

▣ 연구논문

## 시간 가중치를 고려한 연관규칙 Discovering Time Weighted Association Rules

손 승 현\*  
Son, Seung Hyun  
김 재 련\*\*  
Kim, Jae Yearn

### Abstract

Discovery of association rules has been used useful in many fields, especially in the fields of the inventory display, catalog design and cross selling. In previous works, all transactions in the database are treated uniformly. In this paper, we present a method for partitioning transactions in the database using time weights. Transactions are assigned different weights as time goes on. Examples show that these method provides purchasing patterns in the database as well as finding association rules.

### 1. 서론

#### 1.1 연구 배경

대형할인 매장이거나 대규모의 상품을 판매하는 업계에서는 많은 양의 거래 실적을 가지고 있다. 그리고 거래 실적을 통해 매출증진에 관한 여러 가지 정보를 얻을 수 있게 된다. 따라서 기업의 의사결정자는 이러한 많은 양의 거래 실적으로부터 유용한 정보를 얻길 원한다.

연관규칙을 발견하는 작업이란 데이터 안에 존재하는 항목간의 종속관계를 찾아내는 작업이며, 마케팅에서는 손님의 장바구니에 들어있는 품목간의 관계를 알아본다는 의미에서 장바구니 분석(Market Basket Analysis)이라고도 한다. 연관규칙을 통해 여러 항목이나 품목, 또는 현상들이 동시에 발생하는 경향이 있을 경우 그에 대한 패턴을 찾을 수 있으며 이를 통해 데이터에 숨겨져 있는 데이터들간의 관계를 찾아낼 수 있다. 그리고 연관관계가 발견된 품목들에 대해서는 함께 진열한다든지, 카탈로그 제작 시 함께 패키지화하는 마케팅 전략을 수립할 수 있다.

#### 1.2 연구의 목적

기존의 데이터 마이닝에서는 하나의 데이터베이스에서 자료를 추출하여 유용한 패턴 등을 발견하였고 데이터베이스 내에 있는 자료가 최근의 자료나 오래된 자료 모두 동일한 자료로 활용되었다[2][5].

---

\* 한양대학교 산업공학과 박사과정

\*\* 한양대학교 산업공학과 교수

예를 들어, 3년 동안의 대형 할인 매장에서 분기별 매출실적에 대한 정보를 가지고 연관 규칙 기법을 사용한 데이터 마이닝을 한다고 가정해 보자. 어떤 제품들이 과거에는 서로 연관 관계를 가지고 있었지만 최근에는 두 제품간의 연관성이 없어진 제품이 존재할 때, 기존의 데이터 마이닝 기법을 통해서 그러한 매출의 추세(trend)변화를 알 수가 없다. 이 경우, 기존의 연관규칙 기법을 통해서 최근 두 제품간의 연관성이 없어진 제품들이라 할지라도 두 제품이 연관 관계를 가질 수 있기 때문이다.

또한 미래의 예측치를 구하는 방법으로는 이동평균법, 지수평활법 등이 사용되었다[1]. 이러한 방법들은 여러 기간 동안의 자료에 대해 평균을 구하거나 지수적 가중평균을 주는 방법으로 예측치를 구하였다. 하지만 이 방법들을 통해서 각 품목별 예측치를 알 수 있을 뿐, 어떠한 품목들이 서로 연관관계를 가지고 있는지는 알 수가 없었다.

본 연구에서는 마이닝을 하고자 하는 데이터베이스를 구분하여 최근의 데이터 자료에 좀 더 많은 의미(가중치)를 부여하고 오래된 자료일수록 점점 작은 비중을 두게 함으로써 매출에 대한 추세를 반영할 수 있는 연관규칙 기법을 제안한다. 이 방법은 데이터 베이스 안에 있는 자료를 시간에 따라 분할한 후, 연관규칙 기법을 적용한다. 이 방법을 통해 어떠한 항목들이 서로 연관관계를 가지고 있는지를 알 수 있을 뿐만 아니라, 그러한 연관관계가 시간이 지남에 따라 어떠한 추세변화를 가지고 있는지도 알 수 있다.

