TRENDS AND TECHNIQUES IN DATA MINING

Amrita Kumari[1], Anu Sharma[2]
[1]STUDENT, [2]ASSISTANT PROFESSOR

FoE& CS, Computer science Department, Teerthanker Mahaveer University, Moradabad, 244001

Abstract— Data mining is the process of discovering of projecting information from large data sets. It is a collaborative subfield of Computer Science with an overall goal to essence information from a data set and transfer the information into the coherent structure for further use. Websites contains billions of unused raw data. By considering this data new knowledge can be achieved. Since this data is changing and unstructured traditional data mining techniques will not be appropriate.
In this paper we will discuss about trends and techniques in data mining. We will also discuss about works done in the field of social network analysis.

Keywords— Data mining techniques, Social network analysis, Web data mining.

1. Introduction

Data mining is the compelling tool that can help to find patterns and relationship within our data. It discovers hidden information from large databases. The generic goal of the data mining process is to extract information from a dataset and transform into an understandable structure for further use.[7]

Data mining involves following classes of tasks:

a. Anomaly detection
b. Association rule learning
c. Clustering
d. Classification
e. Regression
f. Summarization

There are following data mining techniques:

a. Characterization- It is used to summarize and possibly differentiate data characterstics.

b. Classification- It is the process in which given data is classified into different classes according to classification model.

c. Regression
d. Association
e. Clustering
f. Change detection
g. Deviation detection
h. Link analysis
i. Sequential pattern mining

2. LITERATURE REVIEW:

Anu Sharma [7] suggested that 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, while Twitter now has more than 280 million monthly active users. 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. Recent studies and surveys have revealed an emerging need to continuously collect, monitor, 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. 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 literature. Tomoyuki NANNO[2] 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.

SI.	AUTHOR'S NAME	PAPER TITLE	TECHNIQUES	FINDINGS	YEAR
1	ANU SHARMA	Literature review and challenges of data mining techniques for social network analysis	Clustering	Different interaction pattern can be observed.	2017
2	S.G.S Fernando	Empirical analysis of data mining techniques for social network websites.	Markov models	Hybrid approach by combining social network analysis with content mining would be more useful.	2014
3	Sanjeev Dhawan, Kulvinder Singh, Vandana khanchi	Critical analysis of social network with web data mining.	Web mining techniques	How to utilize the web mining techniques to some real online social networking websites.	2014

4	M.Vedanayaki	A study of data mining and social networking analysis	Knowledge based network analysis	Difficult to collect data	2014
5	Meenu Sharma	Clustering In Data mining: A brief review	Neuro and fuzzy logic approaches	FCM methods responds better in the real-life situations	2014
6	Santosh C. Pawar, Ranjana S. Solanki	Research issues and future directions in web mining: A survey	Web mining	Mining rules from semi-structured as in the semantic web becomes a great challenge	2016
7	Pooja Sikka	DATA MINING OF SOCIAL NETWORKS USING CLUSTERING BASED-SVM	K-Means Clustering Based SVM (KMCBSVM).	SVM have not been favoured for large data sets for mining • K-means micro-clustering technique will be implied with SVM.	2015

3. CONCLUSION

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.

REFERENCES

1. Sankar K. Pal, Varun Talwar, Pavitra 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 Acquisto. John Heinz III, "Information Revelation and Privacy in Online Social Networks (The Facebook case)", Preproceedings 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 Fabrizio"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. Anu Sharma, "Literature review and challenges of data mining techniques for social network analysis" , Advances in Computational Sciences and Technology ISSN 0973-6107 Volume 10, Number 5 (2017) pp. 1337-1354