Hybrid Recommender Systems using Social Network Analysis

Kyoung-Jae Kim, Hyunchul Ahn

Abstract—This study proposes novel hybrid social network analysis and collaborative filtering approach to enhance the performance of recommender systems. The proposed model selects subgroups of users in Internet community through social network analysis (SNA), and then performs clustering analysis using the information about subgroups. Finally, it makes recommendations using cluster-indexing CF based on the clustering results. This study tries to use the cores in subgroups as an initial seed for a conventional clustering algorithm. This model chooses five cores which have the highest value of degree centrality from SNA, and then performs clustering analysis by using the cores as initial centroids (cluster centers). Then, the model amplifies the impact of friends in social network in the process of cluster-indexing CF.

Keywords—Social network analysis, Recommender systems, Collaborative filtering, Customer relationship management

I. INTRODUCTION

IN order to mitigate the information overload, many people use recommender systems that have the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [1]. However, it is needed to develop accurate recommendation engine using proper recommendation algorithm for them. If they do not provide useful information to the customer, the corresponding companies may loose their customers. For this reason, many studies have tried to develop effective and efficient recommendation algorithms. The most popular recommendation algorithm is collaborative filtering (CF). CF predicts user preference score using preference of other similar users who have rated the items [2]. So far, CF has been a very successful approach for building recommender system, but it has a few shortcomings to address. Typical limitations of CF are scalability and sparsity problems. In this study, we suggest a new limitation of typical CF. It has limitation to reflect qualitative and emotional information about users. This means that it unable to distinguish neighbors as friends or strangers with similar taste [3]. Qualitative and emotional information among users may not be considered properly in recommendation process because typical recommender systems do not consider the human relationships among users. However, it may be important to distinguish friends and strangers in recommender systems, since users generally prefer recommendations from his or her friends to those from strangers [4]. To address the limitation, this study proposes cluster-indexing CF with social network analysis (SNA) for recommender systems. Using SNA, the recommender system enables to distinguish friends and strangers of the user. Thus, SNA may provide qualitative and emotional information between members of the same community.

Kyoung-jae Kim is with Dongguk University, Seoul, South Korea (phone: +82-2-2260-3324; fax: +82-2-2260-3684; e-mail: kjkim@ dongguk.edu).
Hyunchul Ahn is with Kookmin University, Seoul, South Korea.

In addition, typical CF often requires a large memory space and computation time to retrieve neighbors of user, but cluster-indexing CF only requires relatively small computation time because it searches neighbors within a subgroup of the user.

This study uses SNA to choose the centroid in subgroups as an initial seed for a conventional clustering algorithm. First, we collect data about users' ratings for several movies and their social network relationships. We develop data collection site for this purpose. Then, we perform social network analysis and find cores of each subgroup of social network. Next, we use these cores as cluster centroid for cluster analysis. Finally, we make recommendations about items using CF within user's cluster, and compared the performances of several comparative models.

This paper consists of six sections. Section 2 describes related studies about recommender systems and social networks. In Section 3, we propose a novel hybrid CF with SNA for recommendations. Section 4 introduces experimental data and design of experiments that have incorporated CF and social networks. Section 5 shows the experimental results, and Section 6 discusses the findings and limitations of our study

II. PRIOR RESEARCH

Recently, many studies proposed application of SNA to CRM and marketing research area. Kautz et al. [5] utilized knowledge in the social networks to recommend public documents found on the Web. They emphasized that although their model did not replace generic search engines, but the model could make their search more focused and effective. In addition, Liu & Maes [6] harvested 100,000 social network profiles and built "InterestMap", a network-style view of the space of interacting interests and identities, by automatically analyzing patterns of correlation between various interests and cultural identities. They suggested that recommendations made in that network space were accurate and highly visually intelligible.