본 연구에서 제안한 방법은 매출 추세까지도 고려한 연관규칙 기법으로, 단순히 하나의 자료를 분석할 때 보다 변화되는 자료의 특성을 잘 반영할 수 있다. 또한 찾아낸 데이터들간의 연관관계뿐만 아니라 자료에 숨어있는 추세변화를 찾아내고 수치화 하여 보여 줌으로써 의사 결정에 더 정확한 정보를 제공해 줄 수 있는 방법을 제시한다.

예를 들어, 우유와 사과 두 제품이 연관성을 가지고 있다고 할 때, 데이터 마이닝을 한 결과로써 단순히 '{우유, 사과}가 연관성이 있다'라고 표현하기 보다는 {(우유, 사과)-35%증가추세}와 같이 두 제품의 매출 경향도 수치로 나타냄으로써 더욱 의미 있는 정보가 된다. 이 방법을 통해 추세 변화를 고려한 연관규칙들을 찾아내고 증가 추세가 큰 연관규칙들을 우선적으로 고려한다. 매장에서 상품 진열 시, 이러한 정보를 이용하여 연관성이 있는 제품들을 근처에 위치시킴으로써 매출 증가를 가져올 수 있다. 또한 제품 홍보를 위한 카탈로그 제작이나 교차 판매 등에 있어서도 연관제품을 같이 위치시킴으로써 효과적인 판매 계획을 수립할 수 있다.

### 1.3 기존 연구

연관규칙 기법의 대표적 방법은 Apriori 방법이다[6]. 이 후 수행 속도를 향상시키는 많은 연구가 이루어졌다[2][5]. 이러한 연구들에서는 모두 하나의 데이터 베이스를 사용하여 연관관계를 구하는 방법들을 제시하였다.

연관규칙 기법에 있어 가중치에 관한 연구는 C.H. Cai, et al., (1998)에 의해 이루어 졌다[4]. 그 논문에서는 가중치를 각 항목(item)에 부여하여 중요한 항목에 대해서는 연관관계를 가질 수 있는 확률을 높게 해주었다. 또한 각 항목들마다 다른 지지도(support)를 줌으로써 항목들의 특성을 고려한 연구도 있었다[3]. 하지만 기존 연구들은 각 항목들의 특성을 고려한 가중치나 지지도를 다르게 준 방법들이었지만 본 연구에서는 시간에 따라 분할된 각 데이터 자료에 대해 다른 가중치를 적용한다. 결국, 시간 변화에 따른 다른 가중치 적용 방법을 통해 기존 연관규칙 기법에서 발견할 수 없었던 매출의 추세 변화를 포함한 연관규칙들을 발견한다.

## 2. 제안 알고리즘

## 2.1 연관규칙

연관규칙은 {'항목A', '항목B'} → {'항목C'}의 형태로 표현하는데, "항목A와 항목 B를 포함한 거래는 항목C도 포함한다" 또는 "항목A와 항목B를 구입한 고객은 항목C 품목도 구입한다" 등으로 해석한다. 각각의 항목들의 집합을  $I = \{i_1, i_2, \dots, i_n\}$ 라고 할 때, 항목집합 (itemset)은 I의 부분집합을 말한다. 트랜잭션 T는 TID라는 유일한 식별자를 가지고 있는 항목집합이며, 검색하고자 하는 데이터베이스는 트랜잭션 T로 구성된 집합이다. 어떤 항목집합을 X라고 할 때, T가 X를 포함하면, 즉  $X \subseteq T$ 를 만족시키면 트랜잭션 T는 항목집합 X를 지지한다고 한다. 그리고 트랜잭션 총 개수에 대한 항목집합 X를 지지하는 트랜잭션 수의 비율을 X의 지지도라고 부른다. 한 트랜잭션에서 구입하는 수량은 고려하지 않기로 가정한다. 그리고 X와 Y를 각각 항목집합이라 한다면, 연관규칙은 다음과 같이 표현할 수 있다.

