Product Value Evaluation Models based on Itemset Association Chain

Yong Sik Chang
Department of e-Business, Hanshin University
(yschang@hs.ac.kr)

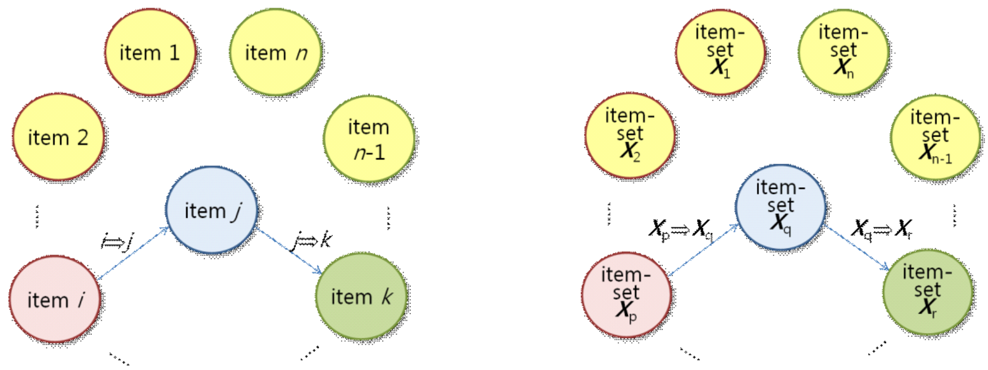
Association rules among product items by association analysis suggest sales effect among products. These are useful for marketing strategies such as cross-selling and product display etc. However, if we evaluate more practical product values reflecting cross-selling effects, they will be also more useful for the decisions of companies such as product item selection for product assortment and profit maximization etc. This study proposes product value evaluation models with the concept of effective value based on single-item association chain and itemset association chain. In addition to that, we performed experiments with transaction data related to clothing of an online shopping mall in Korea to show the performances of our models. In result, we confirmed that some items increased in effective values compared with their pure values while the others decreased in effective values.

Received : June 02, 2010 Revision : June 16, 2010 Accepted : June 21, 2010 Corresponding author : Yong Sik Chang

1. Introduction

By association analysis (Agrawal et al., 1993), some product items have associations among one another. For example, the association between complementary goods such as coffee and sugar is a representative illustration. In this case, the value of coffee has to be regarded as high because coffee contributes to the sales of sugar. <Figure 1> depicts association chains among product items explaining this situation. If product items affect the sales of other items, it can be considered that those contribute to additional profits of other items by association sales as well as their own profits. On the contrary, if their sales are affected by other items, their own contributions to the profits will be regarded as reduced. Product association chains represent

* This work was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD) (KRF-2006-332-B00098).



(a) Single-item association chain (b) Itemset association chain

<Figure 1> Product Association Chains

these associations among product items. <Figure 1(a)> depicts a single-item association chain between an item and another item and <Figure 1(b)> depicts an itemset association chain between an itemset and another itemset.

Like this, the evaluation of a practical product value including cross-selling effects to other products and cross-selling effects caused by other products will be more useful for the decisions of companies such as product item selection for product assortment and profit maximization etc. The purposes of this study are to propose product value evaluation models with the concept of effective value through association analysis and to show the availability of our model through experimentation for performance evaluation. For these purposes, this study is organized as follows. In section 2 we review the related research and in section 3 we propose the product value evaluation models for effective values based on single-item association and itemset association. And then, in section 4, we describe the performance evaluation for the models through experimentation to show this study's availability with the transaction data of an online shopping mall. Finally, we conclude this study with further research.

2. Literature Review

Association analysis is based on network theory. The network theory was started by graph theory published by a mathematician, Leonhard Euler, in 1736. In 1967, Stanley Milgram, a professor of Harvard University in USA, had published a study on social network and he announced that people live in the small world whose social links are 5.5 steps (Barabási, 2002).

Albert et al.(1999) applied this concept to the web and found a phenomenon that the frequencies of web documents to the numbers of web document links follow power-law distribution different from general bell-type distribution through the analysis of web connectivity. They explained that the phenomenon is caused by the growth of web and its preferential connectivity, and they provided a fomula on average distance between two arbitrary web documents. In addition to that, there had been studied on Erdös Number (Grossman et al., http://www.oakland.edu/enp/), nerve tissue network (Watts and Strogatz, 1998), power-law distribution of internet (Faloutsos et al., 1999), collaboration relationship among scientists (Barabási et al., 2002), and industry chains (Kim, 2003) etc.

On the other hand, association analysis among product items was started from the analysis of transaction data by Agrawal et al.(1993), and there were several studies on the improvement of effective algorithm for association analysis (Agrawal and Srikant, 1994; Park et al., 1995; Sarasere et al., 1995; Toivonen, 1996; Brin et al., 1997). These results are being applied to cross-selling and product recommendation etc. Brijs et al.(2000) proposed a product selection mdoel considering association rules to maximize gross margin and Hwang(2004, 2005) proposed an algorithm for the estimation of products' expected profits based on pure sales amount and cross-selling amount. Recently, Kim et al.(2009) studied the characteristics of product networks and Cho and Bang(2009) proposed a method on new product recommendation through social network analysis technique. However, they didn't provide the method for evaluating realistic product values considering multi-item effects based on product associations. Hence, it will be meaningful to propose product value evaluation models with the concept of effective value based on itemset associations.

3. Product Value Evaluation Models

Associations among products can explain the relationships of sales among products. We define the effective value as sum of a pure value and contribution values. The pure value of a product means its own value under the situation without associations among products. On the contrary, the contribution value is generated by associations among products and there are two types. One is the gain value that a product gets from other products when it contributes to the sales of them, and another is the loss value that a part of product value is taken from other products when its sales are caused by them.

