UTM Computing Proceedings
Innovation in Computing Technology and Applications
Volume: 2| Year: 2017 |ISBN: 978-967-0194-95-0

# Multi – Back Propagation (MBP) Algorithm for Big Data Application

Lili Madiha Binti Sulaiman[1], Shafaatunnur Hasan[2] Siti Mariyam Shamsuddin[1.2]

[1]UTM Big Data Centre, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia
[2]Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia

lilimadiha@gmail.com, shafaatunnur@utm.my,mariyam@utm.my

**Abstract.** The Multi-Back Propagation (MBP) algorithm is a simplification of Back-Propagation (BP) algorithm that can be used to train Multi-Feed Forward (MFF) networks to solve classification problem. There is various algorithm in MBP created to solve this problem. However, the overall constraint of every big data algorithm are performance of processing. Many researchers found that Graphics Processing Units (GPUs) can provide incredible performance of processing when compared to CPUs for computationally-intensive applications. In specific, the implementation of multi back propagation (MBP) in GPUs can decrease extremely the long training times during the learning process. However, the implementation of GPU programming in data processing are already widely available and they are no need to duplicate the implementation part. In this research, it focusses on how to implement the big data application using MBP-GPUMLib for classification problem. Although it is only focus on implementation, it is still need to compare the performance of GPU over CPU to ensure the speed and time of training process are significantly improved. Our approach is tested based on the database in experimental result. Morever, the speed up achieved with the GPU greatly compare to CPU.

**Keywords:** Multi-Back Propagation, GPUMLib, Big Data, Neural Network, Multi-Feed Forward.

## 1 Introduction

Big data is dataset that are large or so complex to analyses that the traditional data processing is insufficient. Nowadays, the dataset is become growing rapidly because the increase using the technology such numerous used on mobile devices, software logs, cameras, microphones, radio-frequency identification (RFID) readers and wireless sensor networks and that become a big data problem. Dataset may consist of classification, clustering, prediction and other type pattern of data. This research, will be more focus on classification pattern dataset. According to N. Lopes (2014), the current technology memory original central processing unit (CPU) implementation take a long-time speed processing before the data become a significant consideration. However, since the Graphical Processing Unit (GPU) become more popular in market and started used in many areas and many researchers found that Graphics Processing Units (GPUs) can provide incredible performance of processing when compared to CPUs for computationally-intensive applications. Although the GPU provide incredible performance of processing, the result of speed latency may be different with CPU processing if analyses the different pattern of dataset and different algorithm.

Therefore, the purpose of this research is to is to analyses the classification dataset using the GPU over CPU and compare their speed performance using Multi-Back Propagation (MBP) Algorithm. This research also will come out with the result analyst that already integrate with GPUMLib by implementation in the desktop application. The outcome of this research also will compare the result with related algorithm which is deep learning.

This paper is organized as follows: in Section 1, we present the introduction of this paper. Then, Section 2 explains some of related work of this research. While in Section 3, we provide the research methodology that used in this research project. Next, in Section 4 we discuss the experimental result of this research and in Section 5, we present the discussion of the results that are obtained in our experiments. Finally, in Section 6 provides the conclusion of this research projects.

## 2 Related Work

Due to the increase of data and nowadays the more difficult to process data when using the traditional methods and this affect the performance speed that takes many decades to process it. Several NN that implementations using GPU have been reported in literature review that show the significant performance over the traditional methods. However, the first work that prove the speed up of GPU over CPU already prove by N. Lopes and B. Ribeiro (2011) where a GPU implementation of the Back Propagation (BP) and Multi-Back Propagation (MBP) algorithms. A significant speed performance was achieved compare to CPU implementations. Shafaatunnur Hasan et. al (2014), were together added a GPU implementation of MBP algorithm to GPU Machine Learning Library (GPUMLib) for data classification.

To solve the big data mining in Neural Network (NN), we need the classification technique which is use MBP algorithm to analyses the speed processing compare with Graphical Processing Unit (GPU) and Central Processing Unit (CPU). In this research, explained more details about the NN learning technique which is category by supervised learning and unsupervised learning. Also, will be describe about the detail of MBP Algorithm and how the flow of process data.

## 3. Methodology

### 3.1 MBP-GPUMLib for Big Data Classification Problem

This section presents the methodology adopted in this research.

