PalArch's Journal of Archaeology of Egypt / Egyptology

BIG DATA ANALYTICS WITH MACHINE LEARNING FOR IOT SENSED DATA

K.C. Arun[1], Mubashir Ali[2], Ghulam Irtaza[3], Talha Anwar[4]*

[1*]School of Computing, Asia Pacific University of Technology and Innovation, Malaysia.

[2]Department of Software Engineering, Lahore Garrison University, Lahore, Pakistan.

[3]Department of Information Sciences, University of Education Lahore, Mulan Campus,Pakistan.

[4]Department of Computer Science, National University of Computer and Emerging Sciences,Pakistan.

[1*]kchandran.arun@staffemail.apu.edu.my,[2]dr.mubashirali1@gmail.com,[3]ghulam.irtaza@ue.edu.pk,[4]chtalhaanwar@gmail.com

K.C Arun, Mubashir Ali, Ghulam Irtaza, Talha Anwar. Big Data Analytics with Machine Learning for Iot Sensed Data-- Palarch's Journal of Archaeology of Egypt/Egyptology 17(7), 5293-5301. ISSN 1567-214x

Keywords: Internet of Things, Big Data Analytics,Machine Learning, Sensed Data, Data Processing.

ABSTRACT

Internet of things (IoT) is the most encouraging automation technology and it producing large amount of sensed data. It is not possible to process this data by humans. Big data technology come out as a key data analytics conception that magically process the data for required outputs and actions. Internet of things means connecting things that are present in the world with internet. Big data is important as many public and private organizations gather huge amounts of domain-specific information,which may contain useful information on issues such as intelligence, cyber security, fraud detection, marketing and medical information.Investigation of the enormous data demands more effort for different quantity to take out the awareness for the determination. Machine learning is the best way to know about hidden relationships between data, because it works with a scale machine and works well with a large data set. Big data Analytics and deep learning are two areas that focus heavily on data science. Deep learning algorithms produce high-quality, robust inferences such as data support for a hierarchical learning process. This paper gives comprehensive view about machine learning contributions in processing IoT sensed data with big data analytics. This research work shows the big data algorithms and analytics toolsthat are effectively analyze the data of resource constrained IoT Sensed data.

INTRODUCTION

Internet of things environment requires the standardization that they are accountable, compatible, reliable and very effective on the global scale[1]. In this automation world, the new emerged technology is internet of things (IoT) which is revolutionizing the real-world objects (i.e. cars, home appliances, baby monitor, etc.) into automated virtual objects. It is the reassuring technology in the present age[2]. Internet of things is identified by adopting good and self-build object which can cooperate with each and every one through international chain infrastructure. Data telemetry with different sounds such as pressure, temperature, illumination, and object detection is one of the powerful ways to solve many complex problems by integrating it with Big Data. Hence, these communications between vast amounts of composite item, shows internet of things as an evolving technology which makes easy universal and common computingoperations. The instant growth of IoT sensed data cause the rapid increase of the Big Data[3]. The huge amount of networking sensors continues to collect the transmitted data to store and process in the cloud. This data should be of any kind, for example it should be graphical data, historical data, experimental data, logistic data and the astronomical facts etc.[4]. The primary data practicable are phone accessory, shipment facilities, audience facilities, and the home devices[5]. Management of the rapid growing of bigdata is the big threat for the community in generals[6]. After that, taking out the suitable data from IoT, data can be used to improve our daily lives with condition cognitive environmentfunctions.

Big Data is limited according to the necessary elements which are volume, variety, velocity, precision, and value. These are termed as 4V's[7] to highlight the cause effect relationship. First of all, Volume which revenue the size of the facts, Variety/diversity; which meanscollection of a different type of facts from the many sources, Velocity; which means toward collect the data in the real time, Veracity; which means the uncertainty of data and then Value which it means thatthe profit of the various industrial and the academic fields in the Huge Data. The next figure shows the interpretation and uniqueness of Big data.

