Web-based Product Recommendation System with Probability Similarity Measure

Sang Hyun Choi
Department of ISE, Engineering Research Institute, Gyeongsang National Univ.
(chois@gnu.ac.kr)

Byeong Seok Ahn
College of Business Administration, Chung-Ang Univ.
(bsahn@cau.ac.kr)

This research suggests a recommendation system that enables bidirectional communications between the user and system using a utility range-based product recommendation algorithm in order to provide more dynamic and personalized recommendations. The main idea of the proposed algorithm is to find the utility ranges of products based on user specified preference information and calculate the similarity by using overlapping probability of two range values. Based on the probability, we determine what products are similar to each other among the products in the product list of collaborative companies. We have also developed a Web-based application system to recommend similar products to the customer. Using the system, we carry out the experiments for the performance evaluation of the procedure. The experimental study shows that the utility range-based approach is a viable solution to the similar product recommendation problems from the viewpoint of both accuracy and satisfaction rate.

Keywords : Personalized Recommendation; Similarity Measure; Collaborative Ccommerce; Incomplete Information

Received: November 2006 Accepted: March 2007 Corresponding Author: Sang Hyun Choi

1. Introduction

Firms can maintain the close relationship with the customers and raise customer loyalty by offering differentiated service based on personalization function (Allen et al., 1998). Personalization at the content providing site can be used as a strategy to help users save time and cost of searching the contents, and as a strategy to maximize revenue by increasing the level of user satisfaction about the site. Personalization is now even used in the area of product recommendation, and a number of researches have been conducted on personalized product recommendation system. The prior researches on the personalized product recommendation system, however, lack in the academic rigor on the following issues. First, few recommendation systems explicitly incorporate each user's multi-faceted preferences on the product although they suggest products to the people who have similar preferences, using a number of similarity measures for product recommendation. For example, the price is an important factor in purchasing a computer for some users whereas others consider the brand as important factor in their preferences. Or both of them might be considered important in purchasing decision. Second, current recommendation methods are implemented in a one-sided way, mainly from a system to the user. For example, general purchase process at a store occurs through the communication between the customer and sales clerk. Through this effective communication, a transaction takes a place when a customer acquires information he or she wants and gives the requirements about the product to the sales clerk. But, most systems until now lacked re-handling process based on real time feedback and thus they merely suggested recommended product in a one-sided way.

Therefore, in our research, we developed a recommendation system which enables bidirectional communications between the user and system using a utility range-based product recommendation algorithm in order to provide more dynamic and personalized recommendations. Consequently, we suggest a business process for implementing the business model of electronic commerce that is based on collaboration among the firms and we conduct an empirical experiment to test the system and relevant algorithm.

2. Prior Research on Similarity Measure for Recommendations

Recommendation system is defined as the system which recommends an appropriate product or service after learning the customers' tendency and desire. Today, a number of researchers are conducting a rigorous study on recommendation system, and the most important factor in recommendation system is an ability to correctly filter and analyze the customer preference and behavior, thereby recommending the best product which the customer wants based on accurate estimation approach (Cho et al., 2002; Sarwar et al., 2000; Sarwar et al., 2001). This leads to the research on the personalized recommendation method, and two of the well-known representative methods are content based filtering and collaborative filtering (Mooney and Roy, 2000; Goldberg et al., 1992).

Montaner et al. (2003) suggest advanced classification of Internet-based intelligent recommendation system, and they distinguish the recommendation methods into content-based filtering, collaborative filtering, and hybrid of which previous two methods are combined. And, cosine similarity, naive bayesian (NB) classifier, Pearson r correlation, etc. are being used as key similarity measure for recommendation. But, the key similarity measures, which are used in the recommendation system, such as cosine similarity, Pearson r correlation, NB classifier, Euclidean distance, etc. have the following weaknesses. In case of cosine similarity measure, it is possible to apply the measure when the feature values of products for clustering are separated out from the origin and the clusters are away from (Schalkoff, 1992). In case of Pearson r correlation, the problem of scalability occurs where processing speed slows down when there is a high number of users and the item types due to the scarcity problem especially in case of data with a number of item types (Sarwar et al., 2000; Sarwar et al., 2001; Billsus and Pazzani, 1998). In case of NB classifier, it can be applied when the attributes values are categorical data (Billsus and Pazzani, 1998). And, it does not consider the importance rates of the attributes. With respect to Euclidean distance each data should be matched by point in the space of real values but it is hard to match the data by point using the categorical data (Shyu et al., 2003; Yang and Pedersen, 1997; Schafer et al., 2001). Further, it is hard to apply to the non-numerical value data and it results in the decrease of processing speed since more calculation procedures are needed as the amount of data increases.

