

Recommender System using Implicit Trust-enhanced Collaborative Filtering

Kyoung-jae Kim
Business School, Dongguk University_Seoul
(*kjkim@dongguk.edu*)

Youngtae Kim
Department of Management Information
Systems, Graduate School, Dongguk
University_Seoul
(*kytsky@naver.com*)

.....

Personalization aims to provide customized contents to each user by using the user's personal preferences. In this sense, the core parts of personalization are regarded as recommendation technologies, which can recommend the proper contents or products to each user according to his/her preference. Prior studies have proposed novel recommendation technologies because they recognized the importance of recommender systems. Among several recommendation technologies, collaborative filtering (CF) has been actively studied and applied in real-world applications. The CF, however, often suffers sparsity or scalability problems. Prior research also recognized the importance of these two problems and therefore proposed many solutions. Many prior studies, however, suffered from problems, such as requiring additional time and cost for solving the limitations by utilizing additional information from other sources besides the existing user-item matrix. This study proposes a novel implicit rating approach for collaborative filtering in order to mitigate the sparsity problem as well as to enhance the performance of recommender systems. In this study, we propose the methods of reducing the sparsity problem through supplementing the user-item matrix based on the implicit rating approach, which measures the trust level among users via the existing user-item matrix. This study provides the preliminary experimental results for testing the usefulness of the proposed model.

.....

Received : December 10, 2013 Accepted : December 20, 2013

Type of Submission : Concise Paper Corresponding author : Kyoung-jae Kim

1. Introduction

Although users have enjoyed convenience

with the activation of electronic commerce, they have also suffered from excess information from the Web. Personalization techniques can remedy

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2012S1A5A2A01020119).

these difficulties. The personalization aims for providing customized contents to each user by using his/her personal preferences. Thus, the core parts of personalization include recommendation technologies, which can recommend the proper contents or products to each user by using each user's preference. Prior studies have proposed novel recommendation technologies because they recognized the importance of recommender systems. Among several recommendation technologies, collaborative filtering (CF) has been actively studied and applied in real-world applications.

CF can recommend products to users by using other similar users' results of evaluation for each product. To this end, we prepare the user-item matrix, which includes each user's rating for each product. Finally, we can recommend products through the computation of each user's evaluation similarity based on the user-item matrix. As aforementioned, many researchers have attempted to study CF due to its recommendation performance. However, there are some major limitations in CF, such as the sparsity (or cold-start user or product problem) and scalability problems.

In particular, the sparsity problem is engendered by the sparse distribution of evaluation rating within the user-item matrix because all users cannot evaluate all products. Thus, there exist problems where some customers and products are excluded from the recommendation when data sparsity is severe. This problem is related to the cold-start user or product problem.

The cold-start user or product problem deals with the issue in which the system cannot draw any inferences for users or products as to which it has not yet gathered sufficient information, such as evaluation rating scores (http://en.wikipedia.org/wiki/Cold_start). Due to the problems, CF fails to consider users or products for which no-products or no-one has rated previously. Therefore, the sparsity problem is one of the most important factors for recommendation quality when there are many users and products.

Prior research recognized the importance of these two problems and as a result, proposed many solutions. Many prior studies, however, had suffered from problems, such as requiring additional time and cost for solving the limitations by utilizing additional information from other sources besides the existing user-item matrix.

Some researchers have proposed the concept of trust-based recommender systems, which enable users to understand the sources of recommendation, which were produced via the trust computation process. They usually utilized a trust network, which is a social network with trust ratings, and use the knowledge to generate recommendations. A trust network is a directed graph that includes the nodes and edges. The nodes indicate users and the edges are formed when there is a direct trust relation between two extreme users. In addition, the trust propagation process should be executed when there is no direct trust relation between users, but when there is a possible indirect trust relation. It is possible to establish indirect trust relations between users

who have no direct trust by using direct trust relations.

