OPINION MINING IN BIG DATA: TREND OF THINKING FOR BIG DATA ERA

Mubarak Himmat

Faculty of Cmputer Science, Future University, P.O. Box 10553 Khartoum, Sudan
barakamub@yahoo.com

Mnahel Ahmed Ibrahim

College of Science and Humanities Studies, University, Shaqra,Durma-KSA
mnahel@hotmail.com

ABSTRACT: *This ear with the rapidly growing of internet and network using there are a huge data that have been introduced, Big Data are now on the double expanding rabidly in all domains, including opinion and sentiment analysis, for there are many social media and other websites that offer chances to provide the visitors and customers to post their opinion which usually contains valuable information that could be helpfully for several issues. And there are different methods and techniques that proposed to face this huge data and the big social data to make it more beneficial for several fields. This Paper introduces the big data and the most common it is usage and challenge, and it also investigate the sentiment analysis and it is common techniques and thinking about it is futures. This paper also thinking about the future of big data and opinion mining is clearly discussed and thinking about the future of big data and opinion mining. And the paper will discuss the challenges that facing the big data and opinion mining.*

KEYWORDS: big data, data mining, social media, opinion mining

INTRODUCTION

No doubt that the internet and world wide web(www) s full of data , and every day there is a huge of data and information are increasing, and every day the scale of data and information are increasing ,and according of the internet real statistic [1] that demonstrate the live statistic of the internet user are become more than one and have billion , and websites ,and blog post ,tweets are increasing rabidly every seconds and every days and the thinking about the future become very important and many thousands of researchers are working to Maximum utilization of these data ,the big data [2]for. There are so many things effect and volume of big data such as web and social media, machine ,sensing ,transactions, Internet of Things (IoT), these sources of big data are represented in deferent format such as structured data, semi structured data ,and unstructured data,

Data Mining analyzes and translates data from various viewpoints into useful information that can be used for business solutions and forecasts future trends. Information and data mining lets companies make decisions guided by data. Data mining (DM) is also known as Knowledge Discovery in Databases (KDD) or Knowledge Discovery and Data Mining. Data Mining analyzes and translates data from

various viewpoints and viewpoint into useful information that can be used for business solutions and forecasts future trends.

Big Data is a new term used to describe large-scale, grater complex data sets [4]. Big data is characterized as a large quantity of data that requires new technology and architectures to allow the capture and analysis process to extract value from it. Big data is important because the more information we collect, the more accurate we get and the more company we can automate. Big Data Mining Data is could be considered a new term used to describe large-scale or huge of data that available on digital source, grater complex data sets[3, 4]. Big data is characterized as a large quantity of data that requires new technology and architectures to allow the capture and analysis process to extract value from it. Big data is very relevant and important for the purpose of industry and business and society. The data came from all over the world like sensors used to collect climate information, post or share data on social media sites, video film audio, etc. This data collection is known as "BIG DATA[5, 6]. So with our current methodologies or data mining software tools, we can't store, handle and analyze them. Big data is one of them.

Now there big data sources are increased such as the Financial sources which could be Image, video, twitter and news reporters, Healthcare data and information which contains Health digital documents and images, Internet of Things (IoT) which contains the data that comes from sensors and real time systems, Media/Entertainment sources which Content and user viewing behavior, Social Media which is considered Blog posts, tweets, websites for social networks, log info and forums. Telecommunications sources, and Video Surveillance TV records and recording system for IPTV cameras, etc.

BIG DATA MINING AND ANALYTICS

Data mining and it is processes are now become one of the rich aspect of computer science researches ,the big data mining is an aspect of data mining which is refer to the activity of searching for relevant information through large data sets. Big Data Mining is the ability to extract useful information from these large data sets or data streams that were not previously possible due to their volume, variety, and speed. As it has been mentioned in the introduction the extracted knowledge is beneficial and very useful and the mined knowledge is the representation of different types of patterns and each pattern corresponds to knowledge. The main objectives of large and big data mining techniques go beyond obtaining the information requested or even uncovering some hidden relationships and patterns between numerical parameters. The process of finding useful knowledge that can be used in Big Data insights from large data sources and huge databases called Big Data mining leads to difficulties in Big Data mining. Today, social networks such as search engines, sensors, weblogs, blogs, e-commerce and some other applications generate large volumes of data (zeta bytes) with respect to their routine work.

