International Research Journal of Human Resource and Social Sciences
ISSN(O): (2349-4085) ISSN(P): (2394-4218)
Impact Factor 6.924 Volume 9, Issue 11, November 2022
Website- www.aarf.asia, **Email** : editoraarf@gmail.com

DEEP LEARNING: EFFECTIVE TOOL FOR BIG DATA ANALYTICS

PUNAM PRABHAKER SALUNKHE
Research Scholar, Sunrise University, Alwar, Rajasthan
DR. SUMIT BHATTACHARJEE
Research Supervisor, Sunrise University, Alwar, Rajasthan

ABSTRACT

Lower layers identify simple characteristics and pass that information to higher layers, which may subsequently detect more complex traits. By training each layer independently and feeding its result into the next layer's input, Deep Learning is able to function via a greedy layer-wise unsupervised pre-training. Big Data refers to exceptionally big, heterogeneous data sets characterized by their analysis for patterns and trends due to their features (high volume, diverse formats, speed of processing). In this study, we introduce some of the most advanced neural networks, including the Deep Belief Network, Convolutional Neural Network (CNN), and Recurrent Neural Network, and present a brief history of deep learning for Big Data (RNN). The purpose of this study is to provide a comprehensive overview of deep learning architectures appropriate for handling large datasets. After that, we get into the research and debate around the difficulties and opportunities presented by deep learning for big data analytics. Several unanswered questions and emerging themes in the field are briefly discussed in the conclusion.

Keywords: Deep learning; Big data, Artificial Intelligence, Deep Learning, Machine Learning, Neural Networks, Convolutional Neural Network, Deep Belief Network

INTRODUCTION

For the most part, signal processing and ML methods have relied on shallow-structured architectures up until recently. The raw input signals are converted into a feature space that is unique to the challenge at hand by means of this single-layer architecture. Unfortunately, these designs only work well for issues that can be easily defined and solved. However, they fall short when confronted with issues that are either complicated or lack a clear framework. Therefore, in order to overcome these constraints and tackle increasingly difficult issues, deep architectures are required.

Multi-layered feature detector units are the basis of deep architectures. Lower layers identify simple characteristics and pass that information to higher layers, which may subsequently detect more complex traits. In Deep Learning, the output of one layer is used as input for the next layer, resulting in a greedy layer-wise unsupervised pre-training. As shown in Fig.1, after all iterations have completed, the set of layers with learnt weights may be utilized as initialization for building a deep supervised predictor.

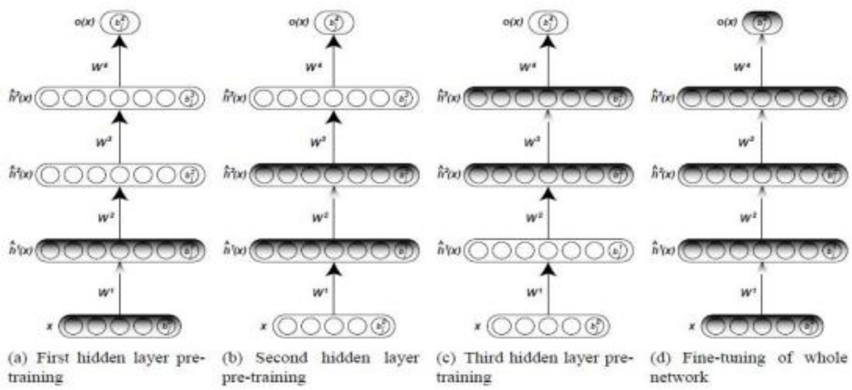

Figure. 1: The greedy unsupervised layer-wise training followed by a supervised fine-tuning stage affecting all layers.

Deep Learning's capacity to create hierarchical representations for both labeled structured input data and unlabeled unstructured input data is a crucial aspect of the technique. In a hierarchical representation, the emphasis is placed on learning higher-order representations from lower-level data. Examples include speech-to-text transcription, image-based object recognition, and video-based stance and motion analysis. The field of computational statistics and the field of machine learning are making the switch to Deep Learning because of this feature. Big Data refers to the large volumes of unstructured, real-time information that are collected from a variety of sources.

The storage, processing, and computational requirements of standard data analysis methods are stretched to the limit by big data analytics. The only way to make sense of Big Data and all of its hidden relationships and complicated patterns is to use fresh structures and technologies. Deep Learning's capacity to extract complicated abstractions has made it a useful tool for the machine learning field in dealing with Big Data. It offers a granular view of the information contained in massive datasets, in particular those that are unlabeled.

