We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Improving performance in recommender system: collaborative filtering algorithm and user's rating pattern

Hee Choon Lee, Seok Jun Lee and Young Jun Chung
Sang-ji University, Sang-ji University, Kangwon National University
Korea

1. Introduction

With the advance of IT technologies, a recommender system in online commerce environment has been introduced as personalized services (Schafer et al., 2006). The recommender system is used in E-commerce for recommending a product, an item or even any web service to each customer based on customer's preference.

Since a recommender system can predict customers' preference and forecast the future degree of customer's fondness for a certain item and services, it is used as a conspicuous service which distinguishes an on-line commerce service from an off-line commerce service. In predicting each user's preference, a recommender system essentially provides enough information of items and users because it is able to predict the specific user's preference for a target item and suggest the result to users.

One of the classic recommender systems is a content-based filtering system which uses textual contents. In the recommender system for an on-line movie rental process, two types of profiles are usually used; movie profile and customer profile. The movie profile describes a movie category, main actors, and performance movie. The customer profile is created with the historical experienced textual information, which is stored in the system, of items or users for seeking the best fits.

This type of approach works well in initial systems, but there are some drawbacks for expanding the scales of recommender systems due to the following reasons. First, there are difficulties in converting features of all traded items into textual data. Additionally, if the number of traded items extremely increases, it is not easy to automatically convert all items' information into textual forms. Second, since content-based systems only recommend items based on the past experience of user, it cannot help the user choose items for specific cases. This problem is called over-specification for recommendations.

Such drawbacks can be eliminated by collaborative filtering recommender systems, which use relationships between users and items that can be represented on numerical scales (i.e. preference rating). This preference rating information can be collected from tracks of clients who surf web and purchase items. Typically, such types of recommender systems utilize neighbour users' data, using a set of data that has similar characteristics for

recommendations of a target item. This concept used for users or item is called user-based or item-based respectively.

Collaborative filtering recommender systems are successfully used in commercial web sites such as Amazon.com, E-bay and Netflix. The item-based approach is generally adapted to commercial web sites because the speed and the range of expansion of user are much higher than items and also the problem of data scarcity is willing to occur in user-based (Linden et al., 2003; popescul et al., 2001).

2. Collaborative Filtering

To make up for shortcomings of content based approach, collaborative filtering approach is adopted in the recommender system. Collaborative filtering approach is the method using only related data between users and items like explicit numerical ratings, and the detailed attributes of both users and items are intentionally ignored. Collaborative filtering can be said that the most popular item is recommended for every user. It is known as the most commercially successful recommender technique and is the base of the studies on the recommender systems algorithms.

Collaborative filtering approach can be grouped into two classes according to algorithms for prediction users' preferences. One is memory-based and the other is model-based. Memory-based algorithms predict the rating of users using the previously rated items by the users and other users who have similar tastes. In contrast to memory-based algorithms, model-based algorithms use the probabilistic approach, such as, cluster models, Bayesian network, and machine learning approach (Adomavicius & Tuzhilin, 2005 ; popescul et al., 2001).

In memory-based collaborative filtering algorithms, to show the similarity of the preferences between the active users and others, the Pearson's correlation coefficient was used in the GroupLens first. Breese et al. researched the ways of improving the prediction accuracy, using the Pearson's correlation coefficient, the Vector similarity, the default voting, the inverse user frequency, and the case amplification (Breese et al., 1998). Also, they researched the collaborative filtering with the use of the Bayesian probability model. Herlocker et al. studied about making the prediction accuracy improved with using both the Pearson's correlation coefficients as the similarity weight and the effect of the number of co-rated items (Herlocker et al., 2004). Memory-based collaborative filtering algorithms can be divided into user-based method using the relations among users and item-based method using the relations between items as the method of algorithm application.

Collaborative filtering system also has some limitations like content-based system have.

→ • New user problem
• New item problem
• Scarcity

New user problem is the same problem as content-based system has. To predict more accurate recommendations, collaborative filtering system has many ratings that users give, because the more ratings are given to the system, the better the user's preferences can be understood. If a new item enters into the system, there are no users who give rating to item. Therefore, this item can not be recommended to users in the system. To solve this problem, it will be possible to make ratings from system manager or some groups of panel users.

In many recommender systems, the numbers of ratings already rated by users are very small to make prediction for recommendation.The success of the collaborative filtering system depends on the available users. Well established system like MovieLens dataset also has the scarcity, 95.8% in 1 million dataset and 93.7% in 100K dataset.

To overcome these problems, diverse approaches are proposed. For example, it is possible to use other information like demographical information and users' behaviour in the web and dimensionality reduction techniques (Adomavicius & Tuzhilin, 2005).

3. Algorithm

To predict the preference of a target user about specific items, the neighbor selection process is firstly carried out. Figure 1 shows the neighbor selection step for predicting the preference of the active user 4 about the specific item 4. The user 1 and the user 3 are selected as the neighbour users of the because they have already rated the item 4. For calculating the prediction value about the preference rating of the user 4 about the item 4, the preference ratings of neighbors are needed and in this figure, the user 1 and the user 4 have already rated about the item 4.

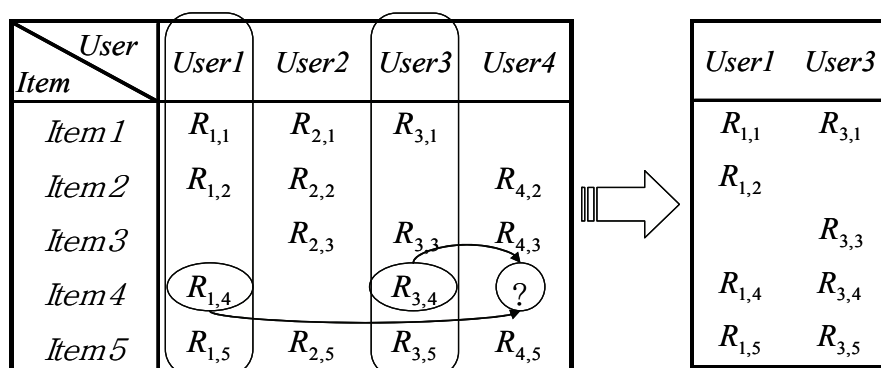| User \ Item	User1	User2	User3	User4		User1	User3
Item 1	$R_{1,1}$	$R_{2,1}$	$R_{3,1}$			$R_{1,1}$	$R_{3,1}$
Item 2	$R_{1,2}$	$R_{2,2}$		$R_{4,2}$		$R_{1,2}$	
Item 3		$R_{2,3}$	$R_{3,3}$	$R_{4,3}$			$R_{3,3}$
Item 4	$R_{1,4}$		$R_{3,4}$	?		$R_{1,4}$	$R_{3,4}$
Item 5	$R_{1,5}$	$R_{2,5}$	$R_{3,5}$	$R_{4,5}$		$R_{1,5}$	$R_{3,5}$

Fig. 1. Neighbour selection step

Before applying algorithms, a preference similarity weight for items between a target user and his neighbors must be defined. Pearson's correlation coefficient is used for the similarity weight between them and equation 1is the similarity weight used in this study.

$$r_{uj} = \frac{\sum_{i=1}^{m}\left(R_{u,i} - \overline{R}_u\right)\left(R_{j,i} - \overline{R}_j\right)}{\sqrt{\sum_{i=1}^{m}\left(R_{u,i} - \overline{R}_u\right)^2 \cdot \sum_{i=1}^{m}\left(R_{j,i} - \overline{R}_j\right)^2}} \tag{1}$$

r_{uj} is the similarity weight between the target user u and neighbor user j and $R_{u,i}$ denotes preference ratings of the target user u for the items i which are already rated by the target user. $R_{j,i}$ denotes preference ratings of neighbor user j, $\overline{R}_u$ and $\overline{R}_j$ are the mean of ratings of user u and j. In this equation, all ratings R must be co-rated by user u and j.

