Proceedings of the International MultiConference of Engineers and Computer Scientists 2013 Vol I,
IMECS 2013, March 13 - 15, 2013, Hong Kong

Recommendation Method Using Bicluster Network Method

Tatsuya Saito, Kohei Kawahara, and Yoshifumi Okada

***Abstract*— In recent years, recommendation technology that predicts and proposes products and information matching user preferences has been gaining attention and is now widely used, particularly by online stores. The collaborative filtering method is the most popular form of recommendation technology. However, collaborative filtering often recommends similar items or items that the user already knows. Our objective is to improve the collaborative filtering method by developing a method that makes a recommendation after performing a wider and more relevant search of items that match the user's preferences in a database. This study proposes a method that expresses the transitivity of preferences between groups of users who like similar items as bicluster networks and recommends items on the basis of these bicluster networks. We performed a simulation experiment using film rating data to compare recommendation accuracy of the proposed method with that of an existing method on the basis of biclusters. The results demonstrate that the proposed method is far superior to existing methods in terms of the *recall* rate showing the degree of coverage of items that should be recommended.**

***Index Terms*— recommendation, collaborative filtering, transitivity of preferences, bicluster network, closed itemset**

I. Introduction

In recent years, there has been a sharp increase in services that provide a wide variety of products and information, such as online retail outlets and movie sharing sites. While this means that it is easy to obtain a wide range of products and information at any time, there is still no simple or efficient method to match information or products to an individual's objectives and preferences. For this reason, technology that recommends products matching user interests and preferences is gaining attention. Currently, the most mainstream form of recommendation technology is a method known as collaborative filtering [1]. This method accumulates user preference information (for example, purchase history and rating values) and recommends products and information (hereafter items) on the basis of this information.

Manuscript received December 26, 2012. This work was supported by Grant-in-Aid for Young Scientists (B) No.24700204 from MEXT Japan.

T. Saito is with the Department of Information and Electronic Engineering, Muroran Institue of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan(e-mail: saito@cbrl.csse.muroran-it.ac.jp).

K. Kawahara is with the Department of Information and Electronic Engineering, Muroran Institue of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan (e-mail: kawahara@cbrl.csse.muroran-it.ac.jp).

Y. Okada is with College of Information and Systems, Muroran Institute of Technology, 27-1, Mizumoto-cho, Muroran 050-8585, Japan (corresponding author to provide phone: +81-143-5408; fax: +81-143-5408; e-mail: okada@csse.muroran-it.ac.jp)

Recently, Symeonidis *et al*. [2] proposed a method of collaborative filtering known as Nearest-Biclusters Collaborative Filtering (NBCF). This method first identifies groups of items that have been rated highly by multiple users. The sets containing user groups, item groups, and rating values are called biclusters. Normally, different user groups will rate item groups differently; consequently, a wide variety and large number of biclusters will exist. Next, this method determines users for which the recommendation is requested and the biclusters where ratings for items are similar, and it proposes items that users belonging to that bicluster will like. NBCF is reported to have higher recommendation accuracy than user-based CF [3] and item-based CF [4], which are the most widely-used recommendation methods.

However, the NBCF method draws its recommendations from biclusters that are limited to ratings provided by users who are similar to the user for whom the recommendation is requested. Therefore, items that the user already knows or items that are similar to each other are often recommended, which limits the effectiveness of the NBCF method.

Our goal is to resolve this limitation by developing a method that makes a recommendation after performing a wider and more relevant search of items matching the user's preferences in a database. This study proposes a method that expresses the transitivity of preferences between groups of users who like similar items as bicluster networks and recommends items on the basis of these bicluster networks. A comparative evaluation of the recommendations generated by the proposed method and the NBCF method using a movie rating data set is performed experimentally.

