[1]Seok-Woo Jang
[2]Sang-Hong Lee

Implementation of Book Recommendation System Using Collaborative Filtering

Journal of Electrical Systems

Abstract: - Recently, with the development of online commerce, many content-based services are being used by many people. In such services, recommendations are of great importance. Through the recommendation system, the user's choice is understood and customized products and content are provided to attract the user's interest. Among the techniques used in the existing system, the most representative one is collaborative filtering. Collaborative filtering provides recommendations using users' rating data. This has the problem that recommendations are not made for items without ratings. In addition, there is a phenomenon in which the performance of the algorithm for big data is greatly reduced. Therefore, in this paper, we develop an algorithm to make recommendations for existing new items and optimize the performance of the algorithm for big data.

Keywords: big data, recommendation system, collaborative filtering, content-based services.

I. INTRODUCTION

Recently, as the use of the Internet has become widespread, service providers have been able to monitor customer information in real time, and various personalized services have been provided. Personalized services refer to providing products or services that customers need without explicitly asking them (Mulvenna et al. 2000). Customization and personalized services are recognized as important success factors for Internet service providers [1]. Since personalized services provide services based on information about individuals, they are effectively implemented when information exchange between service providers and individuals is smooth. Recently, as the use of the Internet has become widespread, service providers have been able to monitor customer information in real time, and various personalized services have been provided. The most important of these personalized services is the recommendation service. Recommendation refers to a visible form of personalization in which information or services are tailored to the needs of individuals or specific groups, and a system that performs this function is called a recommender system. A recommender system not only selects what users should pay attention to, but also filters out unnecessary information by searching through a complex information network. Recommendation systems are already being used in many fields and are considered successful systems [2].

Various companies overseas are actively utilizing recommendation systems. Amazon, the largest online bookstore in the United States, has its own recommendation system called A9. About 35% of Amazon's sales are generated from recommended products, and it is generating a lot of sales through recommendation systems. In addition, YouTube has become the world's largest online streaming site by showing videos that consider individual tastes through personalized recommendation systems and collaborative filtering [3].

Recommendation systems can be implemented through various techniques, and among the techniques recently used in the e-commerce, collaborative filtering is a representative one. Collaborative filtering (CF) is divided into Model-based, Memory-based, and Hybrid, but since it recommends similarity by user rating data, this paper builds a book recommendation system by Memory-based collaborative filtering[4]

II. RELATED RESEARCH

A. Collaborative Filtering

Collaborative filtering (CF) is a technology that predicts on its own based on the taste information collected from many users. It assumes that if a specific person A has the same opinion as person B on a certain issue, there is a

[1] Dept. of Software, Anyang University, Republic of Korea, swjang7285@gmail.com

[2] *Corresponding author: Dept. of Computer Science & Engineering, Anyang University, Republic of Korea, shleedosa@gmail.com

high probability that they will have similar opinions on other issues. There are two kinds of CF: item-based CF and user-based CF[5].

B. Process of User-Based Collaborative Filtering(UBCF)

The process of UBCF is largely divided into three stages, as shown in Figure 1. By organizing the ratings given by p users who experienced q items, a user x item matrix (p x q) is created as shown in Table 1. RA,1 is the rating data value for item 1 by user A, and is the rating data value that the user has not yet left.

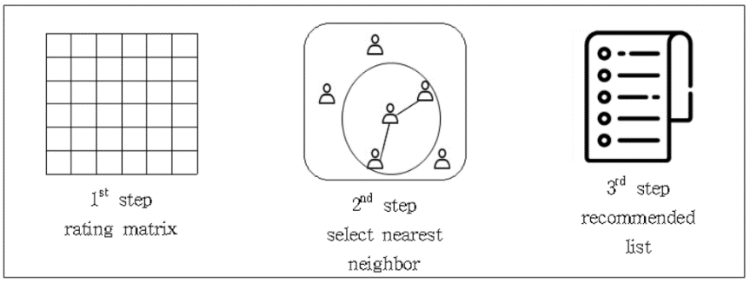

Fig. 1. The process of UBCF

The nearest neighbors are constructed by calculating the similarity using the users' ratings. The user x user similarity matrix is created using the rating matrix created in Step 1. Similarity is measured using Cosine Similarity.

$$\cos(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{\| \vec{A} \|^2 \times \| \vec{B} \|^2}$$

The cos function vectorizes the evaluation values of items commonly evaluated by users A and B into a q-dimensional space, and then calculates the cosine value of the angle as in the equation above to create a user x user similarity matrix. After generating the User x User similarity matrix, select a specific user and its nearest neighbors. The selection method is to sort the User x User similarity matrix in descending order and choose the top N users as the nearest neighbors.

