ISSN: 0975-766X
CODEN: IJPTFI
Available Online through *Research Article*
www.ijptonline.com

A STUDY ON THE ROLE OF BIG DATA ANALYTICS AND SECURITY IN SOCIAL MEDIA FOR BUSINESS DEVELOPMENT

Bhuvaneshwari. P[1], Nagaraja Rao. A[2]
[1]Research scholar, SCOPE, VIT University, Vellore, India.
[2]Associate Professor, SCOPE, VIT University, Vellore, India.
Email: nagarajaraoa@vit.ac.in

Received on: 11-02-2017 *Accepted on: 31-03-2017*

Abstract

Big data analytics in social media platform is in rapid progress with a lot of new tools and technologies emerging every day. Basically, analytics is complicated as it requires experts in cleaning data, understanding the problem, analyzing data by fixing proper tools and technologies and interpreting the accurate results from analysis. This paper presents how social media analytics is used to analyze the heterogeneous big data available in social network sites and provide insights into business development. The sentiment mining, visual analytics, competitive analytics, and business intelligence analytics to develop the enterprise is discussed. Finally, this paper demonstrates the security and trust prediction between users in social media analytics with future enhancements.

Keywords: Big data analytics, Social Media Analytics, Heterogeneous Big Data, Sentiment Mining, Visual Analytics, Competitive Analytics, Business Intelligence Analytics, Security And Trust Prediction.

1. Introduction

Big data plays an important role in all fields and the information is being generated every day by millions of computing machines and collected for future use. Social media is internet based communication technology which now turns into interactive platform technology[1]. The enterprises and their customers generate the huge amount of unstructured data using social media channels and that data is termed as big social data. In web 2.0 technology people use social networks like Facebook, Twitter, Linkedln, Youtube etc., to share their opinions, information, experience and perspective. They share information with others in the form of text, audio, video and image. The tremendous growth of the data has enabled us to uncover the hidden knowledge using analytics tool from accumulated data which produce new knowledge benefitting the society and the enterprise organizations. The innovations in big data tools and technologies lead the business firms to improve their business knowledge, enhancement, change, interoperability,

user-friendly and risk management. Traditionally, firms took a lot of efforts to know about their product by arranging group discussion, interviews and so on. Nowadays social media has become a major source for companies to analyze market analysis, recruiting process, public relations among people, and so on. This big data and social media analytics allow businesses to connect with customers and provide opportunity in the wide range of applications.

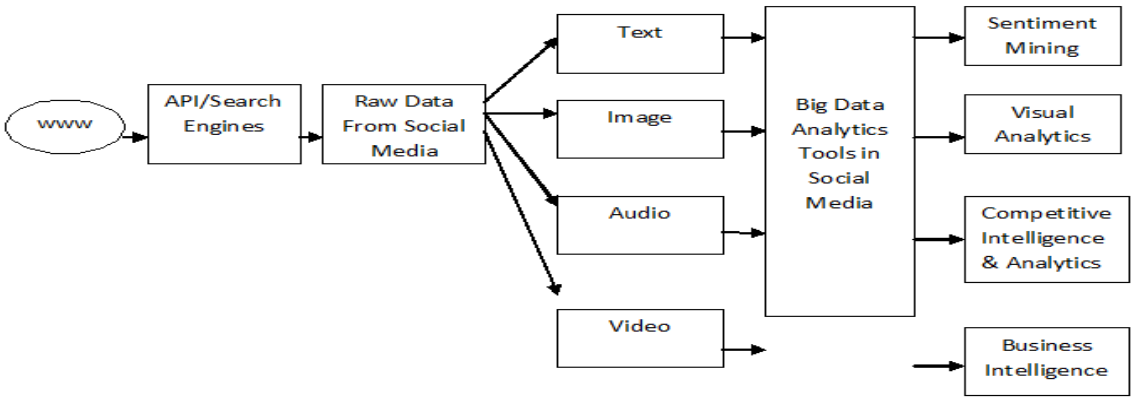

Fig.1 Schematic Representation of Big Data Analytics Role in Social Media.

