Asia Pacific Journal of Information Systems
Vol. 27 No. 2 (JUNE 2017), 126-138

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2017.27.2.126

An Exploratory Study of Collaborative Filtering Techniques to Analyze the Effect of Information Amount

Hyun Sil Moon[a], Jung Hyun Yoon[b], Il Young Choi[c], Jae Kyeong Kim[d,*]

[a] *Research Doctor, School of Management, Kyung Hee University, Korea*
[b] *Researcher, School of Management, Kyung Hee University, Korea*
[c] *Visiting Professor, School of Management, Kyung Hee University, Korea*
[d] *Professor, School of Management, Kyung Hee University, Korea*

A B S T R A C T

The proliferation of items increased the difficulty of customers in finding the specific items they want to purchase. To solve this problem, companies adopted recommender systems, such as collaborative filtering systems, to provide personalization services. However, companies use only meaningful and essential data given the explosive growth of data. Some customers are concerned that their private information may be exposed because CF systems necessarily deal with personal information. Based on these concerns,
we analyze the effects of the amount of information on recommendation performance. We assume that a customer could choose to provide overall information or partial information. Experimental results indicate that customers who provided overall information generally demonstrated high performance, but differences exist according to the characteristics of products. Our study can provide companies with insights concerning the efficient utilization of data.

Keywords: Information Amount, Recommender System, Collaborative Filtering, Personalization Technique, Performance Analysis

Ⅰ. Introduction

Due to the development of the Internet and smart technologies, the area of e-commerce has been expanded from online to offline. Moreover, thanks to the development of various technologies such as transportations and logistics, e-commerce site can offer a variety of products and services, even health-care services. These proliferation of items have made it difficult for customers to find the specific items they want to buy. Therefore, for the increase of customer satisfaction, many companies have tried to

*Corresponding Author. E-mail: jaek@khu.ac.kr Tel: 8229619355

provide personalization services (Adomavicius and Tuzhlin, 2005; Sarwar et al., 2002).

As a kind of personalization services, recommendation services can directly help customers to find what they are looking for. For instance, Amazon.com provides recommendation services to customers based on their purchase history (Smith and Linden, 2017). And Watcha and Neflix recommend movies based on ratings which are provided by other customers. Although there are many techniques for recommendation services, collaborative filtering (CF) techniques are the most popular and extensively used (Herlocker et al., 2004; Kim et al., 2010). However, the proliferation of smart technologies makes it possible to dissolve the clear division between online and offline shopping, resulting in tremendous increases in transaction records (Bandara and Chen, 2011). That is, recently, massive amounts of data including transaction records have been generated (Beath et al., 2012). The growth of data increases the cost of storage, processing, and analysis (Ji et al., 2012). Therefore, the companies want to store only meaningful and essential data in order to reduce these costs (Dong-hui and Guang, 2013). On the other hand, recommender systems necessarily deal with personal and sensitive information, such as customer profiles (Berkovsky et al., 2007). However, some customers have privacy concerns to personalization services (Lee and Kwon, 2010; Smith et al., 1996). For example, according to Berkovsky et al. (2007), 52% of users feel that all their ratings are sensitive information for them. But these explicit ratings are valuable information to infer customers' preferences in recommender systems. And Aimeur et al. (2008) indicated that 59% of users have worried about their privacy when they use a recommendation service. That is, some users were unwilling to provide personal information such as ratings to recommender systems.

Based on these concerns, in this study, we analyze the effect of information amount on the performance of collaborative filtering systems. For this purpose, we assume that a customer could choose to provide overall or partial information. That is, we assume that some customers decided to provide partial information to recommender systems because they worried about their privacy. In contrast, we also assume that the other customers agreed to provide overall information because they expect high quality recommendation services. Based on the assumption, using various data sets, we evaluate the difference in the recommendation performance between customers who provided overall or partial information. The experiment results indicate that the recommendation performance for customers who provided overall information generally shows higher accuracy but there are some differences according to the characteristics of products. Therefore, our study can provide some insights to companies and researchers concerning the efficient utilization of data in recommender systems.

Ⅱ. Related Work

Recommender systems directly help customers to find items that match their interests. These systems also help companies to sell more items by cross-selling (Sarwar et al., 2002). Among them, collaborative filtering (CF) techniques provide item recommendations or rating predictions based on the transactional data (Herlocker et al., 2004). They are generally classified into the user-based and the item-based CF techniques (Sarwar et al., 2000; Sarwar et al., 2002). The user-based CF techniques use the entire user-item matrix to generate predictions. The main idea is to identify other customers that had similar purchasing records with the target customer and recommend items based on their records. The item-based CF techniques also use the entire user-item matrix, but the main idea is to recommend items based on the similarity between items (Jannach et al., 2010).