The contribution value is related to the confidence resulted from association analysis. For example, an association rule $i \Rightarrow j$ means that product item i affects the sales of item j and its sales amount corresponds to Conf(j, i). In this case, an effective value of product item i is the sum of its pure value and the gain value from product item j. This gain value is equal to Conf(j, i)×(pure value of product item j). On the contrary, an effective value of product item j is the value that extracts loss value by product item i from its pure value. This loss value is equal to the gain value of product item i. In each product, there may exist gain values and loss values at the same time. Hence, we can get the effective value of a product item by adding gain values to its pure value and extracting loss values from the pure value. To formulate evaluation models for the effective value of a product item, we define notations as follows.

i, j, k : indexes of items.
p, q, r : indexes of itemsets.
t : index of transaction identification(TID).
$\boldsymbol{X}_p$: pth itemset.
$\mathrm{n}(\boldsymbol{X}_p)$: number of elements in an itemset $\boldsymbol{X}_p$.
$\boldsymbol{I}_t$: itemset of tth TID.
$\mathrm{Conf}(\boldsymbol{X}_p, \boldsymbol{X}_q)$: confidence of an association rule $\boldsymbol{X}_p \Rightarrow \boldsymbol{X}_q$.
$v_t(j)$: pure value of item j in TID t.
$V^0(j)$: pure value of item j in all TIDs, $\sum_t v_t(j)$.
$V^{\mathrm{gain}}(j)$: gain value of item j that gets from other items, $V^{\mathrm{gain}}(j, \boldsymbol{X}_p)$ means a gain value of item j that gets from itemset $\boldsymbol{X}_p$.
$V^{\mathrm{loss}}(j)$: loss value of item j that is taken from other items, $V^{\mathrm{loss}}(j, \boldsymbol{X}_p)$ means a loss value of item j that is taken from itemset $\boldsymbol{X}_p$.
$V^{\mathrm{eff}}(j)$: effective value of item j, $V^{\mathrm{eff,S}}(j)$ is for single-item association chain and $V^{\mathrm{eff,I}}(j)$ is for itemset association chain.
$\boldsymbol{Z}(t, j)$: union of antecedent itemsets to consequent item j by TID t, $\boldsymbol{Z}^{\mathrm{S}}(t, j)$ is for single-item association chain and $\boldsymbol{Z}^{\mathrm{I}}(t, j)$ is for itemset association chain.

Based on the defined notations, we describe an effective value. Equation (1) means an effective value of item j. The first term in the right side is its pure value, the second term is its gain value, and the third term is its loss value. Equation (2) means total effective value for all product items. Effective values of items may differ. However, Total effective value for all product items is equal to total pure value for all product items because total gain value and total loss value for all product items are equal to each other.

$$V^{\mathrm{eff}}(j) = V^0(j) + V^{\mathrm{gain}}(j) - V^{\mathrm{loss}}(j). \qquad (1)$$

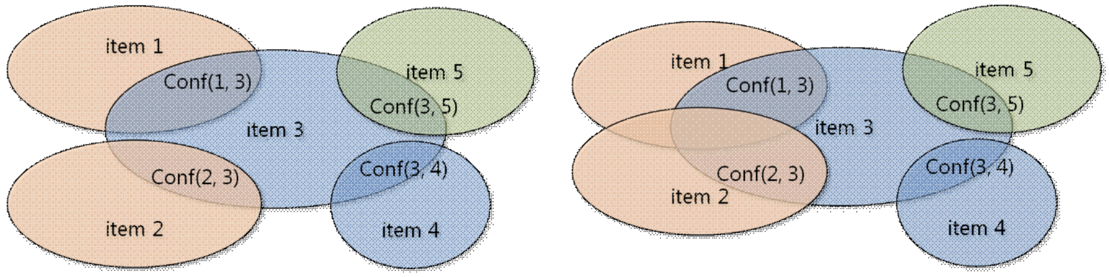
$$\textstyle\sum_j V^{\mathrm{eff}}(j) = \sum_j V^0(j). \qquad (2)$$

The evalaution models for gain values and loss values that determine each effective value differ in single-item association chain and itemset association chain. We describe them with two parts.

3.1 A Product Value Evaluation Model Based on Single-item Associations

If there exist association rules $i \Rightarrow j$ and $j \Rightarrow k$ for some item i, j, and k, those are single-item associations. <Figure 2> depicts diagrams of single-item associations. We assume that there exist four associations with Conf(1, 3), Conf(2, 3), Conf(3, 4), and Conf(3, 5) for item 3. <Figure 2(a)> depicts associations without duplicate TIDs related to Conf(1, 3) and Conf(2, 3). On the contrary, <Figure 2(b)> has associations with duplicate TIDs related to the confidences.

First we think a matter over the case of <Figure 2(a)>. The sales amount by an association rule $1 \Rightarrow 3$ with Conf(1, 3) corresponds to Conf (3, 1). In this case, the effective value of item 3 is reduced by the value that extracts a loss value $V^{\mathrm{loss}}(3, 1)$ from a pure value $V^0(3)$. In addition to that, item 3 affects the sales of item 4. In this case, the effective value of item 3 is in-



(a) Association without duplicate TIDs (b) Association with duplicate TIDs

<Figure 2> Diagrams of single-item associations

creased by the value that adds a gain value $V^{gain}(3, 4)$ to a pure value $V^0(3)$. Equations (3) and (4) describe a loss value of item 3 by items 1 and 2, and a gain value of item 3 from items 4 and 5.

$$V^{loss}(3) = V^{loss}(3,\ 1) + V^{loss}(3,\ 2) \qquad (3)$$
$$= \{\mathrm{Conf}(3,\ 1) + \mathrm{Conf}(3,\ 2)\}\ \ V^0(3).$$

$$V^{gain}(3) = V^{gain}(3,\ 4) + V^{gain}(3,\ 5) \qquad (4)$$
$$= \mathrm{Conf}(4,\ 3)\ \ V^0(4) + \mathrm{Conf}(5,\ 3)\ \ V^0(5).$$

However, in case of <Figure 2(b)> with duplicate TIDs related to Conf(1, 3) and Conf(2, 3), loss value and gain value differ from those of <Figure 2(a)>. We show <Table 1> with 10 TIDs for a concrete illustration. It depicts gain values of items 1 and 2 from item 3 caused by Conf(i, 3). TID groups 1 and 2 are illustrations of <Figure 2(a)> with minimum confidence of 75% and <Figure 2(b)> with minimum confidence of 80% respectively. Two TID groups produce association rules between single items, $1 \Rightarrow 3$ and $2 \Rightarrow 3$. And transaction items $\boldsymbol{I}_t$ in each TID t include items 1, 2, or 3.

As mentioned it before, TIDs in TID group 1 that contribute to Conf(1, 3) and Conf(2, 3) are not duplicated. In this case, total gain value of item 1 is equal to $v_2(3) \sim v_5(3)$ by TIDs {2, 3, 4, 5} that contribute to Conf(1, 3), and total gain value of item 2 is equal to $v_6(3) \sim v_8(3)$ by TIDs {6, 7, 8} that contribute to Conf(2, 3). On the other hand, TIDs in TID group 2 that contribute to Conf(1, 3) and Conf(2, 3) are duplicated by TIDs {5, 6, 7}. In this case, items 1 and 2 share gain values. Hence, we have to divide the TID group 2 into a duplicated part and others. That is, total gain value of item 1 is the sum of $v_2(3) \sim v_4(3)$ by TIDs {2, 3, 4} without duplication and $½v_5(3) \sim ½v_7(3)$ by TIDs {5, 6, 7} duplicated with item 2. And, total gain value of item 2 is the sum of $v_8(3)$ by TID {8} without duplication and $½v_5(3) \sim ½v_7(3)$ by TIDs {5, 6, 7} duplicated with item 1.

