Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354

http://www.ripublication.com

Literature Review and Challenges of Data Mining Techniques for Social Network Analysis

Anu Sharma[1], Dr. M.K Sharma[2] & Dr. R.K Dwivedi[3]

[1]Research Scholar, [2]Associate Professor, [3]Professor

[1]Uttrakhand Technical University, Dehradun, [2]Amrapali institute, Haldwani, India.
[3]Teerthanker Mahaveer University, Moradabad (U.P.), India.

Abstract

Data mining is the extraction of projecting information from large data sets, is a great innovative technology. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Web sites contain millions of unprocessed raw data. By analyzing this data new knowledge can be gained. Since this data is dynamic and unstructured traditional data mining techniques will not be appropriate.

In this paper we discuss about data mining techniques. In this paper a survey of the works done in the field of social network analysis and this paper also concentrates on the future trends in research on social network analysis. This paper presents study about social networks using Web mining techniques.

Keyword: Social Networks, Web Data Mining, Data mining techniques, Social Network Analysis, Clustering.

1. INTRODUCTION

Data mining is a powerful tool that can help to find patterns and relationships within our data. Data mining discovers hidden information from large databases [34]. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Social networks can be used in many business activities like increasing word-of-mouth marketing, marketing research, General marketing, Idea generation & new product development, Co-innovation, Customer service, Public relations, Employee communications and in Reputation management[42].

There are various data mining techniques:

a. Characterization

Characterization is used to generalize, summarize and possibly different data characteristics.

b. Classification

Data classification is a process in which the given data is classified in to different classes according to a classification model.

c. Regression

This process is similar to classification the major difference is that the object to be predicted is continuous rather than discrete.

d. Association

In this process the association between the objects is found. It discovers the association between various data bases and the association between the attributes of single database.

e. Clustering

Clustering involves grouping of data into several new classes such that it describes the data. It breaks large data set into smaller groups to make the designing and implementation process to be simple. The task of clustering is to maximize the similarity between the objects of classes and to reduce the similarity between the classes.

f. Change Detection

This method identifies the significant changes in the data from the previously measured values.

g. Deviation Detection

Deviation detection focuses on the major deviations between the actual values of the objects and its expected values. This method finds out the deviation according to the time as well the deviation among different subsets of data.

h. Link Analysis

It traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques.

i. Sequential Pattern Mining

This method involves the discovery of the frequently occurring patterns in the data.[34]Social networks are important sources of online interactions and contents sharing ,subjectivity , assessments , approaches, evaluation , influences , observations, feelings, opinions and sentiments expressions borne out in text, reviews, blogs, discussions, news, remarks, reactions, or some other documents .[35]

2. LITERATURE REVIEW

In recent years, social media have experienced tremendous growth in their user base. For example, there are more than one billion members belonging to the Facebook network (Facebook 2013), while Twitter now has more than 280 million monthly active users (GlobalWeb- Index 2013). There are a large number of different social media applications or platforms which in general can be categorized as weblogs, microblogs, social network sites, location-based social networks, discussion forums, wikis, podcast networks, picture and video sharing platforms, ratings and reviews communities, social bookmarking sites, and avatar based virtual reality spaces (Zeng et al. 2010). Recent studies and surveys have revealed an emerging need to continuously collect, monitor, analyze, summarize, and visualize relevant information from social interactions and user generated content in various domains such as business, public administration, politics, or consumer decision-making (e.g., Zeng et al. 2010; Kavanaugh et al. 2011; Stieglitz et al. 2012). These activities, however, are considered difficult tasks due to the large number of different social media platforms as well as the vast amount, dynamics, and complexity of social media data. More specifically, social media communication generates an enriched and dynamic set of data and meta data, which have not been treated systematically in the data- and text-mining literature (Zeng et al. 2010). Tomoyuki NANNO present a system that tries to automatically collect and monitor Japanese blog collections that include not only ones made with blog software's but also ones written as normal web pages. This approach is based on extraction of date expressions and analysis of HTML documents. System also extracts and mines useful information from the collected blog pages. This approach obtained 39,272 blogs(pages) and 466,809 entries.[2] David Ediger Karl Jiang[2010] present a GraphCT, a Graph Characterization Toolkit for massive graphs representing social network data. Use GraphCT to analyze public data from Twitter, a microblogging network. analyzing public Twitter streams. applied GraphCT on the Cray XMT to their data set to gather performance data and still are analyzing the graph metric Results[11].Jyoti Shokeen presents the overview of social network analysis methods to represent social network by different means like matrix, formal methods and graphs and then followed by social network metrics [57].