$$X \rightarrow Y \quad (X \subset I, Y \subset I, X \cap Y = \emptyset)$$

연관규칙  $X \rightarrow Y$ 의 지지도는 전체 트랜잭션에서  $X \cup Y$ 를 포함하고 있는 트랜잭션의 비율을 말한다. 그리고 연관규칙에서 미리 설정한 최소지지도(minimum support) 이상을 갖는 항목들의 집합을 빈발 항목집합(large itemset)이라 한다. k개의 항목들로 이루어진 빈발 항목집합을 k-빈발 항목집합이라 한다. 이러한 빈발 항목집합은 연관성을 가지게 된다. 그리고  $X \rightarrow Y$ 의 신뢰도(confidence)는 X의 지지도에 대한  $X \cup Y$ 의 지지도의 비율을 말한다. 또한 미리 지정한 최소 신뢰도(minimum confidence)를 정해둔다. 따라서 연관규칙은 사용자가 미리 지정한 최소지지도를 넘는 항목, 즉 빈발 항목집합들을 구한 후 최소 신뢰도 이상을 만족하는 항목집합들에 대한 연관규칙을 만들어 간다.

## 2.2 제안 알고리즘

본 연구에서는 연관관계의 추세변화를 찾아낼 수 있는 방법을 제안한다. 최근의 자료가 과거의 자료보다 미래를 더 잘 반영한다는 전제하에서 현재에서 과거로 시간이 지남에 따라 지수 적으로 감소하는 가중치를 이용한다. 또한 최근의 자료일수록 더 큰 비중을, 오래된 자료일수록 더 작은 비중을 주기 위해서 데이터베이스에 존재하는 자료를 시간에 따라 적절한 간격으로 분할한다. 분할한 자료에 수요 예측기법인 지수평활법에서 사용하는 가중치를 준다. 그리고 분할한 자료에 대해 각 항목집합의 지지도를 계산한 후, 가중치를 곱하게 된다. 결국 각 항목집합의 전체 지지도는 가중치가 적용된 지지도의 합으로 이루어진다.

예를 들어, 대형할인 매장의 1998~2000년 동안의 분기별 매출 자료 중, 한 항목집합에 대한 전체 지지도는 다음과 같이 구한다.  $\alpha(2000년도의\ 매출자료에\ 대한\ 지지도) + \alpha(1-\alpha)(1999년도의\ 매출자료에\ 대한\ 지지도) + \alpha(1-\alpha)^2(1998년도의\ 매출자료에\ 대한\ 지지도)$ 로 나타낼 수 있다. ( $\alpha$ : 가중치,  $0 \leq \alpha \leq 1$ ) 여기서 사용되는 가중치의 값은  $\alpha \geq (1-\alpha) \geq \alpha(1-\alpha)^2$ 임을 알 수 있으며, 최근의 자료일수록 가중치를 높게 부여한 결과가 된다. 결국 가중치는 시간이 지남에 따라 다음과 같이 변한다.

$\alpha \geq \alpha(1-\alpha) \geq \alpha(1-\alpha)^2 \geq \dots \geq \alpha(1-\alpha)^{n-1}$  이때  $\alpha$ 의 값이 클수록 최근의 자료에 더욱 의미를 많이 주는 효과를 주며, 값이 작을수록 평활(smoothing)의 효과는 더 커지게 된다.

전체 지지도가 사용자가 미리 정의한 최소 지지도를 만족하는지를 조사함으로써 그 항목집합의 빈발 항목집합 여부를 결정하게 된다. 즉 항목집합을 {사과, 우유}로 가정하고, 2000년 매출자료에서 {사과, 우유}의 지지도 = ( 3/5 ), 1999년 매출자료에서 {사과, 우유}의 지지도 = ( 2/5 ), 1998년 매출자료에서 {사과, 우유}의 지지도 = ( 1/5 ) 이라고 가정한다면 전체 지지도

는 다음과 같다. (가중치 = 0.6, 최소 지지도 = 0.5로 가정한다.)

$$0.6(3/5) + 0.6(1-0.6)(3/5) + 0.6(1-0.6)^2(1/5) = 0.5232$$

이 결과 값은 최소 지지도 0.5 이상이므로 (사과, 우유)는 빈발 항목집합이 된다. 즉 '사과와 우유 품목은 연관성을 가지고 있다'라고 볼 수 있다. 또한 최근 연도의 지지도에 가장 높은 가중치(0.6)를 줌으로써 최근 자료가 과거 자료보다 더욱 의미 있는 자료로 활용하였다.

