A survey on Big Data Analytics

D.Sasirega

Department of Computer Science
KG College of Arts and Science, Coimbatore.

*Abstract— T*he concept of electrocardiogram is explained. Then, a problem statement based on manufacturing scenario is presented. Subsequently, the architecture of proposed algorithm called integrated deep denoising auto-encoder (IDDA) and algorithm workflow are provided. Moreover, DECG is compared with traditional factory information system, and the feasibility and effectiveness of proposed algorithm are validated experimentally. The proposed concept and algorithm combine typical industrial scenario and advance artificial intelligence, which has great potential to accelerate the implementation. In the context of Industry 4.0, industrial robotics such as automated guided vehicles have drawn increased attention due to their automation capabilities and low cost. With the support of cognitive technologies for industrial Internet of Things (IoT), production processes can be significantly optimized and more intelligent manufacturing can be implemented for smart factories. The explosive growth in the number of devices connected to the Internet of Things (IoT) and the exponential increase in data consumption only reflect how the growth of big data perfectly overlaps with that of IoT. The management of big data in a continuously expanding network gives rise to non-trivial concerns regarding data collection efficiency, data processing, analytics, and security. The advances in wireless communication technologies, vehicular networks and cloud computing boost a growing interest in the design, development and deployment of Vehicular Cyber-Physical Systems (VCPS) for some emerging applications, which leads to an increasing demand on connecting Mobile Cloud Computing (MCC) users to VCPS for accessing the richer applications and services. Fault diagnosis is an important topic both in practice and research. There is intense pressure on industrial systems to continue reducing unscheduled downtime, performance degradation, and safety hazards, which requires detecting and recovering from potential faults as early as possible.

Keywords: Big Data, Data Analytics.

INTRODUCTION

Big Data Analytics

The volume of knowledge that one should deal has exploded to impossible levels within the past decade, and at constant time, the worth of knowledge storage has consistently reduced. non-public firms and analysis establishments capture terabytes of knowledge concerning their users' interactions, business, social media, and conjointly sensors from devices like mobile phones and vehicles. The challenge of this era is to form sense of this ocean of knowledge. This is often wherever huge knowledge analytics comes into image.

Big knowledge Analytics mostly involves aggregation knowledge from totally different sources, munge it in a very approach that it becomes accessible to be consumed by analysts and eventually deliver knowledge product helpful to the organization business.

In order to produce a framework to prepare the work required by a company and deliver clear insights from huge information, it's helpful to think about it as a cycle with completely different stages. it's by no means that linear, which means all the stages square measure connected with one another. This cycle has superficial similarities with the a lot of ancient data processing cycle as delineated in CRISP.

Fresh DM was framed in 1996 and one year from now, it got present as an Union undertaking underneath the dapperness subsidizing activity. The undertaking was light-radiating diode by 5 organizations: SPSS, Teradata, Daimler AG, NCR Corporation, and OHRA (an insurance agency). The task was at last consolidated into SPSS. The approach is extremely detailed acquainted in anyway an information mining venture should be specific. Let us presently take in somewhat more on every one of the stages engaged with the CRISP-DM life cycle.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

Business Understanding − This underlying stage centers around understanding the venture goals and necessities from a business point of view, and after that changing over this learning into an information mining issue definition. A primer arrangement is intended to accomplish the goals. A choice model, particularly one manufactured utilizing the Decision Model and Notation standard can be utilized.

Information Understanding − The information understanding stage begins with an underlying information gathering and continues with exercises with the end goal to get acquainted with the information, to distinguish information quality issues, to find first experiences into the information, or to recognize fascinating subsets to shape speculations for concealed data.

Information Preparation − The information planning stage covers all exercises to build the last dataset (information that will be sustained into the demonstrating tool(s)) from the underlying crude information. Information planning assignments are probably going to be played out various occasions, and in no recommended request. Undertakings incorporate table, record, and property determination and additionally change and cleaning of information for demonstrating devices.

Demonstrating − In this stage, different displaying methods are chosen and connected and their parameters are aligned to ideal qualities. Ordinarily, there are a few systems for similar information mining issue type. A few procedures have particular prerequisites on the type of information. Along these lines, usually required to venture back to the information readiness stage.

Assessment − At this phase in the task, you have manufactured a model (or models) that seems to have high caliber, from an information investigation point of view. Prior to continuing to definite arrangement of the model, it is imperative to assess the model completely and audit the means executed to build the model, to be sure it appropriately accomplishes the business destinations.

A key goal is to decide whether there is some imperative business issue that has not been adequately considered. Toward the finish of this stage, a choice on the utilization of the information mining results ought to be come to.

Organization − Creation of the model is by and large not the finish of the task. Regardless of whether the motivation behind the model is to build learning of the information, the information picked up should be sorted out and displayed in a way that is valuable to the client.

