# 역 연관규칙을 이용한 타겟 마케팅

황준현* · 김재련**

# Target Marketing using Inverse Association Rule

Jun-Hyun Hwang* · Jae-Yearn Kim**

## 요약

Making traditional plan of target marketing based on Association Rule has brought restriction to obtain the target of marketing. This paper is to present Inverse Association Rule as a new association rule for target marketing. Inverse Association Rule does not use information about relation between items that customers purchase like Association Rule, but use information about relation between items that customers do not pruchase. By adding Inverse Assocition Rule to target marketing, we generate new marketing rule to look for new target of marketing. From new marketing rule, this paper is to show direct marketing about target item and indirect marketing about another item associated with target item to sell target item. The reason is that sales of the item associated with target item have an influence on sales of target item.

Key words : 역 연관규칙(Inverse Association Rule), 타겟 마케팅(Target Marketing)

## 1.서론

### 1.1 연구배경

"데이터마이닝이란 사용 가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 실제 경영의 의사결정 등을 위한 정보로 활용하고자 하는 것이다."

위의 정의에서 보는 바와 같이 데이터마이닝이란 우리가 알지 못하고 있었던 새로운 정보를 뽑아내서 경영에 활용하는 데 그 의의를 두고 있다. 이러한 정보를 뽑아내고자 하는 데이터마이닝 기법 중의 하나가 연관규칙이다. 연관규칙은 데이터 안에 종속하는 항목간의 종속관계를 찾아내는 작업이다. 규칙은 $X \rightarrow Y$ 의 형태로 표현된다. 각 고객이 구매한 항목의 데이터를 가지고 규칙을 생성시키기 때문에 앞의 규칙은 다음과 같이 해석된다. "X를 구매한 사람들은 Y를 구매하는 경향이 많다." 즉, 해석에서 보면 알 수 있듯이 규칙에 표현된 구매 항목들 간에 어떠한 연관관계가 있다는 것을 알려주는 것이다(더 자세한 내용은 기존 고찰 연구에서 설명한다). 이 규칙을 타겟 마케팅(Target Marketing)에 적용시켜 보면(타겟이 Y라고 할 때) X를 구매한 사람들은 Y를 구매하는 경향이 많기 때문에 Y를 판매하기 위해서는 X를 구매한 사람들 중에서 아직 Y를 구매하지 않은 사람들에게 Y를 마케팅할 수 있다는 결론이 나온다. 예와 같이 구매한 자료로 생성된 규칙을 경영에 활용하는 것이다. 하지만 구매한 자료만을 사용하기 때문에 타겟 대상이 제한될 수 있다. 본 논문은 새로운 타겟 대상을 발견하는 방안과 그 방안의 타당성을 밝힐 것이다.

### 1.2 연구목적

본 논문에서는 구매한 자료뿐만 아니라 구매하지 않은 자료를 이용하여 연관규칙에서 생성되지 않은 새로운 규칙을 생성시켜 경영 전략에 도움을 주는 것을 목적으로 한다. 구매하지 않은 항목을 이용하여 규칙을 생성시켜 타겟 마케팅에 적용시키는 방법에 대해서는 본문에 자세히 설명할 것이다. 우선 간단히 설명하면 구매하지 않은 항목을 이용한 규칙은 다음과 같은 형태이다. '$\sim X \rightarrow \sim Y$' 구매하지 않은 항목을 나타내기 위해서 역(inverse; ~)을 사용하였다. 이러한 역 항목집합(inverse itemset)들 간의 관계에 대한 규칙을 역 연관규칙(Inverse Association Rule; IAR)이라고 정의한다. 규칙을 해석해 보면 "X를 구입하지 않은 사람들은 Y도 구매하지 않는 경향이 많다."이다. 역 연관규칙 생성으로 구매하지 않은 항목들 간에 관계를 알 수 있다. 본문에서는 이와 같이 역 연관규칙을 생성하고 생성된 규칙으로 타겟 마케팅에 적용시키는 방법에 대하여 설명한다. 또한 이러한 방법으로 연관규칙에서는 뽑아낼 수 없는 새롭고 유용한 정보를 뽑아 내어 경영에 이용하는 방법을 예를 통하여 밝힐 것이다.

2장에서는 기존고찰연구를 통하여 본 논문의 기초가 되는 알고리즘에 대해 설명한다. 2장에서 설명한 용어가 논문 전체에 걸쳐 계속 사용된다. 3장에서는 타겟 마케팅에 적용하기 위해 제안한 규칙을 설명하고 4장에서는 제안한 규칙을 적용하는 절차에

* 한양대학교 산업공학과 석사과정
** 한양대학교 산업공학과 교수

대하여 설명한다. 5장에서 예제를 통하여 마케팅 전략에 있어서 기존 정보 외에 새로운 정보를 생성하는 결과를 생성함으로써 6장의 결론을 이끈다.

