

# 지수 평활법을 적용한 시간 연관 규칙

변 루 나<sup>†</sup>·박 병 선<sup>††</sup>·한 정 혜<sup>†††</sup>·정 한 일<sup>††††</sup>·임 춘 성<sup>†††††</sup>

## 요 약

전자상거래가 활성화됨에 따라 고객 개인의 관심에 부합하는 개인화된 정보나 상품 서비스를 제공하기 위하여 시간에 따라 분할하여 연관하는 시간 연관 규칙이 최근에 등장하고 있다. 본 논문은 일반적으로 정의된 연관 규칙에 대해 시간의 변화를 고려하기 위하여 최신 데이터에 가중치를 높여 주는 지수 평활법을 적용한 연관 규칙을 정의하고 이를 탐사하는 알고리즘을 제안한다. 시뮬레이션과 적용사례를 통하여 시간에 따라 분할하여 지수 평활법을 적용한 시간 연관 규칙이 기존의 것보다 실행시간은 다소 많지만 시간을 고려한 정확한 탐색률을 갖으므로 전자상점 현장 응용에 효과적임을 확인하였다.

## Temporal Association Rules with Exponential Smoothing Method

Lu Na Byon<sup>†</sup>·Byoung Sun Park<sup>††</sup>·Jeong Hye Han<sup>†††</sup>  
Han Il Jeong<sup>††††</sup>·Choon Seong Leem<sup>†††††</sup>

## ABSTRACT

As electronic commerce progresses, the temporal association rule is developed from partitioned data sets by time to offer personalized services for customer's interest. In this paper, we proposed a temporal association rule with exponential smoothing method that is giving higher weights to recent data than past data. Through simulation and case study, we confirmed that it is more precise than existing temporal association rules but consumes running time.

**키워드** : 전자상거래(E-Commerce), 연관 규칙(Association Rules), 시간 연관 규칙(Temporal Association Rules), 지수 평활법(Exponential Smoothing Method)

### 1. 서 론

현재 전 세계는 인터넷의 등장으로써 전자적인 단일상권으로 묶이는 본격적인 글로벌 전자상거래 시대로 돌입하고 있다. 실제로 국내외에서 이미 수많은 전자 상점이 활발히 운영중이며, 원투원 마케팅 개념을 적용한 CRM(Customer Relationship Management)을 도입하여 서비스를 제공하고 있다[1]. 대부분 전자상점들은 CRM을 위해 다수의 고객에 대한 거래 데이터 웨어하우스의 분석을 통한 고정된 시점(time)하에서 연관 규칙(Association Rule), 클러스터링(Clustering), 신경망(Neural Networks)등을 적용한다[3]. 연관 규칙은 장바구니 분석(Market-basket Analysis)을 통해서 얻어지는 것으로, 현재까지 가장 활발하게 연구 및 활용되고 있는 데이터 마이닝 중의 하나이다[4].

효과적인 CRM을 위하여 많은 전자상점에서는 고객군을 분류하여 관리하는 고객기반 모델과 상품에 대한 아이템 기반 모델을 적용한다. 아이템 기반 모델은 아이템이 유행과 시간에 따라 매우 민감하게 반응하므로, 이를 고려하면 효과적일 것이다. 연관 규칙은 계산과 이해가 편하기 때문에 전자상거래와 같이 많은 거래 데이터를 분석하여 크로스 셀링(Cross Selling)을 하는 경우에 매우 활발히 이용되고 있지만, 시간을 고려하지 않기 때문에 적시성(Just-in-Time)이 떨어지는 단점이 있다. 즉, 전자상점의 매출에 영향을 크게 미칠 수 있는 계절상품 또는 특별상품의 거래 데이터베이스에서 중요한 속성인 시간을 고려하지 못하기 때문에 좋은 추천 성공률을 높게 가질 수 없다. 예를 들어, [발렌타인 데이, 초콜릿], [크리스마스 이브, 케익], [추석, 송편], [추수감사절, 칠면조] 등과 같은 상품은 시간과 밀접하게 관련되어 있다.