We introduce an union of antecedent itemsets to consequent item j by TID t, $\mathbf{Z}^S(t, j)$, to describe these situations. Equation (5) means $\mathbf{Z}^S(t, j)$ which is composed of all items i that satisfy $\mathrm{Conf}(i, j) \geq C_{min}$ for $i \in \boldsymbol{I}_t$, $j \in \boldsymbol{I}_t$.

$$\mathbf{Z}^S(t, j) = \{i\ |\mathrm{Conf}(i, j) \geq C_{min},\ \forall i \in \boldsymbol{I}_t,\ j \in \boldsymbol{I}_t\}. \qquad (5)$$

<Table 1> Gain values of items 1 and 2 from item 3 caused by Conf(*i*, 3)

TID Group 1 (<Figure 2(a)> with Conf(*i*, 3) ≥ 75%)							TID Group 2 (<Figure 2(b)> with Conf(*i*, 3) ≥ 80%)						
TID (t)	Items (I_t)	TID related to items(*i*, 3)		$Z^S(t, 3)$	$V^{gain}(1, 3)$	$V^{gain}(2, 3)$	TID (t)	Items (I_t)	TID related to items(*i*, 3)		$Z^S(t, 3)$	$V^{gain}(1, 3)$	$V^{gain}(2, 3)$
		$i = 1$	$i = 2$						$i = 1$	$i = 2$			
1	1	–	–	–	0	0	1	1	–	–	–	0	0
2	1, 3	A	–	{1}	$v_2(3)$	0	2	1, 3	A	–	{1}	$v_2(3)$	0
3	1, 3	A	–	{1}	$v_3(3)$	0	3	1, 3	A	–	{1}	$v_3(3)$	0
4	1, 3	A	–	{1}	$v_4(3)$	0	4	1, 3	A	–	{1}	$v_4(3)$	0
5	1, 3	A	–	{1}	$v_5(3)$	0	5	1, 2, 3	A	A	{1, 2}	$\frac{1}{2}v_5(3)$	$\frac{1}{2}v_5(3)$
6	2, 3	–	A	{2}	0	$v_6(3)$	6	1, 2, 3	A	A	{1, 2}	$\frac{1}{2}v_6(3)$	$\frac{1}{2}v_6(3)$
7	2, 3	–	A	{2}	0	$v_7(3)$	7	1, 2, 3	A	A	{1, 2}	$\frac{1}{2}v_7(3)$	$\frac{1}{2}v_7(3)$
8	2, 3	–	A	{2}	0	$v_8(3)$	8	2, 3	–	A	{2}	0	$v_8(3)$
9	2	–		–	0	0	9	2	–		–	0	0
10	3	–	–	–	0	0	10	3	–	–	–	0	0
Conf(*i*, 3)		4/5 (80%)	3/4 (75%)				Conf(*i*, 3)		6/7 (86%)	4/5 (80%)			

Remarks : 'A' means an antecedent item in an association rule.

A gain value that item i gets from item j varies according to $n\{Z^S(t, j)\}$ in case that item i is an element of $Z^S(t, j)$. It corresponds to $1/n\{Z^S(t, j)\}$. Hence, we define a step function as equation (6) that discriminate whether item i is an element of $Z^S(t, j)$ or not. It is equal to 1 if item i is an element of $Z^S(t, j)$, 0 otherwise.

$$\Theta\{i, Z^S(t, j)\} \equiv 1 \text{ if } i \in Z^S(t, j), \quad 0 \text{ otherwise.} \tag{6}$$

We use the step function and union of antecedent itemsets to consequent item to describe gain values. The gain value that item i gets from item j is equation (7) and total gain value of item i is equation (8).

$$V^{gain}(i, j) = \sum_t v_t(j)\ \Theta\{i, Z^S(t, j)\}/n\{Z^S(t, j)\}. \tag{7}$$

$$V^{gain}(i) = \sum_j V^{gain}(i, j) = \sum_{tj} v_t(j)\ \Theta\{i, Z^S(t, j)\}/n\{Z^S(t, j)\}. \tag{8}$$

On the contrary, a loss value of item j that is extracted from item i is equal to a gain value that item i gets from item j. Hence, a loss value can be expressed with equations (9) and (10) using equations (7) and (8) related to a gain value.

$$V^{loss}(j, i) = \sum_t v_t(j)\ \Theta\{i, Z^S(t, j)\}/ n\{Z^S(t, j)\}. \tag{9}$$

$$V^{loss}(j) = \sum_i V^{loss}(j, i) = \sum_{ti} v_t(j)\ \Theta\{i, Z^S(t, j)\}/n\{Z^S(t, j)\}. \tag{10}$$

After all, total effective value of item j can be expressed with equation (11) including a pure value, gain values, and loss values.

$$V^{\text{eff,S}}(j) = V^0(j) + \sum_{tk} v_t(k)\ \Theta\{i,\ \boldsymbol{Z}^{S}(t,\ j)\}/\ \text{n}\{\boldsymbol{Z}^{S}(t,\ k)\} - \sum_{ti} v_t(j)\ \Theta\{i,\ \boldsymbol{Z}^{S}(t,\ j)\}/\ \text{n}\{\boldsymbol{Z}^{S}(t,\ j)\}. \quad (11)$$

According to the equation (1), on the other hand, total effective value for all items has to be equal to total pure value for all items. The proof is as follows.

Proof of $\sum_j V^{\text{eff,S}}(j) = \sum_j V^0(j)$: Total effective value for all item j from equation (11) becomes equation (12).

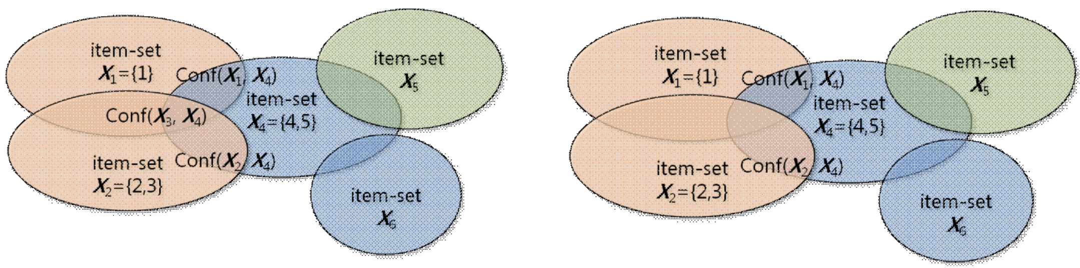
$$\sum_j V^{\text{eff,S}}(j) = \sum_j V^0(j) + \sum_{tkj} v_t(k)\ \Theta\{i,\ \boldsymbol{Z}^{S}(t,\ j)\}/\text{n}\{\boldsymbol{Z}^{S}(t,\ k)\} - \sum_{tij} v_t(j)\ \Theta\{i,\ \boldsymbol{Z}^{S}(t,\ j)\}/\text{n}\{\boldsymbol{Z}^{S}(t,\ j)\}. \quad (12)$$

If i and j in the second term in the right side of equation (12) are replaced with k and i respectively, the second term is equal to the third term and, in result, they are offset by each other. Hence, total effective value and total pure value for all items are equal to each other.

