

연관규칙을 이용한 상품선택과 기대수익 예측

황인수*

Item Selection By Estimated Profit Ranking Based on Association Rule

Insoo Hwang

One of the most fundamental problems in business is ranking items with respect to profit based on historical transactions. The difficulty is that the profit of one item comes from its influence on the sales of other items as well as its own sales, and that there is no well-developed algorithm for estimating overall profit of selected items.

In this paper, we developed a product network based on association rule and an algorithm for profit estimation and item selection using the estimated profit ranking(EPR). As a result of computer simulation, the suggested algorithm outperforms the individual approach and the hub-authority profit ranking algorithm.

Keywords : Item Selection, Data Mining, Association Rule, Cross Selling, Ranking

* 전주대학교 정보기술공학부 부교수

I. 서 론

도매상으로부터 상품을 구입하여 소비자에게 판매하는 소매상은 상품이 판매되면 판매된 수량만큼을 재주문하며, 상품이 판매되지 않으면 이를 폐기 혹은 반품하고 새로운 품목으로 대체하는데, Blischok[1995]는 이를 상품지향경영(Product Oriented Business)이라고 하였다. 그러나 최근에는 기업간의 경쟁이 심화되고 고객의 욕구가 다양화됨에 따라, 고객들의 성향이나 과거의 구매행태에 근거하여 경영활동을 수행하는 고객지향경영(Customer Oriented Business)이 불가피하게 되었다. Kalakota and Robinson[1999]은 이와 같이 고객을 중심으로 이루어지는 마케팅, 세일즈, 서비스 전략 등 일련의 과정을 고객관계관리(Customer Relationship Management, CRM)로 정의하였다.

최근에는 인터넷 등 정보통신기술이 급격히 발달함에 따라 고객 및 거래에 대한 방대한 규모의 데이터베이스가 구축되고 있으며, 이들로부터 정보 및 지식을 추출하여 경영의사결정에 활용하는 데이터마이닝(Data Mining, DM)은 많은 관심의 대상이 되고 있다. 연관규칙(Association Rule)은 거래 데이터로부터 상품간의 관계를 규명하여 상품추천(Product Recommendation)이나 상품진열(Display) 등을 지원하는 데이터마이닝 기법의 하나로서, 관련연구로는 Agrawal et al.[1993, 1994, 1996], Brin et al.[1997], 그리고 Zaki et al.[1997] 등을 들 수 있다. 여기서 상품추천은 장바구니분석(Basket Analysis)을 통해 고객이 관심을 가질 것으로 예상되는 상품을 제안하는 것으로서 Shardanand and Maes[1995], Konstan et al.[1997], Schafer et al.[1999], Kitts et al.[2000], Deshpande and Karypis[2004], 그리고 Herlocker et al.[2004] 등에서 지속적으로 연구되고 있다.

이와 같은 연구에도 불구하고, 오프라인의 소매상이 최적의 마케팅 믹스(Marketing Mix)를

구성하기 위해 데이터마이닝 혹은 연관규칙을 어떻게 활용할 수 있는지에 대한 답은 제시하지 못하고 있다. 즉, 연관분석을 통해 규명된 상품간의 관계가 전체 상품의 판매에 어떠한 파급효과를 일으키며, 각 상품이 전체 수익에 어느 정도의 기여를 하고 있는지에 대한 정확한 분석이 이루어지지 못하고 있다. 이에 따라 연관규칙이 상품의 가격결정이나 교체상품선택 등의 중요한 의사결정에는 사용되지 못하고 있는 실정이다.

교체상품을 선택하는 문제에서 전통적으로는 각 상품의 단위당 수익과 단위기간동안 판매량을 곱한 판매수익을 평가의 기준으로 사용해 왔다. 그러나 이 방법은 교차판매(Cross Selling)의 효과를 고려하지 못하기 때문에, 자체 판매수익은 작지만 다른 상품의 판매에 큰 영향을 미치는 상품이 교체상품으로 선택되어 전체 수익을 저하시키는 문제가 발생할 수 있다. 이에 따라, 상품선택에 연관규칙을 이용하는 방안에 대한 연구가 시작되었는데, 관련연구로는 Brijs et al.[1999, 2000]이 제안한 PROFSET과 Wang and Thomas[2002]가 제안한 HAP(Hub-Authority Profit Ranking) 등을 들 수 있다.

