



Evento	Salão UFRGS 2017: FEIRA DE INOVAÇÃO TECNOLÓGICA DA UFRGS - FINOVA
Ano	2017
Local	Campus do Vale - UFRGS
Título	Tecnologia CMOS
Autores	LARISSA ROZALES GONÇALVES GERALDO FRANCISCO DE OLIVEIRA JUNIOR
Orientador	LUIGI CARRO

RELATÓRIO

ATIVIDADES DO ALUNO DE INICIAÇÃO TECNOLÓGICA E INOVAÇÃO 2016-2017

TECNOLOGIA CMOS

Orientador: Luigi Carro

Aluno: Larissa Rozales Gonçalves

Período integral das atividades: 01/08/2016 a 31/07/2017

RELATÓRIO DE ATIVIDADES

1. Introdução:

A especialização do hardware surge no contexto de uma demanda constante por mais velocidade e maior eficiência da computação enquanto limites de área e de gastos de energia são respeitados. Em vez de construir um chip genérico que busca obter boa performance num caso geral, o que se faz atualmente é incluir unidades de hardware específico para diferentes tipos de computação, como por exemplo: GPUs para computação gráfica, DSP para processamento de áudio e imagem, FPGAs para telecom, etc.

A computação aproximativa é um paradigma que introduz a qualidade como métrica, permitindo que algum erro seja introduzido nos resultados da computação, trocando a precisão por melhor performance, menor gasto de energia e menos área utilizada. Com a especialização do hardware, é possível utilizar a computação aproximativa para tratar grupos de aplicação específicas que toleram imprecisões, como aprendizado de máquina, jogos, multimedia, data mining, etc.

Redes neurais são mecanismos populares para fazer computação aproximativa, utilizando hardware dedicado e substituindo regiões de código por chamadas à rede. As redes neurais, no entanto, exigem muita memória para armazená-las, assim como grandes unidades de ponto flutuante. As árvores de decisão, mecanismo de aprendizado supervisionado que foram utilizadas nesta pesquisa, buscam realizar a computação aproximativa sem a necessidade de tantas operações de ponto flutuante, sem aumentar a complexidade e mantendo o erro relativamente baixo.

2. Atividades realizadas:

Antes de encontrar e escolher as árvores de decisão como substitutas para as redes neurais, outros mecanismos de aproximação foram utilizados: as redes BAM, o algoritmo kmeans e as árvores KD. De todos os métodos testados, as árvores de decisão apresentaram o melhor resultado. Foram feitos testes utilizando as ferramentas gem5 e PAPI para obter as métricas, além de medições de tempo com a biblioteca *chrono* do C++.

Para realizar estes testes foram utilizados os benchmarks disponíveis em <http://axbench.org/>, especificamente as aplicações para CPU. O primeiro trabalho, antes de realizar as substituições pelos outros métodos, foi reproduzir cada um dos benchmarks com as redes neurais disponíveis no link fornecido e utilizadas em outros trabalhos da área.

3. Objetivos atingidos:

- Replicação de resultados obtidos utilizando redes neurais de outras pesquisas, como mencionado anteriormente.
- Substituição das redes neurais por outros mecanismos de aproximação.
- Execução das substituições e realização de testes com as ferramentas PAPI e gem5, além da medição do tempo da execução. Esses testes também foram realizados sobre a rede neural original, com objetivo de fazer comparações entre as redes e o método pesquisado.
- Comparação dos resultados obtidos com todos os métodos testados.

4. Resultados obtidos:

Segue uma tabela com os resultados obtidos para alguns dos benchmarks medidos:

Benchmark	Aceleração (árvore de decisão)	Erro da árvore de decisão (%)	Erro da rede neural (%)
inversek2j	4	16%	13%
jpeg	2	21%	24%
sobel	2	9%	6%
blackscholes	3,5	13%	27%

A coluna aceleração se refere à aceleração obtida das árvores de decisão sobre o benchmark original sem nenhuma aproximação; as redes neurais são sempre mais lentas que a versão sem aproximação, a não ser que seja usado um hardware específico para sua aceleração, que no caso não foi utilizado para as árvores.

5. Conclusão:

A computação aproximativa introduz um novo eixo no design de espaço de processadores: a qualidade. Com este novo eixo, se torna importante saber o quão precisa as aplicações devem ser.

Além disso, a pesquisa mostrou que existem alternativas viáveis para redes neurais, além de ser possível obter melhora de performance, em relação ao código original, em aplicações com cálculos complexos. Existe a possibilidade, também, de acelerar uma área crítica da nova economia, a inteligência artificial.