

## 교차판매효과를 고려한 상품의 가치평가에 관한 연구

황인수\*

### A Study on the Business Value of Products Considering Cross Selling Effect

Insoo Hwang

One of the most fundamental problems in business is to evaluate the value of each product. The difficulty is that the profit of one product not only comes from its own sales, but also its influence on the sales of other products, i.e., the "cross-selling effect".

This study integrates a measure for cross selling and an algorithm for profit estimation. Sales transaction data and post sales survey data from on-line and off-line shopping mall is used to show the effectiveness of the method against other heuristic for profit estimation based on product-specific profitability. We show that with the use of the new method we are able to identify the cross-selling potential of each product and use the information for better product selection.

**Keywords :** Cross-Selling, Product Value, Data Mining, Association Rule

---

\* 전주대학교 정보기술공학부 부교수

## I. 서론

데이터마이닝(data mining)은 대규모의 데이터웨어하우스에 저장되어 있는 데이터를 분석하여 조직의 의사결정을 지원하기 위한 새로운 정보 및 지식을 추출하는 과정이다. 이에 따라, 데이터 혹은 변수간의 상관관계를 분석하거나 시계열 분석 등을 위해 다양한 통계적 방법론을 사용한다[Agrawal *et al.*, 1993, 1994, 1996; Brin *et al.*, 1997; Zaki *et al.*, 1997].

데이터마이닝에서 가장 많이 사용되고 있는 기법의 하나인 장바구니 분석(basket analysis)은 상품추천, 상품진열, 그리고 교차판매(cross-selling) 및 상향판매(up-selling)에 매우 효과적으로 활용되고 있다[Shardanand and Maes, 1995; Konstan *et al.*, 1997; Schafer *et al.*, 1999; Kitts *et al.*, 2000; Deshpande and Karypis, 2004; Herlocker *et al.*, 2004]. 여기서 교차판매는 하나의 상품이 다른 상품의 판매에 영향을 미치는 현상으로서, 기업들은 매출의 증대를 위해 교차판매를 유도하는 다양한 전략들을 수립하여 실행하고 있다. 특히, 다양한 상품을 구비하고 있는 경우 교차판매의 중요성은 더욱 커지며, 최근에는 효과적인 고객관계관리(CRM, Customer Relationship Management)를 위한 필수적인 요소가 되고 있다[김민주, 2005].

시어즈(Sears)는 여러 개의 계열사로 이루어진 기업에서 대규모적인 교차판매가 어떻게 이루어질 수 있는지를 보여주는 대표적인 사례이다[박찬욱, 1996]. 시어즈는 백화점, 카탈로그회사, 은행, 신탁회사, 부동산회사, 보험회사, 신용카드회사 등 세계 최대의 소매조직을 갖고 있다. 시어즈는 이들 기업의 고객정보를 통합하여 데이터베이스로 구축하고 이를 교차판매에 적극 활용하고 있다.

예를 들어, 시어즈 소속의 부동산회사를 통해 주택을 구매한 고객은 시어즈 백화점에서 가전제품 등의 가정용품을 구매할 때 최대 25%의 할인혜택을 받을 수 있으며 시어즈 소속 은행을

통해 대출서비스도 제공받는다. 시어즈는 디스커버 카드를 소지한 고객에게 시어즈 백화점의 자동차 부품이나 서비스의 할인쿠폰을 우송하는 등 교차판매에 많은 관심을 갖고 있다.

교차판매는 일반 소매점 혹은 인터넷 온라인 쇼핑몰에서 거래 데이터를 기반으로 장바구니 분석을 통해 상품을 추천하기 위해 사용되거나, 혹은 은행이나 보험업에서 고객들에게 다양한 금융 상품을 추천하기 위해 사용되고 있다. 그러나 데이터마이닝 혹은 교차판매가 각 상품의 수익기여 가치를 평가하기 위해 어떻게 사용될 수 있는지에 대한 답은 제공하지 못하고 있는 실정이다.

각 상품이 기업의 수익에 기여하는 가치를 평가하기 위해서는 자체 판매량뿐만 아니라 교차판매로 인해 발생하는 수익을 함께 고려해야 한다. 그러나 상품간의 교차판매효과를 평가하는 방법과 교차판매효과가 기업의 수익에 기여하는 메커니즘에 대한 연구가 미진한 상태이다. 따라서 본 논문에서는 기존의 연관규칙이 갖는 문제점을 예제를 이용하여 기술한 후, 황인수[2004]가 제안한 교차판매척도를 소개하고, 수정된 알고리즘의 유의성을 측정하기 위한 인터넷 쇼핑몰 및 오프라인 상점에서의 실험결과를 기술한다.

본 논문의 구성은 다음과 같다. 제II장에서는 연관규칙의 문제점과 교차판매효과의 측정을 위해 새로운 척도에 대해 기술하며, 제III장에서는 교차판매효과를 이용한 각 상품의 판매량 계산 알고리즘을 기술한다. 다음으로, 제IV장에서는 인터넷 쇼핑몰과 오프라인 상점의 거래데이터를 이용한 각 상품의 수익가치평가 결과를 기술하며, 제V장에서는 본 연구의 결과와 향후 연구방향을 기술한다.

## II. 연관규칙과 교차판매효과

### 2.1 연관규칙에 대한 고찰

연관규칙은 Agrawal *et al.*[1993]이 제안한 대

표적인 데이터마이닝 기법의 하나로서, 장바구니 분석을 통한 상품추천 등에 광범위하게 사용되고 있다. 연관분석은 거래내역을 분석하여 각 거래에 동시에 포함되는 상품들을 연관규칙으로 표현하는 것으로서, 연관규칙에 관한 기존의 연구를 정리하면 다음과 같다.