3.2 A product value evaluation model based on itemset associations

Itemset associations that an itemset affects sales of other itemsets can be expressed with Conf($\boldsymbol{X}_p$, $\boldsymbol{X}_q$) as depicted in <Figure 3>. We consider $\boldsymbol{X}_1 = \{1\}$, $\boldsymbol{X}_2 = \{2, 3\}$, $\boldsymbol{X}_3 = \{1, 2, 3\}$ for $\boldsymbol{X}_p$ and $\boldsymbol{X}_4 = \{4, 5\}$ for $\boldsymbol{X}_q$. <Table 2> with samples of 10 transaction data depicts itemsets $\boldsymbol{X}_1$, $\boldsymbol{X}_2$, and $\boldsymbol{X}_3$ that affect sales of an itemset $\boldsymbol{X}_4$.

<Figure 3(a)> depicts a case that Conf($\boldsymbol{X}_p$, $\boldsymbol{X}_4$) is more than or equal to 60% and there exist three association rules $\boldsymbol{X}_1 \Rightarrow \boldsymbol{X}_4$, $\boldsymbol{X}_2 \Rightarrow \boldsymbol{X}_4$ and $\boldsymbol{X}_3 \Rightarrow \boldsymbol{X}_4$. In this case, we consider gain values of items 1, 2, and 3 that get from item 4, $V^{\text{gain}}(1, 4)$, $V^{\text{gain}}(2, 4)$, and $V^{\text{gain}}(3, 4)$. Association rules related to $V^{\text{gain}}(1, 4)$ are $\boldsymbol{X}_1 \Rightarrow \boldsymbol{X}_4$ and $\boldsymbol{X}_3 \Rightarrow \boldsymbol{X}_4$ for $1 \in \boldsymbol{X}_p$. The union of TIDs that include elements of $\boldsymbol{X}_1$ and $\boldsymbol{X}_4$ and elements of $\boldsymbol{X}_3$ and $\boldsymbol{X}_4$ becomes $\{2, 3, 5, 6, 7\} \cup \{3, 6, 7\} = \{2, 3, 5, 6, 7\}$. Here, TIDs {2, 5} are affected by only item 1 and TIDs {3, 6, 7} are affected by three items 1, 2,



(a) Association without duplicate TIDs (b) Association with duplicate TIDs

<Figure 3> Diagrams of itemset associations

<Table 2> Gain values of items in itemset X_p from item 4($\in X_4$) caused by Conf(X_p, X_4)

TID (t)	Items (T_t)	<Figure 3(a)> with Conf(X_p, X_4)$\geq$60%							<Figure 3(b)> with Conf(X_p, X_4)$\geq$70%					
		TID related to X_p			$Z^{I}(t, 4)$	$V^{gain}(1, 4)$	$V^{gain}(2, 4)$	$V^{gain}(3, 4)$	TID related to X_p		$Z^{I}(t,4)$	$V^{gain}(1,4)$	$V^{gain}(2,4)$	$V^{gain}(3,4)$
		X_1	X_2	X_3					X_1	X_2				
1	1, 2, 3	–	–	–	–	0	0	0	–	–	–	0	0	0
2	1, 4	A	–	–	{1}	$v_2(3)$	0	0	A	–	{1}	$v_2(3)$	0	0
3	1, 2, 3, 4	A	A	A	{1, 2, 3}	⅓$v_3(3)$	⅓$v_3(3)$	⅓$v_3(3)$	A	A	{1, 2, 3}	⅓$v_3(3)$	⅓$v_3(3)$	⅓$v_3(3)$
4	2, 3, 4	–	A	–	{2, 3}	0	½$v_4(3)$	½$v_4(3)$	–	A	{2, 3}	0	½$v_4(3)$	½$v_4(3)$
5	1, 4	A	–	–	{1}	$v_5(3)$	0	0	A	–	{1}	$v_5(3)$	0	0
6	1, 2, 3, 4	A	A	A	{1, 2, 3}	⅓$v_6(3)$	⅓$v_6(3)$	⅓$v_6(3)$	A	A	{1, 2, 3}	⅓$v_6(3)$	⅓$v_6(3)$	⅓$v_6(3)$
7	1, 2, 3, 4	A	A	A	{1, 2, 3}	⅓$v_7(3)$	⅓$v_7(3)$	⅓$v_7(3)$	A	A	{1, 2, 3}	⅓$v_7(3)$	⅓$v_7(3)$	⅓$v_7(3)$
8	2, 3, 4	–	A	–	{2, 3}	0	½$v_8(3)$	½$v_8(3)$	–	A	{2, 3}	0	½$v_8(3)$	½$v_8(3)$
9	2, 3, 4	–	A	–	{2, 3}	0	½$v_9(3)$	½$v_9(3)$	–	A	{2, 3}	0	½$v_9(3)$	½$v_9(3)$
10	1, 2, 3	–	–	–	–	0	0		–	–	–	0	0	
Conf(X_p, X_4)		5/7 (71%)	6/8 (75%)	3/5 (60%)					5/7 (71%)	6/8 (75%)				

Remarks : 'A' means an antecedent itemset in an association rule.

and 3. Hence, a gain value of item 1 is composed of $v_2(3)$ and $v_5(3)$ by TIDs {2, 5} and ⅓$v_3(3)$, ⅓$v_6(3)$, and ⅓$v_7(3)$ by TIDs {3, 6, 7}. Association rules related to $V^{gain}(2, 4)$ are $X_2 \Rightarrow X_4$ and $X_3 \Rightarrow X_4$ for $2 \in X_p$. The union of TIDs that include elements of X_2 and X_4 and elements of X_3 and X_4 becoms {3, 4, 6, 7, 8, 9}∪{3, 6, 7}={3, 4, 6, 7, 8, 9}. Here, TIDs {3, 6, 7} are affected by three items 1, 2, and 3, and TIDs {4, 8, 9} are affected by two items 2 and 3. Hence, a gain value of item 2 is composed of ⅓$v_3(3)$, ⅓$v_6(3)$, and ⅓$v_7(3)$ by TIDs {3, 6, 7} and ½$v_4(3)$, ½$v_8(3)$, and ½$v_9(3)$ by TIDs {4, 8, 9}. Association rules related to V^{gain} (3, 4) are $X_2 \Rightarrow X_4$ and $X_3 \Rightarrow X_4$ for $3 \in X_p$. Those are same with the case of $V^{gain}(2, 4)$.

