A Survey on E-Marketing by Behavioral Analysis in Online Social Network

Subhashree K[1], Vijaya Lakshmi T[2]

[1, 2]Assistant Professor, Department of Computer Science and Engineering, Karpagam College of Engineering

Abstract: *Online Social Network has become a great platform for many domains. It usually deals with huge amount of data. Data Mining is an essential factor in analyzing and making decisions in these voluminous data. E- Market is one such Domain, handling these type of huge data. Analyzing the behavior of each individual can create a good difference in profit of retailers in online shopping. The revenue of companies can be increased by integrating the shopping behavior of an individual with a recommendation system. Several challenges arise in the process of constructing an efficient recommendation system. Behavioral Analysis plays a major role in suggesting the right products to the customers. Since the Online Social Network is dynamic, the quality of results obtained should also be focused. This paper deals with the major factors involved in the construction of the Recommendation System such as Social Network Analysis, Behavioral Analysis, Data Mining Tasks and E- Market.*

Keywords: Social Network Analysis, Behavioral Analysis, Data Mining Tasks and E-Market

1. Introduction

A great revolutionary change has been occurred with the advent of Online Social Networks. Facebook, Orkut, LinkedIn, etc are becoming increasingly popular nowadays. Many online activities such as chatting, online shopping, gaming are made easy with the help of these networks. Social networks has been considered as the second fastest growing activity [8]. The data provided by the social networking sites are dynamic. Data mining tasks are performed on these huge data. There are many issues that arise in mining these dynamically growing data.

- Propogation of influence: As millions of users are getting attracted to the social networks, the individual user rely on those social network on making decisions. For example people refer the social network sites to decide which item to purchase, which film to watch, etc., Therefore this strategy has become a vital factor for E-Marketing [12].
- Detection of Community: Here we analyse the behavior of each user and group the users with similar interest. Grouping the individual with similar interest can be useful in making decisions about what product the customer likely to purchase.
- Suggestion system: This will suggest the user about the products they might be interested in. This suggestion system has been used in many websites to suggest items such as movies, gadgets, etc.,
- Predicting trust among user: Trust and Privacy is a major issue online social network. Few users take advantage of this online environment and disturb other users. In order to solve this issue, each user in the online social network must be assessed to predict the level of trust.

2. Data Mining Process

2.1 Tasks of Data Mining:
Depending on the use of data mining result, the various tasks of data mining are classified as follows,

2.1.1 Exploratory Data Analysis
To display the data in such a way that interesting features will become apparent, John Tukey developed a set of techniques called Exploratory Data Analysis (EDA) [3]. EDA techniques are used to encourage the data to suggest models that might be appropriate, whereas classical methods usually begin with an assumed model for the data. Many EDA techniques are provided by the Statpoint Technologies product, which are scattered throughout the statistical procedures. Huge amount of information's are available in the repositories. It analyzes the data, without the knowledge for what the customer is searching for. These techniques are interactive and visual to the customer.

2.1.2 Descriptive Modeling
A mathematical process that describes the real-world events and the relationships between factors responsible for them is a Descriptive modeling. Consumer driven organizations, uses descriptive modeling to help them target their marketing and advertising efforts. According to demographics, expressed interests, purchasing behavior and other descriptive factors, customer groups are clustered in descriptive modeling [1]. Where the customer groups share similarities and where they differ are identified by Statistics. Special attention is received by the most active customers, because they offer the greatest ROI (return on investment). It includes models for partitioning of the p-dimensional space into groups, overall probability distribution of the data and models describing all the data relationships between the variables.

2.1.3 Predictive Modeling
Though Predictive Modeling has many applications in business, it is often associated with meteorology and weather forecasting. For example, to identify the probability that a given message is spam, Bayesian spam filters use predictive modeling. The outliers in a data set that point toward fraudulent activity are identified by predictive modeling [1] in fraud detection. To target messaging to those customers who are most likely to make a purchase, customer relationship management (CRM) uses predictive modeling. Change management, city planning, disaster recovery, digital security management, engineering and capacity planning are several other applications which make use of predictive modeling. The value of one variable to be

predicted from the known values of other variables is permitted by his model.

