



Deep learning and Big Data Analysis: Challenges, Opportunities and Applications

Mehroush Banday

Department of Computer Science and Engineering, School of Engineering
Sciences and Technology, JamiaHamdard, New Delhi, India

ABSTRACT

With the approach of IoT, there are tremendous changes going in the measure of information. Big data and deep learning are the two most critical concentration focuses in this universe of computerized science. Information is considered an indispensable need these days. This structure of open and private information is likewise exceptionally important. The primary favourable position of DL is looking at and gets out taking in measures from it for unsupervised and managed data, by making it a noteworthy gadget for Big Data investigation where crude information is accumulated and is marked and unlabelled and the at long last arranged. In the present paper, we examine with reference to how Deep Learning is exploited for some basic issues in Big Data Analytics, which additionally incorporates extraction of complex information from colossal masses of data, semantic requesting, data marking, speedy information recuperation, and enhancing discriminative errands. Huge information has now ended up being extremely famous as a couple of affiliations are gathering huge measures of room specific information which can be utilized to address the issues related to national learning, computerized security, distortion area, and wellbeing informatics. Profound Learning Big Data permits extraction of not typical condition of information, complex thoughts as data depictions through a different levelled learning process. A key favourable position of Deep Learning is Big Data investigation examination that it can pick up from huge mass of unsupervised data. In this manner, making it an imperative instrument for Big Data Analytics where monstrous measures of data and information are not group. This part weights on the need for immense information, mechanical kinds of

advance, contraptions and strategies being utilized to process huge information are talked about. Mechanical redesigns and hindrances of existing putting away procedures are in like way appeared. This portion shows a structure of titanic information examination, its application, great conditions, and restrictions. Barely any examination issues and future course are introduced in this part. With the digitization of the greater part of the frameworks, change of various social affiliation stages, web journals, plan of various sort of sensors, assurance of hand held motorized gadgets, wearable contraptions and effect in the Internet usage, huge measure of information are being conveyed on steady present. It's unrealistic for anybody to deny that Internet has changed the way affiliations work, working of the association, planning and way of living life.

Keywords: *Internet of things; deep learning; bigdata; semantic ordering; information labelling; discriminative errands*

INTRODUCTION

BIG DATA ANALYTIC

By doing Exceptional study about the deep learning and huge data examination we found two or three difficulties, openings and applications between the DL and Big information. Immense information infers all things considered underwear data that beats the gathering, managing, and figuring motivation behind limitation of standard databases and data examination methodologies [5]. As use, Big Data requires instruments and systems which can be connected with

discrete and remove plots from broad scale data [1]. The move of Big Data is caused by being produced which is extended gathering purposes of imprisonment, and openness of expanded data volumes, which give association a more essential number of data as compared to data they have selecting resources and advances to process. Despite the particular sensational volumes of data, Big Data is likewise associated with other specific complexities, regularly proposed as the 7 Vs
Volume, Variety, Velocity, Veracity, Visualizations, Value and Variability

DEEP LEARNING

The general point of convergence of ML is the depiction of the data and hypothesis of the learnt outlines for use on future subtle data [4]. The goodness of data depiction substantial influences the execution of machine understudies on the data: a poor data depiction is most likely going to decrease the execution of even a pushed, complex machine understudy, while a better than average data depiction can provoke first class for a for the most part less perplexing machine understudy [19]. DL computations are one promising street of research into the robotized extraction of complex data depictions (features) everywhere measures of thought [26]. Such estimations develop a layered, dynamic outline of learning and addressing data, where bigger sum (more unique) features are described with respect to bring down level (less hypothetical) features.

APPLICATIONS

In this fragment we diagram the basic works that has been performed in DL Algorithms. Utilizations of profound learning in huge examination and plans of information, incorporates semantic requesting, discriminative assignments, and data naming. Profound Learning calculations are material to different kinds of data; yet in this segment we centre around its application on picture, artistic, and sound data [12][13].

There are three utilizations of profound learning in huge information examination. They are:

1. Semantic requesting/ordering
2. Information marking
3. Discriminative assignments

Semantic requesting: information recuperation is a key task that is connected with Big Data Analytics. Gainful limit and recuperation of information is a creating issue of Big Data examination, as data in considerable scale sums, for instance, content, picture, video, and sound is being assembled and made open over various zones on the planet. Thusly, strategies and game plans that were at that point used for information accumulating and recuperation are tried by colossal volumes of data. Semantic requesting winds up being a successful framework as it supports learning divulgence and discernment, thusly making web lists work more quickly and capably.