이에 따라 [5]은 타임스탬프(timestamp)된 거래에서 특정 시간 간격 동안의 주기적 연관규칙(Cyclic Association Rules)을 찾아내는 탐사 알고리즘을 제안하였고, [9]는 달력에 기반한 연관 규칙(Calendar-based Association Rules)

† 정 회 원 : 충북대학교 대학원 통계학과  
 †† 정 회 원 : 한국표준과학연구원 연구원  
 ††† 종신회원 : 청주교육대학교 컴퓨터교육과 교수  
 †††† 정 회 원 : 대전대학교 인터넷정보공학 교수  
 ††††† 정 회 원 : 연세대학교 정보산업학과 교수  
 논문접수 : 2003년 9월 30일, 심사완료 : 2004년 1월 7일

을 탐사하는 알고리즘을 제안하였다. 또한 [8]에서는 연관 규칙이 적용되는 시간 구간을 정한 후, 구간에서 연관 규칙의 주기적 패턴을 발견하여 시간적 특성을 갖는 연관 규칙을 찾아내는 방법을 제안하였고, [6]은 연관 규칙을 분할된 데이터셋에 대해 누진적 연관 규칙(Progressive Association Rules)을 제안하여 속도 향상과 시간을 보다 효과적으로 고려할 수 있게 하였다.

본 연구에서는 장기간에 걸쳐 모아진 데이터웨어하우스의 데이터를 토대로, 최근 데이터일수록 높은 가중치를 주고 오래된 데이터일수록 낮은 가중치를 주는 지수 평활법을 적용하여 시간 연관 규칙을 정의하고, 이를 탐색하는 알고리즘을 제안하고자 한다. 또한, 지수 평활법을 적용한 시간 연관 규칙이 분할된 데이터 셋에 대해 증감의 추세를 갖는 데이터 셋, 추세가 없는 데이터 셋에 대하여 어떠한 결과를 보이는지 시뮬레이션을 통해 보이고자 한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 관련 연구들을 살펴보고, 3장에서는 지수 평활법을 적용한 연관 규칙을 정의하고 탐사 알고리즘을 제안하였다. 그리고 4장에서는 실험 및 평가를 통하여 지수 평활법을 적용한 연관 규칙과 탐사 알고리즘의 효율성을 제시하였다. 마지막으로 5장에서는 결론 및 향후 연구에 대하여 논의하였다.

2. 관련 연구

연관 규칙은 거래 데이터베이스에서 아이템간에 발생하는 연관성을 표현하는 것으로[4], 어떤 사건이 발생할 때 그 다음 사건의 관련성을 의미한다. 연관 규칙  $X \rightarrow Y$ 은 데이터베이스의 거래 중  $X$ 라는 항목 집합을 포함하는 거래는 항목 집합  $Y$ 도 함께 포함하는 경향이 있음을 의미한다. 연관 규칙의 척도는 지지도(support)와 신뢰도(confidence)를 이용하여 타당성을 판단한다[4, 5].

2.1 시간 연관 규칙

타임 스템프된 거래 데이터로부터 의미 있는 지식을 탐사하기 위하여 기존의 데이터 마이닝 모델에 대한 여러 연구가 진행되고 있다[8, 9].

여기에 속한 기법으로는 달력으로 표현된 시간 패턴을 가지는 달력 연관 규칙(Calendric Association Rules)[9], 주기적으로 반복되는 주기적 연관 규칙(Cyclic Association Rules)[5], 분할 셋을 기반 연관 규칙(Partitioned Association Rules)[8], 누진적 가중 연관 규칙(Progressive weighted Association Rules)[6] 등이 있으며, 이를 요약하면 다음 <표 1>과 같다.

