A Survey of Data Mining Techniques for Social Media Analysis by UsingClustering Algorithms

Saumya Agarwal[1], Anu Sharma[2]
M.Tech Scholar[1], Assistant Professor[2]
Teerthanker Mahaveer University, Moradabad (U.P.), India[1], TeerthankerMahaveer University,
saumya295991@gmail.com, anu.computers@tmu.ac.in

Abstract--**Data mining is the extraction of projecting information from large data sets. Data mining separate the knowledge in to a form which is very much useful for many real world applications. Social Media in the last decade has gained remarkable attention.It is important to translate sentiment expressed by SM users to useful information using data mining techniques. Data mining techniques are capable of handling the three dominant research issues with SM data which are size, noise and dynamism.Clustering is the task of assigning a set of objects into groups called clusters. Main task of clustering are explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis and information retrieval. Cluster analysis is not an automatic task, but an iterative process of knowledge discovery[2].**
In this paper we discuss about data mining techniques, a survey of the works done in the field of social network analysis and Clustering Algorithms.

Keywords--**Data Mining, Social media, Clustering Algorithms, Survey, Analysis, Techniques**

I. INTRODUCTION

Social Mediais an important source of learning of opinions, sentiments, approaches, evaluation, influences, observations, feelings, reviews, blogs, discussions, news, remarks, reactions, or some other documents [3]. A social network community represents people and connects them. It has provided a way of keeping in touch with friends, create personal profiles, view others profiles, communicate and share personal information. Members of social network communities manage their identity through their profile, they meet new friends and like-minded people in the community, they connect with each other, ask questions, get answers and discuss topics. Facebook, twitter, linkedin and instagram in fig 1. [3] can be taken as most commonly accessed social network sites [4].Social media mining is a process involving the extraction, analysis and representation of useful patterns from data in the social media.

Social media mining is a young field which has been leading research and development by handling enormous amounts of information [5]. Among the information techniques that can be used for the analysis of social networks, Data mining is claimed to be the most suitable one. Therefore, it is more suitable to use the data mining techniques for social networks analysis, and it is also the focus of this paper.

Fig.1 Common Social Media

II. LITERATURE REVIEW

A. *Social Media Analysis*

According to Technorati, about 75,000 new blogs and 1.2 million new posts giving opinion on products and services are generated every day. Massive data are generated every minute on different

common *SM* sites as revealed in Fig.2 [6]. In view of the foregoing, it is necessary to employ tools capable of analysing SMespecially the expression of opinions/sentiments which are main characteristics of *SM*. Data miningtechniques has shown to be capable of mining big data generated on *SM* sites. This is made possible by way of extracting information from large data set generated on *SM* and transforming them into understandable structure for further use.

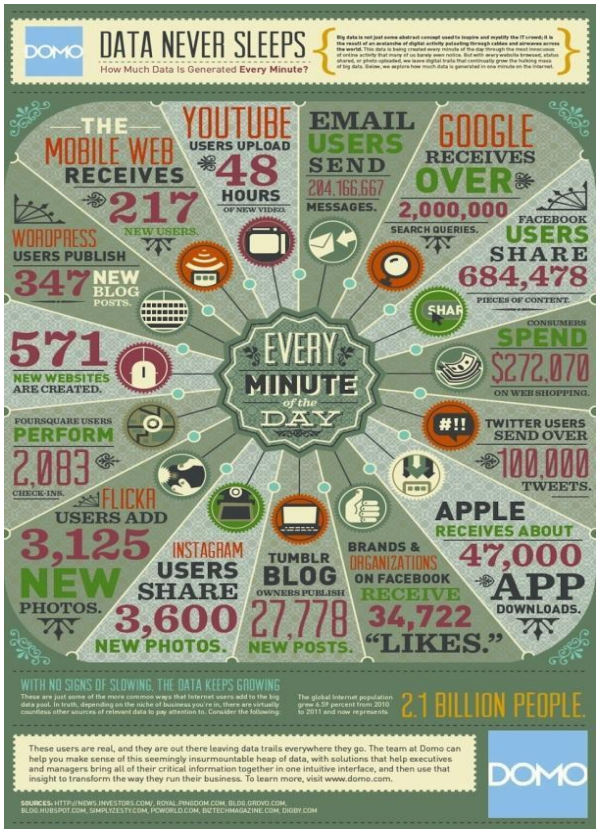

Fig.2. Estimated Data Generated on SM Site Every Minute

1).*Sentiment Analysis on Social Media:*SMhas an unimaginable power on its user by way of expressing their views and sentiments on *SM* sites, be it positive or negative [7]. (e.g. *"The phone does not come in my favourite colour, therefore it is a waste of money"*). It has become necessary to analyse sentiment expressed by *SM* users using data mining techniques in order to generate a meaningful framework [3].
Having given an overview of sentiment analysis on social network, an
overview of some of the data mining tools used for sentiment analysis onsocial network are discussed in Data Mining section of the survey.

B. Data Mining

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 [8]. Zahra ZamaniAlavijeh stated that link mining is becoming a very popular research area for social network analysis [11].PoojaRohilla stated that tools available on the internet mines the data either in semi structured or unstructured format. To retrieve this type of data, we need to define proper patterns and clustering [13].

There are various data mining techniques [1]:

1).*Characterization*:Characterization is used to generalize, summarize and possibly different data characteristics.