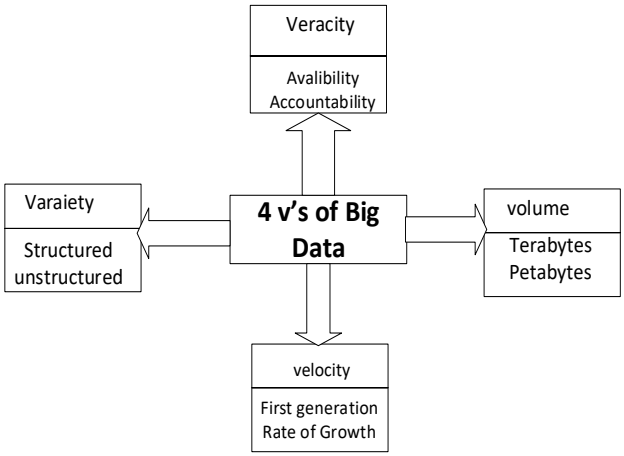

Figure 1. Uniqueness of BigData[8]

Recently the Big Data research has been undertaking original and the real alteration as of its study, collect toward its highest collision and application in many areas blending the Big Data and IoT.Big data to enable knowledge discovery and decision making. This technology has been creating the great opportunity for the progress of the difficult system in the small cities. Moreover, the Big Data technology have emerged the giving out of the huge volume excess of the Internet of things technology.Big data will change how even the smallest companies do business as data archive and explanation become more accessible. New, inventive, and cost-effective advancement and developing that make it very easy for any organization to implement big data strategies. All this data are together as of the many sources in the elegant atmosphere. However, the improvement of web of things and big data system in several domain causes the important elevation of the many types of facts. Big data is the powerful momentum for the later formation of IT management, that are generally manufactured on the tertiary stage, mainly attributing to these i.e.; big data, web of things, cloud computing, etc. Commonly, Data depot have been utilized to handle the wide dataset. In view of this, obtaining the actual observation from feasible big data area big problem[7].Big data depends upon data analytics. Data analysis includes various methods, technologies, and equipment such as text analysis, market intelligence, data visualization, and statistical analysis. The basic issue in the searching of big data is shortage of strategy among the database organization along with search apparatus like data drilling and numerical analysis [8]. The above mentioned challenges mostly appear when we want to execute knowledge analysis along with depiction for functional operations. The fundamental issue is that how to significantly illustrate the fundamental aspects relating big data. Big Data represents the problem areas and techniques used by application domains that collect and store large amounts of raw data for domain-specific data analysis [9]. Modern technologies requiring data and additional data storage resources have contributed significantly to the development of Big Data science. Technology-based companies such as Google, and Amazon have collected and stored big or large data. In addition, social media organizations such as YouTube, and Twitter have billions of users that constantly generate huge amounts of data [10]. Many organizations have invested in product development using Big Data Analytics to monitor their monitoring, evaluation, data analysis, reproduction, and other information and business needs, making it a major topic in data science research.

The clarification of the above mentioned issues is that there is the requirement of epistemological conclusion for interpretation of the data transformation[11]. Moreover, studying the complexity of big data will help us to understand the basic elements along with the development of difficult arrangements in Big data, clear up its portrayal, draw exceptional comprehension, andcomputingsketch is been guided by this and the last, conclusion on big data[12].

Machine learning is a subset of computer science used to analyze information. Machine learning (ML) is the important constituent of data analytics. It has been stated by McKinsey Global Institute that ML will play the role of master key in Big Data revolution. With advancement in technology Machine

learning didn't integrate well with large amount of data. What causes is it its potential to get information from data and provide decisions, insights and predictions based on data.It works with statistical approach and can help predict modes of data, at the same time it doesn't need exact statistical proofs. Learning task is categorized on the basis of two principles that are; Supervised learning is often applied to problems where data must be segmented, and is based on a model of defined classification that the computer has to learn from. Specifically, cross-sectional learning is useful for all problems depending on the program reduction. It may not be necessary for pre-defined separation rules to be defined, Unsupervised learning is very complicated because the computer has to learn how to do the work without getting instructions. This kind of learning is probably about the real world, rationalization it. In last few years, many new machine learning techniques have been introduced which can deal with big data. Some examples from our daily life of how machine learning is being used are:For example, in 2014, the Netflix Recommendation System was just based on restricted Boltzmann machine and kind of matrix factorization.