Some studies tried to use the information social networks to enhance the quality of recommendations. Ziegler & Lausen [7] investigated and analyzed presence or absence of correlation between trust and interest similarity in an online community focusing on books by comparing user similarity among all peers and some trusted peers. They founded there were a positive correlation and also proposed that the level of expressed trust between user and his/her peers could be a good predictor of preference. In addition, Ryu et al. [8] selected the nearest neighbors of target user utilizing Pearson's correlations and used them as the target user's social networks. Moreover, Golbeck [9] proposed FilmTrust, a website that integrates web-based social networking into a movie recommender system. She showed that those recommendations were more accurate than other techniques when the user's opinions about a film were divergent from the average using the FilmTrust system.

Jyun-Cheng & Chui-Chen [10] presented a recommendation system that used trading relationships to calculate level of recommendation for trusted online auction sellers. They demonstrated that network structures formed by transactional histories could be used to expose such underlying opportunistic collusive seller behaviors. Taking a structural perspective by focusing on the relationships between traders rather than their attribute values, they used two social network indicators to create a collaborative-based recommendation system that could suggest risks of collusion associated with an account. They also tested the system against real world ''blacklist'' data published regularly in a leading auction site and found it able to screen out 76% of the blacklisted accounts.

Yang & Dia [11] proposed a framework that utilized the concept of a social network for the targeted advertising. Their approach discovered the cohesive subgroups from a customer's social network as derived from the customer's interaction data, and used them to infer the probability of a customer preferring a product category from transaction data. That information was then used to construct a targeted advertising system. Cho & Bang [12] proposed a novel recommender system which applied centrality concept. They utilized the analysis of the relationships among products by using centrality concept, and then recommended specific product to the customers who were highly like to buy the product. They used product network from user's purchase data, but didn't use real social networks of humans.

Above studies mostly tried to use SNA in the research area of CRM, however, their model did not systematically integrate SNA with CRM systems or with recommendation algorithms. However, Debnath et al. [13] tried to integrate those two components. They proposed a hybridization of CF and content-based recommendation system. In their system, attributes used for content-based recommendations were assigned weights depending on their importance to users. The weight values were estimated from a set of linear regression equations obtained from a social network graph which captured human judgment about similarity of items. However, since it was just an exploratory study so that there was no empirical validation, the implication of their study was quite limited.

In addition, Liu & Lee [3] developed a way to increase recommendation effectiveness by using social network information with CF. They collected data about user preference scores and their social network relationships. And then, they evaluated recommendation performances with diverse neighbor groups combining groups of friends and nearest neighbors. They resulted more accurate recommendations could be made by incorporating social network information into CF.

However, the approach they used was not based on typical social network analysis. The authors did not create 'network model' of all the participants, rather they just reflected the simple relationship between each of users (called 'the distributor' in the study) and their friends. Thus, strictly to say, their model is based on simple social relationship rather than complex social network.

In addition, their recommendation model arbitrary determined the range of friends because there were no guides for determining it.

Under these backgrounds, we propose a novel recommendation model that uses the characteristics of the network model for all the users, which can be obtained by applying SNA. In particular, we propose a novel approach to integrate SNA, cluster analysis and CF for the CRM applications which systematically determine the range of friends by using the clustering algorithm.

III. CLUSTER-INDEXING CF WITH SNA

As mentioned earlier, the purpose of this study is to develop a novel recommender system which reflects qualitative and emotional information. Typical recommender systems have used CF to analyze user profile. However, CF has a limitation to consider qualitative and emotional information about users. It is owing to the fact that CF predicts user preference score utilizing user-rated evaluation score. In this study, we propose novel recommender systems which use SNA in order to consider qualitative and emotional information to recommendation process.

For this end, this study suggests a new recommendation algorithm, cluster-indexing CF based on SNA. In this study, we use the cores in subgroups from social network as the initial seeds for a conventional clustering algorithm. The detailed explanation on the steps of the recommendation algorithm is as follows:

In the first step, we perform SNA. SNA can show social network between users as a graph and centrality of each user. From the graph of social network, we intuitively find subgroups of users. In addition, the centrality value shows who the core of each subgroup is. In this study, we use one measurement of centrality to find the point of centrality in the partitioned subgroup network extracted from the original social network. Subgroups can provide valuable information in understanding the influence of the subgroup on the original network as a whole and on individual characteristics and status within the subgroup, among subgroups, and the entire network. In this study, we evaluate degree of centrality and select top five users of this value.