Contingent upon the necessities, the organization stage can be as basic as creating a report or as mind boggling as executing a repeatable information scoring (e.g. fragment assignment) or information mining process.

The volume of information that one needs to bargain has detonated to incomprehensible dimensions in the previous decade, and in the meantime, the cost of information stockpiling has efficiently lessened. Privately owned businesses and research establishments catch terabytes of information about their clients' collaborations, business, web based life, and furthermore sensors from gadgets, for example, cell phones and cars. The test of this time is to understand this ocean of information. This is the place enormous information investigation comes into picture.

Big Data Analytics to a great extent includes gathering information from various sources, munge it in a way that it winds up accessible to be devoured by experts lastly convey information items helpful to the association business. SEMMA Methodology

- SEMMA is another strategy created by SAS for information mining demonstrating. It remains for Sample, Explore, Modify, Model, and Asses.

Conclusion

Big Data examination as the name recommend is the investigation of examples or extraction of data from huge information. Thus, in enormous information examination, investigation is done on huge information. From the perspective of applicability in the big data era, we point out the limitations of previous diagnostic methods, including the first two approaches. As the data volume increases exponentially, hidden important values lead researchers to explore more advanced and effective approaches to predict and provide diagnosis of device failures.

International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
International Conference on Recent Explorations in Science, Engineering And Technology (ICRESET'19)
Volume- 5, Special Issue- March, 2019

References

[1]   MACHINE LEARNING AND ITS APPLICATIONS: A REVIEW AUTHOR(S) SHEENA ANGRA ; SACHIN AHUJA

{2}   A SURVEY OF DATA MINING AND MACHINE LEARNING METHODS FOR CYBER SECURITY INTRUSION DETECTION
        Author(s) Anna L. Buczak ; Erhan Guven

[3]   A SURVEY OF MACHINE LEARNING FOR BIG DATA PROCESSING JUNFEI QIU, QIHUI WU, GUORU DING, YUHUA XU, SHUO FENG

[4]   A REVIEW OF RELATIONAL MACHINE LEARNING FOR KNOWLEDGE GRAPHS M NICKEL, K MURPHY, V TRESP… - …

[5]   SIGNIFICANCE AND CHALLENGES OF BIG DATA RESEARCH(ARTICLE) JIN, X., WAH, B.W., CHENG, X.,

[6]   Towards a Comprehensive Data Analytics Framework for Smart Healthcare Services Sherif Sakr | Amal ElgammalJune 2016

[7]   Smart4Job: A Big Data Framework for Intelligent Job Offers Broadcasting Using Time Series Forecasting and Semantic Classification Sidahmed Benabderrahmane | Nedra Mellouli | Myriam Lamolle | Patrick ParoubekMarch 2017

[8]   Efficient Resource Management System Based on 4Vs of Big Data Streams Navroop Kaur | Sandeep K. Sood September 2017

[9]   Frequent Itemsets Mining for Big Data: A Comparative Analysis Daniele Apiletti | Elena Baralis | Tania Cerquitelli | Paolo Garza | Fabio Pulvirenti | Luca Venturini September 2017

[10]   Augmenting Data Warehouse with BigData A Survey, Sarala.R, International Journal of Multidisciplinary Educational Research(IJMER), Vol 6, Issue:10 (3), ISSN:2277-7881 October 2017

[11]   J. Wan, S. Tang, Q. Hua, D. Li, C. Liu and J. Lloret, "Context-Aware Cloud Robotics for Material Handling in Cognitive Industrial Internet of Things," IEEE Internet of Things Journal, to be published, DOI: 10.1109/JIOT.2017.2728722, 2017.

[12]   E. Ahmed, I. Yaqoob, I. Hashem, I. Khan, A. Ahmed, M. Imran and A. Vasilakos, "The role of big data analytics in Internet of Things," Computer Networks, vol. 129, pp. 459-471, 2017.

[13]   J. Wan, D. Zhang, Y. Sun, K. Lin, C. Zou and H. Cai, "VCMIA: A Novel Architecture for Integrating Vehicular Cyber-Physical Systems and Mobile Cloud Computing," Mobile Networks and Applications, vol. 19, no. 2, pp. 153-160.

[14]   E. Ahmed, A. Ahmed, I. Yaqoob, J. Shuja, A. Gani, M. Imran, M. Shoaib, "Bringing computation closer towards user network: Is edge computing the solution?" IEEE Communications Magazine, vol. 55, no. 11, pp. 138-144, 2017.

[15]   Y. Xu, Y. Sun, J. Wan, X. Liu and Z. Song, "Industrial Big Data for Fault Diagnosis: Taxonomy, Review, and Applications," IEEE Access, to be published, DOI: 10.1109/ACCESS.2017.2731945, 2017.