Deep Learning In Big Data Analytics For Share Market

Dr. S. Prem Kumar[1], Sahithi .P[2], M. Praneeta Sony[3]
[1, 2, 3] CSE Department,
[1, 2, 3] GPCET (affiliated to JNTUA, Anantapur), Kurnool, India

Abstract- *Big data is an oftenly used technology that is the process for structured, and unstructured data. Deep learning is a new area of Machine learning research, which has been introduced with the objective of moving Machine learning closer to one of its original goals i.e., Artificial Intelligence. Deep learning is an essential hierarchical learning process; it can extract high level complex abstractions. For finding large amount of data we use the concept of deep learning for solving problems in marketing, technology, medical science. Deep learning algorithms are used in analyzing large volumes of unsupervised data, making deep learning tools the most suitable ones for big data analytics. The main objective of this paper is introducing deep learning and big data analytics in stock marketing. Deep learning has an ability to extract features of unsupervised data for attractive stock market predictions at high frequencies. Big data is a trendy technology used in many sectors; one of the sectors is financial market. Day to day evaluations in increase of product rates is predicted by big data analytics.*

Keywords- Deep Learning, Big Data,Share Market,KNN Algorithm.

I. INTRODUCTION

Deep learning is a machine learning technique that explains how computers understand humans. It can be applied in various fields such as speech recognition, information extraction, machine translation, natural language processing. Recent advances in deep learning have improved to the point where deep learning outperforms humans in some tasks like classifying objects in images. There are two main reasons :

1. Deep learning requires large amount of labelled data.
2. Deep learning requires substantial computing power.

Some of the applications of deep learning are automated driving, stock market, aerospace and defence, medical research, industial automation. Most deep learning methods use neural network architectures, in which how deep learning models are referred as deep neural networks. A single neuron in the brain, receives signals as many as 100000 from other neurons. When those other neurons are fired, that exert either an excitatory or inhibitory effect on the neurons they connect to.

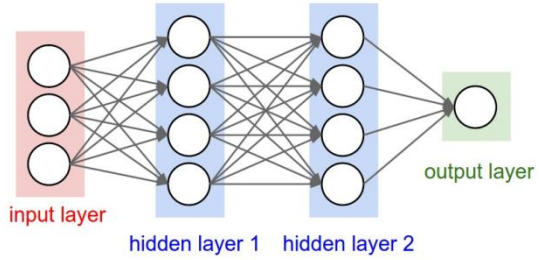

Figure 1.

For the use of big data, we require many programs that use various physical or virtual machines working together. Two largest sources of data in huge quantities are transactional data, includes everything from stock prices to bank data. One of the best methods for changing raw data to useful information is called as Map Reduce. Data analytics is a state of being the importance of big data and big data analytics are available to all the parts of the organization. Some of the big data tools used are Apache Beam, Apache Hive. Industries that have appropriate use of big data include financial services, technology, marketing and health care. The affect of big data continues to redefine the competitive landscape of industries.

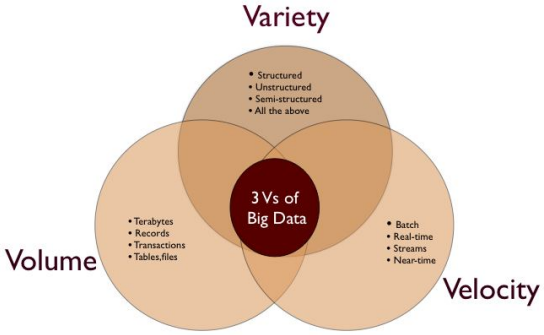

Figure 2.

3V's (volume, varity and velocity) are three defining properties or dimensions of big data. Volume refers to the

amount of data, variety refers to the number of types of data and velocity refers to the speed of data processing.

II. THREE APPLICATIONS OF DEEP LEARNING IN BIG DATA ANALYTICS

For complex representation of data, deep learning algorithms have great potential. Deep learning in big data analytics has become a high focus of data science. Now a day's big data has become important in many organizations as they are collecting massive amounts of domain specific information that can be used to solve problems related to national intelligence, cyber security, fraud detection, and marketing. The key benefit of deep learning in big data analysis can be learnt from massive amounts of unsupervised data. Deep learning algorithms can address the following applications:

A. Semantic Indexing

The key task which is associated with big data analytics is information retrival.The problem in big data analysis is efficient storage and retrieval of information, as data in large scale quantities such as text, image, video and audio is being collected and made available across various domains. Strategies and solutions which are previously used for information storage and retrieval are challenged by massive volumes of data. Semantic indexing proves to be efficient and making search engines work more quickly and efficiently. In indexing of data, data representations play an important role. We should notice that typical level abstract data representations need to be synonymous and demonstrate semantic association in order to understand the comprehension of input.