In this research, we utilize a utility range-based product recommendation algorithm where customer preferences regarding the product specification values can be expressed in numerical values (Choi and Cho, 2004). It is further extended to deal with problems in which different weights have to be attached to each of product specifications. Such weights of the product specifications imply the user preferences. For example, when purchasing a computer which is characterized by multiple specs (or criteria) including CPU type, brand, price, etc., some customers consider the CPU type as the most important factor influencing their purchasing decision but others might consider brand as most important factor. As such, customers in general have different preference information regarding the product specs and it is difficult to measure the weights of product spec as precise numerical values. Here, the basic concept of the algorithm is that we provide the customer with more efficient ways to enter the preference information by which customers can express their judgments on the weights of product specs in terms of mutual relations among the attributes. Refer to the study by Choi et al. (1999) for more detailed descriptions on incomplete information.

3. Similar Product Recommendation Problems and Process

3.1 Problem Definition for Similar Product Recommendation

A utility range-based product recommendation algorithm suggests a way to classify the products in the class of SKUs (Stock Keeping Units), which can be defined by the same set of product specifications, according to the user preference. Using this algorithm, they can be distinguished into the class with the same set of specifications. To be specific, a value of a product with respect to a spec is converted to a utility value by applying one of appropriate normalization methods. The utility values with respect to each of specs are then combined with the weights of specs to obtain a single value representing the whole worth of a product. This notion can be formally described as follows.

The products that belong to the same class can be evaluated by a set of M product specifications at the level of SKU for I products currently owned by a firm. The underlying model is a multiattribute additive value model in order to evaluate the product, $x_i = \{x_{i1}, x_{i2}, \cdots, x_{iM}\}$ in which x_{ij} denotes the original value of i-th product on j-th specification.

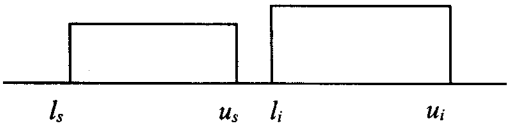
$$v(x_i) = \sum_{j=1}^{M} w_j v_{ij}(x_{ij}) \tag{1}$$

Here, v_{ij} is the marginal value function of specification j such that $v_{ij} : x_{ij} \rightarrow [0, 1]$ and w_j is the importance weight of specification j. After all, the objectives of a utility range-based product recommendation algorithm are to calculate the utility values of the products, which the collaborative company owns and belongs to the same class as the product selected by the customer, and to suggest similar products to the customer which are as close in utility value as the product, x_s, which he or she selected. If both w_j and $v_{ij}(\cdot)$ are given in precise numerical values by a decision maker, then utility values, x_s and x_i for $i \neq s$ can be easily calculated using the Formula 1. But, much effort has to be exerted to obtain precise values on the weights in the real-world problems. In other words, frequent questions and responses between the system and user are required to obtain more precise information on the weights. Such scenario seems to be inappropriate especially in the Web-based system in which the user is usually reluctant to provide more detailed information the system requires to produce results. Instead, it is presumably correct to assume that a decision maker can provide the required information on the weights more easily in the format of incomplete information such as, e.g., ordinal relations among the weights and ranges of weight (Choi and Cho, 2004; Choi et al., 1999). The final solutions (i.e., the worth of each product) subject to such incomplete information inevitably are obtained in the form of not a single value but interval. Therefore, we shall present a probabilistic method for determining similarity between two intervals.

3.2 Similarity Measure

This section suggests the similarity measure that shows the level of similarity between the two products x_i and x_s using the utility value based on the individual's personal preference. As discussed in the previous section, this research tries to calculate the utility ranges of product since it is difficult to calculate the accurate utility value of a product. Let us assumes the ranges for product x_i and product x_s as $[l_i, u_i]$ and $[l_s, u_s]$ respectively. The interval numbers can be ordered by some kinds of probability distributions (e.g., uniform, normal distribution and the like) since the numbers within the interval sometimes does not have the same meaning for the decision maker as is implied by the use of interval ranges. To compare interval numbers by probability distributions, we will define a function d, $d: A \times A \rightarrow [0, 1]$, representing a measure of strength of preference, where A is a finite set of products for comparisons. Let us define $d(i, s) = P(i \geq s)$ where $P(\cdot)$ states that probability of an interval number (i.e., random variable) i is greater than or equal to an interval number s. Further, let $f_i(x)$ and $f_s(y)$ be probability density functions of interval numbers i and s which are defined on $[l_i, u_i]$ and $[l_s, u_s]$ respectively. When one is to compare two interval numbers, one has to consider one of the following three possible cases, depending upon the positions of end points of interval numbers. For simplicity, supposing that interval numbers i and s are uniformly distributed, the heights in the figures below are arbitrarily drawn and thus have no real meaning at all.

