Data Mining for Business Intelligence in Social Network: A survey

Mrs. Smita Bhanap[1], Dr. Seema Kawthekar[2]

Assistant Professor, Fergusson College, Pune, India[1]

Associate Professor, Department of Computer Science & IT,

Dr. Babasaheb Ambedkar Marathwada University Aurangabad, India[2]

Abstract: In the era of digital world, due to the revolution in the field of Information and Communication Technology, Social Networks are providing an accessible platform for users to share and use information in almost all sectors of the society. As a result of this , pervasive usage of social network generates huge amount of social data. Mining this social data has its potential to extract actionable patterns, behave ors which are beneficial business organizations and customers. Different mining tools and techniques can be used to achieve better decision making in the field of Electronic Business Intelligence. This paper reviews basics of Data Mining, Social Network Analysis and its applications in Business Intelligence with data mining techniques and suggest how this survey and study of the data mining approaches can benefit the importance of social network analysis and mining for business intelligence.

Keywords: Social network, social data, business intelligence, data mining, data mining techniques.

I. INTRODUCTION

The term Business intelligence became a popular term in the business and IT communities only in the 1990s. In the late 2000s, business analytics was introduced to represent the key analytical component in BI. Business Intelligence is the ability to dig into data that has been acquired through different channels/sources , business activities and then identify trends , opportunities and areas where efficiency can be improved /(or for better quality improvement.)[1] The Information and Communication Technologies revolution provided us with convenience and ease of access to information, mobile communications and even possible contribution to this amount of information. Moreover, the need of information from these vast amounts of data is even more pressing for enterprises. It is becoming more challenging to handle huge volume of data coming from various sectors. This is where BI comes in.BI compiles , parses , interprets data into useful reports, patterns to be used by business people.[2]

Social Networks are structures consisting of members (represented as nodes on a network graph) that share one or more specific types of special interests, such as values, visions, ideas, financial, exchange, friends, like, dislike, conflict, trade, web links, and so forth (represented as links on a network graph). (Westland, 2009)Online social networks are built on the concept of tradition networks, but without relying on the face-to-face initial. Boyd and Ellison (2007) clearly define social network sites as web based services that allow people: firstly, construct a public or semi-public profile within a bounded system; secondly, articulate a list of other users with whom they share a connection; and thirdly, view and traverse their list of connections and those made by others within the system.[3]

Motivation : Social Network and Business Intelligence are now inseperable. Various social networks along with Business Intelligence allows to reach a huge number of customers/users at a time resulting into variety of data with hidden knowledge. Also, advertising, marketing, cost saving can be achieved .The growing nature and use of current social networks are generating highly dynamic data with increasing volume which is structured , Semistructured or unstructured data. Moreover, the data is not directly translatable into network-based data that would be useful. This is where the ability to perform effective data mining becomes important. Mining information from raw data is an extremely vital and tedious process due to dynamic nature of social networks.[4]

 In order to study the various aspects of social networks it is now required to design some good business model frameworks by using advanced data mining tools and techniques. Analyzing the social network data for information mining has the potential of revealing information of great value. Social network analysis provides a systematic method to identify, examine, visualize and support processes of knowledge sharing in social networks. Social network analysis can help in determining various queries such as how the information flows, important members or links and network dynamics.[3]

 Data Mining (DM) is defined as the process of analyzing large databases, usually data warehouses or internet, to discover new information, hidden patterns and behaviors. It's an automated process of analyzing huge amounts of data to discover hidden traits, patterns and to predict future trends and forecast possible opportunities. One of the main uses of DM is business intelligence and risk management [1]. Enterprises must make business critical decisions based on large datasets stored in their databases, DM directly affect decision-making.

A typical data mining process starts with identifying a data mining problem depending on the goals and interest of a

data analyst. Next, all sources of information are identified and a subset of data is generated from the accumulated data for the data mining application. To ensure quality, the data set is preprocessed by removing noise, handling missing information and transforming to an appropriate format. A data mining technique or a combination of techniques appropriate for the type of knowledge to be discovered is then applied to the derived data set. The discovered knowledge is then evaluated and interpreted, typically involving some visualization techniques. Finally the information is presented to user to incorporate into the company's business strategies.[5]

For example, data mining techniques can identify user sentiments to develop suggestion systems for business of specific products and even to build new friendships or connect certain interest groups.(Facebook uses likes, groups as well as posts of users to recommend users specific ads as well as new pages and groups.