They predict the preference of the target user by the following three steps. The first step is to represent whether users bought items using the binary matrix (user-item matrix), R. Each entry *r(i, j)* in R represents whether the i^{th} user bought the j^{th} item. In the user-based CF techniques, the second step is to compute the similarities between the target user and the others. And they select the best similar users to form the neighbor set. In the item-based CF techniques, they compute the similarity between items. The third step is to recommend the top-*N* items among the items that the target customers did not buy. This step is performed based on the similarities between customers (user-based) or items (item-based).

Although these CF techniques were considered as typical methods of recommender systems and widely used, it is hard to find studies concerning the impact of the information amount to the performance of them. However, some studies for predictive models argued that the consideration of information amount can reduce some costs retaining or improving the performance. For instance, Zaki et al. (1997) indicated that the random sampling of raw transaction data can be an effective method for association rule mining. Their experimental results showed that the performance of association rule mining was different according to the sample sizes. Moreover, when they used the proper sampling method, they could reduce the I/O costs and the length of time required to perform a computational process, maintaining the performance of association rule mining. Leskovec and Faloutsos (2006), meanwhile, tried to find good sample of a large massive graph because too many nodes may be prohibitively expensive. In experimental results, back-in-time sampling method showed better performance than the scale-down sampling method. And they found that the 15% sample of the raw data using the proposed method was enough to show similar patterns with the original graph. Among studies for recommender systems, Huang (2007) proposed a method based on an active-learning sampling technique to improve the performance of rating predictions for newly-introduced movies. As a result, the proposed method showed higher performance than the random sampling method and he also conclude that the number of ratings did not guarantee the performance of recommender systems. According to prior studies, the information amount affects the quality and performance of the predictive models. Although there are few studies related with the effect of information amount to recommender systems, Dong-hui and Guang (2013) proposed a method based on variable precision. As they refine the original data into another three datasets according to variable precision, they found that the recommendation performance showed better under lower variable precision values. However, they just compared the recommendation accuracy in the user-based CF techniques according to variable precision.

In this study, we compare the recommendation performance according to the information amount in the user-based and item-based CF techniques. Moreover, as we assume that the information amount is differ according to the customers' privacy concerns, we also try to analyze the effect of information amount to the recommendation performance with some aspects such as privacy concerns, methods and product characteristics. Through our study, we expect that our study can provide some insight to companies and researchers concerning the information amount of recommender systems.

Ⅲ. Methodology

3.1. Overall Procedure

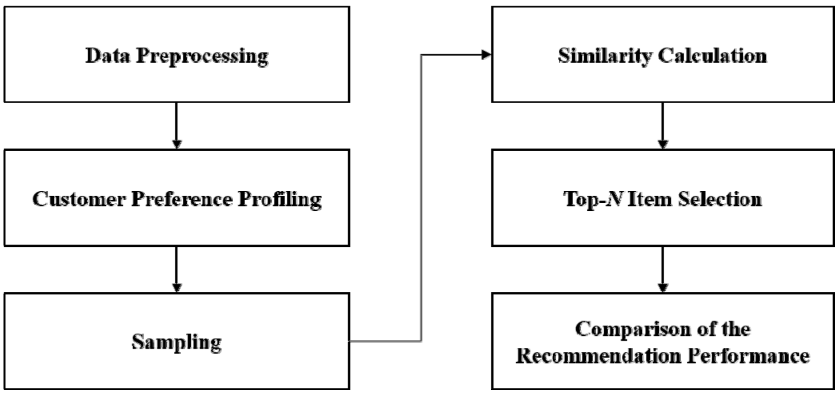
The key underlying purpose of our methodology is to analyze the recommendation performance according to the information amount which can be differ by customers' tendency related with privacy concerns. Based on this purpose, the overall procedure of the proposed methodology is constructed as shown in <Figure 1>. First, we preprocess the experimental data and profile customers' interests from these preprocessed data. Then, we extract samples using stratified random sampling. Through sampling, we classify customers into the *overall* and *partial* groups. In the user-based CF techniques, we use similarity between customers to identify neighbors of the target customer. And we use similarity between items in the item-based CF techniques. After the neighbors and similar items are identified, we select the top-*N* items based on these similarities and compare the recommendation performance between the *overall* and *partial* groups.

