J. Appl. Math. & Informatics Vol. **28***(2010), No. 1 - 2, pp. 439 - 450*
Website: http://www.kcam.biz

CLASSIFICATION FUNCTIONS FOR EVALUATING THE PREDICTION PERFORMANCE IN COLLABORATIVE FILTERING RECOMMENDER SYSTEM

SEOK JUN LEE, HEE CHOON LEE AND YOUNG JUN CHUNG*

ABSTRACT. In this paper, we propose a new idea to evaluate the prediction accuracy of user's preference generated by memory-based collaborative filtering algorithm before prediction process in the recommender system. Our analysis results show the possibility of a pre-evaluation before the prediction process of users' preference of item's transaction on the web. Classification functions proposed in this study generate a user's rating pattern under certain conditions. In this research, we test whether classification functions select users who have lower prediction or higher prediction performance under collaborative filtering recommendation approach. The statistical test results will be based on the differences of the prediction accuracy of each user group which are classified by classification functions using the generative probability of specific rating. The characteristics of rating patterns of classified users will also be presented.

AMS Mathematics Subject Classification : 68P10, 94A13
Key words and phrases : Pre-evaluation, collaborative filtering, classification function

1. Introduction

With the advance of IT technologies, the online commerce environment has introduced a recommender system for personalized services [9]. The recommender system used in E-commerce suggests a product, an item or a web service to each customer based on the customer's preference. Since a recommender system can predict customers' preferences and forecast a customer's fondness for a certain item and services, it is used as a conspicuous service that distinguishes an online commerce service from an offline commerce service. In predicting a user's preference, a recommender system obtains enough information from users

Received July 29, 2009. Revised November 6, 2009. Accepted November 14, 2009.

*Corresponding author. This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-332- B00169) and the items they have historically considered and then predict the specific user's preference for a target item and suggest them to user. One of the classic recommender systems is a content-based filtering system which uses textual contents. For instance, in a recommender system for an online movie rental process, there are typically two types of profiles: a movie profile and a customer profile. The movie profile describes a movie categories, actors and actresses, and performances. The customer profile is created by information of the costumer's movie rental history. In this recommendation system, the textual information of items and users are stored in the system and analyzed for the best fit. This type of approach works well in initial systems, but there are some drawbacks to expanding the scales of recommender systems due to the following reasons. First, there are difficulties in converting the features of all traded items into textual data. Additionally, if the number of traded items dramatically increases, it is not easy to automatically convert all items' information into textual forms. Second, since content-based systems only recommend items based on each user's past experience, it cannot help the user choose items for special cases.

This problem is called over-specification for recommendations. These drawbacks are overcome by collaborative filtering recommender systems, which use relationships between users and items that can be represented on a numerical scale (i.e. preference rating). This preference rating information can be collected from tracking clients who surf the web and purchase items. Typically, such types of recommender systems utilize neighbor data, using a set of data that have similar characteristics for recommendations of a target item. This concept used for users or item is called user-based or item-based. 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 users are much higher than items and also because the problem of data scarcity is willing to occur in user-based systems [6] [7].

In this study, to examine mutual relationship between user preference information and item preference information, we will analyze the interest patterns of a user and then predict the preference of the user for an item. This study also focuses on the pre-evaluation of the prediction accuracy which is predicted by user-based collaborative filtering algorithms, Neighborhood Based Collaborative Filtering Algorithm(NBCFA) and Correspondence Mean Algorithm(CMA). To pre-evaluate and classify users before prediction step, we propose classification functions for selecting high accurate users and low accurate users and analyze their rating pattern and prediction result statistically.