Data Mining Techniques :

Predictive Tasks :
* **Classification**
* **Regression**
* **Deviation Detection**

Descriptive tasks:
* **Clustering**
* **Association**
* **Sequence Patterns**
* **Summerization**

Data mining tasks can be classified into predictive tasks and descriptive tasks. Predictive tasks make predictions based on user data collected whereas the descriptive tasks determine relationships among data, formulate patterns to depict relationship. Under predictive tasks we have classification, regression and deviation detection. Classification function is created by analyzing the relationship between attributes and data objects whose class labels are well known. Eg: Collect various demographic, lifestyle, education related information about customers to target specific group. Regression is a function that maps a data item to a real valued prediction variable. Deviation detection detects deviation from normal behavior. Eg: Change of search history of user.

Descriptive tasks contain clustering, summarization, association learning and sequential patterns. Clustering is the process of making group of abstract objects into similar subclasses.

Eg: find groups of posts that are analogous to each other based on the significant terms in them. Summarization is simple descriptions about subset of data. Association learning is if/then statements that identify associations/relationships between unrelated data. Sequential Patterns finds the complete set of frequent subsequences from a given set of sequences

II. LITERATURE REVIEW

It has been observed that many authors have worked on structured data using different mining techniques in social networks. Some of the observations are explained here. In [6] the authors have provided an overview of Knowledge Discovery in Databases (KDD) approaches. In [7] it has been demonstrated how modern technologies shifted the process of decision-making. A survey of visual data mining techniques is found in [8]. The authors have stated that large data sets with complex dimensions need a better way for representation. In [9],related work on data visualization is reviewed and classified data visualization techniques. In [10], the authors have surveyed interdisciplinary field of Text Mining. Most of the information found in computerized form are textual so it is required to extracts this unstructured text into informative knowledge. Several data mining techniques have been used on traditional data customers, e.g.: regression analysis, Naïve Bayes (NB), Support Vector Machine (SVM), Neural Network (NN). Regression analysis is the most popular technique to predict customer satisfaction [11], and the best technique for data sample less than 1000 compare to any data mining techniques [12]. For example, Mihelis et al. (2001) use regression model to measure customer satisfaction. Liu and Shih [13] develop a product recommender model to increase customer satisfaction, using weighted RFM (WRFM) and K-nearest neighbors with preference-based collaborative filtering.

III. SOCIAL NETWORKING FOR ELECTRONIC BUSINESS INTELLIGENCE USING DATA MINING

Social media tools are used for various purposes. Now a days social networking sites are used for building and maintaining the peer relationships, instant messengers for communication, online forum for expressing your opinions, etc.

Mining social media is very potential and useful for extracting more information and gaining deeper insight about customer..

Fig : A mining process for electronic -business data [5]
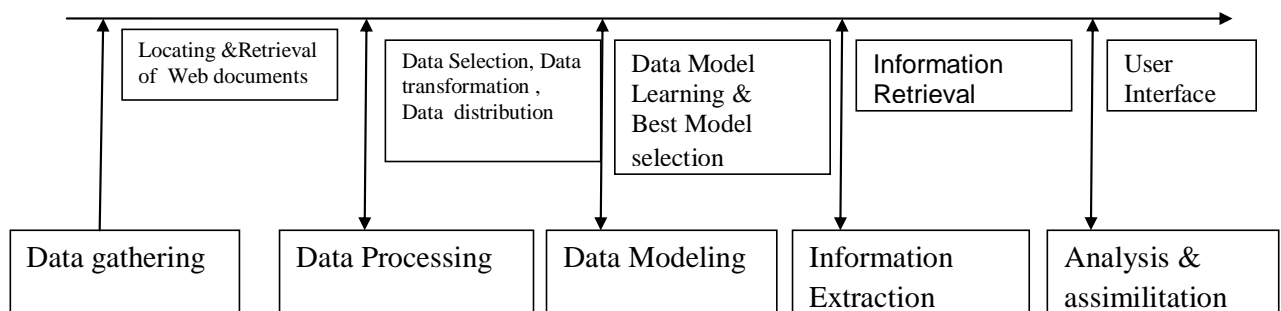
Social network analysis is used extensively in a wide range of applications and disciplines. In the private sector, businesses use social network analysis to support activities such as customer interaction and analysis, information system development analysis,[37] marketing,and business intelligence needs. Also development and analysis of individual and group strategies for community-based problem solving are required.

