INTEGRATING SOCIAL NETWORK ANALYSIS AND DATA MINING TECHNIQUES INTO EFFECTIVE E-MARKET FRAMEWORK

Sukhpreet Dhaliwal Hikmat Jahangirli Riad Aljomai Abdullah Sarhan Wadhah Almansoori Reda Alhajj

Department of Computer Science
University of Calgary
alhajj@ucalgary.ca

Department of Computer Science
Global University
Beirut, Lebanon

Abstract

E-commerce has been considered to be a very important source to share business information and establish a relationship between people on a communication network especially the Web which is driven by the internet. Many retailers have introduced online shopping as a powerful tool to increase revenue either by marketing or analyzing the customers shopping behavior to make the right sales decisions. In this paper, we introduce data mining techniques including clustering, association rules, and social network analysis integrated into a a recommendation system to increase revenue for those companies that use online shopping. Our proposed recommendation system consists of the following recommenders: (a) Shopping List Recommender (b) Checkout Recommender (c) and e-Navigation Recommender. By understanding the customers behavior, we can offer deals to targeted customers which will lead to increase revenue. We have tested our proposed modules using a real online music store and the results are outstanding.

Keywords - Recommendation System, Data mining, Social Network Analysis, E-commerce.

1 INTRODUCTION

All the pages of the paper size should be A4 (21 x 29,7cm). The top and bottom margin should be 2,5 cm the left and right margin. All the text must be in one column format and Arial must be used, including figures and tables. Single-spaced with 10-point interlining spacing. [Arial, 10 point, normal, alignment justify, upper and lower case]

When people connect to an on-line shopping server, they unintentionally leave behind rich information that can be used to understand their shopping behavior. However, this collected data for customers is too complicated to understand by considering its raw version. To best use this data, online shopping server owners analyze this data to extract useful information that can help business owners to make proper sales decisions. This process is called E-commerce (EC). EC is defined as any activity that utilizes some form of electronic communication in the inventory, exchange, advertisement, distribution or payment of goods and services [5], [13]. A good example of collecting online user's data is user's purchase history. Online retailers gather plenty of data about consumers purchase and analyze the online shopping behavior of the consumers to make the right sales decisions to attract more customers and help them decide what to buy using recommendation system techniques [15], [12].

Recommendation systems refer to intelligent sites making recommendations to users to meet individual needs based on users interests, to automate the process of analyzing user purchase trends in order to predict what could be the best subsequent product to be brought to the attention of the user. It is not feasible if at all possible to do the process manually because the data to be analyzed is interrelated and complicated, and hence cannot be easily understood by a human with the intend to capture some useful predictions. Recommendation systems may recommend products to users based on previous purchase behavior of the individual users or based on most popular products purchased with similar consumers patterns [17], [6], [16]. An automated process should be able to provide recommendations like customers who purchased certain products also purchased some other products in the same transaction. Further, it is also important to highlight that friends of a given customer purchased certain products together.

In this paper, we introduce a Recommendation System integrated with social network analysis, data mining, and e-navigation techniques. We will discuss these techniques in more details in the next sections. Our proposed system has the following contributions:

1) Helping businesses to make proper sales decisions in order to increase revenue
2) Identifying communities of users based on their purchase behavior similarity
3) Saving time of users by recommending navigation links
4) Helping users to respond to the current sales by recommending a shopping list
5) Reminding users to purchase items related to their current shopping basket based on trend analysis

Unlike, the existing recommendation systems, our proposed approach takes advantage of the best practices of different methodologies including social network analysis, association rules, clustering, and e-navigation to produce a recommendation system with accurate results.

The rest of the paper is organized as following. Section 2 outlines a background on social network analysis and data mining related to our approach. Section 3 is the related work followed by our proposed recommendation system approach in Section 4. Section 5 is the conclusion and future research directions.

2 BACKGROUND

In our proposed system, we use two main techniques known as Social Network Analysis (SNA) and Data mining which we briefly explain below for convenience.

2.1 Social Network Analysis

Social networks (SN) are defined as the social structure between groups of people or things with a defined relationship. SN consists of nodes representing the actors and links representing the relationship between the actors. A social network is characterized by its mode which reflects the number of disjoint group of actors which could be connected. The most common categories for popular applications are one mode and two mode networks. The former refers to the case where all actors belong to one group and links connect actors within the same group. For instance genes form one group of actors and a link between two genes could reflect the case where the two genes are expressed in the same samples; considering the percentage of samples as label for the link leads to a weighted network. On the other hand, the second category expects two disjoint groups of actors (it is possible to have the two groups as duplicates of the same group) and links are allowed only across the groups. This could be illustrated by considering customers and products as two groups of actors, a link between a customer and a product will refer to the interest of the customer in purchasing the product.

