A Personalized Recommendation Procedure for E-Commerce

Jae Kyeong Kim[a], Yoon Ho Cho[b], Woo Ju Kim[c], Je Ran Kim[d], and Ji Hae Suh[e]

[a, e] *School of Business Administration, Kyung Hee University*
#1, Hoeki-Dong, Dongdaemoon, Seoul, 130-701, Korea
Tel :+ 82-2-961-9355, Fax :+ 82-2-967-0788, e-mail : jaek@khu.ac.kr,

[b] *Graduate School of Management, Korea Advanced Institute of Science and Technology*
207-43 Cheongryangri-Dong, Dongdaemun, Seoul, 130-012, Korea
e-mail: yhcho@dongyang.ac.kr

[c,d] *Department of Industrial Engineering, Chonbuk National University*
664-14, Deokjin-Dong, Deokjin-Ku, 1st St., Chonju, Chonbuk, 560-756, Korea
e-mail: wjkim@chonbuk.ac.kr

Abstract

A recommendation system tracks past actions of a group of users to make a recommendation to individual members of the group. The computer-mediated marketing and commerce have grown rapidly nowadays so the concerns about various recommendation procedures are increasing. We introduce a recommendation methodology by which e-commerce sites suggest new products or services to their customers. The suggested methodology is based on web log analysis, product taxonomy, and association rule mining. A product recommendation system is developed based on our suggested methodology and applied to a Korean internet shopping mall. The validity of our recommendation system is discussed with the analysis of a real internet shopping mall case.

Keywords: Recommendation procedure, Personalization, Association Rule Mining, Web log analysis

Introduction

According to the rapid growth of Internet, there is a keen competition among internet shopping-malls and contents providers. So one-to-one marketing, treating an individual customer as a different marketing strategy, and CRM(Customer Relationship Management) differentiating customers and stressing a long-time relationship with customers, have been given much attention from internet business to gain a competitive power. According to the increase of internet business, customers have more opportunity to select products at a time, but it imposes heavy burden to internet business companies. In this reason, if the internet company knows who is the target customer, and what is the customer really wants, the marketing effect will be increased.

Generally the filtering procedure can be divided by two ways, one is content-based filtering, and the other is collaborative filtering. Content-based filtering recommends products by the similarity degree of customer's purchase history. Collaborative filtering recommends the product by neighbor's (a person who has similar purchase behavior like a customer) preference and interest. Another recommendation procedure is an association rule algorithm, using association rules to recommend products.

This research intends to suggest a recommendation procedure based on collaborative filtering, association rule mining, decision tree, and web log mining. A recommendation system is also developed based on our suggested procedure. The suggested procedure uses the information of customer's purchase behavior and product purchase pattern from the web-log information. Furthermore, the system is not for recommending product by simple product relationship but for recommending efficient product by analyzing behavior and property of

customer. This recommendation system divides a role into data mining specialist and marketer, and applied to the real shopping mall.

Research Methodology

Product recommendation is a service that recommends proper products to customers after analyzing their purchase behaviors. Proper product recommendation service contributes to the increase of sales by suggesting profer products to the customers. In the E-Commerce environment, this research defines product recommendation as follows: product recommendation recommends less than N different products to the customer group purchasing more than P, less than P+I products from the different product class. By this, the customer group can purchase product more than 1. P and N are defined as integers more than 1, and I is defined as the integer more than 0.

This recommendation system uses prediction algorithm to select possible customers who can buy recommended products. Recommended product is acquired by association rule mining. Information about each customer's product preference and sales efficiency is used to the product recommendation. Figure 1 shows the structure of a suggested recommendation system.