PROFSET은 개별상품의 수익보다는 교차판매에 따라 장바구니에 포함되는 전체 상품들의 수익극대화에 초점을 맞추고, 거래내역으로부터 Maximal Frequent Itemset을 식별하여 이를 구매의도로 사용하였다. 그러나 PROFSET은 상품간의 관계정도를 나타내는 연결강도(Strength)를 고려하지 못하였으며, Maximal Frequent Itemset은 거래에서 빈번히 발생하지 않기 때문에 이를 구매의도로 사용할 수 있는지에 대한 의문이 제기되었다. 또한, 선택된 상품간의 상대적인 우선순위에 대한 정보를 제시하지 못하는 단점이 있다[Wang and Thomas, 2002].

Wang and Thomas[2002]는 웹의 연결구조를 이용하여 각 웹페이지의 중요도를 평가하는 Kleinberg[1998]의 HITS(Hyperlink Induced Topic Search) 알고리즘을 변형하여 상품간의 연관

규칙에 따라 네트워크를 구성한 후, 각 상품의 Hub와 Authority값을 계산하여 각 상품의 중요도를 결정하는 HAP 알고리즘을 제안하였다. 이 알고리즘은 상품간의 연결강도를 고려할 뿐만 아니라, 각 상품이 전체 상품에 미치는 영향을 네트워크의 가중치 전파 방법에 따라 계산함으로써 교차판매의 효과를 잘 반영하고 있으며, 컴퓨터 시뮬레이션에서 PROFSET보다 나은 성과를 나타냈다. 그러나 HAP에서 사용한 HITS 알고리즘은 웹페이지를 Hub와 Authority로 구분함으로써 정보검색의 효과를 극대화하기 위한 것으로서, 상품간의 영향의 정도를 표현하기에는 부족함이 있는 것으로 판단된다.

본 논문은 Wang and Thomas[2002]가 제안한 연관규칙을 이용한 상품 네트워크에서 각 상품의 순수판매량과 교차판매량에 따라 상품을 선택한 후, 선택된 상품들의 기대수익을 예측하는 EPR(Estimated Profit Ranking) 알고리즘을 제안한다. 본 논문의 구성은 다음과 같다. 제II장에서는 연관규칙을 이용한 상품 네트워크의 구성과 교차판매량 계산 알고리즘에 대해 기술하며, 제III장에서는 상품선택 및 수익예측에 대해 기술한다. 다음으로, 제IV장에서는 알고리즘의 성과를 측정하기 위한 컴퓨터 시뮬레이션의 결과를 기술하며, 제V장에서는 본 연구의 결과를 요약하고 향후의 연구방향을 제시한다.

II. 연관규칙을 이용한 교차판매량 계산

2.1 연관규칙의 개념

연관규칙은 Agrawal et al.[1993]이 제안한 대표적인 데이터마이닝 기법의 하나로서, 장바구니 분석을 통한 상품추천이나 상품진열 등에 광범위하게 사용되고 있다. 연관분석은 거래내역을 분석하여 각 거래에 동시에 포함되는 상품들을 연관규칙으로 표현하는 것으로서, 지지도(support),

신뢰도(confidence), 그리고 향상도(lift, improvement) 등의 척도를 사용한다.

지지도는 각 상품 혹은 일련의 상품이 전체 거래에 포함되어 있는 비율로서 $Supp(X)$ 혹은 $Supp(X, Y)$ 등으로 표현하며, 신뢰도는 상품 X를 포함하는 거래에서 상품 Y를 포함하는 비율로서 $Supp(X, Y)/Supp(X)$ 를 의미하는 $Conf(X \Rightarrow Y)$ 로 표현한다. 또한, 향상도는 상품 X를 구매한 경우, 이 거래가 상품 Y를 포함하는 경우와 상품 Y가 상품 X에 관계없이 단독으로 구매된 경우의 비율로서 $Supp(X, Y)/(Supp(X) \cdot Supp(Y))$ 로 표현한다. 향상도가 1에 가까우면 상품간의 상관관계가 없으며, 1보다 크면 양의 상관관계가 있고, 1보다 작으면 음의 상관관계가 있음을 의미한다.

