

Review of Challenges in Big Data Mining

Abdullah Alsulami

Monash University, Australia
abu7ala1@hotmail.com

Abstract: Nowadays managing big data is one of the most challenges in data management. Big data mining has become a major focus of data management in the world helping to provide detailed data that informs major decisions, to adequately understand the concept of big data mining. This research paper is analysis four research articles and conferences papers from various aspects of big data mining regarding their approaches, issues, findings and improvements that articles have had in the field of data mining, and addresses the gap in data mining literature. The large volumes of data and the various types, and sources of data that act as barriers to big data mining will be discussed in details. The application of big data mining in clinical medicine will be discussed as well to identify how it improves the clinical decision making and practices. The methods that have been used to analyse and increase the efficiency and accuracy of the data mining process will be identified to provide a critical information to guide industries to make the proper decision.

INTRODUCTION

Big data mining can be described as the process of extracting useful information from large datasets (Alfredo, 2015). Big data mining has become a major focus of data management in the world helping to provide detailed data that informs major decisions, to effectively understand the concept of big data mining the paper analyzed four research articles and conferences papers they highlight the various aspects of big data mining. The paper analyses four research articles regarding their approaches to defining aspects of big data mining, the issues being focused on, the findings and the improvement the paper has had in the field of data mining. The analysis of these articles will help to provide original information that addresses the gap in data mining literature

PROPOSED APPROACHES

Zhang, Guo, Han & Li, (2016) took a theoretical approach to discussing the application and exploration of big data mining in clinical medicine. The study sought to analyze the theoretical framework and the technologies used in big data mining. This approach sought to take an objective look at big data mining theories such as fuzzy theory, rough set theory, and cloud theory.

Wang, Yang, Luo, He & Tan, (2015) took a methodological approach to explaining the application of big data mining and its use in risk warning for food safety. The approach sought to analyze the application of big data mining in the field of food safety through the use of three data mining methods. The three methods were analyzed and each identified on its effectiveness in improving risk warning in the food safety industry. The study, therefore, compared these methodological approaches to data mining and their application to food safety.

Wang & Yuan, (2014) article took an explanatory approach into defining big data mining through the perspective of spatial data mining. The descriptive approached played a key role in explaining spatial data mining its challenges, benefits and its application in various industries. Sebastian, Babu, & Kizhakkethottam (2015) took an analytical approach to determining the major challenges facing big data mining. These approach aimed at analyzing all the factors associated with big data mining such as storage, use, and transfers to determine the extent to which they favor or limit the development of big data mining.

CHALLENGES OR ISSUES

Zhang et al. (2016) focused on a new perspective in big data mining that is the review of basic theories as well as technologies that can be used in data mining for improving clinical medicine. The study focused on the how big data mining can be applied to the various field of clinical medicine such as clinical decision, support, disease risk assessment. Prediction of disease development, evidence-based medicine, medical management and rational use of drugs. The field of big data mining to improve medical practices had been extensively researched however studies into its application in clinical medicine has not. The study, therefore, sought to fill the gap in the literature by addressing the use of big data mining in this field. According to Wang et al. (2015), evaluation and warning are important factors in food safety. Therefore, the authors of this article sought to analyze the application of big data mining to improve measures in food safety. The methodological approach of these study was to analyze the application of big data mining through the use of three main approaches to determining that are Bayesian Network, Decision Tree, and artificial neural network. The authors conducted a comparative study of the application of these there data mining methods on their application in food safety risk warning. Wang & Yuan (2014) took a different approach to studying big data mining, and therefore their study focused on the use of spatial data mining in various aspects of life from academics to the business industry. The study, therefore, aimed at addressing the efficiency of spatial data mining in making data valuable. All these articles identify a different perspective in understanding big data mining. Sebastian et al. (2015) provided a review of challenges facing big data mining. The study focused on the challenges of growing amount of data and its storage, capture and transfer. The study focused on the availability of big data from different sources such as social media being too diverse making the data mining process difficult.

SUMMARY OF FINDINGS AND RESULTS

These articles provide important information to the field of big data mining. According to Zhang, et al. (2016), the main theories supporting big data mining are rough set theory, fuzzy theory, Dempster-Shafer theory, cloud theorem inductive learning theory, genetic algorithm and artificial neural network. The study identified that big data mining can provide adequate information to guide clinical practices and help physicians and health care providers to effectively deliver clinical services. The study concludes that that big data mining plays a critical role in various sectors such a clinical decision supports and evidence-based medicine. These research paper had a major effect on healthcare services since it identifies the use of big data mining as a technology that can be used to inform and improve clinical practices, therefore, opening the field of medicine to embrace the use of big data mining. As noted by Sebastian et al. (2015) the major challenges facing big data mining is the diversity in big data mining, it increasing evolution, its storage, and transfers. The study identified that various types of data are collected by autonomous sources, and the method of recording and coding the information is different for each source making big data mining as the different sources and types of data make mining difficult. The increased volume of data was noted to be one of the top challenges in data mining as the large volumes of data contain a variety of data sets such as texts, graphs, images among others, therefore, limiting the efficiency of big data mining. The impact that this paper has is that it effectively identifies the barriers to big data mining, therefore, helping the organization to set up measures to counter these barriers and improve the big data mining process. Wang et al. (2015) article identifies that the application of big data mining methods such as Bayesian Network, Artificial Neural Network and Decision tree and their application in the providing informative data that guides risk management in food safety. The study identifies that the use of BP neural network in food safety can be effective at identifying risk and providing early warnings improving food safety since it is superior to the other

