IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Study Of Big Data Using Deep Learning

1Saba Tahir and 2Ishrat Jameel
1Saba Tahir and 2Ishrat Jameel
1 Department of computer application GDC Kulgam (contractual), J&K, India
2 Department of computer application GDC Kulgam (contractual),J&K, India

Abstract

Two data science subfields with a lot of focus are big data analytics and deep learning. Big Data has grown in importance as a result of the large-scale collection of domain-specific data by both public and private entities, which can provide useful information regarding issues like national intelligence, cyber security, fraud detection, marketing, and medical informatics. Large data sets are being analysed by businesses like Google and Microsoft for business analysis and decisions that will affect both current and future technologies. Deep learning techniques extract high-level, complex abstractions as a means of data representation through a hierarchical learning process. Based on relatively simpler abstractions created in the previous level of the hierarchy, complex abstractions are learned at a given level. A key benefit of deep learning is its capacity to analyse and learn from massive amounts of unsupervised data.

Keywords: Big data ,deep learning ,database

Introduction

Big data is generally defined as a volume of data that cannot be processed efficiently using standard database techniques. They offered a helpful definition of big data based on the books and journals they read while conducting their research, making observations, and analysing its essential characteristics. Big data, in their definition, is "a set of methodologies and technologies that require new kinds of integration to find significant hidden values from enormous, diverse, complicated, and massive datasets." The following characteristics can also be used to define big data: volume, velocity, veracity. According to a recent definition from Gartner, big data are high-volume, high-velocity, and/or high-variety information assets that call for new forms of processing to enable improved decision-making, insight discovery, and process optimization. Big data is frequently linked to the 4Vs; it has also been referred to as the big data problem. Large, heterogeneous, and frequently unstructured data sets that are challenging to analyse and extract information from using standard data management tools and procedures are another way to define it. Another definition of big data states that it consists of enormous datasets with high data volumes, a wide range of information types, and significant diversity, including structured, semi-structured, and unstructured data that come more quickly (velocity) than conventional datasets. Big data characteristics include: Although different authors of different research have described big data in a variety of ways, the most common classifications of big data centre around the vs. (volume, velocity, variety, veracity, valence, and value). The following four 4V traits have gained widespread acceptance because they define big data precisely: veracity refers to trust and integrity;

volume refers to enormous amounts of data; velocity refers to the speed at which data is generated; variety refers to structured data, unstructured data, and image

Big data issues: The big data issues mentioned in the extra publications and papers analysed in this survey include the following: Data governance, disaster recovery, big data management and storage, big data computing and analysis, scalability and availability, data integrity, data transformation, and data quality. Security, privacy, heterogeneity. How numerous literary works have addressed big data challenges? One significant issue with large data is security, which is the. Because of its size, variety, speed, and vulnerability to threats, big data can be challenging to safeguard. Security is not the only issue to be worried about; rather, all the other issues described above have led to the varied ways that big data can be evaluated, as well as the different technologies that can be used to do it Algorithms that encourage both researchers and practitioners can be employed. Big data analytics and deep learning: To create and forecast models and trends, big data analytics is currently employed in the majority of fields of research and organisations, including corporations, the science and technology sectors, and other fields of study. Other data-related fields of study and disciplines, which have profited from the prior body of knowledge that was generated in previous years, also share concerns with big data analytics. According to research, big data analytics has a substantial impact on firms by generating new ideas, competitive advantages, and ways to make money or raise the likelihood of making money. As a result, numerous analytical methods, algorithms, and tools make it easier to get pertinent data from huge data that is unprocessed and raw. Association rule learning, data mining, cluster analysis, machine learning, text analytics, and business intelligence tools are a few of these methods. We largely concentrated on deep learning algorithms for big data for the purposes of this SLR.

Overview of Deep Learning

A branch of machine learning research that is currently popular in artificial intelligence is deep learning. Deep learning is a component of the machine learning model that "learns hierarchical features for the tasks of classification and pattern recognition" using either the supervised or unsupervised method. Deep learning, according to, attempts to bring machine learning one step closer to artificial intelligence, which was one of its initial goals. Deep learning has been employed in numerous industries, including medicine, botany, image recognition, food processing, mechanics, Web mining, text mining, and data mining. It has also been used in many different research investigations. By constructing and layering many layers, much like neurons, deep learning is a machine learning technique that extracts higher-level representations from datasets. Recent research have shown that deep learning is effective in a variety of tasks, including image processing, speech recognition, web search, recommendation systems, and more. Additionally, machine learning was used to analyse medical data from electronic health records on a sizable patient population that may be related to radiation oncology. Machine learning has been used in this particular area of medicine. Deep belief networks (DBNs) and convolutional neural networks are two well-established deep architectures in deep learning, according to (CNNs). When given with tagged and unlabeled data, DBN uses deep architecture to learn the feature representations, and it also uses supervised and unsupervised fine-tuning procedures to develop the model. CNN is made up of numerous hierarchies, some of which are standard neural networks used for classification and others of which are used for feature representations (or feature maps).

