ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

A Detail Study on Big Data Analytics

Mrs. R.Anusuya[1], Ms. R Vinothini [2]

Assistant professor, Dept. Computer Applications, Pioneer College of Arts and Science, Jothipuram, Coimbatore[1]

Student, Master of Computer Science, Pioneer College of Arts and Science, Jothipuram, Coimbatore[2]

Abstract: Big data is a term for data sets that are so large or complex that traditional data processing applications are inadequate to deal with them. Challenges include analysis, capture, sharing, storage, transfer, visualization, querying, updating and information privacy. The term "big data" often refers simply to the use of predictive analytics, user behaviour analytics, or certain other advanced data analytics methods that extract value from data, user behaviour analytics, or certain other advanced data analytics methods that extract value from data, and seldom to a particular size of data set. The paper focus on the analytic methods used for big data. Big data analytics is the process of collecting, organizing and analysing large sets of data to discover patterns and others useful information. Big data analytics can help organizations to better understand the information contained within the data and will also help identifying the data that is most important to the business and future business decisions. This paper highlights the need to develop appropriate methods to unstructured text, audio, and video formats.

KEYWORDS: Big data, Big data analytics, Big data definitions, Unstructured data analytics, Predictive analytics.

I. INTRODUCTION

This paper documents the basic concepts relating to big data. It attempts to consolidate the hitherto fragmented discussion on what constitutes big data, what metrics define the size and other characteristics of big data, and what tools and technologies exist to harness the potential of big data. From corporate leaders to municipal planners and academics, big data are the subject of attention. Big Data can also play a role for small or medium companies and organizations that recognize the possibilities to capitalize upon the gains. In the past, new technological developments first appeared in technical and academic publications. The fast evolution of big data technologies and the ready acceptance of the concept by public and private sectors left little time for the discourse to develop and mature in the academic domain. For instance, there is little con-sensus around the fundamental question of how big the data has to be to qualify as 'big data'. Which is dominated and influenced by the marketing efforts of large software and hardware developers, focuses on predictive analytics and structured data. It ignore the largest component of big data, which is unstructured and is available as audio, images, video, and unstructured text. It is estimated that the analytics ready structured data forms only a small subset of big data. The unstructured data, especially data in video format, is the largest component of big data that is only partially archived. This paper expand the discussion on various types of big data, namely text, audio, video, and social media. We apply the analytics lens to the discussion on big data.

II. DEFINING BIG DATA

Big data enables organizations to store, manage, and manipulate vast amounts of disparate data at the right speed and at the right time.

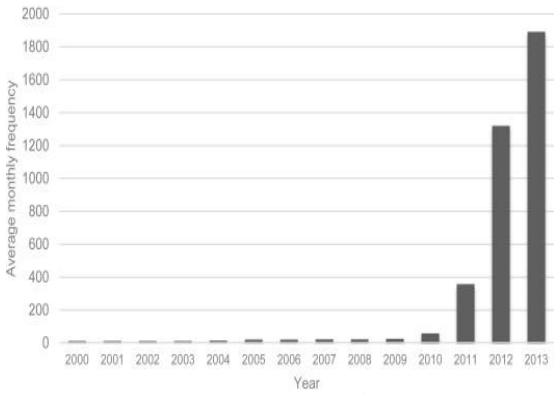

Fig .1 : Frequency distribution of documents Containing the term "big data" in ProQuest Research Library

[1]Historically, a number of the large-scale Internet search, advertising, and social networking companies pioneered Big Data hardware and software innovations. As Google, Yahoo, Oracle, and others have contributed their technology to the open source community, broader commercial and public sector interest took up the challenge of making Big Data work for them. [2]Rather than the data interpreted independently, they see the value realized by adding the new data to their existing operational or analytical systems. So, Big Data describes a holistic information management strategy that includes and integrates many new types of data and data management alongside traditional data.

Big Data has also been defined by the three types: [6][7]Volume, Velocity, and Variety.

VOLUME

The amount of data. While volume indicates more data, it is the granular nature of the data that is unique.

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

VELOCITY

In this context, the speed at which the data is generated and processed to meet the demands and challenges that lie in the path of growth and development. It available in real-time.

VARIETY

The type and nature of the data. This helps people who analyses it to effectively use the resulting insight. In structured and semi-structured data types, such as text, audio, and video require additional processing to both derive meaning and the supporting metadata.

VALUE

Data has inherent value but it must be discovered. There are a range of quantitative and investigative techniques to derive value from data discovering a consumer preference or sentiment, to making a relevant offer by location, or for identifying a piece of equipment that is about to fail.