In Big data analytics examines huge amounts of data to uncover hidden patterns, correlations and other insights. With this era and the current available technology, it

could be said that it's possible to analyze the data and get the results of the analysis easily and perfectly and immediately – an effort that's slower and less efficient with more traditional business intelligence solutions. The past two decades have witnessed the Appearance of concept of big data; most organizations nowadays are concerned on big data and they understand carefully that that if they collect all the data that their businesses produced, they can apply analytics analysis on this data and they can get significant value from it. Big data analytics ' current and emerging focus is on exploring traditional techniques such as rule-based systems, pattern mining, decision trees and other data mining techniques to effectively develop business rules even on large data sets. Either by designing algorithms that use distributed data processing, or by using in-memory computers. The Big data analytic is very important for helps organizations to exploit and exploit their data to identify new opportunities. Currently, there is a lot of research have been done in in data mining and analytic [7-9] This in turn results in more intelligent business activities, more productive processes, higher profits and happy customers. Cloud-based analytics bring significant cost benefits when storing large amounts of data – plus identifying more efficient ways to do business. Through Hadoop's speed and in-memory analysis combined with the ability to evaluate new data sources, companies can quickly analyze knowledge – and make decisions based on what they have found. Different goods and services. Through analytics, the ability to measure customer needs and satisfaction comes the power to give consumers what they want. Davenport points out that more businesses are creating new goods using big data analytics to meet the needs of consumers.

SOCIAL MEDIA EFFECT IN OPINION MINING AND BIG DATA

In the age of social bonding and social colonization, people are to become more excited interaction, sharing, and collaboration through online collaboration media. In recent years, this collective intelligence has spread to many different kinds regions, with particular emphasis on regions related to everyday life such as commerce, tourism, education, and health, causing the size of the social network develop exponentially[10]. However, the process of distillation, without knowing much structural information, is a very difficult task because the content is ideal for human consumption online today, but it is difficult for machines to understand.

Social media and big data have been brought together to create a new study called Social Network Mining, similar to data mining, but limited to the world of Twitter, Facebook, Instagram, and the like. The analysis of social networks is a "process of representation, analysis, and elimination of patterns of information in social networks". The exponential growth of individual users of social Media (SM) and active companies in social networks (Facebook reported over 2.27 billion active accounts at the beginning of 2019) creates an interaction space for customers [11, 12]. SM provides live communication with customers through text, sounds, photos and videos[13], in which the emotional reaction of SM users is to involve the brand on page, advertising, brand loyalty and, indirectly, consumer behavior concerned[14].

The figure 1.shwo the Most popular social networks worldwide as of October 2019, ranked by number of active users that counted by the Stataista web site the provide different types of statistics for scientific researches[15] .
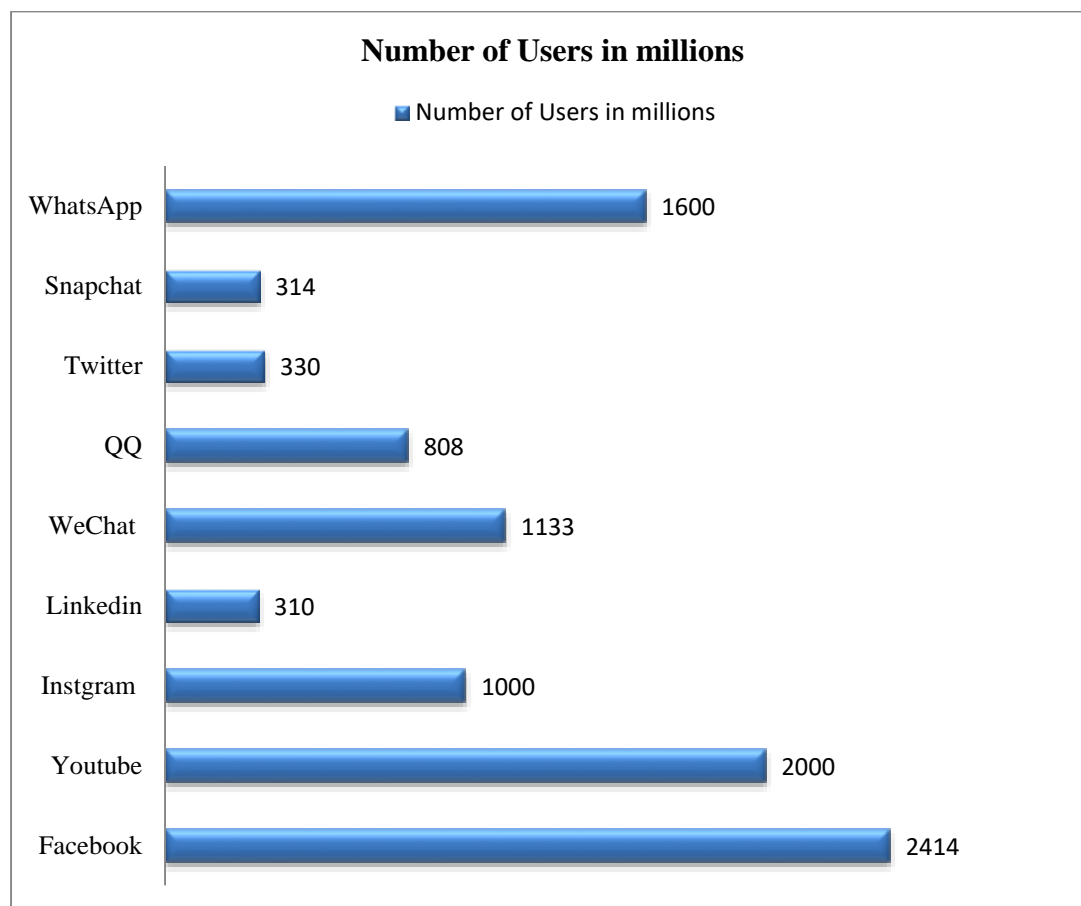