II. Method

A. Basic Idea

In this study, the transitivity of preferences, as shown in Fig. 1, points to the essential fact that "with a user x , a user with similar preferences y, and a user with similar preferences to y as z, there is a possibility that x will like products liked by z". Here it is important to note that the definition of transitivity of preferences used in this study is different from the definition given by Regenwetter *et al*. [5]. We consider the extension of the above-described transitivity of preferences between individuals to the relationship between user groups that share preferences. As shown in Fig. 2, this is realized by replacing user groups with biclusters. That is to say, sets of biclusters with a mutually high overlapping rate (degree of similarity) are seen as user groups demonstrating transitivity of

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

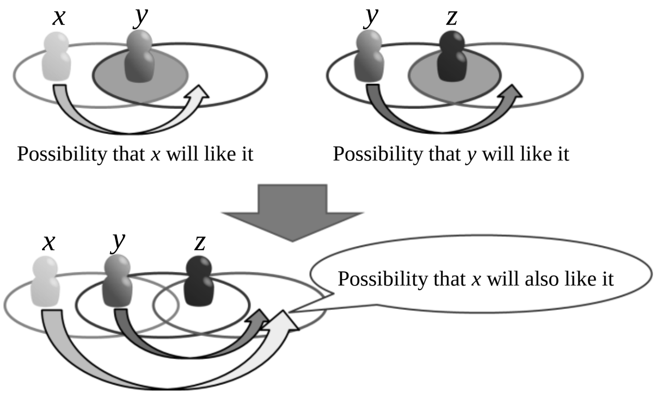

Fig. 1: Transitivity of preferences

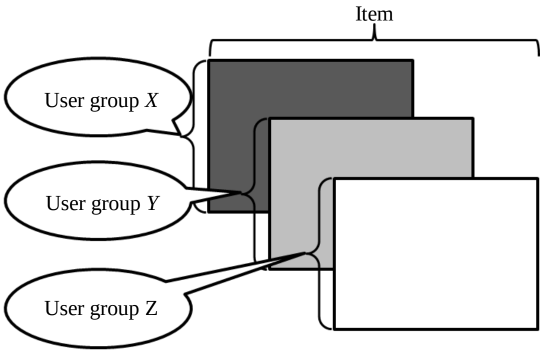

Fig. 2: Bicluster network

preferences. Items are recommended by proposing items that users belonging to the bicluster network would like.

B. Bicluster generation

In this study, we create the biclusters using the Linear time Closed itemset Miner algorithm, which is a high-speed closed itemset enumeration method developed by Uno *et al*. [6] Closed itemset enumeration is a method of comprehensively discovering maximal element sets (closed itemset) from the itemsets (frequent itemsets) that co-occur in the transactions of more than a specified minimum number of users from the transaction database. In Fig. 3, we demonstrate the generation of a bicluster from the transaction database. The generated bicluster refers to the identified user groups, item groups to which a common high rating has been applied, and the set of rating values.

C. Bicluster network detection and recommendation

On the basis of reference document [2], we calculate the degree of similarity between the items liked by the user for whom the recommendation is requested (query) and each bicluster, and consider the bicluster with the highest degree of similarity as the nearest bicluster. The degree of similarity between biclusters X and Y is obtained from the overlapping rate of set C_X of bicluster X elements and set C_Y of bicluster Y elements. The overlapping rate is calculated according to the following formula.

