International Journal of Hybrid Information Technology
Vol. 2, No. 1, January, 2009

The Error Bound of User for Collaborative Recommender System

Uk-Pyo Han
Department of Computer Science, Kangwon National University, Korea
mania2k@kangwon.ac.kr

Gil-Mo Yang
School of Computer IT Engineering, Kangnung National University, Korea
gmyang@wonju.ac.kr

Jae-Soo Yoo
School of Electrical & Computer Engineering, Chungbuk National University, Korea
yjs@chungbuk.ac.kr

Young-Jun Chung
Department of Computer Science, Kangwon National University, Korea
ychung@kangwon.ac.kr

Hee-Choon Lee
Department of Computer Data Information, sangji University, Korea
choolee@sangji.ac.kr

Abstract

We predict accuracy of user's preferences by using memory-based collaborative filtering algorithm in recommender system, and then analyze the results through the EDA approach. The possibilities are presented that prediction accuracy can be evaluated before prediction process by analyzing the results. The classification functions using the generative probability of specific ratings are made, and users are classified by using the classification functions. The prediction accuracies of each classified group are analyzed through statistical tests. The method of setting the Error Bound of users who have high probabilities in low prediction accuracy will be presented.

1. Introduction

The rapid growth of IT infra helps to convert traditional transactions based on the off-line market into e-commerce, internet-based on-line market. Various types of the services, items, and content services sold in off-line come into e-commerce [5][8]. It provides tremendous amount of information and also causes the information overload, which tends to keep people from finding information they want. For mitigating the information overload, e-commerce environment needs some tools to overcome these obstacles. Recommender system is one of alternatives to solve these problems.

This study focuses on collaborative filtering recommender system, especially memory-based collaborative filtering algorithm. In particular, we propose the pre-evaluation method of the user's prediction accuracy before the prediction process of user's preference about the items. To pre-evaluate the prediction accuracy, we use only user's information of already possessed ratings on the items. To show the effect of these methods, we analyze their results through several statistical tests.

2. Recommender system

Recommender system can be defined as a system for automatically suggesting some goods considering customers' interests or tastes [7]. These systems can be adopted on any web sites to make suggestions of items transacted on the web to their customers, who want to find the right things according to their interests or tastes and any types of their interest, for example, books, CDs, contents, places, some information, and so on. Recommender system has several benefits for both businesses and customers. For example, these systems will be a useful tool for businesses because they can automatically find target customers and which items are suit for the customers in the view businesses, and in part of customers, they can find items or goods according to their tastes. So, recommender system is an effective toll for 'customized marketing' for both customers and businesses. Also, they can provide the basic information of customer purchasing patterns on the web to businesses and it will be used for understanding of users' behaviors on their web sites.

Recommender system is usually classified into content-based, collaborative, hybrid approach based on how recommendations are made. There have been many algorithms and implementations of these approaches. Content-based approach systems are used to find the concord textual contents of items transacted on the web with the textual information of users. Collaborative approach systems predict the users' preference, such as explicit rating of the specific items by comparing others who have similar tastes or preferences [1].

2.1. Collaborative filtering

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 predicting users' preferences. One is memory-based and the other is model-based. Memory-based algorithms make prediction 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 networks, and machine learning approaches [1][3][6]. This paper focuses on only memory-based collaborative filtering algorithms.

2.2. Neighbor selection

To predict the preference of active 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 neighbor user of the user 4 because they have been 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.

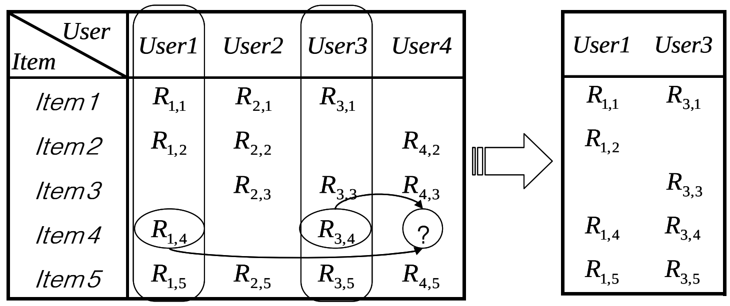


Figure 1. Neighbor selection step: User1 and User3 are selected as neighbors of User4 for calculating the prediction value of the Item4.

2.3. Similarity weight

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 firstly. Breese et al.[2] 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.

Equation 1 is the Pearson's correlation coefficient as similarity weight. It is used for our experiment because the prediction accuracy using this one is more accurate than the other similarity weight as the vector similarity from our previous research.