2.2 교차판매척도

신뢰도는 특정 상품이 다른 상품과 함께 동일한 장바구니에 포함되는 비율로서, 인과관계를 나타내는 영향도, 즉 교차판매의 척도로 사용하기에는 부적합한 점이 있다. 예를 들어, TV와 Video 등 가전제품을 판매하는 상점에서 {TV}, {TV}, {TV, Video}, {TV, Video}, {Video}의 거래가 발생했다면, TV와 Video간의 신뢰도는 다음과 같이 계산된다.

$$\begin{aligned} Conf(TV \Rightarrow Video) &= \frac{p(TV \wedge Video)}{p(TV)} \\ &= \frac{0.40}{0.80} = 0.50 \end{aligned}$$

$$\begin{aligned} Conf(Video \Rightarrow TV) &= \frac{p(Video \wedge TV)}{p(Video)} \\ &= \frac{0.40}{0.60} = 0.67 \end{aligned}$$

신뢰도를 교차판매효과로 이용할 경우, 장바구니를 구성하는 상품중 한 가지라도 존재하지 않으면 해당하는 장바구니의 구입을 모두 포기하는 것으로 가정한다. 즉, 위의 예에서 Video가

TV와 함께 판매된 $Conf(TV \Rightarrow Video)$ 는 0.50로서, 만일 TV가 없을 경우에는 0.50의 신뢰도에 따라 Video의 판매가 2대만큼 감소하는 것으로 계산된다. 이것은 TV를 포함하는 모든 장바구니의 구입을 포기하는 것으로서, TV와 Video를 함께 구입하던 고객이 TV가 없을 경우에는 Video만을 구입할 수도 있는 현실을 반영하지 못하여 불합리한 결과를 초래한다. 이에 따라 본 연구에서는 Hwang[2004]이 신뢰도를 확장하여 제안한 다음의 교차판매척도를 사용하였다.

$$CSE(X \Rightarrow Y) = \frac{p(XY)}{p(X)} \times \frac{p(Y)}{p(X \cup Y)} \leq 1$$

이 척도는 신뢰도에 판매량의 가중치를 부과한 것으로서, 결과가 되는 상품 Y가 전체 상품판매량에서 차지하는 비율이 높을수록 상품 X가 상품 Y에 미치는 교차판매효과 $CSE(X \Rightarrow Y)$ 는 커지게 된다. 이 계산식을 TV와 Video의 예에 적용하면, $CSE(TV \Rightarrow Video)$ 와 $CSE(Video \Rightarrow TV)$ 는 가중치의 영향으로 각각 30%와 53%로 감소되기 때문에, TV와 Video를 함께 구입했던 고객이 상점에 TV 혹은 Video가 없는 경우 장바구니에 담긴 상품의 구입을 모두 포기하기보다는 나머지 상품만을 구입하는 현실을 반영한다. 여기서, TV의 판매가 Video의 판매를 유발하는 정도는 신뢰도를 사용할 때의 50%로부터 30%로 대폭 감소되지만, Video의 판매가 TV의 판매를 유발하는 정도는 신뢰도를 사용할 때의 67%로부터 53%로 상대적으로 감소폭이 작다.

2.3 상품 네트워크 구성

대표적인 네트워크의 하나인 인터넷은 상호 독립적으로 작성된 웹페이지들이 하이퍼 링크를 통해 서로를 참조하거나 혹은 참조되면서 전 세계를 하나로 연결하는 거대한 네트워크이다. 각 웹페이지에 포함되어 있는 콘텐츠는 웹페이지의

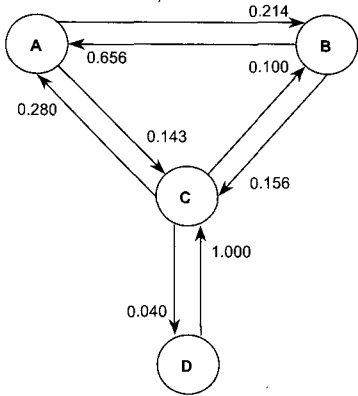
중요도를 결정하는 가장 중요한 정보이지만, 웹페이지간의 연결구조는 웹페이지의 중요도를 결정하는 또 하나의 중요한 정보를 제공한다 [Lempel and Moran, 2001]. 예를 들어, 웹페이지 p에서 웹페이지 q로 연결하는 $p \rightarrow q$ 링크는 웹페이지 p를 방문한 후에 링크를 따라 q를 방문하도록 유도한다. 이러한 네트워크의 연결구조는 상품간의 교차판매효과를 표현할 수 있으며, 네트워크에서 사용되는 가중치 계산방법은 각 상품이 전체 수익에 미치는 영향력의 계산에 적용될 수 있다.