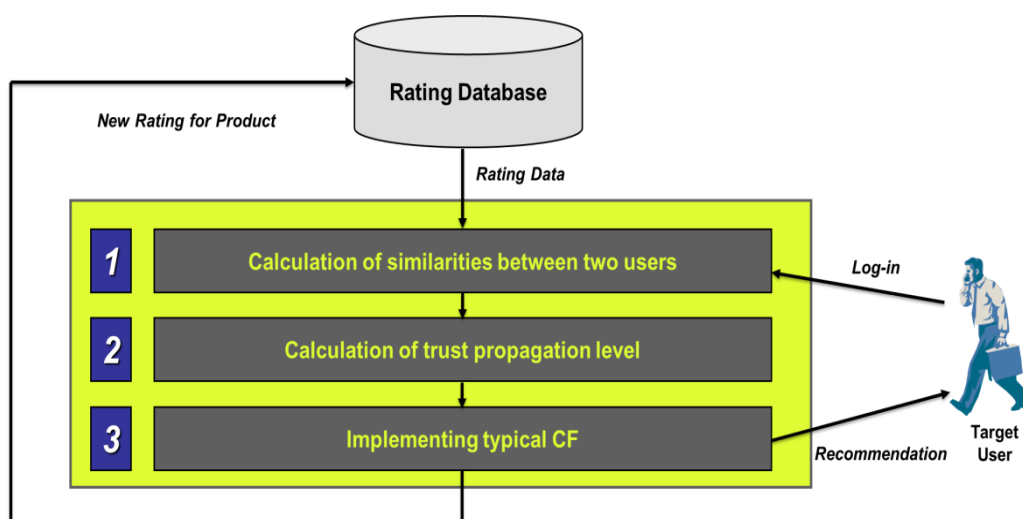
Shambour and Lu (2011) refer this trust enhanced filtering approach to trust-based filtering. The methods of trust-based filtering can be classified into two major classes, including explicit and implicit trust-based filtering approaches. The former usually gets trust values by using an explicit trust network, which is obtained by the direct input of the users. This approach is easier than the latter for obtaining the trust values of the users; however, it cannot consider the implicit trust value because it only uses the explicit one. The latter, however, gets trust values via an implicit trust network, which is deduced by the trust values between users based on their items ratings. Thus, the implicit approach obtains more trust values than the explicit one.

In this study, we propose the methods of reducing the sparsity problem through supplementing the user-item matrix based on the implicit trust score, which measures the trust level among users via the existing user-item matrix.

This paper consists of five sections. In Section 2, we propose a novel implicit trust-based filtering algorithm for the recommendation. Section 3 introduces the experimental data and design of experiments incorporated in the study. The experimental results will also be presented in this section. Section 4 discusses the findings and limitations of our study.

2. Implicit Rating Approach

The purpose of this study is to remedy the sparsity problems in CF using implicit trust-based filtering. As aforementioned, implicit trust-



<Figure 1> Three Stages in the Process of Implicit Trust-Based Filtering

based filtering indirectly estimates the trust score using the existing user-item matrix, whereas explicit trust-based filtering needs additional user rating data. In general, CF is an evaluation method that uses the scores for the co-rated items in order to evaluate the similarity between two users. Thus, we cannot evaluate the similarity between two users when there are no co-rated items. Finally, CF lowers the quality of recommendation because it causes the sparsity problem or the cold-start user problem. Implicit trust-based filtering basically evaluates the trust level between two users by using the similar method of a typical CF; however, it can estimate the similarity between two users when there are no co-rated items through using the scores for different items from two users. It evaluates the implicit trust level for two users and also enables an evaluation of new users by estimating the trust propagation level between two users with no co-rated items for this end. <Figure 1> shows three stages in the process of implicit trust-based filtering.

