Asia Pacific Journal of Information Systems
Vol. 20, No. 1, March 2010

A New Item Recommendation Procedure Using Preference Boundary*

Hyea Kyeong Kim**, Moon Kyoung Jang***, Jae Kyeong Kim****, Yoon Ho Cho*****

Lately, in consumers' markets the number of new items is rapidly increasing at an overwhelming rate while consumers have limited access to information about those new products in making a sensible, well-informed purchase. Therefore, item providers and customers need a system which recommends right items to right customers. Also, whenever new items are released, for instance, the recommender system specializing in new items can help item providers locate and identify potential customers. Currently, new items are being added to an existing system without being specially noted to consumers, making it difficult for consumers to identify and evaluate new products introduced in the markets. Most of previous approaches for recommender systems have to rely on the usage history of customers. For new items, this content-based (CB) approach is simply not available for the system to recommend those new items to potential consumers. Although collaborative filtering (CF) approach is not directly applicable to solve the new item problem, it would be a good idea to use the basic principle of CF which identifies similar customers, i.e. neighbors, and recommend items to those customers who have liked the similar items in the past.

This research aims to suggest a hybrid recommendation procedure based on the preference boundary of target customer. We suggest the hybrid recommendation procedure using the preference boundary in the feature space for recommending new items only. The basic principle is that if a new item belongs within the preference boundary of a target customer, then it is evaluated to be preferred by the customer. Customers' preferences and characteristics of items including new items are represented in a feature space,

---

* This work was supported by a grant from the Kyung Hee University Post-Doctoral fellowship in 2009 (KHU-20090418).

** *The first author, Research Professor at School of Management and Management Research Institute, Kyunghee* University

*** School of Management and Management Research Institute, Kyunghee University

**** Corresponding author, Professor at School of Management and Management Research Institute, Kyunghee University

***** Professor at School of Business Administration, Kookmin University

and the scope or boundary of the target customer's preference is extended to those of neighbors'. The new item recommendation procedure consists of three steps. The first step is analyzing the profile of items, which are represented as k-dimensional feature values. The second step is to determine the representative point of the target customer's preference boundary, the centroid, based on a personal information set. To determine the centroid of preference boundary of a target customer, three algorithms are developed in this research: one is using the centroid of a target customer only (TC), the other is using centroid of a (dummy) big target customer that is composed of a target customer and his/her neighbors (BC), and another is using centroids of a target customer and his/her neighbors (NC). The third step is to determine the range of the preference boundary, the radius. The suggested algorithm is using the average distance (AD) between the centroid and all purchased items. We test whether the CF-based approach to determine the centroid of the preference boundary improves the recommendation quality or not. For this purpose, we develop two hybrid algorithms, BC and NC, which use neighbors when deciding centroid of the preference boundary. To test the validity of hybrid algorithms, BC and NC, we developed CB-algorithm, TC, which uses target customers only. We measured effectiveness scores of suggested algorithms and compared them through a series of experiments with a set of real mobile image transaction data. We spilt the period between 1st June 2004 and 31st July and the period between 1st August and 31st August 2004 as a training set and a test set, respectively. The training set is used to make the preference boundary, and the test set is used to evaluate the performance of the suggested hybrid recommendation procedure. The main aim of this research is to compare the hybrid recommendation algorithm with the CB algorithm. To evaluate the performance of each algorithm, we compare the purchased new item list in test period with the recommended item list which is recommended by suggested algorithms. So we employ the evaluation metric to hit the ratio for evaluating our algorithms. The hit ratio is defined as the ratio of the hit set size to the recommended set size. The hit set size means the number of success of recommendations in our experiment, and the test set size means the number of purchased items during the test period. Experimental test result shows the hit ratio of BC and NC is bigger than that of TC. This means using neighbors is more effective to recommend new items. That is hybrid algorithm using CF is more effective when recommending to consumers new items than the algorithm using only CB. The reason of the smaller hit ratio of BC than that of NC is that BC is defined as a dummy or virtual customer who purchased all items of target customers' and neighbors'.

That is centroid of BC often shifts from that of TC, so it tends to reflect skewed characters of target customer. So the recommendation algorithm using NC shows the best hit ratio, because NC has sufficient information about target customers and their neighbors without damaging the information about the target customers.

Keywords : Preference Boundary, Ramp-up Problem, Recommendation Procedure, New item Recommendation

I. Introduction

With the rapid growth of e-commerce nowadays, customers on the Web are often overwhelmed with choices and flooded with promotional information about new products. A promising technology to overcome such an information overload has been available through recommender systems that help reduce the information overload by filtering out information which may be otherwise inapplicable to an individual or a group of individuals. Customers can choose various items but, on the other side, it is not easy to find the items which they want to purchase among various items. Therefore, item providers and customers need recommender systems with which they can make right choices.

A recommender system uses information filtering or personalization techniques by applying data analysis techniques to help customers find the products they want to purchase by producing a predicted likeness score or a list of recommended products for a given customer [Sarwar *et al.*, 2001]. Two of the most prevalent approaches are content-based (CB) and collaborative filtering (CF) approaches.

In the CB approach recommendations are given to customers by automatically comparing representations of content of an item which has been rated by the customers of an item to be recommended. CB models the content (features) of an item and recommends it by querying its database on features against the preferences of the customer [Krulwich and Burkey, 1996]. The typical characteristic of CB approach lies in its independence from the information provided by other customers. In the CB recommender system, each customer is supposed to be treated independently, and the system requires information about the customers' needs and preferences in order to be able to recommend an item list. The customer profile contains preference information about content of items. Using these items as basic information, the system searches similar items, which are suggested as recommendation list. Since CB recommender system involves parsing (for descriptive features), it works best with text-based documents, and has found limited success in domains such as sound, images, and video [Mirza, 2001]. Thus, CB systems can recommend a customer only items that are similar to those the customer has already rated or purchased. That is, the scope or the boundary of recommendation can be so narrow that the customer will be presented with potentially interesting, but substantially new items. And in order to function effectively, CB systems require customers to have already rated or purchased a sufficient number of items. As a result, they cannot provide proper recommendations for new customers or new items.