In the second step, we perform clustering analysis. There are many kinds of clustering algorithms, this study uses the modified version of k-means clustering algorithm. Typical k-means clustering calculates the similarity between samples using Euclidean distance measure. However, in this study, we need to calculate the similarity between the user preference vectors for items. Many of the elements in the user preference vectors are empty so that it is difficult to calculate Euclidean distance between these vectors. To resolve this problem, this study adopts Pearson correlation rather than Euclidean distance as the measure of the similarity for the k-means clustering.

In addition, we integrate the experimental results of SNA and k-means algorithm by incorporating users with top five degree centrality from SNA as the initial cluster centers in k-means

clustering algorithm. As a result, we can generate five subgroups whose members have the preference scheme that is similar to one of socially leading people. In the final step, we perform CF procedure. This study proposes cluster-indexing CF with SNA. Typical CF computes the similarity between users, and then selects the nearest neighbors for user. This study amplifies impacts of user's friends than those of their strangers in order to reflect qualitative and emotional information to typical CF process. In order to consider social network information into CF, we can simply replace the nearest neighbor from typical CF with new nearest neighbor from SNA. However, this method may be less effective than typical CF because the number of nearest neighbors from SNA is much smaller than that of the other nearest neighbors [3]. Thus, we use method introduced by Liu & Lee [3]. The method amplifies the weights of the social network members' preference (e.g. assigning two-times of the original weight) if they are also in the nearest neighbor set from typical CF, while leaving the weights of other nonsocial network members' preference unchanged.

IV. EXPERIMENTS

A. Experimental Data

In order to validate the usefulness of our proposed recommendation algorithm, we adopt 'empirical validation' that is based on a real-world dataset. Although our proposed model is based on CF, we cannot use public datasets for CF such as MovieLens, or EachMovie because our algorithm requires social network information other than user preference information. Consequently, we build a Web-based system for collecting appropriate data from users.

Our data collection system is designed to collect information on the participants' social relationship with other participants. Especially, we collect the information with whom and how participants have relationship via online social media like instant messenger, social networking web sites like Facebook, Myspace, and Cyworld, and microblog service like Twitter. For the target of the recommendation, we select movies, the most popular item for validating the performance of new recommender systems. So, we collect the participant's preference on top 100 movies of all-time worldwide box office. Also, we collect the basic demographic information on the participants including gender, age, address, and so on.

We collect the data from 91 respondents in one of the major universities in Korea. Among them, we eliminate one case that is seemed to be distorted, and finally select 90 respondents and their online relationships and the ratings for 100 movies as an experimental dataset.

B. Experimental Design

For implementing SNA-based CF algorithm, we combine the results from two independent experiments. The first experiment is done for doing SNA. By using a commercial software – NetMiner version 3.0, we adopt SNA to the experimental dataset, and calculate the degree centrality values of all 90 participants.

The second experiment is designed to implement modified k-means clustering, and CF using the result of the clustering. For this experiment, we develop an experimental system using VBA (Visual Basic for Application) of Microsoft Excel 2007. In the case of modified k-means clustering, we design to perform k-means clustering based on a new similarity measure – Pearson correlation between user preference vectors. For CF, we adopt 'all-but-one' approach. This means all but one rating of the user is given and our algorithm predicts the rating for the left one.

C. Experimental Results

As mentioned, we apply SNA first to the experimental dataset before applying modified k-means and CF. Here, we calculate the degree centrality of all the participants. Based on this result, we succeed to find 5 online social leaders, whose the degree centrality values was bigger than 3.5. And, we use them as the centroids of the modified k-means clustering.

In this study, we set the average MAE (mean absolute error) as the criterion for evaluating performances of the comparative models. The MAE is frequently used in CF literature, and represents the difference between the predicted and actual rating of users [14]–[16]. Average MAE can be defined as (1).

$$Avg.\,MAE = \left\{ \sum_{k=1}^{N} \left(\sum_{i=1}^{n} \left| p_{k,i} - a_{k,i} \right| \middle/ n \right) \right\} \middle/ N \tag{1}$$

where N is the number of users in the dataset T, n is the number of items in the dataset T, $p_{k,i}$ is the predicted ratings of user k for the item i, and $a_{k,i}$ is the actual ratings of user k for the item i.