LITERATURE REVIEW

Many experiments have been carried out related tothe Big Data. Here we will discuss the work of some authors who has been working on Big Data Algorithms. Hind Bangui et al discussed the relation between Big Data, Clustering Algorithms, Internet of things and their Applications. They examine which Big Data technologies can be successfully used withIOT environment[13]. The authors Salimur et al proposed Big Data terminologies based upon Internet of Things and categorization of Big Data analytics. Due to this, the composition of a different data types is important for progress in current era. Furthermore, the combination is mannered in many challenges likewise data arrangement, unnecessary data etc. Mohsen et al proposed what is the interaction of the Internet of things with Big Data. They also examined consultant by discussing different opportunities accomplished by data analytics in internet of things. They further discussed multiple open research tasks regarding Big Data Algorithms. Finally they got result that the current Big Data analytics problems that have been solved out remained in their initial phases of development[14]. Pooya et al implemented big data strategies. They discussed the challenges met by Big data analysts and guided he right maps to riches go through effective execution of Big data policies. They easily presented an understandable narration of significant Big Data algorithms and explained their effective uses through different factual examples[15]. Gameil et al discovered the possible effect of Big Data arouses, Open research problems and the several devices related to Big data and IOT. Riccardo et al experienced a suggested model using bootstrapped qualified mediation analysis. They got the outcome that ambidexterity and agility both are connected tostructural big data analytics (BDA)[16].Nelson et al analyzed the current connection in between Big data analytics skills and co-invention. They used fundamental equations by the fractional least squares method to test hypothesis prototype[17]. Loubna et al put light on Big data analytics and its purpose in different areas like Business Process Management(BPM), Telecommunication, and Human Resources Management(HRM)[18]. Hong-Yen et al proposed a complete review of Privacy-preserving big data analytics.

They introduced well-made classifications which proposed both organized views and a comprehensive categorization of this stimulating research field[19].

BIG DATA ARCHITECTURE

Big data is extensive system developed to organize large amount of data, so that data can be interpreted to be used in business oriented organizations. The architecture can be regarded as a draft for the big data and is drafted in such a way to meet needs of organizations. The volume of the data that is accessible for interpretation, varies daily and now there are more streaming platforms available including the data obtained from different algorithms. But accessing the data doesn't mean you have progressed, you must know how to put that data in right use. Big data architecture is developed to handle the following types of work:

1. Real time processing of the big data.
2. Predictive analytics and internet of things.
3. Batch processing of big data sources.

A well-designedbig data architecture can be help an organization to make profit and to secure their future success. As everything has pros and cons,bigdata gives more access to data, on the other handit might corrupt privacy of your data with activity like cybercriminal could cause loss of your data.

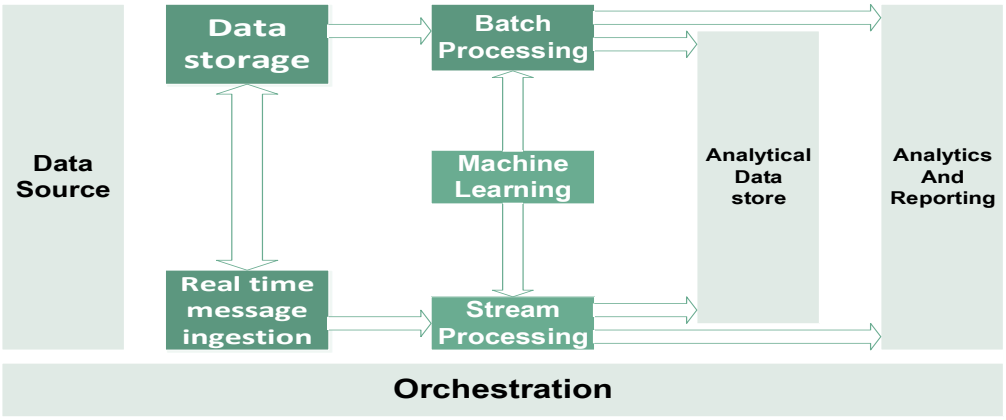

Figure 2. Big Data Orchestration with Machine Learning[18]