## 2. 기존연구고찰

기존에 발표된 연관규칙에 관한 논문들은 구매 항목 간의 연관관계를 찾는 방법을 개선시키는 알고리즘, 즉 구매 항목들의 빈발 항목 집합들을 좀 더 빠르고 효율적으로 찾는 방법에 대해 많이 다루고 있다. Max-miner를 이용하여 가장 긴 빈발 항목 집합들부터 찾아 나가는 방법 (Roberto et al. 1998)과 closure개념을 이용하여 Frequent Closed Itemset(FCI)을 찾아가는 방법(Nicolas et al. 1999)이 그 예이다. 위의 FCI 개념을 이용하여 발전시킨 논문은 Closet(Pei et.al. 2000)과 Charm(Mohammed et.al. 1999)이 있다. 동일한 데이터베이스에서 소개한 각 알고리즘이 생성한 연관관계에 대한 규칙은 같다. 규칙보다는 데이터베이스를 스캔(scan)하는 시간을 줄여 빠르게 빈발 항목 집합들을 찾는 데에 관심을 가지고 있는 것이다. 이 장에서는 먼저 본문 내용의 기초가 되는 Apriori알고리즘에 대해 설명한다. 또한 본 논문이 연관규칙에서 나오는 구매한 정보로부터의 규칙과 새로운 규칙인 구매하지 않은 정보로부터의 규칙을 이용하여 타겟 마케팅에 이용하는 방법을 제안하므로 새로운 규칙을 찾는 한 방법인 Dissociation Rule에 대해서 소개한다.

### 2.1. Apriori 알고리즘을 이용한 연관규칙 (Agrawal et.al. 1994)

$I = \{i_1, i_2, ..., i_m\}$는 항목(item)이라고 하는 문자들의 집합이다. D를 트랜잭션 들의 집합(set)이라고 하고 각 트랜잭션 T는 $T \subseteq I$ 를 만족하는 항목들의 집합이다. 각 T는 TID라고 하는 유일한 식별자가 있다. 만약 $X \subseteq T$ 가 성립하면 I의 부분집합으로 구성되어 있는 X를 트랜잭션 T가 포함한다는 것을 뜻한다. 우리는 연관규칙을 $X \subset I$ , $Y \subset I$ , and $X \cap Y = \phi$ 인 상황에서 $X \Rightarrow Y$ 의 형태로 나타낸다. X를 포함하고 있는 D의 트랜잭션들의 c%가 또한 Y를 포함하고 있다면 $X \Rightarrow Y$ 는 c의 신뢰도(confidence)를 가지고 있다고 말한다.

D에 있는 트랜잭션의 s%가 $X \cup Y$ 를 포함하고 있다면 $X \Rightarrow Y$ 는 s의 지지도(support)를 가지고 있다고 말한다. 트랜잭션 집합 D에서 연관규칙을 찾아내는 문제는 사용자가 정의한 최소지지도(minimum support) min_sup와 최소신뢰도(minimum confidence) min_conf보다 큰 지지도와 신뢰도를 갖는 모든 연관규칙을 찾는 것이다. 최소지지도를 만족하는 항목집합을 빈발 항목집합(frequent or large itemset)이라고 부른다. k개의 항목으로 이루어진 빈발 항목집합을 빈발 k-항목집합(large k-itemset)이라고 한다.빈발 k-항목집합들의 집합을 $L_k$라 하고 이를 생성하기 위한 후보 항목집합들의 집합을 $C_k$라 한다.

알고리즘의 첫 번째 시행에서는 빈발 1-항목집합을 결정하기 위해 데이터베이스를 검색하여 각 항목별로 빈도수를 계산한다. k( $k \geq 2$ )번째 시행부터는 두 단계로 분할하여 알고리즘이 진행된다. 먼저, (k-1)번째 검색에서 발견된 빈발 항목집합 $L_{k-1}$ 으로 후보 항목집합 $C_k$를 만든다. 다음으로, 데이터베이스를 검색하여 $C_k$에 있는 후보 항목집합의 지지도를 계산한다. $C_k$에 있는 후보 항목집합 중에 최소지지도를 만족시키는 항목만 $L_k$에 진입한다.