최소 지지도를 넘는 빈발 항목집합들의 생성 과정은 Apriori 해법에서의 방법과 유사하다 [6]. 빈발 항목집합으로 결정된 항목들에 대해서는 각 기간동안의 증가 또는 감소 등 추세를 파악하게 된다. 예를 들어, 빈발 항목집합 {사과, 우유}의 1998년에서 1999년으로의 매출 성장이 40%, 1999년에서 2000년으로의 매출 성장이 20%로 나타났다고 한다면 {사과, 우유}의 평균 매출 증가율은 30%로 볼 수 있다.

자료를 시간에 따라 분류한 후, 가중치를 부여하고 각 항목집합의 지지도를 계산하여 항목집합의 전체 지지도를 계산하는 알고리즘은 다음과 같다.

#### 알고리즘 단계

[단계 1] 초기화.

- 자료 분할 개수를 결정한다(n).
- 가중치, 최소 지지도, 최소 신뢰도를 결정한다.

[단계 2] 자료 분할, 가중치, 지지도 계산.

- 시간의 흐름에 따라 자료를 분할 개수만큼 분할한다.
- 현재에서 과거로 시간이 지남에 따라 지수 적으로 감소하는 가중치를 사용하여 최근 자료일수록 높은 가중치를 부여한다.
- 분할한 자료에 대해 Apriori 방법으로 각 항목집합의 지지도를 구한다[6].

[단계 3] 분석.

- 각 항목집합에 대해 가중치 값과 지지도를 곱한다.
- 각 항목집합의 전체 지지도는 가중치가 적용된 지지도의 합이다.
- 각 항목 집합들에 대한 빈발 항목집합 여부를 결정한다.
- 빈발 항목집합에 대한 추세 변동률을 계산한다.

[단계 4] 연관규칙 발견.

- 빈발 항목집합으로부터 연관규칙을 찾는 방법은 기존 연관규칙 기법인 Apriori 방법과 동일하다[6].

### 3. 예제 및 분석

#### 3.1 예제

간단한 예제를 통해 기존의 연관규칙 기법과 본 연구에서 제안한 해법과의 차이점을 살펴 보자. 먼저 데이터 베이스에 다음과 같은 자료가 있다고 가정한다.

[표 1] Transaction Database

1998~2000년 매출실적			
TID	Items	TID	Items
1	1 2 4 5	2	1 4 5
3	2 4 5	4	1 2 4 5
5	1 3 5	6	1 2 3 5
7	3 4 5	8	2 3
9	2 3 4 5	10	1 2 4 5
11	1 2 3 4 5	12	2 3 5
13	1 2 3	14	1 2 4
15	2 3 5		

(TID : transaction identifier)

[표 1]의 자료를 가지고 기존의 Apriori 연관규칙 기법을 이용하여 구한 빈발 항목집합의 결과는 다음과 같다. ( 최소 지지도 = 0.5로 가정한다. )

- 1-빈발 항목집합 L1=(1, 2, 3, 4, 5 )
- 2-빈발 항목집합 L2={{2,5}, {4,5}}
- 3-빈발 항목집합 L3={ 항목집합이 없음 }

다음은 위의 자료를 연도별로 나누어 각각의 자료에 대해서 연관규칙 기법을 적용해 보았다. 먼저 연도별로 데이터를 나누어 보면 다음과 같다.