3.2. Data Preprocessing and Profiling

Data processing is defined as the manipulation and conversion of the data into particular forms that yield better results (Provost and Fawcett, 2013). To recommend items, CF techniques need transactional data. Therefore, from raw dataset, we construct transactional dataset which is consist of time, item codes, and customer codes. Based on preprocessed data, we profile the customers' preferences. In CF techniques, profiling describes the customers' interests about items. The results of profiling are represented as a user-item matrix. For profiling customers' preferences, we denote $C = \{c_1, c_2, \cdots, c_n\}$ for the customers and $I = \{i_1, i_2, \cdots, i_m\}$ for the items. Then, we form a $n \times m$ matrix which has binary value r_{ij} and is represented as R. Thus, profiling customers' preferences is defined as:

$$r_{i,j} = \begin{cases} 1, \text{if the customer } c_i \text{ purchases item } i_j \\ 0, \text{otherwise} \end{cases} \quad (1)$$

where c_i means the *i-th* customer, and i_j means



<Figure 1> Overall Procedure

the *j-th* item. Hence, when customer c_i purchases item i_j, the customer's preference $r_{i,j}$ is 1, otherwise, it is zero.

3.3. Sampling

Using the user-item matrix, we extract samples based on the stratified random sampling method, which involve the extraction of a random sample from each stratum. <Figure 2> shows an example of 80% sampling method in our study.

In <Figure 2>, n denotes the number of customers. For sampling, we first randomly divide whole customers into two groups. The one is the *overall group* which provides whole information to recommender system while the *partial group* provides part of their preferences. For example, in <Figure 2>, customers from $c_{\frac{2}{n}+1}$ to c_n are the *partial group*, and the others are the *overall group*. Then, for the *partial group*, we randomly remain some records according to the sampling level. In <Figure 2>, although the customer c_n has 5 records in original dataset, we randomly extract 4 records using 80% sampling because he or she denies to provide overall information.

	i_1	i_2	…	i_m	# of purchases
c_1	1	0		1	10
c_2	0	1		0	12
⋮					
$c_{\frac{n}{2}+1}$	0	1		1	15
⋮					
c_n	1	1		1	5

80% sampling

	i_1	i_2	…	i_m	# of purchases
c_1	1	0		1	10
c_2	0	1		0	12
⋮					
$c_{\frac{n}{2}+1}$	0	1		**0**	**12**
⋮					
c_n	1	**0**		1	**4**

<Figure 2> Example of 80% Samplin

3.4. Similarity Calculation

In order to recommend items, we have to calculate the similarity between target customers and other like-minded customers in the user-based CF techniques and between items in the item-based CF techniques. Although there are many metrics for similarity calculation, we use Pearson's correlation coefficient, which is a common metric in studies for CF techniques. Moreover, Herlocker et al (2004) indicated that CF techniques based on the Pearson coefficient record higher performance than the other metrics. The similarity equation for the user-based CF techniques sim_{user} is:

$$sim_{user}(c,d) = \frac{\sum_{p\in I}(r_{c,p}-\overline{r_c})(r_{d,p}-\overline{r_d})}{\sqrt{\sum_{p\in I}(r_{c,p}-\overline{r_c})^2}\sqrt{\sum_{p\in I}(r_{d,p}-\overline{r_d})^2}} \quad (2)$$

where $r_{c,p}$ means that the rating of the customer c for the item p, and $\overline{r_c}$ means the average rating of the customer c. Then, utilizing the similarity between customers, we determine k-like minded neighbors of the target customer. In the item-based CF techniques, we calculate the similarity sim_{item} following as

$$sim_{item}(p,q) = \frac{\sum_{c\in C}(r_{c,p}-\overline{r_p})(r_{c,q}-\overline{r_q})}{\sqrt{\sum_{c\in C}(r_{c,p}-\overline{r_p})^2}\sqrt{\sum_{c\in C}(r_{c,q}-\overline{r_q})^2}} \quad (3)$$

Likewise, $r_{c,p}$ means that the rating of the customer c for the item p. and means the average rating of the item p.