CONTRIBUTIONS

Sr.No	Authors & Year	Sub-Domains	Contributions
1.	Pooya et al. 2019	Clustering Segmentation	Authors worked on different algorithms i.e.; clustering segmentation for (e.g.; customer, employees, product etc.) management[15].
2.	Weiwei et al. 2019	Ensemble Random Forest Algorithm for Analysis of Big Data Insurance	The random forest algorithm will work with standard estimation, and has the advantage of the automatic selection feature etc. [16].
3.	Loubna et al. 2019	Big Data in Telecom Sector	Author implemented BDA for multiple applications such as Telecommunication, Business Process Management (BPM) and Human Resources

			Management, (HRM) [18].”
4.	Hongjian et al. 2019	Energy-Efficient Scheduling Algorithms	Energy consumption has rapidly increased with the fast development in big data processing in anti-cybercrime. An energy performance design for big data applications is suggested to overcome energy crisis while cooperating with restrains[19].”
5.	Alan et al. 2019	Big Data in Visual Investigation	They emphasized on choice of primary thoughts said byBertin and implemented by many scientists after some time, with explicit thought to how the thoughts identify with developments in cartography, big data, and visual investigation [20].”
6.	Bin et al. 2019	Segmentation Algorithm for Agricultural Image Big Data	Image segmentation is the process of splitting up an image into a number of specific areas and selecting the objects required in the image by computer[21].
7.	Meifanzhang et al. 2019	One-Pass Inconsistency Detection Algorithms for Big Data	Big data has to face data quality problems due to feature like volume. Due to which, this algorithm was proposed by these authors to overcome this issue [22].
8.	Hind et al. 2018	Big Data for Internet Of Things	They explain the similarity and difference between Big Data innovations utilized in various IoT fields and builds up a calculated structure to diagram the risky Big Data advances over all the audited IoT fields[13].”
9.	Marcello et al. 2018	Big data in Hospitality and Tourism	These author analyzed the term wherein Business Intelligence and Big Data by grouping research holes, future advancement and planning for future research[23].
10.	Jingjing et al. 2018	Big data in IoT Tourism	For every kind of tourism, a systemized overview is classified from the parts of research centers, information qualities, investigative strategies, fundamental difficulties and additionaldirections[24].”
11.	Tsan-Mingi et al. 2018	Big Data Analytics inOperationsManagement	They first found the current big data-related investigation strategies, and identified their qualities, shortcomings and furthermore they determined the investigation with a dialog of future research [25].”
12.	Niraj et al. 2018	Big Data Analytics in IoT SensedSupplyChain Management	They delivered a systematic literature review (SLR) of Big Data Analysis (BDA) capacities in store network and build up the abilities experience/development model and gave the certain bearing of research in this field[26],.”
13.	Luca et al. 2018	Train Delay Prediction Systems: A Big Data Analytics Perspective	They proposed information driven Train Delay Prediction System (TDPS) for huge railroad systems with machine learning which built up the most of the present big data advances, studying calculations, and arithmetical instruments[27].
14.	Mohamed et al. 2018	Big Data in IOT Based Health Care	They proposeda reproductive model for protecting textual information from medical images with the combination of Rivets, Shamir, and Adelman algorithms [28].
15.	Radwa et al. 2018	Big Data Systems in Clouds Administration Model	They gave a perceptions about the cutting edge continuous developments and open difficulties in this field[29]”.
16.	Dazhong et al. 2018	Algorithm for Saving Cloud-based Parallel Space-Big Data	Proposed algorithm is applied to the cloud for the analysis of large network data, and then a test number formula for the opposition is extracted to properly store the Top-k objects [30].
17.	Kai Chan et al. 2017	Big Datain Risk Management	They presented the problems and chances of Big Data Analytics(BDA) in this specific application field, their operational risk management(ORM) and Sequential Patterns Discovery Recommendation systems[31]”.
18.	Fergus et al. 2017	Big Data and Data Repurposing	They determined the methodologies of data that can help in defining the causes and analysis of data repurposing.[32].

