



Infrastructure of Data Mining Technique with Big Data Analytics

Manjula Pattnaik¹, Shahidafridi²

Professor, FBE, PNU, Princess Nourah Bint Abdul Rahman University Riyadh, KSA

²Annai College of Arts & Science, Tanjore, Tamilnadu, India

Email: drmanjula23@gmail.com¹, shahidafridics@gmail.com²

Abstract

Data mining is the concept of gathering new information from huge sets of data. It is also known as Knowledgeable Discover Database (KDD). In past few years business development in KDD are rapidly high in the market. Because it's processing is more useful in all kinds of business marketing field. However before the arrival of data mining, business marketing is slightly slow in process, at that time business marketing is more dependent on Television ads, sponsors then marketing executive etc... In late 90's this problem was fulfilled by the development of mining the data's concept. So in this article we are discussing about how the processing way of data mining collides with big data analytics.

Keywords: Big data, Business development, marketing and mining data's..

1. Introduction

In the age of computer science, big data's plays a crucial part in Information Technology (IT) environment. Because in recent period of time most of the people's are used a Smartphone as a personal computer (pc). All the features are available in that smart phone. Also a reach of social media in public mindset was so remarkable. Before the arrival of smart phones, social media names like face book and twitter are the most popular ones in that period. Now the arrival of smart phones social media's like face book, whatsapp, twitter, Instagram, Skype...etc are the most popular one's in the world right now. Across the globe, information's are stored as a data in data sets. Social media's like face book, whatsapp, twitter are used by millions of people, so huge amount of data's are needed for the social medias to store user data's in that application. Big data concept comes to overcome the issue of store large amount of data's while needed.

The process of gaining knowledge from huge set of data's is known as data mining with big data analytics. This data mining concept was included in the fields like business marketing, social media's like face book, twitter id etc... By the work of data mining it creates new information from large sets of data is the most significant one in this concept. This will help the data administrator to evaluate his idea. Online websites like Amazon, flip karts are mostly

dependent on data mining technique. Because once a customer brought one product from that site, a related kind of product has been shown in the notification section of any social media of that customer. Courtesy of data mining algorithm technique had written in that website. The raise of big data has been comes with the nature of using social medias. Face book is the best example of using big data algorithm. Nearly 1.5 billion peoples are following the face book according to the stats of Google. Talk about the concept of big data and algorithms, the size of the big data is evaluated as the range of 10^{15} byte. This kind of range is called big data. The big data's can be classified in to three v's. It is velocity, variety and volume. The only issue in big data algorithm is the large amount of unstructured data's needs to be stored, processed and analyzed. Hadoop Distributed File System (HDFS) overcome the issue of storing data in distributed ways. In today's world the raise of data mining algorithm is comes as a long way. The following fig (1.1) has shown some statistical review.

2. Literature survey

In [1] the status of current linear data techniques conveys the message of storing huge amount of data and volume is very difficult to manage the large data sets. So the evaluation of data mining inherits the way of approaching the big data techniques and its algorithms. The target of the proposed system was managed to support the mining large set of data's which reduced to a less amount of cost. In [2] privacy preserving data mining (PPDM) concept is the proposed model of this paper. In this model, it talks about the concern of security and privacy problems. Because for instance we use data mining algorithm in websites, privacy module is very important one to deliver such a dynamic websites.

In [3] it witnessed a concept of FPM based algorithm to enable the process of data skewness in data mining technique. In that algorithm the proposed model can be composing a one database scanning in data mining algorithm. In [4] in this model it executing the process of data mining with the mobile environment of broadcasting data services. It enforces the huge number of data's to convenient data structure in mobile cache services. The service is made up of cache and request model.

In [5] the process of making an application and techniques in to a trajectory data mining is the proposed model of this paper. The destination of this model was able to find a rhythm of interaction between data mining with mobile applications. In [6] this paper recognizes that the security protocol of social welfare organizations with data mining algorithms. In recent years, people are criticism from the government for social services security problems. In [7] analyze with big data from BP neural network is the proposed target for this paper. In [8] the raise of data mining algorithm in business and network system is the taking point of this paper. Then also discuss about, how the data mining process had done pass through the last four decades.

By the help of related papers, we have witnessed a data mining process had included in most of the business development. The uses of data mining with big data technique have reached the range of high level in Information technology (IT) industries in recent period. Some of the major uses of mining the data's are

- Performing as a advertisement sector in business management field.
- Reduce the time for marketing
- Store a huge set of data's
- Giving a new information from set of data's.

3. Proposed method

The structure of extracting information from large volume of big data's is the proposed model in this paper. The process of data mining was fully supported by the algorithm of Knowledgeable Discovery in Database (KDD).It plays a crucial part in big data technique. Because the history of KDD was discovered in late 1980's.so the capable of doing archive from extracting data's was really remarkable. The structure of data mining with big data technique has show in the Figure (1).Warehousing of data is also the sub part of Data Mining (DM) technique. It is defined as the process of building the integrated data's from different available sources. Data mining algorithm with some of the applications has done a great impact on IT field. The following applications are

- marketing
- Fraud detection
- logics
- Digital marketing

Where A-Website; B-storing big data's; C-Data mining; D-gathering new information;

E-Admin; F-new data;

4. Implementation steps

4.1. marketing

This application provides details of storing a kind of cookies based history checking from any websites. All the data's are handled by one admin server. He is the one who runs the process of data mining in that website. Making a algorithm with big data's from Knowledgeable Discovery in database (KDD) are the only reason of made a impact websites on the internet.

