

O Uso de Redes de Petri na Modelagem de Workflows Científicos e seus Recursos*

Duílio Henrique Haroldo Elias, Kelly Rosa Braghetto

Universidade de São Paulo / Instituto de Matemática e Estatística

duilio.elias@usp.br | kellyrb@ime.usp.br

Resumo

Nas últimas décadas os workflows científicos receberam grande atenção da comunidade acadêmica, devido à sua grande capacidade de manipulação, análise e simulação de experimentos que trabalham com grandes quantidades de dados. No entanto, as ferramentas existentes para tal fim são ainda escassas e muitos cientistas afirmam que esse o é um grande gargalo da produção científica. Neste trabalho, busca-se contribuir com este processo, a partir da identificação e caracterização de linguagens de modelagem de workflows científicos, criando modelos estocásticos capazes de descrever analiticamente alguns workflows bem estudado pela comunidade acadêmica. Com isso, pretende-se extrair índices capazes de prever o impacto do uso de diferentes abordagens de controle de fluxo de dados no desempenho dos workflows e compara-los com os dados das simulações.

Abstract

blablablab
blablablab
blablablab

*Este trabalho foi financiado por uma bolsa de iniciação científica RUSP (processo número: ?????).

Introdução

Falar sobre:

- O que são workflows científicos e porque eles são importantes
- Como eles são modelados? Que ferramentas de apoio eles possuem
- Porque é interessante modelar workflows usando linguagens formais como Redes de Petri (verificação de propriedades dos modelos, extração de índices de desempenho, etc.)
- Porque a predição de desempenho é importante? Onde ela pode ser usada?
- Porque é difícil usar linguagens formais na modelagem?

Objetivos

Detalhar um pouquinho as seguintes ideias:

- O objetivo do trabalho é definir um método de conversão de modelos de workflows baseados em fluxos de dados em modelos em Rede de Petri Estocásticas
- Em modelos em Rede de Petri Estocásticas, pode-se também incorporar informações sobre o tempo de execução das atividades e também sobre os recursos computacionais disponíveis para a execução do workflow; por essa razão, é possível extrair deles predições sobre o desempenho do workflow.

Como a parte de comparar os resultados obtidos com as RdPs com os obtidos por meio do simulador não foi finalizada a tempo, não vamos mencioná-la aqui na seção de objetivos. Mas podemos falar dela na seção de trabalhos futuros.

Materiais e Métodos

Aqui a ideia é explicar os conceitos importantes relacionados ao seu trabalho e descrever de forma sucinta o que você fez:

- Explicar sucintamente os componentes de uma RdP
- Falar sobre as construções de fluxo de dados básicas (pipeline, distribuição, agregação)
- Para cada uma das construções acima, mostrar como ela é mapeada na RdP; também explicar que, no modelo em RdP, uma atividade é modelada como uma transição
- Dizer que, como estudo de caso, foi usado um conjunto de workflows científicos sintéticos (extraídos de workflows reais) que ilustram o uso de diferentes construções de fluxos de dados e que com frequência são empregados como base de comparaçã. Citar a fonte desses workflows.
- Dizer que os workflows estavam descritos na linguagem DAX; seu modelos já trazem informações sobre o tempo médio de execução de cada atividade e também sobre o tamanho dos dados de entrada. Essas informações são importantes na construção do modelo estocástico em RdP
- Explicar como informações sobre os recursos computacionais disponíveis podem ser incluídas no modelo em RdP
- Falar que foi usada a ferramenta PIPE para a análise numérica do modelo.
- Dizer que a análise numérica do modelo nos dá de imediato dois índices de desempenho interessantes: a porcentagem de utilização dos recursos e o rendimento (throughput) das atividades e do workflow.

Resultados

Aqui não temos muita coisa para explicar. Temos que tentar incluir um ou dois daqueles workflows que você fez em Redes de Petri e depois dizer que os demais modelos estão no site (e colocar o link para aquela página que eu criei: <http://www.ime.usp.br/~kellyrb/ic/#duilio>).

Conclusões

Aqui é preciso dizer coisas como:

- No trabalho, foram modelados em RdP um conjunto de workflows científicos sintéticos (extraídos de workflows reais) que ilustram o uso de diferentes construções de fluxos de dados
- Concluiu-se que, por meio das Redes de Petri, é possível representar todas as construções básicas de fluxos de dados
- Recursos computacionais simples (como máquinas, VMs, processadores) são facilmente expressos nos modelos em RdP
- A solução numérica dos modelos em RdP traz uma predição do desempenho do workflow (utilização dos recursos e o rendimento (throughput) das atividades e do workflow)
- Dizer que, como trabalho futuro, intenciona-se comparar os resultados obtidos a partir das redes de petri com resultados obtidos por meio da simulação do workflow em ferramentas como o WorkflowSim.