



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijiemr.org

COPY RIGHT



ELSEVIER
SSRN

2018IJIEMR. Personal use of this material is permitted. Permission from IJIEMR must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. No Reprint should be done to this paper, all copy right is authenticated to Paper Authors

IJIEMR Transactions, online available on 15th Feb 2018. Link :

<http://www.ijiemr.org/main/index.php?vol=Volume-07&issue=ISSUE-02>

Title: **SENTIMENT ANALYSIS OF BIG DATA METHODS APPLICATIONS AND OPEN CHALLENGES**

Volume 07, Issue 02, Pages: 760–769.

Paper Authors

VANAM SRAVAN DR. RASHMI AGAEWAL



USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per **UGC Guidelines** We Are Providing A Electronic Bar Code



SENTIMENT ANALYSIS OF BIG DATA METHODS APPLICATIONS AND OPEN CHALLENGES

¹VANAM SRAVAN ²DR. RASHMI AGAEWAL

¹Research Scholar, Madhav University, Rajasthan. India

²Accosiate Professor, Madhav University, Rajasthan. India

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.

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. 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. Now a days this big data is

used in multiple ways to grow business and to know the world [1,2, 15]. 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. 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 of data. It use some technology like, data fusion and integration, genetic algorithms, machine learning, and signal processing, simulation, natural language processing, time series Analytics and visualization [12,13,16]

1.1. Characteristics of Big Data:-

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 [10, 12,13, 15,16].

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.

Complexity: Today's data comes from multiple sources. And it is still an undertaking to link, match, cleanse and transform data across systems. However, it is necessary to connect and correlate relationships, hierarchies and multiple data linkages or your data can quickly spiral out of control.

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 form Big Data in their application areas is a big issue.

2. BIG DATA ANALYTICS

Big Data Analytics refers to the process of collecting, organizing, analyzing large data sets to discover different patterns and other useful information. 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 [12,13, 15, 16].

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. 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. Analyzing big data allows analysts, researchers, and business users to make better and faster decisions using data that was previously inaccessible or unusable. Using advanced analytics techniques such as text analytics, machine learning, predictive analytics, data mining, statistics, and natural language processing, businesses can analyze previously untapped data sources independent or together with their existing enterprise data to gain new insights resulting in significantly better and faster decisions. It helps us to uncover hidden patterns, unknown correlations, market trends, customer preferences etc. It leads us to more effective marketing, revenue opportunities, better customer service etc. Big Data can be analyzed through predictive analytics, text analytics, statistical analytics and data mining[1,2,4]. Types of big data analytics are: Prescriptive: - This type of analytics help to decide what actions should be taken. 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 analyze the 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, compatains behavioral analysis etc.

By using appropriate analytics organization can increase sales, increase customer service, and can improve operations. Predictive Analytics allow organizations to make better and faster decisions [1, 2, 4, 10].

2.1. Predictive Analytics

Predictive Analytics is a method through which we can extract information from existing data sets to predict future outcomes and trends and also determine patterns. It does not tell us what will happen in future. It forecasts what might happen in future with acceptable level of reliability. It also includes what if-then-else scenarios and risk assessment. Applications areas of Predictive Analytics are [1, 2, 4, 10]:

CRM (Customer Relationship Management): Predictive analytics is useful in CRM in fields such as marketing campaigns, sales, customer services etc. The focus is to put their efforts effectively on analyzeing product in demand and predict customer's buying habits .

Clinical Decision Support: Predictive Analytics helps us to determine that which patients are at risk of developing certain conditions like diabetes, asthma, lifetime illness etc.

Collection Analytics: Predictive Analytics helps financial institutions for the allocation for collecting resources by identifying most effective collection agencies, contact strategies etc. to each customer.

Cross Sell: An Organization that offers multiple products, Predictive Analytics can



help to analyze customer's spendings, their behavior etc. This can help to lead cross sales that means selling additional products to current customers.

Customer Retention: As the number of competing services is increasing, businesses should continuously focus on maintaining customer satisfaction, rewarding loyal customers and minimize customer reduction. If Predictive Analytics is properly applied, it can lead to active retention strategy by frequently examining customer's usage, spending and behavior patterns.

Direct marketing: When marketing consumer products and services, there is the challenge of keeping up with competing products and consumer behavior. Apart from identifying prospects, predictive analytics can also help to identify the most effective combination of product versions, marketing material, communication channels and timing that should be used to target a given consumer.