3.1 Neighborhood Based Collaborative Filtering Algorithm

One of the most famous and well known algorithm is neighborhood based collaborative filtering algorithm(NBCFA) proposed by GroupLens (Resnick et al., 1994).

$$\hat{U}_x = \overline{U} + \frac{\sum_{J \in Raters}(J_x - \overline{J})r_{uj}}{\sum_{J \in Raters} |r_{uj}|}, \ where \ \overline{J} = \frac{\sum_i^n J_i}{n}, i \neq x \tag{2}$$

The $\hat{U}_x$ is the prediction value of the preference of the target user u over the target item x, the $\overline{U}$ is the mean of the all preference ratings of the user u, the J_x is the preference rating of the neighbor user j over the target item x, and the $\overline{J}$ is the mean of the all preference ratings of the neighbor user j except the rating of target item x. Raters are users who rate the preference of the item in the data set. r_{uj} is the similarity weight of both the user u and the neighbor user j and the Pearson's correlation coefficient and the vector similarity are usually used for this but the performance of the Pearson's correlation coefficient is usually better than the vector similarity.

Figure 2 shows the prediction step for the item 4 of the user 4 using the NBCFA. First, the similarity weights between the user 1 and the user 4, the user 1 and the user 3 are calculated. The similarity weight indicates the preference relationship of the two users, and the more similar user 1 and neighbor users are, the more weight will increase in the prediction step. The two most commonly used similarity weights will be described below. Usually the Pearson's correlation coefficient is used for similarity weight of two users but any types of measures, cosine vector and Euclidean distance as similarity weight, are possible if the preference of two users are explained.vector and Euclidean distance as similarity weight. In this chapter we use the Pearson's correlation coefficient as the similarity weight of two users.
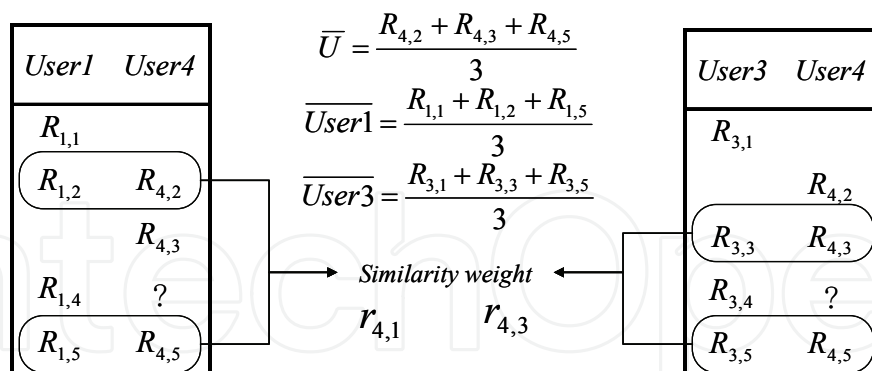

Fig. 2. Neighbourhood Based Collaborative Filtering

3.2 Correspondence Mean Algorithm

In the NBCFA, $\overline{U}$ is the mean of the preferences of the target user u who will take prediction value for the specific item which the target user has never experienced. In this case of calculating $\overline{U}$ with the entire ratings of the target user, the preference of the target user is overestimated, which leads to a possibility that the preference of the target user might be insufficiently or excessively reflected if the numbers of co-rated items with his or her neighbour are small. So, some tuning is needed for alleviating insufficient reflection of target user and his or her neighbour.

This is why $\overline{U}_{match}$ and $\overline{J}_{match}$ are used in the CMA. $\overline{U}_{match}$ is the mean of all the means of the preferences that are rated by both the user u and the neighbour user j.

Equation 3 is the correspondence mean algorithm (Lee et al., 2007a; Lee et al., 2007b)...

$$\hat{U}_x = \overline{U}_{match} + \frac{\sum\limits_{J \in Raters}(J_x - \overline{J}_{match})r_{uj}}{\sum\limits_{J \in Raters}|r_{uj}|} \tag{3}$$

$\overline{J}_{match}$, the mean of the preferences rated by both the user u and the neighbour user j, and it is calculated by the same way of calculating the Pearson's correlation coefficient. For example, if the user u has 10 ratings and one of the neighbour user j has 20 ratings and the other user j has 10 ratings, the preference of the user u must be calculated with the relationship of each neighbour. In the case of this example, if the first user j and user u have only 5 co-rated items and the second user j and user u have 10 co-rated items, it is reasonable to use the only co-rated items to calculate the preferences user u and user j not using all ratings of them. Also, $\overline{U}_{match}$ must be the mean of $\overline{U}_{sub\_match}$s of the user u and user j.
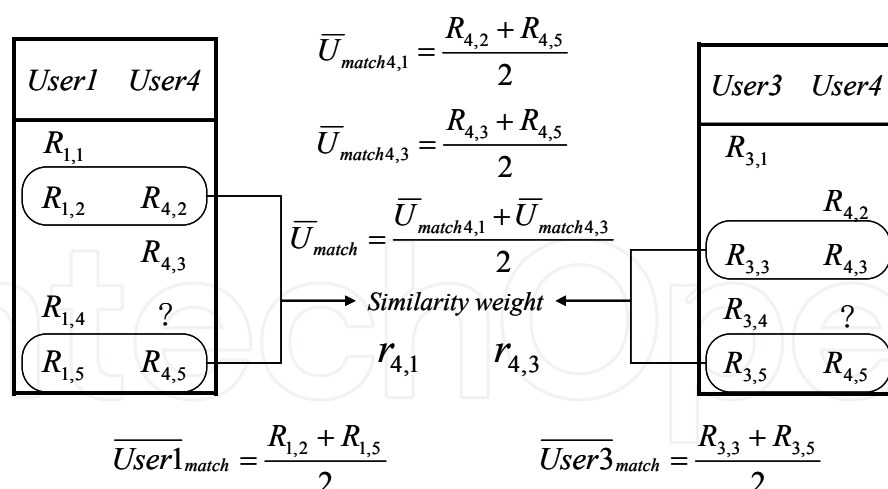

Fig. 3. Correspondence Mean Algorithm

Figure 4 shows the prediction step for item 4 of user 4 using CMA. First, the similarity weights of user 1 and user 4, user 1 and user 3 are calculated. In this step, we calculate the $\overline{U}_{match}$ as the preference of user 4. To compute $\overline{U}_{match}$, $\overline{U}_{match4,1}$ and $\overline{U}_{match4,3}$ are calculated before, which uses the ratings of co-rated by two users.

3.3 Significance Weight and Evaluation Metric

The similarity weight of the target user with the neighbour explains their relationship of preference to items. This similarity weight of both users' preference of items must be considered, so it will be computed with ratings which are rated by both users. If the similarity weight of both users is computed only with small portion of their ratings, it is doubtful of their real relationship of preference to items. For example, the similarity weight using only two pairs of ratings is just 1 or -1, and even it is impossible to compute their relationship of preference. Herlocker et al. adopted the significance weight to devalue the overestimated similarity weight. They showed the improvement of prediction accuracy by reducing the overestimated Pearson's correlation coefficient under the number of co-rated movies as 50 (Herlocker et al., 2004).