정의 1)  $k$ 개의 항목으로 구성된 집합을  $I = \{i_1, i_2, \dots, i_k\}$ 라고 하고,  $I$ 로부터 임의로  $n$ 개의 항목을 선택하여 구성한 집합을  $B = \{b_1, b_2, \dots, b_n\}$ 라고 하면,  $B$ 의 부분집합은  $b_i \in I$ 가 되며, 이를 장바구니라고 한다.

정의 2)  $i_1 \Rightarrow i_2$ 의 연관규칙이 존재하기 위해서는 다음의 조건을 만족해야 한다.

- ①  $n$ 개의 장바구니 중에서 항목  $i_1$ 과  $i_2$ 를 모두 포함하는 장바구니가 최소한  $\text{min-supp}(\%)$  이상 존재해야 한다.
- ② 항목  $i_1$ 을 포함하는 장바구니의  $\text{minconf}(\%)$  이상이 항목  $i_2$ 를 포함하고 있어야 한다.

위의 정의 2)에서 항목 ①은 지지도(support)라고 하며, 항목 ②는 신뢰도(confidence)라고 한다. 지지도는 하나의 상품 혹은 일련의 상품이 전체거래에서 차지하는 비율로서  $\text{supp}(i_1)$  혹은  $\text{supp}(i_1, i_2)$  등으로 표현한다. 신뢰도는 상품  $i_1$ 를 포함하는 거래에 상품  $i_2$ 가 동시에 포함되는 조건부 확률로서  $\text{supp}(i_1, i_2)/\text{supp}(i_1)$ 를 의미하는  $\text{conf}(i_1, i_2)$ 로 표현한다.

정의 3) 상품  $i_1$ 과  $i_2$ 간의 향상도 혹은 관심도(improvement)는 다음과 같이 계산한다[Silverstein et al., 1998].

$$I(i_1, i_2) = \frac{\text{supp}(i_1, i_2)}{\text{supp}(i_1)\text{supp}(i_2)} = \frac{p(i_1 i_2)}{p(i_1)p(i_2)}$$

향상도는 상품  $i_2$ 가 상품  $i_1$ 과 함께 구매된 거래와 상품  $i_2$ 가 상품  $i_1$ 에 관계없이 단독으로 구매된 거래의 비율이다. 따라서 향상도가 1이라는 것은 상품  $i_2$ 가 상품  $i_1$ 에 영향을 받아서 구입되는 비율과 상품  $i_2$ 가 상품  $i_1$ 과 관계없이 단독으로 구입되는 비율이 같음을 의미하기 때문에 이들 두 상품은 독립상품이다. 그러나 향상도가 1보다 크면 교차구매의 효과를 있음을 의미하기 때문에 보완재의 성격을 가지며, 향상도가 1보다 작으면 상품  $i_1$ 의 구매가 상품  $i_2$ 의 구매를 오히려 감소시키기 때문에 대체재의 성격을 갖는다[Brijs et al., 1999].

Ahmed et al.[2000]은 향상도를 변형하여 상품  $i_1$ 에 대한 상품  $i_2$ 의 신뢰도가 상품  $i_2$ 가 독립적으로 구입되는 비율보다 높을 때 정(+의) 상관관계를 갖는다고 주장하였다.

정의 4) 다음의 조건을 만족하면 상품  $i_1$ 과 상품  $i_2$ 는 정(+의) 상관관계를 갖는다.

$$\frac{p(i_1 i_2)}{p(i_1)} - p(i_2) \geq 0$$

## 2.2 연관규칙의 문제점 분석

연관규칙의 신뢰도가 갖는 문제점을 제시하기 위해 Silverstein et al.[1998]이 사용한 예제를 소개하면 <표 1>과 같다. 이 예제는 다양한 상품의 거래 데이터로부터 커피와 홍차에 관한 데이터만을 추출하여 비율로 표현한 것으로서,  $x$ 는 장바구니에 해당 상품이 포함되어 있음을 의미하며  $\bar{x}$ 는 포함되어 있지 않음을 의미한다.

<표 1> 상품 거래데이터의 예(I)

구분	홍차	커피	합 계
커피	0.20	0.70	0.90
$\bar{\text{커피}}$	0.05	0.05	0.10
합계	0.25	0.75	1.00

위의 예에서 (홍차 ⇒ 커피)의 연관규칙을 분석하면 다음과 같다. 장바구니에서 홍차와 커피를 모두 포함하는 비율을 나타내는 지지도는 20%이며, 홍차를 포함하는 장바구니에 커피가 함께 포함되어 있는 비율을 나타내는 신뢰도는 다음과 같은 계산에 따라 80%가 된다. 이는 매우 높은 수준의 신뢰도로서, 홍차 1개를 판매하면 커피 0.8개가 함께 판매됨을 보여준다.

$$\begin{aligned} \text{신뢰도(홍차, 커피)} &= \frac{p(\text{홍차} \wedge \text{커피})}{p(\text{홍차})} \\ &= \frac{0.20}{0.25} = 0.80 \end{aligned}$$

그러나 위의 예에서 커피가 홍차와 관계없이 판매된 비율이 90%에 이른다는 사실을 주목할 필요가 있다. 즉, 홍차의 판매가 커피의 판매에 미친 영향은 80%이지만, 커피가 홍차와 관계없이 판매된 비율은 이보다 높은 90%로서, 결과적으로 홍차의 판매는 커피의 판매에 오히려 부정적인(-) 영향을 미쳤다는 것이다. 이를 향상도의 관점에서 분석하면 다음과 같다.

$$\begin{aligned} \text{향상도(홍차, 커피)} &= \frac{p(\text{홍차} \wedge \text{커피})}{p(\text{홍차})p(\text{커피})} \\ &= \frac{0.20}{0.25 \times 0.90} = 0.89 \end{aligned}$$

$$\begin{aligned} \text{향상도(홍차, } \overline{\text{커피}}) &= \frac{p(\text{홍차} \wedge \overline{\text{커피}})}{p(\text{홍차})p(\overline{\text{커피}})} \\ &= \frac{0.05}{0.25 \times 0.10} = 2.00 \end{aligned}$$