2. Related works

2.1. Collaborative filtering

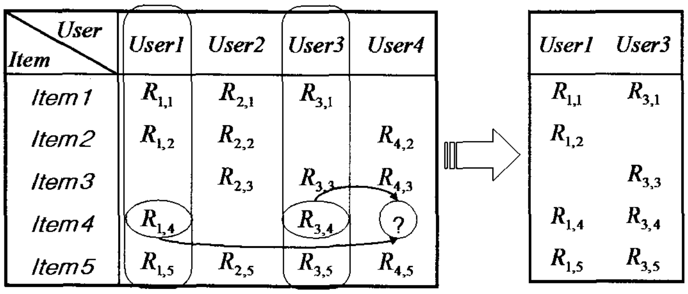
The collaborative filtering approach originated from Tapestry, which is known as the first recommender system. The approach utilizes similar selection criteria that people use when making new choices based on personal experiences. This approach analyzes the preferenential relationship between a target user who will take the recommendation for the target item and his neighbors who have similar preference with the user. This method also analyzes relationships between a target item and neighboring items in the same manner.

2.1.1 Algorithms

There are two approaches to prediction algorithms. One is model-based probabilistic method such as Neural Networks, K-means Clustering, and Association Rules. The other is memory-based approach such as NBCFA and CMA [1] [3] [7].

This study focuses on the performance of classification functions and the characters of classified users' rating patterns predicted by two memory-based algorithms using ratings of each user in hand before applying algorithms.

To predict the preference of a target user about specific items, a neighbor selection process is first carried out. Figure 1 shows the neighbor selection step for predicting the preference of the active user 4 about the specific item 4. Then user 1 and the user 3 are selected as neighbor users of the user 4 because they have been already rated item 4.



Item \ User	User1	User2	User3	User4
Item1	$R_{1,1}$	$R_{2,1}$	$R_{3,1}$	
Item2	$R_{1,2}$	$R_{2,2}$		$R_{4,2}$
Item3		$R_{2,3}$	$R_{3,3}$	$R_{4,3}$
Item4	$R_{1,4}$		$R_{3,4}$	?
Item5	$R_{1,5}$	$R_{2,5}$	$R_{3,5}$	$R_{4,5}$

User1	User3
$R_{1,1}$	$R_{3,1}$
$R_{1,2}$	
	$R_{3,3}$
$R_{1,4}$	$R_{3,4}$
$R_{1,5}$	$R_{3,5}$

FIGURE 1. Neighbor selection step: User 1 and User 3 are selected as neighbors of User 4 for calculating the prediction value of the Item 4.

Before applying the algorithms, preference similarity weights for items between a target user and his neighbors must be defined. In this study, Pearson's correlation coefficient is used to define the weight similarities between them. Equation (1) is the similarity weight used in this study.

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

r_{uj} is the weight similarity 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 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.

To evaluate the classification performance of Pre-evaluation functions and select users with low prediction accuracy or high prediction accuracy, we carry out a prediction process of user preferences for items that are not rated through two prediction algorithms that are NBCFA and CMA. The next equation is the NBCFA which is proposed by the GroupLens [8].

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

In equation (2), The $\hat{U}_x$ denotes the prediction value for the preference of the target user u over the target item x, the $\overline{U}$ is the mean of the all preference ratings that have been rated by 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. The r_uj is the similarity weight of both the user u and the neighbor user j. It is possible however, that some problems might occur, reflecting the tendency of preference for items of target user u and their neighbors during prediction process. Since the preference tendency $\overline{U}$ of the target user u is calculated by all the ratings rated by target user u, the target user shows the same preferences regardless of any other neighbors. Proper adjustment for expressing preference is required since the weight similarity of target user and its neighbor is calculated with only co-rated ratings by them.

The CMA mitigates these problems by using $\overline{U}_{match}$ and $\overline{J}_{match}$, which are the mean of all the $\overline{U}$s and the $\overline{J}$s calculate by co-rated ratings. The following equation is CMA.