- **Stage 1 : Calculation of similarities between two users**

A typical CF can estimate similarities between users by using the evaluation scores from the user-item matrix, cosine vector or Pearson's correlation coefficient. In general, the Pearson's correlation coefficient is popularly used; further, the similarity between user X and user Y can be calculated by using Equation (1).

$$S(X, Y) = \frac{\sum_{i=1}^n (R_{x,i} - \bar{R}_x) \times (R_{y,i} - \bar{R}_y)}{\sqrt{\sum_{i=1}^n (R_{x,i} - \bar{R}_x)^2} \times \sqrt{\sum_{i=1}^n (R_{y,i} - \bar{R}_y)^2}} \quad (\text{where } -1 \leq S(X, Y) \leq 1) \quad (1)$$

In Equation (1), $S(X, Y)$ denotes the similarity between active user X and each of the other user Y, who have the co-rated items with active user X, i is the index of each item that both user X and user Y have rated, $R_{x,i}$ is the rating of user X for item i , $R_{y,i}$ is the rating of user Y for item i , \bar{R}_x is the average rating of user X, and \bar{R}_y is the average rating of user Y. From Equation (1), we can understand that a typical CF basically estimates the similarities based on the evaluation scores when co-rated items exist between users; thus, we cannot estimate the similarities when there is no co-rated item.

- **Stage 2 : Calculation of trust propagation level**

As mentioned earlier, a typical CF can calculate the similarities when co-rated items exist between two users. Thus, the user who has a similar preference with the other user cannot be considered as a similar user because there is no co-rated item. For example, when users X, Y and Z exist, let us suppose that X and Y, and Y and Z have co-rated items, and X and Z does not have a co-rated item. In this case, we can calculate the similarities between two users for users X and Y, and Y and Z respectively; however, we cannot compute them for X and Z be-

cause there is no co-rated item between users X and Z. However, the possibility of similarity between users X and Z is high when users X and Y, and users Z and Y are similar users, respectively, because we can assume user Y as a mediator between users X and Z.

Therefore, we can remedy the limitations of a typical CF and reduce the sparsity of the user-item matrix if we conclude user Z as a similar user for user X, because user Z is similar to mediator Y, who is a similar user of user X. Thus, we need a similarity estimation method using a mediator for two users who have no co-rated items. As mentioned earlier, we can compute each similarity using Equation (1), when there are users X, Y and Z, and when users X and Y, and users Y and Z have co-rated items, respectively, and when there is no co-rated item between users X and Z. Thus, we can calculate the similarity between users X and Y using Equation (1) as well as the similarity between users Y and Z using Equation (2).

$$S(Y, Z) = \frac{\sum_{i=1}^n (R_{y,i} - \bar{R}_y) \times (R_{z,i} - \bar{R}_z)}{\sqrt{\sum_{i=1}^n (R_{y,i} - \bar{R}_y)^2} \times \sqrt{\sum_{i=1}^n (R_{z,i} - \bar{R}_z)^2}} \quad (\text{where } -1 \leq S(Y, Z) \leq 1) \quad (2)$$

Next, we can compute the similarity between users X and Z ($S(X, Z)$) using Equation (3), considering the already computed two similarities ($S(X, Y)$ and $S(Y, Z)$) and the proportions of the number of evaluated items and the number of co-rated ones for each pair of users.

$$S(X, Z) = S(X, Y) \times \frac{n(R_x \cap R_y)}{n(R_x \cap R_y) + n(R_y \cap R_z)} + S(Y, Z) \times \frac{n(R_y \cap R_z)}{n(R_x \cap R_y) + n(R_y \cap R_z)} \quad (3)$$

In Equation (3), $n(R_i \cap R_j)$ denotes the number of co-rated items between users i and j . This implies the value of the weighted sum of similarities for each pair of users according to the proportion of co-rated items for each pair of users.

Although the trust propagation used for users lay across several arcs, however, it must be controlled within the proper level because over-propagation increases computation and decreases efficiency. In addition, multiple trust paths are discovered to estimate the implicit trust level between two users. For example, there may exist two trust paths ($A \rightarrow B \rightarrow C$ and $A \rightarrow D \rightarrow C$) for computing the implicit trust level between users A and C when users A and B, B and C, and A and D have co-rated items. In a real-world application, it is possible to find multiple paths for the trust path. Further, the values of the implicit trust level for multiple trust paths may be different.

