Application of Temporal Data Mining Techniques for Social Network Analysis

Shashi Bhushan

Department of Computer application
Gaya College, Magadh University, Bodhgaya

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. This paper presents study about social networks using Temporal mining techniques.

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

1. INTRODUCTION

Nowadays, a large quantity of data is being accumulated. Usually there is a huge gap from the stored data to the knowledge that could be construed from the data. This transition won't occur automatically, that's where Data Mining comes into picture. In Exploratory Data Analysis, some initial knowledge is known about the data, but Data Mining could help in a more in-depth knowledge about the data. Seeking knowledge from massive data is one of the most desired attributes of Data Mining. Manual data analysis has been around for some time now, but it creates a bottleneck for large data analysis. Fast developing computer science and engineering techniques and methodology generates new demands to mine complex types of data. A number of Data Mining techniques (such as association, clustering, classification) are developed to mine this vast amount of data. 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 [1]. 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 [2].

There are various data mining techniques:

- **Characterization:** Characterization is used to generalize, summarize and possibly different data characteristics.
- **Classification:** Data classification is a process in which the given data is classified in to different classes according to a classification model.
- **Regression:** This process is similar to classification the major difference is that the object to be predicted is continuous rather than discrete.
- **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.
- **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.
- **Change Detection:** This method identifies the significant changes in the data from the previously measured values.
- **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.
- **Link Analysis:** It traces the connections between the objects to develop models based on the patterns in the relationships by applying graph theory techniques.
- **Sequential Pattern Mining:** This method involves the discovery of the frequently occurring patterns in the data. [1]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 [3]

2. Application of Temporal Data Mining Techniques

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 (Global Web- 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 presents 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 [4] 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 [5]. 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 [6].

2.1. Temporal mining Methods and Soft Computing Approaches

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. 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 [7]. 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. Y. K. Mathur¹ stated those Soft Computing techniques and its impact as well as its new emerging trends to suit the changing requirements in the area of Data Mining [3]. 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. Zahra Zamani Alavijeh stated that link mining is becoming a very popular research area not only for data mining and temporal mining but also in the field of social network analysis [8]. 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 [9]. 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. Temporal mining techniques are used for extracting relevant information from various type of websites like online shopping sites. Ritu Mewari presented an opinion mining provides a clear platform to catch public's mood by filtering the noise data [10]. 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 temporal 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. 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 [11].

2.2. Clustering Approach in Temporal Mining Applications

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. Pooja Sikka proposed SVM technique. This new technique is named as K-Means Clustering Based SVM (KMCB-SVM) [12]. 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 Cluster 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].

3. CONCLUSION AND FUTURE TRENDS

This paper provides a more current evaluation and update of temporal data mining techniques in 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 temporal mining for social networks analysis, and reviews the related literature about temporal mining and social networks. Social networks investigation carried out through the techniques of temporal mining is an interesting field of research. However, there are many challenges in this research field to be resolve with improvement.

REFERENCES

1. M. Vedanayaki, "A Study of Data Mining and Social Network Analysis", Indian Journal of Science and Technology, Vol 7(S7), 185–187, November 2014.
2. Esmaeil Fakhimi gheslagh mohammad beig, "Data Mining Techniques for Web Mining: A Review", Applied mathematics in Engineering, Management and Technology 3(5) 2015:81-90.
3. 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.
4. 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.
5. 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.
6. Jyoti Shokeen, Partibha Yadav, "Overview of Social Network Analysis and Tools",August 2016, Volume 3, Issue 8 JETIR (ISSN-2349-5162).
7. 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.

8. 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).
9. 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.
10. 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.
11. 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.
12. 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.