SNA is used to observe the social activities of individuals and find the most important actors that are highly socially connected with the other actors. These important actors can be used to influence other actors to increase their social activities in the same network [11], [3]. In our proposed approach, we form the social network between online shoppers and the products with the purchase history relationship and further analyze the users activities as will be discussed in Section 4.

2.2 Data Mining

Data mining is the process of analyzing a given data in order to discover knowledge which is implicitly present in the data and could not otherwise be retrieved by traditional query models. Several data mining techniques have been proposed in the literature for effective data analysis. Each has its own expectations, methodology and outcome. However, the common requirement for all techniques is a two sets of entities connected by many to many relationship. For instance, customers and products could be considered as the two sets of entities and the relationship is purchase patterns of customers. Association rules mining and clustering are two of the techniques commonly used by researchers to identify frequent and similar patterns, respectively.

2.2.1. Association Rules.

Association rule mining is the process of generating associations or, more specifically, association rules, in transaction databases [10]. Given a set of transactions each consists of a sets of items, the association rules mining process consists of two steps. First, frequent sets of items are found and then possible rules are generated from each frequent set and only rules that satisfy some criteria are returned as interesting.

A pattern is a set of items and is said to be frequent if and only if it is repeating in at least a certain number of transactions. Frequent patterns are important because they reveal certain correlation between items, e.g., items mostly purchased together by certain customers. The latter customers may have some behavior in common and hence may form a community from a social network perspective. Once all frequent patterns are determined, the second stage of the process determines all rules that could be derived from each itemset. A rule has antecedent and consequent which are two disjoint sets of items. A rule is reported to the user if it has high confidence which is measured by considering transactions that contain items in the antecedent to count how many of them contain items from the consequent.

In recommendation systems, association rules are used to remind customers to purchase other items that are frequently purchased with items in the basket [7]. For instance, it is possible to bring to the attention of the customer which other items to consider based on his/her already selected items. This recommendation takes into consideration the purchase patterns of other customers.

2.2.2. Clustering.

Clustering is an unsupervised learning method with no prior knowledge of the classes to be derived. The main idea is to group (cluster) objects with similar characteristics into different groups (clusters) so that each group of objects share the same similarity to a certain degree; and similarity is minimized across the groups. A good clustering would lead to high homogeneity within the individual clusters and high heterogeneity between the clusters. In our work, we intend to find communities of users with similar online purchase behavior and explore the purchase interests in each cluster and find recommendation patterns between users in the same cluster. More details will be discussed in the next sections.

3 RELATED WORK

Adomavicius and Tuzhilin [1] present a classification of the recommender systems into three main categories: collaborative, content-based, and hybrid recommendation approaches. The recommendation system suggested in this paper combines content-based and collaborative recommendation approaches; thus can be considered as a hybrid recommender. This approach allows the system to overcome problems, such as over-specialization, new user, and new item problems that most recommendation systems suffer from.

The recommendation system proposed in this paper makes use of the popular Apriori algorithm for association rules mining. Any other association rules mining algorithm could be used. However, as mentioned by Kumar and Thambidurai [9] Apriori is one of the algorithms that remain almost unaffected by the rapid increase of the data size being processed. This allows keeping the overall complexity of the system under control by offering a resource efficient recommendation system.

The matching algorithm uses the customer preference and association scores of a given product for a user and suggests the items which are highly rated [17]. Customer preferences and associations which are mined from the click streams of customers are similar to our approach, but the use of apriori algorithm in our approach gives more to get the highly frequent items which are related to a given user. The collaborative filtering technology by joining the user clustering and item clustering technology improves the performance of the recommender system [12]. The improved collaborative filtering technology not only improves the performance but makes it more personalized. Using the same idea in our project, we improve the Apriori algorithm by using the transaction records which are related to a given user cluster.

The Meerkat-ED is a toolbox for analyzing students interactions through highlighting the leader/peripheral students, and collaborative students groups using social network analysis techniques [14]. Furthermore, it assists the instructors to find the topic for a debate and the participation of students. Similarly, the use of social network in our project makes the clusters of users which have similarity between them and gives the participation level of the users [14].