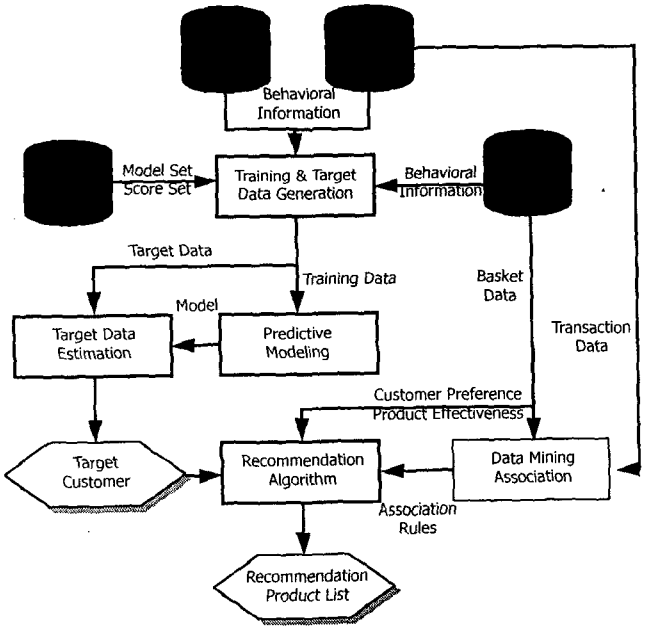

Figure 1 The Structure of Recommendation System

The overall product recommendation process is shown as the following Figure 2.

Step 1. Data Preparation

Data collection is needed to perform product recommendation. Data source consists of customer data, product data, and sales data from the database and web-log files.

Data for product recommendation process can organize data mart through preprocessing. Data mart is constituted as following informations.

(1) Customer attribute information: Customer attribute information is consistent of demographic information, behavior information, and psychological information.

(2) Transaction and basket information: Transaction information extracted from sales data, which is used for association rule mining, and then is transformed into proper form. Basket information extracted from web-log analysis is the information about products in a customer basket.

(3) Other information: This divides into 3 behaviors, 'view of enlarged products', 'products in basket', and 'orders of the products'.

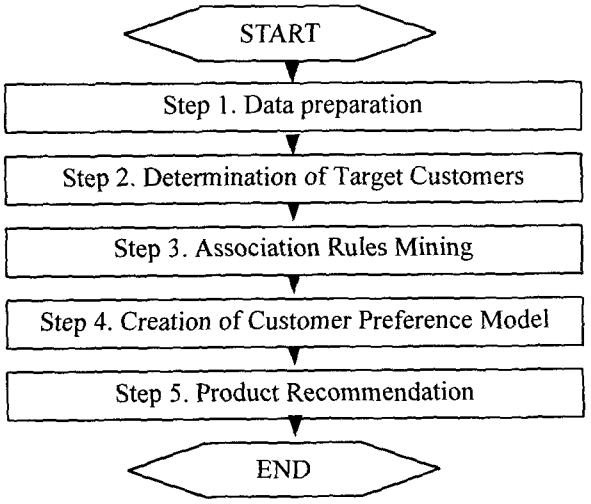

Figure 2 The Overall Flow

Step 2. Determination of Target Customers

Target customer is determined as follows. First, model set and score set is decided. And it is prepared training data and target data made by customer attribute information. Second, the model is trained, and if the model is suitable, then the customer is selected by predicting target data.

As following shows the example of decision model set. Figure 3 shows a product recommendation analysis. Select a product class which is in a product taxonomy level 1, p=1, past period of the model set is between 5 and 8 months, and future period is October.

CID	May	June	July	Aug	Sep	Oct	Model Set
101	.	E	.	B	.	.	T
102	.	.	.	.	.	.	F
103	A	A	.	.	.	E	T
104	.	B, D	F	.	.	C	T

Figure 3 Decision of Model Set

Customer attribute selecting model set makes up training set regardless to purchase. An object variable Y, meaning occurrence of purchase, can be defined as follows.