$$Overlapping(C_X, C_Y) = \frac{|C_X \cap C_Y|}{|C_X| * |C_Y|}$$

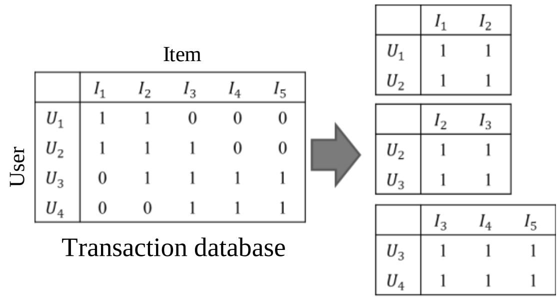

Fig 3: Bicluster generation

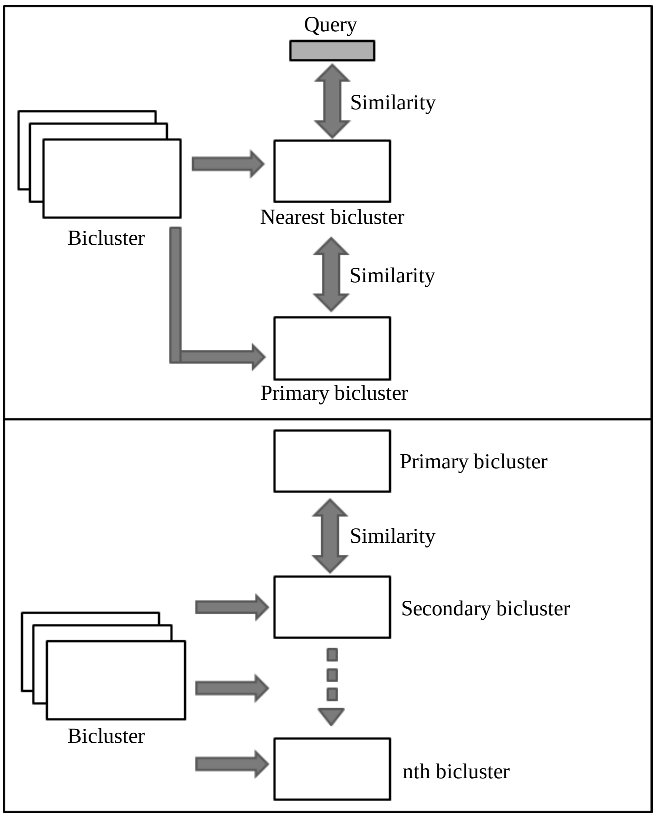

Fig. 4: Bicluster network detection

Fig. 4 outlines the bicluster network detection process. Items are recommended on the basis of bicluster networks according to the following four steps.

1) Obtain the overlapping rate of the nearest bicluster and the other biclusters, and detect the primary bicluster as the bicluster with the highest overlapping rate.
2) Obtain the overlapping rate of the primary bicluster and biclusters other than the nearest bicluster, and detect the secondary bicluster as the bicluster with the highest overlapping rate.
3) Repeat step 2 until the nth bicluster is detected.
4) Propose the items to which the users in the nearest bicluster and primary ~*n*th bicluster have given a high rating to the user for whom the recommendation is requested.