CF is known to be the most successful recommender system and has been widely used in a number of different applications. CF is an information filtering technique based on customers' evaluations of items, i.e. explicit ratings or previous purchasing records, i.e. implicit ratings. It identifies customers whose preferences are similar to those of a given customer, and it recommends contents or products they liked in the past. An important advantage of CF is that it does not need explicit measures of attributes of the items being recommended and explicit descriptions of characteristics of customers. It is also easier to provide good recommendations even when customers' interests are not explicitly given or hidden. CF recommendation algorithm has been known to the most successful recommendation method that has been used in many different applications. Nevertheless, CF-based recommender systems have fundamental weaknesses: scarcity problem i.e. short of item rating; gray sheep problem i.e. customers whose preference is unusually different than others; and a new item ramp-up problem i.e. a new item that has not had enough ratings cannot be easily recommended. Especially the new item 'ramp-up' problem [Avery and Zeckhauser, 1997; Jian *et al.*, 2004] is one of the most serious problems of CF algorithm.

However, the recommender systems for new items have not been much addressed yet [Burke, 2002]. As new items are frequently released nowadays, item providers and customers need the recommender system which is specialized in recommending new items. For example, in the mobile web environment, new images are frequently supplied and their purchasing ratio is considerably high; therefore, image recommender systems need to recommend new items effectively and efficiently.

The new item is added to a system currently, so it cannot be found [Jian *et al.*, 2004] or the new item is having few customers' rating [Burke, 2002]. Because there are no existing records such as rating from customers without connections, it is difficult for customer to get information aboutthe new items. Most of previous approaches for recommender system have to rely on the usage history of customers. So they have suggested CB recommender system using feature values of new items. However it is not sufficient to recommend new items. Although CF is not directly applicable to solve the new item problem, it would be a good idea to use the basic principle of CF which identifies similar customers, i.e. neighbors, and recommends items they liked in the past. Therefore, this research aims to develop a hybrid recommendation procedure for new items. We suggest the hybrid recommendation procedure using the preference boundary in feature space [Kim *et al.*, 2006; Jang *et al.*, 2008] for recommending new items only.

Preference boundary is the scope of customer's actual preference which is defined by centroid and range in the k-dimensional feature space. The basic principle of the suggested procedure is that if new items belong within the preference boundary, then it can be preferred by the target customer. We use the average distance to determine the range among many algorithms, and suggest TC (Target Customer), BC (Big Target Customer), and NC (Target Customer with Neighbors) algorithms to determine centroid. TC is an algorithm developed from contents-based approach, while BC and NC are based on the concept of neighbors, similar customers to a target customer, which are generated from collaborative filtering approach. As this research aims to discover the hybrid procedure (including an algorithm BC and an algorithm NC) that shows better performance than the CB-based procedure (including an algorithm TC), we compare experimental results of suggested procedures.

The rest of this research is organized as follows. Chapter 2 reviews research backgrounds of this research. Chapter 3 illustrates research framework and explains suggested hybrid procedure. Several experimental results are given in chapter 4. Finally, summaries and future works are shown in chapter 5.

Ⅱ. Research Background

2.1 Recommender Systems for New Items

Many researchers have studied hybrid approaches to overcome these limitations of CF and CB approaches, but few of the existing proposed approaches can be a suitable solution for the new item recommendations. Most of previous approaches for recommender system have to rely on the usage history of customers to focus on the current request of customers, so inevitably they are not suitable to recommend new items which have no purchased history. There are several recommender systems which identify merits and demerits for new items.

Kim *et al.* [2006] have supposed the MOBI CORS-music (MOBIle COntents Recommender system for Music) system for recommending new music in the mobile web environment. MOBICORS-music is a hybrid system. Basically it follows the procedure of the CF system, but it uses the data representation of CB system by using the content-based features of music. Based on this data representation, they argue that MOBICORS-music can solve the new item ramp-up problem; further, its performance is better than that of other CF systems. But their methods and experiences are performed for all items, not specified for new items only. Celma *et al.*[2005] have proposed the system that uses the Friend of a Friend (FOAF) and RDF Site Summary (RSS) vocabularies for recommending music to a user, depending on the user's musical preference and listening habits. They expect that the filtering information, which is gathered by the suggested FOAF and RSS, can improve recommender systems. This system, however, needs an additional effort to get individual preference of users to select enormous information which is gathered by the suggested FOAF and RSS. Also, Cornelis *et al.*[2007] have proposed a hybrid recommendation algorithm which involves the fuzzy logic techniques. While it is intuitive to combine the CB and CF into a single formalism, it is hard to set an appropriate value for some parameter to balance the impact of CB and CF contributions and the final recommendation. In sparse rating data, the performance of this algorithm describes a great downward curve. Thus, it applies only in the context of trade exhibition recommendation for e-government since it is not evaluated on real transaction data. Jian *et al.* [2004] have proposed the recommendation algorithms for new items based on indexing techniques. This algorithm presents a different view of a semantic knowledge about the recommendation process based on information retrieval techniques. They cluster all customer transactions by utilizing an index structure X-Features-Tree and using minimum bounding spheres of all leaf nodes in this branch. Before the algorithm performs, it requires specific matching score of the customer transaction and the new item. Melville *et al.* [2002] have proposed the content-boosted collaborative filtering. This method uses a pseudo user-ratings vector for recommending items. The pseudo user-ratings vector consists of the item ratings provided by the user and those predicted by content-based predictor, otherwise. Expectedly, this method suggests poor recommendations if the user rated a few items and the user had a few co-rated items. Thus, this method is performed for all items, not specified for the new item only. Schein *et al.* [2002] have proposed an aspect model which combines collaborative and content information in model fitting for recommending new items. They have suggested several heuristic recommendation methods and tested these methods on implicit rating and rating imputation tasks while evaluating performance under two different methods of recommending embodied by GROC and CROC curve metrics. These GROC and CROC curves measure performance in a specific real world, but as the evaluation goal is to characterize the performance of the recommender system, both GROC and CROC curves are needed since they are often conflicting measures of the recommender system performance.