<표 1> 시간 연관 규칙의 종류 및 특징

달력	<ul style="list-style-type: none"> <li>• 달력으로 표현 가능한 특정 일에 발생하는 아이템간의 관계성</li> <li>• 달력스키마 <math>R=(G_n, D_n, G_{n-1}, D_{n-1}, \dots, G_1, D_1)</math>에 대해, 연관 규칙 <math>(A \rightarrow B) \langle d_n, d_{n-1}, \dots, d_1 \rangle</math>가 있다.</li> <li>• 예) 이승엽 56호 홈런 기간의 야구티켓과 잠자리채</li> </ul>
주기	<ul style="list-style-type: none"> <li>• 주기적 특성을 갖고 발생하는 아이템간의 관계성</li> <li>• 주기 <math>c=(l, o), 0 \leq o \leq l</math>을 갖는 주기적 연관규칙 <math>(A \rightarrow B)</math>이 있다.</li> <li>• 예) 발렌타인데이의 초콜릿과 카드</li> </ul>
분할	<ul style="list-style-type: none"> <li>• 동일기간으로 분할하여 연관 규칙을 탐색</li> <li>• 분할 셋 <math>i</math>에서 <math>k</math>-아이템 셋의 지역 빈발 집합 <math>L_k^i</math>, 지역 후보 집합 <math>C_k^i</math>, 전역 빈발 집합 <math>L_k^G</math>, 전역 후보 집합 <math>C_k^G</math>에 대해 연관 규칙</li> <li>• 시간분할 효과 추가가 있으나, 가중치 개념은 없음</li> <li>• 메인 메모리가 크게 요구되며, 분할 셋의 크기를 메인 메모리 크기로 제한</li> </ul>
누진 가중	<ul style="list-style-type: none"> <li>• 동일기간으로 분할하여 가중함수 <math>W(\cdot)</math>로 최근에 더 큰 가중치를 부여하며, 계산량 감소</li> <li>• <math>s^{PW}(A \cup B) = \frac{S^S(A \cup B)}{\sum  P_i  \times W(P_i)}</math>인 누진적 가중 연관 규칙 <math>(A \cup B)^{PW}</math>이 존재한다.</li> <li>• 가중함수 <math>W(\cdot)</math>의 선택이 임의롭고, <math>W(t) = 1 + \frac{t}{n}</math>의 경우 거래수가 클 경우 적당치 않다. 또한 임의의 가중 함수에 의해 결과해석에 대해서도 해석이 모호</li> </ul>

따라서 본 연구는 거래 데이터 셋  $D$ 를  $k$ 개로 등분할 하여 시간에 대해  $D_i$ 에 대해 시간 변화에 따라 가중치를 주어 연관 규칙을 탐사하는 방법으로, 분할법과 누진 가중기법의 특징을 모두 적용하면서도 메인 메모리를 크게 차지하거나, 사용자가 미리 가중함수를 정해야하는 사전 경험지식을 갖지 않을 수 있는 방법을 제안하고자 한다.

2.2 지수 평활법

일반적인 지수평활 모형은 다음 식 (1)과 같이 정의된다[7].

$$S_t = \alpha y_t + (1 - \alpha) S_{t-1}, \quad 0 < \alpha \leq 1 \quad (1)$$

단,  $y_t$ 는 시점  $t$ 에서의 총 관측값, 초기값  $S_0$ 는 평균값,  $\alpha$ 는 데이터의 영향 정도를 결정하는 지수평활 상수

이 방법은  $\alpha$ 값이 클수록 최근 데이터에 상대적으로 큰 가중치를 준다. 지수평활 상수  $\alpha$  값의 결정은 0~1 사이에서 경험적으로 가장 효과적인 값을 이용한다. 이러한 지수 평활법은 계산이 간단하고 최신자료를 많이 반영하는 적시성 때문에 많은 공학분야에 활발히 활용되고 있으며, 최근에 [2]에서는 협력적 여과 추천 알고리즘에 이를 적용한 이론을 제시하였으나, 전자상거래 데이터에는 아직 적용사례가 없다. 따라서 본 연구에서는 지수 평활법을 적용한 시간 연관 규칙을 제안하여, 최신 정보를 더 많이 고려하는 시간 연관 규칙을 계산할 수 있게 한다.

2.3 문제 제기

전체 트랜잭션 수  $N=1,000$ 이며  $t_1 \sim t_5$ 시점까지 분할된 데이터 셋의 4개 상품에 대하여  $count(P_A, P_B)$ 는 일정한 반면,  $count(P_C, P_D)$ 는 감소하는 경향이 있는 <표 2>를 가정하자.