[표 2] An Annual Transaction Database

1998 매출실적		1999 매출실적		2000 매출실적	
TID	Bar codes	TID	Bar codes	TID	Bar codes
1	1 2 4 5	1	1 2 3 5	1	1 2 3 4 5
2	1 4 5	2	3 4 5	2	2 3 5
3	2 4 5	3	2 3	3	1 2 3
4	1 2 4 5	4	2 3 4 5	4	1 2 4
5	1 3 5	5	1 2 4 5	5	2 3 5

[표 2]의 각 연도별로 나눈 데이터를 기존의 연관규칙 기법을 통해 구한 각 연도의 빈발 항목집합 결과는 다음과 같다.( 최소 지지도 = 0.5로 가정한다. )

- 1998년 자료의 빈발항목집합: 1-빈발 항목집합 L1=(1, 2, 4, 5)
- 2-빈발 항목집합 L2={{1,4},{1,5},{2,4}, {2,5}, {4,5}}
- 3-빈발 항목집합 L3={{1,4,5}, {2,4,5}}
  
- 1999년 자료의 빈발항목집합: 1-빈발 항목집합 L1=(2, 3, 4, 5)
- 2-빈발 항목집합 L2={{2,3}, {2,5}, {3,5}, {4,5}}
- 3-빈발 항목집합 L3={ 항목집합이 없음 }

- 2000년 자료의 빈발항목집합: 1-빈발 항목집합  $L1=\{1, 2, 3, 5\}$   
 2-빈발 항목집합  $L2=\{\{1,2\}, \{2,3\}, \{2,5\}, \{3,5\}\}$   
 3-빈발 항목집합  $L3=\{\{2,3,5\}\}$

지금까지 설명한 예제들에서는 가중치를 고려하지 않았다. 다음 설명은 본 연구에서 제시하는 연도별로 가중치를 다르게 부여한 후 빈발항목을 구하는 방법이다. ( $\alpha = 0.6$ 으로 가정한다.)

[표 2]와 같이 시간의 흐름에 따라 자료를 분할하고 최근 자료일수록 높은 가중치를 부여한 후, 각 항목집합의 지지도를 구하는 단계 2를 거친 후, 단계 3에서 각 항목집합의 빈발 항목집합 여부와 추세 변동율을 구한다.

[표 2]의 자료에서 항목1의 빈발 항목집합 여부를 구하는 방법은 다음과 같다.  $\alpha(2000년\ 매출실적\ 자료에서의\ 항목1에\ 대한\ 지지도) + \alpha(1-\alpha)(1999년\ 매출실적\ 자료에서의\ 항목1에\ 대한\ 지지도) + \alpha(1-\alpha)^2(1998년\ 매출실적\ 자료에서의\ 항목1에\ 대한\ 지지도) > \text{최소 지지도}(0.5)$ 를 만족하는지 조사하면 다음과 같다.

$0.6*(3/5) + 0.6*(1-0.6)*(2/5) + 0.6*(1-0.6)^2*(4/5) = 0.5328 > 0.5$ 를 만족한다. 따라서 항목1은 빈발 항목집합이 된다. 나머지 항목 2, 3, 4, 5에 대해서도 가중치를 고려한 전체 지지도를 구해보면 다음과 같다.

- 항목 2의 지지도: **0.84** > 0.5  
 항목 3의 지지도: **0.73** > 0.5  
 항목 4의 지지도: 0.46 < 0.5  
 항목 5의 지지도: **0.65** > 0.5

항목4를 제외한 모든 항목들이 최소 지지도 0.5를 넘기 때문에 1-빈발 항목집합은  $\{1,2,3,5\}$ 가 된다.

다음 2-항목집합에 대한 빈발항목 집합여부는 다음과 같이 계산할 수 있다. 항목집합  $\{1,2\}$ 에 대한 지지도 계산은  $\alpha(2000년\ 매출실적\ 자료에서\ 항목\{1,2\}\의\ 지지도) + \alpha(1-\alpha)(1999년\ 매출실적\ 자료에서\ 항목\{1,2\}\의\ 지지도) + \alpha(1-\alpha)^2(1998년\ 매출실적\ 자료에서\ 항목\{1,2\}\의\ 지지도)$ 로 계산할 수 있으며, 최소 지지도를 만족하는지를 조사해 보면 다음과 같다.