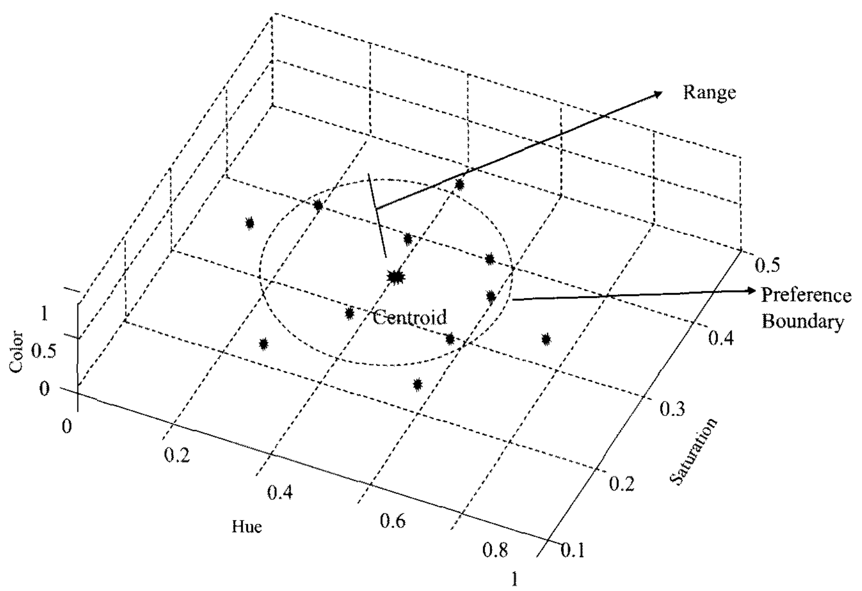
2.2 Representation of Preference Boundary

The accuracy of recommending new items depends, to a large extent, on the ability to represent the customer's actual preference. However, knowing the exact preference of the target customer is not easy. Thus, this research proposes a customer-specific *preference boundary* in multi-dimensional feature space based on his/her past purchase (or rating) history. If a new item is within the target customer's preference boundary, the target customer is assumed to prefer the item because the inner space of the target customer's preference boundary represent his/her preference of items.

Purchased items by a customer include information about the customer's preference on those items. The *personal information set (PIS)*, $P^c = \{p_1, p_2, \cdots, p_L\}$ of customer C consists of items that customer C has purchased. Each item is represented as a vector of features in the k-dimensional feature space that describe its properties such as price, color, and brand. In this research, the scope of a customer's actual preference is represented as *preference boundary*, which is defined by *centroid* and *range* in the k-dimensional feature space. The *centroid vector* O_c is the mean vector of all item vectors in c's personal information set P^c:

$$O_c = \frac{\sum_{i=1}^{L} p_i}{L}.$$

Furthermore, the range is assumed to be constant, so the shape of the preference boundary is round, such as a circle in 2-dimension or sphere in 3-dimension. For an illustrative example, see <Figure 1>, which shows the preference boundary composed of range and centroid vector of a personal information set consisting of 11 images over a 3-dimensional feature space. Each image is represented as a collection of all possible visual features that describe its perceptual properties such as HSV (i.e. hue, saturation, and value of color) based color moment, shape and texture. Six images within the preference boundary are preferred by the target customer, and the other images outside of boundary are not preferred.



<Figure 1> Representation of Preference Boundary

2.3 Neighbor Formation

In this research, each item is represented as a vector of features in the k-dimensional feature space, so we formulated the neighbor set using Centroid Euclidean distance function as a similarity measure [Kim and Chung, 2003; Jiawei *et al.*, 2001; Everitt *et al.*, 2001]. The distance function $d(c, a)$ between the target customer c and other customer a, is calculated as follows;

$$d(c,a) = \sqrt{\frac{\sum_{k=1}^{K}(O_c^k - O_a^k)^2}{K}}$$

where O_c^k and O_a^k are kth feature value of centroid vector O_c and O_a, respectively, and k is the total number of features. The similarity between a target customer c and another customer a, $sim(c, a)$ is calculated using the Weighted Centroid Euclidean distance function;:

$$sim(c,a) = \frac{\underset{b \in H}{Max}[\ d(c,b)\ ] - d(c,a)}{\underset{b \in H}{Max}[\ d(c,b)\ ] - \underset{b \in H}{Min}[\ d(c,b)\ ]}$$

where b implies any customer in neighbor set H, and $d(c, a)$ is a distance function between the target customer c and other customer a, and $Max[d(c, b)]$ and $Min[d(c, b)]$ denote the maximum and minimum distance between two customers c and b, respectively.

Ⅲ. Methodology

In this section, we first set up the research question that we aim to examine in this research. To answer our research question, we carried out the experiments with the intent of answering the following question:

How does the hybrid approach to determine the preference boundary affect the overall performance of a recommender system for new items?