Social network analysis and mining for business applications are becoming popular trends to observe the available information from the e-commerce sites and improve the service based on the results from the analysis [2]. Promotion of products and services in online social networks are based on the user interaction in the social network which focuses on user's preferences more clearly by figuring out the detailed characteristics of the user [2]. In this project, the generation of social network based on the user relationship through the genre of music helps to create the clusters of users who have similar buying patterns which overcome the problem of scarcity of data.

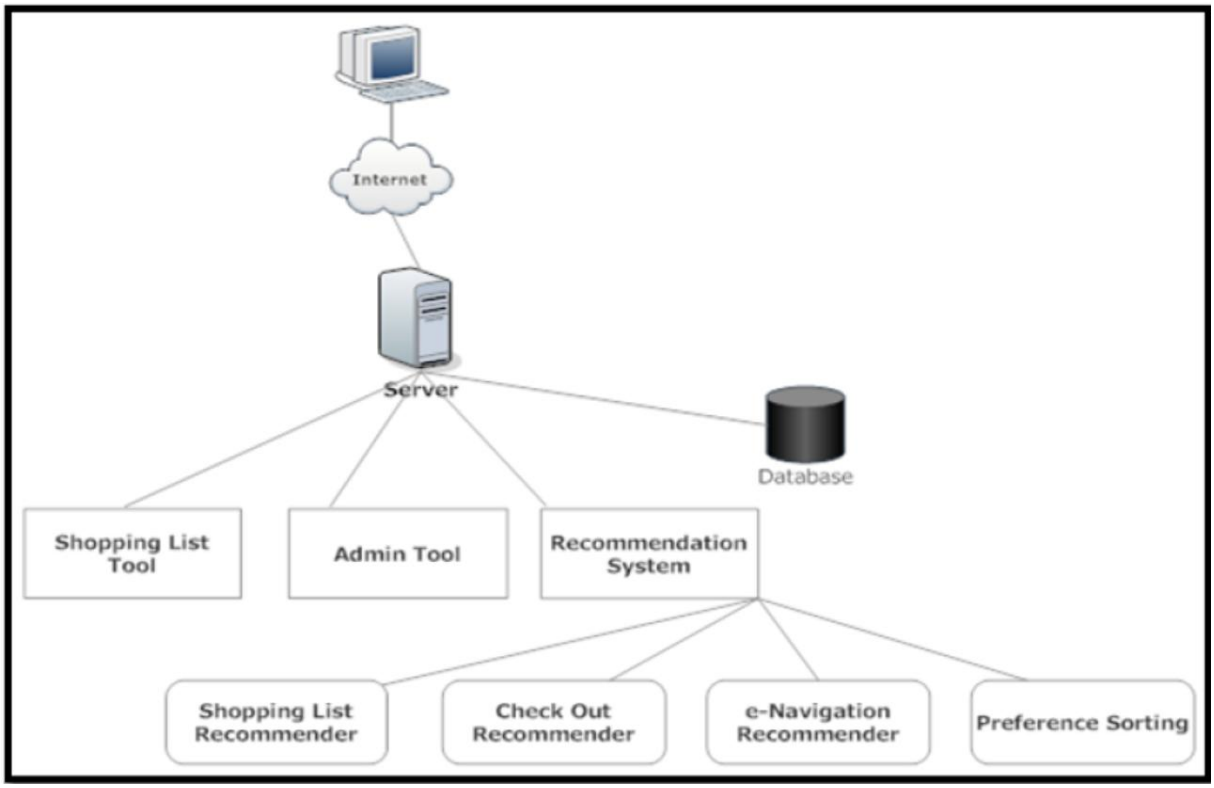

Figure 1. Recommendation System Component.

4 PROPOSED RECOMMENDATION SYSTEM APPROACH

The idea of the recommendation system approach is to increase revenue to online retail sites by increasing the sales. Our main goal is to attract many customers to purchase by:

1) Guiding customers on what products they should buy according to their shopping interests

2) Guiding customers on what products they should buy according to their navigation patterns

3) Guiding customers on what products they should buy based on the popular products and sales

4) Recommending customers to purchase products according to the interests and shopping behavior of the shopping community they belong to. The ideas and solutions presented in this paper have been tested through the development of a prototype application for an online retail site for songs. This online site allows users to view available songs for sale and purchase them through a secured payment system. Users may navigate through the site by songs names, genre, artists, etc. In addition to the common shopping functionality offered by most online retail sites, we propose the following main components and adopted some of those from [4] as part of our proposed recommendation system as shown in Figure 1.