Next thing, we consider <Figure 3(b)> that Conf(X_p, X_4) is more than or equal to 70% and there exist two association rules $X_1 \Rightarrow X_4$ and $X_2 \Rightarrow X_4$ except $X_3 \Rightarrow X_4$ in <Figure 3(a)>. An association rule related to $V^{gain}(1, 4)$ is $X_1 \Rightarrow X_4$ for $1 \in X_p$. The union of TIDs that include elements of X_1 and X_4 becomes {2, 3, 5, 6, 7}. Here, TIDs {2, 5} are affected by only item 1 and TIDs {3, 6, 7} are affected by three items 1, 2, and 3. Hence, a gain value of item 1 is composed of $v_2(3)$ and $v_5(3)$ by TIDs {2, 5} and ⅓$v_3(3)$, ⅓$v_6(3)$, and ⅓$v_7(3)$ by TIDs {3, 6, 7}. An association rule related to $V^{gain}(2, 4)$ is $X_2 \Rightarrow X_4$ for $2 \in X_p$. The union of TIDs that include elements of X_2 and X_4 becomes {3, 4, 6, 7, 8, 9}. Here, TIDs {3, 6, 7} are affected by three items 1, 2, and 3, and

TIDs {4, 8, 9} is affected by two items 2 and 3. Hence, a gain value of item 2 is composed of ⅓ $v_3(3)$, ⅓$v_6(3)$, and ⅓$v_7(3)$ by TIDs {3, 6, 7} and ½$v_4(3)$, ½$v_8(3)$, and ½$v_9(3)$ by TIDs {4, 8, 9}. An association rule related to $V^{gain}(3, 4)$ is $\boldsymbol{X}_2 \Rightarrow \boldsymbol{X}_4$ for $3 \in \boldsymbol{X}_p$. This is the same with the case of $V^{gain}(2, 4)$.

We introduce an union of antecedent itemsets to consequent item j by TID t, $\boldsymbol{Z}^{\mathbf{I}}(t, j)$, to describe these situations. Equation (13) means $\boldsymbol{Z}^{\mathbf{I}}(t, j)$ which is composed of all items i in all itemsets $\boldsymbol{X}_p$ that satisfy $\mathrm{Conf}(\boldsymbol{X}_p, \boldsymbol{X}_q) \geq C_{min}$ for $i \in \boldsymbol{X}_p$, $j \in \boldsymbol{X}_q$, $\boldsymbol{X}_p \in \boldsymbol{I}_t$, $\boldsymbol{X}_q \in \boldsymbol{I}_t$. If we replace $\boldsymbol{X}_p$ and $\boldsymbol{X}_q$ with i and j respectively, we know that $\boldsymbol{Z}^{\mathbf{I}}(t, j)$ becomes same with $\boldsymbol{Z}^{\mathbf{S}}(t, j)$. A gain value that item i gets from item j varies according to $\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, j)\}$ in case that item i is an element of $\boldsymbol{Z}^{\mathbf{I}}(t, j)$. It corresponds to $1/\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, j)\}$. Hence, we define a step function as equation (14) that discriminate whether item i is an element of $\boldsymbol{Z}^{\mathbf{I}}(t, j)$ or not. It is equal to 1 if item i is an element of $\boldsymbol{Z}^{\mathbf{I}}(t, j)$, 0 otherwise.

$$\boldsymbol{Z}^{\mathbf{I}}(t, j) = \{i \mid \mathrm{Conf}(\boldsymbol{X}_p, \boldsymbol{X}_q) \geq C_{min}, \forall i \in \boldsymbol{X}_p, j \in \boldsymbol{X}_q, \boldsymbol{X}_p \in \boldsymbol{I}_t, \boldsymbol{X}_q \in \boldsymbol{I}_t\}. \quad (13)$$

$$\Theta\{i, \boldsymbol{Z}^{\mathbf{I}}(t, j)\} \equiv 1 \text{ if } i \in \boldsymbol{Z}^{\mathbf{I}}(t, j), \; 0 \text{ otherwise}. \quad (14)$$

After all, total effective value of item j can be expressed with equation (15) including a pure value, gain values, and loss values. $V^{eff,\mathbf{I}}(j)$ also includes $V^{eff,\mathbf{S}}(j)$ because $\boldsymbol{Z}^{\mathbf{I}}(t, k)$ includes $\boldsymbol{Z}^{\mathbf{S}}(t, k)$.

$$V^{eff,\mathbf{I}}(j) = V^0(j) + \sum_{tk} v_t(k)\, \Theta\{i, \boldsymbol{Z}^{\mathbf{I}}(t, j)\}/\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, k)\} - \sum_{ti} v_t(j)\, \Theta\{i, \boldsymbol{Z}^{\mathbf{I}}(t, j)\}/\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, j)\}. \quad (15)$$

Under the itemset associations, total effective value for all items has to be equal to total pure value for all items also. The proof is as follows.

Proof of $\sum_j V^{eff,I}(j) = \sum_j V^0(j)$: Total effective value for all item j from equation (15) becomes equation (16).

$$\sum_j V^{eff,I}(j) = \sum_j V^0(j) + \sum_{tkj} v_t(k)\, \Theta\{i, \boldsymbol{Z}^{\mathbf{I}}(t, j)\}/\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, k)\} - \sum_{tij} v_t(j)\, \Theta\{i, \boldsymbol{Z}^{\mathbf{I}}(t, j)\}/\mathrm{n}\{\boldsymbol{Z}^{\mathbf{I}}(t, j)\}. \quad (16)$$

If i and j in the second term in the right side of equation (16) are replaced with k and i respectively, the second term is equal to the third term and, in result, they are offset by each other. Hence, total effective value and total pure value for all items are equal to each other.