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

3. Algorithm

In this study we compare the prediction accuracy of the NBCFA proposed by the GroupLens [7].

3.1. NBCFA

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. 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 users u and the neighbor user j.

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

3.2. Evaluation Metric

The predictive accuracy metrics are measured how close the predicted ratings by algorithm are to the true ratings in the test dataset. Mean absolute error (MAE), one of the predictive accuracy metrics, measures the average absolute deviation between a predicted rating and the user's true rating. Mean absolute error has been used to evaluate recommender systems in several cases [2].

$$MAE = \frac{1}{N}\sum_{j}^{N}\left|R_{uj} - \hat{R}_{uj}\right| \tag{3}$$

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.

4. Pre-evaluation

According to our previous study, the prediction accuracy of user's preference on the item has the close relation with the generative probability of specific ratings which have been already rated by user before prediction process. The generative probabilities of specific ratings as δ_1, δ_2, δ_3 will be used to define the classification functions from the next equations presented by Lee et al[4].

$$\delta_1 = \begin{cases} 1, & P(R_5) \ge P(R_2) \\ 0, & elsewhere \end{cases} \qquad \delta_2 = \begin{cases} 1, & P(R_1) \ge P(R_4) \\ 0, & elsewhere \end{cases} \tag{4}$$

$$\delta_3 = \begin{cases} 1, & P(R_1 \cup R_5) \ge P(R_2 \cup R_3 \cup R_4) \\ 0, & elsewhere \end{cases} \tag{5}$$

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

δ_1, δ_2, δ_3 are the conditions for defining the classification functions on the equation(6) and (7). They have only the values of 1 or 0.

$$f(\delta_1 \cdot \delta_2) = \delta_1 \cdot \delta_2 \tag{6}$$

$$f(\delta_1 \cdot \delta_2 \cdot \delta_3) = \delta_1 \cdot \delta_2 \cdot \delta_3 \tag{7}$$

5. Methodology

In this study, we compose the experiment to grasp the relationship of the MAE with the statistical features of ratings on the items that users have already experienced.

5.1. Experimental dataset

This study uses the MovieLens dataset presented by GroupLens for experiment. The GroupLens presents 2 types of the MovieLens dataset. One is 100K dataset and the other is 1million dataset. We use both the datasets for experiment and analysis. To generate the prediction accuracy, we divide each dataset into 80% of training dataset and 20% of test dataset, and predict the 20% of test dataset through NBCFA using the 80% of training dataset. The prediction accuracy will be evaluated by the MAE. But our study uses the each user's MAE which is calculated especially by using ratings of each user in the test dataset instead of using all the ratings in the test dataset. We study the possibility of the pre-evaluation approach using already possessed preference information of users as ratings on the items before the prediction process for each user's preference.

5.2. 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 technique. To set the prediction error fence, we use the concept of the hinge proposed by Tukey to set the fence[9]. 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 2. 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.

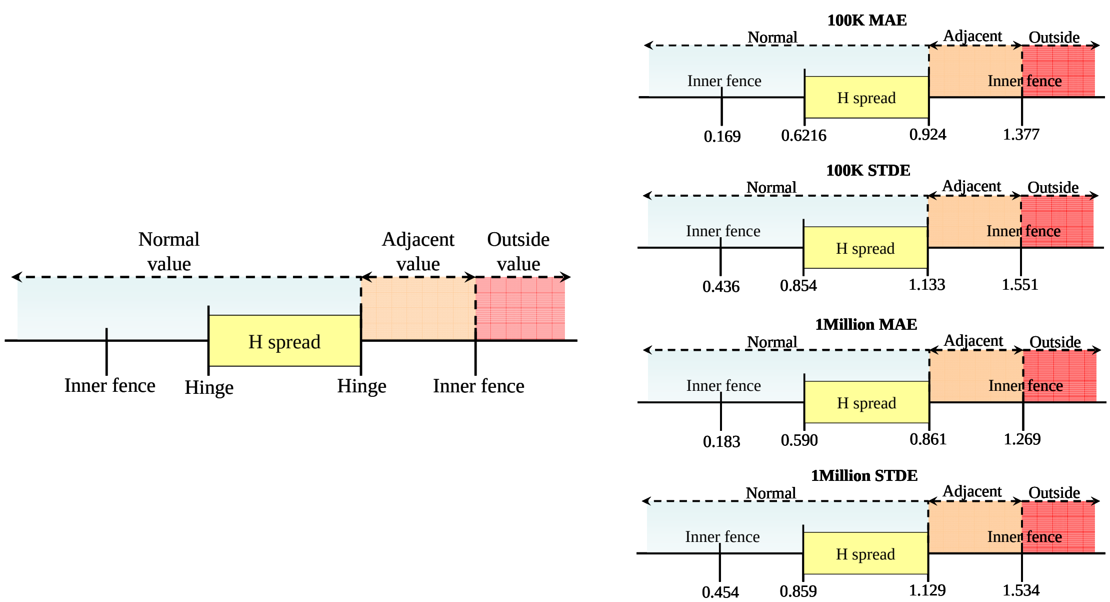


Figure 2. Error bound divided by EDA

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.

