

DATA MINING TECHNIQUES FOR SOCIAL NETWORK SITES

NANDINI SHRIVASTAV¹, Ms. ANU SHARMA²

¹*M.Tech Scholar(CCSIT), TMU, Moradabad, India*

²*Asstt. Professor(CCSIT),TMU, Moradabad, India*

nandinishrivastava7@gmail.com

anu.computers@tmu.ac.in

Abstract- Data mining is the process of analyzing data from different view and sum up it into useful information. Data mining is one of a analytical tools for examine data. The overall goal of the data mining process is to extract information from a data set and convert it into an understandable structure. Social network analysis is a very common field of research as it is much effective for different applications. In this paper we have summarize different data mining techniques used for social network analysis.

Keywords- Social Network, Data Mining, Social Network Analysis.

I. INTRODUCTION

Data mining is sorting through data to identify patterns and establish relationships. It is the process of analyzing data from different view and extracts useful information that can be used to increase revenue, cuts, costs or both. It is the process of finding correlations among dozens of fields in large relational databases^[1].

A social network is an umbrella with nodes of individuals, groups, organizations and related systems that tie in one or more types of interdependency. Social network analysis is focused on detecting the patterning of people's interaction on them^[2-4].

II. LITERATURE REVIEW

In^[5] the author has studied on the issues related to social networks analysis using web mining techniques. And also studied the application of web mining techniques and discussed a general process for social

networks analysis. The paper also highpoints the problems faced in selecting data samples and patterns in social networks and analyzing overlapping communities.

In^[6] the author has studied on a systematical data mining approach to mine intellectual knowledge from social data. The author took Face book as a primary data source and use different data mining techniques to analyze this social networking site and other sites too. An algorithm that the author had discussed was K-nearest neighbor (K-NN). This algorithm characterizes objects based on samples.

In^[7] the author has discussed the application of correlation, clustering, and association analyses to social media. The main focus of this paper was to describe how data mining and text analytics can be applied to social media in order to identify key themes in the data. To describe this more the author took the Twitter posts. Certain issues in terms of accuracy while collecting the data from social media were also highpointed.

In^[8] the authors have studied on the techniques that are currently used to analyze SM. Also, they have accounted those techniques that can be considered further in this field. In this paper the analysis of SM

data has proved to be effective, this is so because of the capacity data mining possess in handling noisy, large and dynamic data. According to the authors, in future to mine the data generated on SM currently used and yet-to-be-explored data mining techniques will be used.

In [9] the authors explain how Social networking site helps a company and its brand to convey messages to massive amount of people very easily. The aim of the authors was to cover up protection strategies for business infrastructure. The authors also discussed about few proactive defensive strategies against the malicious users who could leak the information of a company and get access to the messages that were to be conveyed. The strategies that were mainly discussed included awareness, monitoring and logging, acceptable use policy (AUP) and enforcement.

III. DATA MINING TECHNIQUES

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 suitable to use the data mining techniques for social networks analysis [10], and it is also the focus of this paper. So the data mining techniques are -

A. *Characterization*

Data characterization is used to discover and analyze different data characteristics.

B. *Classification*

Data classification is used to classify the given data into different classes according to a classification model.

C. *Regression*

Regression is similar to classification. The major difference between them is that the object to be predicted is continuous in regression rather than discrete.

D. *Association*

In this process the association or

combination between the objects is found. It discovers the union between different data bases and the association between the attributes of a 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 separates the important changes in the data from the previously measured values.

G. *Deviation Detection*

Deviation detection focuses on the main difference between the actual values objects and expected values of the object. This method finds out the deviation according to the time and also 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 patterns in the data which occurs frequently.

III. SOCIAL NETWORK

Social networks are defined as virtual spaces where people of different ages can create contacts, share ideas, and build a sense of community. It provides a way of keeping in touch with friends, producing personal profiles, view other's profiles and connections, communicate and share personal information.

Members of social network communities take care of their identity through their profile and make new friends. Minded people in the community, connect with

each other and ask questions, get answers and discuss topics.

Face book, twitter, MySpace are the most commonly accessed social network sites. Social networks can also be used in many business activities like marketing research, General marketing, Idea generation & new product development, Customer service, Public relations.

IV. SOCIAL NETWORK ANALYSIS AND DATA MINING

Data mining tools can solve industry questions in less time which were too much time consuming to solve generally. Data mining of social networks can be done using the graph mining methods such as classification/topologies, pattern detection, measurement and metrics, modeling, evolution and structure, data processing, and communities [8].

To abstract the information represented in graphs we need to define metrics that describe the global structure of graphs, then find out the community structure of the network, and define metrics that describe the patterns of local interaction in the graphs, develop efficient algorithms for mining data on networks, and understand the model of graphs generation.

Social network and its analysis is an important field and it is widely spread among young researchers. Social networks research emerged from psychology, sociology, statistics and graph theory. Based on graph theoretical concepts a social networks interprets the social relationships of individuals as points and their relationships as the lines connecting them [11]. The various types of social network analysis are- *Socio Centric (whole) Network Analysis*

- Emerged in sociology.
- Involves quantification of interaction among a socially well defined group of people.
- Focus on identifying global structural patterns.

- Make generalizations of features found in personal networks.
- Difficult to collect data, so till now studies have been rare.

Knowledge Based Network Analysis

- Emerged in Computer Science.
- Involves quantification of interaction between individuals, groups and other entities.
- Discovery based Knowledge on entities associated with actors in the social network.

V. APPLICATIONS

- Viral marketing
- Social analytics
- Expert finding
- Image Analysis
- Fraud detection

VI. CONCLUSION

The rise of social networks gives very strong effects to the set of techniques developed for mining graphs and social networks. Social networks are rooted in many sources of data. Data Mining provides skillful way to execute and make use of database.

In this paper we have briefly reviewed the various data mining techniques which are used for social network analysis and its applications.

REFERENCE

- [1] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, "From data mining to knowledge discovery in databases," AI magazine, vol. 17, no. 3, p. 37, 1996.
- [2] Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences). Cambridge University Press, Cambridge (1994)
- [3] Schtt, J.: Social Network Analysis: A Handbook. Sage Publications, Newbury Park (1991)
- [4] Hanneman Robert, A., Mark, R.: Introduction to Social Network Methods. University of California, Riverside, Riverside (2005)
- [6] Raju, E., & Sravanthi, K, Analysis of Social Networks Using the Techniques of Web Mining. International Journal, 2012.
- [7] Rahman, M. M, Mining Social Data to Extract Intellectual Knowledge. arXiv preprint arXiv: 1209.5345, 2012.
- [8] Mosley Jr, R. C, Social media analytics: Data mining applied to insurance Twitter posts. In Casualty Actuarial Society E-Forum, Winter 2012, vol 2 (p. 1).

- [9] Adedoyin-Olowe, M., Gaber, M. M., & Stahl, F, A Survey of Data Mining Techniques for Social Media Analysis.
- [10] Bassett, R., Chamberlain, T., Cunningham, S., & Vidmar, G. Data mining and social networking sites: protecting business infrastructure and beyond. *Issues in Information Systems*, 11(1), 353, 2010.
- [11] Han J, Kamber M. *Data Mining: Concept and Techniques*, 2nd Edition. Morgan Kaufmann; 2006.
- [12] Mika P. *Social Networks and the Semantic Web*, 1st edition. Springer; 2007.