The Fig.1 describes about the type of data generated by the user and the analytics methods available to process the big data in social media. Big data analytics uses content-based analysis to extract the knowledge pattern from the user generated data in social network sites. The audio analysis uses speech recognition or phonetic approach to extract information from unstructured audio data. Video analysis involves various image processing methods like face recognition and sentiment analysis for analyzing the video streams. Text analysis uses various methods like information retrieval, natural language processing, and text summarization to extract the patterns from textual data. Image analysis and text mining are integrated together and used in most of the analytics. In this work, we discuss the role of big data analytics in social media for business analysis and assessment.

2. Role of Big data Analytics in Business

Big data analytics is growing fast as organizations employ technologies to tap the terabytes of data flowing into their organizations through some sources like social media feedbacks. Usage of advanced analytics is important for Business Intelligent (BI) systems towards analysis of returns of investment. Analytics provide a deeper perspective on data and BI makes analytics actionable by visualizing, reporting, and providing management metrics. The present work discusses some of the analytics method with its applications in business firm using social networking sites.

2.1 Social Media Analytics

Social media is a dynamic platform for information sharing and opinion imprinting by end users in websites. Social network analysis relates and measures the relationships and flows between individuals in a social network. The social

network[11] is a web-based network in which each individual creates their own profile and communicate with others belongs to the same network. It has the characteristics of user generated data and virality, which enable driven applications such as crisis management, firm performance improvement, and viral marketing. The same network people share their opinion[12] about politic issues, cinemas, popular person, entertainment, area of interest, traveling and food experience, shopping, sports, culture, health, and technology. By analyzing the discussion effectively, we can identify the social response about the particular issues. Community detection algorithms like Lead - Follower algorithm[18] is used to identify the internal structure of a community, members present in the same clique, and infer the information from the tweets. It is used to capture the "pulse" of individual emotions and monitor the political or marketing campaigns. Social media analytics shows interest in developing tools which capture, monitor, analyze, predict the result and extract the relevant patterns and intelligence. It has the capability to increase the relationship strength and organizational quality views. As a large amount of data is gathered on daily basis, Data-driven decision making becomes easier. Businesses use social media to promote their product and services by staying touch with customers through Facebook, Twitter, E-mail, Youtube etc. The customer review[2] plays a vital role in finding the flaws and strong point present in the product and marketing intelligence is gained through their feedback. It analysis both structured and unstructured data in online platform and service quality, product quality and fixing the prices are effectively enhanced. Madjid Tavana et al.[8] proposed a novel analytics hybrid framework for choosing the most suitable social media platform. Social media is a powerful tool for marketers to expose their product effectively. Number of social networking sites are available more in online but only very few are popular. Selecting the wrong social media platform leads to the company reputation loss, money waste, and detriment. The Analytic Network Process (ANP) with fuzzy set theory determines the weight of social media platforms and COPRAS-G method selects the most suitable social media platform for the firm.

2.2 Sentiment Mining

Sentiment analysis is one of the important technique used by social media analysis. This is used for computational deduction, the study of opinions, sentiments, emotion recognition and subjective in the text. It is the process of finding out opinions and extract the relevant data contained in textual information by filtering out the irrelevant component. It uses natural language processing, semantic, statistical and computational linguistics to extract the subjective context. Opinion mining analyzes the opinionated text from different sites and generates the meaningful business insight reports. It has been used in various prediction and design applications with the help of positive and

negative feedbacks[2]. Sentiment analysis serves its best way in providing scalability, data mining, risk management, and increase business profits. It relies on machine learning algorithms like support vector, naive Bayes, maximum entropy and matrix factorization. In social media, billions of people share their perspectives in a different manner about various topics. Content clustering method [13] is used to aggregate the related content from a wide variety of sentiments posted by the generalized user. Density based algorithm is implemented to remove the noise from the sentiments and Shared Nearest Neighborhood (SNN) is used to find the similarities between different clusters. SNN provide better performance with high complexity rate of O(N^2), however it is not widely used in web opinions. Christopher C. Yang et al.[13] proposed SDC algorithm which combines the related cluster and filters the noise by measuring, the scalable distance from the initial cluster to the required density. Currently Business organizations using deep sentiment analysis to find the characteristics of user before considering his/her review on certain products in social media. The attitude of a person is evaluated based on the posts, comments, likes, dislikes, personal information given by him in social media. This evaluation is done by using Personality–Value–Attitude (PVA) model and it has more impact . This model classifies the person into three types. (i) utilitarian person who rely on utility (ii) hedonistic person who gives more priority for enjoyment (iii) normative person who focus on social goodness. Based on the type of the person and their emotions, the subjective of the issue is analyzed and the result is predicted.

