Comparação de *softwares* científicos utilizando perfis de desempenho: automatização dos cálculos pela planilha `perfis.xls`

Pedro Augusto Munari Jr*

Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo - Campus de São Carlos

Resumo

Os perfis de desempenho são uma técnica poderosa na comparação de *softwares* científicos. A comparação é apresentada de forma compacta, permitindo a rápida interpretação dos testes computacionais envolvendo um conjunto de *softwares*. Entretanto, compreender como calcular e interpretar os perfis de desempenho pode não ser simples a princípio. Neste relatório, esta técnica é discutida com o intuito de facilitar sua compreensão e utilização. Como contribuição, é proposta a planilha `perfis.xls` que possibilita automatizar a criação do gráfico de perfis de desempenho a partir dos dados de teste.

1 Introdução

A análise da eficiência de *softwares* científicos é de grande interesse pela comunidade acadêmica. Em geral, compara-se um conjunto de diferentes *softwares* de modo a se escolher aquele com o melhor desempenho na resolução de uma classe de problemas. Neste trabalho, o termo *software* refere-se à implementação computacional de um determinado algoritmo. Além da complexidade teórica associada ao algoritmo, o desempenho de um *software* depende de diversos outros fatores como as técnicas de programação utilizadas para sua codificação, linguagem de programação e compilador empregados e a plataforma computacional em que é executado.

A comparação de *softwares* pode ser feita selecionando-se um conjunto de problemas que são resolvidos em cada um dos *softwares* de interesse. Para cada par problema-*software*, pode-se medir o tempo computacional, o erro da solução obtida, o número de falhas, entre outros. A análise dos resultados obtidos na comparação não é uma tarefa simples. O volume de dados pode

*pedro.munari@gmail.com

ser muito grande, dificultando a apresentação e a interpretação dos resultados. Observando este fato, Dolan e Moré (2002) propõem a utilização de *perfis de desempenho*, uma técnica poderosa para a comparação de diferentes *softwares*. De acordo com os autores, os perfis de desempenho combinam as melhores características de outras propostas disponíveis para a avaliação de desempenho.

Desde sua publicação, os perfis de desempenho têm sido cada vez mais adotados na comparação de *softwares* científicos. De acordo com a ferramenta ISI *Web of Knowledge*[1], a proposta de Dolan e Moré (2002) conta, em março de 2009, com 146 citações em trabalhos publicados em periódicos e conferências indexados pela ferramenta. Certamente, este número seria muito maior se fosse possível verificar as dissertações e teses que também utilizaram a proposta para a análise dos resultados obtidos.

É interessante ressaltar que o número de citações é crescente a cada ano, conforme mostrado na Tabela 1, extraída do analisador de resultados da ferramenta citada. Ainda de acordo com as informações do analisador, as citações estão em trabalhos de diversas áreas como matemática aplicada, pesquisa operacional, engenharia de software, entre outras. Pesquisadores reconhecidos têm utilizado os perfis de desempenho na análise de seus resultados, conforme apresentado na Tabela 2 retirada do analisador de resultados.

Ano	Nº de publicações	(%)
2009	13	25,34
2008	37	20,55
2007	30	19,19
2006	28	10,95
2005	16	8,90
2004	12	8,22
2003	8	5,48
2002	2	1,37

Tabela 1: Número de citações a cada ano.

Autor	Nº de publicações	(%)
GOULD, N.I.M.	10	6,85%
NOCEDAL, J.	7	4,79%
HAGER, W.W.	6	4,11%
MARTINEZ, J.M.	6	4,11%
ZHANG, L.	6	4,11%

Tabela 2: Cinco autores com o maior número de citações.

[1] http://www.isiknowledge.com

Compreender como calcular e interpretar a técnica proposta por Dolan e Moré (2002) não é simples e pode exigir um certo esforço inicial. Assim, os perfis de desempenho são discutidos neste relatório, dando-se enfoque para a compreensão dos cálculos necessários e como se interpretar o gráfico correspondente. Como contribuição, é proposta a planilha `perfis.xls` que permite criar de forma automatizada o gráfico de perfis de desempenho, com o intuito de facilitar a utilização da técnica.

Este relatório é dividido em cinco seções. Na Seção 2, tem-se uma breve descrição de como se calcular e interpretar o perfil de desempenho de um *software*. Um gráfico contendo todos os perfis de desempenho calculados possibilita apresentar de forma compacta a comparação entre os *softwares* considerados, conforme apresentado na Seção 3. Na Seção 4, é descrita a planilha `perfis.xls` e, por fim, são feitas algumas considerações finais na Seção 5.