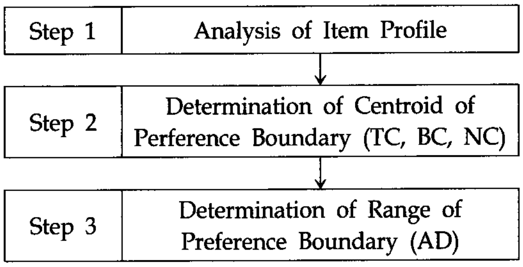
We tested whether the CF approach to determine the centroid of the preference boundary improves the recommendation quality or not. For this purpose, we developed two hybrid algorithms, BC and NC, which use neighbors when deciding the centroid of the preference boundary. To test the validity of hybrid algorithms, BC and NC, we developed CB algorithm, TC, which uses target customers only.

3.1 Overall Procedure

The new item recommendation procedure consists of the following four steps shown in <Figure 2>. The first step is analyzing the profile of items, which are represented as k-dimensional feature values. The feature values of purchased items are used to find the preference boundary of target customer, and those of new items are used to decide whether to recommend the new item or not. In this research, the new items are defined as those which have been just released and needed to be promoted.

The second step is to determine the representative point of the target customer's preference boundary, the centroid, based on a personal information set. To determine the centroid of the preference boundary of the target customer, three algorithms are developed in this research; one is using the centroid of a target customer only (TC), the other is using the centroid of a (dummy) big target customer that is composed of the target customer and his/her neighbors (BC), and another is using the centroids of a target customer and his/her neighbors, respectively (NC). TC is an algorithm developed from a contents-based approach, while BC and NC are based on the concept of neighbors, similar customers to a target customer, which are generated from the collaborative filtering approach.

The third step is to determine the range(radius) of the preference boundary. The suggested algorithm is using an average distance (AD) between the centroid and all purchased items.



<Figure 2> Overall Procedure

3.2 Preference Boundary by Target Customer (TC)

The algorithm TC for finding the centroid of the target customer is used to organize the preference boundary of the target customer based on his/her purchase history only. The preference boundary of the target customer C is determined by the centroid O and the range d of each feature using C's purchase history, $\{O_c^k - \delta_c^k, O_c^k + \delta_c^k\}$, for any k in the feature space. So a new item I is recommended to the target customer C when $O_c^k - \delta_c^k \leq I_c^k \leq O_c^k + \delta_c^k$, for all k. A proper value of the range d is not known exactly so it is determined by the experiments. The preference boundary of the target customer in 3-dimensional feature spaces is shown in <Figure 3>. For example, see the target customer of <Figure 3>. This customer purchased 11 items. In order to locate the centroid of this customer, we calculate the mean value of each figure. In other words, we sum values of each item of each figure, and then divide the number of items, i.e. 11.The algorithm TC for finding the centroid of target customer. This is formally described as follows:

- **TC: Centroid of Target Customer**

  Input:

  T_date: Launching date of items

  Output:

  A centroid of the target customer, O_c

  Algorithm:

  For each target customer C do,

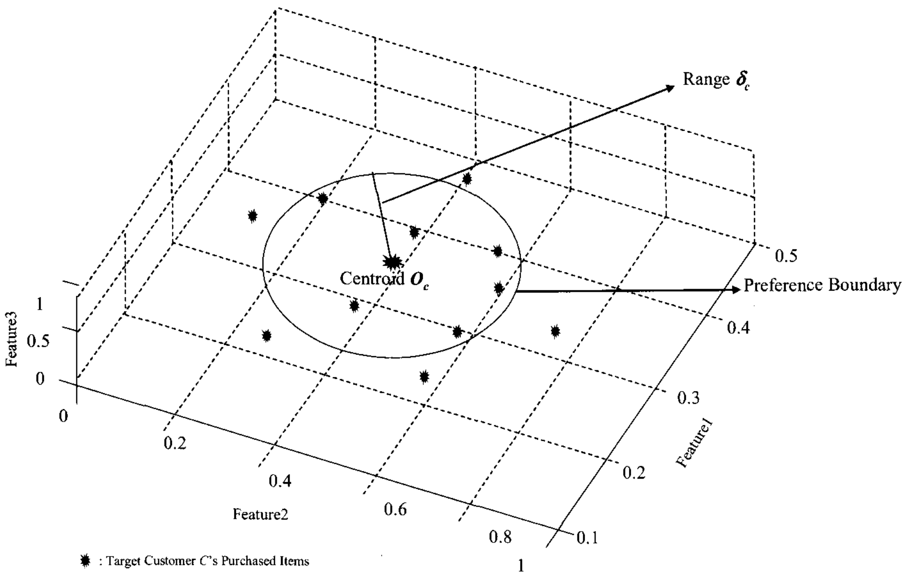
  Find the centroid for each feature based on items which are purchased in the training period by the target customer;

  Endfor;

  Return O_c;

3.3 Preference Boundary by Big Target Customer (BC)

In this research, the 'Big target customer' is defined as a dummy or a virtual customer who purchased all items of the target customer's and its neighbors.' So the preference boundary of the big target customer C is determined by the centroid *O* and the range *d* of each feature using the target customer C's purchase history all together with neighbors' purchase history. TC is an algorithm developed from a contents-based approach, while BC and NC are based on the concept of neighbors, similar customers to the target customer, which are generated from collaborative filtering approach. The preference boundary using the big target customer is shown in <Figure 4>, and the algorithm BC for finding the centroid of the big target customer is formally described as follows:



<Figure 3> Preference Boundary by Target Customer

- **BC: Centroid of Big Target Customer**

  Input:

  T_date: Launching date of items

  N_CNT: Neighbor Size

  Output:

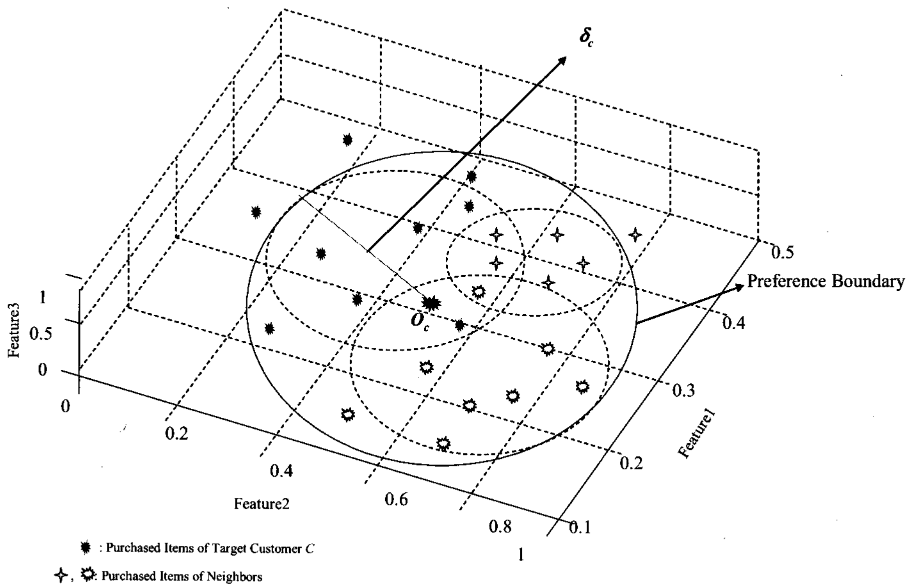
  A centroid of the target customer, O_c

  Algorithm:

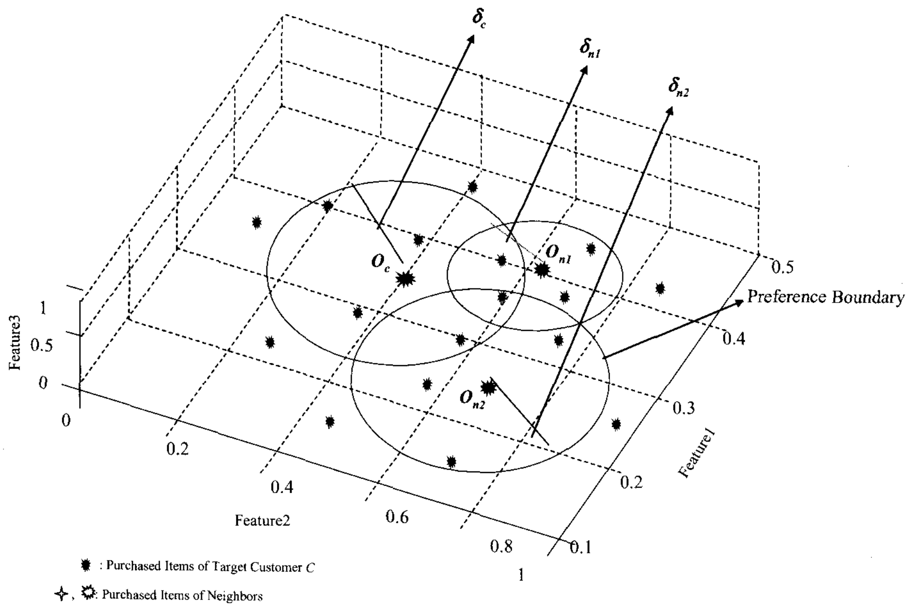
  For each target customer C do,
      Find the centroid for each feature based on items which are purchased in the training period by the target customer and others;
      Find the *N_CNT* neighbor using use Euclidean distance function;
  Repeat;
      Find the centroid based on all purchased data of the target customer and one's neighbor during the training period;
      Until neighbor size <= *N_CNT*;
  Endfor;
      Return O_c;

3.4 Preference Boundary by Target Customer with Neighbors (NC)

As items are represented as points in k-dimensional feature space, neighbors are found by calculating the distance between the target customer and other customers. The personal information set of customers are represented as a cluster in the feature space, so a cluster distance function is used to calculate the distance between centroids of target customer and those of other customers [Han and Kamber, 2006]. The customers are assumed to have similar preferences with that of the target customer if the distance is very close. The idea is to use the similar customers or neighbors derived from CF to determine the representative of the target customer. The Euclidean distance function is used as the cluster distance function in this



<Figure 4> Preference Boundary by Big Target Customer



<그림 5> Preference Boundary by Target Customer with Neighbors

research, because this function is simple, easy to calculate, and generally accepted [Ishikawa *et al.*, 1998].

The algorithm NC is to make the preference boundary of the target customer based on the target customer's and his/her neighbors' purchase history. In other words, the preference boundary of the target customer C is determined by the centroid O and the range d of each feature using C's purchase history and neighbors n's purchase history, separately. So a new item I is recommended to the target customer C, when $(O_c^k - \delta_c^k) \le I_c^k \le (O_c^k + \delta_c^k)$ or $(O_n^k - \delta_n^k) \le I_n^k \le (O_n^k + \delta_n^k)$ for all k and all neighbor n. The preference boundary of the target customer and his/her neighbors is shown in <Figure 5>, and the algorithm NC for finding the centroid of the target customer with neighbors is formally described as follows:

- **NC: Centroid of Target Customer With Neighbors**

Input:

T_date: Launching date of items

N_CNT: Neighbor Size

Output:

Centroids of the target customer O_c

Algorithm:

For each target customer C do,
  Find centroid for each feature based on items which are purchased in the training period by the target customer and others;
  Find the *N_CNT* neighbor using use Euclidean distance function;
  Repeat;
  Fine the centroids for each neighbor;
  Until neighbor size <= *N_CNT*;
Endfor;
Return O_c;

3.5 Determination of Preference Boundary Range

After determining the centroid of the preference boundary, we need to define the range of preference boundary. This research uses an average distance (AD) algorithm, which defines the range of the preference boundary as an average distance between the centroid of the target customer and each purchased item in the personal information set. A new item I is recommended when $O_k^c - d_k^{avg} \le I_k^c \le O_k^c + d_k^{avg}$, for all k. The algorithm AD for determining the range of the preference boundary using the average distance is formally described as follows:

- **AD: Range by Average Distance**

Input:

T_date: Launching date of items

N_CNT: Neighbor Size

Output:

The *proposed new item list* for target customer

Algorithm:

proposed_new_list = {};

For each target customer, C do,

Repeat while new items from I_1 to I_{ni};

Find the average distance (d_{avg}) between the centroid O_c of target customer and transaction data;

Calculate the distance (*CI_DIST*) between the new item I and the centroid of target customer;

Generate proposed recommendation list when *CI_DIST* <= d_{avg};

Until new items from I_1 to I_{ni};

Endfor;

Return *proposed_new_list*;

IV. Experimental Evalution

4.1 Data Set

For our experiment, we use real transaction data and image data in mobile commerce offering character images. The data is provided from *S* content provider, a leading Korean company. The data contains 8,776 images, 1,921 customers, and their 55,321 transactions during the period between 1_{st} June, 2004 and 31_{st} August, 2004.

To characterize images, a data preprocessing task, such as extraction of visual features, is performed. We suggest a recommendation procedure for new items, which has flexibility to use any visual features to characterize images. However, in this research, the HSV (i.e. hue, saturation, and value of color) based color moment is selected over other choices of features such as shape or texture, because the color moment is the most generally used feature and HSV represents human color perception more uniformly than others do[Porkaew *et al.*, 1999]. We obtain the bitmap format files in the 256 colors of those images. For all pixels in images, we translate the values of 3-color channels (i.e. RGB: red, green, and blue) into HSV values. Then, the mean, standard deviation and skewness for HSV values are calculated to represent images as vectors in a 9-dimensional feature space.

We split the period between 1_{st} June and 31_{st} July and the period between 1_{st} August and 31_{st} August as a training set and a test set, respectively. The training set is used to make the preference boundary, and the test set is used to evaluate the performance of the suggested hybrid recommendation procedure.

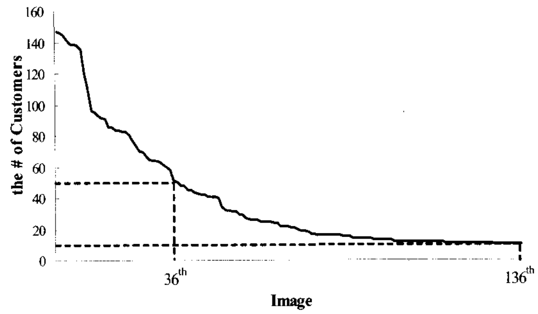
As the target customers, we select 219 customers who have purchased more than 10 images in the training period. In order to make the effective preference boundary, we conduct several heuristic conductings, then the number of items is limited to 10. Generally, the smaller size the purchased items are, the more difficulty we face in making the preference boundary. On the other hand, the larger the size of purchased items, the smaller the size of target customers we get. Finally, the training set consists of 35,436 transaction records, and the test set consists of 19,848 transaction records, each created by the target customers. New images are defined as images released after 1_{st} August and purchased more than 10 times by target customers during the test period. The number of corresponding new images is 136. <Figure 6> shows the overall description of experimental data.

2004.6.1 7.1 8.1 8.31

	Training set	Test set
Total Customers	1,921	
Target Customers	219	
	Purchased more than 10 items	
Transactions	55,284	
	35,436	19,848
Images	8,776	
New Image		136

<Figure 6> Data Set Design

<Figure 7> shows the distribution of customers who purchased new images in the test period. The graph means the number of customers who bought that image. For example, the 36^{th} new items are purchased by the 50 customers, and all new items (i.e. 136^{th} new items) are purchased by the 10 customers in the test period. We use all data except new items purchased by fewer than 10 customers in test period.



<Figure 7> Distribution of New Items

4.2 Experimental Environment

4.2.1 Evaluation Metrics

Recommender system research has various measures for evaluating the effectiveness and efficiency of recommender systems [Sarwar *et al.*, 2001]. The main aim of this research is to compare the hybrid recommendation algorithm with the CB approach. To evaluate the performance of each approach, we compare the purchased new item list during the test period with the recommended item list which is recommended by the suggested approach. So we employ the evaluation metric, *hit ratio,* for evaluating our procedures. The *hit ratio* is defined as the ratio of hit set size to the recommended set size. The hit set size means the number of success of recommendations in our experiment, and the test set size means the number of purchased items in test period. *Hit ratio* can be written as:

$$Hit\ Ratio = \frac{size\ of\ hit\ set}{size\ of\ test\ set}.$$

4.2.2 System for Experiments

A system to perform our experiments is implemented using Visual Basic 6.0, and ADO components. The system consists of two parts: one for image data preprocessing and the other for experiment execution and result analysis. MS-SQL Server 2000 is used to store and process all the data necessary for our experiments. We run our experiments on window XP-based PC with Intel Core 2 Quad CPU having a speed 2.40GHz and 3.24GB RAM.