5.3. Experiment steps

Figure 3. shows the experiment flow diagram for proposing the possibility of the pre-evaluation for the preference prediction errors before the prediction process.

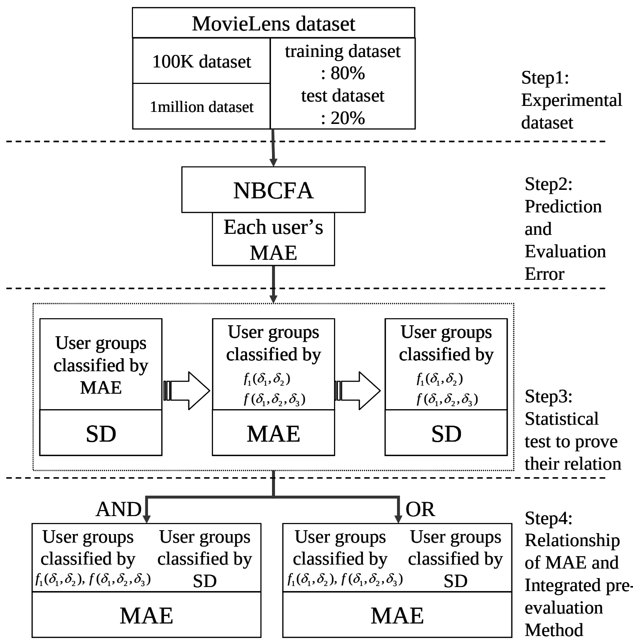


Figure 3. Flow diagram of experiment

In step 1, 100K and 1million MovieLens datasets are divided into 80% of the training dataset and 20% of the test dataset randomly.

In step 2, we evaluate the prediction errors of each user's ratings in the test dataset after the prediction process by using NBCFA.

In step 3, 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.

Second, we use the classification functions as proposed previous work for classifying users as the selected group and non-selected group. We run statistical tests for analyzing the relationships of the classified groups between their prediction errors, and also for analyzing the relationships of their standard deviations calculated from training dataset which is possessed before the prediction step. Through the precedent statistical tests, we propose the possibility of setting the error bound using already-existing information before the prediction of user's preference about the item.

In step 4, we propose the possibility of the pre-evaluation of the user's prediction error through the statistical test of integrating the criteria of classification previously proposed. For integrating the classification criteria, we use 'and' condition and 'or' condition of the classified results of the classification functions and classified by user's standard deviation because these two criteria will come from the already-existing information before prediction step.

6. Experiment and results

6.1. Relationship between MAE and SD

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

Table 1. Basic statistics of each group

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

Table 2. The result of ANOVA test

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 1 shows the result of the basic statistics of each users group classified by the each user's MAE as the prediction accuracy using the prediction results of 100K and 1million MovieLens dataset. Table 2 shows the result of ANOVA test for comparing the means of users' MAE of each group. From the result of the statistical test, it shows that the 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.

6.2. Relationship between MAE and classification functions

Table 3 shows the result of the statistical test between the prediction accuracy of two groups classified by both classification functions as defined in the previous section. Both of all 100K and 1million dataset have the difference in their means of each user's MAE and they have high statistical significance. Especially, 1million dataset has much more statistical significance than 100K dataset. In the first classification function, 23 users are classified in case of 100K dataset and 135 users are classified in case of 1million dataset and in second function 15 users and 91 users respectively are strictly classified.

Table 3. Result of t-test about selected users grouped by classification functions

Function	dataset	Group	N	Mean	Std. Deviation	Mean Difference	t
$f(\delta_1,\delta_2)$	100K	Non-selected	920	0.779	0.242	-0.385	-7.513**
		Selected	23	1.163	0.267		
	1million	Non-selected	5903	0.736	0.220	-0.409	-14.730**
		Selected	135	1.146	0.321		
$f(\delta_1,\delta_2,\delta_3)$	100K	Non-selected	928	0.781	0.243	-0.412	-6.482**
		Selected	15	1.193	0.292		
	1million	Non-selected	5947	0.739	0.222	-0.438	-11.757**
		Selected	91	1.177	0.354		

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

So, the classification functions are statistically useful methods for classifying the users group which are expected that they have lower prediction performance than other groups. The result of the statistical test shows that 1million dataset has much more significance than 100K dataset in statistics.