Discriminative assignments: While performing discriminative assignments in Big Data Analytics, Learning computations empower customers to isolate jumbled nonlinear features from the unrefined data. It is like manner energizes the use of straight models to perform discriminative endeavours using expelled features as information. This approach has two purposes of premium: First by taking out features with Deep Learning adds nonlinearity to the data examination, thusly accomplice discriminative attempt eagerly to AI, what's more applying direct intelligent models on isolated features is more gainful computationally. These two points of interest are fundamental for Big Data since it licenses specialists in accomplishing bewildered errands related to Artificial Intelligence like challenge affirmation in pictures, picture gratefulness, et cetera.

b: Deep Learning strategies help in semantic marking. Significant Learning instruments can empower division and remark of complex picture scenes. Significant Learning can in like manner be used for movement scene affirmation and video data naming. It uses a self-governing variety examination to take in invariant spatiotemporal features from video data. This approach helps in isolating profitable features for performing discriminative errands on picture and video data.

Profound Learning has been powerful in achieving amazing results in isolating accommodating features. In any case, there is so far critical work that outstanding parts to be enhanced the circumstance energize examination that consolidates confirmation of appropriate focuses in adjusting extraordinary data depictions and performing other complex endeavours in Big Data Analytics.

There are many more deep learning applications in big data analytics such as:

- Real life applications
- Social targeting
- Health care
- Education system
- Telecommunication
- Weather forecasting
- Security
- Genomics such as in DNA
- Retail

SOCIAL TARGETING

Deep learning holds the possibility to figure the implicit feelings and occasions in content. It can distinguish between photographs. It can likewise make educated expectations about individuals' conceivable future conduct. Every one of these highlights makes it a hot property in the fields of insight, deals, showcasing, and promoting.

REAL TIME APPLICATION

Does Deep Learning apply to my consistent business?

Genuinely it does

- On the distant chance that your business makes or uses high volumes of variable information
- On the distant chance that time is cash for you
- On the distant chance that you search for works out as expected that propose ensuing stages
- On the distant chance that you are driven and wish to remain before your enemies if, in spite of everything that you can't manage the cost of lethargy and plenitude

HEALTH CARE

Health care and deep learning are so much interlinked with each other these days. Deep learning helps in detecting the diseases which were not before easy to detect or it gives us the knowledge to know about the cure of the disease. Through the research we have mentioned some new inventions about the link of health care and deep learning.

Example

Now your Apple watches can detect the signs of being diabetic, or having high cholesterol with 74% accuracy, or having high blood pressure with 84% accuracy.

"Your heart is associated with your pancreas through the autonomic sensory system. As individuals build up the beginning times of diabetes, their example of heart rate changeability shifts," clarified Johnson Hsieh, prime supporter of Cardiogram[17].

With the assistance of Tesla P100 GPUs and the cuDNN-quickenened TensorFlow profound learning structure, the group prepared their DeepHeart neural system on an informational collection of 200 million heart rate and step check estimations from a clinical investigation of more than 14,000 Apple Watch clients. The prepared system could identify that 462 of the select in think about individuals had diabetes [25].

FORECASTING

As before forecasting was not that easy as it is of now. We were able to do weather forecasting for minimal number of days but now for almost a month we can have weather forecasting and even can know about the whole world. Not only we can forecast about weather but we can also know about the changes in wave forms or in plates tectonic plates beneath the earth , the earthquake forecasting etc. Recently a survey has been done about waveforms and that is :

Utilizing the NVIDIA DGX-1 and cuDNN-quickenened MXNet profound learning structure, the scientists prepared their convolutional neural systems on about 2,500 waveform formats got from the Einstein Toolkit keep running on the Blue Waters supercomputer — and information from the LIGO Open Science Center. Their deep learning method named Deep Filtering accomplishes comparative sensitivities and lower mistakes contrasted with built up gravitational wave location calculations while being much more computationally proficient and stronger to commotion oddities. The technique permits speedier than constant handling of gravitational waves in LIGO's crude information and furthermore empowers new material science since it can distinguish new classes of gravitational wave sources that may run unnoticed with existing discovery calculations [15].

