

Big data Analytics: Challenges and Applications for Social Media Data

Krishnaveni C¹, Udhayakumar U²

¹Department of Computer Science, Shanmuga Industries Arts and Science College, Tiruvannamalai

²Assistant Professor, Department of Computer Science, Shanmuga Industries Arts and Science College, Tiruvannamalai

Abstract

All types of machine automated systems are generating large amount of data in different forms like statistical, text, audio, video, sensor, and bio-metric data that emerges the term Big Data. In this paper we are discussing issues, challenges, and application of these types of Big Data with the consideration of big data dimensions. Here we are discussing social media data analytics, content based analytics, text data analytics, audio, and video data analytics their issues and expected application areas. It will motivate researchers to address these issues of storage, management, and retrieval of data known as Big Data. As well as the usages of Big Data analytics in India is also highlighted.

Keywords — Big Data, Big Data Analytics, Social Media Analytics, Content Based Analytics, Text Analytics, Audio Analytics, Video Analytics.

I. INTRODUCTION

The term big data is used to describe the growth and the availability of huge amount of structured and unstructured data. Big data which are beyond the ability of commonly used software tools to create, manage, and process data within a suitable time. Big data is important because the more data we collect the more accurate result we get and able to optimize business processes [1, 2]. The Big data is very important for business and society purpose. The data came from everywhere like sensors that used to gather climate information, available post or share data on the social media sites, video movie audio etc. This collection of data is called —BIG DATA [1].

In most enterprise scenarios the data is too big or it moves too fast or it exceeds current processing capacity. Big data has the potential to help companies improve

operations and make faster, more intelligent decisions. Big data usually includes data sets with sizes beyond the ability of commonly used software tools to capture, curate, manage, and process data within a tolerable elapsed time. Big data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale [3]. Wal-Mart handles more than 1 million customer transaction every hour. Facebook handles 40 billion photos from its user base. Big data require some technology to efficiently process large quantities.

The 5V's of big data are Volume, Velocity, Variety, Variability, and Value [6].

Volume: Many factors contribute to the increase in data volume. Transaction-based data stored through the years. Unstructured data streaming in from social media. Increasing amounts of sensor and machine-to-machine data being collected. In the past, excessive data volume

was a storage issue. But with decreasing storage costs, other issues emerge, including how to determine relevance within large data volumes and how to use analytics to create value from relevant data.

Velocity: Data is streaming in at unprecedented speed and must be dealt with in a timely manner. RFID tags, sensors and smart metering are driving the need to deal with torrents of data in near-real time. Reacting quickly enough to deal with data generation speed is a challenge for most organizations.

Variety: Data today comes in all types of formats. Structured, numeric data in traditional databases. Information created from line-of-business applications. Unstructured text documents, email, video, audio, stock ticker data and financial transactions. Managing, merging and governing different varieties of data is something many organizations still grapple with.

Variability: In addition to the increasing velocities and varieties of data, data flows can be highly inconsistent with periodic peaks. Daily, seasonal and event-triggered peak data loads can be challenging to manage. Even more so with unstructured data involved.

Value: It includes how we can use this big data for enhancing the business and living style. We know that different types of business or social application generate different types of data. Still identifying values from Big Data in their application areas is a big issue.

II. BIG DATA ANALYTICS

Big Data Analytics refers to the process of collecting, organizing, analysing large data set to discover different patterns and other useful information [4, 5]. Big data analytics is a set of technologies and techniques that require new forms of integration to disclose large hidden values from large datasets that are different from the usual ones, more complex, and of a

large enormous scale. It mainly focuses on solving new problems or old problems in better and effective ways [6].

The main goal of the big data analytic is to help organization to make better business decision, future prediction, analysis large numbers of transactions that done in organization and update the form of data that organization is used [7]. Example of big data Analytics are big online business website like Flipkart, Snapdeal uses Facebook or Gmail data to view the customer information or behaviour. Analysing big data allows analysts, researchers, and business users to make better and faster decisions using data that was previously inaccessible or unusable [8, 9].

Types of big data analytics are: Prescriptive: - This type of analytics help to decide what actions should be taken [10]. It very valuable but not used largely. It focuses on answer specific question like, hospital management, diagnosis of cancer patients, diabetes patients that determine where to focus treatment. Predictive: - This type of analytics help to predict future or what might be happen. For example some companies use predictive analytics to take decision for sales, marketing, production, etc. Diagnostic: - In this type look at past and analyzethe situation what happen in past and why it happen. And how we can overcome this situation. For example weather preadiction, customer behavioral analysis etc. Descriptive: - It describes what is happening currently and prediction near future. For example market analysis, compatainsbehavioral analysis etc [11, 12].