Case 1: $u_s \leq l_i$ (see [Figure 1])



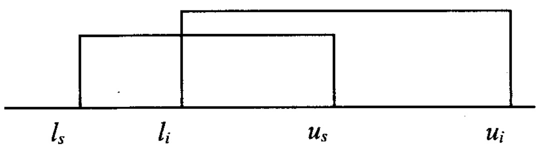
[Figure 1] The case of $u_s \leq l_i$

When two interval numbers do not overlap to each other except a boundary $u_s = l_i$ as it is seen in [Figure 1], an interval number i is certainly positioned prior to an interval number s since a measure of strength of preference becomes $d(i, s) = 1$ in which

$$d(i, s) = \int_{l_i}^{u_i} f_i(x)\, dx = \int_{l_s}^{u_s} f_s(y)\, dy = 1. \quad (2\text{-}1)$$

Obviously, $d(s, i)$, a measure of strength of preference that an interval number s is preferred to an interval number i, becomes zero.

Case 2: $l_s < l_i < u_s \leq u_i$ (see [Figure 2])



[Figure 2] The case of $l_s < l_i < u_s \leq u_i$

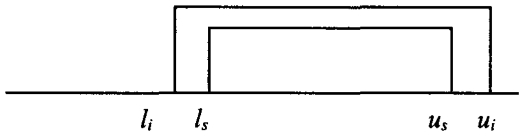
When two interval numbers intersect to each other, a measure of strength of preference, $d(i, s)$ can be derived by the use of the following formula:

$$\begin{aligned} d(i, s) = &\int_{l_s}^{l_i} f_s(y)\, dy + \int_{l_i}^{u_s} f_s(y)\, dy \cdot \\ &\int_{u_s}^{u_i} f_i(x) dx + \int_{l_i}^{u_s} f_s(y) \int_{y}^{u_s} f_i(x)\, dx dy \end{aligned} \quad (2\text{-}2)$$

Similarly, $d(s, i)$, a measure of strength of preference that an interval number s is preferred to an interval number i, is calculated such as:

$$\begin{aligned} d(s, i) &= \int_{l_i}^{u_s} f_s(y) \int_{l_i}^{y} f_i(x)\, dx dy \\ &= \int_{l_i}^{u_s} f_i(x) \int_{x}^{y} f_s(y)\, dy dx \end{aligned} \quad (2\text{-}3)$$

Case 3 : $l_i \le l_s < u_s \le u_i$ (see [Figure 3])



[Figure 3] The case of $l_i \le l_s < u_s \le u_i$

When one interval number is entirely included in the other interval number, a measure of strength of preference, $d(i, s)$ can be derived by the use of the following formula:

$$d(i,s) = \int_{u_s}^{u_i} f_i(x)\,dx + \int_{l_s}^{u_s} f_s(y) \int_{y}^{u_s} f_i(x)\,dxdy \quad (2\text{-}4)$$

Similarly, $d(s, i)$, a measure of strength of preference that an interval number s is preferred to i, is calculated as:

$$d(s,i) = \int_{l_i}^{l_s} f_i(x)\,dx + \int_{l_s}^{u_s} f_s(y) \int_{l_s}^{y} f_i(x)\,dxdy. \quad (2\text{-}5)$$

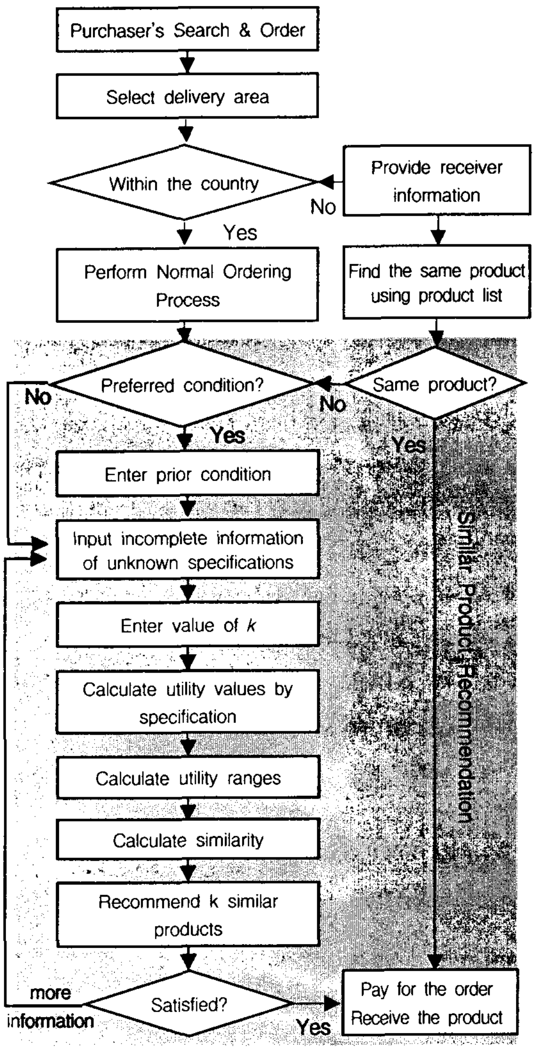
The values of $d(i, s)$ and $d(s, i)$ falls in the range between 0 and 1, and the closer values to ½, the better similarity. The level of similarity between the two products can be calculated as the following.