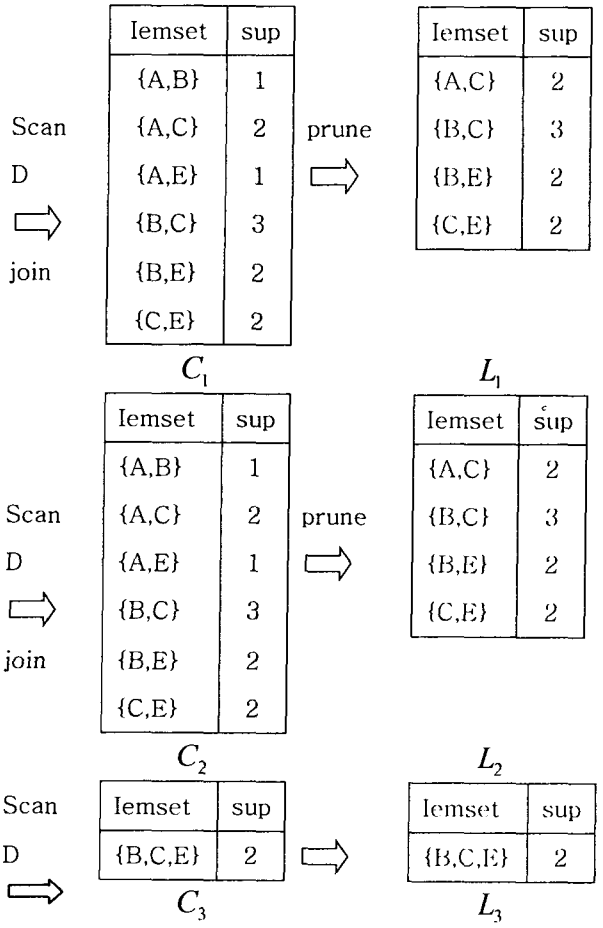
이러한 시행은 $L_k$가 더 이상 발견되지 않을 때까지 반복한다. Apriori에서 가장 중요한 부분이 apriori-gen 함수[8]는 join단계와 prune단계로 구성되어 있다.

[표 1]에는 연관규칙을 찾는 예제 데이터베이스가 주어져 있고, Apriori 알고리즘을 이용해 모든 후보 항목집합과 빈발 항목집합을 찾는 전과정이 [그림 1]에 나타나 있다.

[표 1]을 보면 트랜잭션이 4개이고 항목은 5개이다. 최소지지도는 50%로 한다.

| TID | Item |
|---|---|
| 100 | A C D |
| 200 | B C E |
| 300 | A B C E |
| 400 | B C |

<표 1> Apriori 예제 데이터베이스




[그림 1] Apriori 알고리즘 실행예제

위의 예제를 보면, 첫번째 데이터베이스 검색 시에 각 항목의 지지도를 계산하고 그 중에서 최소지

지도를 만족하는 항목만을 $L_1$ 으로 진입시킨다. Join단계와 Prune단계를 거쳐 $C_2$를 만든다. 다시 데이터 베이스를 검색하여 $C_2$의 지지도를 계산하여 역시 최소지지도를 넘기는 후보 항목집합만을 가지고 $L_2$를 만든다. 이런 과정을 계속 반복하여 $L_3$, $L_4$, $L_5$를 만든다. 이 예제의 경우 $L_4$에서 공집합이 되므로 알고리즘을 종료한다. 최종적으로 사용자가 얻을 수 있는 빈발 항목집합은 $L_1$ = { {A}, {B}, {C}, {E} }, $L_2$ = { {A,C}, {B,C}, {B,E}, {C,E} } 그리고 $L_3$ = { {B,C,E} }이다.

Apriori 알고리즘을 설명한 이유는 본 논문의 기초가 되기 때문이다. 또한 여기서 설명한 용어와 기호들은 본문에서도 똑같이 사용되며 같은 의미로 사용된다.

### 2.2 Dissociation rule (Michael et.al. 1997)

Dissociation rule은 연결자 "and not"을 가진다는 것을 제외하고는 연관규칙과 유사하다. 전형적인 Dissociation rule은 다음과 같다.

"If A and not B then C"

이 규칙을 해석해 보면 'A를 구매하고 B를 구매하지 않으면 C를 구매한다.'이다. 구매하지 않은 정보를 구매한 정보와 같이 고려하여 규칙을 만드는 것이다. Dissociation rule은 시장바구니 분석의 간단한 응용으로 만들어질 수 있다. 그 응용은 각각의 항목(item)의 역(inverse)인 새로운 항목집합을 소개하는 것이다. 전체 항목집합 중에서 각 트랜잭션(transaction)에서 없는 항목들은 역 항목을 취하도록 트랜잭션을 새롭게 구성한다. 예를 들면, 그림 2는 몇몇 트랜잭션들의 변형을 보여준다. 항목 앞에 ~는 역(inverse) 항목을 의미한다.