III. SOCIAL MEDIA ANALYTICS

The Social Media analytics is collecting information or data form the social media websites, blogs etc. and uses it in business purpose or decision making. Now a Days Social Media is the best platform for understand the real-time customer choice or intentions and sentiments, using social media business advertising, product marketing

easily [13]. EBay.com uses two data warehouses at 7.5 petabytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising. Inside eBay's 90PB data warehouse. Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers.

A. Application areas

- a) Behaviour Analytics
- b) Location-based interaction Analytics
- c) Recommender systems development
- d) Link prediction
- e) Customer interaction and Analytics & marketing
- f) Media use
- g) Security
- h) Social studies

B. Challenges of Social Media Analytics

- a) Massive amounts of data require lots of storage space and processing power.
- b) Shifting social media platforms.
- c) Worldwide online accessibility provides more data in many languages.
- d) Evolution of online language.

IV. CONTENT BASE ANALYTICS

Content Base Analytics means whatever data that store in social media back-end site. For example Facebook users store their data, photos, and videos on Facebook storage. For this content they need big amount of storage but now a days number of users increasing rapidly so, social networking sites like Facebook, twitter, WhatsApp need to increase their storage capacity day by day and that's the obstacle because they don't know how much of storage capacity they need to increase [14].

Content-based predictive analytics recommender systems mostly match features (tagged keywords) among similar items and the user's profile to make recommendations. When a user purchases an item that has tagged features, items with features that match those of the original item

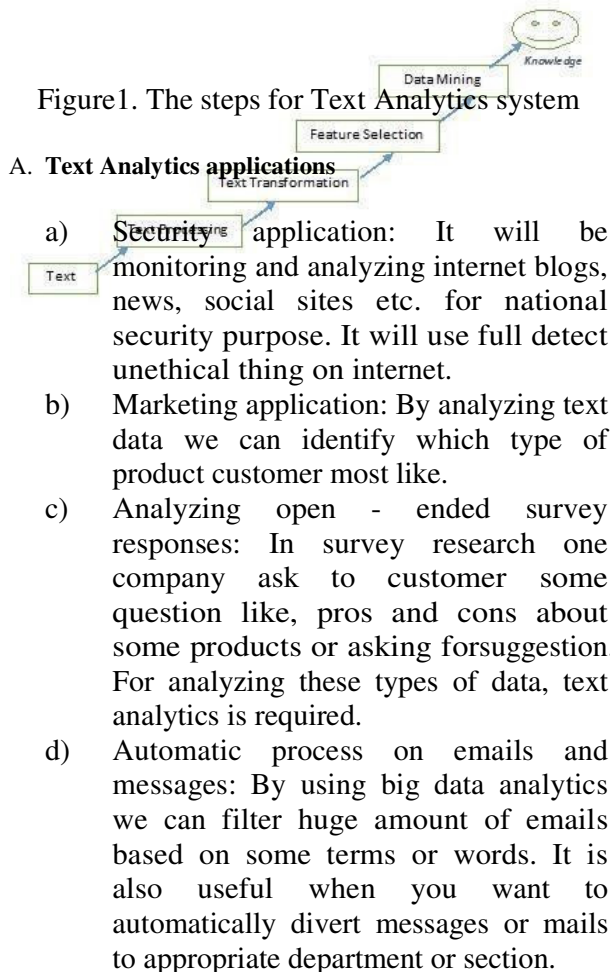
will be recommended. The more features match, the higher the probability the user will like the recommendation. This degree of probability is called precision [15]. User-based tagging, however, turns up other problems for a content-based filtering system (and collaborative filtering) like:

- a) Credibility: Not all customers tell the truth (especially online), and users who have only a small rating history can skew the data. In addition, some vendors may give (or encourage others to give) positive ratings to their own products while giving negative ratings to their competitors' products.
- b) Scarcity: Not all items will be rated or will have enough ratings to produce useful data.
- c) Inconsistency: Not all users use the same keywords to tag an item, even though the meaning may be the same. Additionally, some attributes can be subjective. For example, one viewer of a movie may consider it short while another says it's too long.

V. TEXT ANALYTICS

Most of all information or data is available in textual form in databases. From these contexts, manual Analytics or effective extraction of important information are not possible. For that it is relevant to provide some automatic tools for analysing large textual data [16, 19]. Text analytics or text mining refers process of deriving important information from text data. It will use to extract meaningful data from the text. It use many ways like associations among entities, predictive rules, patterns, concepts, events etc. based on rules [20]. Text analytics widely use in government, research, and business needs. Data simply tells you what people did but text analytics tell you why. From unstructured or semi structured text data all information will retrieve. From all textual data it will extract

important information [17, 18]. After extracting information it will be categorized. And from these categorized information we can take decision for business.