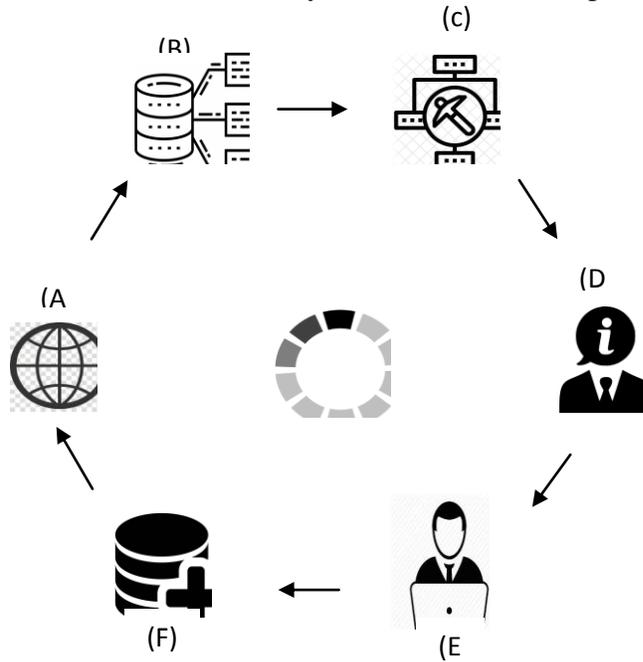


Figure 1 Structure of data mining with big data technique

4.2. Fraud detection

Bilinear Data mining technique is also runs the field of fraud detection applications. It is used to evaluate the process of telecommunication systems. In this system, sometimes telephone calls are comes from the unknown person. At that times, this technique is used to the reach the destination of that call while coming from which area of the city. It also checks the daily status of huge volume of phone calls while coming from other part of the city also.

4.3. logics

The approach of using big data technique with data mining algorithm in business logic is favorable for the administrator to launch his website efficiently. The important uses of data

mining are gathering new information from large sets of database. Once information was gathered, it is easy to raise your profit in that business.

4.4. Digital marketing

The Scopus of online marketing is mostly dependent on data mining techniques. For instance, In online product websites the process of data mining is receiving the product sales data from customer, administrator job is analyze which is the most buying product from that site, then show the product is the best product for our trade, that product visible in all the users of that website. The way of processing is known as online digital marketing.

5. Result and discussion

The contrast of data mining in big data algorithm is arguably best one in the world. The filter of this process has divided, in too many sources of IT industry. One of the parts is mining of constant pattern; it is defined as the process of pattern that occurs constantly in transactional data's. Some of the types are constant item set-it is emerged as a pair of related items, for example: mobile and charger. Constant subsequence-it showed as a consecutive process of buying related product, for example: mobile and head phones. Now the discussion of data mining algorithm is based on privacy problems. In the past decade the data mining algorithm is not popular among the globe. During the time, lesser amount of people was using this algorithm to their websites, nowadays the development of technologies was the reason behind the large number uses of data mining concept in IT industry.(Fig 2) display the growth of KDD, Data science, Big data, Data mining in periods of time.

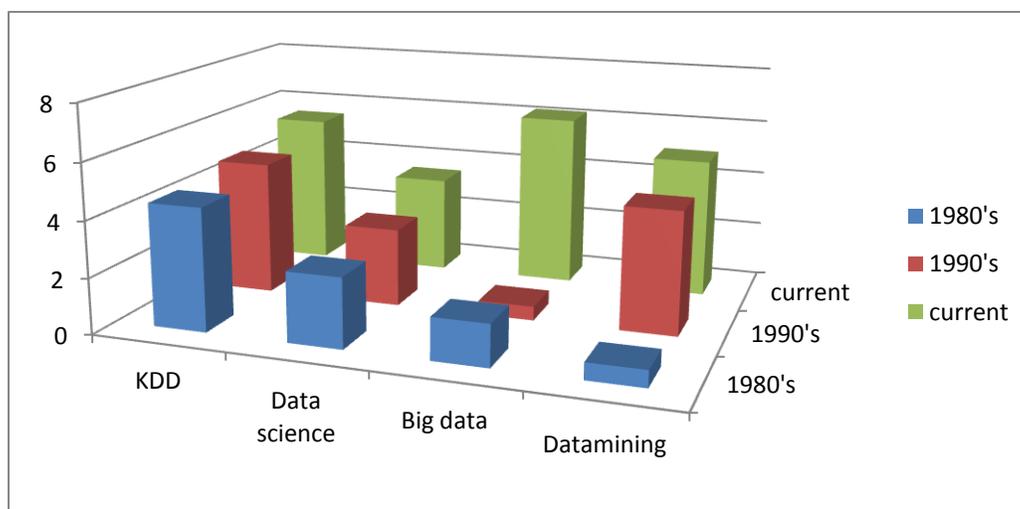


Figure 2 Improvement Growth of Techniques

In the above chart shows the growth of Knowledge Discover in Databases (KDD), data science, big data analytics, data mining in 1980's, 1990's and the current period.