| Customer | Items | | Items |
|---|---|---|---|
| 1 | {A, B, C} | | {A, B, C} |
| 2 | {A} | → | {A, ~B, ~C} |
| 3 | {A, C} | → | {A, ~B, C} |
| 4 | {A} | → | {A, ~B, ~C} |
| 5 | {} | → | {~A, ~B, ~C} |

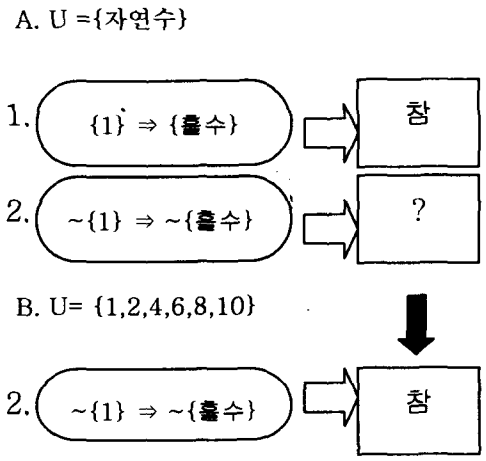
[그림2] Disssociation Rule을 생성하기 위한 트랜잭션의 변형

이러한 트랜잭션의 변형은 구매한 항목과 구매하지 않은 항목을 같이 고려하므로 분석에 사용되어지는 항목의 전체 수가 두 배가 된다. 계산량은 항목의 수에 지수적으로 증가하기 때문에 항목의 수가 두 배가 되는 것은 수행성을 현저하게 떨어뜨리게 된다. 본 논문에서는 Dissociation Rule에서와 같이 혼합된 항목으로 구성하여 정보를 얻어내는 것이 아니라 구매하지 않은 항목만으로 트랜잭션을 구성한다. 또한 목적이 규칙의 생성 그 자체가 아니라 규칙을 생성한 후 타겟 마케팅에 적용시키는 방법과 그 의미이다.

## 3. 타겟 마케팅에 적용하는 새로운 규칙 설명

### 3.1 연관규칙과 역 연관규칙 생성

"1이면 홀수이다" 는 참이 된다. 그 이유는 {1} ⇒ {홀수}에서 1이라는 집합이 홀수라는 집합에 포함되기 때문이다. "1이면 홀수이다"가 참이라고 해서 "1이 아니면 홀수가 아니다"라는 명제가 참이 되는 것은 아니다. 그 이유는 ~{1} ⇒ ~{홀수}에서 {1}의 여집합이 {홀수}의 여집합에 포함되지 않기 때문이다. 어떤 명제가 참이라면 그 "대우"는 참이 되지만 "이" 관계에 있는 명제는 참이라고 할 수 없다. (위의 두 명제는 "이" 관계에 있다)




[그림3] ~{1} ⇒ ~{홀수}이 참이 되는 전체집합

하지만 [그림3]에서 보는 바와 같이 특별한 경우의 전체집합 즉 B의 전체집합에서는 ~{1} ⇒ ~{홀수}의 명제는 참이 된다. [그림3]에서 1번과 2번이 모두 참이 되면 즉, 연관규칙과 역 연관규칙이 모두 생성되면 [그림4]와 같은 규칙이 성립된다. 즉, 전체집합이 U = { 1, 2, 4, 6, 8, 10}일 때에는 우변에 홀수가 되기 위해서 좌변에 {1}을 넣어야 하는 것이다.

이러한 개념을 타겟 마케팅에 응용하려는 것이 본 논문의 취지이다. 여기서 전체집합을 데이터베이스로 확장시키고 규칙은 위의 내용과 같이 생성시키는 데 다만, 신뢰도(confidence)가 위의 예에서와 같이 100%를 만족하는 규칙을 생성하는 게 아니라 최소 신뢰도를 만족하는 규칙을 생성한다.

뒤에서 더 자세히 설명하겠지만 여기서 간단히 위의 개념을 타겟 마케팅에 이용하여 마케팅 전략을 세워 보는 방법에 관하여 설명하면 B라는 품목이 타겟 항목일 때에 A ⇒ B가 성립한다면, 기존의 연관규칙의 정보에서는 A를 구매한 고객 중 아직 B를 구매하지 않은 고객을 대상으로 B를마케팅 하는 전략이 나오겠지만 (위의 예에서 보면 {1} ⇒ ? 에서 우변에 {홀수}를 넣는 의미) A ⇒ B와 ~A ⇒ ~B가 같이 고려하여 둘 다 성립한다면,

A ⇒ B의 마케팅 전략뿐만 아니라 B의 판매를 촉진하기 위해서 A를 먼저 판매 유도하는 전략도 나올 수 있다(? ⇒ {홀수}에서 좌변에 {1}을 넣는 의미). 결국 A의 마케팅이 B의 판매를 유도하는 셈이 된다. 다음 장에서 벤 다이어그램을 통하여 이러한 의미를 더 자세히 설명한다.