$0.6*(3/5) + 0.6*(1-0.6)*(2/5) + 0.6*(1-0.6)^2*(2/5) = 0.4922 < 0.5$ . 항목집합 $\{1,2\}$ 는 최소지지도를 만족시키지 못하기 때문에 빈발 항목집합이 될 수 없다. 다음으로 2-항목집합  $\{1,3\}, \{1,5\}, \{2,3\}, \{2,5\}, \{3,5\}$ 에 대해서도 가중치를 고려한 전체 지지도를 구해보면 다음과 같다.

- 항목  $\{1,3\}$ 의 지지도: 0.31 < 0.5  
 항목  $\{1,5\}$ 의 지지도: 0.29 < 0.5  
 항목  $\{2,3\}$ 의 지지도: **0.62** > 0.5  
 항목  $\{2,5\}$ 의 지지도: **0.56** > 0.5  
 항목  $\{3,5\}$ 의 지지도: **0.52** > 0.5

따라서 2-빈발 항목집합은  $\{\{2,3\}, \{2,5\}, \{3,5\}\}$ 가 된다. 3-항목집합에 대해서도 위와 같은 방법으로 빈발 항목집합을 구한다. 결국, 본 연구에서 제시하고 있는 연도별로 가중치를 다르게 부여한 후 빈발 항목집합을 구한 결과는 다음과 같다.

- 1-빈발 항목집합  $L1=\{1, 2, 3, 5\}$
- 2-빈발 항목집합  $L2=\{(2,3), (2,5), (3,5)\}$
- 3-빈발 항목집합  $L3=\{\text{항목집합이 없음}\}$

### 3.2 예제 분석

예제를 분석한 결과, 다음과 같은 특징이 나타난다.

- 1) 기존의 방법으로 전체 데이터 베이스에서 구한 2-빈발 항목집합은  $\{2,5\}$ ,  $\{4,5\}$ 이다. 그리고 각 연도별로 가중치를 고려한 방법으로 구한 2-빈발 항목집합은  $\{2,3\}$ ,  $\{2,5\}$ ,  $\{3,5\}$ 가 나타났다. 즉 과거 연관관계가 있던 항목집합 $\{4,5\}$ 는 연관성이 줄어들어 최근에는 연관관계를 가지지 않게 되었고, 항목집합  $\{2,3\}$ ,  $\{3,5\}$ 는 가중치를 고려하지 않은 방법으로는 연관관계가 발견되지 않았으나, 시간에 따른 가중치를 고려한 방법에서는 연관관계가 나타났다. 이러한 결과는 최근 자료에 더욱 의미를 둔 가중치를 부여했기 때문이다.
- 2) 최근의 자료만을 살펴보면(2000년 자료만을 가지고 분석한 경우) 항목 $\{1,2\}$ 도 2-빈발 항목집합이 될 수 있으나 전체 데이터를 가지고 분석한다면 의미 있는 연관관계가 될 수 없다.(각 연도별 기존 연관규칙 기법을 통해 구한 빈발 항목집합 결과 참조)
- 3) 2-빈발 항목집합 $\{2,5\}$ 는 1998~2000년까지 꾸준히 연관관계를 유지하고 있다.

이 결과는 최근 자료일수록 높은 가중치를 주고 구한 결과로 최근의 자료가 과거의 자료보다 미래를 더 잘 반영한다는 전제하에서 전체 데이터 베이스 자료에 있는 추세 변화를 나타내고 있다. 하지만 어떤 빈발 항목집합이 가장 높은 증가추세를 보이고 있거나 꾸준히 연관관계를 가지고 있는지를 알 수가 없다. 따라서 빈발 항목집합에 대한 이러한 증가정도를 결과에 나타냄으로써 의사결정 시, 더욱더 많은 도움을 줄 수 있다.