$$\begin{aligned} sim_{is} &= 2 \cdot d(i, s), \text{ for } 0 \le d(i, s) \le 1/2 \\ &= 2 \cdot (1 - d(i, s)), \text{ for } 1/2 < d(i, s) \le 1 \quad (3) \end{aligned}$$

Here, a value, sim_{is} falls in the range between 0 and 1 and the higher value means the better similarity.

3.3 Similar Product Recommendation Procedure

This procedure is applied when a customer searches for a product at the collaborative site, similar to the product he or she selected at the purchasing location. The process is based on the business model where product class and spec information are shared between the shopping malls of two different countries that are under collaborative relationship. Similar product searching process is described in [Figure 4].



[Figure 4] Similar Product Recommendation Procedure

First, a purchasing customer enters order information of the product he or she wants to place at the domestic shopping mall. Then, when the customer selects the delivery location, order information is processed through the normal ordering process if the product recipient is the purchaser or other domestic resident. But, if the recipient is a resident of foreign country, a search for the same product having the same product specification values is conducted using the product list of the site at the delivery location, which is in collaborative relationship with the site where the product search is handled. At the time, if the same product exists, the customer purchases the product at the shopping mall and makes a specific delivery request. But, if it does not, the process of similar product recommendation is carried out.

In the beginning, a customer can select prior search process of entering search conditions of specification values for products he or she wants to buy. The process excludes the products that do not meet the conditions. Then the process of similar product recommendation requires a customer to enter the relations among the weights of product spec based on the five types of incomplete information and to assign the number of similar products to be chosen in the list. Examples of incomplete information are in the form of bounded descriptions: ordinal ranking ($w_i \geq w_j$), strict preference ($w_i - w_j \geq \varepsilon$), preference with a multiple ($w_i \geq \alpha_{ij} w_j$), interval preference ($l_i \leq w_i \leq u_i$) and preference difference ($w_i - w_j \leq w_l - w_m$). Up to this point is the stage of receiving input information from the customers, and we apply a utility range-based product recommendation algorithm based on this information. A utility range- based product recommendation algorithm consists of following four steps.

Step 1 : *Calculation of utility value by specification* - Record the specification values of the products which are selected by the customer and calculate the utility value with respect to each specification based on the recorded specification values. Let x_{ij} be the jth specification value of ith product and $\max_i x_{ij}$ or $\min_i x_{ij}$ the maximum or minimum value of ith specification values. The utility value v_{ij} of ith product on jth specification can be calculated by formula (4), in which it is linearly normalized in the range of [0, 1]. Refer to other normalization methods in Hwang and Yoon (1980). For specification j with better for larger,

$$v_{ij} = (x_{ij} - \min_i x_{ij})/(\max_i x_{ij} - \min_i x_{ij})$$

and for specification j with better for smaller

$$v_{ij} = (x_{ij} - \max_i x_{ij})/(\min_i x_{ij} - \max_i x_{ij}) \quad (4)$$

Step 2 : *Calculation of range of utility value by product* - Calculate the range of utility values for the products in the same class as the customer selected product at the product list in the receiving location. In this case, it is not realistic to assume a customer to enter precise values about weight w_j for each of specifications. Instead, it seems presumably correct to assume that a decision maker can provide the required information on the weights more easily in the format of incomplete information such as, e.g., ordinal relations among the weights and ranges of weight. The final solutions (i.e., the worth of each product) subject to such incomplete information inevitably are obtained in the form of not a single value but interval. Let us define a set of relations among the weights as Φ_W. Using the information obtained, utility value range, $[v_i(\min), v_i(\max)]$ for ith product is calculated using the following mathematical programs.

$$v_i(\min) = \min \ \Sigma_j w_j v_{ij} \text{ subject to } \Phi_W$$
$$v_i(\max) = \max \ \Sigma_j w_j v_{ij} \text{ subject to } \Phi_W \quad (5)$$