III. BIG DATA ANALYTICS

Big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer needs and business information. [9]The analytical findings can lead to more effective marketing, better customer service, improved operational efficiency, competitive advantages over the organizations and other business benefits.

Figure-2: Big data requires High-Performance Analytics

3.1 TEXT ANALYTICS

All information or data is available in textual form in databases. [3]The text analytics solved business problems is called text analytics. 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. Text analytics 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, concepts, events etc. based on rules. From all textual data it will extract important information.

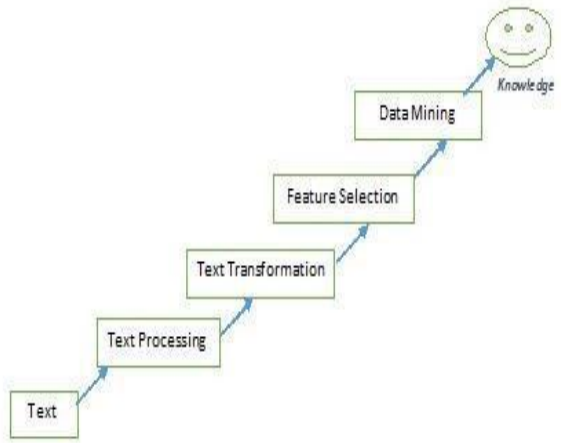

Figure-3: Steps for Text Analytics system

3.2 AUDIO ANALYTICS

Audio analytics is the process of compressing data and packaging the data in to single format called audio analytics.[4] Audio analytics refers to the extraction of meaning and information from audio signals for analysis. Here two ways to represent the audio analytics 1) sound Representation 2) Raw sound files. Audio files format is a format for store digital audio data on a system. There are three main audio formats: Uncompressed audio format, Lossless compressed audio format, Lossy compressed audio format. So that Audio analytic has developed intelligent solutions that involve a unique blend of state of the are knowledge about audio modelling and machine learning, audio capture devices, highly efficient embedded software, and an multitude of other practical aspects sound recognition.

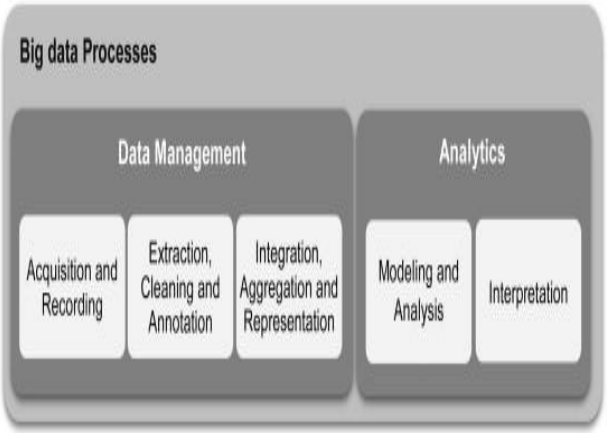

Fig. 4. Processes for extracting insights from big data.

3.3 VIDEO ANALYTICS

Video data analytics, potential terrorists can be identified, tracked, and correlated with other data sources in real time and public safety as well as privacy, can be enabled as well.[5]Digital devices that generate millions of pixels in a flash are in the pockets of billions of people worldwide. CCTV cameras are the one form of digital information and

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

surveillance. All these information is stored and processed for further use, but video contains lots of information. For example YouTube has innumerable videos being uploaded every minute containing a massive information. 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.

When stored on mass storage on secondary storage requires huge amount of space and takes more time retrieving as well as processing. [5]Variety: Videos consisting of various format and variety such as HD videos, Blu-ray copies etc. Velocity: It is speed of data. Now a day, 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.

3.4 SOCIAL MEDIA ANALYTICS

Social media analytics is the practice of gathering data from blogs and social media websites and analysing that data to make business decisions. And uses it in business purpose or decision making. Social media not only provides marketers with a means of communicating with their customers, but also a way to better understand their customers. 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.

There are a number of types of software tools for analysing unstructured data found in tweets and Facebook posts. In addition to text analysis, many enterprise level social media tools will harvest and store the data.

Based on this categorization, [8]the social media analytics can be classified into two groups:

Content-based analytics:

It is focus on the data posted by users on social media platforms, such as customer feedback, product reviews, images, and videos. To keep tags manageable and accurate, an acceptable set of tags may be provided by the websites. Text, audio, and video analytics, as discussed earlier, can be applied to derive insight from such data. Also, big data technologies can be adopted to address the data processing challenges.