B. Conducting Discriminative Tasks

In big data analytics while performing discriminative tasks, learning algorithms allow users to extract complicated non linear features from the raw data. The two advantages of this approach: Firstly, by extracting features with deep learning adds nonlinearity to the data analysis and associating discriminative tasks closely to Artificial Intelligence and secondly applying linear analytical models on extracted features is more efficient, which is important for big data analytics. The major benefit of using deep learning in big data analytics that allows users to accomplish difficult task that is related to Artificial Intelligence such as object recognition, image comprehension etc.Microsoft Research Audio Video Indexing(MAVIS) generates closed captions and keywords that increases discovery of audio and video files with speech content.

C. Image and Video Tagging

This approach helps in extracting useful features for performing discriminative tasks on image and video data. Deep learning can also be used for action scene recognition and video data tagging. Deep learning has been successful in achieving remarkable results in extracting useful features. Deep learning mechanism can facilitate segmentation and annotation of complex image scenes.

III. CHALLENGES OF DEEP LEARNING IN BIG DATA ANALYTICS

The below two challenges are used for modifying and adapting deep learning :

A. Incrementing Real Time Non Stationary Data

Big data analytics deals with fast moving generation of input data, such type of data analysis is useful in monitoring tasks such as fraud detection. The feature of incremental learning and mapping improves generative object functions. By adding features mappings to the existing features are updated and can be merged. This kind of incremental feature is useful in applications where the distribution of data gets changed according to the time in massive online transactions. Deep learning algorithms handle large scale of online real time non stationary data.

B. High Dimensional Multimodal Data

Generally deep learning algorithms in big data are expensive when dealing with high dimensional data. Convolution neural networks are another different method that improves performance on high dimensional data. In this neural network, neurons, in the hidden layer units need not be connected to all the nodes that are in same spatial area. The hidden nodes allow closed form solution with substantial speed up, which greatly simplifies the model selection process. A high dimensional data source contributes the volume of raw data by adding complicating learning to the data.

IV. FIVE PREDICTIONS FOR THE FUTURE OF DEEP LEARNING

The potential of deep learning appear to be limitless, but developers are still resolving how to put in the work.

A. Standard Tools Adopted By The Deep Learning

The deep learning community will intersect on a core set of tool framework, by the end of this decade. Presently, deep learning professionals have a glut of tooling options, most of which are open source. The most popular tooling options are Tensor Flow, MXNet, Torch, Caffe etc.

B. Deep Learning Will Gain The Native Support Within The Spark

In the next 12 to 14 months,the spark community will beef the platform's native deep learning capabilities.From the recent sessions at the recent spark submit,it appeared that the community is learning towards stronger support for TensorFlow and least with Caffe,Torch.

C. Deep Learning Tools Will Incorporate Simplified Programming Frameworks

For fast coding of the core algorithmic capabilities with fewer line of code,the application developers community will insist on APIs and other programming abstractions.Deep learning developers will adopt integrated,open,cloud based development environments.The tools will automate more deep learning development pipeline functions,present a collaboration and sharing paradigm.

D. Deep Learning Toolkits Are Supported for Visual Development

Deep learning toolkits will consolidate interchangeble capabilities for easy of visual design,configuration and training new models from preceeding building models.Many of the reusable components will be derived through "transfer learning".Reuse of deep learning, incorporated into classic libraries and interfaces, will be of future representstions.

E. Deep Learning Toolkits Will Be Embedded in Design Interface

In the upcoming 6 to 12 years deep learning tools like libraries and languages will be the classic components for every software development toolkit .Deep learning development capabilities will be fixed in productive design tools used by artists, designers, architects. As deep learning market overture towards group maintainence it will follow in the track of data visualization, business intelligence.

V. SHARE MARKET

Share market means all the shares or stocks of company are brought together and sold in exchange.A market place where buyers and sellers are met to purchase and sell the company which are listed to them.There are two major exchanges in India,National Stock Exchange(NSE),Bombay Stock Exchange Limited(BSE).

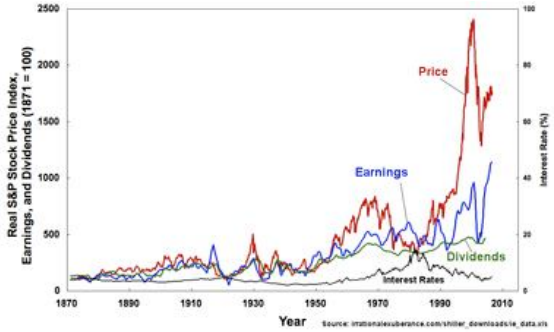

Figure 3.