EDUCATION SYSTEM

Some of the ways in deep learning that can help our education system and make the better future for our generations are to make the school a smart school or a smart college. Education system is growing with the help of deep learning in big data analytics. Learning investigation those tracks understudy information. Dynamic booking matches understudies that need assistance with educators/teachers that have time[16]. There is a smart study going in children. Most of things are helped because of their interest in deep learning, cloud computing and big data analytics.

With the coming of mechanized course modules, it is conceivable to evaluate the scholarly execution continuous. This screens the execution of the understudies after every module what's more, give prompt criticism on their learning design. It likewise causes the instructors to survey their showing teaching method and change in view of the students' execution and necessities. Dropout designs, understudies requiring unique consideration and understudies who can deal with testing assignments can be anticipated (West, 2012). Beck and Mostow (2008) examined the understudy perusing perception utilizing clever mentor programming and watched that perusing botches lessened impressively when the understudies re-read an old story rather than another story.

SECURITY

Digital or cyber-attacks follows the pattern for any kind of attack, more often than not composed by people. They utilized deep learning keeping in mind the end goal to identify factors that could imply that a digital assault is going on or is preparing. On account of consistent learning over weeks they figure out how to get their calculations more grounded after some time. It appears to be extremely encouraging [2].

TELECOMMUNICATION

Low reception of telecommunication services and agitate administration are few of the most widely recognized issues looked by the providers by mobile services (MSPs). The cost of gaining new client is higher than holding the current ones. Client encounter is related with client dependability and income. Keeping in mind the end goal to enhance the client encounter, MSPs break down various factors, for

example, statistic information (sexual orientation, age, conjugal status, and dialect inclinations), client inclinations, family unit structure and use subtle elements (CDR, web utilization, esteem included administrations (VAS)) to show the client inclinations and offer an important customized administration to them. This is known as focused advertising, which enhances the appropriation of versatile administrations, diminishes beat, in this way, expanding the income of MSPs. Ufone, a Pakistan-based MSP, lessened the beat rate by accurately advertising the redid offers to their clients (Utsler, 2013). The organization examines the CDR information to distinguish the call examples to offer distinctive plans to clients. The administrations are showcased to the clients through a call or instant message. Their reactions are recorded for facilitate investigation [12].

RETAIL

Development of web based business, online purchasing volume contribution and the nature of information for information driven customization. Real retail locations may put CCTV not exclusively to watch the examples of burglary yet in addition to track the stream of clients. It makes a difference to watch the age gathering, sexual orientation and buying examples of the clients amid weekdays and ends of the week [3]. In view of the buying examples of the clients, retailers assemble their things utilizing a surely understood information mining strategy called Market Basket Analysis in year 1994 with the goal that a client purchasing diapers may buy beer. This makes a difference to choose the situation of items and settle on the costs. These days, online businesses firms utilize advertise bushel investigation and recommender frameworks to portion and focus on the clients. They gather the snap stream information, watch conduct and suggest items in the constant.

Challenges

7 Vs of big data are also the challenges that deep learning in big data has. The 7 V's of big data are:

- Volume
- Variety
- Veracity
- Value
- Visualisation
- Velocity
- Variability

Volume: Volume is how much information we have – first it was used to be in Gigabytes and now currently being estimated in ZB or YB

Velocity the data which is available for us in how much time or speed.

Variety: depicts one of the greatest difficulties of huge information. This means all sort of structured and un structured data or in formation which is available for us.

Variability: is not the same as assortment. A coffee shop may offer 4 unique mixes of cappuccino; however you get a similar mix each day and it tastes consistently different, that is changeability. The same thing is for data we have that is it may change and can effect the hemogeneity of our data.

Veracity: is with ensuring the information is exact, which has procedures to shield the information from our frameworks. How frequently have you seen Tom and Jerry in your database? It's the exemplary "junk in, rubbish out" test.

Visualisation is normal in this age. Utilizing outlines and charts to picture a lot of complex information is substantially more successful in passing on significance than spread sheets and reports stuffed with numbers and equations.