2.3 Visual analytics in social media

Visual analytics tools are defined as "analytical reasoning facilitated by interactive visual interfaces" which finds the communication between parties in an online forum, classify them, cluster the needful patterns and explore the reason behind every problem[3]. Social media data supports both the network structure and homogeneous data. Traditional visualization tools lack the capability to handle the huge volume of data and later new techniques were evolved to handle the social data. This new analytical tool performs along with statistical tools, machine learning algorithms, and signal processing methods. In online the user generate the surplus amount of data which consists of both useful and non-useful data. To retrieve the relevant data from social media three approaches are practiced (i) keyword based approach (ii) Topic-based approach and (iii) multi-faceted approach. keyword based approach retrieves the information based on the specific keywords or terms. Diakopoulos et al.[21] proposed a visual analytics keyword based system known as "Seriously Rapid Source Review". It is a dictionary based detector used to retrieve the breaking news from tweets. The topic-based approach uses advanced text mining methods, natural language processing techniques and retrieves information based on semantic modeling and clustering. Scatter blog is a visualization topic

based approach which uses Latent Dirichlet Allocation method to extract the inherent topics. It uses z score evaluation method to produce unusual and unexpected topics along with the searching topic result. Multi-faceted approach gets information through various different perspectives and uses advanced data mining techniques to produce the summary of information. TwitInfo is one of the multifaceted approaches which combines and visualizes huge volume of online data and explore an event using novel streaming algorithm. Web opinions are usually disseminate, crisp, less organized text message and keep evolving. For easy and precise representation ontology is used in social media data retrieval, which gives the formal specification of concepts and their relationships. Ontology evaluation[5] involves verification, validation, assessment, consistency, completeness and conciseness. Ontology mining algorithm extracts the inside and outside product characteristics and produces the aspect-oriented analysis[7] using LDA-based computational algorithm. Ontology-based Product Review Miner (OBPRM) is a benchmark evaluator which calculates the procedure taken place in datasets. Ontology Supported Polarity Mining (OSPM) is used to enhance the polarity review by extracting the features, concepts and the relationships among the posted reviews. A fuzzy based ontology captures the uncertainty present in product oriented identification and context sentiment prediction.

2.4 Competitive analytics

Competitive intelligence is the process of examining the competitive environment with the goal of providing actionable intelligence. It is used in the decision-making process and allows the company to identify competitors strength[3], weakness, strategies and other areas that help firm to improve its decision against the competitor. Currently, the companies are keeping eye on finding out in which area they are doing well and in which part they are lagging. A better result is achieved by comparing the customer sentiments with the feedback of other firms belongs to the same industry. Wu He et al.[10] deployed social media competitive analysis framework with sentiment benchmark which gleans industry specific marketing intelligence. It gathers data from both implicit and explicit information sources, analyze it with competitive firms performance and finally produce the meaningful business insight reports.

Financial runners disclose their information in social media, which leads users to gain more knowledge about latest trends, investment information, profitable ratio, corporate news and so on. The voluntary disclosure of the information leads the company to attain its goal and reach the destined profit by increasing company influence. It also helps the users to do a comparison with other companies belongs to the same industry. Information is delivered to users through communication medium like mobile SMS, e-mail alerts, social media, company website and rich set

summary. Juheng Zhang et al.[9] proposed pooled t-test, folded f method and Kruskal-Wallis test to investigate the impact of media adoption in the intensity of information disclosure. K-means cluster analysis methods is also used to find the adoption level of social media in business improvement.

