Personalized Recommendation Based on Collaborative Filtering with Social Network Analysis

Joong Hee Jeong and Jong Woo Kim[+]

School of Business, Hanyang University, Seoul, Korea

Abstract. Collaborative Filtering (CF) is recommendations technique that provides personalized recommendation of services and products to customers by understanding preference through similarity between customers. The network of customers which can be made based on the information about customer's visit information can be used to increase the effect of recommendation. In this study, it suggests a CF based recommendation method that uses network centrality measures of customers with similarities of customers in CF. The usefulness of the proposed method is tested using user visiting log data of a representative UCC (User Created Contents) site. The experimental results show that the combined usage of SNA (Social Network Analysis) measures with similarity measures provides better recommendation performance than traditional CF method.

Keywords: Collaborative Filtering, Recommend Systems, Social Network Analysis, etc.

1. Introduction

Recommendation systems provide personalized recommendation on Internet storefronts and reduce information overload from countless Internet contents which is increasing exponentially [1,3,4]. As a representative recommendation technique, Collaborative Filtering (CF) algorithm is used to recommend personalized services and products by understanding customers' preference from similarities between customers [1,3]. SNA (Social Network Analysis) is emerging area in information systems research and provides metrics to understand the roles of objects in various networks. In the context of personalized recommendation, the information from customers' behavior on Internet storefronts can be used to generate networks of customers, and it can be sources to generate SNA (Social Network Analysis) metrics such as degree centrality, betweenness centrality, and hub centrality which indicate the role of customers in the network [7]. In this paper, we tries to integrate SNA metrics, especially centrality measure with CF algorithm to improve recommendation performance.

2. Related Works

2.1. Collaborative Filtering

Collaborative Filtering (CF) can predict a particular user's evaluation of a particular item (products or services) by using similar user's preference values measured [2]. There are two kinds of CF methods. User-based CF uses preferences between users and Item-based CF uses preferences between items.

In this paper, we focus on User-based CF. In User-based CF, the best products for a target customer are selected based on the neighbor customers' preferences. The neighbors of a target customer are determined

---

[+] Corresponding author. Tel.: + 82-2-2220-1067 fax: +82-2-2220-1169
 E-mail address: kjw@hanyang.ac.kr

based on behavioral similarities such as purchasing or visiting web pages [2,3]. If the data on customer preferences is binary data, the similarity between customers is obtained through the Jaccard measure (1).

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} \tag{1}$$

In general, the recommendations process of User-based CF can be divided into three phases including (1) customer-to-customer matrix generation based on customer behavior data, (2) searching neighborhoods, and (3) prediction of preference scores of alternative items. The prediction of item rating scores of a particular user is calculated through the following equation (2) when the rating scores have numeric values. In the case of binary values, the formula (3) is used instead of equation (2).

$$P_{ik} = \overline{S_i} + \frac{\sum_{l \in Rater(k)}(S_{ik} - \overline{S_i})r_{il}}{\sum_{l \in Rater(k)}|r_{il}|} \tag{2}$$

$$P_{ik} = \frac{\sum_{l \in Rater(k)}J(i,l)}{n(Rater(k))} \tag{3}$$

2.2. Social Network Analysis

A social network is a graph with nodes of individuals and edges which represent specific relationships between the individuals [6]. Social Network Analysis (SNA) provides metrics of social networks or visual representation of social networks [8,9]. Recently, there are some studies on recommendation techniques using social network analysis [5]. However, these have some limits that they require more information for making networks or uses solely network metrics without considering conventional similarity measures of CF. In this study, we propose the use of centralities metrics of nodes with CF user similarity measures.

3. Proposed Approch

In our approach, we try to change prediction formula (3) to formula (4) which include cl is a centrality metrics of a rater l. That is, to predict preference of an item of a particular customer, centrality metrics such degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality are collaboratively used with traditional similarity metrics.

$$P_{ik} = \frac{\sum_{l \in Rater(k)}(w_1 J(i,l) + w_2 c_l)}{n(Rater(k))}, w_1 + w_2 = 1 \tag{4}$$

4. Experimental Design

4.1. Data

To compare the performance of our proposed approach and traditional CF method, YouTube web page visiting data of panels during February 2010 is obtained from KoreanClick©. The web page visiting data consists of panel ID, URL, and visiting duration. The total number of records is 15,496. There are 9,998 UCCs are appeared in the panel data. We filter records of 239 users who had visited more than 10 times in the monthly URL information. After filtering, there are 7,967 UCCs which is appeared more than one time. Proposed methods are implemented by JAVA programming language with MS SQL Server 2005.