19.	Z. Lv et al. 2017	Big Data Analytics in Storage Model Management,	They, reviewed current, examine in information types, storages models, protection, information security, examination techniques, and applications [33].
20.	Gameil et al. 2017	Big Data Analytics in Regression prediction of a variable of interest,	They analyzed the possible effect ofbig datachallengesthrough study of issues. They also worked for the working of the algorithm of,Predicting Interest Rate Variation (e.g., Sales Price, Consumer Performance) [7] ".

CONCLUSIONS

Internet of things is most emerging technology that is producing very huge amount of sensed data. This sensed data is used for taking actions in embedded systems. Quick and efficient data processing needed for embedded systems and it is not possible to process sensed data by humans. Big data analytics is another growing technology that provides automatic data processing and analytics according to given conditions. Machine learning enables the machines and devices to learn from environment and take decisions accordingly. Machine learning techniques are used in data analytics to efficiently process the sensed data. In this paper, we have explored interdependency of machine learning and big data analytics for IoT sensed data. Furthermore, the contributions in this regard are highlighted with domain wise categorization. This paper shows the state of the art contribution of big data analytics to process the IoT sensed data.

REFERENCES

Dorsemaine, B., Gaulier, J.P., Wary, J.P., Kheir, N., &Urien, P. (2015). Internet of things: a definition & taxonomy. *In 9th International Conference on Next Generation Mobile Applications, Services and Technologies*, 72-77.

Ali, M. (2019). IOT Based Architecture for Basketball Supervision. *LGURJCSIT*, *3*(4), 30-38.

Ahmed, M., Choudhury, S., & Al-Turjman, F. (2019). Big Data Analytics for Intelligent Internet of Things. *In Artificial Intelligence in IoT*, *Springer, Cham*, 107-127.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile networks and applications*, *19*(2), 171-209.

Islam, M., & Reza, S. (2019). The Rise of Big Data and Cloud Computing. *Internet of Things and Cloud Computing*, *7*(2), 45.

The Rise and Rise of Big Data Analytics - theceoviews. https://theceoviews.com/the-rise-and-rise-of-big-data-analytics/.

Ali, G.,& Nithya, A. (2017). Challenges and OpenResearch Issues and Tools on Big Data Analytics, *6*(11), 1690–1703.

Acharjya, D.P., & Ahmed, K. (2016). A survey on big data analytics: challenges, open research issues and tools. *International Journal of Advanced Computer Science and Applications*, *7*(2), 511-518.

Das, T.K., & Kumar, P.M. (2013). Big data analytics: A framework for unstructured data analysis. *International Journal of Engineering Science & Technology*, *5*(1), 153.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International journal of information management*, *35*(2), 137-144.

Big Data Software and Analytics Tools and Solutions | Micro Focus.

https://www.microfocus.com/en-us/solutions/big-data-analytics-software.

Huang, Z., & Huang, Z. (1997). A Fast Clustering Algorithm to Cluster Very Large Categorical Data Sets in Data Mining. *Res. ISSUES DATA Min. Knowl. Discov.*, 1-8.

Ge, M., Bangui, H., & Buhnova, B. (2018). Big data for internet of things: A survey. *Future generation computer systems*, *87*, 601-614.

Marjani, M., Nasaruddin, F., Gani, A., Karim, A., Hashem, I.A.T., Siddiqa, A., & Yaqoob, I. (2017). Big IoT data analytics: architecture, opportunities, and open research challenges. *IEEE Access*, *5*, 5247-5261.