Data mining task can be decomposed into many sub tasks when dealing with electronic business data. Figure 1 illustrates a typical data mining process. At first data from required sources is collected. The next and most important task is analysis of data obtained from web. This includes preprocessing, actual mining process and knowledge assimilation. In the end, the discovered knowledge is presented to user. The analysis is useful for decision making ,predicting user behavior, deciding business strategies also.

Future Work :

Data available in social media profiles will be collected which is available across a number of dimensions or characteristics. After preprocessing training data , different classification algorithms will be applied to generate various classes or groups of customers. A comparative study of classification algorithms will be observed based on datasets using R software. The classification model will be tested for its accuracy.

This will help business organizations find useful patterns in better decision making and improve customer satisfaction and relationships. In future , by considering surveys and data sets various mining techniques are to be applied using advanced mining tools for dynamic data.

IV. CONCLUSION

Most of the work has been done on static datasets previously. But the dynamic nature needs to focus on new algorithms and technologies to handle this vast data. This paper focuses on predictive and nonpredictive techniques used for various business related operations. This paper gives an overview of key problems and techniques in social network analysis from the perspective of business applications. Social network analysis and mining constitute a very large, interdisciplinary area of study that is evolving fast. Therefore, analysis is by no means complete. However, goal in this article is to provide an overview of the main technical research areas in relation to business impact. Future work will focus on going deeper into examining the relationship between the techniques described and existing processes. The relevance of BI and social media integrate both domains together and motivate the articulation of a social BI research agenda towards dynamic analysis of social networks for business intelligence.

REFERENCES

[1] Abdul-Aziz Rashid Al-Azmi, DATA, TEXT, AND WEB MINING FOR BUSINESS INTELLIGENCE: A SURVEY, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.3, No.2, March 2013

[2] Javed Mohammed Department of Computer Science, New York Institute of Technology Old Westbury, NY , Business Intelligence and Analytics Evolution, Applications, and Emerging Research Areas, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 4, Issue 2, March 2015 ISSN: 2319-5967

[3] Social Media Business Model Analysis - Case Tencent,Facebook, and Myspace Department of Information and Service Economy Aalto University School of Economics

[4] Social Business Intelligence: a Literature Review and Research Agenda

[5] Richi Nayak , Data Mining for Web-Enabled Electronic Business Applications School of Information Systems Queensland University of Technology Brisbane QLD 4001, Australia

[6] Michael Goebel, Le Gruenwald, (1999) "A Survey Of Data Mining And Knowledge Discovery Software Tools," SIGKDD Explorations, Vol. 1, Issue 1. Pg 20, ACM SIGKDD.

[7] Chidanand Apte, Bing Liu, Edwin P.D. Pednault, Padhraic Smyth, (2002) "Business Applications of Data Mining," Communications of the ACM, Vol. 45, No. 8.

[8] Maria C. Ferreira de Oliveira and H. Levkowitz, (2003) "From Visual Data Exploration to Visual Data Mining: A Survey" IEEE Transactions on Visualization and Computer Graphics, Vol. 9, No. 3.

[9] E.H. Chi, (2000) "A Taxonomy of Visualization Techniques Using the Data State Reference Model," In the Proceedings of the Information Visualization Symposium InfoVis 2000, pp. 69-75.

[10] A. Hotho, A. Nu¨rnberger, G. Paaß, (2005) "A Brief Survey of Text Mining" GLDV-Journal for Computational Linguistics and Language Technologies.

[11] J. Hadden, A. Tiwari, R. Roy, D. Ruta, "Computer Assisted Customer Churn Management: State-Of-The-Art and Future Trends," Computers & Operations Research, October 2007, vol. 34, no. 10, pp. 2902-2917.

[12] I. Ograjensek, "Use of Customer Data Analysis in Continuous Quality Improvement of Service Processes", In Proceedings of the Seventh Young Statisticians Meeting, Metodološki zvezki, 2003, vol. 21, Ljubljana: FDV

[13] D. Liu, Y. Shih. Y, "Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences," The Journal of Systems and Software, 2005, vol. 77, pp. 181–191.