4.1 Components of the Proposed Approach

4.1.1. Database and Server. As in any other system, we store and maintain the data for the online retail site in a secured database connected to a server to manage the system transactions.

4.1.2. Shopping List Component. The shopping list feature operates functions with the basic online shopping functionality such as navigating through products, adding comments, rating the products, search feature, shopping basket, payment services, etc.

4.1.3. Admin Component. The Admin Component feature is responsible for maintaining the site. It controls the sales decisions.

4.1.4. Recommendation System Component. This is the main component that includes our proposed recommendation approach. It contains four recommenders.

1) Shopping List Recommender: The Shopping List Recommender is responsible to recommend a shopping basket list to the customers with recommended items to order. The idea of this recommender came from the fact that some customers do not like to spend too much time navigating through all the items and checking if any items were on sale. This recommender will select the best items to match the customers interests and shopping behavior.
2) Checkout Recommender: The Checkout Recommender assists users during the checkout process to remind them about other items that they might have forgot by promoting some popular products associated to the items listed in the shopping cart that were popularly purchased by others. In addition, this recommender will promote the products on behalf of the retailers [4].
3) E-navigation Recommender: This component keeps track of the product visits by both anonymous users and identified users. This information is used later to guide the customer to the next product to visit based on the tracked behavior.
4) Item Preference Sorting: This feature allows the display of the recommended items in the order of the users interests and shopping behavior.

4.2 Proposed Approach Methodology

As mentioned earlier, the main outcome of our proposed approach is to help attract and guide customers to buy products based on their behavior and preferences in order to increase online retail site revenue.

As shown in Figure 2, our proposed system combines domain experts, data mining, and social network analysis to form an online shopping recommendation system. Most of the recommendation systems focus on one area and develop an algorithm based on the shopping behavior data. However, our system takes advantage of the best practices available to produce a better recommendation system that will lead to increase retailers revenue. The main goal of the domain experts is to lead the business revenue increase and define issues and difficulties of the market. In addition, business domain experts can help in making sales decisions on certain products. For example, if our system targets those sales watcher customers, then we need to consult with the business experts to make the proper attractive sales to satisfy the customer desire. The domain expert also includes the administrators to program several rules we will need to run some of the data mining and social network analysis proposed methods. When people use an online retail sites, they leave behind them a lot of data that we use to feed our model. We store this data in our database as consumer data which contains consumers purchase history, navigation track, shopping interests, and indirect or hidden information known as user preferences that we use to feed our social network component. Every data collected is preprocessed in a format compatible to the recommender component.

4.2.1. Association rules. We define customers as anonymous customers and we defined customers which reflect customers that navigate through the website without logging in using their credentials and customers who navigate through the website after logging in to the system using their credentials, respectively. At this stage, customers may request recommended shopping list; in this case our approach will have two cases:

1) If the user is anonymous and not logged in, then we will run apriori algorithm based on most popular purchased items and reviewed items maintained as part of the tracked navigation data for all users including items that are on sale according to the anonymous users' behavior. For example, if users always respond to items on sale, then new items listed on sale will be considered for the recommendation. A list of all items returned by the algorithm will be recommended to the user in the order of highest to lowest preferences.
2) If the user is defined (logged in), then we retrieve the users profile and tracked clicks and analyze his/her behavior to feed the apriori algorithm for the recommended list similar to the previous case but based on particular users' profiles. We used 90% as the association rules criteria.