Fraud detection: Fraud is a big problem for many businesses and can be of various types: inaccurate credit applications, fraudulent transactions (both offline and online), identity thefts and false insurance. These problems plague firms of all sizes in many industries. Some examples of likely victims are credit card issuers, insurance companies, retail merchants, manufacturers, business-to-business suppliers and even services providers. Predictive analysis can help to identify high-risk fraud candidates in business or the public sector.

Portfolio, product or economy-level prediction: These types of problems can be addressed by predictive analytics using time series techniques. They can also be

addressed via machine learning approaches which transform the original time series into a feature vector space, where the learning algorithm finds patterns that have predictive power.

Risk management: When employing risk management techniques, the results are always to predict and benefit from a future scenario. Predictive analysis helps organizations or business enterprises to identify future risk, Natural Disaster and its effect. Risk management helps them to take correct decision on correct time.

Underwriting: Many businesses have to account for risk exposure due to their different services and determine the cost needed to cover the risk. For example, auto insurance providers need to accurately determine the amount of premium to charge to cover each automobile and driver. For a health insurance provider, predictive analytics can analyze a few years of past medical claims data, as well as lab, pharmacy and other records where available, to predict how expensive an enrollee is likely to be in the future. Predictive analytics can help underwrite these quantities by predicting the chances of illness, default, bankruptcy, etc. Predictive analytics can streamline the process of customer acquisition by predicting the future risk behaviour of a customer using application level data.

2.2. Big Data Analytics usage in India

From predicting ticket confirmations of trains to checking for water supply leakages and even for finding the perfect bride and groom, Big Data is being used in a number of creative ways in India. Following are few

uses of Big Data Analytics in india in last few years [3,9].

- a) Win elections (exit poll).
- b) Finding a perfect match.
- c) Detecting water leakages.
- d) Gaining insights into shopping behavior.
- e) Ensuring proper water supply.
- f) Improve India's financial inclusion ratio.
- g) Improve product development.
- h) Predict ticket confirmations for trains.

3. 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. 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. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB. Facebook handles 50 billion photos from its user base. As of August 2012, Google was handling roughly 100 billion searches per month [8, 9, 14].

3.1. Application areas

- a) Behavior 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

3.2. 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.

4. 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.

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. [4,6,13] 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.

4. 1. Precision with constant feedback

One way to improve the precision of the system's recommendations is to ask customers for feedback whenever possible. Collecting customer feedback can be done in many different ways, through multiple channels. Some companies ask the customer to rate an item or service after

purchase. Other systems provide social-media-style links so customers can —like or —dislike a product.

4.2. Measurement for effectiveness of system recommendations

The success of a system's recommendations depends on how well it meets two criteria: precision (think of it as a set of perfect matches — usually a small set) and recall (think of it as a set of possible matches — usually a larger set). Issues in measurement for effectiveness:

a) **Precision** measures how accurate the system's recommendation was. Precision is

difficult to measure because it can be subjective and hard to quantify.

b) Some recommendations may connect with the customer's interests but the customer may still not buy. The highest confidence that a recommendation is precise comes from clear evidence: The customer buys the item. Alternatively, the system can explicitly ask the user to rate its recommendations.

c) **Recall** measures the set of possible good recommendations your system comes up with. Think of recall as an inventory of possible recommendations, but not all of them are perfect recommendations. There is generally an inverse relationship to precision and recall. That is, as recall goes up, precision goes down, and vice versa.

The ideal system would have both high precision and high recall. But realistically, the best outcome is to strike a delicate balance between the two. Emphasizing precision or recall really depends on the problem you're trying to solve [4,6,13].

5. 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 analyzing large textual data. 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. 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. After extracting information it will be categorized. And from these categorized information we can take decision for business [5, 6].

5.1. Steps for Text Analytics system (Figure -1):

- Text: In initial stage data is unstructured.
- Text processing: All information will transfer in Semantic Syntactic text.
- Text transformation: In it important text will extract for future use.
- Feature selection: In it data is counted and display in Statistics format.
- Data mining: All data is classified and clustered.

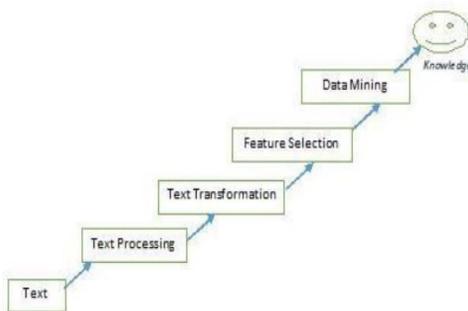


Figure 1. The steps for Text Analytics system

5.2. Text Analytics applications areas:

- Security application: It will we monitoring and analyzing internet blogs, news, social sites etc. for national security purpose. It will use full detect unethical thing on internet.