To devalue the overestimated similarity of active user and their neighbours' preference, the significance weight is set to consider the effect of the number of co-rated movies from both active user and his or her neighbour user and applied as equation 4.

$$\hat{U}_x = \overline{U} + \frac{\sum_{J \in Raters}(J_x - \overline{J})r'_{uj}}{\sum_{J \in Raters}\left|r'_{uj}\right|} \tag{4}$$

$$where, \quad r'_{uj} = r_{uj} \cdot sw$$

The significance weight(sw) gains the weight according to the number of co-rated items as shown below.

$$sw = \frac{\min(n(I_u \cap I_j), C)}{C} \tag{5}$$

In equation 5, the $n(I_u \cap I_j)$ is the number of movies that are rated by both the target user u and neighbor user j, and the C is the number of co-rated movies that are for setting the application range of the significance weight. To get the prediction accuracy more improved, we extend the range of the significance weight according to the number of co-rated movies as the set C below .

$$C = \left\{ \begin{array}{l} 3, 5, 7, 10, 15, \cdots 50, 60, \cdots, 100, 120, 150, 180, 200, \\ 300, \cdots, 500, 700, 1000, 2000, 4000, 7000, 10000 \end{array} \right\} \tag{6}$$

Several techniques have been used to evaluate recommender systems. Those techniques are divided into three categories; predictive accuracy metrics, classification accuracy metrics and rank accuracy metrics. The predictive accuracy metrics measure how close the predicted ratings by algorithm are to the true ratings in the test dataset. In this study, Mean Absolute Error (MAE), one of the predictive accuracy metrics, is used to evaluate the performance of each algorithm, especially measuring each user's MAE, to test the performance of two algorithms and other experimental results.

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|R_{uj} - \hat{R}_{uj}\right| \tag{7}$$

In equation 7, R_{uj} is the true rating of user u given to the item j and $\hat{R}_{uj}$ is the prediction value of user u to the item j.

4. Pre-evaluation

4.1 Error Fence

To find the relationship of the prediction accuracy of users' preference with the pre-evaluation approach, the prediction error fences are set on the each user's MAE by using exploratory data analysis (EDA) technique. To set the prediction error fence, we use the concept of the hinge as proposed by Tukey to set the fence (Tukey, 1977). For classifying the users' groups, we set the range of the normal errors as the H-spread, and the range of abnormal errors is set as the adjacent values and the outside values divided by the inner fence. Figure 4 shows the H-spread and the fences for classifying the normal errors range and the abnormal errors range of MAEs and standard deviations of each user's ratings in the training dataset (Han et al., 2008).
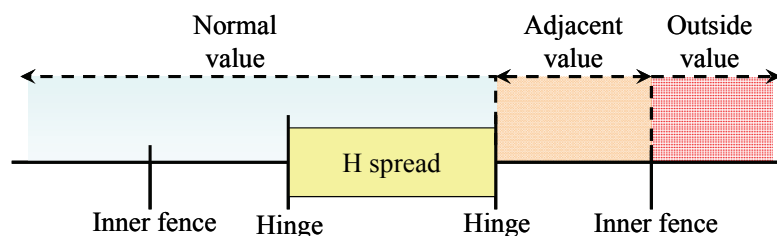
Fig. 4. Error bound divided by EDA

After classifying process, we classify the users' groups according to the classified MAEs and standard deviations and we run the statistical test on the groups to find their relationships.

4.2 Classification Function

According to our previous study, the prediction accuracy of user's preference on the item has a close relationship with the generative probability of specific ratings which have been already rated by the user before_prediction process. The generative probabilities of specific ratings denoted as δ_{u1}, δ_{u2}, δ_{u3} will be used to define the classification functions that select users whose MAE is lower than non-selected users' MAE from the next equations presented by Lee (Lee et al., 2007).

$$\delta_{u1} = \begin{cases} 1, & f_u(R_5) \geq f_u(R_2) \\ 0, & elsewhere \end{cases} \qquad \delta_{u2} = \begin{cases} 1, & f_u(R_1) \geq f_u(R_4) \\ 0, & elsewhere \end{cases}$$

$$\delta_{u3} = \begin{cases} 1, & f_u(\{R_1\} \cup \{R_5\}) \geq f_u(\{R_2\} \cup \{R_3\} \cup \{R_4\}) \\ 0, & elsewhere \end{cases} \tag{7}$$

$$where, \quad R_i = i, \quad i = \{1,2,3,4,5\}$$

δ_{u1}, δ_{u2}, δ_{u3} are the conditions for defining the classification functions and showed in the equation 8. It has only the values of 1 or 0.

$$L(\delta_{u1}, \delta_{u2}, \delta_{u3}) = \delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3} \tag{8}$$

To classify users who have higher prediction accuracy than non-selected users, we propose another classification function in this study. We also define the generative probabilities of specific ratings as θ_{u1}, θ_{u2}, θ_{u3}. Each condition and function is showed in equation 9 and 10.

$$\theta_{u1} = \begin{cases} 1, & f_u(R_2) \geq f_u(R_1) \\ 0, & elsewhere \end{cases} \qquad \theta_{u2} = \begin{cases} 1, & f_u(R_4) \geq f_u(R_5) \\ 0, & elsewhere \end{cases}$$

$$\theta_{u3} = \begin{cases} 1, & f_u(R_3) \geq f_u(\{R_2\} \cup \{R_4\}) \\ 0, & elsewhere \end{cases} \tag{9}$$

$$where, \quad R_i = i, \quad i = \{1,2,3,4,5\}$$

θ_{u1}, θ_{u2}, θ_{u3} are the conditions that classify users who have high prediction accuracy for defining the classification functions and showed in the equation 10. It also has only the values of 1 or 0.

$$H(\theta_{u1}, \theta_{u2}, \theta_{u3}) = \theta_{u1} \cdot \theta_{u2} \cdot \theta_{u3} \tag{10}$$

5. Experiments

5.1 Experimental Dataset

To evaluate the performance of each algorithm and pre-evaluation function, our experiment uses the MovieLens datasets which have been made public by GroupLens for experiment. The GroupLens presents 2 types of the MovieLens dataset. One is a 100K dataset and the other is a 1 million dataset. We use both datasets for our research analysis.

100K dataset was rated by 943 users over 1682 movies and the total ratings are 100,000 while 1 million dataset was rated by 6040 users over 3952 movies and the total ratings are more than 1,000,000.

To test performance of two algorithms and classfication function, we divide each dataset into 80% of training dataset and 20% of test dataset. Generally, training and test datasets are

divided randomly regardless of the number of ratings each user has. In this case, there are biased ratios of ratings belonging to training dataset and test dataset for each user. To balance this discrepancy of the 80% of training and the 20% of test dataset, we divide off training dataset and test dataset from each user's ratings randomly. We then predict the 20% of test dataset through NBCFA and CMA using 80% of training dataset(Han et al., 2008).Figure 5 shows the concept of composing the experimental dataset.

