# Machine Learning – Unsupervised Learning

## Introduction

With the rise of artificial intelligence, machine learning has become a popular topic. The earliest concept of machine learning was introduced in the 1950s. After several decades of development, various methods and algorithms for machine learning have become more mature. With the advent of the era of big data, it is the primary goal for relevant scholars to propose learning methods and learning algorithms that are more suitable for the era of big data. In order to solve many complex problems in real world, unsupervised learning in deep learning algorithms becomes a core part of the research of machine learning. Compared to supervised learning, unsupervised learning can be adopted in more extensive fields.

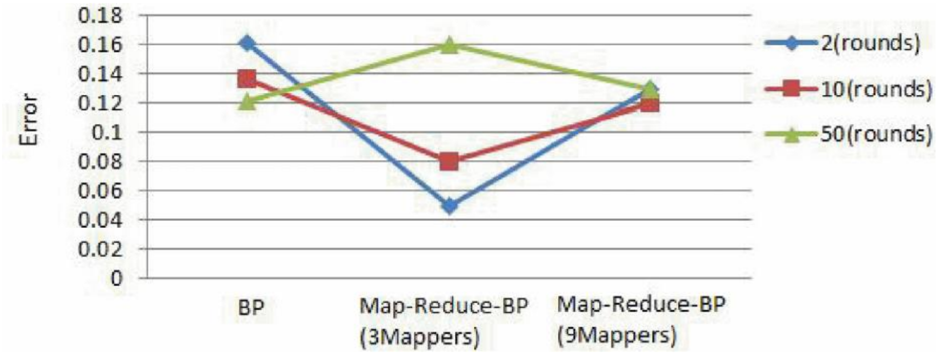
In this report, I mainly focus on some new unsupervised learning methods in a conference paper "Study on Deep Unsupervised Learning Optimization Algorithm Based on Cloud Computing" [1]. This paper was published in 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS). The authors of this paper are Hui Yan, Ping Yu and Duo Long. This paper focuses on the core module of the back-end part of the intelligent agriculture education system project structure - the unsupervised learning neural network module, and studies how to use the effective artificial intelligence technology based on multi-computer cluster for big data analysis. Cloud Computing is a computing model. Its main function is to realize distributed computing of big data, parallel operation of a large number of operations, storage of massive data of the network, virtualization sharing of network resources, resource scheduling and load balancing and data redundancy storage based on hot backup [2]. When the scale of the trained model is relatively large, the training of the model can be accelerated by the data parallelism method [3].

## Analysis

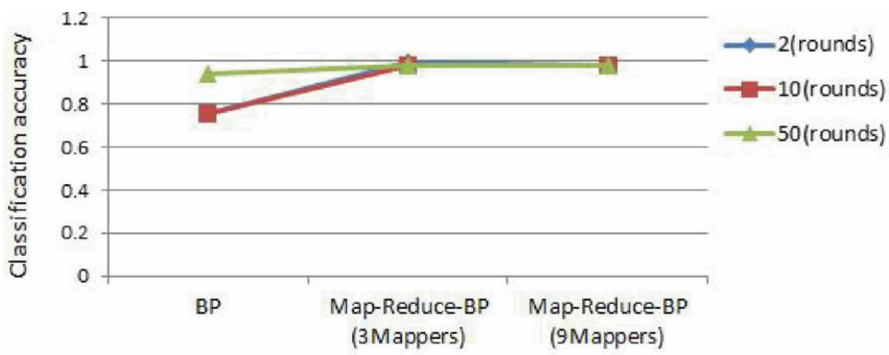
From paper [1], the authors built a CNN-RNN model based on cloud computing technology, and the parallel neural network model divided by training set is adopted to design the batch gradient descent algorithm based on deep unsupervised learning and BP algorithm based on Map-Reduce.

Training neural network with batch gradient descent and data parallel is an efficient way to decrease the cost of time, especially when the number of data is extreme large.

The experiments result in paper [1] is as follow:




Error comparison between serial BP and map-reduce-BP [1]




Classification accuracy comparison between serial BP and map-reduce-BP [1]

## Pros:

From the experiment figures and the results, we can see that overall the error rates of the serial BP algorithm and Map-Reduce-BP algorithm are at the same level and are within acceptable error range. At the same time, the cloud computing clusters costs less time and fewer times of iterations, which means that the Map-Reduce-BP algorithm is more efficient.

## Cons:

(1) Learning accuracy does not increase much compared to serial BP algorithm, and it is not suitable for some high-level-accuracy tasks.

(2) MapReduce in cloud computing also has some problems.

i. The versatility of MapReduce is not very good, and it is not adaptable in many data processing scenarios.

ii. The second is that the operation of Reduce part needs to wait until most parts of Map complete their operations. In some cases, when the distribution of computing is uneven, Map operations will cost a long time and Reduce operation will wait for the complete of Map operation.

Therefore, using Map-Reduce-BP algorithm for learning is not faster than traditional method if the distribution of computing is uneven.

(3) As MapReduce runs on a distributed system, nodes in the system are connected through the network, so a large amount of network message communication is required during the operation of MapReduce. Higher loads can introduce additional communication overhead and even affect the performance and availability of communication between other nodes.

## Recommendation

I recommend this unsupervised learning algorithm to those people who want to analyze a large number of data and the accuracy requirement is not too strict.

## Conclusion

The deep unsupervised learning algorithm based on cloud computing increases the efficiency of learning using neural network. And it is feasible to use cloud computing cluster for unsupervised learning in neural network. But it still has some problems that need further improvement and optimization.

## Summary of group reports

### Jialun Wang's report:

This report focuses on an unsupervised learning algorithm, K-Means algorithm. K-Means is a mainstream algorithm in unsupervised learning. K-Means is convenient and more efficient to process spherical datasets without preset labels. But it requires a appropriate K to get the good results.

### Zeyu Song's report:

This report focuses on object detection and R-CNN methods. This report analyzes the pros of Fast/Faster R-CNN and Mask R-CNN. These methods are keeping improving, together with fast training and inference time.

### Cagri Yoruk's report:

This report focuses on object detection and many models, like YOLO, R-CNN, Faster R-CNN, and Mask R-CNN. Compared them by their architecture, system performance, the implementation styles, and the relation between these models.

### Qing Han's report:

This report focuses on YOLO system. This report analyzes the pros and cons of YOLO compared to other systems, especially in real-time object detection tasks. YOLO is much faster in real-time object detection than other systems.

## Reference


[1] H. Yan, P. Yu and D. Long, "Study on Deep Unsupervised Learning Optimization Algorithm Based on Cloud Computing," 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), Changsha, China, 2019, pp. 679-681.

[2] Yan Hui, "long Duo. 3D scanner-based corn seed modeling[J]", Applied Engineering in Agriculture, vol. 32, no. 2, pp. 181-188.

[3] J Dean, G S Corrado, R Monga et al., "Large scale distributed deep networks[C]", Proc of the 25th International Conference on Neural Information Processing Systems. [S.I.]: Curran Associates Inc, pp. 1223-1231, 2012.