Data indexing/tagging, data simplification, and the discovery of previously unseen patterns in large datasets are all examples of the types of Big Data analytics difficulties that may be included under this umbrella term. When used, Deep Learning is a powerful tool for resolving such challenges. Reflected in the "major agreements and the foundation for an American innovation and economic revolution" are Deep Learning and Big Data. . This paper's remaining sections are structured as follows.

The second section shows examples of the construction and learning of the most popular Deep Neural Networks. In Section III, we provide a high-level overview of Big Data and highlight certain data analysis challenges that can be solved using Deep Learning. In Section IV, we will look at some Deep Learning research that has been done to solve the problem of data analysis. Section V concludes with a discussion of some difficulties encountered by Deep Learning due to the unique requirements of Big Data analysis.

DEEP LEARNING ARCHITECTURES

The goal of Deep Learning is to learn a hierarchical representation of data that is both complex and abstract by iteratively passing data through a series of transformation layers. As a result, Deep Learning algorithms are viewed as Deep architectures of successive layers, where a nonlinear transformation is applied to the input of each layer in order to represent the output of that layer. Auto-Encoders (AEs) and Restricted Boltzmann Machines (RBMs) are two examples of unsupervised single-layer learning algorithms that are used as building blocks for more complex models.

•As unsupervised single-layer learning algorithms, AEs consist of three distinct layers: input, hidden, and output. With the goal of recreating the input in the output layer, AEs aim to record the structure of input data in a way that facilitates this . In the training phase, we utilized a back-propagation algorithm, and the desired result was the input.

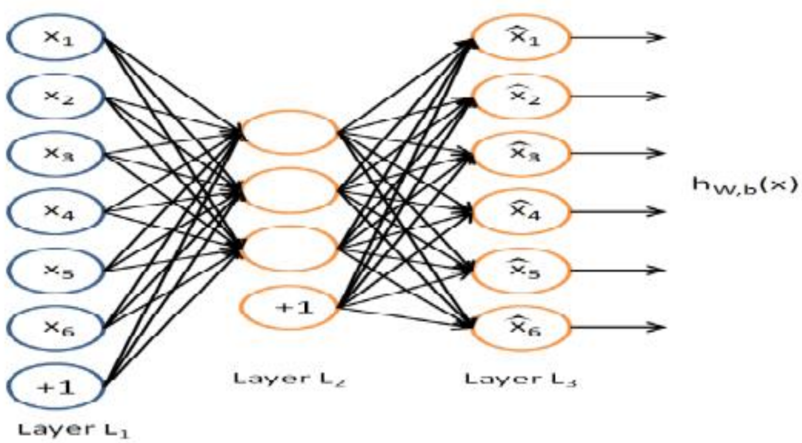

Figure 2: The Auto-Encoder neural network in which the output is similar to the input

•Regular Boltzmann machines (RBMs) are a subclass of BMs in general. BM is a no multilayered network with nodes that are randomly on or off depending on the state of the network as a whole. A RBM is a BM in which no units in the same layer are connected to one another. In the context of BMs, the Contrastive Divergence algorithm has been used most frequently during the training stage.

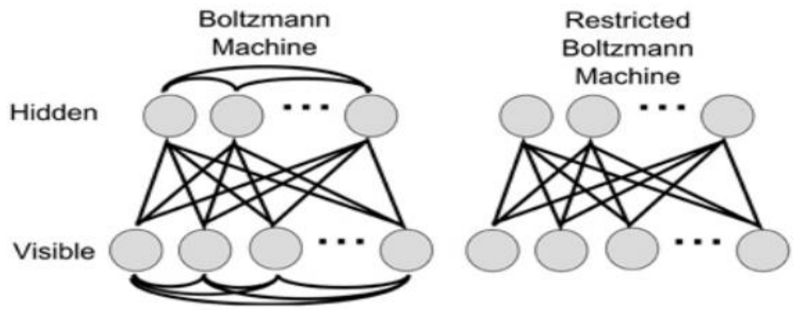

Figure 3: The Restricted Boltzmann Machine as a special case of Boltzmann machine.

However, these unsupervised single layer learning techniques may be used to build Deep unsupervised networks. In this article, we will describe the architecture of the three most popular types of Deep Neural Networks (DNNs) currently in use: Deep Belief Networks, Deep Boltzmann Machines, and Deep Stacking Networks.