4. Performance Evaluation through Illustrative Experiments

We performed illustrative experiments for performance evaluation of our models from the perspective of profit. For practical applications, we used transaction data in the year 2006 related to clothing of an online shopping mall in Korea. Total number of item categories is 106. Total number of TIDs including more than or equal to two item categories reaches 48,044 and average

<Table 3> Association rules

Rule ID	Rule		Confidence(%)	Lift
1	20	⇒ 2	72.727	1.848
2	14	⇒ 13	80.0	15.204
3	1, 15	⇒ 2	75.0	1.906
4	5, 6, 17	⇒ 2	80.0	2.033
5	1, 6, 18	⇒ 2	73.333	1.863
6	6, 10, 18	⇒ 2	72.727	1.848
7	5, 6, 18	⇒ 2	80.0	2.033
8	6, 12, 15	⇒ 2	84.211	2.14
9	1, 6, 15	⇒ 2	84.615	2.15
10	6, 13, 15	⇒ 2	83.333	2.117
11	6, 11, 19	⇒ 2	72.727	1.848
12	2, 11, 19	⇒ 6	72.727	2.367
13	4, 6, 12	⇒ 2	70.588	1.794
14	1, 4, 5	⇒ 2	72.727	1.848
15	1, 6, 13	⇒ 2	75.676	1.923
16	1, 3, 5	⇒ 2	76.471	1.943
17	5, 6, 13, 16	⇒ 2	83.333	2.117
18	6, 7, 8, 9	⇒ 2	81.818	2.097

number of item categories by TID is 4.4. <Table 3> describes 18 assocaition rules found by association analysis with minimum confidence of 70% and minimum support of 0.05%. Because of diverse and similar item categories in the sample set and customers' diverse purchase patterns for clothing, we can't find association rules at high minimum support. The antecedent and consequent item categories in the rules are as follows : casual shirt(1), T-shirt(2), one-piece dress(3), vest (4), skirt(5), pants(6), jacket(7), cardigan(8), jumper(9), blouse(10), belt(11), cap(12), small pieces of clothing(13), string of mobile phone(14), sneakers(15), sweater(16), easy ware(17), casual shoes (18), tote bag(19), and toe slipper(20). We used SPSS Clementine 12.0 for association analysis and implemented JAVA applications for calculating effective values by product evaluation model based on itemset association chain.

<Table 4> shows gain values from item category 2 by TID related to rules R3, R8, and R9. 34 TIDs depict related rule IDs, antecedent and consequent item categories, union of antecedent item category sets to consequent item category 2, profit of item category 2, and gain values from item category 2. The reason why profits of item category 2 are not all same is that either items in the item category or their quantities sold are different.

Based on gain values of <Table 4>, we simulate gain and loss values for 20 item categories related to 18 association rules in <Table 3>. <Table 5> depicts the results. Total gain value of 20 item categories reaches ₩2,325,847. By association rules, item category 2 belongs to consequent item category sets in most of rules among 18 rules. Hence, its loss value $V^{loss}(2)$, ₩2,119,793, is relatively large. Loss values of item categories 6 and 13 are ₩138,348 and ₩67,706 respectively. On the other hand, item category 2 also belongs to an antecedent itemset in rule ID 12 in <Table 3> and its gain value that gets from item category 6 is ₩46,116. Similarly, a gain value of item category 6 that gets from 2 is ₩522,868.5 and a gain value of item category 13 that gets from item category 2 is ₩196,866.7. That is, there exist gain and loss values in item categories 2, 6, and 13 at the same time.

<Table 4> Gain values from item category 2 by TID related to rules R3, R8 and R9

TID (t)	Rule ID	Item category(j) 1	2	6	12	15	$Z^I(t, 2)$	Profit(2)(₩)	$V^{gain}(j, 2)$(₩) $V^{gain}(1, 2)$	$V^{gain}(6, 2)$	$V^{gain}(12, 2)$	$V^{gain}(15, 2)$
1	R3	A	C			A	{1, 15}	4,025	2,012.5	–	–	2,012.5
2	R8		C	A	A	A	{6, 12, 15}	29,820	–	9,940	9,940	9,940
3	R3	A	C			A	{1, 6, 15}	6,670	2,223.3	2,223.3	–	2,223.3
	R9	A	C	A		A						
4	R3	A	C			A	{1, 6, 15}	10,425	3,475	3,475	–	3,475
	R9	A	C	A		A						
5	R8		C	A	A	A	{6, 12, 15}	9,822	–	3274	3274	3274
6	R3	A	C			A	{1, 15}	5,220	2,610	–	–	2,610
7	R3	A	C			A	{1, 6, 15}	4,410	1,470	1,470	–	1,470
	R9	A	C	A		A						
8	R8		C	A	A	A	{6, 12, 15}	30,756	–	10,252	10,252	10,252
9	R8		C	A	A	A	{6, 12, 15}	8,820	–	2,940	2,940	2,940
10	R3	A	C			A	{1, 15}	4,500	2,250	–	–	2,250
11	R3	A	C			A	{1, 6, 15}	22,854	7,618	7,618	–	7,618
	R9	A	C	A		A						
12	R3	A	C			A	{1, 15}	11,440	5,720	–	–	5,720
13	R8		C	A	A	A	{6, 12, 15}	4,320	–	1,440	1,440	1,440
14	R8		C	A	A	A	{6, 12, 15}	32,230	–	10,743.3	10,743.3	10,743.3
15	R8		C	A	A	A	{6, 12, 15}	8,700	–	2,900	2,900	2,900
16	R8		C	A	A	A	{6, 12, 15}	7,840	–	2,613.3	2,613.3	2,613.3
17	R3	A	C			A	{1, 6, 15}	9,430	3,143.3	3,143.3	–	3,143.3
	R9	A	C	A		A						
18	R3	A	C			A	{1, 6, 15}	6,400	2,133.3	2,133.3	–	2,133.3
	R9	A	C	A		A						
19	R3	A	C			A	{1, 15}	10,504	5,252	–	–	5,252
20	R3	A	C			A	{1, 6, 15}	3,450	1,150	1,150	–	1,150
	R9	A	C	A		A						
21	R8		C	A	A	A	{6, 12, 15}	8,690	–	2,896.7	2,896.7	2,896.7
22	R3	A	C			A	{1, 15}	3,450	1,725	–	–	1,725
23	R8		C	A	A	A	{6, 12, 15}	3,990	–	1,330	1,330	1,330
24	R8		C	A	A	A	{6, 12, 15}	5,750	–	1,916.7	1,916.7	1,916.7
25	R3	A	C			A	{1, 6, 12, 15}	17,020	4,255	4,255	4,255	4,255
	R8		C	A	A	A						
	R9	A	C	A		A						
26	R3	A	C			A	{1, 6, 15}	2,310	770	770	–	770
	R9	A	C	A		A						
27	R3	A	C			A	{1, 6, 15}	20,490	6,830	6,830	–	6,830
	R9	A	C	A		A						
28	R8		C	A	A	A	{6, 12, 15}	15,930	–	5,310	5,310	5,310
29	R8		C	A	A	A	{6, 12, 15}	1,470	–	490	490	490
30	R3	A	C			A	{1, 6, 15}	2,700	900	900	–	900
	R9	A	C	A		A						
31	R8		C	A	A	A	{6, 12, 15}	4,500	–	1,500	1,500	1,500
32	R3	A	C			A	{1, 15}	3,750	1,875	–	–	1,875
33	R8		C	A	A	A	{6, 12, 15}	3,150	–	1,050	1,050	1,050
Total								324,836	55,412	92,564	62,851	114,008
									324,836			

Remarks : 'A' means an antecedent item category in an association rule and 'C' means a consequent item category.