Figure1.Number of active users in common social media in millions (October 2019).

Social networks are a rich source of large amounts of data that are difficult to manage without automation. Machine learning (ML) algorithms can be an effective way to manage big data and interpret results to be useful to individual users and businesses. Social and mobile media generate a large amount of data, that is, big data that can be used by the government, companies, and individuals to extract public opinion in order to improve the decision-making process. As a result, extracting big data representations and mining this data plays an important role in the success of decision-making process. The main concepts of public opinion mining and big data are briefly described below: Opinion Mining, is an area of research that uses natural language processing techniques to extract, capture, or determine a person's relationship (if not his feelings) to a particular topic. It is an automatic extraction of attitudes, opinions and emotions from textual, linguistic and source databases through NLP[16]. The opinion mining aimed at identifying, using or extracting subjective information in documents containing textual data to determine the publisher's attitude to the subject or the polarity of feelings of a

particular document that represents his or her assessment, location or judgment on a specific topic, event, topic or product and mood analysis on [17].

DATA MINING AND OPINION MINING TECHNIQUES

Recently, web documents have received great focus within a new medium, which can be used to obtain personal opinions and experiences. This brings more attention to technologies for mining and analyzing opinions, through web documents such as forums, weblogs, and customer review websites [9]. The growth of research in this area is as result of the easily-accessible documents across the web. Another catalyst is the maturity of machine-readable techniques. Furthermore, machine learning methods in Natural Language Processing (NLP), as well as Information Retrieval (IR), have seen a considerable increase in the development of practical methods.

The task of marking documents as positive or negative is called sentiment classification [10]. It is assumed that opinion-based texts express relevance to one objective and that opinions belong to the same decision-maker, which is the case for product evaluations done by clients. Critics of films, books or products written by a critic who expresses his opinion may be positive or negative. This distinction uses classification at the document level.

The classification of opinions and moods uses two broad classification methods. These include machine learning and lexicon-based methods. Machine learning can be subdivided into supervised and semi supervised methods. On the other hand, lexicon - based methods, sometimes called unsupervised methods, are classified as corpus-based methods or dictionary methods. However, any existing supervised learning method can be used to rank the level followed by the document. The most commonly used methods in this regard include the support vector machine, Maximum Entropy, Naive Bayes and Artificial Neural Network [17, 18].

For semi supervised methods, unlabeled documents may be included in the learning process to improve the performance of the sentiment classification process if there is not enough trained corpus to form the model of classification. Graph-based methods and SVM are two semi-supervised methods used to classify sentiment at the document level [17].

The big data challenge is to manage its complexity (speed, volume and diversity) [19],and to process it in a multi-application distributed environment[20]. Traditional classical teaching methods are not adapted to the nature of big data and do not take into account the difference between the number of samples belonging to different classes in a data set[21]. To solve this problem, various methods have been developed that are often divided into two categories. The first extends some methods of binary classification to make them applicable to classification problems of several classes, such as discriminate analysis, decision trees, k-nearest neighbors, naive Bayesian algorithms, neural networks, and support vector machines. The second category is known as the decomposition and ensemble method. It is to subdivide the problem of classifying multiple classes into a series of binary classification problems that can be

solved with binary classifiers and then classifying the new observation by applying an aggregation strategy to binary classifiers predictions [21].

Big Data and data mining and opinion mining challenges .Recently there are many works that focused on big social data [22-25] another work discusses the Combines multidisciplinary analysis for examples such as multimedia management, social media analysis, social media analysis, trend discovery, and opinion mining.