Generally, the prediction accuracy will be evaluated by the MAE which is calculated by the average of the absolute errors of all the real ratings between predicted values in test dataset. But our study uses the each user's MAE which is calculated by using ratings of each user in the test dataset instead of using all the ratings in the test dataset. This study shows the possibility of the pre-evaluation approach using previously possessed preference information of users as ratings on the items before the prediction process for each user's preference.
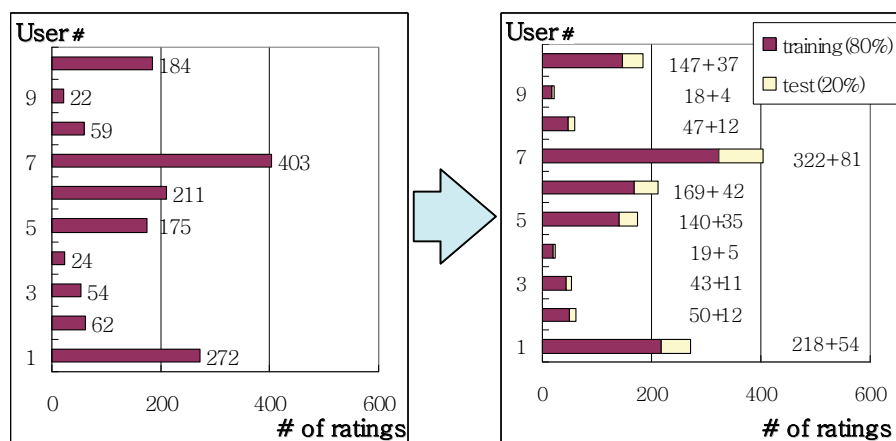
Fig. 5. Experimental dataset formation

5.2 Experimental Design

To compare the prediction accuracy of the result of NBCFA and CMA, the followings are conducted.

First, the Pearson's correlation coefficient is applied to NBCFA and CMA as the similarity weight, to present the preference relation of target users and their neighbours, and the prediction results of that are compared according to the user-based method and the item-based method. The user-based prediction is the way that uses the relations of users to compute the similarity weight and applies them to each algorithm for predicting the test set. And item-based prediction is the way that utilizes the relations of items or goods to compute the similarity weight and applies them. And then we analyze the prediction accuracy statistically in the view of each user's MAE, not using the MAE of all predicted ratings, to confirm the improvement of prediction accuracy using the CMA.

Second, we analyze the effect of the significance weight to the result of each prediction method and prediction algorithm. The similarity weight, which presents the relation of preference between users, might be overestimated if the numbers of co-rated movies are small. To get the prediction accuracy more improved, we extend the range of the significance weight according to the equation 6.

Figure 6 shows the steps for our study and flow of the experiment. Step 1 shows the division of our experiment into user-based and item-based according to the prediction method explained. Step 2 classifies each prediction method by prediction algorithm applied to each dataset. Step 3 divides the prediction algorithm according to similarity weight applied to the each algorithm. To know how much the numbers of co-rated items affect the accuracy of prediction, step 4 is divided according to the significance weight.
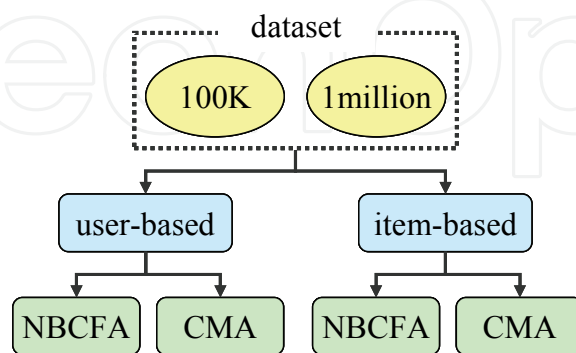

Fig. 6. Experimental flow diagram for algorithm

Figure 7. shows the experiment flow diagram for proposing the possibility of the pre-evaluation for the preference prediction errors before the prediction process.
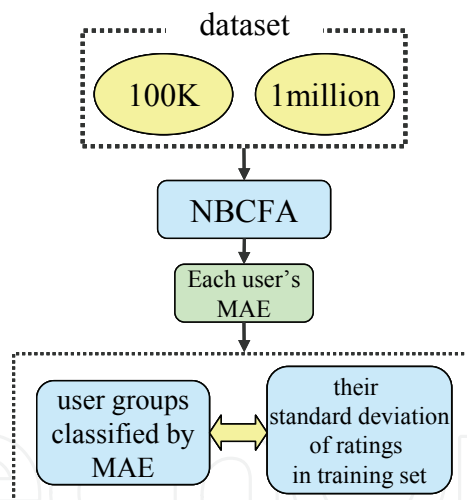
Fig. 7. Experimental flow diagram for error bound

We evaluate the prediction errors of each user's ratings in the test dataset after the prediction process by using NBCFA. And then, we run statistical tests for analyzing the relationship of prediction error of the user's preferred items between information of users before prediction.

First, we classify 3 groups into normal, adjacent, and abnormal users group according to the each user's MAE by applying the exploratory data analysis approach. And then, the analysis of variance test is applied to comparing the means of each user's standard deviation derived from the training dataset. According to the results, users are classified into groups.

Figure 8. shows the experiment flow diagram evaluating the performance of classification functions of the pre-evaluation for the preference prediction errors before the prediction process.
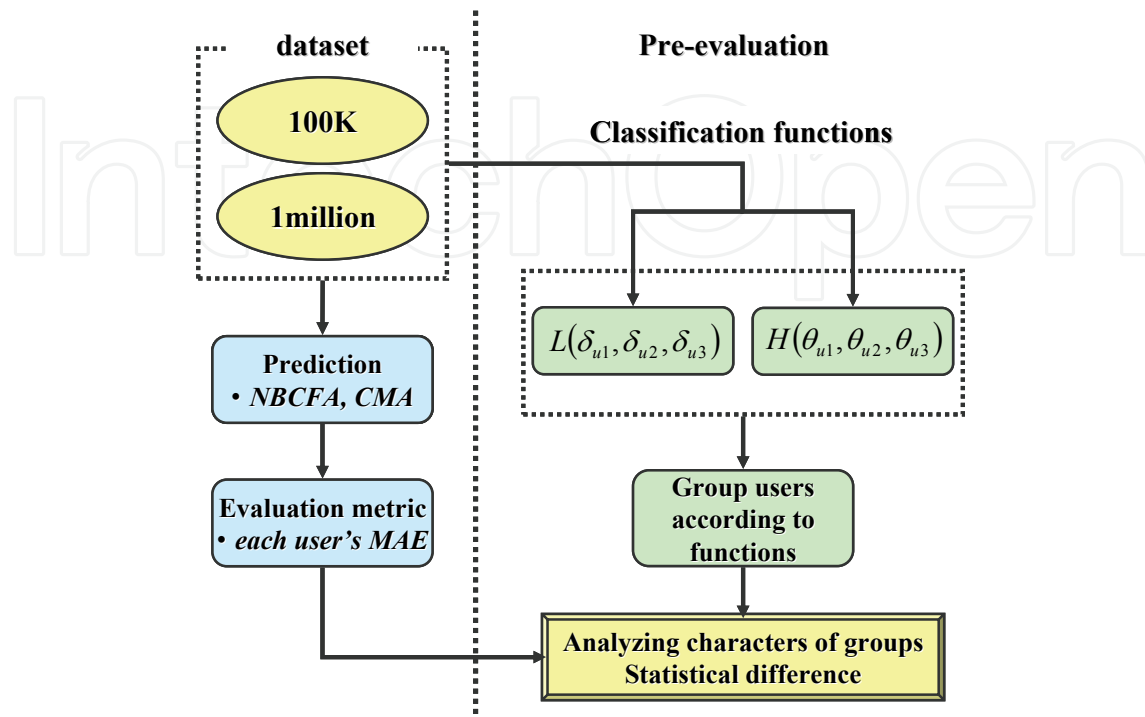

Fig. 8. Experimental flow diagram for classification function

The left side of the vertical dotted line on figure 8 shows the process of prediction domain and two MovieLens datasets which are predicted through NBCFA and CMA, and then the prediction results are evaluated by each user's MAE.