이 결과를 살펴보면, 홍차와 커피간의 향상도는 0.89로서 1보다 작으며, 홍차를 구입하고 커피를 구입하지 않는 거래의 향상도는 2.00으로서 1보다 크다. 즉, 앞의 신뢰도를 이용한 분석에서는 홍차를 구입하는 사람들의 상당수가 커피를 구입하는 것으로 나타났으나, 향상도를 이용한 분석에서는 홍차는 커피와 대체재의 관계로서 홍차의 판매가 커피판매에 부정적인 영향

을 미치는 것으로 나타났다. 따라서 Silverstein et al.[1998]의 연구에 따르면, 신뢰도는 결과를 왜곡시킬 수 있기 때문에 본 연구에서 목표로 하는 교차판매의 효과를 측정하기 위해 신뢰도를 사용하는 것은 부적합하다는 결론에 이르게 된다.

위의 예에서는 커피와 홍차를 모두 포함하지 않는 장바구니의 비율이 5%이며, 커피가 판매된 거래의 비율이 90%로 가정하고 있는데, 이는 판매행태나 상품의 구색에 따라 달라지는 문제가 있다.

예를 들어, 앞의 상점에서 커피와 홍차 외에 인삼차, 녹차, 머그잔 등을 판매하고 있으며 이들의 판매가 전체 거래의 90.5%를 차지하였다면,  $p(\overline{\text{커피}}, \overline{\text{홍차}})$ 는 0.905가 되며, 나머지는 앞의 예에서와 동일한 비율을 갖는다.

<표 2> 상품 거래데이터의 예(II)

구분	홍차	커피	합계
커피	0.020	0.070	0.090
홍차	0.005	0.905	0.910
합계	0.025	0.975	1.000

<표 2>는 앞의 <표 1>에서 제시한 커피와 홍차의 전체 판매량에는 변함이 없으나, 다른 상품의 판매를 고려하여 데이터를 구성했기 때문에 커피와 홍차를 포함하는 장바구니의 비율이 원래의 1/10 수준으로 감소되어 있음을 볼 수 있다. 이 예에서 커피와 홍차간의 향상도를 계산하면 다음과 같다.

$$\begin{aligned} \text{향상도(홍차, 커피)} &= \frac{p(\text{홍차} \wedge \text{커피})}{p(\text{홍차})p(\text{커피})} \\ &= \frac{0.020}{0.025 \times 0.090} = 8.89 \end{aligned}$$

$$\begin{aligned} \text{향상도(홍차, } \overline{\text{커피}}) &= \frac{p(\text{홍차} \wedge \overline{\text{커피}})}{p(\text{홍차})p(\overline{\text{커피}})} \\ &= \frac{0.005}{0.025 \times 0.910} = 0.22 \end{aligned}$$

이 결과는 <표 1>의 예를 이용한 분석과 상이한 결과로서 홍차와 커피를 구성하는 장바구니의 구성에는 변화가 없음에도 불구하고, 분석과정에 다른 상품을 추가함에 따라 홍차와 커피가 상당히 높은 상관관계를 갖는 보완재로 성격이 바뀌었다. 이것은 홍차가 커피에 미치는 신뢰도 비율을 다시 전체 거래에서 커피가 차지하는 비율로 나누었기 때문에 필연적으로 발생하게 되는 문제이다. 결과적으로 향상도는 각 상품이 독립적으로 판매되는 비율을 어떻게 설정하느냐에 따라 그 값이 달라지기 때문에 연관규칙을 평가하는 기준으로 부적합한 것으로 판단된다.

### 2.3 교차판매효과 측정척도

황인수[2004]는 상품 x와 y간 교차판매효과를 측정하기 위해 신뢰도에 각 상품의 판매량을 가중치로 적용하는 새로운 교차판매척도인 CSE (Cross-Selling Effect)를 제안하였다.

$$CSE(x, y) = \frac{p(xy)}{p(x)} \times \frac{p(y)}{p(x \cup y)} \leq 1$$

이 척도를 앞의 <표 1>과 <표 2>의 예제에 적용하면 CSE(커피, 홍차)=5.85%, CSE(홍차, 커피)=75.80%의 교차판매효과를 갖는 것으로 분석된다. 이러한 결과는 신뢰도를 사용했을 때 Conf(커피, 홍차)=22.2%, Conf(홍차, 커피)=80.0%와는 많은 차이를 갖는 것으로서, 홍차의 판매가 커피의 판매에 미치는 영향은 소폭으로 감소하였으나, 커피의 판매가 홍차의 판매에 미치는 영향은 크게 감소함을 알 수 있다.

## III. 교차판매량 계산

### 3.1 교차판매 알고리즘

각 상품의 수익기여가치를 평가하기 위한 교차판매량의 계산에 교차판매척도를 그대로 적

용할 경우, 전체 판매량이 과대 계상되는 문제와 함께 교차판매의 연쇄효과를 고려하지 못하는 문제가 발생하게 된다. 이에 따라, 황인수[2004]는 각 상품의 판매량을 순수판매량과 교차판매량으로 구분하고, 교차판매된 상품이 다시 교차판매를 일으키는 효과를 고려하기 위해 <그림 1>과 같은 알고리즘을 제안하였다.