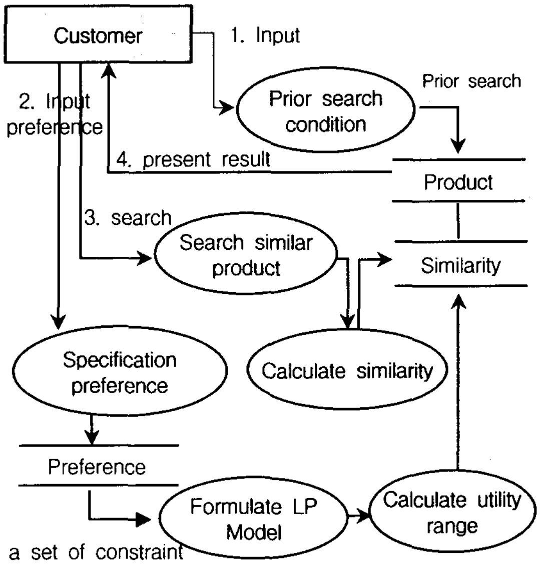
Step 3 : *Calculation of similarity values between products considered* - Calculate how similar the products are by the similarity measure in formula (3). This process is conducted by comparing the utility value ranges of each product from recipient location with the utility of a product the customer selected.

Step 4 : *Suggestion of k most similar products in the order of similarity values* - Define the product as a highly similar product when sim_{is} value for each product at the receiving location is high, and suggest k most similar products in the ascending order of sim_{is} values.

So far, we have examined a utility range-based product recommendation algorithm based on the additive utility theory and similar product searching process. We shall introduce the details of system implementation in the next section.

4. System Implementation

The interactive procedure of recommending the similar product is specified by the following data flow as shown in [Figure 5]. The customer enters prior search condition to look for his/her wanting products using specification values.

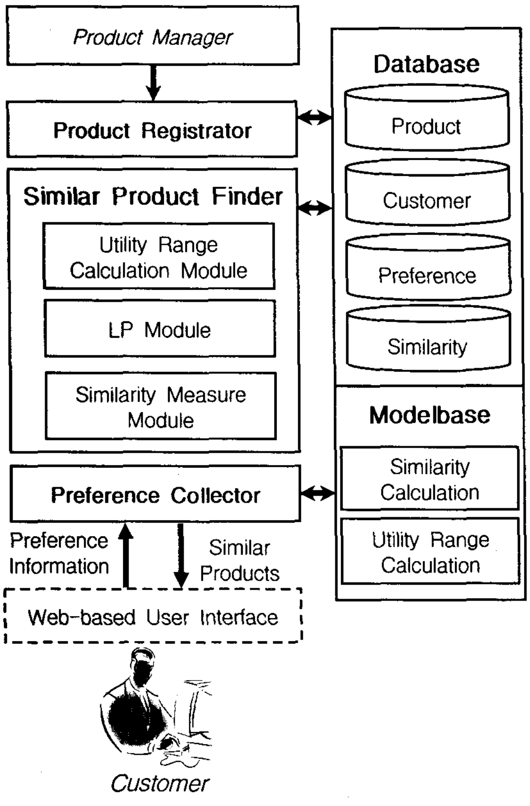


[Figure 5] Data flow of the system

As shown in the system architecture [Figure 6], the system largely consists of Product Registrator, Customer Preference Collector, and Similar Product Finder, and further they are made up of several sub-modules. This system is based on web-based user interface, and the processes in product manager and customer sides are integrated with each relevant module.

4.1 Product Registrator

Product Registrator consists of Product Class Registration Module and SKU-based Product Registration Module. Product manager handles the tasks relating to product registration and he or she defines the new product class and decides the specification of each product class. Product manager must first define the product class of a given product before registering the new product unit. If there is predefined product class, the new product unit should be registered under the appropriate class. But if there isn't a pre-defined product class, a new product class should be created according to the following process.



[Figure 6] System Architecture

First, define the ID and name of the product class and define the spec of the product class. Similarly, refer to "quantitative specs" as specs that can be measured such as price, etc. and "qualitative specs" as specs that are of non-numerical values of specification such as manufacturer name. Next, assign B.F.L.(Better For Larger; for example, computer memory) and B.F.S.(Better For Smaller; for example, price) for each spec. Lastly, the process of assigning the category value for calculating the utility value of qualitative spec is examined. [Figure 7] shows a notebook computer class defined by six specifications.