Table I presents results of the conventional CF model and our proposed model.

TABLE I
THE RESULTS OF THE EXPERIMENTAL MODELS

Model	AVG of MAE	STDEV of MAE
Conventional CF	0.8260	0.2845
The proposed model	0.8224	0.2851

As shown in the Table 1, our proposed model shows small average MAE than the conventional CF model. Thus, we may conclude that our model generates more accurate prediction results in the recommendation for the users.

V. CONCLUSIONS

In this study, we proposed the integrated model of typical CF and social network information from social network analysis. The results show that the proposed approach is effective in CF methodologies for enhancing recommendation performance. The reason of the results may be resulted from that CF reflects qualitative and emotional information from SNA.

Although this study got better recommendation performance than typical CF, the performance may be raised if we incorporated the social network information from social networking Web sites directly because our experiments were executed in a laboratory environment. In addition, we evaluated only 90 users for the proposed model. Although we showed statistically significant difference between the proposed model and typical CF, we might get more reliable results than this if we collected sufficient samples. Moreover, we arbitrarily set the number of online social leaders to five, and then we also set the number of clusters of clustering analysis to five because the online social leaders were used as initial cluster center. Unfortunately, there have been few studies to propose any mechanism to determine the optimal number of leaders in SNA or clusters in clustering algorithm, so it has usually been determined by heuristics. Thus, the attempts to adjust the number of leaders or clusters should be one of the focuses of future research.

REFERENCES

[1] Burke, R (2002) Hybrid recommender systems. User Modeling and User-Adapted Interaction 12(4):331-370
[2] Resnick, P, Iacovou, N, Suchak, M, Bergstrom, P, Riedl, J (1994) GroupLens: An open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina
[3] Liu, F, Lee, HJ (2010) Use of social network information to enhance collaborative filtering performance. Expert Systems with Applications 37:4772-4778
[4] Shinha, R, Swearingen, K (2001) Comparing recommendations made by online systems and friends. In: Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries, Dublin, Ireland
[5] Kautz, H, Selman, B, Shah, M (1997) Referral Web: combining social networks and collaborative filtering. Communications of the ACM 40(3):63-65
[6] Liu, H, Maes, P (2005) InterestMap: Harvesting social network profiles for recommendations. In: Proceedings of the IUI, San Diego, CA
[7] Ziegler, C, Lausen, G (2004) Analyzing correlation between trust and user similarity in online communities. In: Proceedings of the 2nd International Conference on Trust Management
[8] Ryu, Y, Kim, HK, Cho, YH, Kim, JK (2006) Peer-oriented content recommendation in a social network. In: Proceedings of the 16th Workshop on Information Technologies and Systems
[9] Golbeck, J (2006) Generating predictive movie recommendations from trust in social networks. In: Proceedings of the iTrust2006, pp 93-104
[10] Jyun-Cheng, W, Chui-Chen, C (2008) Recommending trusted online auction sellers using social network analysis. Expert Systems with Applications 34:1666-1679
[11] Yang, W, Dia, J (2008) Discovering cohesive subgroups from social networks for targeted advertising. Expert Systems with Applications 34:2029-2038
[12] Cho, Y, Bang, J (2009) Social network analysis for new product recommendation. Journal of Intelligence and Information Systems 15(4):177-193
[13] Debnath, S, Granguly, M, Mitra, P (2008) Feature weighting based recommendation system using social network analysis. In: Proceeding of WWW 2008, pp 1041-1042
[14] Breese, J, Heckerman, D, Kadie, C (1998) Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, San Francisco, CA, pp 43-52
[15] Sarwar, BM, Konstan, JA, Borchers, A, Herlocker, J, Miller, B, Riedl, J (1998) Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system. In: Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW), Seattle, Washington, USA, pp 345-354
[16] Goldberg, K, Roeder, T, Gupta, D, Perkins, C (2001) Eigentaste: A constant time collaborative filtering algorithm. Information Retrieval Journal 4(2):133-151