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

2.1.2 Evaluation metric

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 [9]. The predictive accuracy metrics measures how close the predicted ratings by algorithm are to the true ratings in the test dataset [2]. In this study, Mean Absolute Error (MAE), one of the predictive accuracy metrics, are used to evaluate the performance of each algorithm, measuring each user's MAE to test the performance of the Pre-evaluation functions.

$$MAE = \frac{1}{N} \sum_j^N |R_{uj} - \hat{R}_{uj}| \tag{4}$$

In this equation, 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.

2.2. Pre-evaluation

According to our previous study, the prediction accuracy of user preference on an item has a close relationship with the generative probability of specific ratings which have been already rated by user before the prediction process. The generative probabilities of specific ratings denoted as δ_{u1}, δ_{u2}, δ_{u3} will be used to define the classification functions of select users whose MAE is lower than non-selected users' MAE from the next equations presented by Lee et al [4] [5].

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

$$\delta_{u3} = \begin{cases} 1, & f_u(\{R_1\} \cup \{R_5\}) \geq f_u(\{R_2\} \cup \{R_3\} \cup \{R_4\}) \\ 0, & elsewhere \end{cases} \quad (5)$$

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

δ_{u1}, δ_{u2}, δ_{u3} are the conditions for defining the classification functions and are shown on the equation (6). It only has values of 1 or 0.

$$L(\delta_{u1}, \delta_{u2}, \delta_{u3}) = \delta_{u1} \cdot \delta_{u2} \cdot \delta_{u3} \quad (6)$$

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 (7) and (8).

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

$$\theta_{u3} = \begin{cases} 1, & f_u(R_3) \geq f_u(\{R_2\} \cup \{R_4\}) \\ 0, & elsewhere \end{cases} \quad (7)$$

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

θ_{u1}, θ_{u2}, θ_{u3} are the conditions that classify users who have high prediction accuracy for defining the classification functions and showed on the equation (8). It also has only the values of 1 or 0.

$$H(\theta_{u1}, \theta_{u2}, \theta_{u3}) = \theta_{u1} \cdot \theta_{u2} \cdot \theta_{u3} \quad (8)$$

3. Related works

3.1. Experimental dataset

To evaluate the performance of each 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 1million dataset. We use both datasets for our research analysis.

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

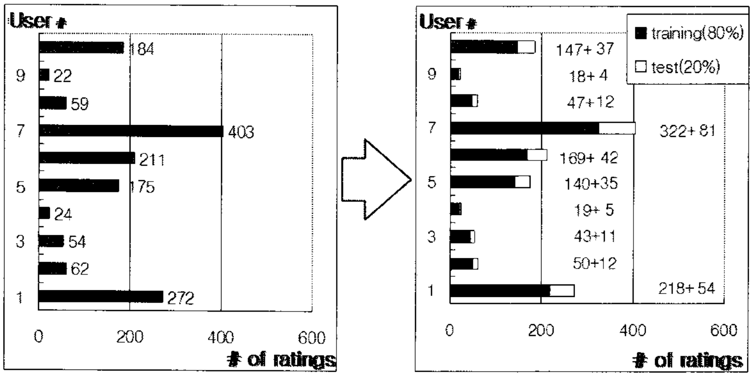


FIGURE 2. Experimental dataset formation.

To generate the prediction accuracy, we divide each dataset as 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 and test dataset for each user. To balance the 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 20% of the test dataset through NBCFA and CMA using 80% of training dataset. Figure 2 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. However, 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 such as ratings on the items before the prediction process for each user's preference.