3.5. Selection of Top-N Items

After the calculation of similarities, we can predict the probability that a customer will buy an item. In the user-based CF techniques, the probability is estimated based on the ratings of *k*-like minded neighbors. In contrast, it is estimated based on the similarity between items in the item-based CF techniques. Therefore, in this study, it is calculated using a weighted average of the ratings by the neighbor sets or similar items. The equations are following as:

$$\hat{r}(c,p) = \overline{r_c} + \frac{\sum_{d \in N} sim(c,d) * (r_{d,p} - \overline{r_d})}{\sum_{d \in N} sim(c,d)} \quad (4)$$

$$\hat{r}(c,p) = \overline{r_p} + \frac{\sum_{q \in I} sim(p,q) * (r_{c,q} - \overline{r_q})}{\sum_{q \in I} sim(p,q)} \quad (5)$$

where *c* means the target customer, *p* means an item, and *N* is referred to the nearest neighbor set of customer c. Therefore, the equation (4) is for the user-based CF techniques and the equation (5) is of the item-based CF techniques. Based on these probabilities, we select the top-*N* items that the target customer is likely to prefer.

3.6. Comparison of the Recommendation Performance

To evaluate the recommendation performance, we first consider two metrics, i.e., recall and precision, which are used extensively for evaluating the quality of information retrieval in general (Herlocker et al., 2004; Jannach et al., 2010; Sarwar et al., 2000). They are represented as follows:

$$Recall = \frac{|\{recommended\ items\} \cap \{purchased\ items\}|}{|\{purchased\ items\}|} \quad (6)$$

$$Precision = \frac{|\{recommended\ items\} \cap \{purchased\ items\}|}{|\{recommended\ items\}|} \quad (7)$$

These two metrics present the extent to which the recommender system supports customers' decision making. Although they are easy to compute and understand, increasing the number of recommended items increases recall and decreases precision. These facts are critical for evaluating the recommendation performance. Thus, to precisely evaluate the recommendation performance, we use the F1 metric, as shown below, which is the harmonic mean of recall and precision (Herlocker et al., 2004; Jannach et al., 2010; Sarwar 2000).

$$F1 = \frac{2 * recall * precision}{recall + precision} \quad (8)$$

Ⅳ. Experiments

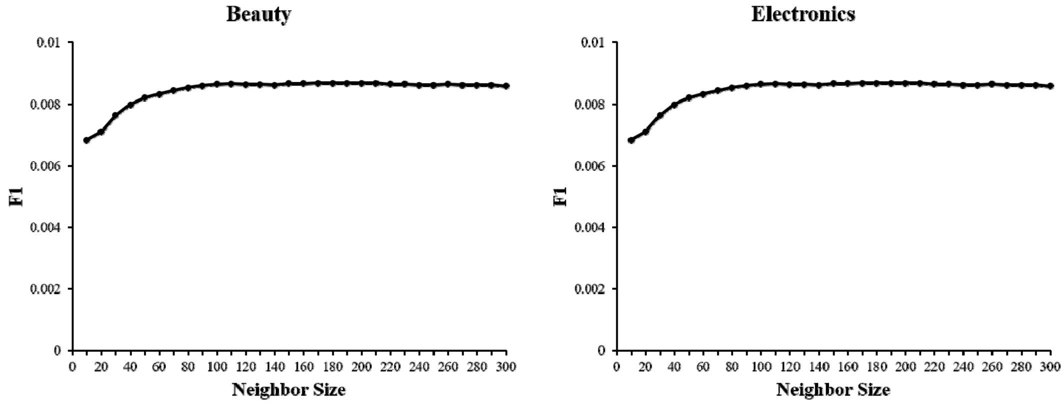
4.1. Data Set

For our experiment, we use two kinds of real transactional data. Detailed descriptions of the data are provided in <Table 1>.

First, we use dataset which consists of reviews Amazon.com collected by McAuley et al. (2015). Because customers who purchase the item through Amazon.com can only write reviews for it, we consider review records as purchase history (Kim and Srivastava, 2007). Next, to compare the differences according to the level of product involvement as a kind of the product characteristics which affects customers' perceptions and behavior, we select the beauty and the electronics categories (Beneke et al., 2016). In general, the beauty products are considered as the low involvement products because they are easy to buy with inexpensive price (Gbadamosi, 2009). In contrast, the electronics are generally high

<Table 1> Data Descriptions

	Beauty	Electronics
Trans. time periods	2012.06.01~2014.05.31 (2 years)	
no. of customer	8,162	36,481
no. of products	3,910	4,013
avg. purchases per customer	4.7175	3.8797
avg. sales per product	9.8476	35.2691
no. of transactions	38,504	141,535