2.1.4 Discovering Patterns and Rules:

The process of partitioning a set of data (or objects) into a set of meaningful sub-classes (clusters) is called Clustering. It helps the users to understand the natural grouping or structure in a data set. Used either as a preprocessing step for other algorithms or as a stand-alone tool to get insight into data distribution, which is primarily used to find the hidden pattern as well as to discover the pattern [4] in the cluster. A number of patterns of different size and clusters are available are available in a cluster whose aim is "how best we will detect the patterns", which in turn can be accomplished by using rule induction and many more techniques in the data mining algorithm and thus are called as the clustering algorithms.

2.1.5 Retrieval by Content

The efficiency of an information retrieval system is crucial because, the information retrieval systems usually deal with large user communities and/or large information collections. Matching of information needs with information items: which is the algorithmic task of computing similarity of information items and information needs. It constitutes the heart of the information retrieval model [5]. Relevance of information for users is modeled by the similarity of the structured representations and as a result a ranked result or a selection of relevant information items can be presented to the user and it imposes the fundamental constraints on the retrieval model. Those retrieval models that would capture relevance very well, but are computationally prohibitively expensive are not suitable for an information retrieval system. To find the data sets frequently used in the audio/video as well as images and in turn finding the pattern similar to the pattern of interest in the data set is the primary objective of this task.

2.2 Data Mining System Types

Data mining systems can be categorized based on various criteria's as follows,

- Huge amount of data are available in an organization where we need to classify these data which are available most of times in a similar fashion and hence it is required to classify these data according to its type (ex. text format or audio/video).
- Hierarchical data model, Object model, Object Oriented data model, Relational data model are the various data mining models available in real time. Different data is used in each and every model based on which every model is been classified.
- Based on the data mining functionalities or the kind of knowledge discovered, such as characterization, association, discrimination, classification, clustering [5] etc. some systems tend to be comprehensive systems offering several data mining functionalities together.
- According to the data analysis approach used such as neural networks, machine learning, genetic algorithms, visualization, statistics, database oriented or data warehouse-oriented, etc. the degree of user interaction involved in the data mining process such as autonomous systems, interactive exploratory systems, or query-driven systems [7] can also be taken into account in the classification. A wide variety of data mining techniques to fit different situations and options are provided by comprehensive systems which also offer different degrees of user interaction.

2.3 Life Cycle of Data Mining

There are six phases [2] in the life cycle of a data mining project. The sequence of the phases is not rigid, moving back and forth between different phases is always required and it depends on the outcome of each phase. The data mining life cycle process is depicted in the following figure 2.1
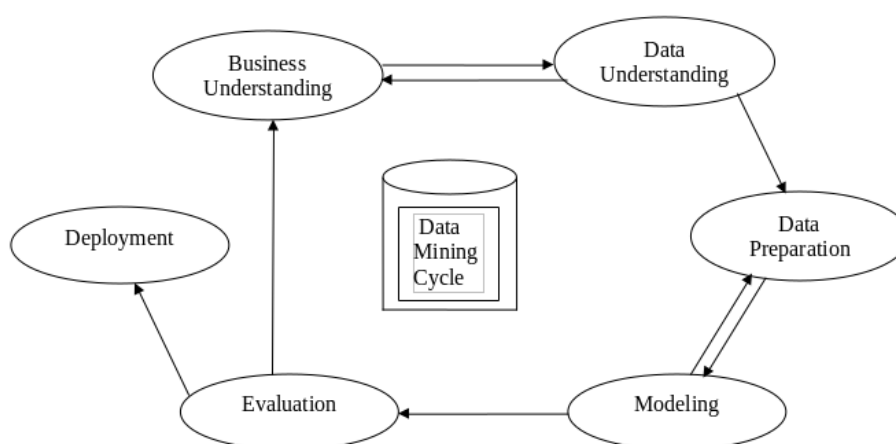

Figure 2.1: Life Cycle of Data Mining

The main phases are:

2.3.1 Business Understanding:
- Understanding the requirements from a business perspective and the project objectives

- Converting this knowledge into a data mining problem definition and a preliminary plan designed to achieve the objectives.

2.3.2 Data Understanding

Starts with an initial data collection

- To get familiar with the data
- To identify data quality problems
- To discover first insights into the data
- To detect interesting subsets to form hypotheses for hidden information.

2.3.3 Data Preparation:

- Collects all the different data sets
- Based on the initial raw data, constructs the varieties of the activities.

2.3.4 Modeling

- Various modeling techniques are selected and applied and their parameters are calibrated to optimal values.