<표 2> 시간 연관 규칙의 지지도 비교

	과거	←	...	→	최근	합	기준 지지도	누진적 지지도	지수 평활 지지도
	$t_1$	$t_2$	$t_3$	$t_4$	$t_5$				
$Count(P_A, P_B)$	50	45	55	40	50	240	0.240	0.240	0.239
$Count(P_C, P_D)$	100	70	40	20	10	240	0.240	0.239	0.212

이러한 경우에 Apriori 연관 규칙  $P_A \rightarrow P_B$ 와  $P_C \rightarrow P_D$ 의 지지도는 0.240로 동일하며, 주기적 또는 달력기반 연관 규칙은 이를 반영하지 못한다. 반면 분할 셋으로 누진적 가중 지지도의 경우  $W(t) = 1 + \frac{5t}{N}$ 로 가정하면 0.240와 0.239이며,  $\alpha=0.2$ 로 가정한 지수 평활법의 경우에 각각 0.239와 0.212로 계산되어, 추세가 있는 경우에 대하여 보다 정확히 근접해 가는 것을 알 수 있다. 누진적 지지도의 경우 기준의 지지도보다는 좋지만,  $N$ 이 매우 큰 대부분의 경우에 효과적이지 못하다. 또한 가중치의 해석이 사용자에게 의존적이며 모호하여 연관 규칙들이 많을 때 비교 평가가 쉽지 않다.

이와 같이 최근 시점에서의 거래 데이터에 대하여 지수 평활법을 적용함으로써 사전 가중치를 주지 않으면서도 추천 성공률을 보다 높일 수 있을 것이다.

3. 지수 평활법을 적용한 시간 연관 규칙

3.1 지수 평활법을 적용한 시간 연관 규칙

본 논문에서 제안하는 지수평활 시간 연관 규칙은 <표 1>의 분할과 누진적 가중 연관 규칙과 같이 데이터 셋을 분할함으로써 정의된다. 타임 스탬프된 거래 데이터 셋  $D$ 를 동일 구간을 갖는  $k$ 개의 서브셋으로 분할한  $DS_t, t = 1, 2, \dots, k$ 이 있다고 하자. 이 때  $D = \bigcup_{t=1}^k DS_t, \emptyset = \bigcap_{t=1}^k DS_t$ 을 만족하며, 상품집합을  $\{p_1, \dots, p_m\}$ 라 하자.

각  $DS_t$ 에 대하여 지역 지지도(local support)  $s_t$ , 지역 신뢰도(local confidence)  $c_t$ ,  $D$ 에 대하여 전역 지지도(global support)  $s$ , 전역 신뢰도(global confidence)  $c$ 가 존재한다.

본 논문에서 제안하는 지수 평활 시간 연관 규칙은 지수 평활 전역 지지도(global support)  $s^{ES}$ , 지수평활 신뢰도(global confidence)  $c^{ES}$ 이 각각 사용자가 미리 정한 최소 지지도  $ms$ , 최소 신뢰도  $mc$ 보다 큰 모든 연관 규칙으로

다음 정의와 같다.

[정의] 어느 시점  $t$ 에 대하여, 상품 크기가  $j \leq m$ 일 때, 지지도와 신뢰도  $s^{ES}, c^{ES}$ 를 갖는  $\{p_1, \dots, p_{j-1}, p_{j+1}, \dots, p_m\} \rightarrow p_j(s^{ES}\%, c^{ES}\%)$ 은 지수평활 시간 연관규칙이다. 단,  $0 < \alpha \leq 1$ 에 대하여  $s^{ES}, c^{ES}$ 는 각각 다음을 만족한다.

$$\frac{\sum_{t=1}^k count(\{p_1, \dots, p_m\}, DS_t) \cdot \alpha(1-\alpha)^{k-t+1}}{|D|} \geq ms\%,$$

$$\frac{\sum_{t=1}^k [count(\{p_1, \dots, p_m\}, DS_t)]}{\sum_{t=1}^k [count(\{p_1, \dots, p_{t(i-1)}, p_{t(i+1)}, \dots, p_m\}, DS_t)]} \times \frac{\alpha(1-\alpha)^{k-t+1}}{\alpha(1-\alpha)^{k-t+1}} \geq mc\%. \quad \square$$

[6]의 누진적 연관 규칙과 본 연구의 지수평활 시간 연관 규칙간의 관계를 다음의 따름 정의와 같이 쉽게 얻을 수 있다.