Structure-based analytics Referred to as social network analytics, this type of analytics are concerned with synthesize the structural attributes of a social network and extracting intelligence from the relationships among the participating entities. The structure of a social network is modelled through a set of nodes and edges, representing participants and relationships, respectively. The model can be visualized as a graph composed of the nodes and the edges. There two types of network graphs, namely social graphs and activity graphs. In social graphs, an edge between a pair of nodes only signifies the existence of a link between the corresponding entities.

3.5. PREDICTIVE ANALYTICS

Predictive analytics can be applied to almost all disciplines from predicting the failure of jet engines based on the stream of data from several thousand sensors, to predicting customers' next moves based on what they buy, when they buy, and even what they say on social media. At its core, predictive analytics seek to uncover patterns and capture relationships in data. Predictive analytics techniques are subdivided into two groups. [8]Some techniques, such as moving averages, attempt to discover the historical patterns in the outcome variable and extrapolate them to the future. Others, such as linear regression, aim to capture the interdependencies between outcome variable and explanatory variables, and exploit them to make predictions. Based on the underlying methodology, techniques can also be categorized into two groups: regression techniques (e.g., multi monologist models) and machine learning techniques (e.g., neural networks). Another classification is based on the type of outcome variables: techniques such as linear regression address continuous outcome variables (e.g., sales price of houses), while others such as Random Forests are applied to discrete outcome variables (e.g., credit position).Predictive analytics techniques are primarily based on statistical methods. Several factors call for developing new statistical methods for big data. First, conventional statistical methods are rooted in statistical significance: a small sample is obtained from the population and the result is compared with chance to examine the significance of a particular relationship. The third factor corresponds to the distinctive features inherent in big data: We describe these below.

- Heterogeneity
- Noise
- Spurious correlation
- Incidental endogenetic

IV. CONCLUSION

The paper first defined what is meant by big data to consolidate the divergent discourse on big data. We presented various definitions of big data, highlighting the fact that size is only one dimension of big data. Other dimensions, such as velocity and variety are equally important. The paper focus on analytics to gain valid and valuable insights from big data. The conclusion is then generalized to the entire population. In contrast, big data sample are massive and represent the majority of, if not the entire, population. The paper highlight the point that predictive analytics, which deals mostly with structured data, overshadows other forms of analytics applied to unstructured data, which constitutes 95% of big data. We reviewed analytics techniques for text, audio, video, and social media data, as well as predictive analytics. The paper makes the case for new statistical techniques for big

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

data to address the peculiarities that differ4.entiate big data from smaller data sets. Most statistical methods in practice have been devised for smaller data sets comprising samples.

Technological advances in storage and computations have enabled cost-effective capture of the informational value of big data in a timely manner. Consequently, one observes a proliferation in real-world adoption of analytics that were not economically feasible for large-scale applications prior to the big data era. The processing of unstructured text fuelled by the massive influx of social media data is generating business value by adopting conventional (pre-big data) sentiment analysis techniques, which may not be ideally suited to leverage big data. Although major innovations in analytical techniques for big data have not yet taken place, one anticipates the emergence of such novel analytics in the near future.

REFERENCES

[1] Diebold, F. X. (2012). A personal perspective on the origin(s) and develop-ment of "big data.Retrieved fromhttp://papers.ssrn.com/sol3/papers.cfm?abstract id=2202843Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National ScienceReview, 1(2), 293–314.

[2] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al.(2011). Big data: The next frontier for innovation, competition, and produc-tivity. McKinsey Global Institute. Retrieved from http://www.citeulike.org/group

[3] Parthasarathy, S., Ruan, Y., & Satuluri, V. (2011). Community discovery in socialnetworks: Applications, methods and emerging trends. In C. C. Aggarwal (Ed.),Social network data analytics (pp. 79–113). United States: Springer.

[4] Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Retrievedfrom http://www-03.ibm.com/systems/hu/resources/the real word use ofbig data.pdf

[5] Small and midsize companies look to make big gains with "bigdata," according to recent poll conducted on behalf of SAP. In C. C. Aggarwal (Ed.), Social network data analytics (pp. 177–214).United States: Springer.

[6] Web content available on: http://searchbusinessanalytics.techtarget.com/definition/big-dataanalytics on the dated: 16-08-2015

[7] Web content available on the link: https://www.statsoft.com/textbook/text-mining on the dated: 16-08-2015

[8] Jump up to:[a] [b] Khan G. F., 2015.

[9] Jump up^ Tera, Data. "Capitalize On Social Media With Big Data Analytics". www.forbes.com. Retrieved 27 May 2015.