$Y =$ { 1: Customer who buy a product in different product class more than 1 during the future period
0: Customer who doesn't buy a product during the future period

In training data N, which is organized of customer attribute data, the record of m^{th} Customer is expressed as follows.

$$record_m = [X_1, X_2, X_3, X_4, ..., X_N, Y]$$

Figure 4 shows training data which is composed of customer attribute data, age, sex, job, sum of total purchase, and frequency of purchases. In this, Y is '1' if customer buy a product, otherwise Y is '0'.

CID	Age	Sex	Job	Sum of total Purchase	Frequency of Purchases	Recent Visit Date	...	Y
101	22	M	Student	64	4	0827	...	0
103	36	F	Government Employ	57	6	1018	...	1
104	23	F	Student	128	10	1104	...	1

Figure 4 Training Data

Score set is a customer set that buys a product from the different product class more than P and less than P+I up to now. Customer attribute information in score set makes up a target data. Figure 5 is an example of decision score set, I=4, past period is from July to October, future period is on December.

CID	July	Aug	Sep	Oct	Nov	Dec	Score Set
001	A	A	E	·		?	T
002	·		·	·		?	F
003	·	B	·	B		?	T
...	...	...	...	...		...	...

Figure 5 Decision of Score Set

From the target data composed of customer attributes, a record of mth customer presents as follows.

$$record_m = [X_1, X_2, X_3, X_4, ..., X_N]$$

Step 3. Association Rule Mining

Association rule mining is researched as follows. Product class set that the purchase of customer 'm' is defined as $Purset_m$. Product class set that associated with purchasing product class is defined as $AssoSet_m$. Each product class about product association rule mining $A^{(m)}$ is defined as follows.

$$A^{(m)} = [A_{m1}, A_{m2}, ..., A_{ms}, ..., A_{mS}]^T, m = 1, ..., M$$

$A_{ms} =$ { $Conf(s)$: product class s exist in $AssoSet_m$ but doesn't exist in $PurSet_m$
0 : Other cases

Conf(s) means a confidence of association rule leading from product class s. If the association rule is more than 1, the rule with the highest confidence value is selected. Figure 6 shows the $AssoSet_m$ about each customer m when association rule is discovered to the customer in Figure 5 ,

Product Association Rule
A⇒C(0.8), A⇒D(0.6), B⇒E(0.8), B⇒F(0.3), C⇒A(0.7), C⇒F(0.4)

CID	$PurSet_m$	$AssoSet_m$
001	A, E	C(0.8), D(0.6)
003	B	E(0.8), F(0.3)
012	B, C, E	A(0.7), F(0.4)
017	B, D	E(0.8), F(0.3)

Figure 6 Each Customer's Associated Product Class

Product association rule is a process of finding association rules from the customer product class. But it is hard to find rules in the actual purchase data, so we investigate the basket product class on the web. The process of product association rule is as follows.

(1) Set analyzing period.

(2) Classify products as a transaction classification criteria, and aggregate it for a class unit, and find rules. Through this process, rule set $Ruleset_P$ is created.

(3) Classify products which is sold on the analysis period as a transaction classification criteria, and aggregate it as a class unit, minimum support applying to basket analysis is higher than step (2). Through this process, rule set $Ruleset_B$ is created.

(4) Combine $Ruleset_P$ with $Ruleset_B$ and make final $Ruleset_{all}$. If the same rules exist in $Ruleset_P$ and $Ruleset_B$, the Rule with the highest confidence value will put in $Ruleset_{all}$.

(5) Produce association rules

Figure 7 shows an example of combining $Ruleset_P$ with $Ruleset_B$ and the association rule $Ruleset_{all}$.

Rule Set	Association Rule about Product Class
$Ruleset_P$ (ms=0.05)	A⇒C(0.8), B⇒E(0.4), B⇒F(0.5), C⇒A(0.2)
$Ruleset_B$ (ms=0.1)	A⇒C(0.6), B⇒E(0.8), C⇒A(0.7), C⇒F(0.4), A⇒D(0.6)
$Ruleset_{all}$	A⇒C(0.8), A⇒D(0.6), B⇒E(0.8), B⇒F(0.5), C⇒A(0.7), C⇒F(0.4)

Figure 7 Integration of Association Rules

Step 4. Creation of Customer Preference Model

Customer preference model is constructed by three behavior patterns during buying product at the shopping

mall. These behaviors are about how many times they click each class product, how many times they put the product in their baskets, and how many times they buy the products.