Let  $s_i$  is the amount of product sold,  
 $\rho$  is the cross reselling factor, and  
 $\epsilon$  is the allowed error level.

$$c_{ij} = \frac{p(ij)}{p(i)} \times \frac{p(j)}{p(i \cup j)}$$

$$q_{ij} = s_i \times c_{ij}$$

do {

$$\forall i=j, q_{ij} = \frac{q_{ij}}{\sum_k q_{ki}} \times s_i$$

$$\forall i \neq j, q_{ij} = (q_{ij} + \rho \sum_{k \neq i} (q_{kk} \times c_{ki})) \times c_{ij}$$

$$\Delta = \sum_i (s_i - \sum_k q_{ki})^2$$

} while( $\Delta > \epsilon$ )

<그림 1> 교차판매량 계산 알고리즘

여기서,  $s_i$ 는 각 상품의 총판매량,  $\rho$ 는 교차판매된 상품이 다시 교차판매를 일으키는 비율, 그리고  $\epsilon$ 는 계산을 종료하는 오차의 수준을 나타낸다. 상품  $i$ 의 순수판매량  $q_{ii}$ 는 상품  $i$ 의 총판매량과 다른 상품들로부터 교차판매척도  $c_{ki}$ 에 따라 계산된 교차판매량의 합을 정규화하여 계산된다. 상품  $i$ 로부터 상품  $j$ 로의 교차판매량  $q_{ij}$ 는 상품  $i$ 의 순수판매량  $q_{ii}$ 와 다른 상품들  $k$ 로부터 유발된 상품  $i$ 의 교차판매량  $\sum_{k \neq i} (q_{kk} \times c_{ki})$ 에  $\rho$ 를 곱한 값으로부터 계산된다.

위의 알고리즘에서  $\rho$ 를 0.1,  $\epsilon$ 를 0.0001로 설정한 후, {A, C}, {A}, {C}, {A, B, C}, {B}, {A}, {A, B}, {A, B}, {C}, {A}, {C, D} 등의 장바구니를 갖는 거래문제에 적용하여 교차판매량을 계산하면 <표 3>과 같다. 여기서, 대각선에 위치한 셀들은 각 상품의 순수판매량이며, 기타의

셀은 교차판매량이다.

<표 3> 순수판매량과 교차판매량

From \ To	A	B	C	D
A	4.239	0.965	0.643	0.000
B	1.860	2.713	0.443	0.000
C	0.901	0.322	3.031	0.129
D	0.000	0.000	0.883	0.871
합계	7.000	4.000	5.000	1.000

### 3.2 교차판매 알고리즘의 수정

본 연구에서는 위에서 제시한 알고리즘이 다음과 같은 두 가지의 문제점을 갖고 있음을 발견하였다.

첫째, 이 알고리즘에 따르면 상품 x로 인해 상품 y가 교차판매된 경우, 교차판매된 상품 y가 상품 x에 다시 영향을 미치는 불합리한 재귀효과가 발생한다. 둘째, 이 알고리즘은 대체재의 존재에 따른 교차판매효과 감소현상을 고려하지 못하고 있다. 이에 따라 본 연구에서는 재귀효과를 방지하고 대체재 효과를 반영할 수 있도록 알고리즘을 수정·보완하였다.

Let  $s_x$  is the quantity of sold for product  $x$ ,  
 $p_x$  is the possibility of competitive goods,  
 $\rho$  is the cross reselling factor, and  
 $\epsilon$  is the allowed error level

$$c_{xy} = \frac{p(xy)}{p(x)} \times \frac{p(y)}{p(x \cup y)}$$

$$q_{xy} = s_x \times c_{xy}$$

do {

$$\forall_{x=y}, q_{xy} = \frac{q_{xy}}{\sum_i q_{iy}} \times s_x$$

$$\forall_{x \neq y}, q_{xy} = (q_{xx} + \rho \sum_{i \neq x, i \neq y} (q_{ii} \times c_i)) \times c_{xy} \times (1 - p_x)$$

$$\Delta = \sum_x (s_x - \sum_i q_{ix})^2$$

} while ( $\Delta > \epsilon$ )

<그림 2> 수정된 교차판매량 계산 알고리즘

앞에서 제시한 동일한 예제에 대해 수정된 교차판매량 계산 알고리즘을 적용하여 순수판매량과 교차판매량을 계산하면 <표 4>와 같다. 여기서,  $\rho$ 와  $\epsilon$ 은 위에서와 동일하게 각각 0.1과 0.0001로 설정하였으며, 각 상품에 대한 대체재의 존재 가능성  $p_x$ 는 0.10으로 모두 동일한 것으로 가정하였다.

<표 4>에서 제1단계는 각 상품의 실제 판매량에 따라 교차판매되는 상품들의 교차판매량을 계산한 것으로서, 이러한 경우 각 상품의 합계가 실제 판매량을 초과한다. 이에 따라 제2단계에서는 대각행렬로 나타나는 각 상품의 순수판매량은 합계를 이용하여 정규화하고, 나머지의 교차판매량은 본 연구에서 제시하는 수식에 따라 계산한 것이다.

<표 4> 수정된 순수판매량과 교차판매량

A) 제 1 단계

From \ To	A	B	C	D
A	7.000	1.500	1.000	0.000
B	2.625	4.000	0.625	0.000
C	1.400	0.500	5.000	0.200
D	0.000	0.000	1.000	1.000
합계	11.025	6.000	7.625	1.200

B) 제 2 단계

From \ To	A	B	C	D
A	4.444	0.875	0.594	0.000
B	1.594	2.667	0.388	0.000
C	0.837	0.308	3.279	0.118
D	0.000	0.000	0.750	0.833
합계	6.875	3.850	5.011	0.951

C) 제 10 단계

From \ To	A	B	C	D
A	4.505	0.886	0.603	0.000
B	1.680	2.813	0.409	0.000
C	0.815	0.301	3.192	0.115
D	0.000	0.000	0.796	0.885
합계	7.000	4.000	5.000	1.000

<표 4>에서 제10단계는 위의 과정을 10회 반복한 결과로서, <표 1>의 결과와 비교할 때 교차판매량이 약간씩 감소되는데, 이는 대체재의 효과와 재귀효과를 방지한 결과이다.