Value In the wake of tending to volume, speed, assortment, inconstancy, veracity, and representation – which takes a considerable measure of time, exertion, and assets – you need to make certain your association is getting an incentive from the information [14]

Besides all these 7 V's there are many more challenges for deep learning in big data and they are:

- Data mining
- Variety of data format
- Unstructured data
- Real time analysis
- Support for new technology like IoT, IoE
- Non-stationary information incremental learning
- Information with high dimensions
- Expansive scale model

Data Mining

Mining of data as defined by Wikipedia is "mining of information(KDD), is the subfield of software

engineering, which is interdisciplinary and is the computational procedure of finding designs in expansive indexes of information ("enormous information") which includes techniques at the convergence of computerized reasoning, machine learning, measurements, and frameworks of database [9]."

Deep Learning is an arrangement of Machine Learning calculations which have at least one shrouded layers in Neural Networks. Presently Deep Learning frameworks are utilized for any assignments other Machine Learning calculations are utilized for. These can be Classification, Dimensionality Reduction, Object Recognition, Clustering and so forth.

Unstructured data format

It is the type of data which does not have any conventional data models. On account of the development of elective stages for putting away and overseeing such information, it is progressively predominant in IT frameworks and is utilized by associations in an assortment of business knowledge and examination applications. Unstructured information or data isn't suited to exchange preparing applications, which are the region of organized information. Rather, it's fundamentally utilized analytics and examination [23]. One well known application is client examination. Retailers and different organizations investigate unstructured information to enhance client relationship administration forms and empower more-focused on advertising; they likewise do supposition examination to distinguish both positive and negative perspectives of items, client benefit, and corporate substances, as communicated by clients on interpersonal organizations and in different gatherings.

Prescient support is a rising examination utilizes case for unstructured information. For instance, makers can break down sensor information to endeavour to identify hardware disappointments before they happen in plant-floor frameworks or completed items in the field. Vitality pipelines can likewise be observed and checked for potential issues utilizing unstructured information gathered from IoT sensors. Unstructured information investigation additionally helps administrative consistence endeavours, especially in helping associations comprehend what corporate archives and records contain.

Support for new technology like IoT and IoE

The Internet of Things vision is to change ordinary items in being brilliant by misusing an extensive variety of propelled advancements, from implanted devices/gadgets and correspondence advancements to Internet conventions, information examination, and so forward. The potential monetary effect of IoT is normal to bring numerous business openings and to quicken the monetary development based on IoT administrations. As of late, numerous IoT applications emerged in various vertical spaces, i.e., wellbeing, transportation, shrewd home, savvy city, farming, instruction, and so forth. The principle component of most of these apps are wise learning instrument for expectation (i.e., grouping or relapse), or bunching. Among the numerous machine learning approaches, Deep learning has been effectively used in numerous IoT applications in late years [22]. The reason for this escalated reputation for Deep Learning alludes to the way that customary ML approaches don't address the rising investigative needs of IoT frameworks. Rather, IoT frameworks require distinctive current information explanatory methodologies and simulated knowledge (AI) techniques as indicated by the pecking order of IoT information age and administration and is represented in the below figure. The developing enthusiasm for the Internet of Things (IoT) and its subsidiary huge information require partners to unmistakably get it their definition, building pieces, possibilities, and difficulties. IoT also, enormous information have a two way relationship [18]. On one hand, IoT is a principle maker of enormous information, and then again, it is an vital focus for huge information examination to enhance the procedures furthermore, administrations of IoT[10]

Non-stationary information incremental learning

Among all Big Data Analytics viewpoints one is to manage spilling and information moving quickly. This information investigation is useful in assignment observation, example, and location for extortion. It is not that easy to adjust DL to deal with information which is gushing, as there is a need for calculations that manages a lot of persistent info [8].

Information with high dimensions

A part of the Deep Learning counts can wind up being restrictively computationally-extravagant while administering information with high dimensions, example, images, mostly in perspective of the

sometimes coordinate learning process related with a noteworthy layered pecking solicitation of taking in reflections and delineations from a lower level layer to a greater entirety layer. Figuratively, these DL estimations can be hindered when working with Big Data those introductions broad Volume, one among the seven Vs related with Analytics of Big Data [6].

Expansive scale models

With a calculation and investigation perspective, the current achievements of DL to significantly bigger scale models and huge datasets can be scaled? The outcomes shown are accurate and the adequacy of expansive scale models, with specific spotlight on models with an extensive number of model parameters which can extricate more confounded highlights and portrayals [7]. Not with the issue of taking care of volumes i.e gigantic volumes of information, huge scale DL models for Big Data Analytics needs additionally to battle with many more different Big Data issues, example, adjustment of space and spilling information.