[따름정의]  $\alpha=1$ 이면, 지수평활 시간 연관규칙

$$\{p_1, \dots, p_{j-1}, p_{j+1}, \dots, p_m\} \rightarrow p_j(s^{ES}\%, c^{ES}\%) \\ \equiv \{p_1, \dots, p_{j-1}, p_{j+1}, \dots, p_m\} \rightarrow p_j(s^{PW}\%, c^{PW}\%) \quad \square$$

이와 같이 지수 평활 시간 연관 규칙은 누진 가중 시간 연관 규칙의 특수한 경우이다. 그러나 본 연구의 지수평활 시간 연관 규칙을 계산하기 위한 탐사 알고리즘은 3.2절에서 제안하며, 이는 누진 가중 연관 규칙과 다르다.

3.2 탐사 알고리즘

다음 (그림 1)은 3.1절 정의와 식 (1)에 의한 지수평활 시간 연관 규칙 집합을 계산하는 탐사 알고리즘이다.

```

Input : D, N, k, a, ms
Output : ES_AR // 지수평활 시간 연관규칙 집합
procedure ES_AR(D, N, k, a, ms)
1 while D ≠ ∅ and N > k do
2   sort(D) by transaction_time
3 for t = 1 to k begin
4   read_in_partition(DS_t ∈ D)
5   ES_count(t) = average(DS_t)
6   Lt = gen_large_itemsets(DS_t)
7 for(i=2; Lt_i ≠ ∅; i++) begin
8   C_i^G = ∪_{j=1,2,...,k} Lt_j^i
9 for all candidates c ∈ C_i^G
10  ES_count(t+1)
    ← a * gen_count(c, DS_{t+1}) + (1-a) * ES_count(0,t)
11 end
12 s^ES = c.count / N
13 ES_AR = {c ∈ C_i^G | s^ES ≥ ms}
14 return ES_AR
    
```

(그림 1) 지수 평활법을 적용한 연관 규칙 알고리즘

이 때,  $N$ 은  $D$ 의 거래 총 수,  $k$ 는 분할 셋 수,  $DS_t$ 는  $t$  시점에서의 분할 셋,  $a$ 는 지수평활 상수,  $ES\_AR$ 은 지수평활 시간 연관 규칙,  $L_i^k$ 는 분할 셋  $t$ 에서  $i$ -아이템 셋의 지역 빈발 집합,  $C_i^k$ 은 지역 후보 집합,  $C^k$ 은 전역 후보 집합을 나타낸다.

#### 4. 실험 및 평가

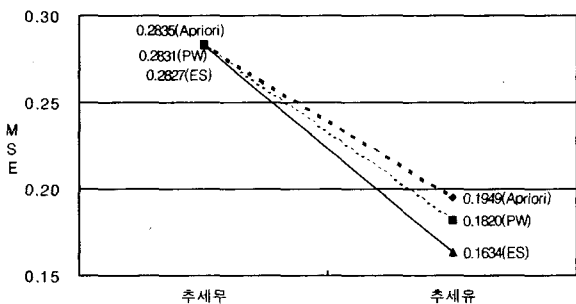
##### 4.1 실험 설계

본 절에서는 제안한 지수 평활법을 적용한 시간 연관 규칙을 기존의 일반 연관규칙 및 분할 셋에 대한 누진가중 연관규칙과 비교하였다.

전자상점 A마트에서 취급하는 32개 항목에 대한 1998년 1월~2002년 3월까지 266,854개의 거래 데이터를 분기별 시점( $k=17$ )으로 구분한 후, 연관 규칙을 사전 계산한다. 구해진 연관 규칙을 대상으로 회귀분석(Regression Analysis)에 의해 시간에 따른 추세 변화를 갖는 구매 상품과 추세가 없는 상품 집합으로 구분하여 지지도에 대한 평균오차 제곱합(Mean Squared Error)과 계산시간을 비교 실험을 하였다. 추세 유무별 상품 집합을 분류한 후, 각 집합에 대해 75%의 13개 분기 레코드를 훈련용(training)으로, 나머지 25%는 검정용(testing)으로 최소 지지도 값 0.1~0.9에 대해 누진적 가중 시간 연관 규칙과 지수평활 시간 연관 규칙을 JAVA로 구현하여 실험하였다.