네트워크의 연결구조에서는 웹페이지간을 연결하는 링크의 강도가 모두 동일한 것으로 가정한다 [Page, 1998; Kleinberg, 1999]. 하나의 웹페이지에 존재하는 모든 링크는 동일한 확률로 선택될 것이라는 가정으로서, 현실적으로는 웹페이지마다 선택되는 정도가 다르기 때문에 링크마다 서로 다른 연결강도를 부여하는 것이 바람직하다. 이에 따라, 소규모의 웹사이트에서는 웹 로그(Web Log) 분석을 이용하여 웹페이지를 연결하는 링크들의 가중치를 서로 다르게 부여함으로써 각 웹페이지의 중요도를 정확히 평가하기 위한 연구도 이루어지고 있다 [Xue et al., 2003].

본 연구에서는 각 상품을 웹페이지로 보고, 상품간의 연관규칙을 네트워크의 링크로 설정함으로써 상품 네트워크를 구성한다. 각 상품이 다른 상품에 영향을 미치는 연결강도에는 교차판매척도를 사용했다. 여기서, 교차판매척도는 상호 비대칭이기 때문에, 이로부터 구성되는 상품 네트워크는 이중연결(Doubly Linked)구조를 갖는 방향성 그래프이다.

예를 들어, 상품 A, B, C, D가 총 11회의 거래에서 각각 {A, C}, {A}, {C}, {A, B, C}, {B}, {A}, {A, B}, {A, B}, {C}, {A}, {C, D}의 장바구니를 구성했다고 가정해 보자. 여기서, 상품 A를 포함하는 7회의 거래에서 상품 B를 포함하는 거래는 3회로서, 앞의 교차판매척도에 따르면 $CSE(A \Rightarrow B)$

는 0.214가 되고, $CSE(B \Rightarrow A)$ 는 0.656으로 계산된다. 이에 따라 상품 네트워크를 구성하면 <그림 1>과 같다.



<그림 1> 연관규칙으로 구성한 상품 네트워크

연관규칙을 이용하여 구성한 상품 네트워크는 상품간의 연결관계, 각 상품의 종속성, 그리고 교차판매 영향력의 정도를 보여준다. 즉, 다른 노드(Node)로부터 들어오는 입력링크(In Link)의 값이 클수록 종속적이며, 다른 노드로 연결되는 출력링크(Out Link)의 값이 클수록 교차판매의 영향력은 크게 나타난다. 위의 예에서 상품 A의 입력링크의 합은 0.936이고 출력링크의 합은 0.357로서, 상품 A는 교차판매를 일으키기보다는 단독으로 구매되거나 혹은 다른 상품의 구매에 따라 교차판매되는 정도가 높은 상품인 것으로 판단할 수 있다.

2.4 교차판매량

앞에서 소개한 교차판매척도를 교차판매량 계산에 적용하기 위해서는 교차판매된 상품이 유발하는 2단계 교차판매 효과를 고려해야 한다. 즉, 상품 A가 상품 B와 C의 판매를 유발했다면, 교차판매된 상품 B와 C가 유발하는 상품 A와 C 등의 판매를 고려해야 상품 A의 수익창출력을 정확히 계산할 수 있다. 그러나 앞의 상품

네트워크에서 상품 A가 상품 C를 통해 상품 D와 연결되어 있으나, 상품 A가 D와 단 한번도 함께 판매되지 않았기 때문에, 상품 A의 판매가 상품 D의 교차판매를 일으키는 것으로는 볼 수 없다. 이에 따라 본 연구에서는 상품 네트워크에서 직접적인 관계를 갖는 상품, 즉 신뢰도가 0보다 큰 경우에만 교차판매가 발생하는 것으로 가정하여 교차판매량 계산 알고리즘을 개발하였다.

