Deep Learning Methods for Big Data Analytics

Renuka Devi.D[1], Diana Judith.I[2]
[12]*Assistant Professor,*
[12]*Department Of Computer Science, Stella Maris College, Chennai, India*

Abstract- Deep Learning (DL) is one of the most prominent methods in machine learning. DL has acquired its application in the field of image mining, text, speech, pattern recognition and so on. In the past few years, DL has played an imperative position in the field of big data analytics. The conventional data processing methods are not suitable for handling voluminous data. To process and analyze big data, there is a need for highly optimized algorithms and methods. DL techniques are used to process huge real-time data with higher efficiency. Recent researches on DL, have thrown a light on different hybrid mechanisms to process real-time big data in a most competent way. This paper outlines the applications of DL in the field of big data analytics and further research directions.

Keywords- Big data, Deep Learning, real-time data, analytics, data processing

I. INTRODUCTION

Big data is defined as a huge voluminous data with higher acquisition rate and higher velocity [1]. Every day, a huge volume of data is generated by multiple sources like transactions, social networks, mobile devices, sensor data etc., thus, the variety of different data sources gathered at a very high frequency in unknown and unpredictable ways.

Big data is identified by its characteristics such as velocity, volume, and variety. There is a need for new technologies and methods to handle big data. Nowadays, big data analytics has fascinated more and more research development in the field of analytics [2].

The big data is not processed by the conventional methods, as the dataset size is beyond the processing capacity of an ordinary processing system. Four core components are emphasized in the big data analytics, which are gathering data, storage space, administration, and analysis [3].

DL is a kind of machine learning (ML) methods which enables the computers to learn from experiences and recognize the objects by using a hidden functionality. Since, it gathers the knowledge from the experiences during the training phase of data collection, and applies to the unknown data in future.

This minimizes the need for a human computer operator properly to detail all of the knowledge required by the computer [5].Big data analytics is intended to derive meaningful insights from a huge volume of data. Many recent researches proved that the analytical methods perform well in terms of efficacy such as increased processing speed, accuracy, and adaptability to real-time streaming date. Various research studies were carried out in the domains of image, data analytics, natural language processing (NLP) and so on. The recent development in computer visions, DL methods are effectively applied to large (CNN)-scale visual recognition challenge (ILSVRC) by ImageNet [6].This paper emphasizes the need for emerging trends in DL and its applications in the domain of big data analytics. The concepts of big data analytics also outlined in this work.

II. BIG DATA ANALYTICS

A. Characteristics

The characteristics of big data are defined in terms of 5 V's, Volume, Variety, Velocity, Variability, and Value. It is shown in the fig.1.

Fig.1: Characteristics of Big data

Volume:

In the contemporary era of big data, trillions of data are generated by different applications that include social media, high- resolution images, genomic data, geospatial data and video applications.

Variety:

The different types of data include structured (conventional databases), unstructured (media tweets, tags), semi-structured (e-mail) are termed as variety.

Velocity:

It is the rate at which data are captured. A diverse type of data arrives at a different speed at different intervals of times. The streaming real-time data are generated at a rapid speed and they have to be processed as soon as they are captured [7].

Value:

The creation of value from Big Data –This is the expected outcome after processing and analyzing huge data. The derived insights are converted into meaningful decisions or strategies.

Veracity: This is uncertainty that exists in data.

B. Architecture

The Big data architecture is shown in the Fig.2. The components are storage, processing methods, analyze and visualization.

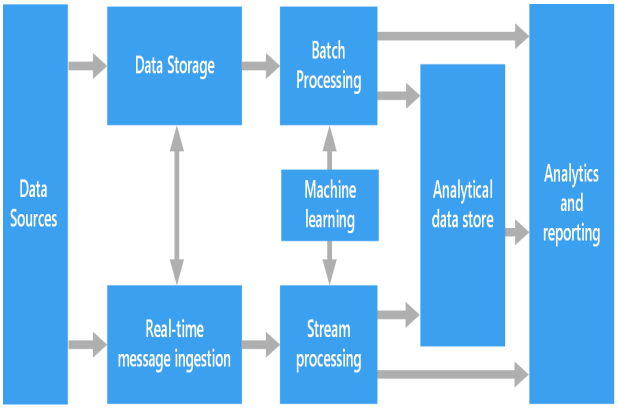

Fig.2: Big data Architecture

Data sources:
Big data are captured from different sources like social media data, health data, real-time streaming data, sensor data and so on.