The right side of the line shows the pre-evaluation process using $L(\delta_{u1}, \delta_{u2}, \delta_{u3})$ and $H(\theta_{u1}, \theta_{u2}, \theta_{u3})$ function for classifying users who have low prediction performances or high prediction performances. These functions classify users into three groups; lower performance group, higher performance group and non-selected group. Non-selected group has normal performance. In order to analyze characters of users for each group, we show their rating pattern graphically and their statistical features through statistical tests.

6. Experimental Results

6.1 Prediction Accuracy
The followings are the results of the NBCFA and the CMA that don't consider the number of co-rated items in the user-based method.

Figure 9 shows the prediction results of 100K dataset and 1 million dataset in the user-based according to similarity weight as the Pearson's correlation coefficient. In the results of experiment, the results of prediction accuracy predicted by the CMA are more accurate than those of the NBCFA. The results of 1 million dataset are more accurate than those of 100K dataset and the improvements of the prediction accuracy are similar to all cases.
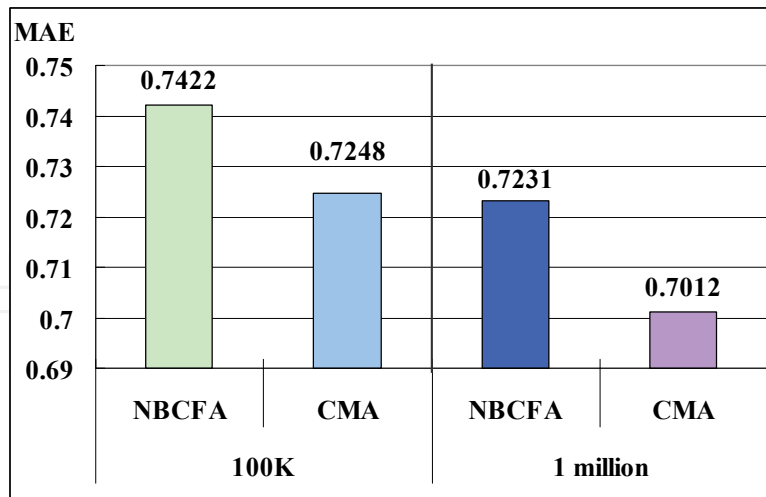
Fig. 9. Prediction results of 100K and 1 million dataset in the user-based

Figure 10 shows the prediction results of 100K dataset and 1 million dataset in the item-based according to similarity weight as the Pearson's correlation coefficient. In the results of experiment, the results of the prediction accuracy predicted by the CMA are more accurate than those of the NBCFA, but the improvements degree of prediction accuracy is less than that of the user-based.
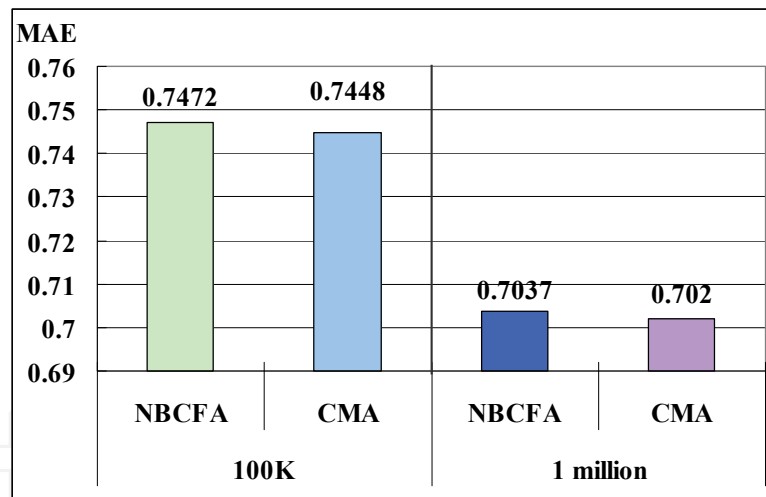

Fig. 10. Prediction results of 100K and 1 million dataset in the item-based

For more analysis, we classify the prediction errors by each user and calculate their MAE. Usually, the MAE measures the accuracy of algorithms used to predict user's preference to items, so the MAE using all predicted ratings means the systematic accuracy. But in our experiment, we take statistical approach for more analysis, so we classify 943 users' prediction values in 100K dataset and 6040 users' prediction values in 1 million dataset and then compute their each MAE. We compare the means of each user's MAE by the result of each algorithm with paired-samples t-test. Table 1 shows the result of paired
Table 1 shows the result of paired-samples t-test of two algorithms with 100K dataset and 1 million dataset.

From the table 1, it is found that the results of user-based have significant differences in the mean statistically, and t-value of the 1 million dataset is bigger than that of 100K dataset, As a result, the prediction accuracy of the CMA is better than that of NBCFA, especially in 1 million dataset. In case of item-based, the result of each algorithm has not statistically significant difference.

Method	dataset	Algorithm	Mean	t-value	Sig.
user-based	100K	NBCFA	0.7691	5.6	0.000**
		CMA	0.7562		
	1million	NBCFA	0.7465	22.417	0.000**
		CMA	0.7285		
item-based	100K	NBCFA	0.761	1.34	0.18
		CMA	0.759		
	1million	NBCFA	0.7171	1.868	0.062
		CMA	0.7164		

    * : $p<0.05$, ** : $p<0.01$

Table 1. Results of paired-samples t-test

Below shows the degrees of the changes depending on the number of co-rated items in two datasets.
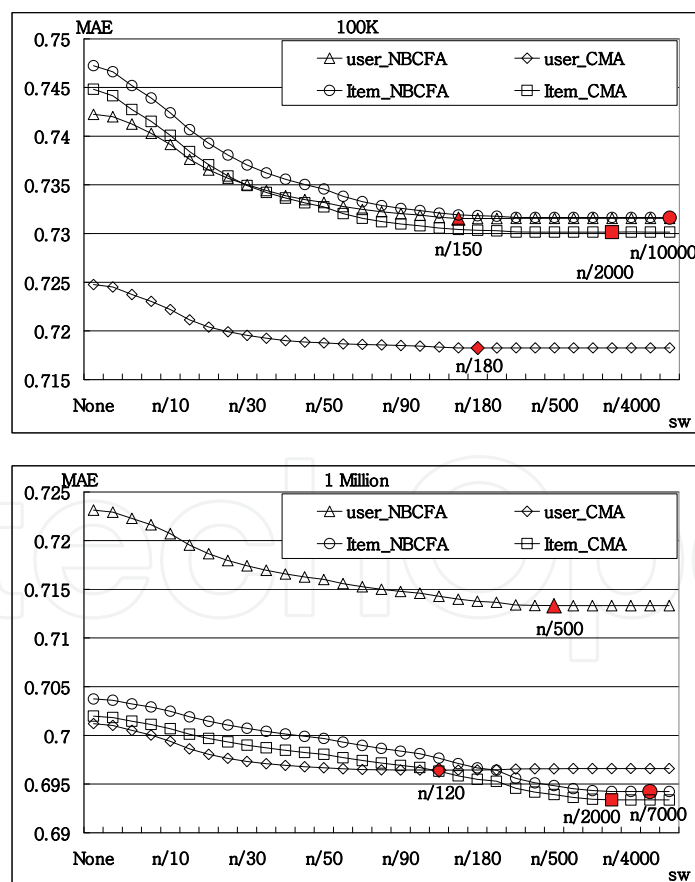