2.1 Web Mining Methods and Soft Computing Approaches

Rupam Some focuses on the analysis of the pattern of relationships among people, organizations, states and such social entities. Recent trends on research are in area of link analysis, dark web analysis, and spam behavior detection. [24] Mariam Adedoyin-Olowe1 stated that accessing social network sites such as Twitter, Facebook LinkedIn and Google+ through the internet and the web 2.0 technologies has become more affordable. People are becoming more interested in and relying on social network for information, news and opinion of other users on diverse subject matters. Data mining provides a wide range of techniques for detecting useful knowledge from massive datasets like trends, patterns and rules. Data mining techniques are used for information retrieval, statistical modelling and machine

learning. These techniques employ data pre-processing, data analysis, and data interpretation processes in the course of data analysis. [35].Y. K. Mathur1[2014] stated that Soft Computing techniques and its impact as well as its new emerging trends to suit the changing requirements in the area of Data Mining.Soft computing methodologies, involving fuzzy sets, neural networks, genetic algorithms, rough sets, and their hybridizations, have recently been used to solve data mining problems. Recently, several commercial data mining tools have been developed based on soft computing methodologies. These include Data Mining Suite, using fuzzy logic.[39] Zahra Zamani Alavijeh stated that link mining is becoming a very popular research area not only for data mining and web mining but also in the field of social network analysis.[43] Pooja Rohilla stated that there are various tools available on the internet which mines the data according to their types like whether they are in a structured format or semi structured or unstructured data. Data is in semi structured format or it is totally unstructured data. To retrieve this type of data, we need to define proper patterns and clustering. In this, there are no tables in proper ordering. It can contain audio, videos, images etc in non structured format. Web Mining techniques are used for extracting relevant information from various type of websites like online shopping sites. [48]

Ritu Mewari presented an opinion mining provides a clear platform to catch public‟s mood by filtering the noise data. It also provides computational techniques used to extract and consolidate individual's opinion from unstructured and noisy text data. Opinion mining is a burning field of web mining. There exist a lot of benefits of opinion mining at customer and business level. A bulk of data is daily posted on web sites like face book ,twitter e.t.c. User post their sentiments in the form of comments, reviews and feedback daily .An opinion mining process gives us the way to extract pearl knowledge from it.[54]

2.2 Clustering

Clustering is a data mining technique used to place data elements into related groups. Clustering data in databases is an important task in real applications of data mining and knowledge discovery. It is the process of partitioning a data set into clusters so that similar objects are put into the same cluster while dissimilar objects are put into different clusters.Farhat Roohi provides a framework for neurofuzzy cluster analysis. Neural networks and the fuzzy set theory has emerged as a great breakthrough in the field of clustering, which is a process of grouping data items based on a measure of similarity. neurofuzzy system as it is a self learning system and generates patterns and rules automatically.[28] Pooja Sikka proposed SVM technique. The new technique is named as KMeans Clustering Based SVM (KMCB-SVM). The complexity of SVM depends on no. of input variables and support vectors. The proposed model will apply a clustering algorithm that scans entire data set only to provide the high quality samples that will carry statistical data. This will provide finer description closer to boundary and farther to boundary. KMCB-SVM would be used for classifying large data sets of relatively low dimensions in large warehouses.[52] Thai Le, Phillip Pardo

and William Claster stated that Artificial Neural Network (ANN) is an area of extensive research. ANN architecture in creating a Self-Organizing Map (SOM) to cluster all the textual conversational topics being shared through thousands of management tweets .ANN facilitated SOM is powerful in analyzing social media data to gain competitive knowledge in the field of tourism [60].

2.3 Classification Mining Tools

Classification used in web as well data mining technique to predict group membership for data instances (Fernandez et al 2010). Popular classification techniques include decision trees and neural networks (Rajen & Gopal 2006). Classification tree (also known as decision tree) methods were a good choice when the data mining task is classification or prediction of outcomes. Classification tree labels records and assigns them to discrete classes. A classification tree may also provide the measure of confidence that the classification is correct.

Classification tree is built through a process known as binary recursive partitioning. This is an iterative process of splitting the data into partitions and then splitting it up further on each of the branches. In social data mining, classification is considered as the most effective decision making techniques among all the human activities. A problem that occurs in classification is when a person or an entity has to be put into a class based on the predefined properties of the person or entity. Traditional methods available in statistical for classification including discriminate analysis were used in the past for decision making under uncertaining using Baye's theorem (Kazienko & Kajdanowicz2010). In this work, a new probability based model has been proposed by the authors to compute the posterior probability which is used for effective classification.Neural networks provide a number of advantages in the effective classification of data. First, neural networks based classification techniques are flexible and based on weight assignment. Second, they use function approximations with respect to the activation functions. Third, they are useful to form rules for learning and decision making. Finally, training and testing for neural networks are easy to implement. Gauri Joshi[2015] stated multi-label classification algorithm that is Naive Bayes Multi-label Classifier algorithm and memetic algorithm is applied to categorize tweets presenting students problems. Memetic algorithm is a population based approach with separate individuals learning for problem search.[50] Remya R S proposed multi-label classification algorithm that is used to classify comments in which attributes are divided into multiple categories reflecting student's problems. The proposed system implemented a naive bayes multi-label classifier, which allowed one comment to fall into multiple categories at the same time. This study is beneficial in learning analytics, educational data mining, and learning technologies. It provides a workflow for analyzing social media data for educational purposes that overcomes the major limitations of both manual qualitative analysis and large scale computational analysis of user-generated textual content.[55]

Table 1 [Literature Review Summary Table]