<Table 5> Gain and loss values between item categories

(Unit : ₩)

Item category(i)	Item categiry(j)	2	6	13	Total
	Loss value / Gain value	$V^{loss}(2,i)$	$V^{loss}(6,i)$	$V^{loss}(13,i)$	$V^{gain}(i)$
1	$V^{gain}(1, j)$	331,913.5	0	0	331,913.5
2	$V^{gain}(2, j)$	0	46,116	0	46,116
3	$V^{gain}(3, j)$	68,077	0	0	68,077
4	$V^{gain}(4, j)$	105,033.7	0	0	105,033.7
5	$V^{gain}(5, j)$	172,008.7	0	0	172,008.7
6	$V^{gain}(6, j)$	522,868.5	0	0	522,868.5
7	$V^{gain}(7, j)$	20,306.5	0	0	20,306.5
8	$V^{gain}(8, j)$	20,306.5	0	0	20,306.5
9	$V^{gain}(9, j)$	20,306.5	0	0	20,306.5
10	$V^{gain}(10, j)$	30,084.5	0	0	30,084.5
11	$V^{gain}(11, j)$	31,947	46,116	0	78,063
12	$V^{gain}(12, j)$	153,402	0	0	153,402
13	$V^{gain}(13, j)$	196,866.7	0	0	196,866.7
14	$V^{gain}(14, j)$	0	0	67,706	67,706
15	$V^{gain}(15, j)$	188,583.5	0	0	188,583.5
16	$V^{gain}(16, j)$	25,715	0	0	25,715
17	$V^{gain}(17, j)$	19,713	0	0	19,713
18	$V^{gain}(18, j)$	86,231.4	0	0	86,231.4
19	$V^{gain}(19, j)$	31,947	46,116	0	78,063
20	$V^{gain}(20, j)$	94,482	0	0	94,482
Total	$V^{loss}(j)$	2,119,793	138,348	67,706	2,325,847

Finally, we get pure values, gain values, effective values, and differences of 20 item categories as depicted in <Table 6>. Pure values were calculated from sample transaction data. And gain and loss values were reorganized from <Table 5>. In percentage rates of gain values, the rate of item category 14 is 105.59% and largest. Next thing, item category 20 reaches 83.37%. There exist loss values in only three item categories 2, 6, and 13. Those are 0.76%, 0.05%, and 0.25% respectively. As a result, the effective value to pure value of item category 2 was reduced while effective values of other item categories were increased. Item category with the largest difference is item category 14 and its value is 205.59%. On the contrary, effective value of item category 2 was reduced by 99.26%. Difference varies according to item category, but total effective value

<Table 6> Pure and effective values by item category

(Unit : ₩)

Item category (j)	Pure value[A] ($V^0(j)$)	Gain value ($V^{gain}(j)$)		Loss value ($V^{loss}(j)$)		Effective value[B] ($V^{eff,I}(j)$)		Difference[B−A]	
1	29,924,626	331,913.5	(1.11)	0	(0)	30,256,539.5	(101.11)	331,913.5	(1.11)
2	278,786,388	46,116	(0.02)	2,119,793	(0.76)	276,712,711	(99.26)	−2,073,677	(−0.74)
3	219,121,541	68,077	(0.03)	0	(0)	219,189,618	(100.03)	68,077	(0.03)
4	28,124,190	105,033.7	(0.37)	0	(0)	28,229,223.7	(100.37)	105,033.7	(0.37)
5	229,528,960	172,008.7	(0.07)	0	(0)	229,700,968.7	(100.07)	172,008.7	(0.07)
6	305,606,344	522,868.5	(0.17)	138,348	(0.05)	305,990,864.5	(100.13)	384,520.5	(0.13)
7	335,300,399	20,306.5	(0.01)	0	(0)	335,320,705.5	(100.01)	20,306.5	(0.01)
8	100,407,644	20,306.5	(0.02)	0	(0)	100,427,950.5	(100.02)	20,306.5	(0.02)
9	90,760,771	20,306.5	(0.02)	0	(0)	90,781,077.5	(100.02)	20,306.5	(0.02)
10	183,350,269	30,084.5	(0.02)	0	(0)	183,380,353.5	(100.02)	30,084.5	(0.02)
11	17,145,244	78,063	(0.46)	0	(0)	17,223,307	(100.46)	78,063	(0.46)
12	10,831,742	153,402	(1.42)	0	(0)	10,985,144	(101.42)	153,402	(1.42)
13	26,757,458	196,866.7	(0.74)	67,706	(0.25)	26,886,618.7	(100.48)	129,160.7	(0.48)
14	64,120	67,706	(105.59)	0	(0)	131,826	(205.59)	67,706	(105.59)
15	8,253,371	188,583.5	(2.28)	0	(0)	8,441,954.5	(102.28)	188,583.5	(2.28)
16	197,115,035	25,715	(0.01)	0	(0)	197,140,750	(100.01)	25,715	(0.01)
17	2507,351	19,713	(0.79)	0	(0)	2,527,064	(100.79)	19,713	(0.79)
18	7,898,501	86,231.4	(1.09)	0	(0)	7,984,732.4	(101.09)	86,231.4	(1.09)
19	14,018,478	78,063	(0.56)	0	(0)	14,096,541	(100.56)	78,063	(0.56)
20	113,322	94,482	(83.37)	0	(0)	207,804	(183.37)	94,482	(83.37)
Total	2,085,615,754	2,325,847	(0.11)	2,325,847	(0.11)	2,085,615,754	(100.0)	0	(0.0)

Remarks : ( ) means a percentage rate(%) to pure value.

and total pure value for all items are same each other.

5. Conclusion

In a company, final goal is to pursuit profit. Hence, it is meanful to evaluate product values according to the degree that products contribute to their company from the perspective of profit. This study measured more practical product values with the concept of effective value that is composed of pure value, gain values, and loss values. To achieve this goal, we proposed two mathematical product value evaluation models based on single-item associations and itemset associations.