Fig. 11. The accuracy of 100K (above) and 1 million(below) dataset is changing according to the significance weight

Figure 11 shows decreasing curves of MAE according to the significance weight of 100K dataset and 1 million dataset using correlation coefficient as the similarity weight. In the results, it is found that all the results of CMA are better than those of NBCFA. In case of user-based, the decreasing width of MAE is bigger than the result of item-based. The results of user-based are more accurate than those of item-based in 100K dataset, but the results of 1 million dataset show vice versa.

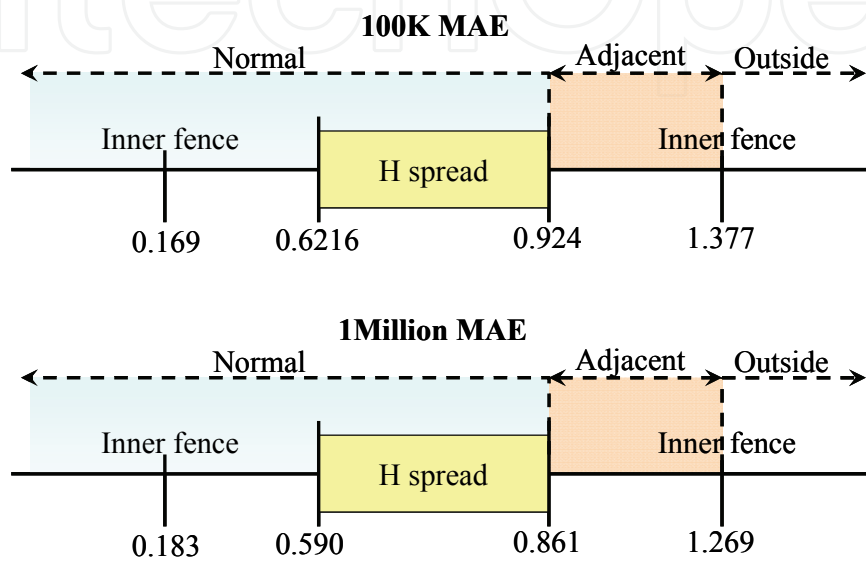## 6.2 Relationship between MAE and SD

Fig. 12. Error bound of 100K and 1million dataset

To classify users who have low prediction accuracy, the EDA approach is applied. We define the users who have low prediction accuracy as outside values, adjacent values, and H-spread who have superior prediction accuracy as normal user groups. And then, abnormal user group is divided into two groups, one is an adjacent values group within inner fence, and the other is an outside values group beyond the inner fence (Fig. 12). Table 2 and table 3 show the results of ANOVA test to compare the prediction accuracy of each group in 100K and 1 million MovieLens dataset.

dataset	Group	N	Mean	Std. Deviation	Min	Max
100K	Normal	707	0.954	0.176	0.314	1.541
	Adjacent	213	1.138	0.212	0.492	1.723
	Outside	23	1.304	0.227	0.681	1.561
	Total	943	1.004	0.207	0.314	1.723
1million	Normal	4380	0.943	0.172	0.139	1.726
	Adjacent	1453	1.134	0.200	0.484	1.719
	Outside	205	1.269	0.247	0.687	1.823
	Total	6038	1.000	0.206	0.139	1.823

Table2. Basic statistics of each group

dataset	-	Sum of Squares	df	Mean Square	F	Duncan
100K	Between	7.670	2	3.835	110.555**	{1}{2}{3}
	Within	32.608	940	0.035		
	Total	40.278	942			
1million	Between	55.319	2	27.660	833.505**	{1}{2}{3}
	Within	200.270	6035	0.033		
	Total	255.589	6037			

*: $p<0.05$, **: $p<0.01$

Table 3. The result of ANOVA test

Table 2 shows the result of the basic statistics of each user group classified by the each user's MAE as the prediction accuracy using the prediction results of 100K and 1million MovieLens dataset. Table 3 shows the result of ANOVA test to compare the means of users' MAE of each group. From the result of the statistical test, it shows that each group has the difference in the means of standard deviations and they are clearly grouped by the multiple comparison with their means of SD as Duncan test. So, it will be possible to use the standard deviations from training dataset for classification criterion of the users who have low prediction performance.

6.3 Rating Pattern

Figure 13 and 14 show rating patterns of users who are classified by $L(\delta_{u1},\delta_{u2},\delta_{u3})$ function applied to 100K and 1 million experimental dataset modified MovieLens dataset.
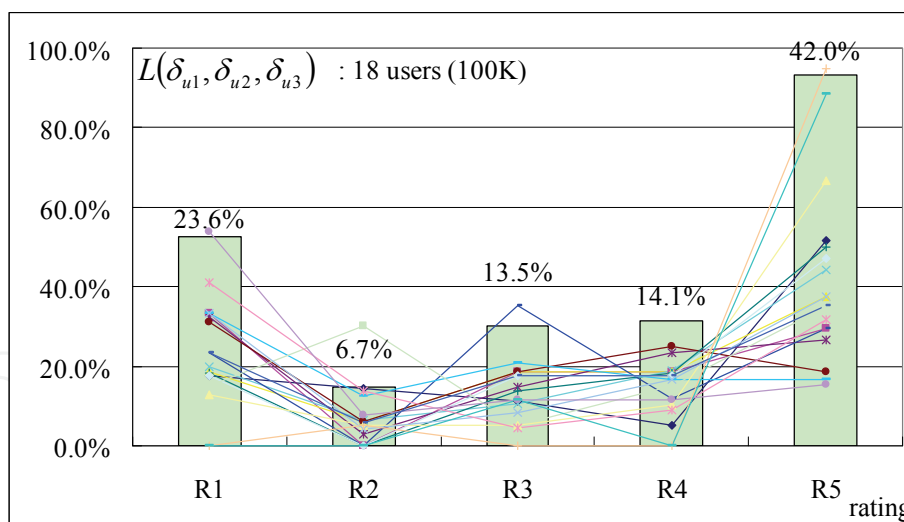

Fig. 13. Rating patterns of classified users by $L(\delta_{u1},\delta_{u2},\delta_{u3})$ function in 100K dataset

The function $L(\delta_{u1},\delta_{u2},\delta_{u3})$ classifies 18 users with 100K dataset and 90 users classified with 1 million dataset. Lines on each chart represent the ratio of each rating rated by every selected user, and bars represent the average ratio of each rating.