B. Applying Text Analytics to Social Media

- a) **Event Detection**: It aims to monitor a data source and detect the occurrence of an event that is to be captured within that source. These data sources include

images, videos, audios, text documents.

- b) **Collaborative Question Answering**: As social networking websites has emerged, the collaborative question answering services have also emerged. It includes several expert people to answer the questions posted by the people. A large number of questions and answers are posted on the social networking websites
- c) **Social Tagging**: Tagging of the data has also increased to a great extent.

VI. AUDIO ANALYTICS

Audio analytics is the process of compressing data and packaging the data in to single format called audio. Audio Analytics refers to the extraction of meaning and information from audio signals for Analysis. There are two way to represent the audio Analytics is 1) Sound Representation 2) Raw Sound Files. Audio file format is a format for store digital audio data on a system. There are three main audio format: Uncompressed audio format, Lossless compressed audio format, Lossy compressed audio format.

Application Area of Audio Analytics:

The audio is the file format that used to transfer the data to one place to another. Audio analytics is used to check whether given audio data is available in proper format or in similar format that sender send. The Application of audio Analytics are many:

- a) **Surveillance application**: Surveillance application is based on approach for systematic choice of audio classes for detection of crimes done in society. A surveillance application is based on audio Analytics framework is the only way to detect suspicious kind of activity. The application is also used to send some important information to surveillance at some crisis situation urgently.

b) Detection of Threats: The audio mechanism is used to identify the thread that take place between sender and receiver.

c) Tele-monitoring System: New technology have camera with the facilities to record the audio also. Audio Analytics may provide effective detection of screams, breaking glass, gun sound, explosions, calling for help sound etc. Combination of audio Analytics and video Analytics in single monitoring system result as a good threat detection efficiency.

d) Mobile Networking System: The Mobile networking system is used to talk or transfer information to one place to another place. Sometimes due to some network problem the audio sound is not work properly at that time Audio Analytics is used to find the information that not send properly due to some problems.

VII. VIDEO ANALYTICS

Video is a major issue when considering big data. Videos and images contribute to 80 % of unstructured data. Now days, CCTV cameras are the one form of digital information and surveillance. All these information is stored and processed for further use, but video contains lots of information and is generally large in size. For example YouTube has innumerable videos being uploaded every minute containing a massive information. Not all video are important and viewed largely. This creates a situation where videos create a junk and hard-core contribution to big data problems. Apart from videos, surveillance cameras generate a lot of information in seconds. Even a small Digital camera capturing an image stores millions of pixel information in mille seconds.

VIDEO Data Analytics dimensions -
Volume: Size of video being more, takes the network as well as the server, time for processing. Low bandwidth connections create traffic on network as these videos deliver slowly. When stored on mass storage on

secondary storage requires huge amount of space and takes more time retrieving as well as processing. Variety: Videos consisting of various format and variety such as HD videos, Blu-ray copies etc. Velocity: It is speed of data. Now a days, Digital cameras process and capture videos at a very high quality and high speed. Video editing makes it to grow in size as it contains other extra information about the videos. Videos grow in size faster as they are simply nothing but collection of images.

Application of video analytics:

a) Useful in accident cases: With the use of CCTV cameras we can identify what happened at the time of accident it's also used for security reason and parking vehicles etc.

b) Useful in schools, traffic police, business, security etc.

c) Video Analytics for investigation (Video Search): Video analytics algorithms isimplemented to analyze video, a task that is challenging and its very time consuming for human operator especially when there is large amount of data are available using video analytics we can search particular video when we required.

d) Video analytics for Business Intelligence: It uses to extracts statistical and operational data. Rather than having operator that review all the video and tally all the people or cars moving in certain area, or checking which traffic routes are most commonly taken, video analytics can do it automatically.

e) Target and Scene Analytics: Video Analytics for business Intelligence involves target and scene Analytics. Target Analytics provides details information about the target movement, patterns, appearance and other characteristics which can be used for identification of target.

f) Direction Analytics: Direction Analytics is the ability to distinguish behaviour by assigning specific values (low to high) to areas within a camera's field of view.

g) Remove the human equation through the

automation: It removes the tedium involved in giving one or more set of eyes on a monitor for an extended period of time. The automation of video analytics allows the insertion of human judgment at the most critical time in the surveillance process.