ISBN: 978-988-19251-8-3
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

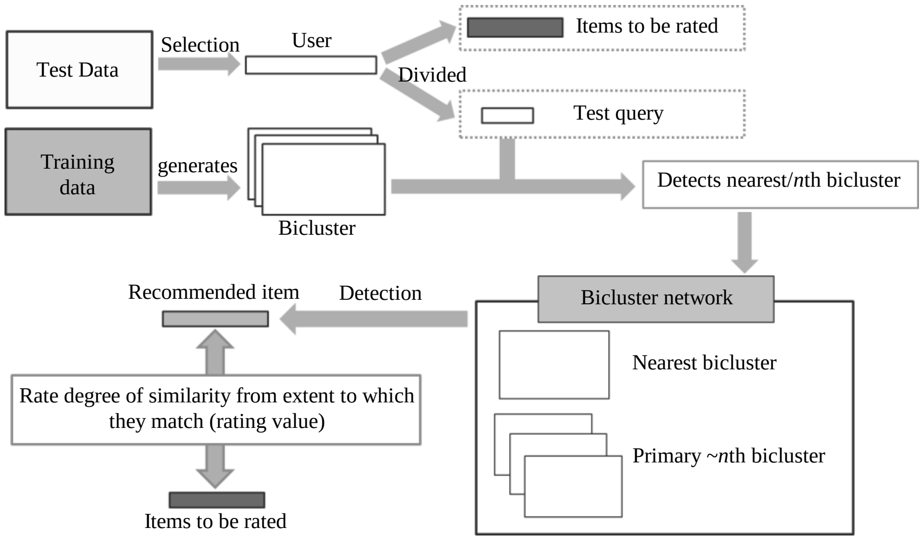

Fig. 5: Rating experiment

III. Rating Experiment

A. Data set

This study uses MovieLens, a movie rating data set, which was published by GroupLens [7].

MovieLens contains rating data from 943 users for 1,682 movies. Movies are ranked from one to five, with one being the lowest rating. Each user rates a minimum of 10 movies and there are a total of 100,000 ratings.

B. Pre-processing

The rating data is binary-processed and the transaction database is constructed in the following way. Ratings of four or five (highly-rated movies) are converted to one and all other ratings (movies with lower ratings) are converted to zero. The transaction database contains matrix data: rows are users, columns are items (movies), and the elements are binary rating values (1, 0).

C. Cross validation

Fig. 5 provides an outline of the rating experiment. In this experiment, three-fold cross-validation is used to rate the accuracy of recommendation methods using bicluster networks. First, the transaction database is divided into three sets. The first two sets are training data sets, and the third is a test data set. The training data sets are used to generate the bicluster network.

The test data set is divided into items for rating and test queries for each user. At this time, a fixed number of movies (determined in advance) are used as items for rating, and the remaining movies are used as test queries. The item recommendation accuracy is rated in terms of the extent to which the items for rating were recommended in relation to the test queries. With the above operation, the rating experiment was conducted on a total of three patterns, substituting the training data set for the test data set.

D. Rating index

Rating with this method is conducted using *precision* rate, *recall* rate, and F-measure. Here *precision* rate refers to the share of the movies recommended that were actually rated highly. The *recall* rate expresses the share of the highly-rated movies in the transaction database that were actually recommended. F-measure is an index that takes the balance between the *precision* rate and *recall* rate into consideration. The *precision* rate, *recall* rate, and F-measure are calculated according to the following formulae.

$$Precision = \frac{|A \cap B|}{|B|}$$

$$Recall = \frac{|A \cap B|}{|A|}$$

$$F-measure = \frac{2 * Recall * Precision}{(Recall + Precision)}$$

Here A is the set of movies rated highly by the user and B is the set of recommended movies. In addition, the *precision* rate, *recall* rate, and F-measure were also calculated in relation to NBCF, which is an existing method based on biclusters, for the purpose of comparison. In this experiment, the minimum number of users was set to 30, the maximum degree of bicluster network detection was set to three, and the number of items for rating was set to 10 and 20.

IV. Results and Discussion

The results of setting the number of items for rating to 10 and 20 are shown in Fig. 6 and Fig. 7, respectively. The figures show the average value of the index value calculated using three-fold cross-validation.

The F-measure calculated using the proposed method is a little higher than that calculated using the NBCF method, which demonstrates a high level of recommendation accuracy. In addition, the proposed method is far superior to NBCF in terms of *recall* rate, regardless of whether the number of items