REFERENCES

1. Agneeswaran, V., 2012. Big-data - Theoretical, engineering and analytics perspective. In: Srinivasa, S., Bhatnagar, V. (Eds.), Big Data Analytics. Vol. 7678 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, pp. 8-15.
2. J. Manyika, M. Chui, J. Bughin, R. Dobbs, P. Bisson, and A. Marrs, Disruptive technologies: Advances that will transform life, business, and the global economy. McKinsey Global Institute San Francisco, CA, 2013, vol. 180.
3. K. Panetta. (2016) Gartner's top 10 strategic technology trends for 2017.[Online]. Available: <http://www.gartner.com/smarterwithgartner/gartners-top-10-technology-trends-2017/>
4. M. Mohammadi and A. Al-Fuqaha, "Enabling cognitive smart cities using big data and machine learning: Approaches and challenges," IEEE Communications Magazine, vol. PP, no. 99, pp. 1-8, 2017.
5. M. Chen, S. Mao, Y. Zhang, and V. C. Leung, Big data: related technologies, challenges and future prospects. Springer, 2014.
6. X.-W. Chen and X. Lin, "Big data deep learning: challenges and perspectives," IEEE Access, vol. 2, pp. 514-525, 2014.

7. M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. Mccauley, M. Franklin, S. Shenker, and I. Stoica, "Fast and interactive analytics over hadoop data with spark," *USENIX Login*, vol. 37, no. 4, pp. 45–51, 2012.
8. Margette Rouse (27 April 2010) "Managing Hadoop Project"
9. C.-W. Tsai, C.-F.Lai, M.-C. Chiang, L. T. Yang et al., "Data mining for internet of things: A survey." *IEEE Communications Surveys and Tutorials*, vol. 16, no. 1, pp. 77–97, 2014.
10. C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Context aware computing for the internet of things: A survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 414–454, 2014.
11. M. A. Alsheikh, S. Lin, D. Niyato, and H.-P. Tan, "Machine learning in wireless sensor networks: Algorithms, strategies, and applications," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1996– 2018, 2014.
12. Z. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art deep learning: Evolving machine intelligence toward tomorrow's intelligent network traffic control systems," *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, 2017.
13. Maryam M Najafabadi, Flavio Villanustre, Taghi M Khoshgoftaar, Naeem Seliya, Randall Wald and Edin Muharemagic, "Deep Learning Applications and Challenges in Big Data Analytics" *Springer* February 2015
14. M. Hilbert, "Big data for development: a review of promises and challenges," *Development Policy Review*, vol. 34, no. 1, pp. 135–174, 2016.
15. W. Fan and A. Bifet, "Mining big data: current status, and forecast to the future," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 2, pp. 1–5, 2013.
16. AbhinavAyalur, Isaac Wilcove, Lynn Dang, Ricky Avina "Speed to Safety: Autonomous RC Car Aids Emergency Evacuation" August 2017
17. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing big data issues in scientific data infrastructure," in *Collaboration Technologies and Systems (CTS)*, 2013 International Conference on. *IEEE*, 2013, pp. 48–55.
18. M. Strohbach, H. Ziekow, V. Gazis, and N. Akiva, "Towards a big data analytics framework for iot and smart city applications," in *Modeling and processing for next-generation big-data technologies*. Springer, 2015, pp. 257–282.
19. G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
20. L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1–29, 2014.
21. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2347–2376, 2015.
22. Agrawal, R., Srikant, R., 1994. Fast algorithms for mining association rules in large databases. In: *Proceedings of the 20th International Conference on Very Large Data Bases. (VLDB '94)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, pp. 487-499.
23. QingchenZhang , Laurence T. Yang ,ZhikuiChen,Peng Li, 2017 . "A survey of deep learning for big data" July 2018 in Elsevier.
24. Mehdi Mohammadi, Ala Al-Fuqaha, SamehSorour, Mohsen Guizani "Deep Learning for IoT Big Data and Streaming Analytics: A Survey" December 2017.
25. "Apple watch Detects Diabetes" February 2018 News Developer
26. Vincent Granville on January 2, 2017 "Difference between Machine Learning, Data Science, AI, Deep Learning, and Statistics"