VI. MARKET PARTICIPANTS

A. FII

This term is used to refer to outside companies investing in the financial markets of India.FIIs play very important role in the market.FIIs have more financial power,this has the power to change the direction of stock market.

B. Propritetary

This term indicates firms,who trade on their own account like brokers etc.

C. DII

Domestic Institutional Investors which undertake investment in securities and other financial assets.

D. Clients

Retail investors like individuals.

VII. ANALYSIS AND PREDICTION OF STOCK MARKET

Stock market prediction is to estimate the future value of a company stock or other financial exchanges. The efficient-market hypothesis suggests that stock prices reflect all currently available information and any price changes that are not based on newly revealed information thus are inherently unpredictable.

There are two basic types of stock analysis: Fundamental analysis and Technical analysis. Fundamental analysis concentrates on data from sources including financial records, economic reports, company assets, and market share. a stock analyst may use the debt ratio, which is calculated by dividing total liabilities by total assets. Technical analysis focuses on the study of past market action to predict future price movement. Technical stock analysis is effective only when the price trend analyzed is influenced by supply and demand forces.

Figure 4.

A. APPLICATION ANALYSIS OF STOCKS

Stock market includes routine activities like sensex calculation,exchange of shares.The software will be examine the sensex based on companies stock value.It depends on various factors

a) Demand and supply:

The major reason is changing of prices amd stocks.When demand increases and supply is less,price raises and vice versa.

b) b) Corporate results:

This will be regarding to the profits or progress of the company over a span of time.

c) c) Popularity:

Popularity of a company can effect on buyers.Stock value depends on other factors as well,but we are taking into consideration only these main factors.

VIII. STOCK MARKET PREDICTION BY USING KNN ALGORITHM

k-Nearest Neighbor Classifier (kNN) K-nearest neighbor technique is a machine learning algorithm that is considered as simple to implement (Aha et al. 1991). The stock prediction problem can be depicted into a similarity based classification. The historical stock data and the test data is mapped into a set of vectors.N represents vector for stock features. Euclidean distance is calculated to take a decision.kNN is considered a indolent learning which do not build a model or function previously, but provides the closest k records of the training data set that have the highest similarity to the test . Then, a majority value is performed among the selected k records to estimate the class label and then allocate it to the query record. The analysis of stock market closing price is calculated by using kNN as follows:

1) Calculate the no. of nearest neighbors.
2) b)Estimate the distance between the training samples and the query record.
3) Arrange all training records according to the distance values.
4) Use a majority value for the class labels of k nearest neighbors, and allocate it as a prediction value of the query record.

A. Mathematical Calculations and Visualizations Models

This represents an overall equations that were applied to forecast next day price. The calculations include error estimation, total sum of average error, cumulative closing price when arranged using predicted values, k-values and training RMS errors.

a) Root Mean Square Deviation (RMSD) is precision that computes the differences between the estimated values, Y, and the actual values, X. The total of RMSD is grouped into a single value measure.
RMSD = SQRT(Y-X) 2 .

b) Sum of Squares (ESS) is computed as follows:

$ESS=\sum_{i=1}^{n}(\hat{y}i-\overline{y})$
Where yi is the predicted variable, and y is the actual value.

c) Average Estimated Error (AEE) AEE is the total sum of RMS errors for all variables in stock records divided by the total number of the records.
$AEE=\sum_{i=1}^{n} RMSi/n$

IX. CONCLUSION

There are many areas of bigdata where deeplearning requires more survey to cover them,such as learning with

streaming data distributed computing etc.The hierarchial learning and removal of different levels of data abstraction in deeplearning provides definite degree of simplification for bigdata analytics.In this paper we discussed about how deep learning and bigdata analytics is used in financial market mainly in stock exchange.

REFERENCES

[1] Dumbill E (2012) What Is Big Data? An Introduction to the Big Data Landscape. In: Strata 2012: Making Data Work. O'Reilly, Santa Clara, CA O'Reilly.

[2] https://www.infoworld.com/article/3172554/analytics/6-predictions-for-the-future-of-deep-learning.html

[3] https://www.researchgate.net/publication/316373382_Deep_Learning_Networks_for_Stock_Market_Analysis_and_Prediction_Methodology_Data_Representations_and_Case_Studies.

[4] https://link.springer.com/article/10.1186/s40537-014-0007-7

[5] https://www.allerin.com/blog/3-applications-of-deep-learning-in-big-data-analytics.

[6] http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data.

[7] https://www.technologyreview.com/s/513696/deep-learning/

[8] http://stepupmoney.com/how-to-invest-in-shares-in-india-beginners-guide-to-stock-market-investment-in-india/

[9] http://www.ijbhtnet.com/journals/Vol_3_No_3_March_2013/4.pdf