<Figure 3> Sensitivity of the Neighbor Size

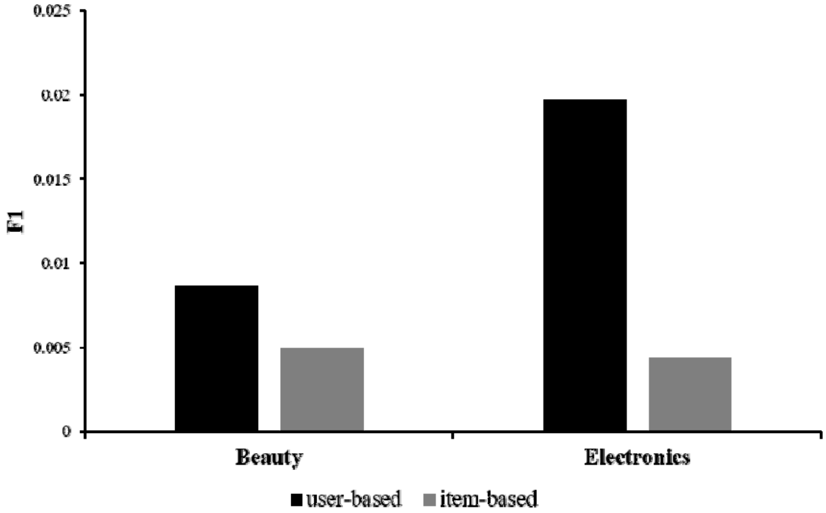
involvement products due to their higher prices and brand commitment (Lockshin et al., 1997; Michaelidou and Dibb, 2008). Therefore, using high and low involvement datasets, we can also compare the effect of information amount with the level of involvement.

To precisely compare each recommendation performance, we adopted the holdout method. In the holdout method, the raw data is divided into a training set and a test set. The training set is used to train CF techniques and the test set is used to evaluate the performance of each recommendation (Herlocker et al., 2004). In this study, the training set contains the first 70% periods of the data set's entire time, and the test set contains the remaining. We design an application that we programmed in R, and all of the experiments are implemented using a personal computer that had an Intel Core i5 (1.80 GHz) processor and 4 GB RAM running Windows 10 (64-bit) OS.

4.2. Experimental Results

Before evaluating the recommendation performance, we first conduct experiments to determine the number of neighbors *k*, which is a critical parameter in the user-based CF techniques. In the user-based CF techniques, if the number of neighbors *k* is too small or large, the accuracy of the recommendation may be decreased (Jannach et al., 2010). In all the experiments, we recommend 5 items because average purchases per customer are approximately 5 in both data sets. The F1 values according to the neighbor sizes are shown in <Figure 3>.

As shown in <Figure 3>, the performance of rec-



<Figure 4> Comparison of Overall Accuracy

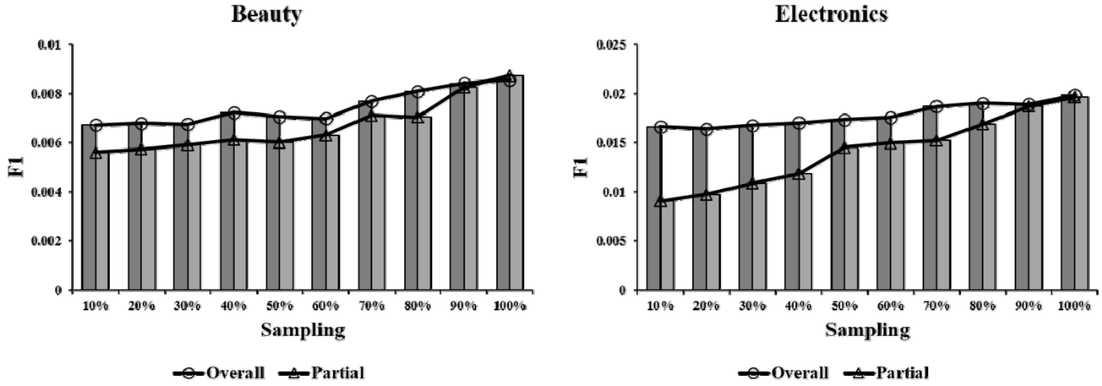
ommendation is gradually increased according to the number of neighbors. However, the recommendation performance in the beauty products is imperceptibly decreased after the neighbor size of 170. In other words, the opinions of many neighbors over 170 are noise to recommender systems. In contrast, the performance in the electronics shows stable increases until the neighbor size of 300. Moreover, until the neighbor size of 100, its recommendation performance is dramatically increased. Therefore, we conclude that 100 neighbors are needed for stable recommendations of electronics. Through experiments, we select the neighbor size of 170 for the beauty products and 300 for the electronics.