2 Perfis de desempenho

O conteúdo desta seção é baseado no artigo de Dolan e Moré (2002). São apresentados apenas os pontos principais, de forma sucinta, com o objetivo de facilitar a compreensão da técnica.

Considere que um conjunto $\mathcal{P}$ de n_p problemas seja utilizado no teste de um conjunto $\mathcal{S}$ de n_s *softwares*. Para cada problema $p \in \mathcal{P}$ e *software* $s \in \mathcal{S}$, seja $t_{p,s}$ o tempo computacional necessário para a resolução de p utilizando-se s. Para utilizar outra métrica de desempenho, basta definir $t_{p,s}$ de forma adequada. Por exemplo, pode-se considerar a quantidade de memória utilizada, o *gap* da solução obtida, entre outras medidas.

Para um *software* $s \in \mathcal{S}$, a comparação de seu desempenho na resolução de um problema $p \in \mathcal{P}$, em relação ao desempenho do melhor *software*, é feita utilizando-se a *razão de desempenho*:

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : \forall\ s \in \mathcal{S}\}}.$$

Observe que esta razão é sempre maior ou igual a 1. Se o *software* s falhar na resolução do problema p então $r_{p,s}$ é definido como r_M, tal que

$$r_M \geq r_{\bar{p},\bar{s}},\ \forall\ \bar{p} \in \mathcal{P},\ \forall\ \bar{s} \in \mathcal{S}.$$

Neste trabalho, considera-se $r_M = 1 + \max\{r_{\bar{p},\bar{s}} : \bar{p} \in \mathcal{P} \text{ e } \bar{s} \in \mathcal{S}\}$. Conforme mostrado por Dolan e Moré (2002), a escolha de r_M não influencia na avaliação de desempenho, desde que satisfaça a desigualdade acima.

A razão de desempenho mostra o comportamento de um *software* na resolução de um determinado problema. Entretanto, o que se deseja é uma avaliação geral do desempenho do *software*. Para isso, é definido o *perfil de desempenho* do *software*, dado por:

$$\rho_s(\tau) = \frac{1}{n_p} \left| \{p \in \mathcal{P} : r_{p,s} \leq \tau\} \right|,$$

com $|.|$ representando o número de elementos no conjunto. O perfil de desempenho $\rho_s(\tau)$ é uma função que associa a um dado valor $\tau \in \mathbb{R}$, a fração de problemas resolvidos pelo *software* com um desempenho dentro de um fator τ do melhor desempenho obtido. Tem-se que $\rho_s : \mathbb{R} \mapsto [0, 1]$ é uma função constante por partes, não-decrescente e contínua à direita de cada ponto de descontinuidade.

Um perfil de desempenho também pode ser visto como a probabilidade de que a razão de desempenho $r_{p,s}$ associada ao *software* s esteja dentro de um fator τ da melhor razão obtida. Com isso, $\rho_s(\tau)$ corresponde a uma função de distribuição acumulada para a razão de desempenho, dada por:

$$\rho_s(\tau) = P(r_{p,s} \leq \tau : 1 \leq s \leq n_s).$$

Pode-se dizer, também, que $\rho_s(\tau)$ é a probabilidade de que o *software* s resolva um problema de $\mathcal{P}$ em não mais do que τ vezes o tempo mínimo levado por qualquer *software* em $\mathcal{S}$. Se o conjunto $\mathcal{P}$ é suficientemente grande e capaz de representar os problemas provavelmente encontrados na prática, então os *softwares* com grande probabilidade $\rho_s(\tau)$ são preferidos.

Para se analisar um perfil de desempenho, alguns pontos principais devem ser observados. O primeiro deles é o valor associado a $\tau = 1$. Para um dado *software* $s \in \mathcal{S}$, o valor $\rho_s(1)$ corresponde à fração de problemas para os quais s obteve melhor desempenho que os demais *softwares*, ou seja, é a probabilidade de que o *software* ganhe dos demais. Assim, se o interesse é saber o número de vezes que um *software* s foi o melhor, basta analisar o valor de $\rho_s(1)$.