CONCLUSION

In this paper, the definition of Big Data in data mining was discussed and the challenges of big data and opinion mining are demonstrated. The paper studied the evolution of a social network as a graph, for example, is very limited because it does not take into accounts the information flowing between network nodes. It is a challenge to store and retrieve vast amounts of structured as well as unstructured data at a desirable time lag. Some of these limitations in handling and processing vast amounts of data using traditional storage techniques led to the emergence of the term Big Data. Due to the emergence of the Internet, though big data has gained attention, but cannot be compared with it. From this study we can found that the mainly challenges that face the big data could be summarized in the following Challenges of Internet of Things and Big Data Integration, Energetic Provisioning, Big manipulations in data mining, and the other important issue is Big data security and privacy, the analyzing and getting the advantages of   sentiment is expressed about a product or concept in a blog or forum or in social media , The difficulties of statistical calculation of the utility of various authority measures and the impact of the enormous amount of this data. the development of algorithms for large data mining is also a challenge for developer Development of a standard to big data mining unifying framework. The huge amount rabidly growth of the social data. .

REFERENCES

1.   https://www.internetlivestats.com/.
2.   Ward, J.S. and A. Barker, *Undefined by data: a survey of big data definitions.* arXiv preprint arXiv:1309.5821, 2013.
3.   Bifet, A., *Mining big data in real time.* Informatica, 2013. **37**(1).
4.   Thakur, B. and M. Mann, *Data mining for big data: A review.* International Journal of Advanced Research in Computer Science and Software Engineering, 2014. **4**(5).
5.   Daniel, B., *B ig D ata and analytics in higher education: Opportunities and challenges.* British journal of educational technology, 2015. **46**(5): p. 904-920.
6.   Furht, B. and F. Villanustre, *Introduction to big data*, in *Big data technologies and applications*. 2016, Springer. p. 3-11.
7.   Provost, F. and T. Fawcett, *Data Science for Business: What you need to know about data mining and data-analytic thinking*. 2013: " O'Reilly Media, Inc.".
8.   Kannan, N., et al., *Predictive big data analytic on demonetization data using support vector machine.* Cluster Computing, 2019. **22**(6): p. 14709-14720.

9.   Barkhordari, M. and M. Niamanesh, *Arvand: a method to integrate multidimensional data sources into big data analytic structures.* J. Inf. Sci. Eng., 2018. **34**(2): p. 505-518.

10.   Cambria, E., et al., *Computational intelligence for big social data analysis [guest editorial].* IEEE Computational Intelligence Magazine, 2016. **11**(3): p. 8-9.

11.   Ladhari, R., et al., *Consumers' motives for visiting a food retailer's Facebook page.* Journal of Retailing and Consumer Services, 2019. **50**: p. 379-385.

12.   Chang, S.-H., et al., *The influence of web aesthetics on customers' PAD.* Computers in Human Behavior, 2014. **36**: p. 168-178.

13.   Gangadharbatla, H., T. Hopp, and K. Sheehan, *Changing user motivations for social networking site usage: implications for internet advertisers.* International Journal of Internet Marketing and Advertising, 2012. **7**(2): p. 120-135.

14.   Gutiérrez-Cillán, J., C. Camarero-Izquierdo, and R. San José-Cabezudo, *How brand post content contributes to user's Facebook brand-page engagement. The experiential route of active participation.* BRQ Business Research Quarterly, 2017. **20**(4): p. 258-274.

15.   https://www.statista.com/, *Statista* 2019.

16.   Kumar, A. and A. Sharma, *Systematic Literature Review on Opinion Mining of Big Data for Government Intelligence.* Webology, 2017. **14**(2).

17.   Ibrahim, M.A., *An Aspect-Oriented Sentiment Analysis Of Arabic Tweets Using Language Pattern and Semantic Analysis Techniques*, 2018, Sudan University of Science & Technology.

18.   Himmat, M. and N. Salim. *Survey on product review sentiment classification and analysis challenges.* in *Proceedings of the First International Conference on Advanced Data and Information Engineering (DaEng-2013).* 2014. Springer.

19.   Khan, N., et al., *Big data: survey, technologies, opportunities, and challenges.* The Scientific World Journal, 2014. **2014**.

20.   Oussous, A., et al., *Big Data technologies: A survey.* Journal of King Saud University-Computer and Information Sciences, 2018. **30**(4): p. 431-448.

21.   Del Río, S., et al., *On the use of MapReduce for imbalanced big data using Random Forest.* Information Sciences, 2014. **285**: p. 112-137.

22.   Manovich, L., *Trending: The promises and the challenges of big social data.* Debates in the digital humanities, 2011. **2**: p. 460-475.

23.   Zikopoulos, P. and C. Eaton, *Understanding big data: Analytics for enterprise class hadoop and streaming data.* 2011: McGraw-Hill Osborne Media.

24.   Hussain, A. and E. Cambria, *Semi-supervised learning for big social data analysis.* Neurocomputing, 2018. **275**: p. 1662-1673.

25.   Del Vecchio, P., et al., *Creating value from social big data: implications for smart tourism destinations.* Information Processing & Management, 2018. **54**(5): p. 847-860.