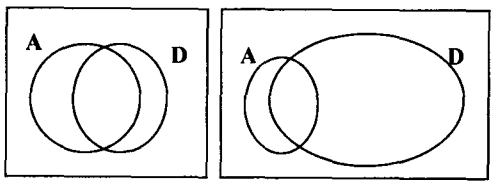

위의 그림이 성립하기 위하여 좌변에 넣어야 할 집합은 무엇인가? 답 => 1


[그림4] 새로운 규칙 생성

## 3.2 벤 다이어그램을 통해 역 연관규칙 첨가의 의미 이해

A ⇒ B가 규칙으로 생성되었을 때에 이 규칙만 보고서 B를 판매하기 위하여 A를 먼저 판매해야 한다는 마케팅전략을 세울 수는 없다. 반드시 A ⇒ B와 ~A ⇒ ~B 규칙이 같이 생성되어야만 A를 먼저 판매해야 한다는 전략이 타당하게 된다. 이 절에서 벤 다이어그램을 이용하여 이 가정이 타당하다는 것을 증명한다.

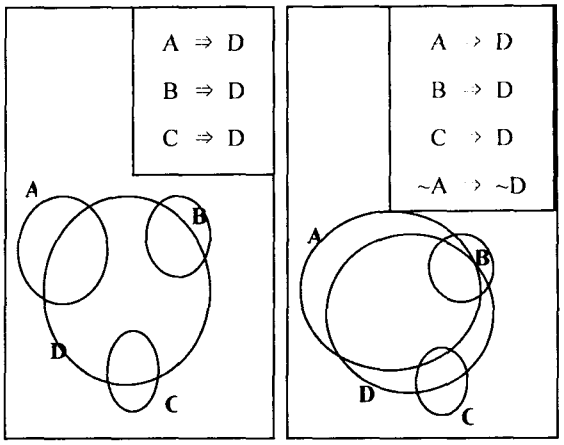
A ⇒ B가 규칙으로 생성되었을 경우, 그 규칙은 A를 구매한 사람들이 B를 구매하는 경향이 많다라는 것이지 B에 대한 확실한 정보가 나온 것은 아니다. [그림 5]에서 보듯이, A ⇒ B 규칙이 나오는 경우는 여러 가지가 될 수 있다. 그림은 그 중 두 가지 경우를 보여준다.




[그림 5] A ⇒ D가 규칙으로 나올 수 있는 두 가지 경우

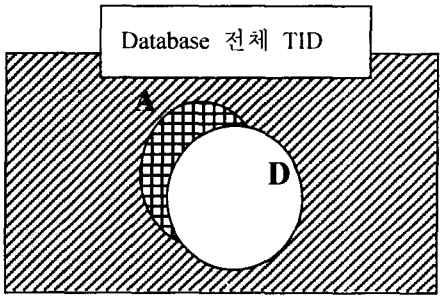
즉, 위의 오른쪽 그림을 보면 A를 구매한 사람들이 D를 구매하는 경향이 있는 것은 확실하지만 D를 구매한 사람들이 A도 구매하는 경향이 많다라고는 말할 수 없는 것이다. 다시 말하면 D의 판매가 A의 영향을 받는다고 말할 수 없는 것이다. 다음 그림을 보자.

[그림6]의 왼쪽 그림에서 A와 B와 C의 구매는 D의 영향을 받는다는 것을 알 수 있지만 연관규칙에서 나온 규칙으로는 D가 어떤 항목에 의해 영향을 많이 받는지는 알 수 없다




[그림6] D를 타겟으로 세 가지 규칙이 생성되었을 경우

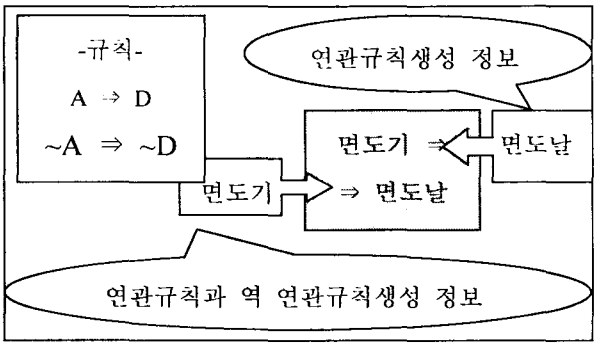
기존의 규칙이외에 ~A ⇒ ~D의 규칙 추가로 새로운 정보를 얻을 수 있다. 즉, D의 판매는 A의 판매에 의해 종속 받는다는 정보이다. 그러므로 연관규칙에 역 연관규칙을 첨가 시켜 생성시킬 수 있는 마케팅 전략은 A와 B와 C를 구입하고 아직 D를 구입하지 않은 고객에게 타겟 항목인 D를 바로 마케팅 하는 것뿐만 아니라 A와 D의 종속성으로 인해 타겟 항목인 D를 판매하기 위해서는 A의 판매가 이루어져야 하므로 먼저 A를 마케팅 하는 전략도 가능하게 된다. 결국 A의 마케팅전략이 B의 판매를 유도한다. 연관규칙에서 생성되지 않는 정보인 ~A ⇒ ~D 규칙 첨가로 우리는 타겟 마케팅에 새로운 타겟 대상을 다음 그림과 같이 찾을 수 있다. ( 편의상 A ⇒ D 와 ~A ⇒ ~D 만 고려했을 때)