Data storage:
The challenging issue in big data analytics is the storage space. Normally the huge volume of data is divided across different nodes using the distributed file system. It can hold the data of different file formats and variety.

Batch processing:
The big data are processed as a batch of multiple jobs and executed for a long run. It is appropriate only to the enormous data but not the streaming data. Typically these jobs engage interpretation of source files, processing, and deliver the output to original files.

Stream processing:
The data streams from different sources arrive at a very high speed, they are actually processed as soon as they arrive without storing them in the system.

Analytics and reporting:
Analytics involves extracting the meaningful information from the voluminous data. Many data tools are available for both processing the data and for visualizing the data. Some of the tools are hadoop, spark, flink, storm, and samza.

III. OPEN ISSUES IN BIG DATA

The various challenges related to big data analytics is still emerging as par with recent developments around the world, where every second million of data were generated. To address these challenges, significant research efforts are required to advance the competence of data collection, data storage, data analysis and visualization.

IV. DEEP LEARNING MODELS

DL is an emerging branch of supervised learning method in machine learning. For example, in neural network, for a given neuron function, the activation function is applied on the deep layers to extract the abstractions from voluminous data. This is similar to a hierarchical structure where deep learning models are applied [8]. In the research domain, DL is a very promising and evolving area. DL has many highly developed algorithmic models such as convolution neural networks (CNNs), deep Boltzmann machines (DBMs), deep belief networks (DBNs), deep representation, recursive auto-encoders and restricted boltzmann machines (RBMs)[9].
Generally, DL models are applied to the huge volume of unsupervised data, to automate the extraction from data. DL has combined with allied domains such as artificial intelligence, which simulate the human brain function to analyze, learn and extract the meaningful insights from the data gathered. The research works addressed towards this challenge, has been a key objective to develop DL algorithmic models [10].

V. APPLICATIONS OF DEEP LEARNING IN BIG DATA

The Deep Belief Network (DBN) [11] has been applied to both labelled and un-labelled data. DL models are constructed using both supervised and unsupervised learning techniques. There are 3 layers namely the input layer, hidden layer, and output layer. The nodes of each layer are connected to the nodes of the next layer but not connected to the nodes of the very same layer. DBN is used to solve many big data application issues. The overall architecture is shown in the fig.3.

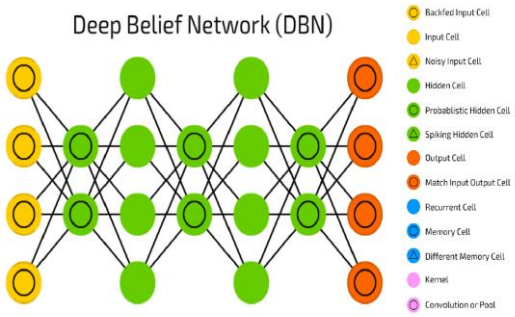

Fig.3: DBN

A Convolutional Neural Network (CNN) has many hidden layers. Each layer performs certain functions; either it is map layer or classification layer as shown in the fig.4. The convolution layer applies the convolution function to the input given to the layer and forwards the output to the next layer. In CNN, a large number of deep learning methods are connected locally.