Applications of deep learning in big data analytics

As previously mentioned, Deep Learning algorithms extract valuable abstract representations from the raw data by using a hierarchical multi-level learning approach, wherein higher-level more abstract and complex representations are learned based on the less abstract concepts and representations in the lower level(s) of the learning hierarchy. While Deep Learning can be used to learn from labelled data if it is accessible in sufficient quantities, it is mostly attractive for learning from vast amounts of unlabelled/unsupervised data, making it attractive for extracting meaningful representations and patterns from Big Data. More traditional discriminative models can be trained using comparatively fewer supervised/labelled data points once the hierarchical data abstractions have been learned from unsupervised data using deep learning, where the labelled data is normally supplied through human/expert input. In comparison to relatively shallow learning architectures, deep learning algorithms have been demonstrated to perform better in extracting non-local and global correlations and patterns from the data. Other beneficial properties of the learned abstract representations by deep learning include: (1) relatively simple linear models can work effectively with the knowledge from the more complex and more abstract data representations; (2) increased automation of data representation extraction from unsupervised data enables its broad application to different data types, such as image, textural, audio, etc.; and (3) relational and semantic knowledge can be obtained a relatively simple linear model and relational and semantic knowledge can be obtained from more complex and more abstract representations of data. While Deep Learning-based representations of data have additional beneficial aspects, the above-mentioned properties are particularly crucial for Big Data Analytics. Volume, Variety, Velocity, and Veracity—the four Vs of Big Data—are taken into account by Deep Learning architectures and algorithms are better equipped to handle problems relating to the volume and variety of big data analytics. While algorithms with shallow learning hierarchies struggle to explore and comprehend the higher complexities of data patterns, deep learning by its very nature takes use of the availability of vast volumes of data, or volume in big data. Additionally, since Deep Learning is concerned with data abstraction and representations, it is probably well-suited for analysing unstructured data that is presented in a variety of formats and/or from a variety of sources, or Variety in Big Data. This could reduce the need for input from human experts to extract features from every new data type observed in Big Data. Big Data Analytics gives a significant potential for the development of unique algorithms and models to handle particular Big Data-related difficulties, while also posing a number of challenges for more traditional data analysis methodologies. For data analytics specialists and practitioners, one such solution venue is provided by deep learning principles. For instance, when complicated data is represented in greater levels of abstraction, simple linear modelling techniques can be taken into consideration for Big Data Analytics as a useful source of knowledge for decision-making, semantic indexing, information retrieval, and other applications. The remaining portions of this section provide summaries of several significant research findings in the area of deep learning algorithms and architectures, including data tagging, discriminative tasks, and semantic indexing. Because some of the application domains in the works presented include vast amounts of data, our goal is for professionals to see how these Deep Learning works can be used in unique ways to Big Data Analytics. Although deep learning techniques can be used with a variety of input data, in this section we specifically focus on their use with image, text, and audio data.

Conclusion

We put out a development setting that makes it possible to combine various big data sources with deep learning architectures. Additionally, it enables quick and simple experiment preparation and execution so that users can draw meaningful conclusions from data. We examined the prior research in the big data and deep learning sectors. Numerous methods currently in use either employ large-scale deep learning systems or distributed deep learning models on big data architectures. We also looked at the tools that may be used to manipulate data. Our suggested environment aims to bridge the gap between deep learning systems and large data systems. It gives deep learning systems the capacity to integrate different huge data streams. Contrary to other strategies, ours is independent of the large deep learning model or huge data stream. We have also constructed a prototype to demonstrate the viability and efficacy of our suggested environment to support our recommended method.

Future scope

The recent achievements of deep learning in self-driving cars and in extremely difficult games like go are a good sign for future achievements when deep learning is used in other fields. Deep learning algorithms and data from big data systems together will make it simpler to move into new fields of study and product creation. By lowering the bar for developing such systems, the development environment presented in this thesis makes the process accessible to both programmers and non-programmers.

Bibliography

1.Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference On. IEEE Vol. 1. pp 886–893

3. Lowe DG (1999) Object recognition from local scale-invariant features. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference On. IEEE Computer Society Vol. 2. pp 1150–1157

4. Bengio Y, LeCun Y (2007) Scaling learning algorithms towards, AI. In: Bottou L, Chapelle O, DeCoste D, Weston J (eds). Large Scale Kernel Machines. MIT Press, Cambridge, MA Vol. 34. pp 321–360. http://www.iro.umontreal.ca/~ lisa/pointeurs/bengio+lecun_chapter2007.pdf

5. Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. Pattern Analysis and Machine Intelligence, IEEE Transactions on 35(8):1798–1828. doi:10.1109/TPAMI.2013.50

6. Arel I, Rose DC, Karnowski TP (2010) Deep machine learning-a new frontier in artificial intelligence research [research frontier]. IEEE Comput Intell 5:13–18

7. Hinton GE, Osindero S, Teh Y-W (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554

8. Bengio Y, Lamblin P, Popovici D, Larochelle H (2007) Greedy layer-wise training of deep networks, Vol. 19

9. Larochelle H, Bengio Y, Louradour J, Lamblin P (2009) Exploring strategies for training deep neural networks. J Mach Learn Res 10:1–40

10. Salakhutdinov R, Hinton GE (2009) Deep boltzmann machines. In: International Conference on, Artificial Intelligence and Statistics. JMLR.org. pp 448–455

11. Goodfellow I, Lee H, Le QV, Saxe A, Ng AY (2009) Measuring invariances in deep networks. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. pp 646–654 12. Dahl G, Ranzato M, Mohamed A-R, Hinton GE (2010) Phone recognition with the mean-covariance restricted boltzmann machine. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. pp 469–477

13. Hinton G, Deng L, Yu D, Mohamed A-R, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath T, Dahl G, Kingsbury B (2012) Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. Signal Process Mag IEEE 29(6):82–97

14. Seide F, Li G, Yu D (2011) Conversational speech transcription using context-dependent deep neural networks. In: INTERSPEECH. ISCA. pp 437–440

15. Mohamed A-R, Dahl GE, Hinton G (2012) Acoustic modeling using deep belief networks. Audio Speech Lang Process IEEE Trans 20(1):14–22

16. Dahl GE, Yu D, Deng L, Acero A (2012) Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. Audio Speech Lang Process IEEE Trans 20(1):30–42 17. Krizhevsky A, Sutskever I, Hinton G (2012) Imagenet classification with deep convolutional neural network