[그림 7] A ⇒ D 와 ~A ⇒ ~D 규칙 생성시 고려해야 할 대상들

위의 빗금 중 ⊞ 부분은 연관규칙 생성으로 고려되어지는 타겟 대상이 되고 연관규칙과 역 연관규칙을 같이 고려하면 타겟 대상이 ⊞ 부분뿐만 아니라 ▨ 부분으로 확장된다. 즉, 연관규칙과 역 연관규칙이 같이 생성되었을 때는 빗금 친 부분에서

▦ 부분에 있는 고객에게는 D품목을 판매하기 위해서 D를 바로 마케팅 하고 ▨부분에 있는 고객에게는 D품목을 판매하기 위해서 우선 A품목을 마케팅 한다. 그런 후에 A를 구매한 고객을 대상으로 D품목을 마케팅 한다. A를 구매한 고객은 D를 구매하는 경향이 많기 때문에 A의 판매가 결국은 D의 판매를 촉진시키는 결과를 발생시킨다.




[그림8] 역 연관규칙 첨가한 예제

"면도기를 구매한 사람은 면도날을 구매하는 경향이 많다"라는 정보(A ⇒ D)가 의미 있는 정보이고 "면도기를 구매하지 않은 사람은 면도날을 구매하지 않는 경향이 많다"라는 정보( ~A ⇒ ~D)가 의미 있는 정보라면 면도날을 팔기 위해서는 면도기를 구매한 사람들 중 아직까지 면도날을 구매하지 않은 사람들에게 면도날을 마케팅 하여 구매 유도하는 것뿐만 아니라 면도기를 구매치 않은 사람에게 우선 면도기를 구매케 한 다음 면도날을 구매케 하는 마케팅 전략도 필요하다.

연관규칙은 좌변에 있는 항목을 고정시키고 우변에 있는 항목을 마케팅 하는 반면 역 연관규칙을 첨가하면 우변에 있는 항목을 고정시키고 좌변에 있는 항목을 마케팅 하는 전략도 가능하게 되는 것이다. 즉, 어떤 제품을 판매하기 위해서는 그 제품만 마케팅 하는 것 뿐만 아니라 더 나아가 그 제품이 아닌 다른 어떤 제품을 마케팅 하는 것이 필요할 수 있는 것이다. 이러한 관계를 역 연관규칙의 첨가로 알아낼 수 있다.

## 3.3 연관규칙과 역 연관규칙 신뢰도 비교

연관규칙과 역 연관규칙의 신뢰도를 비교하기 위해 다음 간단한 예제를 이용한다. A, B 두 항목에 대해서만 데이터베이스를 스캔하여 표를 구성하면 아래와 같다. 여기서 O는 트랜잭션에서 항목을 구입한 경우이고 X는 항목을 구입하지 않은 경우이다.

| 구매여부 | A | B | count |
|---|---|---|---|
| 1 | O | O | 20 |
| 2 | O | X | 10 |
| 3 | X | O | 20 |
| 4 | X | X | 50 |

<표2> 구매 여부에 따라 TID수를 합한 테이블

B를 타겟으로 할 때 연관규칙에서 나올 수 있는 규칙은 A ⇒ B 이고 신뢰도는 다음과 같이 구한다.

$$\text{conf}(A \Rightarrow B) = \frac{\sup(A \cup B)}{\sup(A)} = 20/30$$

여기서 sup(A)는 A를 구매한 트랜잭션의 수이기 때문에 [표 . 3]에서 구매여부 1, 2를 합한 지지도(count) 30이고 $\sup(A \cup B)$ 는 A, B를 모두 구매한 트랜잭션의 수이기 때문에 구매여부 1의 지지도 20이 된다.