Defined Product Class Spec

No.	Spec Name	Category Value	Determine B.F.L. or B.F.S.	Delete
1	Price	Define	B. F. S.	Delete
2	Brand	Defined	B. F. L.	Delete
3	RAM	Defined	B. F. L.	Delete
4	HDD	Defined	B. F. L.	Delete
5	ODD	Defined	B. F. L.	Delete
6	Display	Defined	B. F. L.	Delete

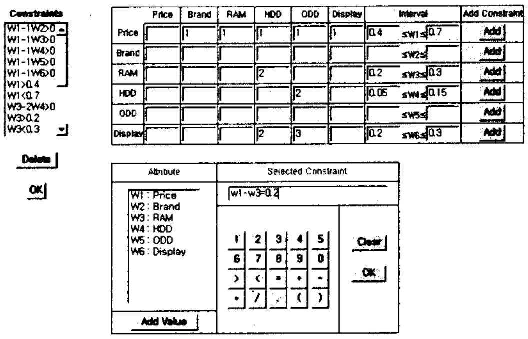
※ B.F.L. : Better For Larger
B.F.S. : Better For Smaller

[Figure 7] Defined Product Class Spec

After product class and detailed spec are defined, new unit products begin to be regularly registered. Then, product registration is divided into the registration of products with known specification values and those with partial or entirely unknown specification values. First, registering the product, of which its specification value is already known, involves entering the numerical value when the specification value under consideration is quantitative specification value. But, if it is qualitative specification value then it selects the corresponding category value. Lastly, specification value can be expressed by setting the relationship information among other products as 5 types of incomplete information when accurate specification values of a product are not known. The relationship is used as the constraint when calculating the maximum and minimum values of specification value through LP Module. The relative utility value of each product, which has accurate specification value, is computed using Formula (3) when the registration of all the unit products is completed.

4.2 Customer Preference Collector

In order to recommend a similar product, preference information about the degree of importance for each specification should be entered by the customer. But, as mentioned in previous section, preference information is usually incomplete and is stated inaccurately. Therefore, we developed and utilized the tool in this system in order to effectively express the incomplete information of specification weights.



[Figure 8] Screen of Incomplete Information Entry

The incomplete information is encoded in the input screen of the software through the use of a user-friendly interface as shown in [Figure 8]

[Similar Product Finding-Result]

You selected the following product at the L mall:

Product ID	Model Name	Price	Brand	RAM	HDD	ODD	Display
54	ThinkPad G40-2388 LK2	1585000	LG-IBM(2)	DDR512(5)	60GB 이상(5)	COMBO(5)	14inch이상~16inch 미만(3)

We have found 5 similar products at the B mall respectively by Utility Range-based Approach.

•Utility Range-based Approach

Product ID	Model Name	Price	Brand	RAM	HDD	ODD	Display
13	PCG-GRX770	1990000	Sony(5)	DDR512(5)	30GB 이상(5)~60GB 미만(3)	COMBO(5)	16inch 이상(3)
4	EVO N1000C 470051-341	1409000	Compaq(3)	DDR256(4)	30GB 이상(5)~60GB 미만(3)	COMBO(5)	14inch이상~16inch 미만(3)
6	EVO N1020V 470059-789	1630000	Compaq(3)	DDR512(5)	30GB 이상(5)~60GB 미만(3)	COMBO(5)	14inch이상~16inch 미만(3)
23	드림북 N7600.4	1435000	TG(2)	DDR256(4)	30GB 이상(5)~60GB 미만(3)	COMBO(5)	14inch이상~16inch 미만(3)
24	드림북 N7620.4	1435000	TG(2)	DDR256(4)	30GB 이상(5)~60GB 미만(3)	COMBO(5)	14inch이상~16inch 미만(3)

[Figure 9] Result of Similar Product Finding

The capture screen displays the portion of a participant's imprecise value judgment about relations between the specifications. For example, the screen shows that price spec is more important than each spec of brand, RAM, HDD, ODD, and display type. Then, the relations between the specifications are transformed into the constraints in formula (5) and the value range of each computer is obtained by solving the program.

4.3 Similar Product Finder

This module helps the system go through the process of maximization and minimization through LP Module in order to calculate the range of utility value which is used for measuring product similarity. This system, enabled by LP Module, supports the automated formulization of linear programming problem. The solution for the formulized linear programming problem is searched through Lindo API which is an internal application. And, calculated maximum value and minimum value are stored in the database by each product. According to the similar product searching process as shown in [Figure 4], we specify the number of similar products with the product which the customers want and similar products are suggested through internal processing of each module. [Figure 9] presents the screen showing the result of similar products found.

5. Experimental Study

We conducted an empirical experiment to measure the level of user satisfaction and accuracy of a utility range-based product recommendation algorithm.