We performed practical experiments using transaction data of clothing of an online shopping mall in Korea to show our models' availability. As a result of analysis to 18 association rules, 20 item categories affect 3 item categories and their effective values differ from their pure values. That is, we confirmed that some items increased in effective values compared with their pure values while the others decreased in effective values.

Association rules depend on transaction data. The sample data used in this study are classified into diverse and similar item categories from characteristics of clothing. This produces the results that we can't find association rules with high minimum support. To find useful association rules, it needs more effective categorization. It will enable us to find more accurate effective values.

Companies already apply association analysis to marketing strategies such as cross-selling and product display etc. In addition to that, this study's effective value can be used for evaluating more practical product values and selecting optimal products that contribute to high profit. This study provides the method for evaluating effective values of product items with 1st order association chain between itemsets. The effective values considering higher order cross-selling effects will be more useful.

References

Agrawal, R., T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases", *In Proceedings of the ACM SIGMOD Conference on Management of Data*, (1993), 206～216.

Agrawal, R. and R. Srikant, "Fast Algorithms for Mining Association Rules", *The International Conference on Very Large Database*, (1994), 487～499.

Albert, R., H. Jeong, and A. L. Barabási, "Diameter of the World Wide Web", *Nature*, Vol.401(1999), 130～131.

Barabási, A.-L., Linked : The New Science of Networks, Perseus Books Group, 2002.

Barabási, A.-L., H. Jeong, Z. Neda, E. Ravasz, A. Schubert, and T. Vicsek, "Evolution of the Social Network of Scientific Collaborations", *Physica A*, Vol.311(2002), 590～614.

Brijs, T., B. Goethals, G. Swinnen, K. Vanhoof, and G. Wets, "A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data : The Generalized PROFSET Model", *In Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, (2000), 300～304.

Brin, S., R. Motwani, J. D. Ullman, and S. Tsur, "Dynamic Itemset Counting and Implication Rules for Market Basket Data", *SIGMOD Record(ACM Special Interest Group on Management of Data)*, Vol.26, No.2(1997), 255～264.

Cho, Y. H. and J. H. Bang, "Social Network Analysis for New Product Recommendation", *Journal of Intelligence and Information Systems*, Vol.15, No.4(2009), 177～194.

Faloutsos, M., P. Faloutsos, and C. Faloutsos, "On Power-Law Relationships of the Internet Topology", *Computer Communications Review*, Vol.29(1999), 251～262.

Grossman, J. et al., The Erdös Number Project, http://www.oakland.edu/enp.

Hwang, I. S., "Item Selection by Estimated Profit

Ranking based on Association Rule", *Asia Pacific Journal of Information Systems*, Vol. 14, No.4(2004), 83～97.

Hwang, I. S., "A Study on the Business Value of Products Considering Cross Selling Effect", *Asia Pacific Journal of Information Systems*, Vol.15, No.3(2005), 210～221.

Kim, Y. H., Social Network Theory, Parkyoungsa, Korea, 2003.

Kim, H. K., J. K. Kim, and Q.-Y. Chen, "A Network Approach to Derive Product Relations and Analyze Topological Characteristics", *Journal of Intelligence and Information Systems*, Vol.15, No.4(2009), 155～176.

Park, J. S., M.-S. Chen, and P. S. Yu, "An Effective Hash based Algorithm for Mining Association Rules", *In M. J. Carey and D. A. Schneider, editors, Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data*, San Jose, CA(1995), 175～186.

Sarasere, A., E. Omiecinsky, and S. Navathe, "An Efficient Algorithm for Mining Association Rules in Large Databases", *In Proceedings of 21st International Conference on Very Large Data Bases*, (1995), 432～444.

Toivonen, H., "Sampling Large Databases for Association Rules", *In Proceedings of 22nd International Conference on Very Large Data Bases*, (1996), 134～145.

Watts, D. J. and S. H. Strogatz, "Collective Dynamics of Small-World Networks", *Nature*, Vol.393(1998), 440～442.

Abstract

상품군 연관망 기반의 상품가치 평가모형

장용식*

연관분석에 의한 연관규칙은 상품 간 연관성을 나타내고 있으며, 교차판매와 상품진열 등의 마케팅 전략에 활용되고 있다. 그러나, 교차판매 효과를 반영하는 더 실질적인 상품가치를 평가한다면, 상품구색과 이윤극대화 등의 기업 의사결정에 더 유용하게 활용될 수 있을 것이다. 본 연구는 단일상품 간의 연관망과 상품군 간의 연관망 기반에서 상품의 순수가치, 이득가치, 손실가치로 구성되는 유효가치를 바탕으로 상품의 가치를 평가하는 수학적 모형을 제시하고, 두 모형에 대해 각각 예시를 통한 평가과정을 기술하였다. 이 경우, 상품군은 단일상품을 포함하기 때문에 상품군 간 연관망 기반에서의 상품가치 평가모형은 단일상품 기반의 평가모형을 포함하고 있다.

모형의 실질적인 유용성을 보이기 위하여, 국내 한 온라인 쇼핑몰의 과거 1년 간 의류 관련 거래 데이터 표본을 이용하여 상품분류군 간의 연관규칙을 발견하고 상품분류별 유효가치를 평가하는 실험을 하였다. 표본은 총 106개 상품분류와 48,044건의 거래 데이터로 이루어져 있다. 먼저, SPSS Clementine 12.0을 이용하여 상품분류군 간 18개의 연관규칙을 발견하였다. 한편, 순수가치와 연관규칙을 바탕으로 이득가치, 손실가치를 계산하고 유효가치를 평가하는 JAVA 어플리케이션을 구현하였다. 유효가치 평가의 실험결과, 순수가치보다 큰 유효가치를 갖는 상품분류가 있는 반면, 순수가치보다 작은 유효가치를 갖는 상품분류를 확인하였다.

본 연구는 상품 연관망에서 일차적인 관계만을 고려한 유효가치를 평가하였다. 향후, 다단계 연관성의 교차판매 효과를 반영하는 고차적인 평가모형 연구는 보다 효과적인 유효가치 평가를 가능케 할 것이다.

Keywords : 연관분석, 상품가치 평가모형, 유효가치

* 한신대학교 경상대학 교수

저 자 소 개

장용식

현재 한신대학교 경상대학 e-비즈니스학과에 재직 중이다. KAIST에서 경영공학 박사학위를 취득하였으며, POSDATA 등 기업에서 MIS와 전자상거래 관련 개발 및 연구과제를 수행하였다. *Decision Support Systems*, *OMEGA*, *Expert Systems with Applications*, *Journal of Organizational Computing and Electronic Commerce* 등의 국제학술지와 그 외 국내학술지에 다수 논문을 게재하였으며, 주요 관심분야는 지능형 정보시스템을 위한 의사결정 모형화이다.