So the problem was come in to this technique, because of the reach of data mining in large number of users. Large amount of data's has been stored in several databases of their websites. In online process and current period of technologies, hackers are easily extracting their data's from that websites. Cloud computing services likes the process is only way to avoid the security problems in data mining technique. On the other side problem is storing big data's. Because sometimes, while the large number of users had run their sites on same time, bigger issue is to handle the hanging problems in websites. At that time (HDFS) Hadoop Distributed File System is used to overcome this issue to solve the storage problems.

S.No	Storage	Data processing	Advantage	Limitation
1.	HDFS	Map Reduce/yarn	Scale out architecture, Fault tolerance, optimized scheduling, high availability.	Problem with Resource utilization. Not suitable for real time analytics
2	HDFS	Spark	High computation, speed, In-memory features, data locality.	Best suitable for iterative processing
3	Beowulf	MPI/open Mp	High Speed parallel	No Fault tolerance.

Table 1: Technologies and their integration to support Big data analytics

6. Conclusion

Big data algorithm is the capable of storing huge volume of data's in Knowledgeable Discovery in database (KDD).With the preference of mining data's concept is so effective for business logic and digital marketing. Gathering new information from user data's is the biggest advantage of data mining algorithm to increase the sales any company. Also it sends the notification for related kind of data's to the user for more communication with that website. At

user's point of view it is used to reduce the process of searching a related data's. It is worked on all over the world. It is storing huge volumes of data's day by day. One of the biggest applications used by big data technique was facebook. Huge margin of data's can be stored and retrieved in this application frequently.

References

- [1] Wu, X., Zhu, X., Wu, G.Q. and Ding, W., "Data mining with big data. IEEE transactions on knowledge and data engineering," in IEEE transactions on knowledge data engineering. 2013, pp. 97–107.
- [2] Xu, L., Jiang, C., Wang, J., Yuan, J. and Ren, Y., "Information security in big data: privacy and data mining. Ieee Access," in Ieee Access. 2014, vol, pp. 1149–1176.
- [3] Cheung, D.W., Lee, S.D. and Xiao, Y., "Effect of data skewness and workload balance in parallel data mining", IEEE IEEE Transactions on Knowledge and Data Engineering, pp. 498-514, 2002.
- [4] Saygin, Y. and Ulusoy, Ö., "Exploiting data mining techniques for broadcasting data in mobile computing environments," in IEEE Transactions on Knowledge and Data Engineering, 2002, pp.1387-1399.
- [5] Feng, Z. and Zhu, Y.,, "A survey on trajectory data mining: Techniques and applications," in IEEE Access 2016, pp. 2056–2067.
- [6] Cao, L., "Social security and social welfare data mining: An overview IEEE Transactions on Systems, Man, and Cybernetics, 2012,pp.837-853
- [7] Ming, J., Zhang, L., Sun, J. and Zhang, Y., "Analysis models of technical and economic data of mining enterprises based on big data analysis", IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), 2018, pp 224-227.
- [8] Kalegele, K., Sasai, K., Takahashi, H., Kitagata, G. and Kinoshita, T., "Four decades of data mining in network and systems management," in . IEEE Transactions on Knowledge and Data Engineering, 27(10), 2015, pp. 2700-2716.
- [9] Hajian, S. and Domingo-Ferrer, J., "A methodology for direct and indirect discrimination prevention in data mining.," in IEEE transactions on knowledge and data engineering, 25(7), 2012, pp. 1445-1459.



- [10] Zhang, N. and Zhao, W., “Privacy-preserving data mining systems” in *Computer*, 40(4), 2007, pp. 52-58.
- [11] Musolesi, M., “Big mobile data mining: Good or evil?.” in *IEEE Internet Computing*, 18(1), 2007, pp.78-81.
- [12] Li, Y., Chen, M., Li, Q. and Zhang, W., “Enabling multilevel trust in privacy preserving data mining” in *IEEE Transactions on Knowledge and Data Engineering*, 24(9),2011, pp.1598-1612..
- [13] Feng, M., Zheng, J., Ren, J., Hussain, A., Li, X., Xi, Y. and Liu, Q., “Big data analytics and mining for effective visualization and trends forecasting of crime data,” in *IEEE Access*, 2013, pp.106111-106123.
- [14] Akay, A., Dragomir, A. and Erlandsson, B.E., “A novel data-mining approach leveraging social media to monitor consumer opinion of sitagliptin.” in *IEEE journal of biomedical and health informatics*, 19(1),2012, pp.389-396.
- [15] Thuraisingham, B., “A primer for understanding and applying data mining..” in *It Professional*, 2(1), 2000, pp.28-31.