Yao Jianyu

Residence
Beijing, China

Email
yaojianyu19f@ict.ac.cn; yaojianyu89@gmail.com

Education

Masters degree. Institute of Computing Technology, Chinese Academy of Sciences [2019 - 2022]

Bachelors degree in Software Engineering. Shandong University [2015 - 2019]

Research

IAAT: A Input-Aware Adaptive Tuning framework for Small GEMM

November 2020 - May 2021. The paper is under review in ICPADS'21

GEMM with the small size of input matrices is becoming widely used in many fields like HPC and machine learning. Although many famous BLAS libraries already supported small GEMM, they cannot achieve near-optimal performance. This is because the costs of pack operations are high and frequent boundary processing cannot be neglected. We proposes an input-aware adaptive tuning framework(IAAT) for small GEMM to overcome the performance bottlenecks in state-of-the-art implementations. IAAT consists of two stages, the install-time stage and the run-time stage. In the run-time stage, IAAT tiles matrices into blocks to alleviate boundary processing. This stage utilizes an input-aware adaptive tile algorithm and plays the role of runtime tuning. In the install-time stage, IAAT auto-generates hundreds of kernels of different sizes to remove pack operations. Finally, IAAT finishes the computation of small GEMM by invoking different kernels, which corresponds to the size of blocks. The experimental results show that IAAT gains better performance than other BLAS libraries on ARMv8 platform.

IAAT for different platforms

June 2021 - now

Our previous work only focused on ARMv8 platform, and was not suitable for other platforms. Therefore, we reaserch the common method and implementation of small GEMM for different platforms. Besides, we plan to utilize JIT(just-in-time) to optimize small GEMM.

High-performance implementation and optimization of trigonometric functions based on SIMD

August 2020 - April 2021. The paper is published in 计算机科学(Computer Science)

As basic mathematical operations, the high-performance implementation of trigonometric functions is significant to the basic software ecology of the processor. The current processors have adopted the SIMD architecture. Therefore, the high-performance implementation of trigonometric functions based on SIMD has important research significance and application value. We use numerical analysis method to implement and optimize five commonly used trigonometric functions sin, cos, tan, atan, atan2. We design efficient trigonometric function algorithms according to the analysis of floating-point IEEE754 standard. Then, we imporve algorithm's accuracy by the application of polynomial approximation algorithm. eg. Taylor formula, Pade approximation and Remez algorithm. Finally, we promote the algorithm performance by using features of hardware. The experiment shows that we gain 1.77-6.26 times faster than libm library, and 1.34-1.5 times faster than ARM_M library on the ARMv8 platform.

C	++++	C++	++++	Assembly	+++	SIMD	++++	CUDA	+++	MPI	+++
pthread	+++	Makefile	+++	CMake	+++	Latex	++++	Linux	++++	Git	++++
Python	+++	Java	+++	JavaFX	++++	Android	++++	SQL	++		

Python Java JavaFX Android SQL