3.2. Experiment step

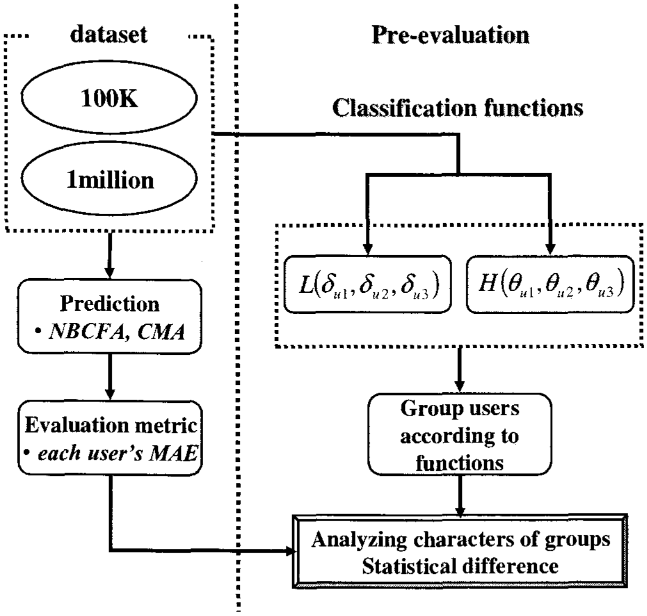

dataset
100K
1million
Prediction
• NBCFA, CMA
Evaluation metric
• each user's MAE
Pre-evaluation
Classification functions
$L(\delta_{u1},\delta_{u2},\delta_{u3})$
$H(\theta_{u1},\theta_{u2},\theta_{u3})$
Group users according to functions
Analyzing characters of groups
Statistical difference

FIGURE 3. Flow diagram of experiment.

Figure 3 shows the experiment flow diagram evaluating the performance of classification functions of the pre-evaluation for the preference prediction errors before the prediction process.

The left side of the vertical dotted line on figure 3 shows the process of prediction domain and two MovieLens datasets are predicted through NBCFA and CMA. The prediction results are then evaluated by each user's MAE.

The right side of the line on figure 3 shows the pre-evaluation process using $L(\delta_{u1},\delta_{u2},\delta_{u3})$ and $H(\theta_{u1},\theta_{u2},\theta_{u3})$ functions 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-selective group. Non-selected group has normal performance. We analyze the characteristics of users for each group, showing their rating pattern graphically and showing their statistical features through statistical tests.

4. Experimental results

4.1. Rating patterns

Figure 4 and 5 shows rating patterns of users who are classified by $L(\delta_{u1},\delta_{u2},\delta_{u3})$ function applied to 100K and 1million experimental dataset modified MovieLens dataset.

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

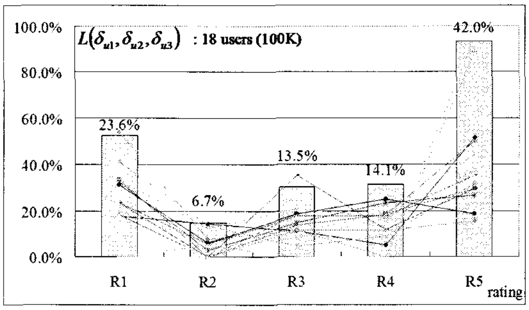


FIGURE 4. Rating patterns of classified users by $L(\delta_{u1}, \delta_{u2}, \delta_{u3})$ function in 100K dataset.

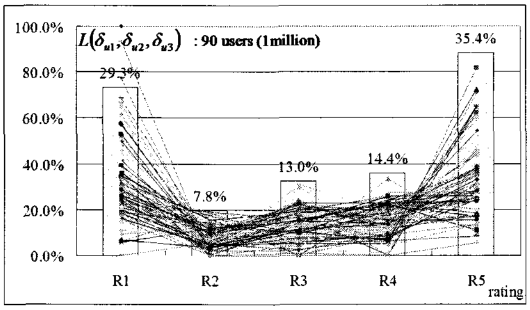


FIGURE 5. Rating patterns of classified users by $L(\delta_{u1}, \delta_{u2}, \delta_{u3})$ function in 1million dataset.

As shown in Figure 4 and Figure 5, both rating patterns show that the average ratio of rating forms a "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.

Figure 6 and 7 show rating patterns of users classified by $H(\theta_{u1}, \theta_{u2}, \theta_{u3})$ applied to 100K and 1million 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 1million dataset.