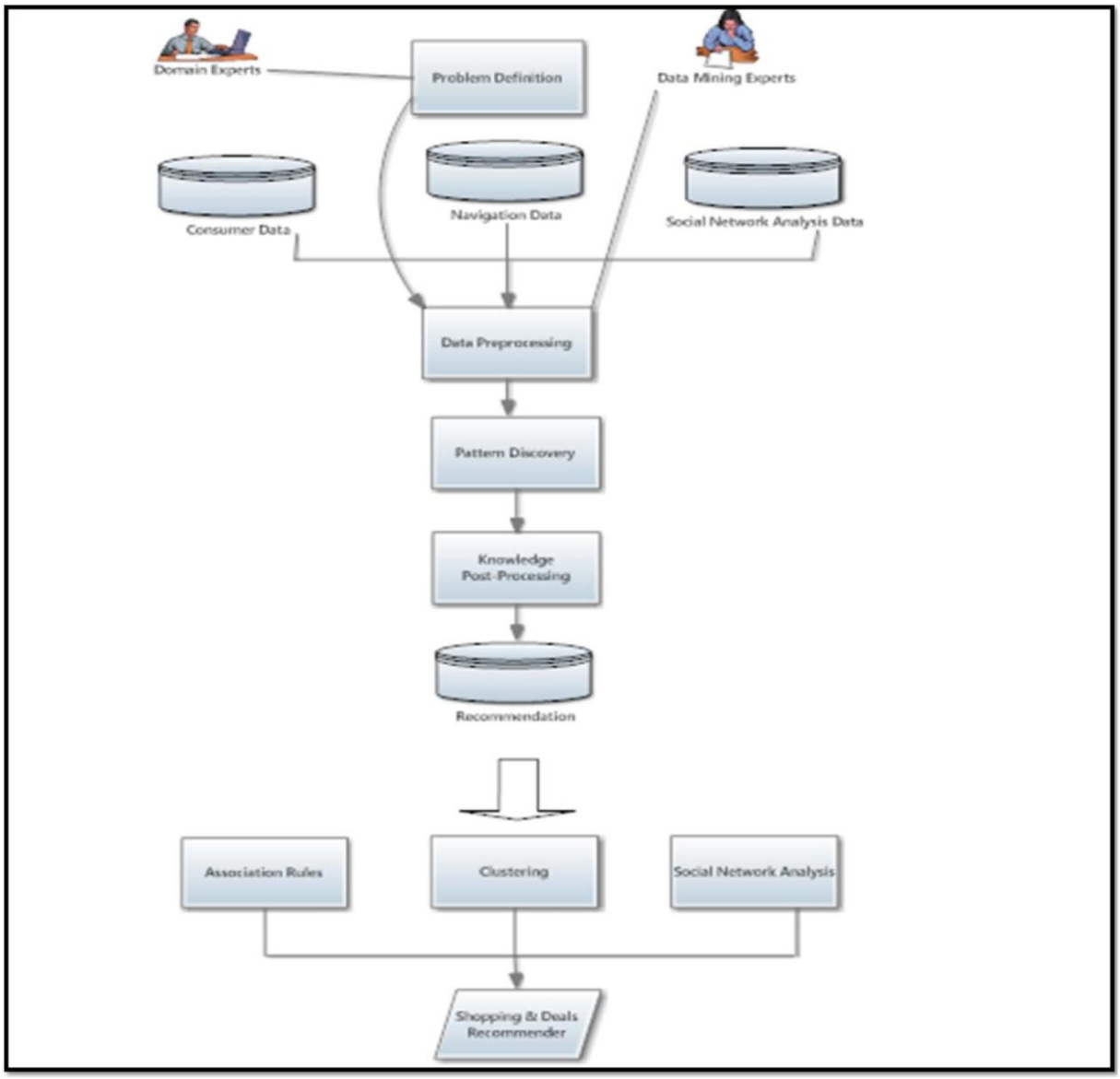

Figure 2. Proposed Approach Methodology and Results

4.2.2. Clustering. Consumers are grouped into communities based on genre they like based on the following algorithm:

1) Define a matrix where rows represent consumers and columns represent genre they purchased.
2) Cluster the matrix using EM algorithm so that all users with similar genre interest will be in the same community.

3) Within the members of each cluster, check the songs purchased between each pair say Alex and Bob and recommend to Alex songs purchased by Bob but not by Alex.

The idea behind this step is to use popular consumers that may influence others as resources for recommendation. The results produced from the clustering technique are added to the shopping list recommender to recommend items purchased by others in the same cluster. As shown in Figure 4, we have a good distribution of cluster members with similar shopping behavior with lowest 2% and highest 13% of members assignment percentage.

4.2.3. Social Network Analysis. The benefit of the social network analysis is to analyze and explore the social relationship between communities that may affect one another for purchase behavior. We use the following steps to build our social network model: 1) Define a matrix were rows represent consumers and columns represent genre they purchased and the data is either zero or one. This network is two-mode network (bipartite graph) and must be converted to two one-mode networks as in step 2. 2) Transpose the network built in step 1 into two one-mode networks one represents users with common genre and the other one represents genre with common purchased patterns. 3) Extract the cliques from the two networks in step 2 and in every pair of cliques with the highest common members, recommend products purchased by the non-common customers. Finding cliques in a network is to find smaller groups (three or more nodes) within a larger group who all choose each other. The benefits of finding cliques is to see how smaller groups within the social network share common interests and this may lead to predict future connections between nodes. Figure 4 shows the clique names and clique members of some selected cliques for space issues, and Figure 5 shows the size of the clique which is basically the number of nodes in a clique where nodes represent customers. We analyze the cliques by comparing the members of each clique with the other cliques members. For instance, if we compare clique2 and clique3 in Figure 4, we can find that the cliques have the same 15 members except customer30 and customer38. This means that these two customers do not have links to each other at all. However, in clique2 and clique3 they share a large number of other customers meaning that those common customers like the same genre. In the shopping interests domain, this can predict connections between customer 30 and 38 in the future through the common 15 customers. In addition, having a link between customers 30 and 38 with the other 15 customers but not with both of them gives an indication that there is one genre in common between the 15 customers and customer 30 which we will use to recommend it to customer 38 since this customer has common interests with the same 15 customers.