##### 4.2 실험 결과

최소 지지도를 0.2, 0.5, 0.75에 대해 시뮬레이션한 결과 모두 유사한 결과를 보였다. 그중 추세 유무 상품군에 따른 지수 평활법 시간 연관규칙과 누진가중 시간 연관규칙에 대하여 검정용 셋의 최소 지지도 0.2에 대한 MSE가 (그림 2)와 같다.

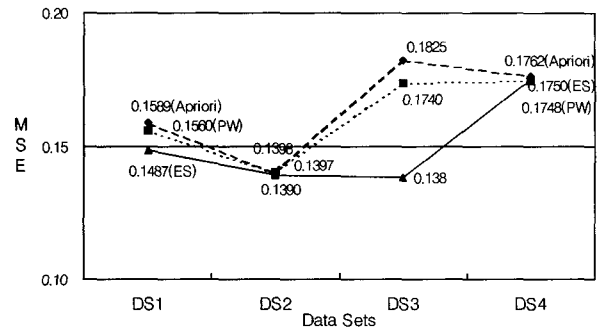


(그림 2) 추세 유무군에 따른 MSE( $ms=0.2$ )

추세가 없는 상품군에 대해서는 Apriori법, 누진 가중법, 지수 평활법들간에 차가 매우 미미하나, 추세가 있는 상품군에 대해서는 지수 평활법이 가장 작은 MSE를 갖음으로

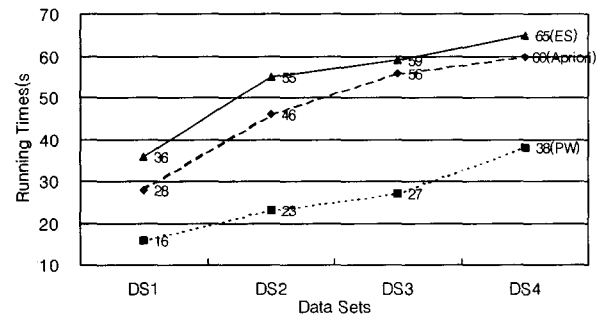
써 가장 우수한 정확도를 보였다.

$DS_1$ (2001년 7월~2001년 12월),  $DS_2$ (2001년 1월~2001년 12월),  $DS_3$ (2000년 7월~2001년 12월),  $DS_4$ (2000년 1월~2001년 12월)로 누적 서브셋으로 나누어 누적 데이터 서브셋에 따른 시간 연관 규칙의 매칭 성공에 대한 MSE 비교 결과를 나타내는 (그림 3)를 보면, (그림 2)의 결과와 유사하게 시간 연관 규칙이 전체 데이터 셋에 대해서도 여전히 다른 연관 규칙보다 적은 MSE를 가짐으로써 가장 효과적임을 보였다.



(그림 3) 누적 데이터 셋에 따른 MSE

데이터 셋의 크기에 따른 지수평활 시간 연관 규칙과 기존 연관 규칙의 실행시간을 (그림 4)과 같이 비교하였다.



(그림 4) 데이터 셋에 따른 실행시간 비교

직관적으로 누적 데이터 셋이 커질수록 실행시간은 세 가지 모두 증가하는 것으로 나타났으며, 지수평활 시간 연관 규칙 탐사가 Apriori 알고리즘에 비해 다소 많은 실행시간이 소요되는 것으로 나타났고 누진 가중 알고리즘이 가장 효과적으로 나타났다.

따라서 본 연구에서 정의한 지수평활 시간 연관 규칙이 기존의 시간 연관 규칙보다 효과적이나, 계산 비용을 다소 더 들을 보였다. 이는 분할 셋에 대한 지수평활 필터의 추가계산으로 인한 당연한 결과이다. 최근 컴퓨터들의 발달로 인해 계산 비용의 중요성을 고려하지 않는다면, 지수 평활 시간 연관 규칙은 시간을 고려하지 않는 경우 탐색되지 않

을 수도 있는 연관 규칙을 정확하게 추가 탐색할 수 있다. 또한 이렇게 얻어진 연관규칙을 일반 연관규칙 탐색결과와 비교해봄으로써, 시간을 고려한 직관적 해석도 용이하며 효과적이라 할 수 있다.

