

---

# Parallel Computing For Machine Learning

## PARALLEL COMPUTING FOR MACHINE LEARNING

In the new global economy, Machine Learning has become a central issue for most research field as it offer a techniques to deal with tremendous demanding real-world problem. The study of machine learning is important for addressing fundamental question related to scientific and engineering. There is three subdomain of machine learning whereas; Supervised learning in which the training will be held only if the data have been labelled and consist of input and desired output, Unsupervised learning where the training data didn't need any labelled and the environment only produced input without specific targets and lastly reinforcement learning which the characteristic of information available in the training data stand between supervised and unsupervised and this kind of learning is happening from feedback received through interactions with external environments.

Both supervised and unsupervised learning is suitable for data analysis while reinforcement learning is suite to handle problem that consist of decision making process. As the rapid emerging of machine learning trends, the need for improving the traditional machine learning to the modern machine learning is very important. The improvement in term of software(algorithm) and hardware is the key in order to come out with advanced machine learning that able to deal with current machine learning issued. In several advanced learning method have been mention in order to improve the traditional machine learning including; Representation learning, Deep learning, Distributed and Parallel learning, Transfer learning, Active learning and Kernel-based learning.

The parallel learning is basically based on the parallel computing environment. Parallel computing defined as a set of interlink process between processing elements and memory modules. In machine learning, parallel computing have improved the traditional machine learning by implemented the used of multicore processor instead of single processor[2]. Some researcher have discussed and applied the parallel computing in order to deal with machine learning issued. Qiu et. al. come with a review paper that discussed on how big data have been processed by using machine learning. As data become big and complex, traditional machine learning face difficulty to train those data. Therefore, six advanced learning methods have been introduced as stated previously. Then, five issued of machine learning in big data have been discussed.

One of the issues including understanding large scale of data. In order to solve this issued, distributed framework based on parallel computing is suggested. Alternating direction method of multipliers(ADMM), the framework that can produce algorithm with capability to scatter and scaling is very suitable. ADMM able to split multiple problems and helps to identify solution by coordinating those solutions into smaller group of problems. Next, the used of parallel programming methods have also been mentioned to solve the large-scale of data sets issued. After that, Memeti et. al. review on two techniques for scheduling parallel computing systems which is Machine Learning and Meta-heuristic techniques. As parallel computing usually involved with problem that is very complex and resource intensive so, the used of multiple processing unit is needed.

---

The combination of CPUs, GPUs and FPGAs is always being done by researcher to come out with higher performance, and energy efficiency environments. In this paper, Memeti et. al. also discussed on several popular algorithms for machine learning; Linear Regression(LR), Decision Tree(DT), Support Vector Machine (SVM), Bayesian Inference (BI), Random Forest (RF), and Artificial Neural Network (ANN). Then, Memeti et. al. explains characteristic of problem that can be handle by either machine learning and meta-heuristic techniques based on suitable approaches.

Next, the framework that based on parallel computing architecture have been developed. Chiao et. al. come with paper that aim to train real-time trading model by collecting a set of similar historical financial.

## Current method used

AdaBoost algorithm have been improved to joint-AdaBoost algorithm and applied with Open Computing Language (OpenCL) parallel computing platform. OpenCL is an open standard framework that designed for multiple platforms. Its provided an architecture for parallel computation and build with API, completed programming language, programming libraries and a runtime. Then, it allowed implementation of complex algorithms by user. OpenCL consist of CPU and GPGPU (General-purpose computing in processing unit) that support data and task parallel programming model. Besides, Pratama et. al. come with paper that help reorganizing Japanese handwritten image using combination of Auto-Encoder algorithm and Residual Block framework with CUDA architecture that support programming library and rely on parallel computing. The used of OpenCL in Chiao et.al and Residual Block in Pratama et. al. has proved that parallel computing framework will helps to reduce the time taken for process of training data. Next, both paper also find out that the used of CPU is more efficient compared to GPU.

Then, Cybenko analyse on how parallel computing been applied in machine learning for handling social network. Cybenko suggested three critical ingredients in order to identify either applying machine learning is compulsory:

- dealing with large number of data sets,
- types of software that available to implement those learning process,
- related computing cycles.

As social network dealing with large number of data sets, therefore applying machine learning is compulsory. Furthermore, current machine learning used single core GPU which is time consuming. Therefore, Cybenko suggested the used of multiprocessor either CPU or GPU is very helpful in order to reduce the training time as social network have a lot of data and need to be handle quickly and efficiently. In the experiment, Cybenko also figure out that the used of CPU is more efficient compared to GPU. Table 1 below will conclude all off five paper that have discussed regarding the application of parallel computing in machine learning.

## References:

1. G. Cybenko, "Parallel Computing for Machine Learning in Social Network Analysis," 2017.
2. M. I. Jordan and T. M. Mitchell, "Machine learning: Trends, perspectives, and

---

prospects,” *Science* (80-. ), vol. 349, no. 6245, pp. 255–260, 2015.

3. J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, “A survey of machine learning for big data processing,” *EURASIP J. Adv. Signal Process.*, vol. 2016, no. 1, 2016.
4. J. Byrne et al., “A Review of Cloud Computing Simulation Platforms and Related Environments,” *Proc. 7th Int. Conf. Cloud Comput. Serv. Sci.*, no. April, pp. 679–691, 2017.
5. S. Memeti and J. Kołodziej, “A Review of Machine Learning and Meta-heuristic Methods for Scheduling Parallel Computing Systems,” 2018.
6. C. Chen, S. Huang, and Y. Chang, “Decision Support System for Real-Time Trading based on On-Line Learning and Parallel Computing Techniques,” 2016.
7. M. O. Pratama and P. Kareen, “Reconstructing Japanese Handwritten Images Using Auto-Encoder with Residual Block in Parallel Computing,” no. ICELTICs, pp. 231–234, 2017.

gradesfixer.com