According to these neighbor sizes, we first compare the recommendation performance in beauty products and electronics between user-based and item-based CF techniques. <Figure 4> shows the graphs of the experimental results.

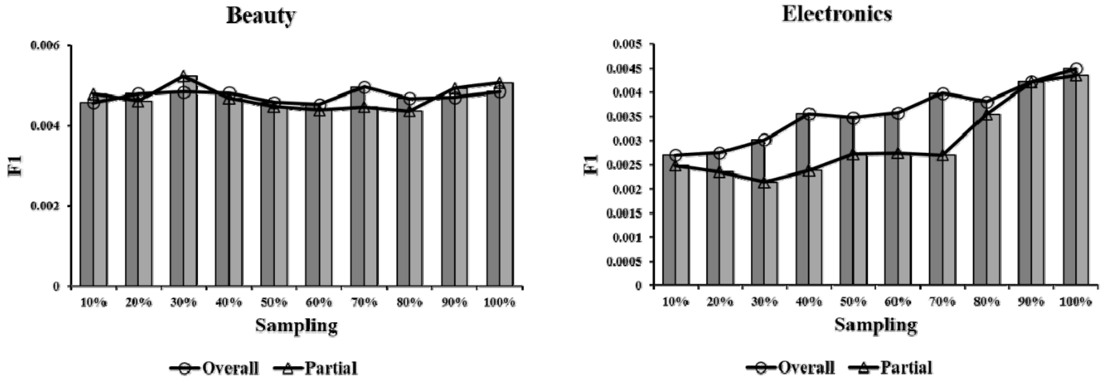
In <Figure 4>, the F1 value of the user-based CF techniques always show higher than the item-based CF techniques. However, there are some differences between methods and the product characteristics. According to <Figure 4>, the performance of the user-based CF techniques is higher in the electronics but the item-based CF techniques perform better in the beauty products. Through these results, we find that the customers in the electronics are sensitive to others' opinions because the user-based CF techniques originally use the opinions of nearest neighbors. In contrast, we also find that the customers in the beauty products consider both own purchase history and the others' opinions.

Next, we analyze and compare the recommendation performance in order to evaluate the impact of the information amount. <Figure 5> shows the graphs of the experimental results using the user-based CF techniques according to the sampling sizes, i.e., the information amount.

<Figure 5> shows that the F1 value of 100% sampling is always the highest in both product categories. This result is a natural outcome because the other sampling sizes do not provide sufficient information to recommend items. Moreover, through experiments, we find that customers who provided sufficient information to recommender systems can re-



<Figure 5> The User-Based CF Performance According to Sample Size



<Figure 6> The Item-Based CF Performance According to Sample Size

ceive more accurate and robust personalization services. Although the user-based CF techniques require other customers' opinions to recommend items, the recommendation performance for the *overall group* is always higher than the *partial group*. That is, even if the other customers provide little information, the customers who provided sufficient information to recommender systems can receive proper recommendations. On the other hand, the performance differences between the *overall* and *partial* groups are increased in electronics category. That is, the smaller information amount they provided the user-based CF techniques for electronics, the poorer recommendations they received. Although there are some reasons for this result, the main is that the user-based CF techniques can not find the proper neighbors for the target customers when he or she provides insufficient information. Therefore, we conclude that it is hard to infer the preferences and recommend items with small information of the target user.

In the results of the item-based CF techniques, they show different tendencies. <Figure 6> shows the experiment results using the item-based CF techniques.

In the left side of <Figure 6>, the recommendation performance between the *overall* and *partial* group is indifferent. These results indicate that the item-based CF techniques can provide stable recommendations regardless of the information amount. The item-based CF techniques try to find similar items using overall customers' opinions. Therefore, if there were sufficient purchase records per product, they can recommend items. Likewise, in electronics category, although there are differences between two groups, the recommendation performance for the *overall* group is also decreased as the sampling size is decreased. Therefore, we can conclude that the performance of item-based CF techniques is affected by the whole data size rather than the information amount per customer.