#	Author's Name	Paper Title	Techniques	Findings	Year	References
1	Marcelo Maia, Jussara Almeida, Virgílio Almeida	Identifying User Behavior in Online Social Networks	Clustering Algorithm	• Different interaction patterns can be observed for different groups of users. • characterizing and identifying user profiles in online social networks. • exploit the user be-havior to display more apropriate advertisements	2008	R6
2	Mohammad Al-Fayoumi, Soumya Banerjee, Jr., P. K. Mahanti	Analysis of Social Network Using Clever Ant Colony Metaphor	Clever Ant Colony Metaphor.	• to cluster social network structure through maximum clique and sub grouping criteria. • impressive performance on test run using the standard clique sequence data particularly for the clustering of the instance named as Keller 6.	2009	R10
3	David Ediger Karl Jiang, Courtney Corley Rob Farber, William N. Reynolds	Massive Social Network Analysis:Mining Twitter for Social Good	Graph CT SNA algorithm	• GraphCT to analyze public data from Twitter, a microblogging network. • analysts focus attention on a much smaller data subset. • more work on sampling is needed.	2010	R11
4	A. Selman Bozkır , S. Güzin Mazman , and Ebru Akçapınar Sezer	Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case	Decision Tree Algorithms, ANN and SVM.	• SVM exhibits the most accurate results	2010	R13
5	Stefano Baccianella, Andrea Esuli, and Fabrizio Sebastiani	SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining	SENTIWORDNET 3.0	• SENTIWORDNET 3.0 is an improved version of SENTIWORDNET 1.0. • SENTIWORDNET 3.0 is substantially more accurate than SENTIWORDNET 1.0, with a 19.48% relative improvement for the ranking by positivity and a 21.96% improvement for the ranking by negativity.	2010	R15
6	Alexander Pak, Patrick Paroubek	Twitter as a Corpus for Sentiment Analysis and Opinion Mining	Opinion Mining and Sentiment Analysis	• When the dataset is large enough, the improvement may be not achieved by only increasing the size of the training data.	2010	R14
7	Rojalina Priyadarshini, Nillamadhab Dash, Tripti Swarnkar, Rachita Misra	Functional Analysis of Artificial NeuralNetwork for Dataset Classification	Back Propagation Algorithm	• highly effective tool for dataset classification with appropriate combination of training, learning and transfer functions. • combination of TRAINLM, LEARNGDM	2010	R12

				and LOGSIG works better for comparatively smaller datasets • combination of TRAINSCG LEARNGDM and LOGSIG is effective for larger datasets.		
8	E.Raju, K.Sravanthi	Analysis of Social Networks Using the Techniques of Web Mining	Web Mining Techniques- Clustering, Association rule Mining	• data sampling is a big issue when using web mining for social networks analysis. • it becomes a difficult task to select suitable samples representative of the real social networks. • Other challenges include finding communities in social networks, finding patterns in social networks and analyzing overlapping communities. • how to apply the web mining techniques to some real on-line social networking websites, such as blogs and on-line photo albums	2012	R19
9	G.Vinodhini, RM.Chandrasekaran	Sentiment Analysis and Opinion Mining : A Survey	Sentiment Classification	• Sentiment detection has a wide variety of applications in information systems, including classifying reviews, summarizing review and other real time applications.	2012	R20
10	Xi Long , Wenjing Yin, Le An, Haiying Ni, Lixian Huang, Qi Luo, and Yan Chen	Churn Analysis of Online Social Network Users Using Data Mining Techniques	Decision Tree Classifier and K-Means Algorithms	• analyze the potential churn of users. • A DT-based and a k-means algorithms allow us to process large amount of data in a timely manner	2012	R22
11	Rupam Some	A Survey on Social Network Analysis and its Future Trends	Social network models: Statistical model for analysis	• Recent trends on research are in area of link analysis, dark web analysis, and spam behavior detection.	2013	R24
12	Paridhi Jain , Ponnurangam Kumaragur , Anupam Joshi	Identifying Users across Multiple Online Social Networks	Finding Nemo	• Finding Nemo, can be used by analysts to find flagged user identities (e.g. spammers) across networks	2013	R26
13	Neha Mehta , Mamta Kathuria, Mahesh Singh	Comparison of Conventional & Fuzzy Clustering Techniques: A Survey	Web Document Clustering	• improve the efficiency in information finding process • Fuzzy clustering are suitable for handling the issues related to understandability of patterns, incomplete/noisy data, mixed media information and human interaction, and can provide approximate	2013	R27