38회, 구입하지 않겠다는 응답이 18회, 그리고 무응답이 24회였다. 여기서, 관련 상품을 구입하지 않겠다는 것은 이에 대한 대체품이 존재하지 않는 것으로 해석할 수 있다.

## IV. 데이터 수집 및 분석

### 4.1 인터넷 쇼핑몰의 거래분석

본 연구에서 제시한 교차판매척도의 타당성을 검증하기 위해서는 실존하는 상점에서 특정 상품을 제거하였을 때의 각 상품의 판매량 변화에 관한 데이터를 수집하거나, 혹은 상품을 구입하는 모든 사람들의 상품구매에 관한 설문조사를 실시해야 한다.

그러나 오프라인의 상점에서 특정상품을 제거한 판매 데이터를 수집하는 것은 현실적으로 많은 어려움이 존재하므로, 본 연구에서는 연구목적의 인터넷 쇼핑몰을 구축하여 거래데이터를 수집하였다. 웹 사이트는 톰캣(Tomcat) 웹서버에서 자바(Java)와 JSP 프로그래밍 언어로 개발하였으며, JDBC를 이용하여 MySQL 데이터베이스와 연동하였다.

<그림 3>은 본 연구의 대상이 되는 상품들의 목록을 보여주며, <그림 4>는 상품간의 교차판매영향 유무를 파악하기 위해 생성한 인터넷 설문문의 예를 보여주고 있다. 예를 들어, 유자차와 컵을 구매했다면, 유자차가 없을 경우에도 컵을 구매할 의도가 있는지, 그리고 이와는 반대로 컵이 없을 경우에도 유자차를 구매할 의도가 있는지를 파악함으로써 상품간의 교차판매정도를 파악한다. 여기서, “잘 모르겠음”을 선택한 경우에는 동일한 상품간의 관계에 대해 응답한 다른 응답자의 응답비율에 따라 계산한다.

<그림 4>의 마지막 열은 각 상품에 대한 소비자의 구매의도로부터 대체재의 존재비율을 추정하기 위한 설문이다. 유자차가 없을 경우에도 종이컵 등 관련 상품을 구입하겠다는 응답이



<그림 3> 인터넷상에 구축한 쇼핑몰의 예



<그림 4> 관계를 파악하기 위한 설문문의 예

다음으로, 관련 상품을 구매할 경우에는 유자차가 관련 상품의 구매에 영향을 미치지 않거나 혹은 유자차에 대한 대체상품이 존재함을 의미한다. 여기서, 관련 상품을 구입하겠다는 38회의 응답중에서 다른 상품을 구입하겠다고 응답은 19회에 달했다. 따라서 총 56회의 유효한 응답중에서 19회의 응답이 유자차에 대한 대체상품을 구입할 의사를 갖고 있으므로 유자차에 대한 대체재의 존재율은 33.9%로 추정할 수 있다. <그림 3>의 각 상품에 대해 차례대로 대체재 존

재율을 계산하면 23.6%, 33.9%, 35.4%, 40.5%, 41.8%, 36.4%, 47.6%, 24.4%, 25.0%, 그리고 19.7%가 된다. 여기서, 종이컵의 대체품이 존재하지 않음에도 불구하고, 19.7%를 나타낸 것은 종이컵이 불필요한 상품으로의 전환구매를 반영한 설문결과로 해석된다.

#### 4.1.1 상품의 교차판매량

본 연구에서는 인터넷 쇼핑몰에서 발생한 총 223건의 거래에 대해 분석을 실시하였으며, 이로부터 상품간의 교차판매량을 계산하면 <그림 5>와 같다. 여기서, 교차판매된 상품이 다시 교차판매를 일으키는 비율은 수차례에 걸친 시뮬레이션을 통해 1/3로 설정하였다.

구분	상품	N(x)	IDQ(x)	Sub-Product	N(y)	CSE	CSQ(y)	SVQ(y)
1	CF	38	20.7	UT	59	0.102	2.055	0.0
				RT	43	0.059	1.201	0.0
				LE	43	0.029	0.590	2.0
				VT	67	0.070	1.431	0.0
				CM	29	0.012	0.239	0.5
				GT	61	0.016	0.339	0.0
				SG	33	0.233	4.505	3.0
				PM	34	0.277	5.278	14.2
				CP	49	0.149	2.993	3.4
				CF	38	0.027	0.527	3.4
2	UT	59	35.3	CF	38	0.042	1.249	1.5
				RT	43	0.126	3.582	1.5
				LE	43	0.038	1.106	0.0
				VT	67	0.120	3.465	2.2
				CM	29	0.017	0.512	3.0
				GT	61	0.194	5.536	3.5
				SG	33	0.032	0.949	1.7
				PM	34	0.026	0.766	2.7
				CP	49	0.094	2.736	6.0
				UT	59	0.194	5.536	3.5

<그림 5> 상품의 순수판매량과 교차판매량

<그림 5>에서  $N(x)$ 는 상품  $x$ 를 포함하고 있는 장바구니의 개수,  $IDQ(x)$ 는 상품  $x$ 의 순수판매량,  $N(x \wedge y)$ 는 상품  $x$ 와  $y$ 를 모두 포함하고 있는 장바구니의 개수,  $N(x \vee y)$ 는 상품  $x$  또는  $y$ 를 포함하고 있는 장바구니의 개수,  $CSE(x,y)$ 는 상품  $x$ 가 상품  $y$ 의 교차판매에 미치는 영향의 정도,  $CSQ(y)$ 는 상품  $x$ 로 인한 상품  $y$ 의 교차판매량, 그리고  $SVQ(y)$ 는 상품  $x$ 가 없을 경우 구매를 포기하는 상품  $y$ 의 개수로서 설문조사를 통해 결정된 것이다. 주요 상품간의 교차판매량

을 정리하면 <표 5>와 같다.