LITERATURE REVIEW

Ed Crego, (2016) The future course and potential of innovation in the United States will be affected by two fundamental trends: big data and deep learning. Both have received a lot of attention over the last year, so it's reasonable to assume that they will play an important role in the expansion of the American economy in the years to come. We can't say that we're feeling very optimistic. The future seems bright for both of these movements, in our view. The problem is that they have certain restrictions that need to be lifted before they can fulfill that promise. So, let's have a look at the rationale behind this. There is no shortage of either Big Data or the people who make their profession storing and mining it. Web traffic, tweets, and public transportation data can all be analyzed with Big Data to make predictions about consumer behavior and inform the development of more targeted marketing and service strategies. In an essay for the May/June issue of Foreign Affairs, Kenneth Cuvier, Data Editor of The Economist, and Viktor Mayer-Schoenberger, Professor of Internet Governance and Regulations at the Oxford Internet Institute, praise the potential of Big Data. From their latest book, Big Data: A Revolution That Will Transform How We Live, Work, and Think, comes this piece.

Maryam M Najafabadi (2015) Deep Learning and Big Data Analytics are two areas of data science that get a lot of attention. Due to the proliferation of data collection initiatives in sectors as diverse as national intelligence, cyber security, fraud detection, marketing, and medical informatics, the concept of "Big Data" has emerged as a crucial tool for addressing complex issues in these fields. Tech giants like Google and Microsoft are affecting present and future developments by evaluating massive amounts of data for business analysis and decision making. Deep Learning methods use a hierarchical learning process to extract high-level, complicated abstractions as data representations. Abstraction is learned in a hierarchical fashion, with each level building on the knowledge and understanding of the level below it. One of Deep Learning's strongest points is that it can analyze and learn from large amounts of unsupervised data, making it a useful tool for Big Data Analytics, where the raw data is often not categorized or labeled. In this research, we investigate how Deep Learning can be used to simplify discriminative tasks, index and retrieve information quickly, and extract complex patterns from large data sets, all of which are pressing issues in the field of Big Data Analytics. We also look into some uncharted territory in Deep Learning study, such as how to deal with streaming data, high-dimensional data, model scalability, and distributed computing, all of which are unique challenges brought about by Big Data Analytics. We wrap up with some questions that provide glimpses into potential directions for future research, such as better semantic indexing, semi-supervised learning, active learning, and the definition of criteria for obtaining useful data abstractions.

Min Chen (2017) Early disease detection, patient care, and community services all gain from precise analysis of medical data, made possible by the proliferation of big data in the biomedical and healthcare sectors. But when medical data is of poor quality, the precision of the analysis suffers. Furthermore, certain regional diseases exhibit unique characteristics across geographic regions, which may hinder the ability to predict disease outbreaks. In this paper, we present a streamlined method for using machine learning algorithms to predict the spread of chronic diseases in areas where such outbreaks are common. Using actual hospital data collected in the heart of China between 2013 and 2015, we put the updated prediction models through their paces. We use a latent factor model to reconstruct the missing data and get around the problem of incomplete data. Brain infarction is a chronic disease that we study experimentally. Utilizing both structured and unstructured hospital data, we propose a novel convolutional neural network (CNN)-based multimodal disease risk prediction algorithm. Our research showed that no previous work in the field of medical big data analytics had dealt specifically with both types of data. Our proposed algorithm outperforms the CNN-based unimodal disease risk prediction algorithm in terms of convergence speed and prediction accuracy, reaching 94.8% in both cases.

Qingchen Zhang (2018) As one of the most impressive machine learning approaches at the moment, deep learning has been widely successful in areas including image analysis, audio recognition, and text interpretation. It employs both supervised and unsupervised learning techniques to construct hierarchical representations of features for use in classification and pattern recognition. Due to recent advancements in sensor networks and communication technology, massive amounts of data may now be collected. Big data's qualities of high volume, large diversity, large velocity, and huge veracity provide numerous tough challenges on data mining and information processing while also offering enormous prospects for a wide range of fields, such as e-commerce, industrial control, and smart medicine. Deep learning has emerged as a key component of modern big data analytics platforms. In this work, we summarize recent studies on deep learning models for feature learning in large datasets. Furthermore, we highlight the remaining difficulties of large data deep learning and offer potential avenues for further research.