Ⅴ. Conclusion

In this study, we analyze and compare the recommendation performance according to customers' intentions to provide their information. Consequently, in the user-based CF techniques, customers who provided overall information to recommender systems are guaranteed high recommendation quality. Moreover, they may not be affected by customers who provided partial information to recommender systems. However, in the item-based CF for beauty products, the *partial* group can receive a similar quality of recommendations with the *overall* group. And, in electronics, the performance of *overall* group is affected by the others' information amount including the *partial* group. Therefore, we can conclude that the recommendation performance in the item-based CF is determined by the whole data size, not by the size of each customer's data.

These results provide some significant implications and insights. First, through our experiments, we find that the item-based CF techniques show robust performance regardless of the information amount per customer. Therefore, the companies which want to provide recommendation services with the item-based CF should consider the way to increase the whole data size, i.e., transaction records. Second, because the user-based CF techniques can provide better recommendation to the *overall* group, in spite of the privacy concerns, the companies should encourage customers to provide their personal information. Moreover, using these experimental results, they can explain why users should provide personal information for recommender systems. Third, although the user-based CF techniques always show higher performance than the item-based CF techniques, we find that customers who consider the low involvement products are also affected by own purchases history not only the others' opinions. Therefore, we suggest that studies for recommender system to improve the recommendation performance should consider the product characteristics such as the level of product involvement. Lastly, based on these results, we conclude that the time and cost of recommender systems can be decreased because recommender systems can be operated using only meaningful and loyal customers' information. As mentioned above, some previous studies have adopted sampling techniques for these advantages (Huang, 2007; Leskovec and Faloutsos, 2006). However, as the effect of information amount is different according to some factors, researchers and companies should consider the characteristics of methods and products when they try to reduce the costs related with recommender systems.

Although our study can provide some insight to researchers and companies concerning the efficient utilization of data for personalization services, it has some limitations. We only test the recommendation performance of CF techniques based on accuracy metrics. And we only consider the level of involvement as the product characteristics. In future work, we will test other recommender systems with both accuracy and non-accuracy metrics such as diversity. And we will expand our study to consider other factors such as prices, brand loyalty, and so on.

Acknowledgement

This work was supported by a grant from Kyung Hee University in 2013.(KHU-20130544)

<References>

[1] 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, 17*(6), 734-749.

[2] Aimeur, E., Brassard, G., Fernandez, J. M., Onana, F. S. M., and Rakowski, Z. (2008). Experimental Demonstration of a Hybrid Privacy-Preserving Recommender System. *Proceedings of the Third International Conference on Availability, Reliability and Security*, 161-170.

[3] Bandara, U. and Chen, J. (2011). Ubira: A Mobile Platform for an Integrated Online/offline Shopping Experience. *Proceedings of the 13th International Conference on Ubiquitous computing*, 547-548.

[4] Beath, C., Becerra-Fernandez, I., Ross, J., and Short, J. (2012). Finding Value in the Information Explosion. *MIT Sloan Management Review, 53*(4), 18-20.

[5] Beneke, J., de Sousa, S., Mbuyu, M., and Wickham, B. (2016). The effect of negative online customer reviews on brand equity and purchase intention of consumer electronics in South Africa. *The International Review of Retail, Distribution and Consumer Research, 26*(2), 171-201.

[6] Berkovsky, S., Eytani, Y., Kuflik, T., and Ricci, F. (2007). Enhancing Privacy and Preserving Accuracy of a Distributed Collaborative Filtering. *Proceedings of the 2007 ACM conference on Recommender systems*, 9-16.

[7] Dong-hui, Y. and Guang, Y. (2013). Bigger Data Set, Better Personalized Recommendation Performance? *Proceedings of the 2013 International Conference on Management Science and Engineering*, 28-35.

[8] Gbadamosi, A. (2009). Cognitive dissonance: The implicit explication in low-income consumers' shopping behaviour for “low-involvement” grocery products. *International Journal of Retail and Distribution Management, 37*(12), 1077-1095.

[9] Herlocker, J., Konstan, J., Terveen, L., and Riedl, J. (2004). Evaluating Collaborative Filtering Recommender Systems. *ACM Transactions on Information Systems, 22*(1), 5-53.

[10] Huang, Z. (2007). Selectively acquiring ratings for product recommendation. *Proceedings of the Ninth International Conference on Electronic Commerce*, 379-388.

[11] Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender Systems: An Introduction.* Cambridge University Press, New York.

[12] Ji, C., Li, Y., Qiu, W., Awada, U., and Li, K. (2012). Big Data Processing in Cloud Computing Environments. *Proceedings of the 12th International Symposium on Pervasive Systems, Algorithms and Networks*, 17-23.