<표 5> 주요 상품간의 교차판매량

x	N(x)	IDQ(x)	y	N(y)	CSE	CSQ(y)	SVQ(y)
CF	38	20.7	SG	33	0.233	4.505	9.0
			PM	34	0.277	5.278	14.2
			CP	49	0.149	2.993	3.4
CM	29	22.8	LB	43	0.064	1.028	2.0
			CP	49	0.167	2.629	5.2
UT	59	35.3	VT	67	0.120	3.465	2.2
			GT	61	0.194	5.536	3.5
GT	61	35.4	VT	67	0.103	2.423	3.7
			CP	49	0.117	2.723	4.7
RT	43	26.4	LB	43	0.178	3.657	4.3
			GT	61	0.168	3.526	4.9
CP	49	27.7	CF	38	0.089	2.549	4.0
			SG	33	0.073	2.079	4.0
SG	33	17.5	PM	34	0.350	5.633	11.8
			CF	38	0.308	5.070	5.4
PM	34	18.9	SG	33	0.330	5.543	9.1
			CF	38	0.345	5.839	6.6

여기서, 커피(CF)는 설탕(SG), 프림(PM), 그리고 종이컵(CP) 등의 판매에 영향을 미치고 있으며, 커피가 없을 경우 이들의 판매량은 각각 4,505, 5,278, 2,993개 감소되는 것으로 계산되었다. 따라서 커피가 기업의 수익에 기여하는 가치는 커피 자체의 판매량뿐만 아니라 이로 인한 교차판매량을 함께 고려해야 한다.

각 상품의 수익기여가치를 평가하기 위해 본 연구에서 사용한 인터넷 설문결과에 따르면 커피가 없을 경우 설탕과 프림의 판매량이 각각 9.0개와 14.2개 감소하는 것으로 나타났다. 이는 본 연구의 계산 결과와는 약간의 차이를 보이고 있지만, 다른 상품들의 경우에는 차이가 크지 않으며 전체적인 방향성도 일치하고 있다.

IDQ(x)는 상품  $x$ 가 교차판매의 영향을 받지 않고 순수하게 판매된 수량으로서, 전체판매량



에서 차지하는 비율이 낮을수록 단독으로 판매 되기보다는 다른 상품과 함께 판매될 가능성이 높은 것으로 해석된다. 즉, 커피믹스(CM)는 총 29개의 판매량 중에서 약 78.6%인 22.8개가 순수판매량으로 계산된 반면, 설탕(SG)은 총 33개의 판매량 중에서 약 53.0%인 17.5개가 순수판매량으로 계산되었다. 따라서 커피믹스는 다른 상품과 관계없이 개별적으로 판매되는 것이 일반적이지만, 설탕은 교차판매의 정도가 상대적으로 높은 상품임을 알 수 있다.

#### 4.1.2 상품의 가치평가

각 상품이 순수판매와 교차판매를 통해 기업의 수익창출에 기여한 정도는 해당 상품이 없었을 경우에 발생하는 손실과 동일한 의미를 갖는다. 따라서 본 연구에서는 각 상품이 제거되었을 경우 수익의 감소를 개별상품의 판매량 감소, 신뢰도에 따른 관련 상품의 판매량 감소, 그리고 본 연구에서 제시한 CSE에 근거한 판매량 감소의 세 가지에 대한 분석을 실시하였다. 이를 표로 정리하면 <표 6>과 같다. 여기서, 각 상품의 단위당 판매수익은 <그림 3>에 나타나 있는 각 상품가격의 10%로 가정하였다.

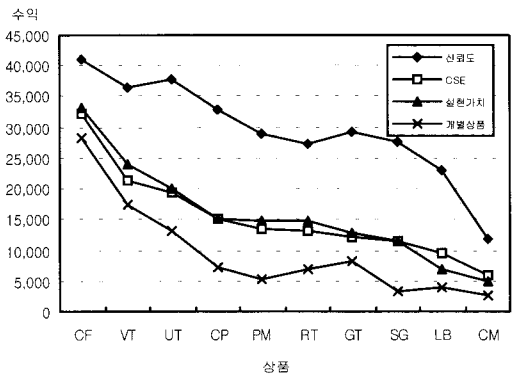
개별상품을 기준으로 가치를 평가하는 것은 교차판매량을 고려하지 않고 대체재 효과만을 고려할 경우 해당 상품의 판매가 감소되는 정도를 계산하는 전통적인 방법으로서 상품가치를 평가하는 하한값(lower bound)이다. 신뢰도를 기준으로 할 경우, 해당상품이 포함되어있는 장바구니의 구입을 모두 포기하는 것으로서, 상품가치를 평가하는 상한값(upper bound)이 된다.

<표 6>에서 실현가치는 설문을 통해 조사한 각 상품의 실현가치로서 개별상품기준과 신뢰도기준 내에 포함되어 있음을 알 수 있다. 또한, CSE기준은 상품에 따라 실현가치와 다소간의 차이를 나타내고 있으나 상당히 근접하게 평가하는 것으로 나타났다.

<표 6> 각 상품의 수익기여 가치

상품	수익가치 예측기준			실현가치
	개별상품	신뢰도	CSE	
CF	28,161.0	40,906.0	32,276.4	33,326.0
VT	17,547.3	36,332.3	21,425.7	23,876.1
UT	13,259.7	37,724.7	19,354.3	19,972.4
CP	7,082.5	32,777.5	15,125.6	15,095.2
PM	5,355.0	28,785.0	13,372.1	14,830.7
RT	6,944.5	27,384.5	13,050.2	14,697.8
GT	8,150.8	29,160.8	12,287.4	12,713.0
SG	3,368.0	27,728.0	11,610.6	11,583.4
LB	3,837.8	22,877.8	9,524.3	7,014.3
CM	2,489.9	11,889.9	6,067.5	4,787.9
합계	96,196.5	295,566.5	154,094.1	157,896.8

상품의 실현가치와 비교할 때, 개별상품을 기준으로 예측한 각 상품의 가치는 평균적으로 44.5%, 그리고 신뢰도 기준의 경우에는 110.5%의 높은 오차를 나타냈으나, CSE기준은 상대적으로 낮은 10.4%의 오차를 보였다. 또한 CSE로 예측한 수익가치의 합은 실현가치의 합으로부터 약 2.4%의 오차 내에 있음을 볼 수 있다.