4.3 적용 사례

2001년 1월~2001년 3월(12주)까지 전자상점 B 마트에서 취급하는 식품 및 잡화, 의류품목 50개 품목에 대한 거래 데이터( $k=12$ )로 분할하여 지수 평활법 시간 연관 규칙을 계산하였다. 총132,346개 레코드를 103,231개의 훈련용과 29,115개의 검정용으로 분할하여 기존의 연관 규칙과 지수 평활 시간 연관 규칙에 의해 각각 상품을 추천한 후, MSE를 <표 3>과 같이 제시했다.

도출된 총 13개의 연관 규칙의 구매 상품에 대해 회귀분석을 실시하여 유의한 추세를 갖는 경우(8개)와 그렇지 않은 경우(5개)로 분류하여 MSE를 제시하였는데, 음영부분은 차이가 작게 나타난 우수한 결과를 표시한다.

따라서 적용 사례에 대한 실험 결과인 <표 3>과 같이, 추세가 있는 구매상품의 경우에 지수평활 시간 연관 규칙이 기존 연관 규칙보다 추천 성공률을 보다 높일 수 있음을 보여준다.

<표 3> 두 연관규칙의 MSE(단위 : %)

추세	연관 규칙	연관 규칙		지수평활 시간 연관 규칙	
		지지도	신뢰도	지지도	신뢰도
두	$P_{50} \rightarrow P_{11}$	<b>1.29</b>	<b>3.01</b>	1.39	3.25
	$P_{47} \rightarrow P_{11}$	<b>0.74</b>	<b>1.23</b>	0.83	1.44
	$P_{12}, P_1 \rightarrow P_{11}$	<b>7.60</b>	<b>2.19</b>	7.63	2.27
	$P_5, P_1 \rightarrow P_{11}$	<b>7.69</b>	<b>1.60</b>	7.86	2.03
	$P_7, P_5, P_1 \rightarrow P_{12}$	6.63	2.06	<b>6.49</b>	<b>1.63</b>
유	$P_{12} \rightarrow P_{11}$	3.65	1.95	<b>3.19</b>	<b>1.32</b>
	$P_5 \rightarrow P_{11}$	3.97	1.06	<b>3.79</b>	<b>0.79</b>
	$P_7, P_{12} \rightarrow P_{11}$	4.27	1.98	<b>3.83</b>	<b>1.18</b>
	$P_5, P_{12} \rightarrow P_{11}$	4.30	0.89	<b>3.91</b>	<b>0.13</b>
	$P_5, P_{11} \rightarrow P_{12}$	4.30	2.24	<b>3.91</b>	<b>1.50</b>
	$P_7, P_5 \rightarrow P_{11}$	4.65	1.34	<b>4.42</b>	<b>0.87</b>
	$P_7, P_5 \rightarrow P_{12}$	5.25	2.89	<b>4.72</b>	<b>1.80</b>
	$P_7, P_5, P_{11} \rightarrow P_{12}$	4.73	2.52	<b>4.37</b>	<b>1.63</b>

5. 결 론

본 연구에서는 전자상점의 매출에 영향이 크게 미칠 수 있는 계절상품이나 기획상품과 같이 시간의 변화에 민감한 추세를 갖는 상품들에 대하여 지수평활 시간 연관 규칙을 정의하고, 이를 탐사하는 알고리즘을 제안하였다. 그리고 시블레이션을 통하여 제안된 방법이 기존의 시간 연관 규

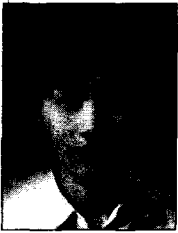
칙에 비해 실행시간이 다소 소요되는 것으로 나타났지만, 보다 작은 오류를 갖는 것으로 나타나 좋은 예측력을 갖음을 보여주었다. 특히 추세가 없는 상품의 경우보다 추세가 있는 항목의 경우에 더 작은 오류를 갖는 것으로 나타났으며, 이는 적용 사례를 통해서도 같은 결과를 얻었다.