Abhay Narayan Tripathi (2020) Deep Learning (DL) is increasingly being used in favor of machine learning (ML), despite ML's longstanding status as a synonym for AI. While supervised and unsupervised machine learning approaches are being actively developed, deep learning maintains its drive to model the human nervous system via the use of complex Neural Networks (NN).. By applying Deep Learning to Big Data, hitherto unattainable patterns may be discovered. Self-driving vehicles, image recognition, healthcare, transportation, and more all use Deep Learning. Companies nowadays are beginning to understand that having access to massive volumes of data is crucial to making informed decisions and bolstering their plans. Big Data refers to exceptionally big, heterogeneous data sets characterized by their analysis for patterns and trends due to their features (high volume, diverse formats, speed of processing). In this study, we introduce some of the most advanced neural networks, including the Deep Belief Network, Convolutional Neural Network (CNN), and Recurrent Neural Network, and present a brief history of deep learning for Big Data (RNN).

APPLICATION OF DEEP LEARNING IN BIG DATA

When considering Deep Learning's role in Big Data, it's important to keep in mind that DL primarily addresses the volume and veracity of Big Data. What this implies is that DL are well-suited for evaluating and extracting relevant information from both massive volumes of data and data obtained from many sources [20]. Microsoft's (MAVIS) speech recognition system uses Deep Learning to facilitate the search of audio and video data using human voices and talks [21]. [22]. The Google image search service is another use of DL in a Big Data setting. So that pictures might be annotated and tagged for use in image search engines, image retrieval systems, and image indexing, they turned to DL. There are a few obstacles that must be overcome if we are going to use Dl, and they are as follows:

1) Deep Learning for High Volumes of Data

The first is whether or not all of the Big Data input should be used. From one perspective, the issue is how much amount of data is required for training, while from another, we use DL algorithms in a subset of accessible Big Data for training aim and utilize the remainder of the data for extracting abstract representations.

A second unsolved issue is domain adaptation, which arises in situations when the distribution of training data and test data are dissimilar. A further angle of attack would be to consider how we may expand DL's ability to generalize learned patterns when there is a difference between the input domain and the target domain.Criteria for permitting data representations to give future-useful semantic meanings is another challenge. In layman's terms, this means that any individual representation of the retrieved data should not be permitted to have any meaningful weight. Better data representations require that we adhere to certain standards.Another challenge is that, in the Big Data setting, it may be tough to make heads or tails of the complexities of most DL algorithms, which need a loss to be defined and require us to know what it is we want to extract. Also, most of them include analytical findings that are too complex for laypeople to grasp. That is to say, the technique is too complicated to study in detail. In a Big Data setting, the problem just becomes worse.Due to its ability to learn abstract representations, Deep Learning appears to be a good fit for the integration of heterogeneous data with multiple modalities.Last but not least, they have a serious issue in needing labeled data. They will perform poorly if we are unable to provide labeled data. We can use reinforcement learning, in which the system automatically collects data and our only involvement is to provide rewards, to solve this problem.

2) Deep Learning for High Variety of Data

These days, information can be obtained from a wide range of sources, in a plethora of formats, and likely with a few different distributions. Multimedia data from the web and mobile devices, for instance, includes a vast assortment of images, videos, audio streams, graphics, animations, and unstructured text, all of which have their own unique characteristics and are growing at an alarming rate. Some of the unanswered questions in this area are as follows.

In light of the fact that many sources may provide contradictory information, the question arises as to how best to reconcile these discrepancies and integrate data from several sources.If significantly increased modalities improve system performance, how much?At what granularity do deep learning architectures become useful for feature fusion in heterogeneous data?

3) Deep Learning for High Velocity of Data

It's crucial to be able to quickly process the massive amounts of data being generated. An option for processing such high-velocity data for learning is online learning strategies, which can be implemented using deep learning. There hasn't been much development in online deep learning in recent years. Problems can be encountered in this context due to factors like:

In many cases, data are not stationary; rather, data distributions evolve over time.The key issue is whether or not Big Data and deep architectures can help with transfer learning.

RESULTS AND DISCUSSION

According to studies, substantial strides have been made in the area of Big data analytics using deep learning algorithms. Problems in Big data analytics, such as large-scale data analysis, semantic indexing, data tagging, information retrieval, classification, and prediction, are significantly simplified by DL. However, deep learning has only made modest strides in areas such as model scaling, distributed computing, and high-scale data processing, where large amounts of data must be handled quickly and accurately. Several questions that still need answers and emerging areas of study are outlined below.

1) The ever-increasing volume of big data necessitates the development of ever-more-complex deep learning models. Depending on the methods and resources at hand, it may no longer be possible to effectively train such massive deep learning models that can be used for Big Data. The future of this issue relies on the development of novel learning structures and computing infrastructures.