2.3.5 Evaluation

- The model is thoroughly reviewed and evaluated.
- The steps executed to construct the model to be certain it properly achieves the business objectives.
- At the end, a decision on the use of the data mining results should be reached.

2.3.6 Deployment

- Purpose is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it.
- The deployment phase [14] can be simple or complex for generating a report and implementing a repeatable data mining process across the enterprise.

3. Social Network Analysis

Online Social Network (OSN) is nothing but a social structure among different types of user with a defined relationship [9]. It contains a number of nodes, which represents the individual user and links, which represent the relationship among the user. The social network analysis phases are depicted in the following figure 2.2.

Figure 2.2: Social Network Analysis Phases

Social Network Analysis (SNA) is nothing but the process of mapping and measuring the relationship and link among the individuals, group of users, an organization, URLs and many other information entities which are connected. Mathematical as well as visual analysis of the relationship among the entities of the network can be done with the help of Social Network Analysis. Some people name this process as Organizational Network Analysis (ONA) when they use this analysis with the business clients [15]. Organizational Network Analysis allows to get the complete details of the organization and reveals all relationships which connects every entity of the organizational Network.

In order to get a clear view about the network and the entities of the network, the evaluation of the location of the user and grouping of the user in the network is mandatory. After this evaluation process one can easily identify the users who act as a connector, who acts as a leader. We can easily spot out the clusters, individual users inside the clusters. On the whole, we can easily identify the core part of the network.

According to the Kite Network, two entities are related if they regularly in touch with each other and share information. The Kite Network clearly depicts the differences among the famous three centrality measures:

- Degree centrality: The measurement of network activity can be done in a Social Network with the help of degree of node. The degree of a node is number of direct links it has with other nodes [6]. We can say a node with highest degree as most active node in the network. These nodes can be treated as a connector of the network.
- Betweenness centrality: The nodes which have lesser degree when compared with a node with an average degree can also play an important role in the network. In this case, their location has to be considered. If these nodes occur in between two great groups, they play a vital role in communication of the groups. These nodes act as an intermediate. But this may also cause single point failure. Care should be taken such that if the central nodes fail, the communication between the two groups should not be affected.
- Closeness centrality: There may be certain nodes in the network which can access all nodes in the network very quickly. These nodes can be used to monitor the network flow since they have a better visibility about the happening in the network.

Certain other measures are also available like

- Network Centralization which gives an insight of individual user's location in the network [13] and it will also provide the relationship among the centralities of all nodes. A less centralized network may be less prone to single point failure
- Shortest path to reach a particular node is also one major factor which decides the faster reachability of a node.
- Boundary spanners – nodes that have a link to all connectors. They can combine the ideas and suggestions that come from all clusters.
- Client feedback is also an important factor, which analyzes the social network based on the comments and feedback obtained from the clients and stake holders.

There are two common categories [16] of social network. They are one mode and two mode networks. The one mode network refers to the network in which all the users are in one group and they are related together in the same group to which they belong. The two mode network refers to the network in which the users come in two different groups and the relation should only be between the two groups. No communication should be done inside the same group. E-Market deals with the two mode networks, where the customers come under one group and the products come under another group. The relationship between the customer and the product will be useful in analyzing the behavior of the customer.

4. Behavioral Analysis of Customers

It is possible to link the behavior of the individuals with the availability of social network data. Predicting the behavior of individual is the key objective in business. Online marketers used the data such as age, location and website usage pattern to predict the behavior [11] of the individual. But today, classical marketing strategy has replaced the individual level predictions. Now baseline modeling to predict the behavior of the customers has evolved. According to this strategy the data obtained from social networks such as Facebook, Twitter, Google, etc., can give better predictions about the customer behavior.

Many challenges arise while handling large amount of data. The major issue is predicting the human behavior in the social networks. The human behavior [10] revolves around a number of factors that may be far beyond computer science. So, predicting an individual's behavior and recommending the products interested to them is really a touch task to handle. Not only in behavior prediction, these issues have their impact on many domains like community discovery, trend prediction, information retrieval and so on.

Behavior Analysis is a great platform for industries to create new innovative ideas that focus on the new generation social network and its services. Sociology and anthropology researches have given an evidence to the advantage of demonstrating the advantage of using the behavior analysis. These researches have given way towards what the customer need, in terms of applications and services. In case of behavior analysis in E-Market to analyze the behavior of customer [16], one strategy that can be used is response to advertising.