4.2. Experiment procedure

The experiment procedure of this study is shown in Fig.1. Starting from User-UCC matrix, User-to-User matrix is generated. Based on the matrix, prediction of user preferences is performed using traditional CF method. Also, from User-UCC matrix, User Network is generated. The nodes of User Network is users and the edges are defined when two users saw the same UCC. From the User Network, centrality metrics are calculated, the centrality metrics in the study include degree centrality, closeness centrality, betweenness

centrality, and eigenvector centrality. We try to calculate predicted preferences using formula (4) with various weights.
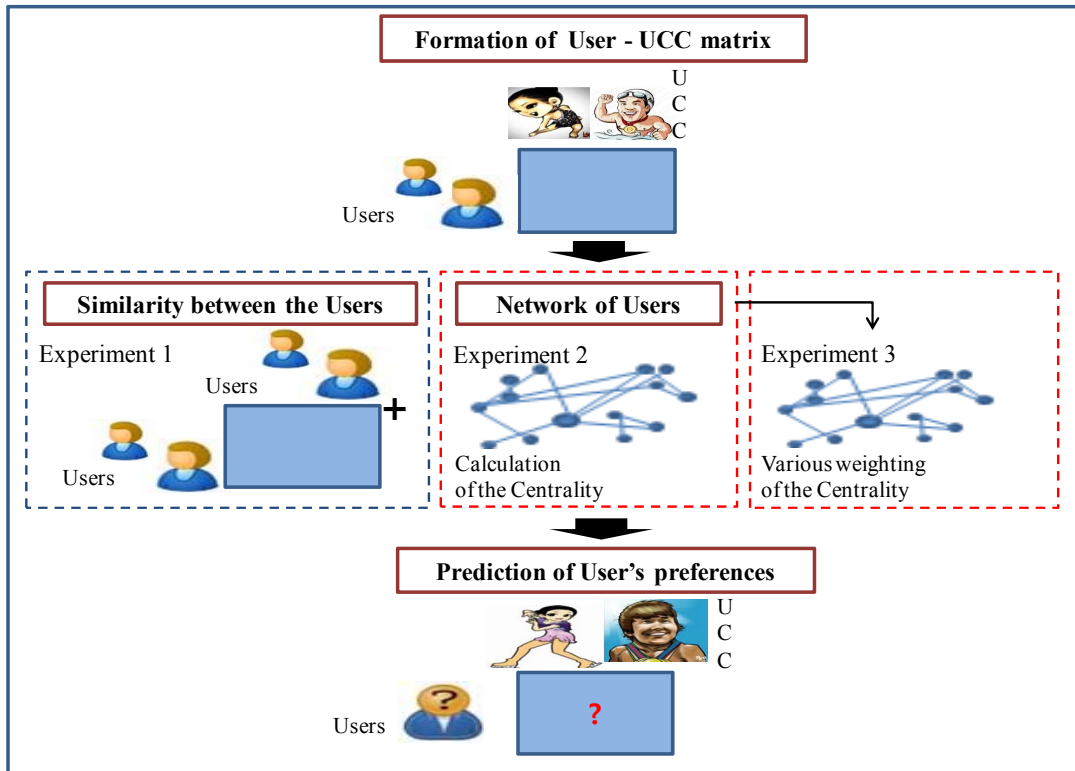

Fig. 1: Experimental Procedure

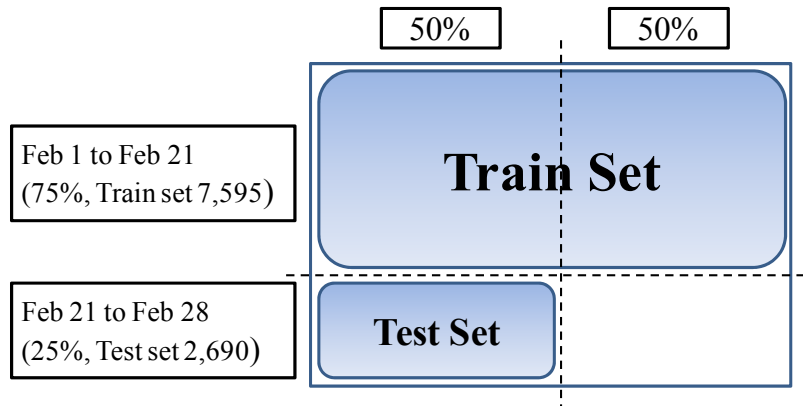The experimental data set is decomposed as training data set and test dataset as shown in Fig.2.

Fig. 2: Train set and Test set

5. Results

Tab.1 shows the experimental results of 6 methods, (1) random selection, (2) traditional CF, (3) CF with degree centrality, (4) CF with closeness centrality, (5) CF with betweenness centrality, and (6) CF with eigenvector centrality. The results is came with weight 0.1 in formula (4). Four results of propose approach are better than traditional CF and random selection. CF with degree centrality shows the highest performance. The next is CF with eigenvector centrality and CF with closeness centrality.

Tab.2 shows the changes of performance depending on the change of weights. When the weight is 0.1, all of four methods show the best performance.