2) Some recent multi-modal deep learning models only add the characteristics learnt for each modality in a linear fashion. Sometimes this works, but usually it doesn't provide the desired outcomes. To increase the efficiency of multi-modal deep learning models, it is necessary to explore the most efficient fusion methods of learnt features. However, the high computational complexity of deep computational models is the result of the vast number of parameters in these models. Computational complexity reduction research for deep computational models is needed.

3) Most parameter and structure-based integrated learning algorithms work well with classic, hidden-layer learning models. Studies on the potential of integrated learning algorithms in deep learning models and deep architectures are required.

4) As the volume of poor-quality data continues to rise, it will soon be crucial to explore trustworthy deep learning models for this kind of data.

5) New parallel and distributed algorithms/frameworks are needed to make deep learning models scalable.

CONCLUSION

In order to get superior abstract knowledge, it is now essential to take on Big Data. Deep Learning (Hierarchical Learning) is one method that may be used to achieve this goal since it gives a more abstracted view of the underlying data. There are benefits and drawbacks to using Deep Learning in the Big Data setting, like any other technology. More information means more high-level abstract information, but we still have a long way to go. The purpose of this study was to provide a high-level overview of deep learning, beginning with its origins and development. To that end, we want to expand our existing understanding of deep learning to include a wider variety of niche applications, particularly those that make use of convolutional and recurrent neural networks. New approaches and game-changing solutions are needed to solve technological problems if we are to fully fulfill the promise of Big Data. Since this is the case, more research into the topic of deep learning is necessary.

REFERENCES

1. E. Crego, G. Munoz, and F. Islam, (2016). 'Big Data and Deep Learning: Big deals or big delusions?'[Online] Available at: http://www.huffingtonpost.com/george-munoz frank-islam-and-ed-crego/big-data-anddeep-learnin_b_3325352.html. (Accessed on 2 /1/ 2016)

2. Maryam M Najafabadi (2015) “Deep learning applications and challenges in big data analytics” Najafabadi et al. Journal of Big Data (2015) 2:1 DOI 10.1186/s40537-014-0007-7

3. Chen M, Hao Y et al (2017) Disease pre-diction by machine learning over big data from healthcare com-munities. IEEE Access 5:8869–8879

4. Zhang Q, Yang LT et al (2018) A survey on deep learning for big data. Inf Fusion 42:146–157

5. Abhay Narayan Tripathi (2020) “A Depth of Deep Learning for Big Data and its Applications” International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org ENCADEMS - 2020 Conference Proceedings

6. L. Deng, “An Overview of Deep-Structured Learning for Information Processing”. In Proceedings of Asian-Pacific Signal and Information Processing Annual Summit and Conference (APSIPAASC), 2011. [2]

7. H. Lee, R. Gross, R. Ranganath, and A. Ng, “Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations”. In ICML, 2009. [3]

8. R. Socher, Y. Bengio, and C. Manning, “Deep Learning for NLP”. Tutorial at ACL,http://www.socher.org/index.php/DeepLearnngTutorial/DeepLearningTutoril, 2012. [4]

9. H. Larochelle, Y. Bengio, J. Louradour, and P. Lamblin, “Exploring strategies for training deep neural networks”, The Journal of Machine Learning Research, vol. 10, pp. 1–40, 2009. [5]

10. B. Tannahill, C. Maute, Y. Yetis, M. Ezell, A. Jaimes, R. Rosas, A. Motaghi, H. Kaplan, and M. Jamshidi, ”Modeling of system of systems via data analytics — Case for Big Data in SoS”. In System of Systems Engineering International Conference, pp.177, 183, “doi: 10.1109/SYSoSE.2013.6575263”, 2013. [6]

11. S. Sagiroglu, and D. Sinanc, ”Big Data: A review”. In Collaboration Technologies And Systems (CTS),2013 IEEE International Conference on San Diego,CA: IEEE, 2013. [7]

12. X. Chen, and X. Lin, “Big Data Deep Learning: Challenges and Perspectives”. In Access, IEEE, vol.2, pp.514, 525, doi: 10.1109/ACCESS.2014.2325029, 2014. [8]

13. M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N, Seliya, R. Wald, and E. Muharemagic, “Deep Learning applications and challenges in Big Data analytics”. Journal of Big Data, vol.2, doi: 10.1186/s40537-014-0007-7, 2015.

14. L. Deng, “A tutorial survey of architectures, algorithms, and applications for Deep Learning”, In APSIPA Transactions on Signal and Information Processing, 2013.