### 3.1.1 MBP-GPUMLib framework

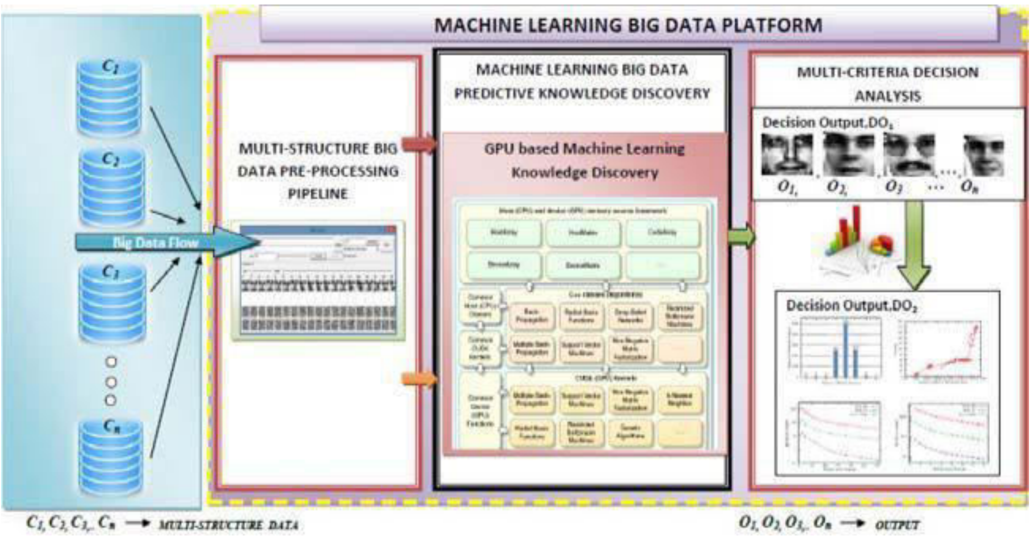
MACHINE LEARNING BIG DATA PLATFORM
$C_1$
$C_2$
$C_3$
$C_n$
Big Data Flow
MULTI-STRUCTURE BIG DATA PRE-PROCESSING PIPELINE
MACHINE LEARNING BIG DATA PREDICTIVE KNOWLEDGE DISCOVERY
GPU based Machine Learning Knowledge Discovery
MULTI-CRITERIA DECISION ANALYSIS
Decision Output,$DO_1$
$O_1$, $O_2$, $O_3$ … $O_n$
Decision Output,$DO_2$
$C_1, C_2, C_3, C_n$ → MULTI-STRUCTURE DATA
$O_1, O_2, O_3, O_n$ → OUTPUT

**Figure 1** Machine Learning Big Data Framework (S. Hasan et. al ,2014)

From the Figure 1 above is a framework of machine learning for big data. The framework is started from structure of the various input from classification pattern data and then multi structure big data pre-processing pipeline follows by post-processing machine knowledge discovery which is will come out the result or the output from the process of data.

### 3.2.1 İnput-Process-Output Flow

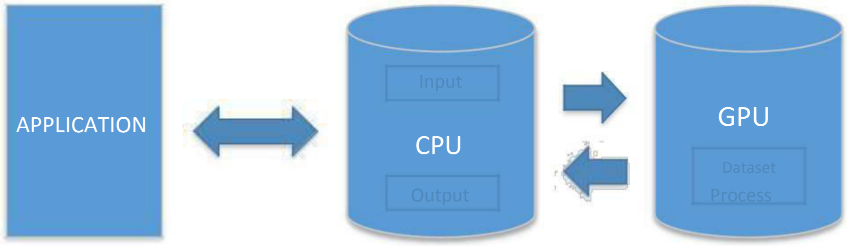

**Figure 2** Framework of Big Data on MBP-GPUMLib

From the Figure 2 above is about the general framework of big data when pre-processing and post-processing in MBP-**GPUMLib. İnitially,** the input data is the dataset that we want to analyze from CPU send to GPU which to process the dataset and then will send back the output to CPU. Basically, it is the input process and then output will the result of process data.

## 4 Experimental Result

In this research, we conduct several experiments using 322 of image processing using GPUMLib and testing in desktop application that already integrate with GPUMLib. The result is based on Table 1, the speed performance to process data using the GPUMLib - MBP algorithm which take time only 0.245 s for 1000 cycles. The accuracy (83.20%) where accurate each of data. The (20%) not accurate which is may because of the noise of the data. The input data which 49 features of data and for the layers to train and test is set to 10 nodes with one hidden layers. The training set samples consists of 276 samples and the testing set consist of 46 samples.

**Table 1:** Experimental Setup and Result

| | | | |
|---|---|---|---|
| Database | Animal Image Database | | |
| Type of Devices | GeForce GTX 960 (1291 Mhz) - supports CUDA 5.2 | | |
| Experimental Setup | | Experimental Result | |
| Topology | 49-10-1 | Time (s) | 0.245s |
| Training Sample | 276 | Accuracy (%) | 83.20% |
| Testing Sample | 46 | F-Measure | 82.07% |
| Epoch | 1000 | | |

## 5 Discussion

One of the main challenge in performing image processing where managing and analyzing the processing of big data. Data may contains a large number of redundant and irrelevant feature that will influences the accuracy of the result during the testing. For our experiments we use Animal Image dataset that contains 322 items of each data represent **the features. İn thisexperiments, we use 49 features to test the accuracy of image processing.**

Based on the result, overall we found the speed performance of the experiment using features selection method which is GPU is more speed up that CPU. Also, the GPU is more accurate that CPU. Furthermore, we also implement and develop dekstop application that integrated with GPUMLib and it integrate using the command that have provided.