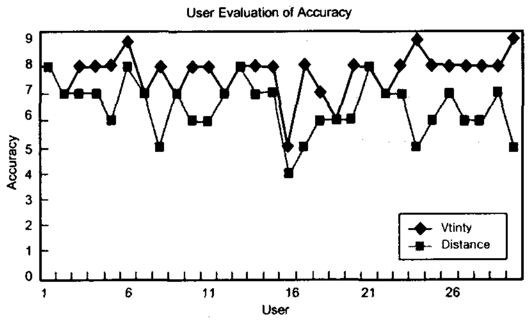
5.1 Experiment Setting

The experiment was conducted with 30 students who did not have prior knowledge about the algorithms used in the system and who had different preferences. And, we compared the result of the utility range-based algorithm with that of Euclidean distance measure-based algorithm. We selected the Euclidean distance method for comparison because it could deal with the multiple attribute values and be commonly used at various similarity research areas. For performance evaluation of the recommendation method, we used the accuracy and the satisfaction rate of the result of recommendation. We firstly showed the list of computers, ask the students to keep the most desirable computers in mind, and enter the information about specification values describing the reason that they want them. After the system found the products using the specification values and the weights of specifications, each user was asked to directly select the choice using the 9-point Likert scale (1 : Extremely inaccurate, 5 : Average, 9 : Extremely accurate) to indicate the degree of identicalness between the user selected product and recommended similar product. And, using the 9-point Likert scale (1: Extremely dissatisfied, 5 : Average, 9 : Extremely satisfied), the user was also asked to indicate the level of satisfaction in the input and output process of using the system while executing the similar product recommendation process. The product class used in the experiment is notebook computer and the relevant product spec considered are limited to price, brand, memory, hard disk capacity, display size, and CD-ROM drive type. For the ease of conducting the experiment, registering the product and defining the category values regarding the product spec are previously implemented by system developer.

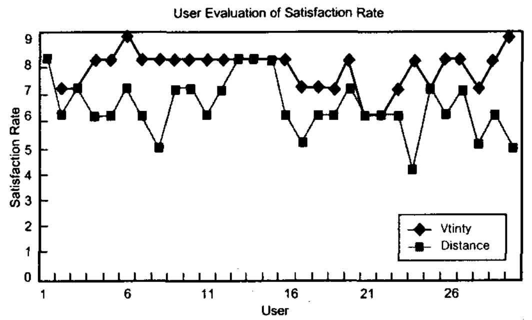
In the first part, problems and relevant issues were explained to the experiment participants, and the actual experiment was conducted in the order of the similar product recommendation as shown in [Figure 4]. Here, we identified the number of similar products which are to be recommended as five. First, an experiment participant enters the personal preference information for each product spec using an appropriate input entry format. And, experiment indicates the accuracy and satisfaction of the suggested similar products when the system, using the entered personal preference information, suggests 5 similar products by each internal module and a utility range-based product recommendation algorithm. Next, we also proceeded in the comparable manner to indicate the accuracy and satisfaction of the suggested similar products by Euclidean distance measure.

5.2 Results of Experiment

According to the result of accuracy and satisfaction evaluation as shown in [Figure 10] and [Figure 11] relatively, we were able to verify that the recommendation results of a utility range-based product recommendation algorithm were higher in both accuracy and satisfaction than the results by Euclidean distance measure method.



[Figure 10] Evaluation Result in terms of Accuracy



[Figure 11] Evaluation Result in terms of Satisfaction

We conducted t-test of a paired sample in order to statistically verify the effectiveness of the two recommendation methods. As shown in <Table 1> and <Table 2>, both the accuracy and satisfaction results were statistically valid within significance level of 0.05 in terms of average differences between two methods. The results came out in such that the method was designed to incorporate the individual preference in the interviews of experiment participants.

<Table 1> Statistical Verification on Accuracy

Accuracy	Num. of Users	Mean	Std. Dev.	Std. Err.	T-Value	P-Value
Utility Range	30	7.733	0.828	0.151	5.92	0.000
Euclidean Distance	30	6.467	1.008	0.184		
Paired Difference	30	1.267	1.172	0.214		

<Table 2> Statistical Verification on User Satisfaction Rate

Satisfaction Rate	Num. of Users	Mean	Std. Dev.	Std. Err.	T-Value	P-Value
Utility Range	30	7.667	0.711	0.130	6.49	0.000
Euclidean Distance	30	6.333	0.994	0.182		
Paired Difference	30	1.333	1.124	0.205		

6. Conclusion and Future Research Directions

In this research, we developed a system based on a utility range-based product recommendation algorithm and conducted an experimental study to examine the functions of the developed system. The system characteristics can be explained from the perspectives of both the firms that are selling the product and customers who are the users of the system. From the firm's perspective, it is possible to register and search the similar products even for the product where its precise specification value is not known. And, mutually complimentary similar product recommendation is possible by composing the product map which shares among the collaborating firms based on this system. From the customer's perspective, different preferences of each individual can be fully considered by entering them by each spec, and more accurate recommendation can be obtained by entering additional preference information based on the recommended result. Further, the key characteristic is its ability to efficiently enter the personal preference information which is hard to accurately express. Lastly, the overall process of product registration and product recommendation also improved the convenience of both the product manager and customers through the utilization of web-based graphic user interface and automated module.