Tabesh, P., Mousavidin, E., & Hasani, S. (2019). Implementing big data strategies: A managerial perspective. *Business Horizons*, *62*(3), 347-358.

Rialti, R., Zollo, L., Ferraris, A., & Alon, I. (2019). Big data analytics capabilities and performance: Evidence from a moderated multi-mediation model. *Technological Forecasting and Social Change*, *149*, 119781.

Lozada, N., Arias-Pérez, J., & Perdomo-Charry, G. (2019). Big data analytics capability and co-innovation: an empirical study. *Heliyon*, *5*(10), e02541.

Rabhi, L., Falih, N., Afraites, A., & Bouikhalene, B. (2019). Big data approach and its applications in various fields. *Procedia Computer Science*, *155*, 599-605.

Tran, H.Y., & Hu, J. (2019). Privacy-preserving big data analytics a comprehensive survey. *Journal of Parallel and Distributed Computing*, *134*, 207-218.

MacEachren, A.M. (2019). (re) Considering Bertin in the age of big data and visual analytics. *Cartography and Geographic Information Science*, *46*(2), 101-118.

Liu, B., He, S., He, D., Zhang, Y., & Guizani, M. (2019). A Spark-Based Parallel Fuzzy $ c $-Means Segmentation Algorithm for Agricultural Image Big Data. *IEEE Access*, *7*, 42169-42180.

Zhang, M., Wang, H., Li, J., & Gao, H. (2019). One-pass inconsistency detection algorithms for big data. *IEEE Access*, *7*, 22377-22394.

Mariani, M., Baggio, R., Fuchs, M., & Höepken, W. (2018). Business intelligence and big data in hospitality and tourism: a systematic literature review. *International Journal of Contemporary Hospitality Management,Emerald Group Publishing Ltd*, *30*(12), 3514–3554.

Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). Big data in tourism research: A literature review. *Tourism Management*, *68*, 301-323.

Choi, T.M., Wallace, S.W., & Wang, Y. (2018). Big data analytics in operations management. *Production and Operations Management*, *27*(10), 1868-1883.

Arunachalam, D., Kumar, N., & Kawalek, J.P. (2018). Understanding big data analytics capabilities in supply chain management: Unravelling the issues, challenges and implications for practice. *Transportation Research Part E: Logistics and Transportation Review*, *114*, 416-436.

Oneto, L., Fumeo, E., Clerico, G., Canepa, R., Papa, F., Dambra, C., & Anguita, D. (2018). Train delay prediction systems: a big data

analytics perspective. *Big data research*, *11*, 54-64.

Elhoseny, M., Ramírez-González, G., Abu-Elnasr, O.M., Shawkat, S.A., Arunkumar, N., & Farouk, A. (2018). Secure medical data transmission model for IoT-based healthcare systems. *IEEE Access*, *6*, 20596-20608.

Elshawi, R., Sakr, S., Talia, D., & Trunfio, P. (2018). Big data systems meet machine learning challenges: towards big data science as a service. *Big data research*, *14*, 1-11.

He, D., Yang, Y., & Liu, J. (2018). A Cloud-Based Parallel Space-Saving Algorithm for Big Networking Data. *IEEE Access*, *6*, 45886-45898.

Choi, T.M., Chan, H.K., & Yue, X. (2016). Recent development in big data analytics for business operations and risk management. *IEEE transactions on cybernetics*, *47*(1), 81-92.

Doubal, F.N., Ali, M., Batty, G.D., Charidimou, A., Eriksdotter, M., Hofmann-Apitius, M., & Ritchie, C.W. (2017). Big data and data repurposing-using existing data to answer new questions in vascular dementia research. *BMC neurology*, *17*(1), 1-10.

Lv, Z., Song, H., Basanta-Val, P., Steed, A., & Jo, M. (2017). Next-generation big data analytics: State of the art, challenges, and future research topics. *IEEE Transactions on Industrial Informatics*, *13*(4), 1891-1899.