[13] Kim, H.K., Jang, M.K., Kim, J.K., and Cho, Y.H. (2010). A New Item Recommendation Procedure using Preference Boundary. *Asia Pacific Journal of Information Systems, 20*(1), 81-99.

[14] Kim, Y. and Srivastava, J. (2007). Impact of Social Influence in e-commerce Decision Making. *Proceedings of the Ninth International Conference on Electronic Commerce*, 293-302.

[15] Lee, Y. and Kwon, O. (2010). Information Privacy Concern in Context-Aware Personalized Services: Results of a Delphi Study. *Asia Pacific Journal of Information Systems, 20*(2), 63-86.

[16] Leskovec, J. and Faloutsos, C. (2006). Sampling from large graphs. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631-636.

[17] Lockshin, L. S., Spawton, A. L., and Macintosh, G. (1997). Using product, brand and purchasing involvement for retail segmentation. *Journal of Retailing and Consumer services, 4*(3), 171-183.

[18] McAuley, J., Pandey, R., and Leskovec, J. (2015). Inferring Networks of Substitutable and Complementary Products. *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

[19] Michaelidou, N. and Dibb, S. (2008). Consumer involvement: a new perspective. *The Marketing Review, 8*(1), 83-99.

[20] Provost, F. and Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc., California.

[21] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2000). Analysis of Recommendation Algorithms for e-commerce. *Proceedings of the 2nd ACM conference on Electronic commerce*, 158-167.

[22] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. (2002). Recommender Systems for Large-scale e-commerce: Scalable Neighborhood Formation using Clustering. *Proceedings of the fifth International Conference on Computer and Information Technology*.

[23] Smith, B. and Linden, G. (2017). Two Decades of Recommender Systems at Amazon. com. *IEEE Internet Computing, 21*(3), 12-18.

[24] Smith, H.J., Milberg, S.J., and Burke, S.J. (1996). Information Privacy: Measuring Individuals' Concerns about Organizational Practices. *MIS quarterly*, 167-196.

[25] Zaki, M. J., Parthasarathy, S., Li, W., and Ogihara, M. (1997). Evaluation of Sampling for Data Mining of Association Rules. *Proceedings of the Seventh International Workshop on Research Issues in Data Engineering*, 42-50.

◆ About the Authors ◆

Hyun Sil Moon

Hyun Sil Moon obtained his M.S. and Ph.D. in Management Information Science (MIS), and his B.S. in Business Administration from Kyung Hee University. His current research interests focus on big data analysis, recommender systems, text mining, and social network analysis. He has published numerous papers which have appeared in International Journal of Information Management, Asia Pacific Journal of Information System, Journal of Intelligence and Information Systems, Journal of Information Technology Services, and Journal of Information Technology Applications and Management.

Jung Hyun Yoon

Jung Hyun Yoon is obtained her M.S. in the Management Information System (MIS) at Kyung Hee University. She obtained her B.S. in Cultural Art from Global Cyber University. Her current research interests include data mining, recommender system, and big data analysis. She has published a paper which has appeared in Korea Society of IT Services.

Il Young Choi

Il Young Choi(choice102@khu.ac.kr) obtained his MS and PhD at School of Management, Kyunghee University and his BS in Economics from Kyung Hee University. His current research interests focus on Recommender Systems, green business/IT, and business intelligence. He has published numerous papers which have appeared in International Journal of Information Management, Information Technology and Management, International Journal of Internet and Enterprise Management, Journal of the Korean Society for Management, Korean Management Science Review, Journal of Intelligence and Information Systems, and Information Systems Review.

Jae Kyeong Kim

Jae Kyeong Kim(jaek@khu.ac.kr) is a professor at School of Management, Kyunghee University. He obtained his MS and PhD in Management Information Systems (MIS) from KAIST (Korea Advanced Institute of Science and Technology), and his BS in Industrial Engineering from Seoul National University. His current research interests focus on business intelligence, network management, and green business/IT. He has published numerous papers which have appeared in Artificial Intelligence Review, Electronic Commerce Research and Applications, European Journal of Operational Research, Expert Systems with Applications, Group Decision and Negotiations, IEEE transactions on services computing, International Journal of Human-Computer Studies, International Journal of Information Management, Technological Forecasting and Social Change.

Submitted: May 6, 2017; 1st Revision: June 27, 2017; Accepted: June 28, 2017