본 연구에서는 교차판매효과를 보다 정확히 기술하기 위해 특정 상품의 판매를 순수판매량과 교차판매량으로 구분하였다. 교차판매량은 순수판매량으로부터 교차판매척도에 따라 계산된 판매량과 교차판매된 상품이 연쇄적으로 유발하는 교차판매량을 합산하여 계산된다. 여기서, S_i 는 각 상품의 총판매량, ρ 는 교차판매된 상품이 다시 교차판매를 일으키는 비율, 그리고 ϵ 는 계산을 종료하는 오차의 수준을 나타낸다.

Let S_i is the amount of product sold,
 ρ is the cross reselling factor, and
 ϵ is the allowed error level.

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

$$Q_{ij} = c_{ij} * S_i$$

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 2> 교차판매량 계산 알고리즘

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 ρ 를 곱한 값으로부터 계산된다. 앞에서 소개한 상품 A, B, C, D의 문제에서 각 상품의 교차판매량을 계산하면 <표 1>과 같다. 여기서, 대각선에 위치한 셀들은 각 상품의 순수판매량이며, 기타의 셀은 교차판매량이다.

<표 1> 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
합 계	7.000	4.000	5.000	1.000

III. 상품선택과 수익예측

소매점에서 상품교체나 상품진열을 위해 상품을 선택할 때, 전통적으로는 각 상품의 단위당 판매수익과 판매량을 곱한 총판매수익에 근거하였다. 앞에서 소개한 예에서, 상품 A, B, C, D의 단위당 판매수익이 각각 13, 24, 16, 95라고 하면, 각 상품은 차례대로 91, 96, 80, 95의 총판매수익을 창출한다. 따라서 이들 중에 2개의 상품을 선택해야 한다면 총판매수익이 작은 상품 A와 C를 제거하고 상품 B와 D를 선택하여 96+95=191의 총수익을 기대하게 된다. 그러나 제거되는 상품이 선택된 상품의 판매에 영향력을 미쳤다면, 실제 총수익은 감소된다.

<표 2> 교차판매에 따른 상품별 예상손실

상품	단위 수익	판매 수량	판매 수익	교차판매량				예상 손실
				A	B	C	D	
A	13	7	91	4.239	0.965	0.643	0.000	124.4
B	24	4	96	1.860	2.713	0.443	0.000	127.3
C	16	5	80	0.901	0.322	3.031	0.129	111.1
D	95	1	95	0.000	0.000	0.883	0.871	109.1
합계	-	17	362	7.000	4.000	5.000	1.000	-

이에 따라, 본 연구에서는 각 상품간의 교차판매량을 고려하여 각 상품의 예상손실을 계산한다. <표 2>는 교차판매가 다시 교차판매를 일으키는 비율 ρ 를 0.1, 종료조건 ϵ 을 0.0001로 설정했을 때의 계산결과로서, 예상손실이 가장 낮은 상품 C와 D를 제거하고 상품 A와 B를 선택한다.

하나의 상품만을 제거하고 나머지 상품을 모두 선택하는 문제라면, 전체 판매수익에서 해당 상품의 예상손실을 차감하여 계산될 수 있다. 그러나 두 개 이상의 상품을 제거할 경우, 각 상품의 예상손실을 모두 차감하면 제거되는 상품간의 교차판매량이 이중으로 계산되기 때문에 전체수익이 음(-)으로 계산되는 문제가 발생한다. 이에 따라 본 연구에서는 제거되는 상품들의 자체판매 손실과 이들이 남아 있는 상품의 판매에 미치는 영향을 통해 다음과 같이 예상손실을 계산하였다.

$$L(R) = \sum_{i \in R} P_i \times S_i + \sum_{i \in R, j \in R} P_j \times Q_{ij}$$

여기서, R 은 제거된 상품들의 집합, P_i 는 상품 i 의 단위당 수익, S_i 는 상품 i 의 총판매량, 그리고 Q_{ij} 는 제거된 상품 i 가 남아있는 상품 j 에 미치는 교차판매량으로서 교차판매척도를 이용하여 계산한다. 위의 예에 이 계산식을 적용하면, 상품 C와 D를 제거할 경우의 예상손실은 194.5로서, 전체 판매수익에서 이를 차감하면 167.5의 수익을 기대할 수 있다. 이는 교차판매효과를 고려하지 않았을 때의 예상수익 187보다 19.5만큼 감소된 것이다.