2). *Classification:*Data classification is a process in which the given data is classified in to different classes according to a classification model.

3).*Regression:*This process is similar to classification the major difference is that the object to be predicted is continuous rather than discrete.

4).*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.

5).*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.

C. Analysis of Data Mining

In social network data mining, existing data mining algorithms cannot be used directly because of the dynamic behaviour. When analyzing the literature on social network data mining techniques; it was found that each algorithm has strengths and weaknesses [4].

1).*Graph Mining Algorithm*:Most popular data mining technique in Social Network Analysis is using Graph mining algorithms. World Wide Web including social networks is a collection of interconnected hypertext documents. These are interconnected by hyperlinks. So we can be considered as a directed graph, where nodes will be the hypertext documents and edges will be hyperlinks.
The Algorithm is based on frequent pattern mining in transactional and graph databases with periodic pattern mining in unidimensional and multidimensional sequences.

2).*Classification:*Most popular classification algorithms in data mining are decision trees, naïve Bayesian classifier and neural networks. Surma and Furmanek (2010) introduced an interesting algorithm called C&RT, combining classification and regression tree algorithms to determine rules to identify target groups to market. This can be used in real social network data.

3).*Clustering*: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. PoojaSikka proposed SVM technique [1]. It is also known as K- Means Clustering Based SVM (KMCB-SVM). The complexity of SVM depends on no. of input variables and support vectors. It scans entire data to provide the 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 [9]. Thai Le, Phillip Pardo and William Claster stated Artificial Neural Network (ANN) which cluster all the textual conversational topics being shared through thousands of management tweets.ANN gives knowledge in the field of tourism [10].
Clustering is mainly used in information retrieval. Graph based clustering is commonly used in structure mining. Text based clustering is most commonly used in web content mining whether you create clusters based on the content of the web document. Bartal et.al (2009) introduced an interesting method combining social network analysis and text based clustering to predict the nodes of a social network would be linked next.

4).*Associations:* Association rule mining is used to find frequent patterns and correlation among data set. Nancy et.al (2013) had use association rules to mine social network data using 100 Facebook university pages. The research focused on the formulation of association rules using which decisions can be made and uses Apriori Algorithm to derive association rules.

Table 3.1
[Literature Review Summary Table]

#	**Author's Name**	**Paper Title**	**Techniques**	**Findings**	**Year**
1.	M. Vedanayaki	A Study of Data Mining and Social Network Analysis	Knowledge Based Network Analysis	Focus on identifying patterns.	2014
2.	Mariam Adedoyin-Olowe1	A Survey of Data Mining Techniques for Social Media Analysis	Sentiment Analysis	Focus on finding opinions/Sentiments	2014
3.	Zahra ZamaniAlavijeh	The Application of Link Mining in Social Network Analysis	Link Mining	Focus on finding patterns by modelling the link among data sets	2015

4.	PoojaSikka	Data Mining Of Social Networks Using Clustering Based-Svm	K- Means Clustering Based SVM (KMCB-SVM)	SVM is not for mining large data sets. K- Means clustering used with SVM	2015
5.	Thai Le, Phillip Pardo and William Claster	Application of Artificial Neural Network in Social Media Data Analysis	Artificial Neural Network (ANN)	Gives knowledge in the field of tourism	2016
6.	R.Adaikkalam and Dr. A. Shaik Abdul Khadir	A Survey on Data Mining Techniques for Analysis of Social Network	Data Mining Technique	Display Graph Mining Algorithm	2016

D. Analysis of Clustering Algorithms

III. Result & Conclusion

This paper provides overview of Data Mining techniques for social Media by using Clustering Algorithm. Literatures have been reviewed based on different aspects of social network analysis. This paper studies the application of the techniques for social networks analysis.

References

[1] Anu Sharma, Dr. M.K Sharma &Dr.R.KDwivedi, "Literature Review and Challenges of Data Mining Techniques for Social Network Analysis", ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354 Advances in Computational Sciences and Technology. Anoop Kumar Jain and Satyam Maheswari, "Survey of Recent Clustering Techniques in Data Mining**",** Volume 3

[2] June 2012: 68 – 75 ISSN: 0976-4828 International Archive of Applied Sciences and Technology.

[3] Mariam Adedoyin-Olowe1, Mohamed Medhat Gaber1 and Frederic Stahl, "A Survey of Data Mining Techniques for SocialMedia Analysis**".**

[4] R.Adaikkalam and Dr. A. Shaik Abdul Khadir, "A Survey on Data Mining Techniques for Analysis of Social Network*",* ISSN: 2321-7782 Volume 4, Issue 3, March 2016 International Journal of Advance Research in Computer Science and Management Studies.

[5] ThabitZatari, "Data Mining in Social Media", International Journal of Scientific & Engineering Research, Volume 6, Issue 7, July-2015 ISSN 2229-5518

[6] Tepper, A.: How Much Data Is Created Every Minute? [INFOGRAPHIC]. 2012, http://mashable.com/2012/06/22/datacreated-every-minute/. Retrieved on 16/102013 at 19.00.

[7] Aggarwal, C.: An introduction to social network data analytics. Springer US, 2011.

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

[9] PoojaSikka, "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.

[10] 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.

[11] Zahra ZamaniAlavijeh, 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.

[12] PoojaRohilla, 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.