As shown in Figure 13 and Figure 14, both rating patterns show that the average ratio of rating forms 'W' in shape. In other words, the number of users classified by R1 and R5 is more than R2, R3, and R4. Users who have lower prediction performance are generally apt

to rate either R1 or R2. As a result, they have bigger deviations of their ratings to items than non selected users.
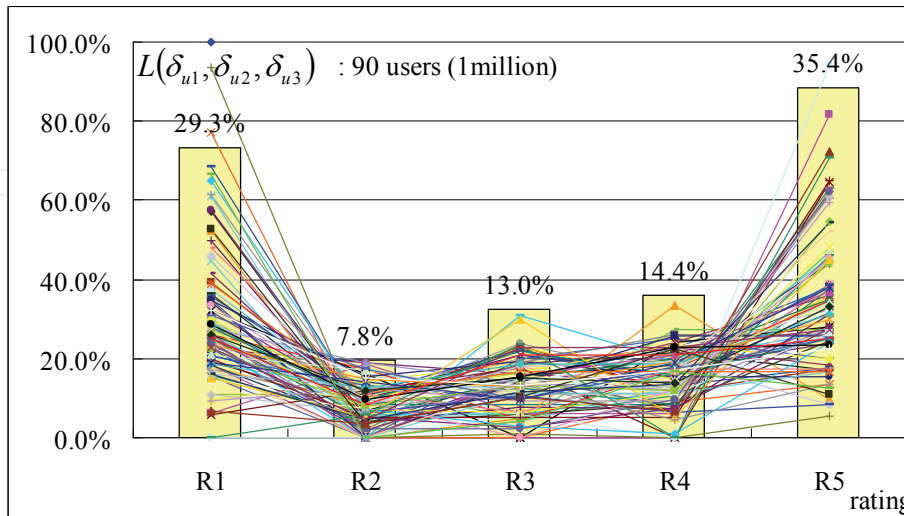

Fig. 14. Rating patterns of classified users by $H(\theta_{u1},\theta_{u2},\theta_{u3})$ function in 100K dataset

Figure 15 and 16 show rating patterns of users classified by $H(\theta_{u1},\theta_{u2},\theta_{u3})$ applied to 100K and 1 million experimental dataset modified MovieLens dataset. The function $H(\theta_{u1},\theta_{u2},\theta_{u3})$ classifies 63 users with 100K dataset and 260 users classified with 1 million dataset.

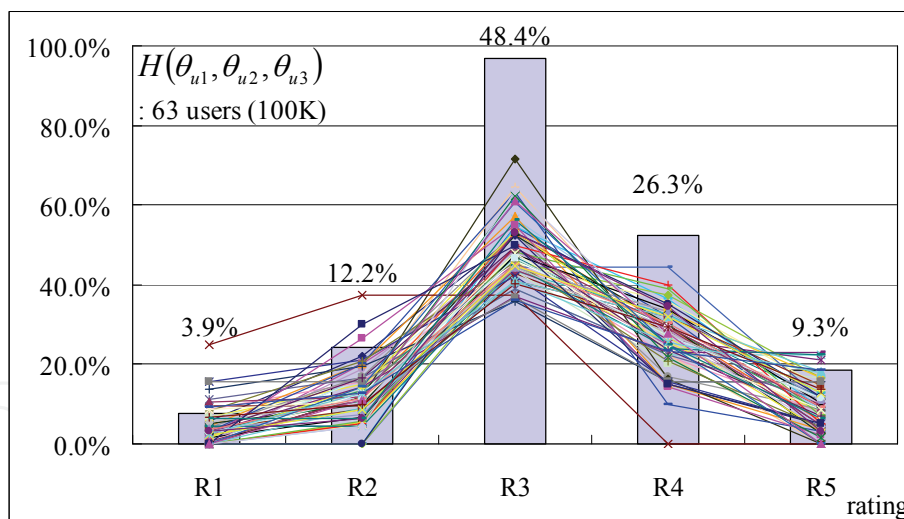Fig. 15. Rating patterns of classified users by $H(\theta_{u1},\theta_{u2},\theta_{u3})$ function in 100K dataset

As shown in Figure 15 and Figure 16, both rating patterns show that the average ratio of rating forms 'hat' in shape. In other words, the number of users classified by R3 and R4 is more than R1 and R5. Users who have higher prediction performance are generally apt to rate one of R2, R3 and R4. As a result, they have smaller deviations of their ratings to items than non-selected users.
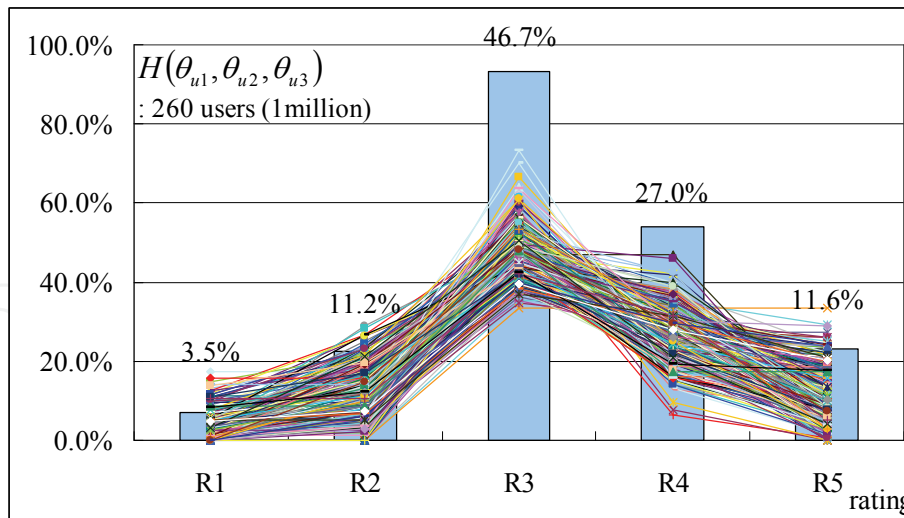
Fig. 16. Rating patterns of classified users by $H(\theta_{u1}, \theta_{u2}, \theta_{u3})$ function in 1million dataset

Table 4 shows the result of ANOVA test over the each user's MAE grouped by two classification functions and non-classified users' group. From Table 4, F values show a meaningful significance statistically. Thus, the means MAE of three groups (H, Non, L) have some differences, but their variances are not so big. As the result of Duncan's Multiple Range Test, we have some difficulties in discriminating the means of users' MAE between group H and group Non, but we can easily distinguish group L from other groups

100K	Group	N	mean	SD	F value	Duncan.
NBCFA	H	63	0.710	0.234	15.99**	{H,Non}{L}
	Non	862	0.780	0.247		
	L	18	1.082	0.283		
	total	943	0.781	0.251		
CMA	H	63	0.707	0.241	12.60**	{H,Non}{L}
	Non	862	0.763	0.242		
	L	18	1.031	0.322		
	total	943	0.764	0.246		

*: p<0.05, **: p<0.01

Table 4. The result of ANOVA over 100K dataset

Similarly, Table 5 shows that three groups are well distinguished statistically than the result of 100K dataset. This result shows that our classification functions can be used as useful tools for detecting or pre-evaluating before prediction process.

1million	Group	N	mean	SD	F value	Duncan.
NBCFA	H	260	0.693	0.187	218.59**	{H}{Non}{L}
	Non	5690	0.744	0.222		
	L	90	1.228	0.314		
	total	6040	0.749	0.231		
CMA	H	260	0.674	0.190	201.20**	{H}{Non}{L}
	Non	5690	0.727	0.218		
	L	90	1.182	0.347		
	total	6040	0.731	0.226		

*: $p<0.05$, **: $p<0.01$

Table 5. The result of ANOVA over 1million dataset

7. Conclusion

As the extensive use of e-commerce through web-site increases, the need for other marketing approach is also increasing more than ever before. Increased concern by on-line company and academia has led to the development of numerous method and techniques that improve the performance of recommender system and promote customers' interests.