IV. 컴퓨터 시뮬레이션

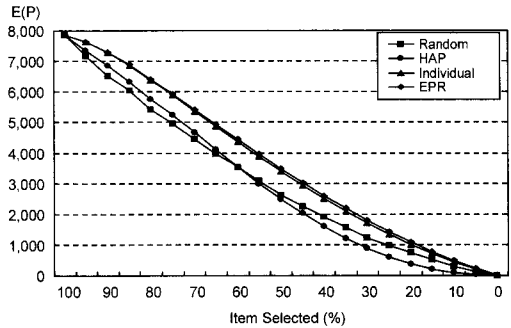
본 연구에서 제안한 예상손실을 이용한 상품 선택 알고리즘의 성과를 측정하기 위해 다음과 같이 컴퓨터 시뮬레이션을 실시하였다. 먼저, 컴퓨터 시뮬레이션 환경 및 데이터 세트의 구성에

대해 기술하면 다음과 같다. 본 알고리즘은 2.66GHz CPU를 갖는 윈도우즈 2000서버 운영체제에서 객체지향언어인 자바(Java)로 구현되었으며, 데이터관리는 오라클 데이터베이스를 이용하였다. 또한 실제 경영환경에서 판매상품의 구성을 변화시키면서 실험을 하는 것은 현실적으로 불가능하기 때문에, 이를 모형화하여 상품 및 거래 데이터를 생성하였다.

상품 및 거래 데이터의 생성방법은 다음과 같다. 컴퓨터의 처리속도를 고려하여 상품의 개수는 100개, 거래횟수는 1,000회로 하여 총 30개의 데이터 세트를 생성하였다. 각 상품의 단위당 판매수익은 10%는 0.1~0.999, 80%는 1.0~4.999, 그리고 나머지 10%는 5.0~9.999로 하였다. 이는 Wang and Thomas[2002]가 유사 정규분포(Normal Distribution)의 데이터 세트를 생성하기 위해 사용한 방법을 준용한 것이다. 각 장바구니에는 1에서 10개의 상품이 포함되도록 하였는데, 앞에서 단위당 판매수익을 설정하는 방법과 유사하게 10%는 1개, 80%는 2~6개, 그리고 나머지 10%는 7~10개의 상품이 포함되도록 하였다. 끝으로, 장바구니에 포함되는 상품을 선택할 때, Wang and Thomas[2002]는 모든 상품이 동일한 확률로 선택되도록 하였으나, 이는 단위당 수익이 낮은 저가의 상품이 단위당 수익이 높은 고가의 상품보다 상대적으로 많이 판매되는 현실을 적절히 반영하지 못하기 때문에, 본 연구에서는 상품의 단위당 수익이 증가할수록 판매량은 지수적으로 감소되도록 하였다.

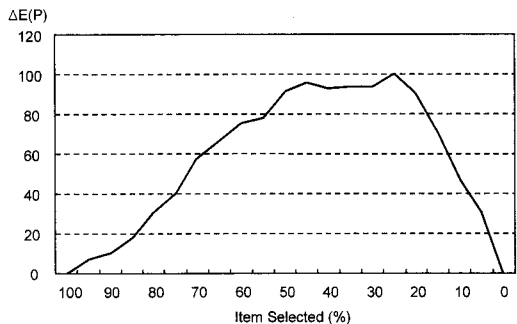
위에서와 같은 방법으로 생성한 30개의 데이터 세트에 대해 선택상품의 비율을 100%에서부터 0%까지 변화시키면서 컴퓨터 시뮬레이션을 실시하였다. 본 연구에서는 교체상품을 임의로 선택하는 Random, 교차판매의 효과를 고려하지 않고 매출수익이 큰 상품을 선택하는 Individual Rank, Brijis et al.[1999]이 제안한 PROFSET 보다 나은 성과를 나타내었던 Wang and Thomas [2002]의 HAP, 그리고 본 연구에서 제안한 기대

수익을 이용하여 상품을 선택하는 EPR 등 네 가지 알고리즘의 성과를 측정하였다.