Response to online advertisement is one analysis strategy to predict the behavior of the customer. This method uses the strategy that the person who are in interest with a particular product are more likely to click on to the online advertisement shown on their wall. These data are highly beneficial in predicting the customer behavior. The results of the analysis play a vital role in predicting the customer behavior. Based on the results the shopping interest of the individual user can be traced. The important factor to be considered is the growing size of data and changing nature of the customer's interest. No guarantee can be given that the shopping interest of the customer will always remain same.

5. Conclusion

Therefore a suggestion system which considers all the factors discussed above like the data mining tasks, Social Network Analysis, Customer behavior in social networks like online shopping can bring a great revenue difference in the organization. There are many obstacles like increasing data size, dynamic change in the customer behavior that can reduce the accuracy of the results obtained from the suggestion system. So, a special care should be taken to monitor the data size and dynamic change in the customer behavior. By considering all the factors, an efficient recommendation system can be built.

References

[1] Introduction to Data Mining and Knowledge Discovery, Third Edition ISBN: 1-892095-02-5, Two Crows Corporation, 10500 Falls Road, Potomac, MD 20854 (U.S.A.), 1999.

[2] The Survey of Data Mining Applications And Feature Scope, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), Vol.2, No.3, June 2012

[3] Baazaoui, Z., H., Faiz, S., and Ben Ghezala, H., "A Framework for Data Mining Based Multi-Agent: An Application to Spatial Data, volume 5, ISSN 1307-6884," Proceedings of World Academy of Science, Engineering and Technology, April 2005.

[4] Kay, J., Maisonneuve, N., Yacef, K., Zaiane O., "Mining patterns of events in students' teamwork data", Proceedings of the ITS (Intelligent Tutoring Systems) 2006 Workshop on Educational Data Mining, pages 45-52, Jhongli, Taiwan, 2006.

[5] Schultz, M. G., Eskin, Eleazar, Zadok, Erez, and Stolfo, Salvatore, J., "Data Mining Methods for Detection of New Malicious Executables". Proceedings of the 2001 IEEE Symposium on Security And Privacy, IEEE Computer Society Washington, DC, USA , ISSN:1081-6011, 2001.

[6] Bartok J., Habala O., Bednar P., Gazak M. & Hluchy L. (2010). Data mining and integration for predicting significant meteorogical phenomena. International Conference on Computational Science, (ICCS 2010)

[7] Liu, L., Kantarcioglu, M., Thuraisingham, B.M.: A Novel Privacy Preserving Decision Tree. In: Proceedings Hawaii International Conf. on Systems Sciences (2009)

[8] Sukhpreet Dhaliwal Hikmat Jahangirli Riad Aljomai Abdullah Sarhan Wadhah Almansoori Reda Alhajj-Integrating Social Network Analysis and Data Mining Techniques into Effective E-Market Framework, ICIT 2013 The 6 th International Conference on Information Technology

[9] Songjie Gong. A collaborative filtering recommendation algorithm based on user clustering and item clustering. Journal of Software, 5(7), 2010.

[10] Zhao Qiu, Mingrui Chen, and Jun Huang. Design of multi-mode e-commerce recommendation system. In Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on, pages 530 –533, april 2010.

[11] Reihaneh Rabbany k., Mansoureh Takaffoli, and Osmar R. Za ̇ïane. Social network analysis and mining to support the assessment of on-line student participation. SIGKDD Explor. Newsl., 13(2):20–29, May 2012.

[12] Guo Yan-yan and Liu Qi-cheng. E-commerce personalized recommendation system based on multi-agent. In Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference on, volume 4, pages 1999 –2003, aug. 2010.

[13] Ziming Zeng. An intelligent e-commerce recommender system based on web mining. International Journal of Business and Management. Vol 4, No 7, 2009.

[14] Songjie Gong. A collaborative filtering recommendation algorithm based on user clustering and item clustering. Journal of Software, 5(7), 2010.

[15] Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes. Social network analysis and mining for business applications. ACM Trans. Intell. Syst. Technol., 2(3):22:1–22:37, May 2011.

[16] Kumar and P. Thambidurai. Collaborative web recommendation systems based on association rule mining. International Journal of Computer Science and Information Security., 8(3):222–227, June 2010.

[17] Ziming Zeng. An intelligent e-commerce recommender system based on web mining. International Journal of Business and Management. Vol 4, No 7, 2009.