Papagelis et al. (2005) suggested that the choice of trust path is possible with *Path Composition* and *Path Selection*. They classified the path composition approach as *Average Composition* and *Weighted Average Composition*. In addition, they distinguished two approaches for the path selection as *Selection Based on Path Maximum Confidence* and *Selection Based on Mini-*

imum Mean Absolute Deviation (MAD). *Average Composition* uses the average of the trust calculated by each of the alternative paths. *Weighted Average Composition* computes the weighted average of all the trust values that are inferred by the alternative paths with weights of the propagated confidence of each calculated association between two users. *Selection Based on Path Maximum Confidence* uses a path which has the highest confidence value. *Selection Based on Minimum Mean Absolute Deviation* chooses a path which has the lowest mean absolute deviation. A more detailed explanation can be found in Papagelis et al. (2005).

- **Stage 3 : Implementing typical CF**

After the implicit trust level is computed, the implicit neighbor is determined as the nearest neighbor of a specific user and is included for the product recommending process if it is determined as a high level of determination of the nearest-neighbor process.

3. Experiments

3.1 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. Our experiments are run on a famous open dataset, the MovieLens2 datasets. MovieLens is originally a web-based recommender system for movies, which was developed by the GroupLens Research Project at the

University of Minnesota. However, it is also an experimental data source and a framework for researching user interface issues related to recommender systems. The dataset contains simple demographic information for users (age, gender, occupation and zip code) as well as movie ratings. The ratings are on a numeric five-point scale, with 1 and 2 representing negative ratings, 4 and 5 representing positive ratings, and 3 indicating ambivalence. We sample a user-item matrix to test the feasibility of the proposed model. By applying the ‘random sampling’ technique, we extract 50 users and 20 items from the dataset.

3.2 Experimental Design

For implementing the proposed algorithm, we develop an experimental system using VBA (Visual Basic for Application) of Microsoft Excel 2010. For CF, we adopt the ‘all-but-one’ approach. This means that all but one rating of the user is given; moreover, our algorithm predicts the rating for the remained one. We implement two implicit rating based models as well as a conventional CF. We use the Average Composition Method from the Path Composition Approach and Selection based on the Path Maximum Confidence Method from the Path Selection Approach.

3.3 Experimental Results

As mentioned, we apply the proposed implicit rating models and the conventional CF al-

<Table 1> Results of the Experimental Models

	Conventional CF	Average composition	Maximum confidence
Top 3	1.050998	0.722247	0.660794
Top 5	0.723111	0.669036	0.610706
Top 10	0.663145	0.642443	0.593323
All-inclusive	0.585337	0.622255	0.582468

gorithm to the experimental dataset. For the nearest neighbor selection process in CF, we use four different approaches, such as “All-inclusive,” “Top 3,” “Top 5” and “Top 10.”

In this study, we set the average MAE (mean absolute error) as the criterion for evaluating the 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 (Breese et al., 1998; Sarwar et al., 1998; Goldberg et al., 2001). Average MAE can be defined as Equation (4).

$$Avg.MAE = \left\{ \sum_{k=1}^N \left(\sum_{i=1}^n |p_{k,i} - a_{k,i}| / n \right) \right\} / N, \quad (4)$$

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 item i .

<Table 1> presents the results of the conventional CF model and our proposed model.

As shown in the <Table 1>, our proposed model (Maximum confidence with All-inclusive) shows the smallest average MAE among the

comparative models. In addition, Maximum confidence model always outperforms other models regardless of the various number of nearest neighbors. Moreover, Average composition model, the other proposed model in this study, always performs better than Conventional CF with various number of nearest neighbors except for case of “All-inclusive.” Thus, we may conclude that our model generates more accurate prediction results in the recommendation for the users.

4. Conclusions

In this study, we propose a novel implicit rating approach for collaborative filtering in order to mitigate the sparsity problem as well as to enhance the performance of recommender systems. In this study, we propose the methods of reducing the sparsity problem through supplementing the user-item matrix based on the implicit trust score, which measures the trust level among users, by utilizing the existing user-item matrix. The preliminary experimental results demonstrate the usefulness of the proposed model.

As aforementioned, we evaluated only 50 users and 20 items for the proposed model. Al-

though this study attained a better recommendation performance compared to the conventional CF, we might be able to obtain more reliable results if we can collect a sufficient amount of samples.