<그림 5> 선택상품의 변화에 따른 기대수익의 변화

시뮬레이션 결과, <그림 5>에서 보는 바와 같이 Random의 성과가 가장 낮게 나타났으며, HAP은 선택상품의 비율이 높을 때 Random보다 약간 더 나은 성과를 나타냈다. 다음으로 각 상품의 판매수익에 근거한 Individual Rank는 상대적으로 좋은 성과를 나타냈으나, 본 연구에서 제안한 EPR이 가장 좋은 성과를 나타냈다. 이러한 결과는 상품의 종류, 가격, 장바구니의 구성 등에 따라 달라질 수 있기 때문에, 그 차이의 정도나 행태를 일반화하기에는 좀 더 많은 연구가 필요할 것으로 판단된다.



<그림 6> 선택상품의 변화에 따른 기대수익의 차이

<그림 6>은 EPR과 Individual Rank 알고리즘 간의 성과차이를 보다 명확히 제시하기 위해 선

택상품의 개수가 100%로부터 0%까지 감소될 때의 기대수익의 차이 $\Delta E(P)$ 를 그림으로 나타낸 것으로서, 선택상품의 비율이 100%로부터 약 50% 수준에 이르기까지 지속적으로 증가되다가 약 25% 수준 이후에는 점차적으로 감소되는 것으로 나타났다.

V. 결 론

기업간의 경쟁이 심화되고 정보통신기술이 급격히 발달함에 따라, 기업들은 고객 및 거래에 대한 방대한 규모의 데이터베이스로부터 정보 및 지식을 추출하여 경영의사결정에 활용하는 데이터마이닝에 많은 관심을 갖게 되었다. 특히, 거래 데이터로부터 상품간의 관계를 규명하는 연관규칙은 광범위하게 사용되고 있으나, 오프라인의 소매상에서 최적의 마케팅 믹스를 구성하기 위해 연관규칙을 어떻게 활용할 수 있는지에 대한 답을 제시하지는 못하고 있다. 또한, 각 상품이 전체 상품의 판매에 어떠한 파급효과가 있으며, 수익에 어느 정도의 기여를 하고 있는지에 대한 정확한 분석이 이루어지지 못하기 때문에 상품의 가격결정이나 교차상품선택 등의 중요한 의사결정에는 사용되지 못하고 있는 실정이다.

교차판매의 영향을 고려하여 상품을 선택하기 위한 연구들이 진행되고는 있으나, 교차판매

의 파급효과를 반영하지 못할 뿐만 아니라, 선택된 상품의 기대수익을 정확하게 예측하지 못하는 문제가 있다. 이에 따라 본 연구에서는 연관규칙을 기반으로 구성된 상품 네트워크에서 각 상품의 순수판매량과 교차판매량에 근거하여 상품을 선택하며 기대수익을 예측하는 EPR 알고리즘을 제안하였다. 컴퓨터 시뮬레이션 결과, 전통적으로 사용되는 Individual Rank나 HAP 알고리즘보다 더 좋은 성과를 나타냈다.

본 연구의 한계점으로는 현실의 데이터를 사용하는 대신 컴퓨터를 이용하여 생성한 데이터 세트를 이용하였기 때문에, 현실의 문제에 어느 정도 일치하는지에 대해 검증할 수 없었다는 것과 장바구니에 포함되는 각 상품의 수량을 고려하지 못했다는 것이다. 그러나 특정제품의 교차판매효과를 측정하기 위해서는 해당 상품을 제거했을 때의 거래데이터를 수집해야 하는 데, 이는 현실적으로 거의 불가능한 일이다. 이에 따라, 웹에 가상의 상점을 개설한 후 다양한 실험설계를 통해 거래데이터를 수집하여 분석하기 위한 후속 연구를 진행하고 있다. 또한, 관련 학문분야의 연구자들과의 공동연구를 통해 본 연구의 결과를 이론적으로 검증할 뿐만 아니라, 대체재 및 보완재의 유무에 따른 교차판매효과의 변화, 그리고 상품의 가격탄력성과 교차판매효과를 고려한 가격결정방안 등으로 연구의 범위를 확대해 나갈 계획이다.