In the product class s, when V_{ms} is the number of total click, and B_{ms} is the number of putting the products in the basket, customer preference model $C^{(m)}$ can be defined as follows.

$$C^{(m)} = [C_{m1},...,C_{ms},...,C_{mS}]^T, m = 1,...,M \ ,$$

$$where \quad C_{ms} = \frac{\hat{V}_{ms} + \hat{B}_{ms}}{2}$$

$$\hat{V}_{ms} = \frac{V_{ms}}{\sum_{s'=1}^{S} V_{ms'}} \ , \qquad \hat{B}_{ms} = \frac{B_{ms}}{\sum_{s'=1}^{S} B_{ms'}}$$

When six product classes from A to E are given, customer preference model is illustrated like Figure 8.

CID	A	B	C	D	E	F
001	0.2	0	0.15	0	0.2	0
003	0	0.2	0.1	0	0.1	0.08
012	0.2	0	0.1	0	0.15	0.22
017	0	0.12	0	0.2	0.1	0.15

Figure 8 Customer Preference Model

Step 5. Product Recommendation

After the previous 4 steps, the last step is continued as follows.

(1) Calculate the matching score.

$$MS(m,s) = \frac{2 \times A_s^{(m)} \times P_s^{(m)}}{A_s^{(m)} + P_s^{(m)}}$$

Figure 9 shows a matching score with product association rules of Figure 6 and customer preference model of Figure 8.

CID	A	B	C	D	E	F
001	0	0	.253	0	0	0
003	0	0	0	0	.178	.126
012	.311	0	0	0	0	.284
017	0	0	0	0	.178	.2

Figure 9 Match Score Table

(2) Decide top-N product classes with the highest matching score.

(3) Calculate the purchase efficiency about each class product. Efficiency value of product P is the ratio of the number of click p product to the number of purchase p product. This can be called as an *Click-to-Buy rate.* Figure 10 shows the product efficiency.

(4) To top-N product class, n product which has the highest purchase efficiency value in product class is recommend to the customer m.

Figure 11 shows recommendation products to each customer, which is obtained from the matching score value of Figure 6, and applying the Figure 10 product purchase efficiency.

Class	A	B	C	D	E	F
Products	A_1(.01) A_2(.05) A_3(.02) A_4(.03)	B_1(.02) B_2(.02) B_3(.04)	C_1(.03) C_2(.02) C_3(.00) C_4(.03) C_5(.04)	D_1(.01) D_2(.01) D_3(.00) D_4(.02)	E_1(.02) E_2(.01) E_3(.01)	F1(.00) F2(.015) F3(.01) F4(.01)

Figure 10 Purchase Efficiency of Product

CID	Purchase Product Class	*Top*-2 Product Class	Recommendation Product
001	A, E	C, D	C_5
003	B	E, F	E_1
012	B, C, E	A, F	A_2
017	B, D	E, F	F_2

Figure 11 Customer's Recommendation Product

Realization of Recommendation System

The recommendation system is developed based on our suggested procedure. Figure 12 shows the overall structure of the system. User interface of recommendation system is realized with HTML and JavaServlet. The centeral process of mining algorithm is realized with Java. Cache 4.0 is used for the database systems. This product recommendation system runs on Window 2000 server and JRUN environment. The product recommendation system is divided into two modes. One is the product recommendation of expert mode. And the other is product recommendation of marketer mode. Expert mode is a process about creating model and discovering product association rule mining. Marketer mode selects recommendation customers and proper products by using association rule mining obtained at expert mode.