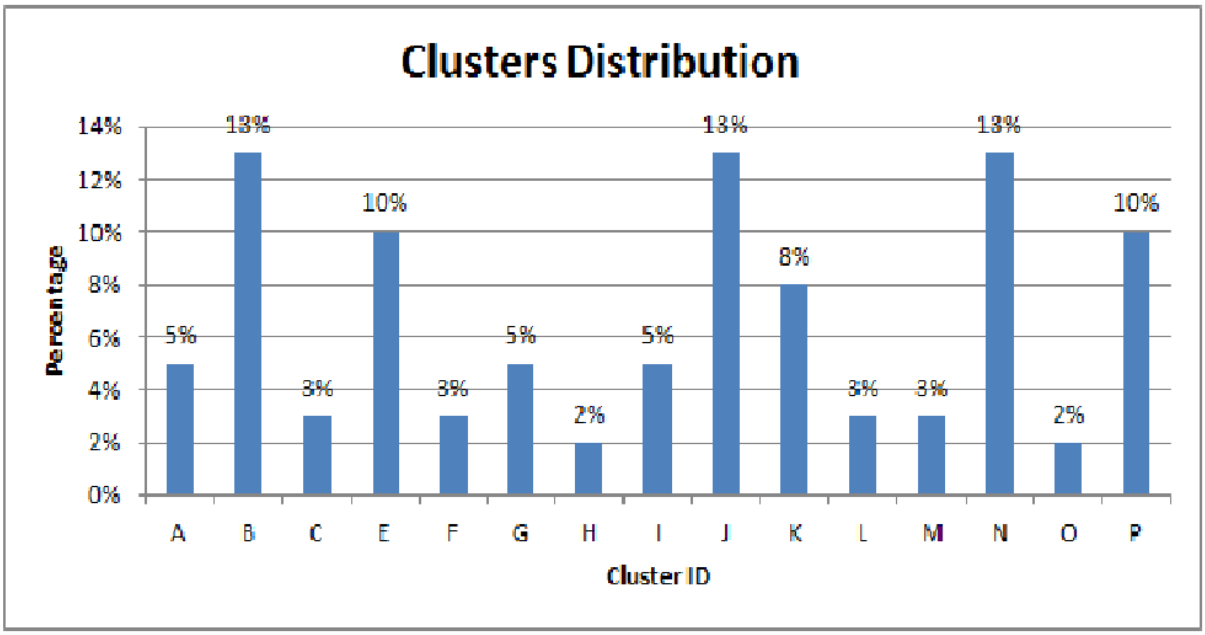

Figure 3. Clustering Results

Clique1	Node14	Node15	Node18	Node10	Node9
	Node13	Node60	Node6	Node7	Node5
	Node17	Node16	Node19	Node4	Node3
	Node12	Node11	Node2	Node8	Node1
Clique2	Node14	Node15	Node18	Node10	Node9
	Node13	Node60	Node6	Node7	Node5
	Node17	Node16	Node19	Node4	Node30
Clique3	Node14	Node15	Node18	Node10	Node9
	Node13	Node60	Node6	Node7	Node5
	Node17	Node16	Node19	Node4	Node38