In this work we presented our research results in the area of collaborative filtering algorithm and other techniques to improve the performance of recommender systems which are one of the most important tools for the on-line marketing.

From our experimental results, it can be summarized as two main parts. One is algorithmic improvements for prediction accuracy and the other is possibilities of pre-evaluation methods using each user's rating pattern which is already collected in the system.

In the view point of algorithmic improvements, the followings are the results of this study.

First, the prediction performance of CMA on the view of accuracy is superior to that of NBCFA compared to all the results of user-based and item-based approaches. Second, the significance weight which makes up for overestimated preference relationships between target user and his or her neighbours, where the number of co-ratings is so small, contributes greatly to the accuracy of prediction. Also it is necessary to set the extended weighted range rather than existing N/50 ratings. Third, under the extending scale of recommender system, it is more efficient to run the recommender system controlling the increasing numbers of items than to control the increasing numbers of customers. Item-based approach which controls the numbers of items has the more accurate prediction results than those of user-based approach, but our another research which isn't presented on this work shows that the rank correlations between predicted values and real values of user-based approach are more accurate than those of item-based approach. This means that it would be needed to decide one of the two approaches between accurate prediction for rating and customer's preference rank for trade-off. It will be needed that the further research on this topic follows.

In the view point of pre-evaluation, the followings are the results of this study.

This work presents experimental results about setting the error bound for classifying the users who have lower prediction performance before prediction process using collaborative filtering algorithm in the recommender system. Through the statistical analysis, we have

significant results from the prediction result of the error bound. This result is not the approach of improving the prediction performance of algorithm, nor is the method of decreasing the prediction error, but it will be a useful basis for improving algorithms and also understand users' rating pattern better only by using already-existing ratings as pre-information before the prediction of users preferences about items.

Also, we show the evaluation performances of classification functions which classify users with lower or higher prediction accuracy before prediction processes using collaborative filtering algorithm in the Recommender System. With our statistical analysis, we show that applying classification functions before prediction process to the users' preference data would get meaningful results for pre-evaluating users' prediction accuracy. This is especially useful to detect users who have lower prediction accuracy before time consuming prediction process. Additionally, it would be helpful to protect recommender system from malicious attacker. However, this result also does not suggest the way to improve the users who have been classified by proposed classification functions or make clear the reason why these results are produced. It will be expected that further studies must be made in the near future.

In conclusion, it seems that in the near future, the field of recommender systems in e-business will attract even more interest from the research community. The increasing adoption of recommender system as a main tool for on-line marketing by prominent company and diverse field denotes its strategic role in on-line shopping environment.

8. References

Adomavicius, G. and Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 6, pp. 734-749, ISSN: 1041-4347.

Breese, J., Heckerman, D. and Kadie, C. (1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering, *Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence*, ISBN: 1-55860-555-X, pp. 43-52. July, San Francisco, CA.

Condliff, M., Lewis, D., Madigan, D. and Posse, C. (1999). Bayesian Mixed-Effects Models for Recommender Systems, *ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation*, Aug.

Han, U., Yang, G., Yoo, J., Chung, Y., Lee, H. (2008). A Proposal on the Error Bound of Collaborative Filtering Recommender System, *Proceeding of the 2nd International Conference on Multimedia and Ubiquitous Engineering*, ISBN: 978-0-7695-3134-2, pp. 348-354, Apl., Busan, Korea.

Herlocker, J., Konstan, J., Terveen L. and Riedl, J. (2004). Evaluating collaborative filtering recommender systems, *ACM Transactions Information Systems*, vol. 22, no. 1, pp.5-53, ISSN: 1046-8188.

Lee, H., Lee, S., Chung, Y. (2007a). A Study on the Improved Collaborative Filtering Algorithm for Recommender System, *Proceeding of the 15th International Conference on Software Engineering Research, Management and Applications*, ISBN: 0-7695-2867-8, pp. 297-304, Aug., Busan, South Korea

Lee, H., Lee, S., Chung, Y. (2007b). An Improved Algorithm for Collaborative Filtering Recommender System, *ACIS International Journal of Computer & Information Science*, vol. 8, no. 3, pp. 444-453, ISSN: 1525-9293.

Lee, S., Kim, S. and Lee, H. (2007). Pre-Evaluation for Detecting Abnormal Users in Recommender System, *Journal of the Korean Data & Information Science Society*, vol. 18, no. 3, pp. 619-628, ISSN: 1598-9402.

Linden, G., Smith, B. and York, J. (2003). Amazon.com Recommendations: Item-to-Item Collaborative Filtering, *IEEE Internet Computing*, vol. 7, no. 1, pp.76-80, ISSN: 1089-7801.

Popescul, A., Ungar, L., Pennock, D. and Lawrence, S. (2001). Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments, *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, ISBN: 1-55860-800-1, pp. 437-444, Aug., Seattle, WA .

Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P. and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work*, ISBN: 0-89791-689-1, pp. 175-186, Oct., Chapel Hill, NC.

Schafer, J., Konstan, J. and Riedi, J. (1999). Recommender systems in e-commerce, *Proceedings of the 1st ACM conference on Electronic commerce*, ISBN: 1-58113-176-3, pp.158-166, Nov., Denver, Colorado.

Tukey, J. (1977). *Exploratory Data Analysis*, Addison-Wesley, ISBN: 0-201-07616-0.
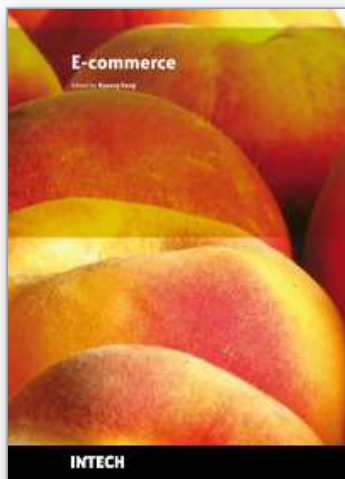
E-commerce

Edited by Kyeong Kang

E-commerce provides immense capability for connectivity through buying and selling activities all over the world. During the last two decades new concepts of business have evolved due to popularity of the Internet, providing new business opportunities for commercial organisations and they are being further influenced by user activities of newer applications of the Internet. Business transactions are made possible through a combination of secure data processing, networking technologies and interactivity functions. Business models are also subjected to continuous external forces of technological evolution, innovative solutions derived through competition, creation of legal boundaries through legislation and social change. The main purpose of this book is to provide the reader with a familiarity of the web based e- commerce environment and position them to deal confidently with a competitive global business environment. The book contains a numbers of case studies providing the reader with different perspectives in interface design, technology usage, quality measurement and performance aspects of developing web-based e-commerce.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Hee Choon Lee, Seok Jun Lee and Young Jun Chung (2010). Improving Performance in Recommender System: Collaborative Filtering Algorithm and User's Rating Pattern, E-commerce, Kyeong Kang (Ed.), ISBN: 978-953-7619-98-5, InTech, Available from: http://www.intechopen.com/books/e-commerce/improving-performance-in-recommender-system-collaborative-filtering-algorithm-and-user-s-rating-patt

INTECH
open science | open minds