Further, we achieved more improved accuracy and satisfaction level than previous recommendation method through user experiment. Future research interest can be described as such that system development is needed where specification value searching is limited in order to recommend a product that satisfies an unique specification value, apart from current function of selecting a product and searching its similar product. In order for the system development to be successful, the test should be conducted in an organizational setting through practical application.

References

[1] Allen, C., Kania D. and Yaeckel B., *Internet world guide to one-to-one web marketing*, New York, John Wiley & Son, 1998.

[2] Billsus, D. and Pazzani M. J., "Learning collaborative information filters", *International Conference on Machine Learning*, Vol.15 (1998), 46~54.

[3] Cho, Y. H., Kim J. K. and Kim S. H., "A personalized recommender system based on web usage mining and decision tree induction", *Expert Systems with Applications*, Vol.23, No.3(2002), 329~342.

[4] Choi, S. H. and Cho Y. H., "A utility range-based similar product recommendation algorithm for collaborative companies", *Expert Systems with Applications*, Vol.27, No.4 (2004), 549~557.

[5] Choi, S. H., Kim S. H. and Kim J. K., "An interactive procedure for multiple attribute group decision making with incomplete information", *European Journal of Operational Research*, Vol.118, No.1(1999), 139~152.

[6] Goldberg, D., Nichols D., Oki B. M. and Terry D., "Using collaborative filtering to weave an information tapestry", *Communications of the ACM*, Vol.35, No.12(1992), 6170.

[7] Hwang, C. L. and Yoon, K. S., *Lecture notes in economic and mathematical systems*, Springer-Verlag, New York, 1980.

[8] Mooney, R. and Roy, L., "Content-based book recommending using learning for text categorization", *ACM Conference on Digital Libraries*, Vol.5(2000), 195~204.

[9] Montaner, M., Lopez B. and Rosa J. L. D., "A taxonomy of recommender agents on the internet", *Artificial Intelligence Review*, Vol. 19(2003), 285~330.

[10] Sarwar, B., Karypis G., Konstan J. and Reidl J., "Analysis of recommendation algorithms for e-Commerce", *ACM Conference on Electronic Commerce*, (2000), 158~167.

[11] Sarwar, B., Karypis G., Konstan J. and Reidl J., "Item-based collaborative filtering recommendation algorithms", *International WWW Conference*, Hong-Kong, 2001.

[12] Schafer, J. B., Konstan J. A. and Riedl J., "E-commerce recommendation applications", *Data Mining and Knowledge Discovery*, Vol.5(2001), 114~153.

[13] Schalkoff, R., *Pattern recognition: statistical, structural and neural approaches*, John Wiley and Sons, 1992.

[14] Shyu, M. M., Haruechaiyasak C. and Chen S. C., "Category cluster discovery from distributed WWW directories", *Information Sciences*, Vol.155(2003), 181-197.

[15] Yang, Y. and Pedersen J. P., "A comparative study on feature selection in text categorization", *Proc. of the Fourteenth Int. Conf. on Machine Learning*, (1997), 412~420.

요약

확률 유사성척도를 활용한 웹 기반의 상품추천시스템

최상현* · 안병석**

본 연구에서는 사용자와 시스템간 양방향 의사소통 방법을 가능하게 하는 확률 유사성척도 기반의 추천시스템을 제안한다. 본 시스템에서 활용한 알고리즘의 주요한 아이디어는 사용자가 제시한 상품 사양에 대한 선호정보를 사용하여 상품의 효용 범위를 구하고, 두 상품의 효용범위 값들 간의 유사도의 값으로서 겹침 확률을 계산한다. 앞에서 구해진 상품 간 유사도의 값을 사용하게 되면, 유사 정도가 가장 높은 상품들을 유사상품으로 등록하게 된다. 본 추천시스템은 개별 사용자 별로 제시된 정보를 사용하므로 차별화된 추천이 가능해진다. 본 시스템을 활용하게 되면 상품 정보를 공유하는 기업들이 협업 전자상거래 프로세스를 수행할 수 있다. 협업 기업들은 협업을 위한 상품들을 등록하고 해당 상품들과 유사한 상품들을 각 기업의 유사상품 데이터베이스에 저장하여, 향후 유사상품 추천에 활용할 수 있다. 본 연구에서 제시된 유사상품 추천시스템은 웹 기반의 응용시스템으로서 인터넷이 가능한 어떠한 환경에서도 수행될 수 있다. 본 연구에서 제시된 절차의 효과를 검증하기 위하여 사용자 실험을 수행하였다. 실험결과를 살펴보면 효용기반의 방법론이 정확도와 만족도 측면에서 상품추천 문제에 대한 하나의 효과적인 해결책으로 사용될 수 있다는 것이 검증되었다.

* 경상대학교 산업시스템공학부
** 중앙대학교 경영대학 경영학부