VIII. CONCLUSIONS

Now, computer industry accepts Big Data as a new challenge for all types of machine automated systems. There are many issues in storage, management, and retrieval of data known as Big Data. The main problem is how we can use this data for increasing business and improvement in living standard of people. In this paper we are discussing the issues, challenges, application as well as proposing some actionable insight for Big Data. It will motivate researchers for finding knowledge from the big amount of data available in different forms in different areas.

ACKNOWLEDGMENT

I would like to convey my heartfelt thanks to Mr.Udhayakumar, my research guidance for completion of my project. He helped me to understand and remember important details of the project. He helped me and gave his advice on completing my project successfully.

REFERENCES

1. Akoka, J., Comyn-Wattiau, I., &Laoufi, N. (2017). *Research on Big Data—A systematic mapping study*. *Computer Standards & Interfaces*, 54, 105-115.
2. Bechini, A., Marcelloni, F., &Segatori, A. (2016). *A MapReduce solution for associative classification of big data*. *Information Sciences*, 332, 33-55.
3. Berger, M.L., &Doban, V. (2014). *Big data, advanced analytics and the future of comparative effectiveness research*. *Journal of comparative effectiveness research*, 3 2, 167-76.
4. *Big Data for dummies* – Judith Hurwitz et al., Wiley, 2013.

5. *Big Data, Big Analytics* – Michael Minelliet al., Wiley, 2013.
6. Bradlow, E. T., Gangwar, M., Kopalle, P., &Voleti, S. (2017). *The role of big data and predictive analytics in retailing*. *Journal of Retailing*, 93(1), 79-95.
7. Chen, M., Mao, S., & Liu, Y. (2014). *Big data: A survey*. *Mobile Networks and Applications*, 19(2), 171-209.
8. Choi, T. M., Chan, H. K., &Yue, X. (2017). *Recent development in big data analytics for business operations and risk management*. *IEEE transactions on cybernetics*, 47(1), 81-92.
9. Ducange, P., Pecori, R., &Mezzina, P. (2017). *A glimpse on big data analytics in the framework of marketing strategies*. *Soft Computing*, 1-18.
10. Elgendy, N. and Elragal, A., 2014, July. *Big data analytics: a literature review paper*. In *Industrial Conference on Data Mining* (pp. 214-227). Springer, Cham.
11. D. Ranjith, J. M. Balajee and C. Kumar, "Trust computation methods in mobile ADHOC network using glomosim: A Review" *International Journal of Scientific Research and Modern Education*, Vol. 1, Issue I, pp. 777-780, Nov.2016.
12. Janarthanan Y, Balajee J.M, and SrinivasaRaghava S. "Content based video retrieval and analysis using image processing: A review." *International Journal of Pharmacy and Technology* 8, no.4 (2016): 5042-5048.
13. Jeyakumar, Balajee, MA SaleemDurai, and Daphne Lopez. "Case Studies in Amalgamation of Deep Learning and Big Data." In *HCI Challenges and Privacy Preservation in Big Data Security*, pp. 159-174. IGI Global, 2018.
14. Kamalakannan, S. "G., Balajee, J., SrinivasaRaghavan., "Superior content-based video retrieval system according to query image"." *International Journal of Applied Engineering Research* 10, no. 3 (2015): 7951-7957.
15. Priya, V., Subha, S., &Balamurugan, B. (2017). *Analysis of performance measures to handle medical E-commerce shopping cart abandonment in cloud*. *Informatics in Medicine Unlocked*.
16. Rangith. D Lakshmi narayanan. J and Balajee. J, " A study of behavior on information system in a university

campus by analysis of people mobility” *International journal of research in computer application & management*, Vol. 6, Issue 7, pp. 29-31, Jul.2016.

17. Ranjith, D., J. Balajee, and C. Kumar. "In premises of cloud computing and models." *International Journal of Pharmacy and Technology* 8, no. 3 (2016): 4685-4695.
18. Sethumadahavi R Balajee J “Big Data Deep Learning in Healthcare for Electronic Health Records,” *International Scientific Research Organization Journal*, vol. 2, Issue 2, pp. 31–35, Jul. 2017.
19. Ushapreethi P, Balajee Jeyakumar and BalaKrishnan P, Action Recongnition in Video Surveillance Using Hipi and Map Reducing Model, *International Journal of Mechanical Engineering and Technology* 8(11), 2017,pp. 368–375.
20. Zhang, W., Zhao, Q., Deng, J., Hu, Y., Wang, Y., &Ouyang, D. (2017). Big data analysis of global advances in pharmaceuticals and drug delivery 1980–2014. *Drug Discovery Today*.