2.5 Business intelligence and analytics

Big data analytics helps the firms to exploit big data for developing the business in various forms[6]. Companies receive a vast amount of data like transactional data, voice data, video data and social media data. BDAC takes decision making by considering management, technology and personnel talent with the feedbacks received from customers. The unstructured and noisy data present in social media comments makes companies face difficulties in the area of improvement. The main focus of Business intelligence[4] is to improve the quality of the product as well as profit. Business intelligence collectively refers to software and systems that import data streams of any size and use them to generate informational displays that point towards specific decisions. In business activities, the operations like gathering data, examining, measuring and inferring exploits lead the decision-making process. The analysis of consumer review is beneficial for both the customers and owners but still, difficulties exist because of their huge volume, veracity, and velocity of data. Nowadays consumer shares their opinion about the product and the services genuinely, this leads the company to leverage about the product resource in the competitive world. Mohammad Salehan et al.[2] proposed a sentiment analysis software called SentiStrength Software. It analyses the emotions in a text, debug the recurrent words and reproduce the positive result in negative expressions. Consumers have a perception to buy the product and this analytics gives the reassurance to the quality of the product. Sentiment analysis helps the customer to gain more knowledge about the product before they buy. Feature-based summarization extract the product features and apriori association rule mining algorithm is used to highlight the frequently occurring feature in the product review. To identify the context sensitive in sentiments, linguistic rules are applied and semantic orientation analysis is used to identify the polarity of sentiments. To optimize business operations and to take better decision making, business people need to track up-to-date data and analyze it effectively. In Business analytics nowadays ETL(Extract-Load-Transform) process is implemented in all firms to improve the profit gains and enhance the performance. ETL process[17]extract data from various data sources (Business Intelligence), convert it to the predefined syntactic rule, process the data and store the result in the new model which can be used for business analysis or monitoring. Thus quality management is done by filtering the unwanted data and reconciled into syntactic models. As big data grows day by day, the processing speed can be improved by following event-centric approaches

which access the data in distributed multi-thread fashion and deliver the result without any time delay. Business Intelligence has the drawback of processing only the historical data, to overcome that drawback Business analytics monitoring (BAM) was evolved. BAM is an enterprise event driven solution which collects the data from in and out of an organization, analyze it from web services, customers, partners, message queues etc., summarize the real-time data activities with solutions, and produce the result to management.

3. Trust and security in social networks

In social media, various attackers try to steal others identity in the motivation of gaining their own profits. They use attacks [14] like spam attack, malware attack, Sybil attack, social phishing, impersonation, hijacking and sending a fake request. It threatens the user privacy and trustworthiness of the network by affecting the completeness, confidentiality, accessibility and availability. The enhanced access control mechanism[20] is introduced initially by implementing scalable fine grain access control model. This model performs authentication, filtration, management of network and improves the security of social media network. For specific multimedia analysis and storage, fine-grained multimedia access control model with bit-vector transform domain is implemented. Kappa- Fuzzy based trust interference mechanism[19] to measure the trustworthiness of the same person belongs to different community proposed an effective. Wu et al.[20] developed a framework for sending recommendations based on user-based interest. This GCCR framework divides the user into small groups based on content-based clustering and utilizing graph summarization. Recommended friend of known friends improves the quality of service and trust in social media networks. For predicting the trust in an online social network[15] Synthetic Minority over sampling Technique Boost (SMOTE) algorithm is applied and accurate result is obtained. In virtual environment participants takes decisions based on the communication between anonymous members. It leads to Face to Face communication lagging and trust-based issues. Golbeck defined trust in social web as "trust in a person is a commitment to an action based on a belief that the future actions of that person will lead to a good outcome''. The prediction of trust between users in an online application can be done through classification and machine learning algorithms. The interpersonal trust prediction[16] can be done through syntactical approach by combining both contextual and structural data. The trust built between users in social media is based on four topics (i) Trust based on policy (ii) Trust based on reputation (iii) general models based on trust and (iv) information based resources. This framework employs five qualitative factors like relationship, reputation, knowledge, same perception and personality based trust. It is then mapped with some specific measurable datasets and result is retrieved.

4. Conclusions: Social media plays an important role in enhancing the business by promoting their product and services. This paper provides insight to select the right social platform to analyze and trustable user feedbacks with different analytics method. The interpersonal trust between online users are evaluated with a syntactic approach and in future, the time varying trust prediction can also be evaluated with privacy - preserving collaborative filtering. We find that customer sentiments have more influence in real time applications. Competitive analysis framework collects feedback from the customer every day, finds its strengths and weakness, and tries to improve their product, profit, and performance. In addition, the relationship between the organization versus customer relationship, business performance versus customer sentiments can also be done in future. Visual analytics analyze, explore and provide meaningful information by using various approaches and algorithms. However, it has an issue in handling overloaded social media data from various sources.

