

DNN paper outline

Marco Maohua Zhu

March 6, 2014

1 Introduction

Stochastic gradient descent is a widely used method for deep neural network training. [1]
Left blank because this is an outline.

2 Background

DNN is hot in machine learning research. SGD is the most common training algorithm but it is difficult to tune and parallelize. CG and L-BFGS, though more complicated, now can be used thanks to the availability of high performance multi-core CPUs, GPUs and computer clusters.

Challenge: too many network parameters, hard to compute.

We need an optimized implementation on suitable (heterogeneous) platforms.

3 Motivation

There are several available architectures to build a heterogeneous system for deep learning.

It is not trivial to map deep learning algorithms on heterogeneous systems.

Researchers and companies need an analysis of

advantages and bottlenecks of each platform given the deep learning task.

4 Mainstream architectures

An introduction to multi-core CPU(SIMD in CPU, SSE/AVX), APU and discrete GPU architecture.

5 Implementation and optimization

How we map the algorithm to APU and discrete GPU.

6 Evaluation results and analysis

Performance, power and performance per watt. Present bottlenecks of each platform.

7 Further analysis

If there is an alternative choice which seems better, do an experiment.

8 Conclusion

Insights.

References

- [1] Le Q. V., Ngiam J., Coates A., Lahiri A., Prochnow B., and Ng A. Y. On optimization methods for deep learning. *Proceedings of the 28th International Conference on Machine Learning*, 2011.