<그림 6> 상품의 평가가치 비교(1)

<그림 6>은 실현가치가 큰 상품으로부터 실현가치가 작은 상품으로 정렬한 후, 수익가치를 예측하는 기준에 따른 예측치를 그래프로 나타낸 것이다.

이 그림으로부터, 본 연구에서 제시한 CSE에 따른 상품의 가치평가가 실현가치에 매우 근접함을 알 수 있다. 또한, 개별상품기준 및 신뢰도 기준과는 달리 상품의 중요도 순서가 실현가치와 정확히 일치하고 있기 때문에, 상품을 선택하는 문제에도 활용될 수 있을 것으로 기대된다.

#### 4.1.3 기존의 알고리즘과의 성과비교

본 연구에서 제시한 재귀효과의 방지와 대체재에 대한 고려가 상품의 수익가치 평가에 어떻게 영향을 주는지를 평가하기 위해 두 가지 요인을 이용한 네 가지의 실험을 실시하였다.

<표 7> 수정 알고리즘의 성과 측정(1)

		대체재	
		고려 안함	고려함
재귀효과	허용	14.5	11.0
	불허	12.2	10.4

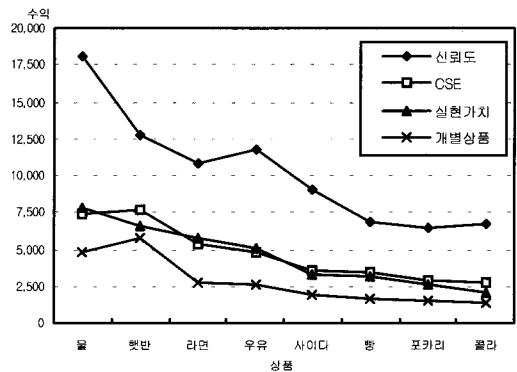
기존의 알고리즘에서는 상품간 교차판매의 재귀효과를 허용하며 대체재를 고려하지 않고 있는데, 이때의 실현가치에 대한 오차는 14.5%이다. 여기서, 재귀효과를 방지하거나 혹은 대체재를 고려할 경우에는 오차가 각각 12.2%와 11.0%로 감소되며, 이들 모두를 고려할 경우에는 10.4%로 낮아지는 것으로 나타났다. 결과적으로 본 연구에서 제시한 수정알고리즘은 상품의 수익가치의 평가에 있어서 기존의 알고리즘보다 높은 정확도를 갖는 것으로 나타났다.

#### 4.2 오프라인 상점의 거래분석

앞에서는 가상적으로 구축한 인터넷 쇼핑몰을 이용한 분석을 실시하였으나, 이를 검증하기 위해 오프라인 상점에 대한 분석을 추가적으로 실시하였다. 그러나 기존의 거래데이터만으로는 각 상품간의 교차판매효과를 검증할 수 없기 때

문에, 인터넷 쇼핑몰에서와 동일한 방법으로 상품을 구매한 사람들에게 대해 설문을 실시하였다.

데이터의 수집대상은 M 편의점을 대상으로 하였으며, 비교적 구매빈도가 높은 햇반, 라면, 빵, 물, 콜라, 사이다, 포카리스웨트, 우유 등 8가지의 상품을 포함하는 318건의 거래에 대해 데이터를 수집하여 분석하였다.



<그림 7> 상품의 평가가치 비교(2)

<그림 7>은 인터넷 쇼핑몰에 대한 분석방법을 오프라인 상점에 적용한 결과로서, 앞에서와 동일하게 실현가치에 근접하는 결과를 도출하고 있다. 또한, 재귀효과와 대체재 요인에 따른 분석을 실시한 결과, <표 8>에서 보는 바와 같이 본 연구에서 제시한 수정알고리즘이 의미있는 결과를 도출하였다.

<표 8> 수정 알고리즘의 성과 측정(2)

		대체재	
		고려 안함	고려함
재귀효과	허용	19.9	18.8
	불허	13.5	13.1

## V. 결 론

인터넷 등 정보통신기술이 급격히 발달함에 따라 고객 및 거래에 대한 방대한 규모의 데이터베이스가 구축되고 있으며, 이들로부터 정보

및 지식을 추출하여 경영의사결정에 활용하는 데이터마이닝은 많은 관심의 대상이 되고 있다. 데이터마이닝의 여러 가지 기법 중에서 연관규칙은 거래 데이터로부터 상품간의 관계를 규명하는 것으로서, 상품추천이나 상품진열 등에서 광범위하게 사용되고 있다.

그러나 오프라인의 소매상이 수익을 극대화하기 위한 최적의 마케팅 믹스를 구성하기 위해 데이터마이닝 혹은 연관규칙이 어떻게 활용될 수 있는지에 대한 답은 제시하지 못하고 있다. 즉, 연관분석을 통해 규명된 상품간의 관계가 전체 상품의 판매에 어떠한 파급효과를 제공하며, 각 상품이 전체 수익에 어느 정도의 기여를 하고 있는지에 대한 정확한 분석이 이루어지지 못하고 있다.

이에 따라 본 연구에서는 인터넷 쇼핑몰과 오프라인 상점의 거래데이터를 이용하여 교차판매효과 척도의 타당성을 검증하고, 현실의 문제에서의 활용 가능성을 제시했다. 특히, 기존의 교차판매량계산 알고리즘을 수정하여 재귀효과를 방지하고 각 상품에 대한 대체재의 존재비율을 고려함으로써 상품 가치평가의 정확도를 향상시킬 수 있었다.