- Marketing application: By analyzing text data we can identify which type of product customer most like.

- Analyzing open – ended survey responses: In survey research one company ask to customer some question like, pros and cons about some products or asking for suggestion. For analyzing these types of data, text analytics is require.

- Automatic process on emails and messages: By using big data analytics we can filter huge amount of emails based on some terms or words. It is also useful when you want to automatically divert messages or mails to appropriate department or section.

5.3. Distinct Aspects of Text in Social Media:

- Time Sensitivity: An important feature of the social media services is their real-time nature. With the rapid growth of the content and communication styles, text is also changing. As the time sensitivity of the textual data the people’s thoughts also changes from time to time.

- Short Length: Successful processing of the short texts is essential for the text analytics method. As the messages are short, it makes people more efficient with their participation in social networking websites. Short messages are used in social media which consists of few phrases or sentences.

- Unstructured Phrases: An important difference between the text in social media and traditional media is the difference in the quality of content. Different people posts different things according to their knowledge, ideas, and thoughts. When composing a message also many new abbreviations and acronyms are used for e.g.

How r u? —Gr8! are actually not words but they are popular in social media.

5.4. 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 includes 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. For example when any particular user is looking or searching for a recent event like —Bihar Election! then the system will return the results that are tagged as —Bihar! or !Election!. Textual data in social media provides lots of information and also the user-generated content provides diverse and unique information in forms of comments, posts and tags. [5,6]

6. 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. [11]

6.1. 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 indentify 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.

7. VIDEO ANALYTICS



Video is a major issue when considering big data. Videos and images contribute to 80 % of unstructured data. Now a 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.[7]

7.1. 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 is implemented 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 behavior 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.

7. CONCLUSION AND FUTURE WORK Now, computer industry accept 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.

REFERENCES

- [1] Web content available on the link: —http://www.sas.com/en_us/insights/big-data/what-is-big-data.html on the dated: 16-08-2015
- [2] Web content available on the link: —<http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html> on the dated: 16-08-2015
- [3] Web content available on the link: —<http://www.dqindia.com/8-innovative-examples-of-big-data-usage-in-india/> on the dated: 16-08-2015
- [4] Web content available on the link: —<http://searchbusinessanalytics.techtarget.com/definition/big-dataanalytics> on the dated: 16-08-2015
- [5] Web content available on the link: —<https://www.statsoft.com/textbook/text-mining> on the dated: 16-08-2015
- [6] Web content available on the link: —<https://www.predictiveanalyticstoday.com/text-analytics> on the dated: 16-08-2015
- [7] Web content available on the link: —<https://gigaom.com/2014/01/24/why-video-is-the-next-big-thing-in-big-data/> on the dated: 16-08-2015
- [8] Web content available on the link: —<http://searchbusinessanalytics.techtarget.com/definition/socialmedia-analytics> on the dated: 16-08-2015
- [9] Web content available on the link: http://en.wikipedia.org/wiki/Ecommerce_in_India#cite_noteOnline_shopping_touched_new_heights_in_India_in_2012-1 on the dated: 16-08-2015.
- [10] Amir Gandomi , Murtaza Haider, —Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management 35 (2015) 137–144, journal homepage: www.elsevier.com/locate/ijinfomgt
- [11] Radhakrishnan, Ajay Divakaran and Paris Smaragdis, —AUDIO ANALYSIS FOR SURVEILLANCE APPLICATIONS, 2005 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, October 16-19, 2005, New Paltz, NY.
- [12] C.L. Philip Chen, Chun-Yang Zhang, 2014, —Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Contents lists available at ScienceDirect Information Sciences, 275 (2014) 314–347
- [13] Stephen Kaisler, Frank Armour, J. Alberto Espinosa, William Money, 2013, —Big Data: Issues and Challenges Moving Forward, 2013 46th Hawaii International Conference on System Scien 1530- 1605/12, 2012 IEEE,
- [14] Web content available on the link: —<http://analyticstraining.com/2014/the-power-of-social-mediaanalytics/> on the dated: 17-10-2015



International Journal for Innovative Engineering and Management Research

A Peer Reviewed Open Access International Journal

www.ijemr.org

[15] Edmon Begoli, James Horey, 2012, —Design Principles for Effective Knowledge Discovery from Big Data, 2012 Joint Working Conference on Software

Architecture & 6th European Conference on Software Architecture, 978-0-7695-4827-2/12, IEEE