Table 1. User x item matrix

	Item 1	Item 2	Item 3
User A	$R_{A.1}$	$R_{A.2}$	ϕ
User B	$R_{B.1}$	ϕ	$R_{B.3}$
User C	$R_{C.1}$	$R_{C.2}$	$R_{C.2}$

The steps for constructing a recommendation list are as follows. First, a list of the closest neighbors who are most similar to a specific user is selected. The top N items preferred by the closest neighbors are selected. Among them, items that the specific user has not yet experienced are selected to construct a recommendation list.

C. Process of Item-Based Collaborative Filtering(IBCF)

IBCF is largely divided into four stages, as shown in Figure 2 below. In the IBCF process, ratings are predicted for items that the user has not yet experienced, and recommendations are made for items with high ratings. The reason for the difference from the UBCF process described above is that the user x item matrix is likely a sparse matrix, so recommendations using predicted ratings are made only in IBCF.

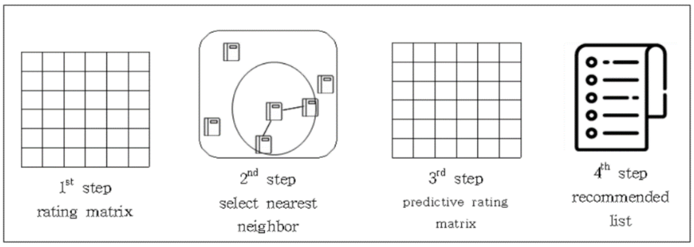

Fig. 2. The process of IBCF

By organizing the ratings given by p users who experienced q items, a user x item matrix of p x q is created as in Table 1. In order to measure the item-based similarity, an item x user matrix is created with the rows and columns swapped using Table 2, as in Table 3. $R_{A,1}$ is user A rating for item 1, and $R_{1,A}$ is user A rating for item 1 is the rating data that the user has not yet left.

The nearest neighbors are constructed by calculating the similarity using the users' ratings. The rating matrix generated in the first step is used to generate an item x item similarity matrix. The similarity is measured using Cosine Similarity. The cos function vectorizes the ratings commonly evaluated for items A and B into a q-dimensional space, and then calculates the cosine value to generate an item x item similarity matrix. After generating the item x item similarity matrix, the top N items with high similarity to a specific item are chosen as the nearest neighbors.

$$P_{u,i} = \sum N(S_{i,N}{}^{*} R_{u,N}) / \sum N(|S_{i,N}|$$

For items that a specific user has not experienced, a nearest neighbor list is generated, and then a new predicted rating for all movies is calculated based on the item similarity of the nearest neighbor lists and the rating data of items that the user has experienced, and a recommendation list is generated for high predicted ratings. The formula is predicted as follows. The meanings of the variables are as follows. The evaluation value $P_{u,i}$ is the predicted rating value of user u for item i, $S_{i,N}$ is the similarity vector of the Top-N items with the highest similarity to item i, and $R_{u,N}$ is the actual rating vector for the Top-N items with the highest similarity to user u's item i.

In the prediction rating matrix generated in step 3, the prediction rating matrix of a specific user is sorted and the top N that the specific user has not yet experienced are recommended to the user.

Table 2. User x item matrix

	Item 1	Item 2	Item 3
User A	$R_{A,1}$	$R_{A,2}$	ϕ
User B	$R_{B,1}$	ϕ	$R_{B,3}$
User C	$R_{C,1}$	$R_{C,2}$	$R_{C,2}$

Table 3. Item x user matrix

	User A	User B	User C
Item 1	$R_{1,A}$	$R_{2,A}$	ϕ
Item 2	$R_{1,B}$	ϕ	$R_{3,B}$
Item 3	$R_{1,C}$	$R_{2,C}$	$R_{2,C}$

D. Collaborative Filtering Performance Evaluation

The performance evaluation of collaborative filtering can be divided into a method for measuring recommendation accuracy and a method for evaluating the usability of the system [6]. The accuracy measurement method is divided into prediction accuracy, classification accuracy, and ranking accuracy.

Table 4. How to measure recommendation accuracy

Classification of accuracy	Measurement method	Measurement tool
Predictive accuracy	Measure the difference between predicted preference and actual preference	MAE

Classification accuracy	The frequency with which the correct judgment is made about whether an item is appropriate or not	Precision Recall F-measure
Rank accuracy	The difference between the ranking of items by preference between the user and the recommendation system	Item ranking

In this paper, items are recommended using the predicted ratings of users, so the Mean Absolute Error (MAE) as a performance measurement index of the recommendation system. MAE is the average error between the actual ratings and the predicted ratings of users. The MAE index is an index that indicates how similar the actual user ratings are to the predicted user ratings on average, and a recommendation system with a low MAE is evaluated as a recommendation system with good performance.

III. IMPLEMENTING A RECOMMENDATION SYSTEM

In order to build a book recommendation system, book information was collected from Yes24. Approximately 27,428 books were collected, and approximately 9,312 review data were collected from Aladdin. The reason for collecting book information and review data from different platforms is that Yes24 blocked part of the member IDs, such as yje948 -> y****8, making it impossible to distinguish members.