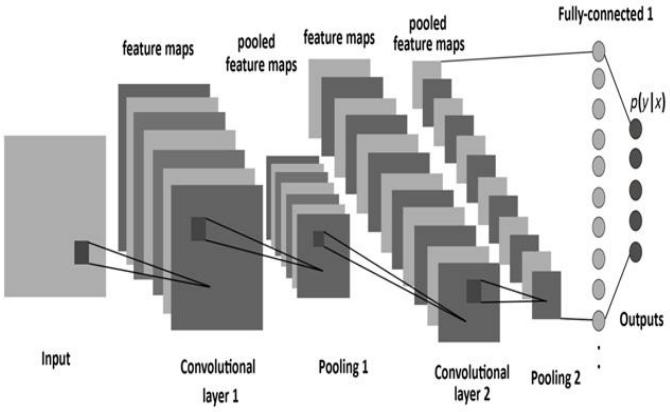

Fig.4: CNN

Pooling methods are applied to the outputs of neuron to combine as the one cluster[12]. In a fully connected network, every neuron is connected to other neurons in the next layer. The weights are assigned to each neuron [13].

It computes the output function depending upon the weights assigned to it [14]. CNN is potentially used for handling a large volume of data with higher accuracy and speed. Correspondingly, most of the research work proves that the efficacy of this algorithm in the field of big data analytics.

VI. CONCLUSION

Deep Learning and Big data analytics are significant facets of Data Science [15]. The knowledge acquired from DL algorithms is successfully applied to big data analytics. Here two aspects of potential issues are discussed, one how DL methods are applied to specific problems of big data and how it could be improved to address the future challenges. The allied domains of big data, such as computer vision and speech recognition, have seen the application of Deep Learning in the larger extent to produce higher accuracy classification results within an optimized time complexity.

VII. REFERENCES

[1]. H J Watson, Tutorial: Big data analytics: Concepts, technologies, and applications, Communication of the Association for Information Systems, Vol.34, Article 65, pp.124–168, 2014.

[2]. F. J. Alexander, A. Hoisie, and A. Szalay, "Big data," Computing in Science & Engineering, vol. 13, no.6, pp. 10-13, 2011.

[3]. J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition, and productivity," McKinsey Global Institute, Tech. Rep., May 2011.

[4]. Renuka Devi, D. Sangeetha.T.G.R, "A Review on Big Data Analytics: An Eminent Approach for Handling an Outsized Data", Vol. 6, Special Issue 11, pp. 515- 521, September 2017.

[5]. ILSVRC2012. Large Scale Visual Recognition Challenge 2012 [Internet]. [Updated 2013-08-01]. Available from: http://www.imagenet. Org/challenges/LSVRC/2012/

[6]. Good fellow, Ian, et al. Deep learning. Vol. 1. Cambridge: MIT press, 2016.

[7]. I. O'Reilly Media, Big Data Now: 2014 Edition, O'Reilly Media, 2014.

[8]. Hinton, G. E., Osindero, S., and Teh, Y. W. "A fast learning algorithm for deep belief nets",Neural Computation. 2006;18(7):1527–1554

[9]. Fong, Simon, Suash Deb, and Xin-she Yang. "How meta-heuristic algorithms contribute to deep learning in the hype of big data analytics." Progress in Intelligent Computing Techniques: Theory, Practice, and Applications. Springer, Singapore, 2018. 3-25.

[10]. Najafabadi, Maryam M., et al. "Deep learning applications and challenges in big data analytics." Journal of Big Data 2.1 (2015): 1.

[11]. Jan, Bilal, et al. "Deep learning in big data Analytics: A comparative study." Computers & Electrical Engineering (2017).

[12]. Wei, Yunchao, et al. "Hcp: A flexible cnn framework for multi-label image classification." IEEE transactions on pattern analysis and machine intelligence 38.9 (2016): 1901-1907.

[13]. Hoo-Chang, Shin, et al. "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning." IEEE transactions on medical imaging 35.5 (2016): 1285.

[14]. Zheng, Zhedong, Liang Zheng, and Yi Yang. "A discriminatively learned cnn embedding for person reidentification." ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) 14.1 (2017): 13.

[15]. Chen, Xue-Wen, and Xiaotong Lin. "Big data deep learning: challenges and perspectives." IEEE access 2 (2014): 514-525.