위 예제의 경우, 2-빈발 항목집합  $\{2,3\}$ 은 1998년에서 1999년으로 42%증가, 1999년에서 2000년으로 14% 매출이 증가했음을 알 수 있다. 따라서 평균 증가율은  $(42+14)/2 = 28\%$ 로 나타낼 수 있다. 다른 2-빈발 항목집합들도 같은 방법으로 평균 증가율을 구해보면, 항목집합  $\{2,5\}$ 는 0%, 항목집합  $\{3,5\}$ 는 14.5%의 평균 증가율을 구할 수 있다. 본 연구에서 제시한 해법을 통해 얻어진 2-빈발 항목집합은 다음과 같이 표현할 수 있다.  $\{(2,3)-28\%증가\}$ ,  $\{(2,5)-0\%증가\}$ ,  $\{(2,3)-14.5\%증가\}$

따라서 가장 큰 증가추세를 가지고 있는 빈발 항목집합은  $\{2,3\}$ 이다.

### 4. 결론

본 연구는 시간 변화에 따라 데이터 베이스에 있는 자료를 분할한 후, 최근 자료에 높은 가중치를 줌으로써 최근 자료가 과거 자료보다는 미래에 대한 예측을 더 잘 반영할 수 있게 하였다. 시간적으로 나눈 분할 영역에서 항목집합들의 지지도를 각각 구한 후, 차등화 된 가중치를 지지도에 부여한다. 가중치가 적용된 각 분할영역의 지지도들을 더함으로써 전체 지지도를 구한다. 그리고 각 항목집합의 전체 지지도가 사용자가 미리 정의한 최소 지지도를 넘는지를 조사함으로써 빈발 항목집합 여부를 조사하게 된다. 빈발 항목집합은 서로 연관성이 높은 항목들으로써 우리가 관심을 가지게 되는 항목집합이 된다.

본 연구결과를 통해서 기존 연관규칙 기법으로는 발견할 수 없었던 자료의 추세 변화를 파악할 수 있었을 뿐 아니라, 경향을 수치로 표현함으로써 더 명확한 자료를 의사결정자에게 제공할 수 있었다. 의사결정자는 추세 변화를 지닌 이러한 연관규칙들을 통해 매출 증가 추세가 큰 빈발 항목집합들을 우선적으로 고려한 마케팅 전략을 수립할 수 있다. 상품 진열 시, 연관성이 있는 제품들을 근처에 위치시킴으로써 매출 증진을 가져올 수 있으며 카탈로그 디자인, 교차 판매 등에 있어서도 제품간의 연관성을 강조한 판매 전략을 세울 수 있다.

또한 현재에서 과거로 시간이 지남에 따라 지수 적으로 감소하는 가중치를 이용하여 미래 예측을 위한 소비자의 구매 패턴을 찾아낼 수 있었으며 이러한 결과는 기존의 연관 규칙 기법으로는 발견될 수 없었던 결과이다. 분석하고자 하는 자료의 추세를 제대로 파악하지 못한다면, 아무리 좋은 연관규칙 기법을 통해 마이닝을 한다 하더라도 예측 자료로서의 정확성이 떨어지게 된다. 따라서 각 기간동안 자료의 중요성을 나타내주는 가중치를 어떻게 정할 것인가도 중요한 문제가 된다. 앞으로 실험을 통해 효과적인 가중치 결정과 데이터 베이스 자료를 몇 개로 분할하여 가중치를 주는 것이 효과적인지에 관한 연구가 필요하다.

## 참고문헌

- [1]윤재홍, 생산 운영 관리론, 형설출판사, 1997.
- [2]A. Savasere, E. Omiecinski, and S. Navathe. "An efficient algorithm for mining association rules in large databases," In proceedings of the 2th International Conference on Very Large Data Bases, pp.432-444, 1995.
- [3]Bing Liu, Wynne Hsu and Yiming Ma, "Mining Association Rules with Multiple Minimum Supports," KDD-99, pp.337-341, 1999.
- [4]C.H. Cai, Ada W.C. Fu, C.H. Cheng and W.W. Kwong, "Mining Association Rules with Weighted Items," Proceedings of the International Database Engineering & Applications Symposium, pp.68-77, 1998.
- [5]J.S. Park, M-S. Chen, and P.S. Yu. "An effective hash-based algorithm for mining association rules," In Proceedings of ACM SIGMOD, pp.175-186, 1995.
- [6]R. Agrawal and R. Srikant., "Fast algorithm for mining association rules," In Proceedings of the 20th VLDB Conference, pp.487-499, 1994.