O valor assumido no outro extremo também deve ser verificado. Tem-se que $\rho_s(r_M) = 1$ já que $r_{p,s} = r_M$ quando o problema p não é resolvido pelo *software* s. Com isso, a probabilidade de que o *software* s resolva um problema p qualquer é dada por ρ_s^*, tal que $\rho_s^* = \rho_s(r_M - \varepsilon)$, com $\varepsilon \to 0$ e $\varepsilon \in (0, \infty)$. Em outras palavras, ρ_s^* é o maior valor assumido por $\rho_s(\tau)$, quando τ está o mais próximo possível de r_M, à esquerda, porém ainda é diferente de r_M. Observe que ρ_s^* não depende de um parâmetro τ e, assim, indica apenas se o *software* é capaz ou não de resolver um problema, sem considerar o desempenho que terá na resolução. Se o interesse está em escolher um *software* com grande probabilidade de sucesso, então deve-se comparar os valores de ρ_s^* para cada $s \in \mathcal{S}$.

Conforme demonstrado por Dolan e Moré (2002), os perfis de desempenho não são sensíveis aos resultados em um pequeno número de problemas, pois o desempenho obtido em um problema particular do conjunto não afeta de modo significativo os perfis de desempenho associados. Além disso, os perfis são dificilmente afetados por pequenas mudanças nos resultados associados a muitos problemas.

3 Gráfico de perfis de desempenho

Para a comparação dos *softwares*, os perfis de desempenho calculados devem ser apresentados em um único gráfico, cuja abcissa represente o parâmetro τ variando no intervalo $[1, r_M)$. Para cada *software*, tem-se uma curva que representa seu perfil de desempenho conforme a variação de τ. Como r_M pode ser muito maior do que 1, recomenda-se utilizar escala logarítmica para a apresentação dos perfis de desempenho no gráfico. Assim, será utilizado o mapeamento:

$$\tau \mapsto \frac{1}{n_p} \left| \{ p \in \mathcal{P} : \log_2(r_{p,s}) \leq \tau \} \right| .$$

Agora, o fator τ varia no intervalo $[0, r_M)$, com $r_M = 1 + \max\{\log_2(r_{\bar{p},\bar{s}}) : \bar{p} \in \mathcal{P} \text{ e } \bar{s} \in \mathcal{S}\}$. Com isso, deve-se ter cuidado na interpretação dos perfis de desempenho baseando-se em τ, devido à utilização da escala logarítmica.

Na Figura 1, tem-se um gráfico contendo os perfis de desempenho de três *softwares*, referenciados como *Software* 1, *Software* 2 e *Software* 3. Para a criação deste gráfico, um mesmo conjunto de problemas foi resolvido por cada um dos *softwares*, medindo-se os tempos de execução associados e o número de falhas ocorridas. A partir destes dados, calculou-se as razões de desempenho e, então, os perfis de desempenho.

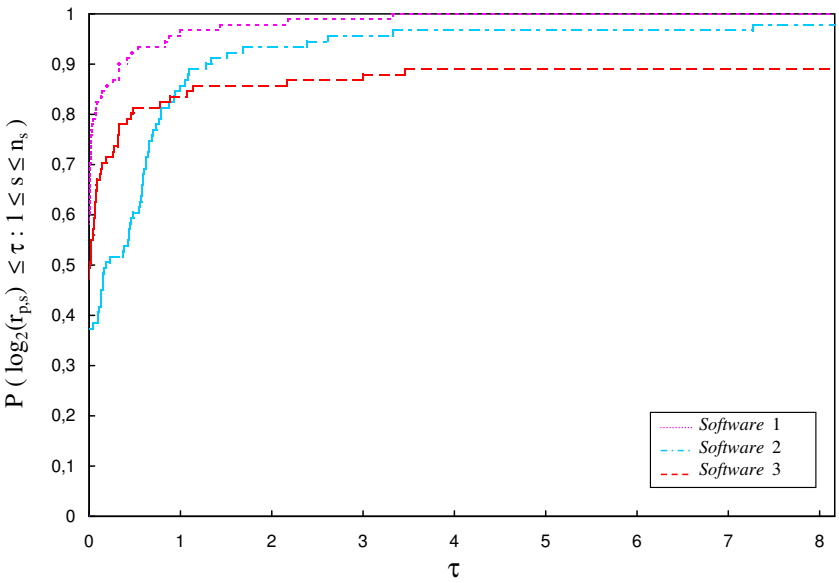

Figura 1: Gráfico de perfis de desempenho para 3 *softwares*.