				solutions faster.		
14	Farhat Roohi	NEURO FUZZY APPROACH TO DATA CLUSTERING: A FRAMEWORK FOR ANALYSIS	Neurofuzzy Cluster Analysis.	• neurofuzzy system as it is a learning system and generates patterns and rules automatically.	2013	R28
15	Mita K. Dalal , Mukesh A. Zaveri	Automatic Classification of Unstructured Blog Text	Naïve Bayesian Model Artificial Neural Network Model	• NaïveBayesian classification model clearly out-performs the basic ANN based classification model for highly domain-dependent • unstructured blog text classification	2013	R25
16	G Nandi , A Das	A Survey on Using Data Mining Techniques for Online Social Network Analysis	Graph Theory	• Analysis of OSN data though has its solid basis in graph theory.	2013	R29
17	Mita K. Dalal and Mukesh A. Zaveri	Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews	Semisupervised Approach	• successfully identify opinionated sentences from unstructured user reviews and classify their orientation with acceptable accuracy. .	2013	R 30
18	Asad Bukhari, Um-e-Ghazia and Dr. Usman Qamar	CRITICAL REVIEW OF SENTIMENT ANALYSIS TECHNIQUES	Sentiment Analysis Techniques	• new automated mechanisms are required that help in drilling down to top microposts. • improve the sentiment analysis classifiers that can better analyze the user sentiments	2014	R 36
19	S.G.S Fernando	Empirical Analysis of Data Mining Techniques for Social Network Websites	Markov Models	• hybrid approach by combing social network analysis with content mining would be more useful. • The statistical methods like Markov Models can be adopted to resolve the temporal behavior of web data .	2014	R31
20	Sanjeev Dhawan, Kulvinder Singh, Vandana Khanchi	Critical Analysis of Social Networks with Web Data Mining	Web Mining Techniques	• finding communities in social networks structure, searching patterns in social networks and examining overlapping communities. • how to utilize the Web mining techniques to some real on-line social networking Websites, such as on-line photo albums, comments and blogs.	2014	R33
21	Sanjeev Dhawan, Kulvinder Singh, Deepika Sehrawat	Emotion Mining Techniques in Social Networking Sites	Sentiment Analysis (or opinion mining)	• Emotion generation and analysis have a number of practical applications including managing customer dealings, human machine interaction,	2014	R40

				information retrieval, natural text-to-speech systems, and in social and literary analysis.		
22	M. Vedanayaki	A Study of Data Mining and Social Network Analysis	Knowledge Based Network Analysis	• Focus on identifying global structural patterns. • Difficult to collect data	2014	R34
23	Md. Ansarul Haque , Tamjid Rahman	SENTIMENT ANALYSIS BY USING FUZZY LOGIC	Sentiment Analysis	• focus and analyze the extracted opinions (sentiments or emotional contents) from the posted comments	2014	R 38
24	Mariam Adedoyin-Olowe, Mohamed Medhat Gaber1 and Frederic Stahl	A Survey of Data Mining Techniques for Social Network Analysis	TRCM. (Transaction-based Rule Change Mining)	• conducted on social network analysis.	2014	R35
25	Ana¨ıs Collomb, Crina Costea, Damien Joyeux	A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation	Sentiment Analysis.	• models tend to target a simple global classification	2014	R 32
26	Meenu Sharma	Clustering In Data Mining : A Brief Review	Neuro and Fuzzy Logic Approaches	• ANFIS and FCM methods respond better to real life situation	2014	R37
27	Y. K. Mathur, Abhaya Nand	Soft Computing Techniques and its Impact in Data Mining	Soft Computing Techniques	• Soft computing methodologies- solve data mining problems. • it is easy to discover a huge number of patterns in a database	2014	R39
28	Esmaeil Fakhimi gheshlagh mohammad beig	Data Mining Techniques for Web Mining: A Review	Web Mining	• traditional data mining techniques is not practical because of the unstructured and the dynamic behavior of web data. • hybrid approach by combining social network analysis (web structure mining) with content mining would be more useful.	2015	R42
29	Zahra Zamani Alavijeh, Isfahan, Iran	The Application of Link Mining in Social Network Analysis	Link Mining	• focused on finding patterns in data by exploiting and explicitly modeling the links among the data instances.	2015	R 43
30	Pooja Rohill , Ochin Sharma	Web Content Mining: An Implementation on Social Websites	Web Mining Techniques	• used for extracting relevant information from various type of websites like online shopping sites. • Web Mining has a huge advantage to extract the data as per the user's criteria.	2015	R 48
31	Shilpa Radhakrishnan	A Survey on Text Filtering in Online Social Networks	Machine Learning Based Approach	• Support Vector Machine algorithm is mostly used for text classification. • The SVM algorithm can operate in large feature	2015	R 49