6.3. Relationship between classification functions and SD

In the result and of the previous experiments, we compare the each user's MAE as prediction accuracy with statistical features of ratings in training dataset from 100K and 1million dataset. Table 4 shows the result of the relationships of the classification functions between the standard deviations of users in training dataset.

Table 4. The result of t-test about the relationship of classification functions and standard deviations from training dataset

Function	dataset	Group	N	Mean	Std. Deviation	Mean Difference	t
$f(\delta_1,\delta_2)$	100K	Non-selected	920	0.993	0.194	-0.456	-11.101**
		Selected	23	1.449	0.200		
	1million	Non-selected	5903	0.988	0.191	-0.515	-31.241**
		Selected	135	1.503	0.189		
$f(\delta_1,\delta_2,\delta_3)$	100K	Non-selected	928	0.996	0.197	-0.489	-9.498**
		Selected	15	1.485	0.236		
	1million	Non-selected	5947	0.992	0.194	-0.549	-24.030**
		Selected	91	1.541	0.217		

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

From the result of table 4, the user groups classified by functions have more differences in their SD and statistically significance compared with MAE. Also, the result of 1million dataset has much more significance than 100K dataset. The relationship between the classification functions with the standard deviations has close correlations. So, these two pre-evaluation methods are useful for classifying the users who have lower prediction performance.

6.4. Relationship between MAE and *'and'* condition

To extend the result of table 4, we analyze the classification performance using both integrated pre-evaluation approaches. Table 5 shows the result of new groups classified by intersected users with classification functions and standard deviation divided by EDA. In the result, the numbers of selected users are smaller than those of the classification functions only applied. Also the statistical significance is lower, but the mean of selected group is higher than the result of the classification functions for itself. It means that this integrated approach classifies the users better who show much lower performance. But the statistical significance is lower because the numbers of selected users are small.

Table 5. Integrated classification functions with standard deviation grouped by EDA under the *'and'* condition

Integrated condition	dataset	Group	N	Mean	Std. Deviation	Mean Difference	t
$f(\delta_1,\delta_2)$ and Adj. values of S.D.	100K	Non-selected	926	0.781	0.244	-0.384	-6.428**
		Selected	17	1.165	0.254		
	1million	Non-selected	5965	0.741	0.226	-0.397	-11.583**
		Selected	73	1.137	0.292		
$f(\delta_1,\delta_2)$ and Out values of S.D.	100K	Non-selected	938	0.786	0.247	-0.439	-3.951**
		Selected	5	1.224	0.314		
	1million	Non-selected	5980	0.741	0.225	-0.469	-11.978**
		Selected	58	1.210	0.297		
$f(\delta_1,\delta_2,\delta_3)$ and Adj. values of S.D.	100K	Non-selected	934	0.784	0.246	-0.434	-5.269**
		Selected	9	1.218	0.284		
	1million	Non-selected	6009	0.743	0.227	-0.480	-7.277**
		Selected	29	1.223	0.355		
$f(\delta_1,\delta_2,\delta_3)$ and Out values of S.D.	100K	Non-selected	938	0.786	0.247	-0.439	-3.951**
		Selected	5	1.224	0.314		
	1million	Non-selected	5980	0.741	0.225	-0.469	-11.978**
		Selected	58	1.210	0.297		

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

6.5. Relationship between MAE and *'or'* condition

Table 6 shows the result of *'or'* condition other than *'and'* condition in table 5 to verify the differencs as integrated groups.

Compared the result of *'and'* condition in table 5, the numbers of the selected users are increase and the means of non-selected group are decrease. It means that this integrated approach classifies the users better who have much higher performance. And the statistical significance is increased because the numbers of selected users increase. From the experimental result, the pre-evaluation approach using already-existing before the prediction process is a very useful way, and the classification functions have very close relations with the standard deviations of ratings which have been already rated by each user, and also their integrated criteria will be expected that they select users who have lower prediction performance well.