따라서 시간에 민감한 상품을 취급하는 전자상점의 추천 시스템 구축시 본 연구에서 제안하는 지수평활 시간 연관 규칙을 적용함으로써, 상품검색, 상품 선택에 의한 고객만족의 극대화할 뿐만 아니라 매출증대를 높일 수 있을 것이다.

향후 연구로는 전자상점 거래 데이터에 대한 경험적 최적 지수평활 상수  $\alpha$  값의 계산과 분할갯수  $k$ 에 따른 탐사 알고리즘의 효율성 분석 및 최적 분할 셋 값 추정이 될 것이다.

참 고 문 헌

- [1] 돈페퍼스, 마사로저스, "평생고객을 만드는 1대 1 마케팅혁명, CM비즈니스", 1995.
- [2] 이경희, 한정혜, 임춘성, "지수적 가중치를 적용한 협력적 상품 추천 시스템", 정보처리학회논문지B, 제8-B권 제6호, 2001.
- [3] Agrawal R., Imielinski T. and Swami A., "Mining Association Rules between Sets of Items in Large Database," Proceedings of ACM SIGMOD Conference on Management of Data, Washington D. C., pp.207-216, 1993.
- [4] Agrawal R. and Strikant R., "Fast Algorithms for Mining Association Rules," Proceedings of the 20th International Conference on Very Large Databases, Santiago, Chile, 1994.
- [5] Banu Ozden, Sridhar Ramaswamy, Abraham Silberschatz, "Cyclic Association rules," Proceeding of International Conference on Data Engineering, pp.412-421, 1998.
- [6] Chang-Hung Lee., "Mining Association Relationship in a Temporal Database," Ph. D. Thesis, Department of Electrical Engineering National Taiwan University Taipei, Taiwan, ROC, 2003.
- [7] Hunter, J. S., "The Exponentially Weighted Moving Average," Journal of Quality Technology, Vol.18, pp.203-210, 1986.
- [8] X. Chen and I. Petrounias, "A framework for temporal data mining," In Proc. Ninth International Conference on Database and Expert Systems Applications, DEXA, 1998.
- [9] Y. Li, P. Ning, X. S. Wang, and S. Jajodia, "Discovering Calendar-based Temporal Association Rules," Proceedings of the 8th International Symposium on Temporal and Reasoning, 2001.



**변 루 나**

e-mail : lnbyon@hanmail.net  
1992년 충북대학교 컴퓨터공학과(석사)  
2004년 충북대학교 통계학과 전산통계  
전공(박사)  
관심분야 : 전산통계, 데이터마이닝



**박 병 선**

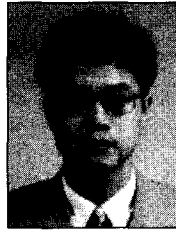
e-mail : bspark@kriss.ac.kr  
2003년 충북대학교 전자계산학과(박사)  
2001년~2003년 극동정보대학 전산정보  
처리과 겸임교수  
2003년~현재 한국표준과학연구원 연구원  
관심분야 : 데이터마이닝, 전자상거래



**한 정 혜**

e-mail : hanjh@cje.ac.kr  
1998년 충북대학교 전자계산학과(박사)  
1998년~1999년 연세대학교 산업시스템  
공학과 포닥연구원  
1999년~2001년 행정자치부 국가전문행정  
연수원 통계연수부 전산교육  
전임교수

2001년~현재 청주교육대학교 컴퓨터교육과 조교수  
관심분야 : 멀티미디어, 전자상거래, 데이터마이닝



**정 한 일**

e-mail : hijeong@dju.ac.kr  
1991년 서울대학교 산업공학과(석사)  
1996년 서울대학교 산업공학과(박사)  
1997년~현재 대전대학교 인터넷정보공학  
전공 부교수

관심분야 : 전자상거래, 기업정보화방법론, 공급망관리



**임 춘 성**

e-mail : leem@yonsei.ac.kr  
1987년 서울대학교 산업공학과(석사)  
1992년 UC Berkeley 산업공학과(박사)  
1993년~1995년 미국 Rutgers University  
산업공학과 조교수  
1995년~현재 연세대학교 정보산업전공 교수

관심분야 : 전자상거래, 기업정보화방법론