References

- Breese, J., D. Heckerman and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, (1998), 43~52.
- Goldberg, K., T. Roeder, D. Gupta and C. Perkins, "Eigentaste: A constant time collaborative filtering algorithm," *Information Retrieval Journal*, Vol.4, No.2(2001), 133~151.
- Papagelis, M., D. Plexousakis and T. Kutsuras, "Alleviating the sparsity problem of collaborative filtering using trust inferences," *Proceeding iTrust'05 Proceedings of the Third international conference on Trust Management*, (2005), 224~239.
- Sarwar, B. M., J. A. Konstan, A. Borchers, J. Herlocker, B. Miller and J. Riedl, "Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system," *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work*, (1998), 345~354.
- Shambour, Q. and J. Lu, "A hybrid trust-enhanced collaborative filtering recommendation approach for personalized government-to-business e-services," *International Journal of Intelligent Systems*, Vol.26, No.9(2011), 814~843.

Abstract

내재적 신뢰가 강화된 협업필터링을 이용한 추천시스템

김경재* · 김영태**

개인화는 개인적인 기호를 바탕으로 각 사용자에게 맞춤형 콘텐츠를 제공하는 것을 목표로 한다. 이러한 관점에서, 개인화의 핵심적인 부분은 각 사용자의 기호에 적합한 콘텐츠나 상품을 추천할 수 있는 추천기술이라 할 수 있다. 선행연구들은 추천시스템의 중요성을 인지하고 새로운 추천기술을 제안하여 왔다. 여러 추천기술들 중에서 협업필터링은 실무에서 활발하게 연구되고 활용되어 왔다. 그러나, 협업필터링은 종종 희박성 또는 확장성 문제를 겪게 된다. 선행연구들 역시 이 두 가지 문제점의 중요성을 인지하고 그에 대한 여러 가지 해결방안들을 제안하였다. 하지만, 여러 선행연구들은 기존의 사용자-상품 매트릭스 외에 다른 원천들로부터 생성된 추가적인 정보를 이용함으로써 문제점들을 해결하려 함으로 인하여 추가적인 시간과 비용을 요하는 다른 문제를 야기하였다. 본 연구에서는 희박성 문제를 완화하고 추천시스템의 성능을 개선하기 위하여 협업필터링을 위한 새로운 내재적 평가방법을 제안한다. 즉, 본 연구에서는 기존 사용자-상품 매트릭스를 이용하여 사용자 간의 신뢰수준을 측정할 수 있는 내재적 평가법에 기반한 사용자-상품 매트릭스의 보완을 통해 희박성 문제를 완화할 수 있는 방안을 제안한다. 또한, 본 연구에서는 제안하는 방안의 유용성을 평가하기 위한 탐색적 실험 결과를 제공한다.

Keywords : 내재적 평가, 희박성, 추천시스템, 협업필터링, 고객관계관리

* 동국대학교_서울 경영학부 교수

** 동국대학교_서울 일반대학원 경영정보학과 석사과정

저 자 소개



김경재

현재 동국대학교 경영대학 경영학부 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공으로 박사학위를 취득하였으며, *경영학연구*, *지능정보연구*, *Annals of Operations Research*, *Applied Intelligence*, *Applied Soft Computing*, *Asia Pacific Journal of Information Systems*, *Computers and Operations Research*, *Computers in Human Behavior*, *Expert Systems*, *Expert Systems with Applications*, *Information, Intelligent Data Analysis*, *International Journal of Electronic Commerce*, *Intelligent Systems in Accounting, Finance and Management*, *Neural Computing and Applications*, *Neurocomputing* 등의 학술지에 논문을 게재하였다. 연구 관심분야는 고객관계관리, 데이터마이닝, 비즈니스 인텔리전스, 지식경영 등이다.



김영태

현재 동국대학교 일반대학원 경영정보학과 석사과정에 재학 중이다. 동국대학교 컴퓨터멀티미디어학부에서 공학사를 취득하였다. 연구 관심분야는 고객관계관리, 사회연결망분석, 데이터마이닝, 소셜네트워크 서비스(SNS) 등이다.