B를 타겟으로 할 때 역 연관규칙에서 나올 수 있는 규칙은 ~A ⇒ ~B 이고 신뢰도는 또한 다음과 같이 구한다.

$$\text{conf}(\sim A \Rightarrow \sim B) = \frac{\sup(\sim A \cup \sim B)}{\sup(\sim A)} = 50/70$$

여기서 sup(~A)는 A를 구매하지 않은 트랜잭션의 수이기 때문에 [표 3]에서 구매여부 3, 4를 합한 지지도(count) 70이고 $\sup(\sim A \cup \sim B)$ 는 A, B를 모두 구매하지 않은 트랜잭션의 수이기 때문에 구매 여부 4의 지지도 50이 된다.

연관규칙에서 나온 규칙에 대한 신뢰도가 최소신뢰도를 만족하면 그 규칙을 유의한 정보라고 생각하고 마찬가지로 역 연관규칙에서 나온 규칙에 대한 신뢰도도 최소신뢰도(연관규칙의 최소신뢰도 값과 같을 필요가 없다)를 만족하면 그 규칙을 유의한 정보라고 생각한다.

# 4. 제안하는 알고리즘

먼저 새로운 용어에 대해 용어 정의를 하고 앞에서 소개한 연관규칙과 역 연관규칙 생성으로 타겟 마케팅에 적용시키는 방법에 대해 알고리즘을 전개한다.

## 4.1 용어 정리

알고리즘을 전개하기 위하여 필요한 새로운 개념에 대해 용어를 정의한다.

### 4.1.1 역 데이터베이스(Inverse Database)

기존 데이터베이스에서는 TID(트랜잭션 번호)의 항목집합을 구매한 항목으로 나타낸다. TID의 항목집합을 구매하지 않은 항목으로 나타낸 데이터베이스를 역 데이터베이스라고 정의한다. 즉 [그림9]와 같다.

### 4.1.2 역 항목(Inverse Item)

구매하지 않은 항목을 역 항목(Inverse Item)이라 한다. A를 항목이라고 하면 A의 역 항목은 ~A라 표시한다.

### 4.1.3 역 연관규칙(Inverse Association Rule)

연관규칙(Association Rule)은 구매한 항목에 관한

연관관계를 규칙으로 표시한 것이다. 이 규칙과 구별이 될 수 있도록 역 연관규칙을 정의하여 사용한다. 역 연관규칙이란 역 데이터베이스에서 나온 비 구매항목에 관한 규칙이다. 즉, 구매하지 않은 품목간의 패턴을 규칙으로 표시한 것이다. 역 항목간의 관계 ~A => ~B의 형태로 표현된다.

| TID | Itemset |
|---|---|
| 1 | ABCF |
| 2 | ABCDEF |
| 3 | CDF |
| 4 | ACD |
| 5 | DE |
| 6 | ABDF |
| 7 | BF |
| 8 | BCD |
| 9 | DF |
| 10 | ABCDF |

| TID | Itemset |
|---|---|
| 1 | ~D~E |
| 2 | |
| 3 | ~A~B~E |
| 4 | ~B~E~F |
| 5 | ~A~B~C~F |
| 6 | ~C~E |
| 7 | ~A~C~D~E |
| 8 | ~A~E~F |
| 9 | ~A~B~C~E |
| 10 | ~E |

[그림9] 역 데이터베이스로의 전환

## 4.1 타겟 마케팅에 적용하기 위한 절차

본 논문에서 제안하는 바가 타겟 마케팅에 적용되기 위해서는 연관규칙과 역 연관규칙이 같이 생성되어야 한다. 그러므로 [단계2]와 [단계3]을 통해서 연관규칙과 역 연관규칙을 생성한다. 절차를 요약하면 다음과 같다.

[단계1] 데이터베이스를 역 데이터베이스로 전환한다.

[단계2] 역 데이터베이스에서 빈발항목집합을 생성한다.
이 단계는 연관규칙을 생성하기 위한 단계로 빈발항목집합을 생성하기 위해서는 역 항목으로 구성된 역 데이터베이스에서 고려하는 항목의 역 항목이 없는 트랜잭션의 수를 세면 된다. 예를 들면 A의 지지도는 ~A가 없는 트랜잭션의 수를 세면 되는 것이다.

[단계3]역 데이터베이스에서 역 빈발항목집합을 생성한다(Inverse Frequent Itemset).
이 단계는 역 연관규칙을 생성하기 위한 단계로 역 빈발항목집합을 생성하기 위해서는 역 항목으로 구성된 역 데이터베이스에서 고려하는 역 항목이 있는 트랜잭션의 수를 세면 된다. 단, 단계 2에서 빈발항목집합으로 생성된 집합만을 고려한다. 그 이유는 타겟 마케팅에 적용 시에 어느 정도 구매력이 있는 항목들만을 고려하기 위해서다. 빈발 항목 집합들을 생성시키는 단계2는 구매력이 있는 항목집합 들을 찾아나가는 과정이다.