Table. 1: Comparison of recommendation performance

Experiment		Precision		
		Top 5	Top 10	Top 20
1	Random	0.000196	0.000196	0.000196
2	CF	0.006101	0.005469	0.003507
3	CF + Degree Centrality	0.013474	0.009766	0.005962
4	CF + Closeness Centrality	0.008576	0.005813	0.004419
5	CF + Betweenness Centrality	0.007929	0.006101	0.003969
6	CF + Eigenvector Centrality	0.010415	0.007959	0.005644

(Each centrality's weight is 0.1)

Table. 2: Recommendation performance changing with weights

Experiment	Precision	Experiment	Precision
	Top 10		Top 10
CF×0.9 + D.C×0.1	0.00833	CF×0.9 + C.C×0.1	0.00417
CF×0.8 + D.C×0.2	0.00694	CF×0.8 + C.C ×0.2	0.00417
CF×0.7 + D.C ×0.3	0.00556	CF×0.7 + C.C×0.3	0.00000
CF×0.6 + D.C ×0.4	0.00417	CF×0.6 + C.C×0.4	0.00000
CF×0.5 + D.C ×0.5	0.00417	CF×0.5 + C.C×0.5	0.00000
CF×0.4 + D.C ×0.6	0.00000	CF×0.4 + C.C×0.6	0.00000
CF×0.3 + D.C ×0.7	0.00000	CF×0.3 + C.C×0.7	0.00000
CF×0.2 + D.C ×0.8	0.00000	CF×0.2 + C.C×0.8	0.00000
CF×0.1 D.C ×0.9	0.00000	CF×0.1 + C.C×0.9	0.00000
Only D.C	0.00000	Only C.C	0.00000
CF×0.9 + B.C×0.1	0.00556	CF×0.9 + E.C×0.1	0.00694
CF×0.8 + B.C×0.2	0.00417	CF×0.8 + E.C×0.2	0.00694
CF×0.7 + B.C×0.3	0.00417	CF×0.7 + E.C×0.3	0.00139
CF×0.6 + B.C×0.4	0.00417	CF×0.6 + E.C×0.4	0.00139
CF×0.5 + B.C×0.5	0.00417	CF×0.5 + E.C×0.5	0.00000
CF×0.4 + B.C×0.6	0.00278	CF×0.4 + E.C×0.6	0.00000
CF×0.3 + B.C×0.7	0.00278	CF×0.3 + E.C×0.7	0.00000
CF×0.2 + B.C×0.8	0.00000	CF×0.2 + E.C×0.8	0.00000
CF×0.1 + B.C×0.9	0.00000	CF×0.1 + E.C×0.9	0.00000
Only B.C	0.00000	Only E.C	0.00000

(D.C : Degree Centrality, C.C : Closeness Centrality,
B.C : Betweenness Centrality, E.C : Eigenvector Centrality)

6. Conclision

In this study, we try to integrate SNA (Social Network Analysis) measures with similarity measure of CF to improve recommendation performance. From the experiments using UCC visiting panel data, the proposed approach show better performance than traditional CF. Especially, CF with degree centrality shows better performance than CF with other centrality metrics. Also, we try to compare the performance of methods with changes of weighting factor of centrality and similarity. Lower weighting on centrality shows best performance than higher weighting on centrality.

To generalize the result, we need to repeat experiments in order to test experimental results statically. Also, we need to investigate the effect of changes of weighting factors more detail. Other experiments with other data set are also benefit to generalize the results.

7. References

[1]   H. Kautz, B. Selman and M. Shah, "Referral Web: Combining Social Networks and Collaborative Filtering," Communications of the ACM, Vol. 40, No. 3, 1997, pp.63-65.

[2] J. K. Kim, J. Seo, D. H. Ahn and Y. H. Jo, "A Study on Personalized Product Recommendation Method Development Using Cooperative Filtering Method", Korea Intelligent Information System Society, Vol. 8, No. 2, 2002, pp.139-157.

[3] J. Moody and D. R. White, "Structural Cohesion and Embeddedness: A Hierarchical Concept of Social Groups", American Sociological Review, Vol.68, 2003, pp.103-127.

[4] J. Mostafa, S. Mukhopadhyay, W. Lam and M. Palakal, "A Multilevel Approach to Intelligent Information Filtering: Model, System, and Evaluation", ACM Transactions on Information Systems, Vol.15, No.4, 1997, pp.368-3993.

[5] P. Morville, Ambient Findability, Hanbit Media, Inc., 2006.

[6] P. Y. K. Chau, M. Cole, A. P. Massey, M. Montoya-Weiss and R. M. O'Keefe, "Cultural Differences in The Online Behavior of Consumers", Communications of the ACM, Vol.45, Vol. 10, 2002, pp.138-143.

[7] U. Brandes and D. Wagner, "Analysis and Visualization of Social Networks", http://www.inf.uni-konstanz.de/algo/publications/

[8] Y. H. Kim, Social Network Analysis, PAKYOUNGSA, 2003.

[9] Y. J. Jeong, Yu-Jin, Jeong's Web 2.0 planning theory, Hanbit Media, Inc., 2006.