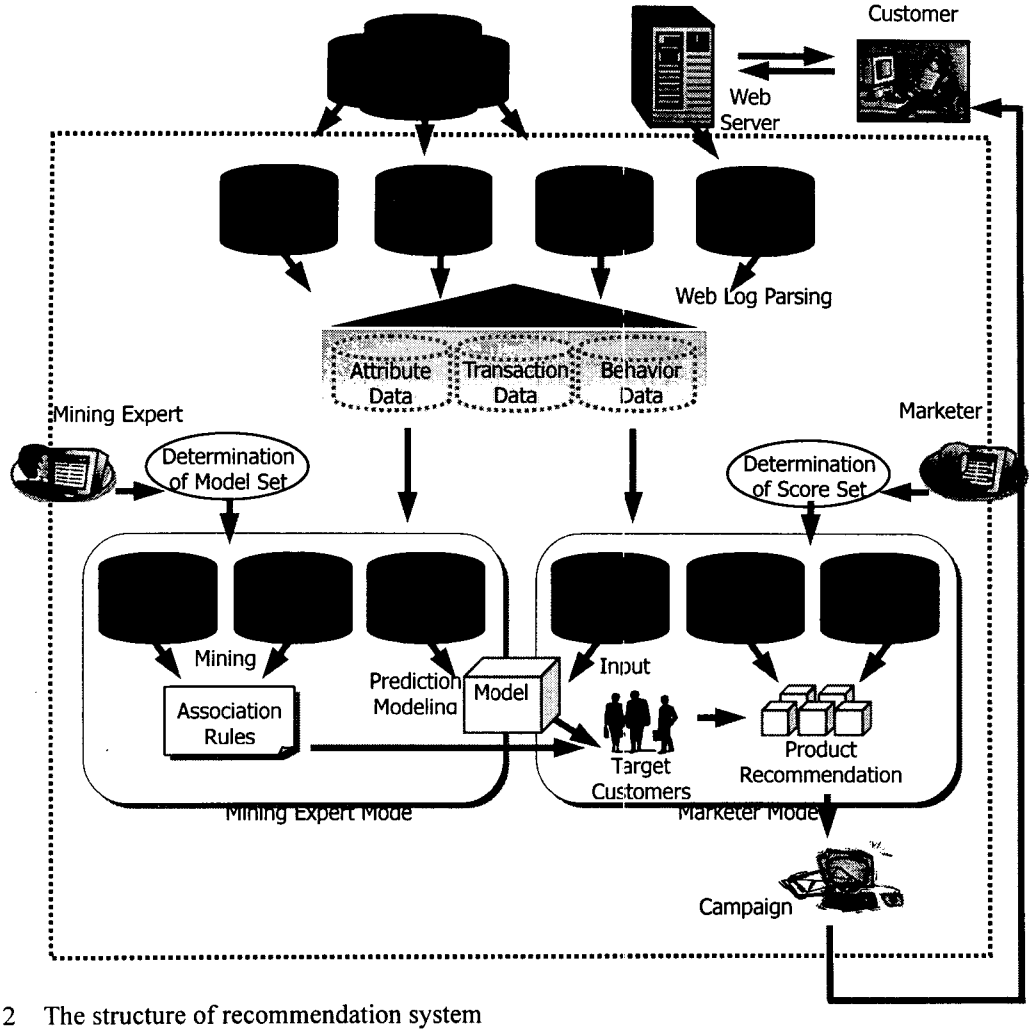

Figure 12 The structure of recommendation system

Case Study

We apply this recommendation system to K Corporation. K Co. has internet sites treating cosmetics, clothing, and lotteries. K Co. uses MS SQL Server 7.0 as database system, and services web-service by IIS5.0. Each site is managed by 4 web-servers and creates 16 log-files, compressing 300M web-log files.

Recommendation system is conducted by sales data and web-log data of K Co. The web-log data is from Jan. 2001. to May 2001. K Co. has 360 thousand customers, and sells about 3000 cosmetics on cosmetic site, and these products are organized by 3-levels. This research selects 3 level as an analyzing unit.

