados climáticos e meteorológicos que servem de base para a execução de modelos de simulação de várias culturas, como, por exemplo, trigo, soja, morango e maça, e de doenças que atacam estas culturas.

Por isso, este trabalho buscará alternativas paralelas para a execução destes modelos, buscando melhorar o seu tempo de execução e ampliar sua aplicabilidade. Obtendo resultados satisfatórios, a continuidade desta pesquisa possibilitará a utilização destes modelos com uma série mais ampla de dados observados históricos, que podem compreender de meses até algumas dezenas de anos ou na sua execução para um conjunto ampliado de áreas, possibilitando, assim, utilizar os dados na melhoria de sistemas de alertas de pragas e doenças utilizados em várias partes do país.

Referências

EUGSTER, M. J. et al. (2011). Hands-on tutorial for parallel computing with R. *Computational Statistics*, 26(2):219–239.

PAVAN, W. (2007). *Técnicas de Engenharia de Software Aplicadas à Modelagem e Simulação de Doenças de Plantas.* PhD thesis, UPF - Faculdade de Agronomia e Medicina Veterinária, Programa de Pós Graduação em Agronomia., www.ppgagro.upf.br/download/Willingthon_Pavan.pdf.

SCHMIDBERGER, M. et al. (2009). State of the art in parallel computing with R. *Journal of Statistical Software.*, 31:1–27.

ZHAO, G. et al. (2013). Large-scale, high-resolution agricultural systems modeling using a hybrid approach combining grid computing and parallel processing. *Environmental Modelling & Software*, 41(0):231 – 238.