				sets .		
32	Gauri Joshi1, Mr. Samadhan Sonawane	Filtering and Classification Of User Based On Social Media Data Using Memetic and Naïve Bayes Methods	Naive Bayes Multi-Label Classifier algorithm and Memetic Algorithm	• to examining social medium statistics.	2015	R 50
33	Hilal Ahmad Khanday, Dr. Rana Hashmy	Exploring Different Aspects of Social Network Analysis Using Web Mining Techniques	Data Mining Techniques	• Lot of research needs to be done by comparing adjacent matrices and lists with incidence matrices using graphs obtained from online social networks as the input data.	2015	R 51
34	Pooja Sikka	DATA MINING OF SOCIAL NETWORKS USING CLUSTERING BASED-SVM	K-Means Clustering Based SVM (KMCB-SVM).	• SVM have not been favored for large data sets for mining • K-means micro-clustering technique will be implied with SVM.	2015	R 52
35	Faris Kateb, Jugal Kalita	Classifying Short Text in Social Media: Twitter as Case Study	Text Classification Techniques	• challenge in classifying social media text is that the data is streamed	2015	R 53
36	G. Besiashvili, T. Bliadze and Z. Kochladze	Application of Adaptive Neural Networks for the Filtration of Spam	Multi-layer Neural Network	• multilayer neural network learn to differentiate wanted and unwanted e-mails from one another	2015	R 46
37	Ritu Mewari, Ajit Singh, Akash Srivastava	Opinion Mining Techniques on Social Media Data	Opinion Mining	• extract pearl knowledge	2015	R 54
38	Remya R S, Smitha E S	Text Categorization using Data Mining Technique on Social Media Data	Naive Bayes Multi-Label Classification	• allowed one comment to fall into multiple categories at the same time. • beneficial in learning analytics, educational data mining, and learning technologies.	2015	R 55
39	Kanika Mathur	Online Social Network Mining	Naive Bayes Text Classifier	• it becomes more challenging when the textual information is not structured according to the grammatical convention.	2016	R 56
40	Rajeev Kumar, M.K. Sharma	Advanced Neuro-Fuzzy Approach for Social Media Mining Methods using Cloud	Hybrid Neuro-Fuzzy	• solve data mining problems	2016	R 58
41	R. Adaikkalam, Dr. A. Shaik Abdul Khadir	A Survey on Data Mining Techniques for Analysis of Social Network	Data Mining Techniques	• displays four features: structural intuition, systematic relational data, graphic images and mathematical or computational models.	2016	R 59
42	Thai Le, Phillip Pardo and William Claster	Application of Artificial Neural Network in Social Media Data Analysis: A Case of Lodging Business in Philadelphia	ANN, SOM	• ANN facilitated SOM is powerful • analyzing social media data in the field of tourism.	2016	R 60

43	Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, Sweta Tiwari	Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier	K-Nearest Neighbour(KNN) and Naïve Bayes	• in case of movie reviews Naïve Bayes' gave far better results than K-NN • for hotel reviews these algorithms gave lesser, almost same accuracies.	2016	R 61
44	Mukta Patkar, Pooja Pawar Mony Singh, Ashwini	A New way for Semi Supervised Learning Based on Data Mining for Product Reviews	Semi Supervised Approach	• sentiment analysis of product reviews gives positive and negative reviews • gives neutral and constructive opinion	2016	R 62
45	Dr. S. P. Victor, Mr. M. Xavier Rex	Analytical Implementation of Web Structure Mining Using Data Analysis in Educational Domain	Web Structure Mining	• identifying the specified URL structure content analysis. • identified the university web portal is more emphasized on educational links rather than with the individual college links.	2016	R63
46	Hemant Kumar Soni, Sanjiv Sharma, Pankaj K. Mishra	ASSOCIATION RULE MINING: A DATA PROFILING AND PROSPECTIVE APPROACH	Association Rule Mining	• there is a need to shift the paradigm form single objective to multi-objective association rule mining	2016	R64
47	Heling Jiang, An Yang , Fengyun Yan and Hong Miao	Research on Pattern Analysis and Data Classification Methodology for Data Mining and Knowledge Discovery	Pattern Analysis and Data Classification Methodology	• Analyze the organizational structure of network graphical pattern with the knowledge of machine learning methodology and graph theory.	2016	R65
48	Saravanan Suba and T. Christopher	AN EFFICIENT DATA MINING METHOD TO FIND FREQUENT ITEM SETS IN LARGE DATABASE USING TR- FCTM	TR-FCTM (Transaction Reduction- Frequency Count Table Method)	• finding direct frequency count and total frequency count for an itemsets.	2016	R66
49	V. A. Chakkarwar, Amruta A. Joshi	Semantic Web Mining using RDF Data	Semantic web mining	• minimizing extraction of number of pages by ranking technique • this research can be extended to sentence search as well as image search and video search.	2016	R67
50	Kuldeep Singh Rathore, Sanjiv Sharma	A Review on Web Usage Mining For Web Personalization Using Clustering Techniques	Clustering Techniques	• finds hidden information from large amount of log data, namely web log (also called as database)	2016	R68
51	Santosh C. Pawar, Ranjana S. Solapur	Research Issues and Future Directions in Web Mining: A Survey	Web Mining	• mining rules from semi structure and unstructured as in the semantic web becomes a great challenge.	2016	R 69

3. CHALLENGES:

SMA faces a number of challenges relating to the nature of social media data, their collection, and existing analysis and mining methods. First, social media data are generated in very large quantities and are highly dynamic and complex in their nature. Therefore, they cannot be processed easily using traditional data processing applications or database management tools as well as desktop statistics and visualization packages.