Figure 4. Cliques Members

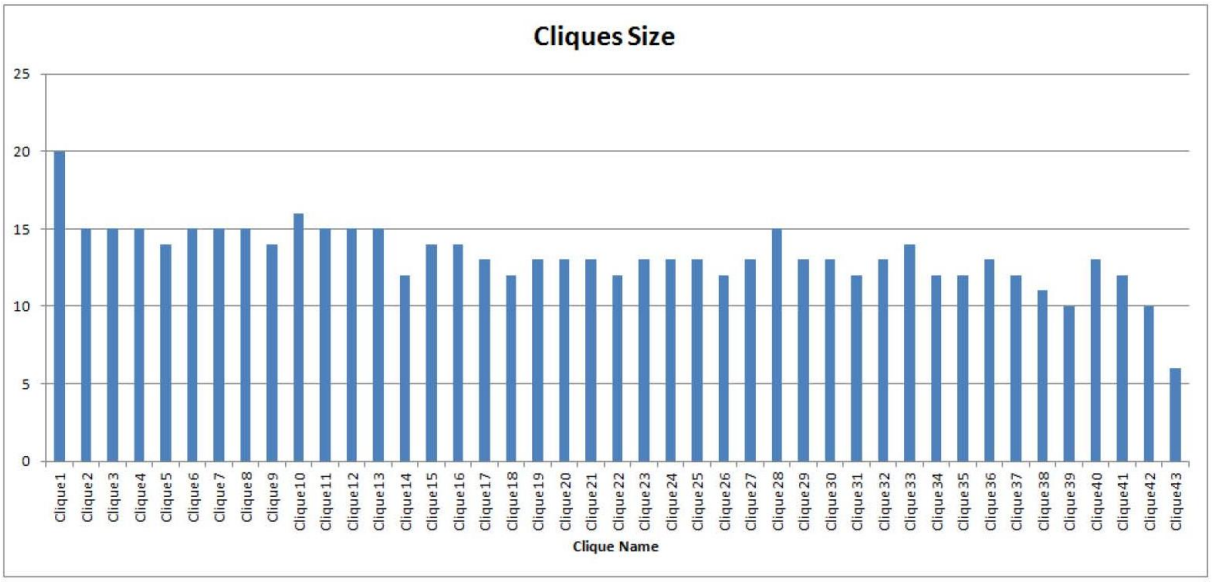

Figure 5. Cliques Sizes

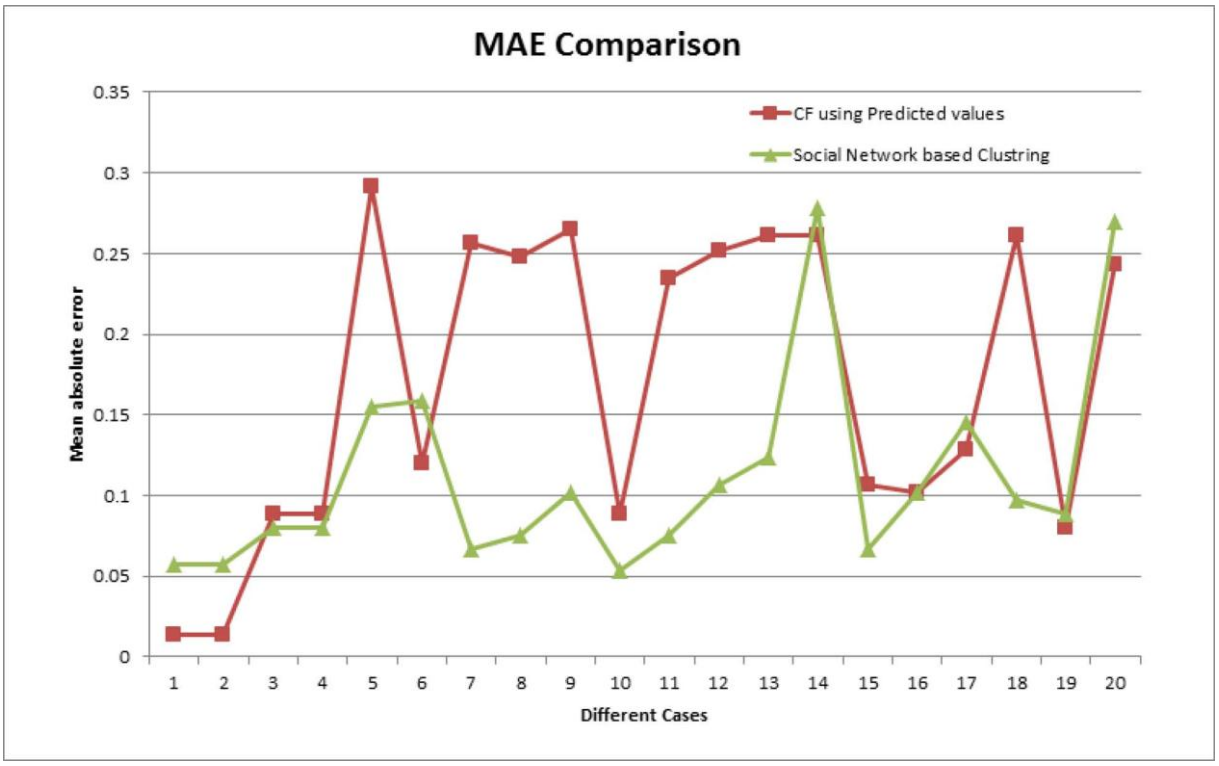

Figure 6. MAE Comparison.