Observando a curva do perfil de desempenho do *Software* 1, vê-se que ele obteve melhor desempenho que os demais *softwares* em quase 60% dos

problemas ($\tau = 0$). Além disso, este *software* conseguiu resolver todos os problemas, dado que a curva de seu perfil de desempenho atingiu o valor 1, como pode ser observado para τ próximo de 3. Assim, pode-se dizer que este *software* foi capaz de resolver cada um dos problemas, dentro de um fator próximo de 2^3 do melhor desempenho obtido entre todos os *softwares*.

Uma observação importante é que o *Software* 1 se destaca em relação aos demais *softwares*, dada a curva de seu perfil de desempenho dominar as demais curvas. Isto não indica que este *software* obteve melhor desempenho que os demais e, sim, que ele conseguiu resolver mais problemas dentro de um fator τ do desempenho de qualquer outro *software*.

Os demais *softwares* não foram capazes de resolver todos os problemas, dado que as curvas de seus perfis de desempenho não atingem o valor 1 para $\tau \in [0, r_M)$. Pode-se verificar que o *Software* 2 resolveu cerca de 98% dos problemas, enquanto o *Software* 3, quase 90%. Em outros termos, a probabilidade de se resolver um problema é em torno de $0,98$ para o *Software* 2 e $0,90$ para o *Software* 3.

É interessante notar como o gráfico de perfis de desempenho apresenta de forma compacta a comparação entre os *softwares* de acordo com a métrica adotada e o número de falhas. Pode-se ver claramente o número de problemas nos quais um *software* teve desempenho melhor que os demais e quantos problemas ele foi capaz de resolver dentro de um determinado fator τ do melhor desempenho possível.

Para facilitar a interpretação do gráfico, recomenda-se a utilização de uma tabela que contenha os valores extremos dos perfis de desempenho de cada *software*. Um exemplo considerando o gráfico apresentado na Figura 1 é dado na Tabela 3.

	$\tau = 0$	$\tau \to r_M$
Software 1	0,58	1,00
Software 2	0,37	0,98
Software 3	0,48	0,89

Tabela 3: Tabela com os valores extremos dos perfis de desempenho.

4 A planilha `perfis.xls`

Como apresentado na seção anterior, a criação do gráfico de perfis de desempenho envolve diversos cálculos e pode ser trabalhosa. Assim, deve-se ter uma maneira de automatizar a criação do gráfico, a partir dos dados que se deseja utilizar na comparação dos *softwares*.

O *script* `perf.pl` está disponível no *site* COPS[2] para a criação automatizada do gráfico de perfis de desempenho, utilizando-se o *software* Matlab.

[2] http://www.mcs.anl.gov/~more/cops

Entretanto, este *software* é comercial, restringindo o uso do *script*. Além disso, o usuário pode não ser familiarizado com o Matlab e ter dificuldades para entrar com os dados obtidos em seus resultados computacionais.

Mittelmann e Pruessner (2006) apresentam um aplicativo *web*, denominado PAVER, que pode ser utilizado para o cálculo do gráfico de perfis de desempenho, além de outras medidas de desempenho disponíveis. Entretanto, o aplicativo exige que a entrada de dados seja feita por meio de um *trace file*, um arquivo texto que possui um formato particular especificado pelo autores. Esta exigência prejudica a facilidade de utilização do PAVER, já que o usuário deve converter seus dados para o formato especificado.

Considerando estes fatos, a planilha `perfis.xls` é proposta neste trabalho, cujo objetivo é a criação automatizada do gráfico de perfis de desempenho. Por estar no formato `.xls`, esta planilha pode ser utilizada no Microsoft Excel ou em qualquer outro *software* não-comercial equivalente, o que torna a entrada de dados bastante simples. A planilha `perfis.xls` foi desenvolvida pelo autor e está disponível no *site*[3] do Laboratório de Otimização do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.

Para gerar o gráfico de perfis de desempenho utilizando a planilha, o usuário precisa apenas entrar com os resultados dos testes realizados com cada *software*, considerando uma determinada métrica de desempenho e o número de falhas. É importante que o usuário tenha bem definido o que será considerado como falha em suas comparações.