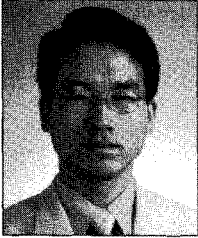
본 연구에서는 두 가지 상품간의 관계만을 고려하였다는 점과 보다 대규모의 데이터를 이용한 검증이 이루어지지 못했다는 한계점을 갖고 있다. 따라서 향후의 연구에서는 두 개 이상의 상품이 다른 상품에 미치는 영향을 고려하는 방안과 대규모의 데이터를 이용한 알고리즘의 검증을 연구하고자 한다.

## 〈참 고 문 헌〉

- [1] 김민주, *마케팅어드벤처*, 미래의창, 2005.
- [2] 박찬욱, *데이터베이스 마케팅*, 연암사, 1996.
- [3] 황인수, "연관규칙을 이용한 상품선택과 기대수익 예측," *경영정보학연구*, 제14권 4호, 2004. 12, pp. 87-97.
- [4] Agrawal, R. and Srikant, R., "Fast Algorithm for Mining Association Rules," *Proceedings of the 20th International Conference on Very Large Databases*, 1994, pp. 487-499.
- [5] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A., "Fast Discovery of Association Rules," *In Advances in Knowledge Discovery and Data Mining*, AAAI Press, 1996, pp. 307-328.
- [6] Agrawal, R., Imielinski T., and Swami, A., "Mining Association Rules Between Sets of Items in Large Databases," *Proceedings of ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
- [7] Ahmed, K., El-Makky, N., and Taha, Y., "A note on Beyond Market Baskets: Generalizing Association Rules to Correlations," *SIGKDD Explorations*, Vol. 1, No. 2, 2000, pp. 46-48.
- [8] Brijs, T., Goethals, B., Swinnen, G., Vanhoof, K., and Wets, G., "A Data Mining Framework for Optimal Product Selection in Retail Supermarket Data: The Generalized PROFSET Model," *Proceedings on ACM KDD-2000*, Boston, MA, 2000, pp. 300-304.
- [9] Brijs, T., Swinnen, G., Vanhoof, K., and Wets, G., "Using Association Rules for Product Assortment Decisions: A Case Study," *Proceedings on KDD-99*, ACM, San Diego, CA, USA, 1999, pp. 254-260.
- [10] Brin, S., Motwani, R., Ullman, J., and Tsur, S., "Dynamic Itemset Counting and Implication Rules for Market Basket Data," *Proceedings ACM SIGMOD International*

- Conference on Management of Data*, 1997, pp. 255-264.
- [11] Deshpande, M. and Karypis, G., "Item-Based Top-N Recommendation Algorithms," *ACM Transaction on Information Systems*, Vol. 22, No. 1, 2004, pp. 143-177.
- [12] Ewald, R., Keynote Address, *The 3rd International Conference on Information and Knowledge Management*, 1994.
- [13] Herlocker, J., Konstan, J., Terveen, L., and Riedl, J., "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, Vol. 22, No. 1, 2004, pp. 5-53.
- [14] Kalakota, R. and Robinson, M., *e-business: Roadmap for Success*, Addison Wesley, 1999.
- [15] Kits, B., Freed, D., and Vrieze, M., "Cross-sell: A Fast Promotion-tuneable Customer-Item Recommendation Method based on Conditional Independent Probabilities," *Proceedings of ACM SIGKDD International Conference*, New York, 2000, pp. 437-446.
- [16] Kleinberg, J., "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM*, Vol. 46, No. 5, September 1999, pp. 604-632.
- [17] Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordron, L., and Riedl, J., "GroupLens: Applying Collaborative Filtering to Usenet News," *Communications of ACM*, Vol. 40, No. 3, 1997, pp. 77-87.
- [18] Schafer, J., Konstan, J., and Riedl, J., "Recommender Systems in e-Commerce," *Proceedings of ACM E-Commerce*, New York, 1999.
- [19] Shardanand, U. and Maes, P., "Social Information Filtering: Algorithms for Automating, Word of Mouth," *Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems*, New York, 1995, pp. 210-217.
- [20] Silverstein, C., Brin, S., and Motwani, R., "Beyond Market Basket: Generalizing Association Rules to Dependence Rules," *Data Mining and Knowledge Discovery*, Vol. 2, 1998, pp. 38-68.
- [21] Wang, K. and Thomas, S.M., "Item Selection By Hub-Authority Profit Ranking," *Proceedings on ACM SIGKDD*, Edmonton, 2002, pp. 652-657.
- [22] Xue, G., Zeng, H., Chen, Z., Ma, W., Zhang, H., and Lu, C., "Implicit Link Analysis for Small Web Search," *Proceedings on ACM SIGIR'03*, Jul 28-Aug 1, 2003, Toronto, pp. 56-63.
- [23] Zaki, M., Partasarathy, S., Ogihara, M., and Li, M., "New Algorithm for Fast Discovery of Association Rules," *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, 1997, pp. 283-286.

◆ 저자소개 ◆



황인수 (Hwang, Insoo)

현재 전주대학교 정보기술공학부 정보시스템 전공의 부교수로 재직하고 있다. 고려대학교 경영학과를 졸업하고 동 대학원에서 경영정보시스템을 전공하여 석사 및 박사학위를 취득하였으며, 산업연구원(KIET) 물류·유통연구센터의 연구원을 역임하였다. 주요 관심분야는 e-Business, CRM, 데이터마케팅, 웹 에이전트 등이다.

◆ 이 논문은 2005년 5월 2일 접수하여 1차 수정을 거쳐 2005년 8월 30일 게재확정되었습니다.