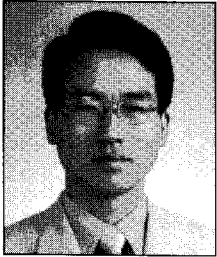
〈참 고 문 헌〉

- [1] Agrawal, R. and Srikant, R., "Fast Algorithm for Mining Association Rules," *In Proceedings of the 20th International Conference on Very Large Databases*, 1994, pp. 487-499.
- [2] 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.
- [3] Agrawal, R., Imielinski, T. and Swami, A., "Mining Association Rules Between Sets of Items in Large Databases," *In Proceedings of ACM SIGMOD Conference on Management of Data*, 1993, pp. 207-216.
- [4] 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," *In Proceedings on ACM KDD-2000*, Boston, MA, 2000, pp. 300-304.
- [5] Brijs, T., Swinnen G., Vanhoof, K. and Wets, G., "Using Association Rules for Product Assortment Decisions: A Case Study," *In Proceedings on KDD-99*, ACM, San Diego, CA, USA, 1999, pp. 254-260.
- [6] Brin, S., Motwani, R., Ullman, J. and Tsur, S., "Dynamic Itemset Counting and Implication Rules for Market Basket Data," *In Proceedings ACM SIGMOD International Conference on Management of Data*, 1997, pp. 255-264.
- [7] Deshpande, M. and Karypis, G., "Item-Based Top-N Recommendation Algorithms," *ACM Transaction on Information Systems*, Vol. 22, No. 1, 2004, pp. 143-177.
- [8] 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.
- [9] Hwang, I., "Association Rule for The Analysis of Cross-Selling Effect," *Working Paper*, University of South Carolina, 2004.
- [10] Kalakota, R. and Robinson M., *e-business: Roadmap for Success*, Addison Wesley, 1999.
- [11] Kits, B., Freed, D. and Vrieze, M., "Cross-sell: A Fast Promotion-tunable Customer-Item Recommendation Method based on Conditional Independent Probabilities," *In Proceedings of ACM SIGKDD International Conference*, New York, 2000, pp. 437-446.
- [12] Kleinberg, J., "Authoritative Sources in a Hyperlinked Environment," *In Proceedings of the 9th ACM-SIAM Symposium on Discrete Algorithm*, 1998, pp. 668-677.
- [13] Kleinberg, J., "Authoritative Sources in a Hyperlinked Environment," *Journal of the ACM*, Vol. 46, No. 5, September 1999, pp. 604-632.
- [14] 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.
- [15] Lempel, R. and Moran, S., "SALSA: The Stochastic Approach for Link-Structure Analysis," *ACM Transactions on Information Systems*, Vol. 19, No. 2, April 2001, pp. 131-160.
- [16] Page, L., Brin, S., Motwani, R. and Winograd, T., "The PageRank Citation Ranking: Bringing Order to the Web," *Technical Report*, Stanford University Database Group, 1998.
- [17] Schafer, J., Konstan, J. and Riedl, J., "Recommender Systems in e-Commerce," *In Proceedings of ACM E-Commerce*, New York, 1999.
- [18] Shardanand, U. and Maes, P., "Social Information Filtering: Algorithms for Automating, Word of Mouth," *In Proceedings of the ACM CHI'95 Conference on Human Factors in Computing Systems*, New York, 1995, pp. 210-217.
- [19] Wang, K. and Thomas, Su M., "Item Selection By Hub-Authority Profit Ranking," *In Proceedings on ACM SIGKDD*, Edmonton, 2002, pp. 652-657.
- [20] 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.~Aug. 2003, Toronto, pp. 56-63.
- [21] Zaki, M., Partasarathy, S., Ogihara, M. and Li, M., "New Algorithm for Fast Discovery of Association Rules," *In proceedings of the Third International Conference on Knowledge Discovery and Data Mining*, 1997, pp. 283-286.

◆ 저자소개 ◆



황인수 (Hwang, Insoo)

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, 데이터마이닝, 웹 에이전트 등이다.

◆ 이 논문은 2004년 5월 4일 접수하여 2차 수정을 거쳐 2004년 10월 4일 게재확정되었습니다.