Table 6. Integrated classification functions with standard deviation grouped by EDA under the *'or'* condition

Integrated condition	dataset	Group	N	Mean	Std. Deviation	Mean Difference	t
$f(\delta_1,\delta_2)$ and Adj. values of S.D.	100K	Non-selected	707	0.725	0.211	-0.251	-13.355**
		Selected	236	0.976	0.261		
	1million	Non-selected	4529	0.686	0.192	-0.237	-34.260**
		Selected	1509	0.923	0.245		
$f(\delta_1,\delta_2)$ and Out values of S.D.	100K	Non-selected	919	0.778	0.241	-0.403	-8.066**
		Selected	24	1.180	0.274		
	1million	Non-selected	5899	0.736	0.220	-0.402	-14.647**
		Selected	139	1.138	0.322		
$f(\delta_1,\delta_2,\delta_3)$ and Adj. values of S.D.	100K	Non-selected	707	0.725	0.211	-0.251	-13.355**
		Selected	236	0.976	0.261		
	1million	Non-selected	4529	0.686	0.192	-0.237	-34.260**
		Selected	1509	0.923	0.245		
$f(\delta_1,\delta_2,\delta_3)$ and Out values of S.D.	100K	Non-selected	927	0.780	0.242	-0.437	-7.125**
		Selected	16	1.217	0.298		
	1million	Non-selected	5943	0.739	0.222	-0.426	-11.663**
		Selected	95	1.164	0.355		

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

7. Conclusions

This study experiments about setting the error bound for classifying the users who have lower prediction performance before prediction process using memory-based collaborative filtering algorithm in recommender system. Through the statistical analysis, we have significant results from that. This study is not the approach of improving the prediction performance of algorithm, neither is the method of decreasing the prediction error but it will be a useful basis for improving algorithms and the more understanding of users' rating pattern only by using already-existing ratings as pre-information before the prediction of users preferences about items.

8. 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, vol. 17, no. 6, June, 2005, pp. 734-749.

[2] J. Breese, D. Heckerman and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", in Proceedings of the 14th Annual Conference on Uncertainty in Artificial Intelligence, July, 1998, pp. 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", Journal of the Korean Data & Information Science Society, vol. 18, no. 3, Aug. 2007, pp. 619-628.

[5] G. Linden, B. Smith, and J. York, "Amazon.com Recommendations: Item-to-Item Collaborative Filtering", IEEE Internet Computing, vol. 7, no. 1, Jan. 2003, pp.76-80.

[6] A. Popescul, L. Ungar, D. Pennock and S. Lawrence, "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments", 17th Conference on Uncertainty in Artificial Intelligence, Seattle, Washington, Aug. 2001, pp. 437-444.

[7] P. Resnick, N. Iacovou, M. Suchak, P. Bergstorm and J. Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews", In Proceedings of ACM 1994 Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, Oct. 1994, pp. 175-186.

[8] J. Schafer, J. Konstan and J. Riedi, "Recommender systems in e-commerce", In Proceedings of the 1st ACM conference on Electronic commerce, ACM Press, Denver, Colorado, Nov. 1999, pp.158-166.

[9] J. Tukey, "Exploratory Data Analysis", Addison-Wesley, 1977.

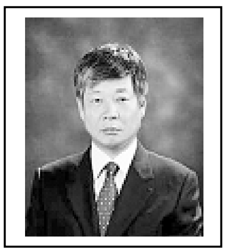
Authors

Uk-Pyo Han received his graduate degree in Computer science from Kangwon National University, Korea, in 1993 and his M.S. and Ph.D. degrees in Computer science from Kangwon National University, in 1996 and 2007, respectively. From 1997 to 2003, he was a manager in Kangwon Internet University, where he was involved in system administration and network design projects. His research interests include the ubiquitous computing, performance evaluation, and control of distributed computer communication systems and wireless sensor networks.

Gil-mo Yang received his BS degree in Computer Science from Hongik University, Korea, in 1983 and his M.S. degree in Mathmatics & Information Engineering from Tokyo University of Agriculture and Technology, Japan, in 1990. Since 1997, he has been doctoral course in Information & Communication from ChungBuk National University and he is currently a professor in School of Computer & Information Engineering, kangnung National University. his research interests are include database, artificial intelligent and ubiquitous computing.

Jae-Soo Yoo received the B.S. degree in Computer Engineering in 1989 from Chonbuk National University, South Korea. And he received the M.S. and Ph.D. degrees in Computer Science in 1991 and 1995 from KAIST(Korea Advanced Institute of Science and Technology), South Korea. He is now a professor in the department of Computer and Communication Engineering, Chungbuk National University, South Korea. His main research interests include database system, sensor data management, location based services, distributed computing and storage management system.

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.



(Corresponding Author) Hee-Choon Lee is a professor in the department of Computer Data Information at Sang-ji University. His research interests are in the areas of information retrieval, e-commerce and wireless network. He has investigated issues in recommender system and collaborative filtering algorithm.