- Furthermore, they exhibit both structured and unstructured characteristics. While structured data (or meta-data) comprise user profile characteristics, spatial, temporal, and thematic data as well as attention related data (e.g., number of "likes", comments, retweets, mentions etc.), unstructured data include user-generated textual content ranging from relatively context-sparse microblogs through Facebook comments to context-rich blogs and audiovisual materials. This information overload represents a significant challenge that calls for large computing capacities and sophisticated sampling, extraction, and analysis methods.
- As boyd and Crawford (2012) point out, large data sets from Internet sources are often unreliable because of their potential incompleteness and inconsistency,
- Privacy issues are always present when data are collected. Researchers and other parties interested in gathering data may face questions such as whether it is ethical to collect, process, use, and report on social media data even if these are actually "public" in principle.
- E.Raju stated about how to use web mining techniques for social networks analysis. Data sampling is a big issue when using web mining for social networks analysis. In social networks analysis, it becomes a difficult task to select suitable samples representative of the real social networks.
- Other challenges include finding communities in social networks, finding patterns in social networks and analyzing overlapping communities. [19]

4. CONCLUSION AND FUTURE TRENDS

This paper provides a more current evaluation and update of social network analysis research available. Literatures have been reviewed based on different aspects of social network analysis. This paper studies the application of the techniques and concept of Web mining for social networks analysis, and reviews the related literature about Web mining and social networks. Social networks investigation carried out through the techniques of Web mining is an interesting field of research. However, there are many challenges in this research field to be resolve with improvement.

REFERENCES

[1] Sankar K. Pal, Varun Talwar, Pabitra Mitra, "Web Mining in Soft Computing Framework:Relevance, State of the Art and Future Directions", IEEE TRANSACTIONS ON NEURAL NETWORKS, VOL. 13, NO. 5, SEPTEMBER 2002.

[2] Tomoyuki NANNO, Toshiaki FUJIKI, "Automatically Collecting, Monitoring, and Mining Japanese Weblogs", WWW2004, May 17–22, 2004, New York, New York, USA.ACM1-58113-912-8/04/0005.

[3] Ralph Gross, Alessandro AcquistiH. John Heinz III, "Information Revelation and Privacy in Online Social Networks (The Facebook case)", Pre-proceedings version. ACM Workshop on Privacy in the Electronic Society (WPES), 2005.

[4] Huan Liu and Lei Yu, "Toward Integrating Feature Selection Algorithms for Classification and Clustering",IEEE Transactions on Knowledge and Data Engineering Volume 17 Issue 4,April 2005.

[5] Andrea Esuli_ and Fabrizio Sebastiani†, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining", Proceedings of the 5th Conference on Language Resources and Evaluation, 2006.

[6] Marcelo Maia, Jussara Almeida, Virgílio Almeida, "Identifying User Behavior in Online Social Networks", SocialNets'08, April 1, 2008 , Glasgow, Scotland, UK Copyright 2008 ACM ISBN 978-1-60558-124-8/08/04.

[7] http://www.openparenthesis.org/wp-content/uploads/2008/05/bcb3-retweeter.pdf

[8] Ai Ho, Abdou Maiga, Esma Aïmeur, "Privacy Protection Issues in Social Networking Sites", 978-1-4244-3806-8/09/$25.00 © 2009 IEEE.

[9] L. Dey and Sk. M. Haque, "Opinion mining from noisy text data," International Journal on Document Analysis and Recognition, vol. 12, no. 3, pp. 205–226, 2009.

[10] Mohammad Al-Fayoumi, Soumya Banerjee, Jr., and P. K. Mahanti, "Analysis of Social Network Using Clever Ant Colony Metaphor", PROCEEDINGS OF WORLD ACADEMY OF SCIENCE, ENGINEERING AND TECHNOLOGY VOLUME 41 MAY 2009 ISSN: 2070-3740.

[11] David Ediger Karl Jiang , Courtney Corley Rob Farber,William N. Reynolds, "Massive Social Network Analysis:Mining Twitter for Social Good", 2010 39th International Conference on Parallel Processing.

[12] Rojalina Priyadarshini; "Functional Analysis of Artificial Neural Network for Dataset Classification", Special Issue of IJCCT Vol. 1 Issue 2, 3, 4; 2010 for International Conference [ACCTA-2010], 3-5August 2010.

[13] A. Selman Bozkır1 , S. Güzin Mazman2 , and Ebru Akçapınar Sezer1, " Identification of User Patterns in Social Networks by Data Mining Techniques: Facebook Case", IMCW 2010, CCIS 96, pp. 145–153, 2010. © Springer-Verlag Berlin Heidelberg 2010.

[14] Alexander Pak, Patrick Paroubek, "Twitter as a Corpus for Sentiment Analysis and Opinion Mining", Proceedings of the International Conference on Language Resources and Evaluation, LREC 2010, 17-23 May 2010, Valletta, Malta.

[15] S. Baccianella, A. Esuli, and F. Sebastiani, "SentiWordNet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining," in Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC '10), pp. 2200–2204, 2010.

[16] Sitaram Asur, Bernardo A. Huberman, " Predicting the Future With Social Media", arXiv:1003.5699v1 [cs.CY] 29 Mar 2010.

[17] Kuan-Yu Lin, Hsi-Peng Lu, "Why people use social networking sites: An empirical study integrating network externalities and motivation theory", Computers in Human Behavior 27 (2011) 1152–1161.

[18] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," Computational Linguistics, vol. 37, no. 2, pp. 267–307, 2011.

[19] E.Raju K.Sravanthi, "Analysis of Social Networks Using the Techniques of Web Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012.

[20] G. Vinodhini and RM. Chandrashekharan, "Sentiment Analysis and Opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Volume2, Issue 6, June 2012.