4.3 Evaluation

We have evaluated our approach to an existing collaborative filtering recommendation algorithm based on user clustering and item clustering [8]. Basically this algorithm uses the statistical approach based on MAE values to evaluate the system. We have implemented this algorithm and applied our dataset to it and then compared the results generated by our proposed approach versus the collaborative filtering approach using training and testing datasets by calculating the statistical methods MAE to evaluate the user preference closer to the actual preference. As shows in Figure 3, our approach produces closer preferences to the actual.

The existing collaborative filtering approach is summarized as following: 1) Overcome the problem of sparsity 2) Predicting the values for the user preference to an item still based on the user cluster but can be incorrect if the user ends in wrong cluster because of no data about user preferences yet. 3) Then predict the item based on item clustering We believe our approach is better because of the following: 1) Integrating the data mining and social network analysis techniques 2) Data mining techniques: apriori algorithm which uses association rules to offer the items to a given user which offers the items to new user by predicting most frequent items and for existing user offers the items based on his/her history (purchasing record). 3) Will help to learn user preferences more quickly. 4) Using social network matrix based on the genres of music a user like helps to create clusters. Based on the cluster a user has, the recommendation system offers the items which make it more personalized and also increase the performance since we do not have to evaluate every item in the database which is closely related to a user but grabs the similar users from clusters and finds out the their preferred items and sort the items from most recent to oldest.

5 CONCLUSION AND FUTURE WORK

In this project, we introduced a recommendation system approach by using clustering, association rules, navigation, and social network analysis to increase revenue for those companies that use online shopping. We have applied our approach to online retail site for songs. We compared our approach to the collaborative filtering approach that uses MAE values. Our approach shows better results. We are

currently trying to apply our approach to different online site domains and use prediction features to predict the behavior of current and new customers.

6 REFERENCES

[1] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE Trans. on Knowl. and Data Eng., 17(6):734–749, June 2005.

[2] 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.

[3] Lianhong Ding and Peng Shi. Social network analysis application in bulletin board systems. In Intelligence Science and Information Engineering (ISIE), 2011 International Conference on, pages 317 –320, aug. 2011.

[4] Olga; Pramataris Katherine; Doukidis, Georgios; Papaemmanouil and George. Prassas. A recommender system for online shopping based on past customer behaviour. In Proceedings of the 5th international conference on Intelligent user interfaces. BLED 2001 Proceedings Paper 5, 2001.

[5] Oren Etzioni. The world-wide web: quagmire or gold mine? Commun. ACM, 39(11):65–68, November 1996.

[6] Yongjian Fan, Yanguang Shen, and Jianying Mai. Study of the model of e-commerce personalized recommendation system based on data mining. In Electronic Commerce and Security, 2008 International Symposium on, pages 647 –651, aug. 2008.

[7] Xiaobin Fu, Jay Budzik, and Kristian J. Hammond. Mining navigation history for recommendation. In Proceedings of the 5th international conference on Intelligent user interfaces, In Proceedings of IUI ’00, pages 106–112, New York, NY, USA, 2000. ACM.

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

[9] A. 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.

[10] Jiuyong Li, Hong Shen, and Rodney W. Topor. Mining the smallest association rule set for predictions. In Proceedings of the 2001 IEEE International Conference on Data Mining, ICDM ’01, pages 361–368, Washington, DC, USA, 2001. IEEE Computer Society.

[11] R.S. Monclar, J. Oliveira, F.F. de Faria, L. Ventura, J.M. de Souza, and M.L.M. Campos. Using social networks analysis for collaboration and team formation identification. In Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on, pages 562 –569, June 2011.

[12] Maurice D. Mulvenna and Alex G Buchner. Data mining and electronic commerce. In Overcoming Barriers to Electronic Commerce (OBEC’97), Málaga, Spain, 1997.

[13] 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.

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

[15] Luo Ya. E-commerce personalized recommendation technology based on expert system. In Digital Content, Multimedia Technology and its Applications (IDC), 2010 6th International Conference on, pages 342 –345, aug. 2010.

[16] 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.

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