References

1. Vatrapu R, Mukkamala RR, Hussain A, Flesch B, " Social Set Analysis: A Set Theoretical Approach to Big Data Analytics," 2016, Vol 4, pp 2542-2571.
2. Salehan M, Kim DJ, "Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytic," Decision Support Systems, 2016 Jan 31,Vol 81, pp 30-40.
3. Ribarsky W, Wang DX, Dou W, "Social media analytics for competitive advantage," Computers & Graphics, 2014 Feb 28, Vol 38, pp 328-31.
4. Holsapple C, Lee-Post A, Pakath R, " A unified foundation for business analytics," Decision Support Systems, 2014 Aug 31, Vol 64, pp 130-141.
5. Chauhan A, Vijayakumar V, Vincent R, Pradeep KV, " Towards the Development of a Framework for Socially Responsible Software by Analyzing Social Media Big Data on Cloud Through Ontological Engineering," Procedia Computer Science, 2015 Dec 31, Vol 50, pp 524-30.
6. Ram J, Zhang C, Koronios A, " The Implications of Big Data Analytics on Business Intelligence: A Qualitative Study in China," Procedia Computer Science, 2016 Dec 31, Vol 87, pp 221-226.
7. Lau RY, Li C, Liao SS, " Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis," Decision Support Systems, 2014 Sep 30, Vol 65, pp 80-94.

8. Tavana M, Momeni E, Rezaeiniya N, Mirhedayatian SM, Rezaeiniya H, " A novel hybrid social media platform selection model using fuzzy ANP and COPRAS-G," Expert Systems with Applications. 2013 Oct 15,Vol 40(14), pp 5694-5702.
9. Zhang J," Voluntary information disclosure on social media, " Decision Support Systems. 2015 May 31, Vol 73, pp 28-36.
10. He Wu, Yan G, Akula V, Shen J, " A novel social media competitive analytics framework with sentiment benchmarks," Information & Management, 2015 Nov 30, Vol 52(7), pp 801-812.
11. Peng S, Wang G, Xie D, "Social influence analysis in social networking big data: opportunities and challenges," IEEE Network, 2016 Oct.
12. Sapountzi A, Psannis KE, "Social networking data analysis tools & challenges," Future Generation Computer Systems, 2016 Oct 21.
13. Yang CC, Ng TD, "Analyzing and visualizing web opinion development and social interactions with density-based clustering," IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans," 2011 Nov, Vol 41(6), pp 1144-1155.
14. Zhang Z, Gupta BB, "Social media security and trustworthiness: overview and new direction," Future Generation Computer Systems, 2016 Oct 12.
15. Raj ED, Babu LD, "An enhanced trust prediction strategy for online social networks using probabilistic reputation feature," Neurocomputing, 2017 Jan , Vol 219, pp 412-421.
16. Zolfaghar K, Aghaie A, "A syntactical approach for interpersonal trust prediction in social web applications: Combining contextual and structural data," Knowledge-Based Systems, 2012 Feb 29, Vol 6, pp 93-102.
17. Longo A, Giacovelli S, Bochicchio MA, "Fact–Centered ETL: A Proposal for Speeding Business Analytics up," Procedia Technology, 2014 Dec 31, Vol 16, pp 471-480.
18. Shah D, Zaman T, "Community detection in networks: The leader-follower algorithm," arXiv preprint arXiv:1011.0774, 2010 Nov 2.
19. Chen S, Wang G, Jia W, "K-FuzzyTrust: efficient trust computation for large-scale mobile social networks using a fuzzy implicit social graph," Information Sciences, 2015 Oct 10,Vol 318, pp 123-143.
20. Wu J, Chen L, Yu Q, Han P, Wu Z, "Trust-aware media recommendation in heterogeneous social networks," World Wide Web, 2015 Jan 1, Vol 18(1), pp 139-57.