[21] Fatemeh Khoshnood, Mehregan Mahdavi and Maedeh Kiani sarkaleh, "Designing A Recommender System Based On Social Networks And Location Based Services", International Journal of Managing Information Technology (IJMIT) Vol.4, No.4, November 2012.

[22] Xi Long, Wenjing Yin, Le An, Haiying Ni, Lixian Huang, Qi Luo, and Yan Chen, "Churn Analysis of Online Social Network Users Using Data Mining Techniques", International MultuConference of Engineers and Computer Scientists 2012 Vol 1,IMECS.

[23] Simona Vinerean1, Iuliana Cetina1, Luigi Dumitrescu2 & Mihai Tichindelean, "The Effects of Social Media Marketing on Online Consumer Behavior", International Journal of Business and Management; Vol. 8, No. 14; 2013 ISSN 1833-3850.

[24] Rupam Some, "A Survey on Social Network Analysis and its Future Trends", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 6, June 2013.

[25] M. K. Dalal and M. A. Zaveri, "Automatic classification of unstructured blog text," in Journal of Intelligent Learning Systems and Applications, vol. 5, no. 2, pp. 108–114, 2013.

[26] Paridhi Jainy, Ponnurangam Kumaraguruy, Anupam Joshi, "19wole01-Jin--@I seek 'fb.me':Identifying Users across Multiple Online Social Networks", 22th International World Wide Web Conference, 2013.

[27] Neha Mehta, Mamta Kathuria, Mahesh Singh, "Comparison of Conventional & Fuzzy Clustering Techniques: A Survey", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 1, April 2013.

[28] Farhat Roohi, "NEURO FUZZY APPROACH TO DATA CLUSTERING: A FRAMEWORK FOR ANALYSIS", European Scientific Journal March 2013 edition vol.9, No.9 ISSN: 1857 – 7881 (Print) e - ISSN 1857- 7431.

[29] G Nandi, A Das, "A Survey on Using Data Mining Techniques for Online Social Network Analysis", IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 6, No 2, November 2013.

[30] Mita K. Dalal and Mukesh A. Zaveri, "Semisupervised Learning Based Opinion Summarization and Classification for Online Product Reviews", Hindawi Publishing Corporation Applied Computational Intelligence and Soft Computing Volume 2013.

[31] S.G.S Fernando, "Empirical Analysis of Data Mining Techniques for Social Network Websites", COMPUSOFT, An international journal of advanced computer technology, 3 (2), February-2014 (Volume-III, Issue-II).

[32] Anaïs Collomb , Crina Costea , Damien Joyeux , Omar Hasan , Lionel Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Laboratoire d'InfoRmatique en Image et Systèmes d'information,2014.

[33] Sanjeev Dhawan, Kulvinder Singh, Vandana Khanchi, "Critical Analysis of Social Networks with Web Data Mining", IJITKM Special Issue (ICFTEM-2014) May 2014 pp. 107-111 (ISSN 0973-4414).

[34] M. Vedanayaki, "A Study of Data Mining and Social Network Analysis", Indian Journal of Science and Technology, Vol 7(S7), 185–187, November 2014.

[35] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber and Frederic Stahl, "A Survey of Data Mining Techniques for Social Network Analysis", Journal of Data Mining & Digital Humanities, 2014.

[36] Asad Bukhari, Um-e-Ghazia and Dr. Usman Qamar, "CRITICAL REVIEW OF SENTIMENT ANALYSIS TECHNIQUES", Proceeding of the International Conference on Artificial Intelligence and Computer Science (AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA.

[37] Meenu Sharma, "Clustering In Data Mining : A Brief Review", International Journal Of Core Engineering & Management (IJCEM) Volume 1, Issue 5, August 2014.

[38] Md. Ansarul Haque1, Tamjid Rahman , "SENTIMENT ANALYSIS BY USING FUZZY LOGIC", International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol. 4,No. 1, February 2014.

[39] Y. K. Mathur, Abhaya Nand , "Soft Computing Techniques and its Impact in Data Mining", International Journal of Emerging Technology and Advanced Engineering (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 4, Issue 8, August 2014).

[40] Sanjeev Dhawan, Kulvinder Singh, Deepika Sehrawat, "Emotion Mining Techniques in Social Networking Sites", International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 12 (2014).

[41] Haseena Rahmath P, Tanvir Ahmad, "Fuzzy based Sentiment Analysis of Online Product Reviews using Machine Learning Techniques", International Journal of Computer Applications (0975 – 8887) Volume 99 – No.17, August 2014.

[42] Esmaeil Fakhimi gheshlagh mohammad beig, "Data Mining Techniques for Web Mining: A Review", Applied mathematics in Engineering, Management and Technology 3(5) 2015:81-90.

[43] Zahra Zamani Alavijeh, Isfahan, Iran, "The Application of Link Mining in Social Network Analysis", ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 3, No.15 , May 2015.

[44] Thabit Zatari, "Data Mining in Social Media",International Journal of Scientific & Engineering Research, Volume 6, Issue 7, July-2015.

[45] Zohreh Madhoushi, Abdul Razak Hamdan and Suhaila Zainudin, "Sentiment Analysis Techniques in Recent Works", Science and Information Conference (SAI), 2015. IEEE, 2015.