As shown in Figure 6 and Figure 7, both rating patterns show that the average ratio of rating forms a "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 performances are generally apt to rate one of R2, R3 and R4. As a

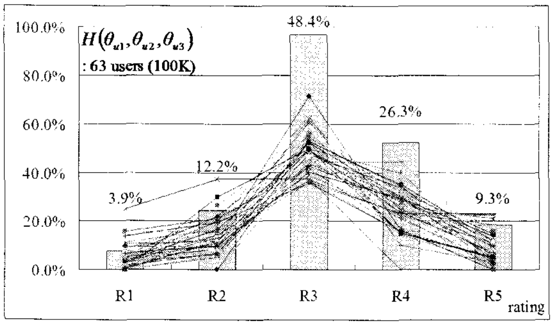


FIGURE 6. Rating patterns of classified users by $H(\theta_{u1},\theta_{u2},\theta_{u3})$ function in 100K dataset.

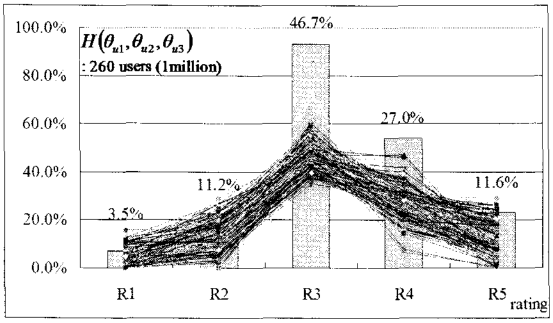


FIGURE 7. Rating patterns of classified users by $H(\theta_{u1},\theta_{u2},\theta_{u3})$ function in 1million dataset.

result, they have smaller deviations of their ratings to items than non selected users.

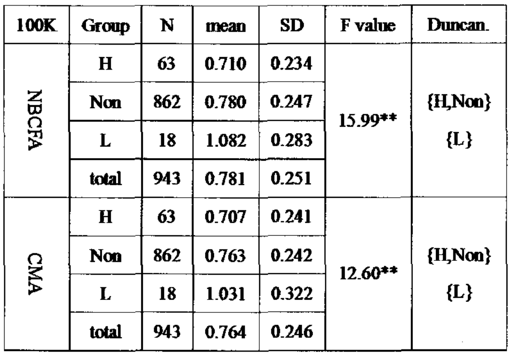
4.2. Statistical characteristics

Table 1 shows the result of the ANOVA test over each user's MAE grouped by two classification functions and non-classified users' group. From Table 1, we find that F values show a meaningful statistical significance. Thus, the means (MAE) of the three groups(H, None, 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 mean of users' MAE between group H and group Non, but we can easily distinguish group L from other groups.

Similarly, Table 2 shows that the three groups are well distinguished statistically than the result of 100K dataset. This result shows that our classification functions can be used as a useful tool for detecting or pre-evaluating before prediction process.

5. Conclusions

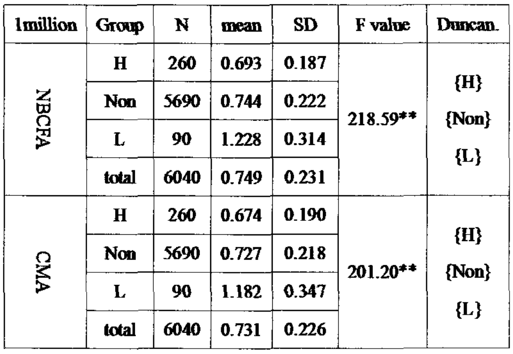
TABLE 1. The result of ANOVA over 100K dataset.



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 2. The result of ANOVA over 1million dataset.