Also we studied recommendation system both expert mode and marketer mode. But now, we will examine Expert Mode's Process of Recommendation System. First, set the past period of model set from January to March 2001, and select customer who buys different class products more than 1 during the past period. Among customers, if they purchase product on May, the object variable is '1', if not the object variable is '0.' Training data is organized by age, job, recent visiting date, sum of price, and frequency of visiting. The model is created by using decision tree algorithm. Next, from January to May, sales data and basket data are used to find association rules. The screen to obtain input variables is shown at Figure 13

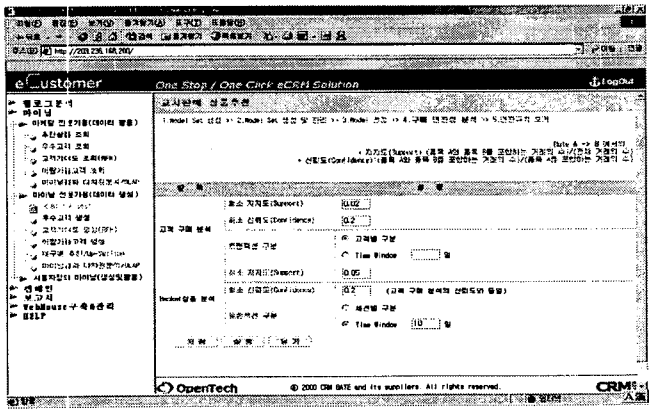

Figure 13 The Screen of Purchase Association Analysis

Minimum support of association rule mining of

purchased products is 0.02, and minimum support of basket product is 0.05. Figure 14 is an association rule of cosmetic products.

No.	Condition	Result	Support	Confidence
1	Mask	Foam cleansing	0.053	0.5
2	Basic Set	Whitening	0.153	0.65
3	Sun Block	Body lotion	0.07	0.54
4	Two way cake	Essence	0.025	0.46
...	...	...	...	...
32	Treatment	Styling	0.03	0.8

Figure 14 Association rule of cosmetic products

All of 52 association rules are selected and are used to marketer mode product recommendation.

Conclusion

This research shows a recommendation procedure and product recommendation system. The characteristics of suggested system is as follows. First, existing system recognized customers from the preference data, but our system uses customer purchase behavior pattern and purchase pattern information for the more exact information of customer preference. Second, we select not all customers but possible customers using prediction modeling technology. Third, when generating association rules, we can't get plenty of rules only by purchasing data, so rules are added from the basket analysis. Finally, we recommend high efficiency product by measuring product efficiency from the web-log.

This research realizes actual product recommendation system and applies to real shopping mall case. Also, we analyze the validity of product recommendation system. We apply product recommendation system at K Co. cosmetics site, but we didn't measure the effect of our suggested recommendation procedure, so it is a further research area to measure it. Also, it will be interesting to compare the performance of collaborative filtering association recommendation system and rule based association recommendation system.

Reference

[1] R. Agrawal, T. Imielinski, and A. Swami: Mining Association Rules between Sets of Items in Large Database, *In Proceedings. of the ACM SIGMOD Conference on Management of Data, pp. 207-216, Washington, D.C., May 1993.*

[2] R. Agrawal and R. Srikant: Fast Algorithms for Mining Association Rules, *In Proceedings of the 20th VLDB Conference, Santiago, Chile, Sept, 1994.*

[3] Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. GroupLens: An Open Architecture for Collaborative Filtering of Netnews, *In Proceedings of CSCW '94, Chapel Hill, NC.*

[4] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. GroupLens: Applying Collaborative Filtering to Usenet News, *Communications of the ACM, 40(3), pp. 77-87, 1997.*

[5] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl: Analysis of Recommendation Algorithm for E-Commerce, *In Proceeding of EC'00 Conference, October 17-20, 2000, Minneapolis, Minnesota*