1	userID	isbn
2	탐이푸르다	9.79E+12
3	탐이푸르다	9.79E+12
4	탐이푸르다	9.79E+12
5	탐이푸르다	9.79E+12
6	dodododel	9.79E+12
7	dodododel	9.79E+12
8	dodododel	9.79E+12
9	dodododel	9.79E+12
10	dodododel	9.79E+12
11	농업코코	9.79E+12
12	하늘처럼	9.79E+12
13	하늘처럼	9.79E+12
14	하늘처럼	9.79E+12
15	하늘처럼	9.79E+12
16	하늘처럼	9.79E+12
17	고슴도치	9.79E+12
18	고슴도치	9.79E+12
19	고슴도치	9.79E+12
20	고슴도치	9.79E+12
21	고슴도치	9.79E+12
22	박재현	9.79E+12

Fig. 3. UBCF recommendation list

1	userID	isbn
2	파티를꿈꾸는둔재	9.78893E+12
3	파티를꿈꾸는둔재	9.78893E+12
4	파티를꿈꾸는둔재	9.78896E+12
5	파티를꿈꾸는둔재	9.78896E+12
6	파티를꿈꾸는둔재	9.79119E+12
7	파티를꿈꾸는둔재	9.79119E+12
8	파티를꿈꾸는둔재	9.789E+12
9	파티를꿈꾸는둔재	9.79116E+12
10	파티를꿈꾸는둔재	9.79116E+12
11	파티를꿈꾸는둔재	9.79116E+12
12	크와트로	9.79119E+12
13	크와트로	9.7912E+12
14	크와트로	9.78896E+12
15	크와트로	9.78896E+12
16	크와트로	9.78897E+12
17	크와트로	9.78896E+12
18	크와트로	9.78896E+12
19	크와트로	9.78893E+12
20	크와트로	9.78896E+12
21	크와트로	9.78896E+12
22	moonghong	9.789E+12

Fig. 4. IBCF recommendation list

The collaborative filtering developed in this paper was implemented in the jupyter notebook environment. Item-based collaborative filtering calculates predicted ratings to recommend books that match the user's taste, and user-based collaborative filtering recommends books based on user similarity rather than calculating predicted ratings because there are many sparse matrices. Once the recommendation list is created, it is saved in a CSV file as shown in Figure 3 and Figure 4. It is about which book to recommend to the user using the isbn through the user ID.

Recommendation performance was only performed on the recommendation system that applied item-based collaborative filtering. In this paper, the performance of composing the recommendation list based on items that match the user's personal taste was about 0.126 better than that of composing the recommendation list based on item similarity.

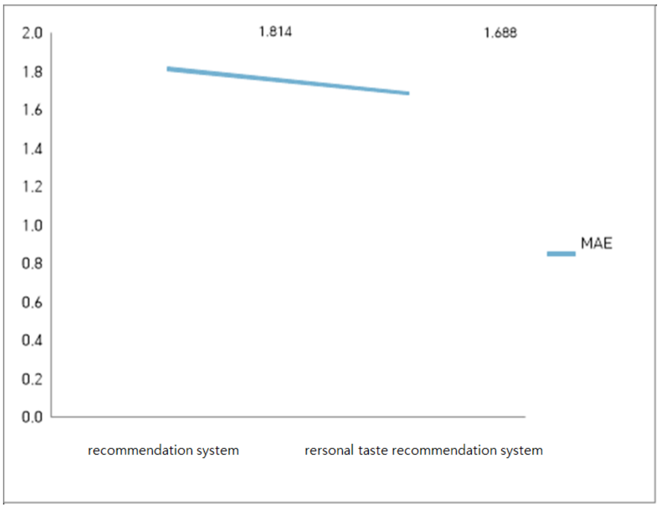

Fig. 5. Performance evaluation

IV. CONCLUDINGS REMARKS

This paper builds a recommendation system by users' ratings. The number of times a user experiences an item is much greater than the number of times he or she leaves a review for the item. Most users do not leave a review data even if they have a good feeling about the item. Therefore, if we build a recommendation system using data such as users' purchase experience, usage time, and number of views rather than the existing method, we can improve the sparsity problem of the existing collaborative filtering. As a future research task, we will collect more data than the currently collected data and compare and analyze the performance evaluation of the recommendation system according to the amount of data to improve the recommendation performance.

REFERENCES

[1] Y. S. Kim, “Research trend recommend service for personalization service,” Korean Institute of industrial Engineers, ie magazine, Vol. 19, No 1, pp. 37-42, May 2012.

[2] Application of Research Paper Recommender System to Digital Library.

[3] An Empirical Study on Hybrid Recommendation System Using Movie Lens Data.

[4] Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering.

[5] A system for recommending university liberal arts courses using collaborative filtering.

[6] A Comprehensive Performance Evaluation in Collaborative Filtering.