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

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

In this study, we show the evaluation performances of classification functions which classify users with lower or higher prediction accuracy before prediction processes using a memory-based 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 meaningful pre-evaluate users' prediction accuracy. This is especially useful in detecting users who have lower prediction accuracy before time consuming prediction process. Additionally, it would be beneficial in protecting recommender systems from malicious attackers. This study however, does not suggest a way to improve the users who have been classified by proposed classification functions or make clear the reason why these results are produced. It is expected that further studies will made in the near future to elucidate this matter.

References

1. G. Adomavicius and A. Tuzhilin, *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, **17** (2005), 734-749.
2. J. Breese, D. Heckerman and C. Kadie, *Empirical Analysis of Predictive Algorithms for Collaborative Filtering*, Proc. the 14th Annual Conference on Uncertainty in Artificial Intelligence, July, 1998, 43-52.
3. M. Condliff, D. Lewis, D. Madigan, and C. Posse, *Bayesian Mixed-Effects Models for Recommender Systems*, Proc. ACM SIGIR '99 Workshop Recommender Systems: Algorithms and Evaluation, Aug. 1999.
4. S. Lee, S. Kim, H. Lee, *Pre-Evaluation for Detecting Abnormal Users in Recommender System*, J. the Korean Data & Information Science Society, **18** (2007), 619-628.
5. H. Lee, *Enhancement of Collaborative Filtering in Electronic Commerce Recommender Systems*, PhD thesis, Kangwon National University, 2009.
6. G. Linden, B. Smith, and J. York, *Amazon.com Recommendations: Item-to-Item Collaborative Filtering*, IEEE Internet Computing, **7** (2003), 76-80.
7. A. Popescul, L. Ungar, D. Pennock and S. Lawrence, *Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments*, Proc. 17th Conference on Uncertainty in Artificial Intelligence, Aug. 2001, 437-444.
8. P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm and J. Riedl, *GroupLens: An Open Architecture for Collaborative Filtering of Netnews*, Proc. of ACM 1994 Conference on Computer Supported Cooperative Work, Oct. 1994, 175-186.
9. J. Schafer, J. Konstan and J. Riedle, *Recommender systems in e-commerce*, Proc. of the 1st ACM conference on Electronic commerce, Nov. 1999, 158-166.

SEOK JUN LEE received M.S. degree in Industrial Engineering and Ph.D. degree in Business Aministration from Sangji University, Korea. He is currently a full time lecture at Sangji University since 2008. His research focus is e-commerce and ubiquitous manufacturing and knowledge mining in the web. He has investigated issues in recommender system and collaborative filtering algorithm.

Department of Management Information System, Sangji University, Wonju 220-702, Korea.
e-mail: `digitaldesign@sangji.ac.kr`

HEE CHOON LEE received M.S. degrees in Mathematics Education and Statistics from Kyung Hee University, Korea, in 1981 and 1983, respectively. And received Ph.D. degree in Statistics from Kyung Hee University, Korea, in 1987, also received Ph.D. degree in Computer Science from Kangwon National University, Korea, in 2009. He is currently a professor at Sangji University since 1991. His research focus is information retrieval and e-commerce. He has investigated issues in recommender system and collaborative filtering algorithm.

Department of Computer Data Information, Sangji University, Wonju 220-702, Korea.
e-mail: `choolee@sangji.ac.kr`

YOUNG JUN CHUNG received the BS degree in Electrical Engineering from Seoul National University, Korea, in 1974 and the MS and Ph.D. degrees in electrical and computer engineering from the University of Kansas, Lawrence, Kansas, in 1983 and 1988, respectively. Between May 1982 and May 1988, he worked with Advanced Computer Division at KGS, Lawrence, Kansas, USA. Between May 1988 and August 1991, he worked with Computer Division at RANPAC, Rancho California, California, USA. Since August 1991 he has been a faculty at the Department of Computer Science at Kangwon National University, Korea. His research interests are in the areas of wireless and mobile communications systems, sensor networks, Internet applications and services, network security.

Department of Computer Science, Kangwon National University, Chuncheon 200-701, Korea.
e-mail: ychung@kangwon.ac.kr