[46] G. Besiashvili, T. Bliadze and Z. Kochladze1, "Application of Adaptive Neural Networks for the Filtration of Spam", EPiC Series in Computer Science Volume 36, 2015, Pages 42–50 GCAI 2015.

[47] Smita Bhanap, Dr. Seema Kawthekar, "Data Mining for Business Intelligence in Social Network: A survey", International Advanced Research Journal in Science, Engineering and Technology Vol. 2, Issue 12, December 2015.

[48] Pooja Rohilla, Ochin Sharma, "Web Content Mining: An Implementation on Social Websites", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 7, July 2015.

[49] Shilpa Radhakrishnan, "A Survey on Text Filtering in Online Social Networks", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (4) , 2015, 3874-3876.\

[50] Gauri Joshi, Mr. Samadhan Sonawane, "Filtering and Classification Of User Based On Social Media Data Using Memetic and Naïve Bayes Methods", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4795-4798.

[51] Hilal Ahmad Khanday, Dr. Rana Hashmy, "Exploring Different Aspects of Social Network Analysis Using Web Mining Techniques", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 4, Issue 2, March-April 2015.

[52] Pooja Sikka, "Data Mining Of Social Networks Using Clustering Based-Svm", Ijrrest International Journal Of Research Review In Engineering Science & Technology (ISSN 2278–6643) VOLUME-4, ISSUE-1, APRIL-2015.

[53] Faris Kateb,Jugal Kalita, "Classifying Short Text in Social Media: Twitter as Case Study", International Journal of Computer Applications (0975 8887) Volume 111 - No. 9, February 2015.

[54] Ritu,Ajit Singh,Akash Srivastava, "Opinion Mining Techniques on Social Media Data", International Journal of Computer Applications (0975 – 8887) Volume 118 – No. 6, May 2015.

[55] Remya R S, Smitha E S, "Text Categorization using Data Mining Technique on Social Media Data", International Journal of Advanced Research in Education & Technology (IJARET), Vol. 2, Issue 4 (Oct. - Dec. 2015)

[56] Kanika Mathur, "Online Social Network Mining", International Journal of Computer Trends and Technology (IJCTT) – Volume 35 Number 4 - May 2016.

[57] Jyoti Shokeen, Partibha Yadav, "Overview of Social Network Analysis and Tools",August 2016, Volume 3, Issue 8 JETIR (ISSN-2349-5162).

[58] Rajeev Kumar, M.K. Sharma, "Advanced Neuro-Fuzzy Approach for Social Media Mining Methods using Cloud", International Journal of Computer Applications (0975 – 8887) Volume 137 – No.10, March 2016.

[59] R. Adaikkalam, Dr. A. Shaik Abdul Khadir, "Data Mining Techniques for Analysis of Social Network", Volume 4, Issue 3, March 2016 International Journal of Advance Research in Computer Science and Management Studies.

[60] Thai Le, Phillip Pardo and William Claster , "Application of Artificial Neural Network in Social Media Data Analysis: A Case of Lodging Business in Philadelphia", © Springer International Publishing Switzerland 2016.

[61] Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, Sweta Tiwari, " Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier", I.J. Information Engineering and Electronic Business, 2016, 4, 54-62.

[62] Mukta Patkar, Pooja Pawar Mony Singh, Ashwini Save, "A New way for Semi Supervised Learning Based on Data Mining for Product Reviews", 2nd IEEE International Conference on Engineering and Technology (ICETECH), 17th & 18th March 2016, Coimbatore, TN, India.

[63] Dr. S. P. Victor, Mr. M. Xavier Rex, "Analytical Implementation of Web Structure Mining Using Data Analysis in Educational Domain", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016)

[64] Hemant Kumar Soni1, Sanjiv Sharma2, Pankaj K. Mishra3, "ASSOCIATION RULE MINING: A DATA PROFILING AND PROSPECTIVE APPROACH", International Conference on Futuristic Trends in Engineering, Science, Humanities, and Technology (FTESHT-16) ISBN: 978-93-85225-55-0, January 23-24, 2016, Gwalior.

[65] Heling Jiang1,2, An Yang1 , Fengyun Yan1 and Hong Miao1, "Research on Pattern Analysis and Data Classification Methodology for Data Mining and Knowledge Discovery", International Journal of Hybrid Information Technology Vol.9, No.3 (2016)

[66] Saravanan Suba1 and T. Christopher2, "An Efficient Data Mining Method To Find Frequent Item Sets In Large Database Using TR- FCTM" ICTACT Journal On Soft Computing, January 2016, Volume: 06, Issue: 02

[67] V. A. Chakkarwar, Amruta A. Joshi, "Semantic Web Mining using RDF Data", International Journal of Computer Applications (0975 – 8887) Volume 133 – No.10, January 2016.

[68] Kuldeep Singh Rathore, Sanjiv Sharma, "A Review on Web Usage Mining For Web Personalization Using Clustering Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 5, May 2016.

[69] Santosh C. Pawar, Ranjana S. Solapur, "Research Issues and Future Directions in Web Mining: A Survey", International Journal of Computer Applications (0975 – 8887) National Seminar on Recent Trends in Data Mining (RTDM 2016)