Os dados devem ser inseridos na aba `Dados` (canto inferior esquerdo), nas colunas `B` a `H`. Não é necessário preencher todas as colunas. Entretanto, uma vez que uma coluna for utilizada, todas as suas linhas devem ser preenchidas até o número total de problemas considerado. Para indicar que um *software* falhou na resolução de um problema, deve-se entrar com o texto `NaN` na célula correspondente. A coluna `A` é utilizada para se associar nomes aos problemas, mas seu preenchimento é opcional. As demais colunas, a partir da coluna `I`, são calculadas automaticamente e não devem ser alteradas.

A variável `Epsilon`, definida no canto superior esquerdo da aba `Dados`, é utilizada quando o melhor desempenho para um dado problema p é igual a zero. Para permitir que a razão de desempenho $r_{p,s}$ associada seja calculada, o valor $t_{p,s}$ de cada *software* é deslocado uma quantidade `Epsilon` à direita, isto é, soma-se `Epsilon` a $t_{p,s}$. O valor definido como padrão é `Epsilon` = 1, porém o usuário pode ajustá-lo conforme desejar. Definindo-o como 0, considera-se que todos os softwares tiveram o mesmo desempenho para o problema, caso o melhor desempenho seja nulo. Considerando esta variável, o cálculo da razão de desempenho é dado pelo Algoritmo 1.

Tem-se também, na mesma aba, a variável `USAR_LOG` que determina se o cálculo das razões de desempenho deve ser realizado utilizando-se loga-

[3] http://www.otm.icmc.usp.br/aplicativos

Algoritmo 1: Razão de desempenho calculada pela planilha `perfis.xls`

```
1  Se t_{p,s} = ∅ então r_{p,s} ← ∅
2  Senão Se t_{p,s} = NaN então r_{p,s} ← NaN
3          Senão Se min{t_{p,s}, ∀ s ∈ S} > 0 então
4                  r_{p,s} ← t_{p,s} / min{t_{p,s}, ∀ s ∈ S}
5                Senão Se min{t_{p,s} + ε, ∀ s ∈ S} > 0 então
6                        r_{p,s} ← (t_{p,s} + ε) / min{t_{p,s} + ε, ∀ s ∈ S}
7                      Senão r_{p,s} ← 1
```

ritmo. Ao se definir `USAR_LOG`= 1, o gráfico de perfis de desempenho será apresentado usando escala logarítmica. Para não utilizá-la, deve-se definir `USAR_LOG`= 0. A etiqueta que descreve o eixo das ordenadas deve ser ajustada.

Após entrar com os dados, o usuário pode visualizar o gráfico de perfis de desempenho, gerado automaticamente, clicando na aba `Gráfico`. As abas `Perfis` e `Plotagem` são responsáveis pelos cálculos necessários para a criação das curvas em forma de escada, correspondentes a cada um dos perfis de desempenho. Como esse cálculo é automatizado, as células destas abas não devem ser alteradas.

O gráfico gerado pela folha de cálculo pode ser exportado para a utilização em um documento de texto. Para utilizá-lo no Microsoft Word, ou similar, basta copiá-lo utilizando-se o botão Copiar da barra de ferramentas e, então, colá-lo no documento de destino. Para utilizá-lo em um documento `.tex`, pode-se copiar o gráfico e colá-lo em um editor de imagens, como o Microsoft Paint. Após salvar a imagem contendo o gráfico, é possível inserí-la no documento `.tex`. Uma outra maneira, é imprimir o gráfico usando uma impressora `pdf`. Com isso, o documento `.pdf` contendo o gráfico pode ser inserido no documento `.tex`.

5 Considerações finais

Os perfis de desempenho possibilitam a exposição compacta dos resultados obtidos em testes computacionais, mesmo quando o volume de dados é grande. Além disso, a utilização do gráfico de perfis de desempenho permite que a interpretação seja feita de forma simples e direta. A planilha `perfis.xls` permite que o usuário crie, de forma automatizada, o gráfico correspondente aos seus dados de testes e utilize-o em documentos de texto.

É interessante ressaltar que a utilização do gráfico de perfis de desempenho não descarta a utilização das tabelas contendo os dados utilizados para criá-lo. Outros cálculos também devem ser feitos de modo a complementar a interpretação dos dados, como a média, o desvio padrão, entre outros.

Referências

Dolan, E. D.; Moré, J. J. Benchmarking optimization software with performance profiles. *Mathematical programming*, v. 91, p. 201–213, 2002.

Mittelmann, H. D.; Pruessner, A. A server for automated performance analysis of benchmarking data. *Optimization Methods and Software*, v. 21, n. 1, p. 105–120, 2006.