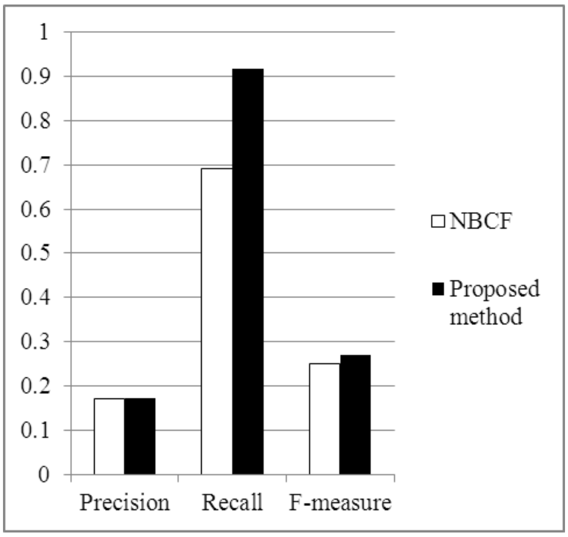

Fig. 6: Results of 10 rated items

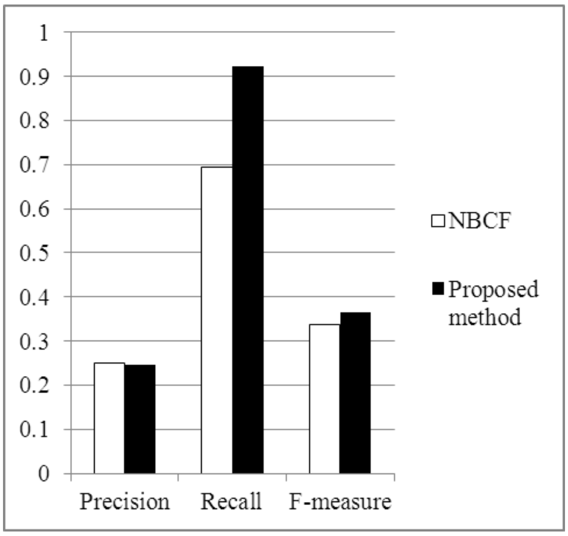

Fig. 7: Results of 20 rated items

is set to 10 or 20. This is not only because the number of movies recommended has increased as a result of bicluster networks but also because the test query users are actually applying a high rating for those movies. This result means that, with bicluster networks, the transitivity of preferences is working efficiently and is contributing to improvements in the *recall* rate. On the other hand, no significant difference was seen between the two methods in terms of *precision* rate. However, this demonstrates that a reduction in the recommendation quality for proposed movies is being suppressed, regardless of the increase in the number of recommended movies.

By realizing transitivity of preferences using bicluster networks, recommendations can be made after performing a wider and more relevant search of items matching the user's preferences in a database.

V. CONCLUSION

In this study, we have expressed the transitivity of preferences using bicluster networks and used this to propose a new information recommendation method. As an experiment, we used a movie rating data set and compared the recommendation accuracy of the proposed method and a popular method NBCF. The results demonstrated that the proposed method exhibits a higher level of recommendation accuracy than the NBCF method. In addition, the *recall* rate can be significantly increased without lowering the *precision* rate. In other words, it is possible to make recommendations after performing a wider and more relevant search of items matching the preferences of the user, while maintaining recommendation accuracy.

In future, we will conduct rating experiments with different minimum numbers of users and different data sets.

ACKNOWLEDGMENT

This work was supported by Grant-in-Aid for Young Scientists (B) No.24700204 from MEXT JAPAN.

REFERENCES

[1] X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques", Advances in Artificial Intelligence, Vol.2009, Article ID 421425.
[2] P. Symeonidis, A. Nanopoulos, A. Papadopoulos, and Y. Manolopoulos, "NearestBiclusters Collaborative Filtering,"WEBKDD'06, 2006.
[3] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews.", In Proceedings of CSCW '94, Chapel Hill, NC, 1994.
[4] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms", In Proceedings of the 10th international conference on World Wide Web, pp.285-295, 2001.
[5] M. Regenwetter, J. Dana and CP. Davis-Stober. "Transitivity of preferences", Psychol Rev. Vol.118, No.1, pp.42-56, 2011.
[6] T. Uno, T. Asai, Y. Uchida and H. Arimura, "An Efficient Algorithm for Enumerating Closed Patterns in Transaction Databases," Lecture Notes in Artificial Intelligence, 3245, 2004, pp. 133-141.
[7] GroupLens. Available: http://www.grouplens.org