

Estudo de Abordagens Para Redução do Tempo de Avaliação de Desempenho de Programas Paralelos

Elisa Gabriela Machado de Lucena¹, Samuel Xavier-de-Souza¹

¹Departamento de Engenharia da Computação e Automação
Universidade Federal do Rio Grande do Norte (UFRN)

elisa.lucena.127@ufrn.edu.br, samuel@dca.ufrn.br

Abstract. *Scalability analysis in parallel applications can be costly due to the need for multiple executions. This work proposes a comparison between an analytical modeling and a neural network (MLP) approach to estimate execution time. The goal is to evaluate the trade-off between computational cost, accuracy, and generalization. The expected outcome is to enhance the Pascal Analyzer by integrating predictive techniques that speed up performance analysis.*

Resumo. *A análise de escalabilidade em aplicações paralelas pode ser custosa devido à necessidade de múltiplas execuções. Este trabalho propõe a comparação entre uma modelagem analítica e uma abordagem com redes neurais MLP para estimar o tempo de execução. A meta é avaliar o equilíbrio entre custo computacional, precisão e generalização. Espera-se que os resultados contribuam para a evolução do Pascal Analyzer, com a incorporação de técnicas preditivas que acelerem a análise de desempenho.*

1. Introdução

A análise de desempenho de aplicações paralelas é essencial, especialmente em ambientes de alto desempenho. Ferramentas como o *Pascal Analyzer* [da Silva et al. 2017] — desenvolvida pelo LAPPS/UFRN — permitem avaliar a escalabilidade de algoritmos paralelos, mas exigem múltiplas execuções sob diferentes configurações, o que torna o processo custoso.

Diante desse desafio, surgem soluções baseadas em aprendizado de máquina, como o uso de redes neurais do tipo MLP para prever o tempo de execução [Rodrigues 2025]. Apesar da acurácia, essas abordagens demandam alto custo computacional e grandes volumes de dados [Sze et al. 2017].

Este trabalho propõe a comparação de soluções baseadas em aprendizado de máquina com alternativas de menor custo computacional, como uma modelagem analítica do tempo de execução. A proposta procura equilibrar custo, precisão e aplicabilidade prática na análise de escalabilidade.

2. Metodologia

A metodologia do trabalho consiste inicialmente na construção de um modelo analítico baseado em curvas de tempo de execução, com aplicação de modelagem semelhante à utilizada por Furtunato et al [Furtunato et al. 2020], adaptando para a métrica de tempo de execução.

Posteriormente, os resultados obtidos com essa modelagem serão comparados a uma abordagem baseada em aprendizado de máquina, com algoritmos treinados com subconjuntos amostrados dos dados de execução.

Os experimentos serão realizados com três algoritmos - todos paralelizados com diretivas do OpenMP: multiplicação de matrizes, busca em largura (BFS) e o traçado de raio (ray trace). Para cada algoritmo, serão variadas o número de núcleos utilizados e o tamanho da entrada. O tempo de execução será registrado para diferentes combinações desses parâmetros, servindo como base para a construção e validação dos modelos.

A comparação entre eles será feita com base em três métricas: erro relativo absoluto entre o tempo estimado e o tempo real, tempo total de análise e custo computacional.

3. Resultados Esperados

Espera-se que a abordagem analítica reduza o tempo de análise sem comprometer a acurácia, principalmente em cenários onde a relação entre parâmetros de entrada e tempo de execução possa ser descrita por funções simples. Por outro lado, a abordagem com MLP tende a ser mais precisa em situações mais complexas, embora demande maior esforço computacional.

Porém existem desafios, como a limitação da abordagem analítica em modelar comportamentos não lineares e o risco de overfitting nos modelos de aprendizado de máquina, caso o conjunto de dados de treino não represente adequadamente todo o domínio de execução. Ao final, pretende-se elaborar uma abordagem que oferece o melhor equilíbrio entre custo computacional e precisão, possibilitando uma junção das metodologias para a evolução do Pascal Analyzer.

4. Considerações Finais

Este estudo propõe uma investigação sobre estratégias alternativas para otimizar o processo de análise de escalabilidade de aplicações paralelas. A abordagem combina modelagens analíticas e técnicas de aprendizado de máquina, visando alcançar maior eficiência com menor custo computacional. As etapas seguintes envolvem o desenvolvimento experimental e a validação das propostas. Espera-se que os resultados contribuam para tornar as análises de desempenho mais rápidas, econômicas e viáveis em diferentes contextos de aplicação.

References

- da Silva, A. B., Cunha, D. A., Silva, V. R., de A. Furtunato, A. F., and Xavier-de Souza, S. (2017). Pascal viewer: A tool for the visualization of parallel scalability trends. In *International Workshop on Extreme-Scale Programming Tools*, pages 250–264. Springer.
- Furtunato, A. F. A., Georgiou, K., Eder, K., and Xavier-De-Souza, S. (2020). When parallel speedups hit the memory wall. *IEEE Access*, 8:79225–79238.
- Rodrigues, M. d. S. L. (2025). Estimação adaptativa baseada em acurácia para tempo de execução de aplicações paralelas. B.S. thesis, Universidade Federal do Rio Grande do Norte.
- Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329.