## 6 Conclusion

In this research, we investigate the speed performance between CPU and GPU based on GPUMLib machine library using the big data which is animal image processing. Each image contains each attribute and features that contrast each data. To test this speed performance, we analyze this data for data cleaning which is pre processing using NMF algorithm. After achieve the result, we compare the speed performance between CPU and GPU.

Based on the result that obtained form the experiments, we found that GPU processing is much better compare to CPU when to analyze the image processing based on the speed performance and the accuracy using MBP algorithm. The result also depend on how much layer to take during the training and testing. The more layers set between input and output which relate to deep neural networks or deep learning. Thus, the more deep layers, the decrease the accuracy of data because of when training and testing it become more complex.

## References

Brandstetter, and A. Artusi (2008), *IEEE Transactions on Neural Networks* Volume 19 Issue 12, December 2008.

Daniel del Hoyo Rodriguez, (2012). "*Optimization of Back-propagation Learning Algorithm on MLP Network*s" D.1.3, I.2.6,

Emilio Corchado HujunYin (Eds.) (2009), *Intelligent Data Engineering and Automated Learning -IDEAL 2009*, 10th International Conference, Burgos, Spain, September, pp. 23-26, 2009, Proceedings. ISSN 0302-9743

GPUMLib Website, May 16, 2016, http://gpumlib.sourceforge.net.

J. T. Lalis1, B. D. Gerardo2 and Y. Byun (2014), *An Adaptive Stopping Criterion for Backpropagation Learning in Feedforward Neural Network*,International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.8 (2014), pp.149-156, http://dx.doi.org/10.14257/ijmue.2014.9.8.13.

J¨urgen Schmidhuber (2014), The Swiss AI Lab IDSIA Istituto Dalle Molle di Studi sull'Intelligenza Artificiale, University of Lugano & SUPSI, Galleria 2, 6928 Manno-Lugano Switzerland, 8 October 2014 "*Deep Learning in Neural Networks: An Overview*". IDSIA-03-14 / arXiv:1404.7828v1 [cs.NE].pp 2553-2561.

Li Deng (2014). *A tutorial survey of architectures, algorithms, and applications for deep learning* . APSIPA Transactions on Signal and Information Processing, 3, e2 doi:10.1017/atsip.2013.9 Pages e2,29 pages.

N. Lopes and B. Ribeiro (2011), International Journal of Computer Information Systems and Industrial Management Applications, ISSN 2150-7988 Volume 3 (2011) pp. 355-362 © MIR Labs, http://www.mirlabs.net/ijcisim/index.htm.

Noel Lopes and Bernardete Ribeiro (2011). "*An evaluation of multiple feed-forward networks on GPUs*". International Journal of Neural Systems (IJNS), 21(1): pp. 31-47, 2011.

N. Lopes, B. Ribeiro and R. Quintas (2010), "*GPUMLib: A new Library to combine Machine Learning algorithms with Graphics Processing Units," Hybrid Intelligent Systems (HIS)*, 2010 10th International Conference on, Atlanta, GA, 2010, pp. 229-232. doi: 10.1109/HIS.2010.5600028

Noel Lopes and Bernardete Ribeiro (2009). *GPU implementation of the multiple back-propagation algorithm.* In 10th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL 2009), LNCS 5788 , pages 449-456. Springer, 2009.

Noel Lopes and Bernardete Ribeiro. *Hybrid learning in a multi-neural network architecture.* In INNS-IEEE International Joint Conference on Neural Networks (IJCNN 2001) , volume 4, pages 2788-2793, 2001

Noel Lopes and Bernardete Ribeiro (2011), *GPUMLib: An Efficient Open-Source GPU Machine Learning Library*, ISSN 2150-7988 Volume 3 (2011) pp. 355-362 © MIR Labs.

Shafaatunnur Hasan, Siti Mariyam Shamsuddin, and Noel Lopes (2014), *Machine Learning Big Data Framework and Analytics for Big Data Problems*, Int. J. Advance Soft Compu. Appl, Vol. 6, No. 2, July 2014, Copyright

NVIDIA CUDA Programming Guide Version 2.2 (2009).

Wu and Zheng Yi (2014), "*Portable GPU-Based Artificial Neural Networks For Data-Driven Modeling*" (2014). CUNY Academic Works. http://academicworks.cuny.edu/cc_conf_hic/92

Wei, Xun-Kai Li, Ying-Hong Li, Yu-Fei Zhang, Dong-Fang (2008), **"***Enclosing machine learning: concepts and algorithms***"***, Neural Computing and Applications*, ISSN- 1433-3058. DOI- 10.1007/s00521-007-0113-y pp 273-243.