<Table 1> Detail Result Descriptions of Each Customer

Customer Id	# of Purchased New Items	# of Hit Items in Preference Boundary			Hit Ratio		
		TC	BC	NC	TC	BC	NC
622	78	3	2	8	0.038	0.026	0.103
1595	70	1	1	1	0.014	0.014	0.014
1331	51	0	0	0	0.000	0.000	0.000
1460	50	7	14	15	0.140	0.280	0.300
827	49	7	9	9	0.143	0.184	0.184
1232	49	1	1	1	0.020	0.020	0.020
1012	48	5	5	7	0.104	0.104	0.146
1256	45	1	2	2	0.022	0.044	0.044
1485	43	8	8	10	0.186	0.186	0.233
204	43	3	3	3	0.070	0.070	0.070
787	42	2	2	3	0.048	0.048	0.071
1194	41	4	5	4	0.098	0.122	0.098
869	40	3	4	6	0.075	0.100	0.150
784	39	0	0	1	0.000	0.000	0.026
1244	38	2	2	2	0.053	0.053	0.053
757	11	1	1	4	0.091	0.091	0.364
1473	11	2	2	2	0.182	0.182	0.182
843	11	2	3	3	0.182	0.273	0.273
1689	11	2	2	2	0.182	0.182	0.182
1777	11	1	2	1	0.091	0.182	0.091
372	10	3	1	4	0.300	0.100	0.400
770	10	1	1	3	0.100	0.100	0.300
317	10	2	2	2	0.200	0.200	0.200
1200	7	0	0	0	0.000	0.000	0.000
1418	6	3	3	3	0.500	0.500	0.500
Average	19.918	4.945	5.032	5.443	0.291	0.295	0.315

4.3 Experimental Result

In this section, we compare the result of the BC and NC with that of TC. <Table 1> shows the detailed result descriptions of 219 target customers and an average value of target customers. In order to help better understanding of results generated by suggested algorithms, we present the result of a sampled 25 target customers in <Table 1>. For an example, see one of the target customers whose id is 869. This customer purchased 40 new items during the test period. The number of items within him/her preference boundary using TC is 42, that using BC is 52, and that using NC is 64. The number of hit items using TC is 3, that using BC is 4, and that using NC is 6. As the hit ratio defined as the ratio of hit set size to the test set size, that using TC is 0.075, i.e. 3 divided by 40, that using BC is 0.1, i.e.4 divided by 40, and that using NC is 0.15, i.e. 6 divided by 40.

So we calculate the average of each result, which is the number of purchased items, the number of hit items, and the hit ratio. The average number of purchased new items of target customers is 19.92 during the test period. NC shows the largest number of items within the preference boundary, i.e. NC recommends the largest number of items among suggested algorithms, and thus NC shows the largest number of the hit set. Accordingly, the best hit ratio is generated by NC. In contrast, TC shows the smallest number of items within the preference boundary. TC recommends the smallest number of items among suggested algorithms; as a result, TC shows the smallest number of the hit set. Accordingly, the worst hit ratio is generated by TC.

The number of purchased new items means the actual purchased new items in the test period. And the number of hit items in the preference boundary means the number of actual purchased new items among suggested items by recommendation procedures. As dividing the number of hit items in preference by the number of purchased new items is the hit ratio, the fact that the hit ratios of BC and NC are bigger than the hit ratio of TC means using neighbors is more effective to recommend new items. That is, the hybrid approach using CF is more effective to recommend new items than the approach using only CB.

The reason for the smaller hit ratio of BC than that of NC is that BC is defined as a dummy or virtual customer who purchased all items of the target customer's and his/her neighbors.' This means that the centroid of BC often shifts from that of TC, so it tends to reflect skewed characters of the target customer. So the recommendation procedure using NC shows the best hit ratio, because NC has sufficient information of the target customer and his/her neighbors without information damage of the target customer.

V. Conclusion

This research proposes the preference boundary-based procedure for recommending new items to customers who are highly likely to buy them. The key concept of the suggested procedure is the preference boundary which presents customer's preference in multi-dimensional feature space.

This research makes the following contributions to the recommender system research area. Traditional CF approach, known as one of the most effective recommendation approaches, has fundamental weakness in recommending new items. This is referred to as a new item ramp-up problem. Although many researchers have used the CB recommendation approach to solve the new item ramp-up problem, we suggest the hybrid recommendation procedure combining CB approach and CF approach to recommend new items (i.e. BC and NC). Experimental test results showing that the hit ratios of BC and NC are bigger than that of TC means using neighbors is more effective to recommend new items. This result also means the suggested hybrid procedure is more effective than the procedure using only CB. The analysis of performance results of these algorithms generates some interesting facts which could be further analyzed.

Experiments of the suggested procedure give promising results to recommend new items. But the suggested procedure is evaluated on particular mobile image transaction data. Accordingly, future researches are needed to comparatively evaluate the procedures by expanding further to other domain of data (e.g. offline department transaction data). Thus, it is advisable to increase the size of the target customer set to improve the reliability of the suggested hybrid procedure. It is also important to conduct experiments with a different number of a target customer set and a new item set to effectively compare the robustness of suggested procedures and independency of changed experimental data. Also in this research, we use the Euclidean distance to calculate the similarity of neighbors without reflecting on the weight or similarity. Therefore, in the future studies, it is important to adopt other algorithms to define neighbors who purchase a more number of same items than the target customer does.

⟨References⟩

[1] Avery, C. and Zeckhause, R., "Recommendation systems for evaluating computer messages," *Communications of the ACM*, Vol. 40, No. 3, 1997, pp. 88-89.

[2] Burke, R., "Hybrid Recommendation Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, Vol. 12, No. 4, 2002, pp. 331-370.