two methods as it provides a better rating of risks enabling accuracy in the risk management process. This study has filled the gap in literature while at the same time identifying the most suitable data mining methodology to be used in the food safety industry to improve early risk detection. Wang & Yuan, (2014) concluded that spatial data mining is a useful approach to helping organizations determine interesting and unknown patterns in large spatial datasets. The study identified the use of spatial data in breaking down a large volume of data into smaller portions easily understandable and useful information that can be used to inform decisions. The paper improved on earlier perspectives into the various methods of big data mining identifying an effective and beneficial approach to big data mining that is through the use of spatial data mining.

JUDGMENT

The four articles portrayed a deep understanding of big data mining and approached the topic from different angles such as the application of big data mining in clinical medicine, spatial data mining, challenges of big data mining and its application in risk warning in food safety industry. On its own each article provides an in-depth analysis of the topic helping to provide accurate and useful results in addition to filling the gap in big data mining literature. According to Zhang, et al. (2016), big data mining can provide information that guides the decision-making process and improving the clinical practices in any health organization. Big data mining is a valuable tool especially for an industry such as the healthcare industry which is dependent on precise information. Big data mining, therefore, has the ability to improve the entire healthcare industry not just in clinical medicine helping to improve decision-making and provide information that guides the activities of healthcare practitioners (Sajana, Sheela Rani & Narayana, 2016). Wang et al. (2015) identified the use of BP neural network as an effective method in big data mining in food safety. The findings of the article are accurate as BP neural networks are more precise and timely helping to effectively identify risks and implement warning measures in food safety. Sebastian et al. (2015) identified that the increasing complexity of the different data sources, large volumes of data and storage act as barriers to effective big data mining. The article identifies the issue of large volumes of data increasing change and their different types as a major challenge in big data mining which is true since each day more data is collected and added to the large volumes already present means that big data mining methods must be adapted to overcome this challenges. Wang & Yuan, (2014) identified that the best approach to big data mining is the use of spatial data mining since it helped to categories data into spatial categories. The findings are adequate and sufficient since spatial data mining can reduce data to multi-tables, therefore, increasing the speed of big data mining.

CONCLUSION

The four articles identified various aspects of big data mining. The paper identified that the large volumes of data and the various types, as well as sources of data, act as barriers to big data mining. The paper identified that the application of big data mining in clinical medicine helps to improve clinical practices by providing critical information to guide clinical decisions and practices. The use of spatial data method in big data mining is an effective method of mining data since it increases the efficiency and accuracy of the data mining process. An analysis of the applicability of various data mining methods in risk warning for food safety identified that BP neural network is most effective data mining method due to its precision. Big data mining has potential in various industries from food safety, healthcare and business environment since it provides critical information that helps to transform the current activities in this environments and guides decision making enabling sustainability and development of these industries. Therefore, the study of its various aspects helps to fill gap in its literature and increase the understanding of the topic.

REFERENCES

1. Alfredo, C. (2015). Big data mining or turning data mining into predictive analytics from large-scale 3vs data: The future challenge for knowledge discovery. Lecture Notes in Computer Science 2014, Vol.8748, pp.4-8
2. Sajana, T, Sheela Rani, C., & Narayana, K. (2016). A survey on clustering techniques for big data mining. Indian Journal of Science and Technology, 9(3), 1-12.
3. Sebastian, L., Babu, S., & Kizhakkethottam, J. (2015). Challenges with big data mining: A review. Soft-Computing and Networks Security (ICSNS), 2015 International Conference on, 1-4.
4. Wang, Y, Yang, B, Luo, Y, He, J., & Tan, H. (2015). The Application of Big data Mining in Risk Warning for Food Safety. Asian Agricultural Research, 07, Asian Agricultural Research, August 2015, Vol.07.
5. Wang, S., & Yuan, H. (2014). Spatial Data Mining: A Perspective of Big Data. International Journal of Data Warehousing and Mining (IJDWM), 10(4), 50-70.
6. Zhang, Y, Guo, S., Han, L., & Li, T. (2016). Application and Exploration of Big Data Mining in Clinical Medicine. Chinese Medical Journal, 129(6), 731-738.

Citation: Abdullah Alsulami "Review of Challenges in Big Data Mining". American Research Journal of Computer Science and Information Technology, Volume 1, 2016; pp:1-4.

Copyright © 2016 Abdullah Alsulami, This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.