[3] Celma, O., Ramirez, M., and Herrera, P., Foafing the music: A music recommendation system based on RSS feeds and user preferences, London, UK, 6th International Conference on Music Information Retrieval, 2005, pp. 464 - 457.

[4] Cornelis, C., Lu, J., Guo, X., and Zhang, G., "One-and-only item recommendation with fuzzy logic techniques," *Information Sciences*, Vol. 177, No. 22, 2007, pp. 4906-4921.

[5] Everitt, B.S., Landau, S., and Leese, M., Cluster Analysis, London, UK, Oxford University Press: A Hodder Arnold Publication, 2001.

[6] Ishikawa, Y., Subramanya, R., and Faloutsos, C., MindReader: Querying databases through multiple examples, the 24th International Conference on Very Large Data Bases, 1998, pp. 218-227.

[7] Jang, M.K., Kim, H.K., and Kim, J.K., "An Extended Content-based Procedure to Solve a New Item Problem," *Korea Intelligent Information Systems Society*, Vol. 14, No. 4, 2008, pp. 201-216.

[8] Jian, C. and Jian, Y, Jin, H., Recommendation of New Items Based on Indexing Techniques, Shanghai, China, the 3rd International Conference on Machine Learning and Cybernetics, Vol. 2, 2004, pp. 1168-1172.

[9] Han, J. and Kamber, M., Data Mining: Concept and Techniques (2nd ed.), San Francisco, USA, Morgan Kaufmann Publishers, 2006.

[10] Kim, D. and Chung, C., Qcluster: Relevance Feedback Using Adaptive Clustering for Content-Based Image Retrieval, ACM SIGMOD international conference on Management of data, San Diego, USA, 2003, pp. 599-610.

[11] Kim, J.K., Cho, Y.H., Kang, M.Y. and Kim, H.K., "A Hybrid Multimedia Contents Recommendation Procedure for a New Item Problem in M-commerce," *Korea Intelligent Information Systems Society*, Vol. 12, No. 2, 2006, pp. 1-15.

[12] Krulwich, B. and Burkey. C., Learning User Information Interests Through Extraction of Semantically Significant Phrases, the AAAI Spring Symposium on Machine Learning in Information Access, 1996, pp.

110-112.

[13] Melville, P., Mooney, R. and Nagarajan, R., Content-boosted Collaborative Filtering for Improved Recommendations, the 18th National Conference on Artificial Intelligence, Edmonton, Canada, 2002, pp. 187-192.

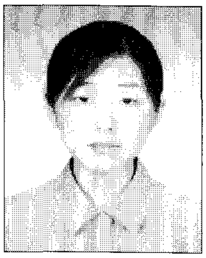
[14] Mirza, B.J., Jumping Connections: A Graph-Theoretic Model for Recommender Systems Master's Thesis, Virginia Polytechnic Institute and State University, 2001.

[15] Porkaew, K., Chakrabarti, K., and Mehrotra, S., Query Refinement for Multimedia Similarity Retrieval in MARS, the 7th ACM Multimedia Conference, Orlando, USA, 1999, pp. 235-238.

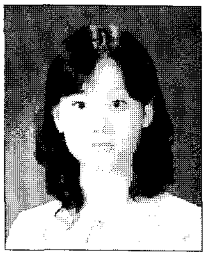
[16] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., Item-based collaborative filtering recommendation algorithms, the 10th international conference on World Wide Web, Hong Kong, China, 2001, pp. 285-295.

[17] Schein, A.I., Popescul, A., Ungar, L.H., and Pennock, D.M., Methods and Metrics for Cold-Start Recommendations, the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Tampere, Finland, 2002, pp. 253-260.

◆ About the Authors ◆

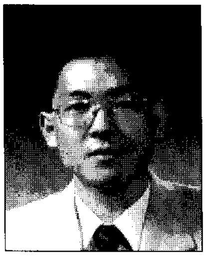


Hyea Kyeong Kim
Hyea Kyeong Kim(kimhk@khu.ac.kr) is a research professor at School of Management, Kyung Hee University. She obtained her MS and PhD in MIS, and her BS in Physics from Kyung Hee university. Her current research interests focus on Recommender Systems and Complex Systems. She has published numerous papers which have appeared in IEEE Transactions on Services Computing, International Journal of Information Management, Expert Systems: The International Journal of Knowledge Engineering and Neural Networks, Expert Systems With Applications, Lecture Notes in Computer Science, and Lecture Notes in Artificial Intelligence, among others.



Moon Kyoung Jang
Moon Kyoung Jang(jmoonk25@gmail.com) earned her BS and MS in MIS from Kyunghee university. She has worked a number of years for Lotte Data Communication Company, Seoul, Korea. Her current research interests focus on CRM, recommender systems, data mining, and e-business.

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. He is a vice president of KMIS (the Korea Society of Management Information Systems), a BK21 project manager of HRM based on network science, and a member of Green IT department, Public forum under the low carbon green growth. 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. He is also a chief editor of JIIS (Journal of Intelligence and Information Systems), and an AE (associate editor) of Information Technology and Management (SSCI).



Yoon Ho Cho

Yoon Ho Cho(www4u@kookmin.ac.kr) is an associate professor of School of Business Administration at Kookmin University in Korea. He received a BS in Computer Science and Statistics from Seoul National University and a MS and a PhD degree in MIS from KAIST. His research interests include recommender systems, mobile business, customer relationship management, and data mining. He has published numerous papers which have appeared in IEEE Intelligent Systems, Information Sciences, Electronic Commerce Research and Applications, Expert Systems with Applications, Computers and Industrial Engineering, and International Journal of Internet and Enterprise Management, among others.