[단계4] 타겟 마케팅에 필요한 규칙을 생성한다.
타겟 항목에 필요한 연관규칙과 역 연관규칙을 생성하고 이러한 규칙을 이용하여 마케팅전략을 세운다.

단계 2의 역 데이터베이스에서 빈발항목집합을 찾는 과정을 좀 더 자세히 살펴 보면 [그림10]과 같다.

단계 1. [$L_1$ 을 발견]
- 역 데이터베이스를 액세스하여 1의 항목 중 고려하는 항목의 역 항목이 각 트랜잭션에 포함되지 않은 지지도(miss count)가 최소지지도보다 같거나 큰 지지도를 갖는 항목을 $L_1$에 등록한다.
- k ←2

단계 2. [$C_k$ 를 생성]
- $L_{k-1}$로부터 apriori-gen (Han et. Al. 2000)을 사용하여 $C_k$를 생성한다.
- 만약 $C_k$가 $\phi$이면 단계 4로 간다.

단계 3. [$C_k$의 지지도를 계산]
- 역 데이터베이스를 액세스하여 후보 k-항목집합의 지지도를 계산할 때 후보 k-항목집합의 역 항목집합이 각 트랜잭션에 포함되지 않은 지지도(miss count)를 계산한다. (여기서 후보 k-항목집합의 역 항목집합이 각 트랜잭션에 포함되지 않는다는 것은 후보 k-항목집합의 역 항목집합이 역 데이터베이스의 각 트랜잭션의 항목집합과 교집합이 없다는 것을 의미한다.)

단계 4. [$L_k$를 발견]
- 최소지지도보다 같거나 큰 지지도(miss count)를 갖는 후보 k-항목집합을 $L_k$에 등록한다.
- k ← k+1로 설정하고 단계 1로 간다.

단계 5. [끝냄]
- 끝낸다.

[그림10] 역 데이터베이스에서 빈발항목집합 발견하는 절차

단계3의 역 데이터베이스에서 역 빈발항목집합을 찾는 과정을 좀 더 자세히 살펴 보면 그림과 같다.

단계 1. [inverse $L_1$ 을 발견]
- 역 데이터베이스를 액세스하여 1의 항목 중 고려하는 역 항목이 각 트랜잭션에 포함된 지지도가 최소지지도보다 같거나 큰 지지도를 갖는 항목을 inverse $L_1$에 등록한다. 단, 단계 2을 만족하면 inverse $L_1$에 등록하지 않는다.
- k ←2
- 단계 3으로 간다.

단계 2 [빈발항목집합을 만족하는 항목의 역 항목만을 고려](다음 조건을 만족하는 역 항목 제외시킴)
역 데이터베이스에서 역 항목집합의 지지도
≥ 전체 TID 수 -(원 데이터베이스 최소지지도 -1)

단계 3. [inverse $C_k$를 생성]
- inverse $L_{k-1}$로부터 apriori-gen을 사용하여 inverse $C_k$를 생성한다.
- 만약 inverse $C_k$가 $\phi$이면 단계 4로 간다.

단계 4. [inverse $C_k$의 지지도를 계산]
- 역 데이터베이스를 액세스하여 후보 inverse k-항목집합이 각 트랜잭션에 포함된 지지도를 계산한다.

단계 5. [inverse $L_k$를 발견]
- 최소지지도보다 같거나 큰 지지도를 갖는 후보 inverse k-항목집합을 inverse $L_k$에 등록한다. 단, 단계 2를 만족하면 inverse $L_k$에 등록하지 않는다.
- k ← k+1로 설정하고 단계 3으로 간다.

단계 6. [끝냄]
- 끝낸다.

[그림11] 역 데이터베이스에서 역 빈발항목집합 발견하는 절차

## 5. 예제

본 장에서는 연관규칙과 역 연관규칙을 생성하고 생성된 규칙들을 타겟 마케팅에 적용시키는 방법을 예제를 통하여 설명한다. 5.1에서는 빈발항목집합을 생성하여 타겟 항목에 대한 연관규칙을 생성하는 방법에 대하여 설명한다. 4.2에서 설명한 타겟 마케팅을 수립하기 위한 두 번째 단계에 해당한다. 5.2에서는 역 빈발항목집합을 생성하여 타겟 항목에 대한 역 연관규칙을 생성하는 방법에 대하여 설명한다. 4.2에서 설명한 타겟 마케팅을 수립하기 위한 세 번째 단계에 해당한다.

### 5.1 역 데이터베이스에서 빈발항목집합과 연관규칙 생성

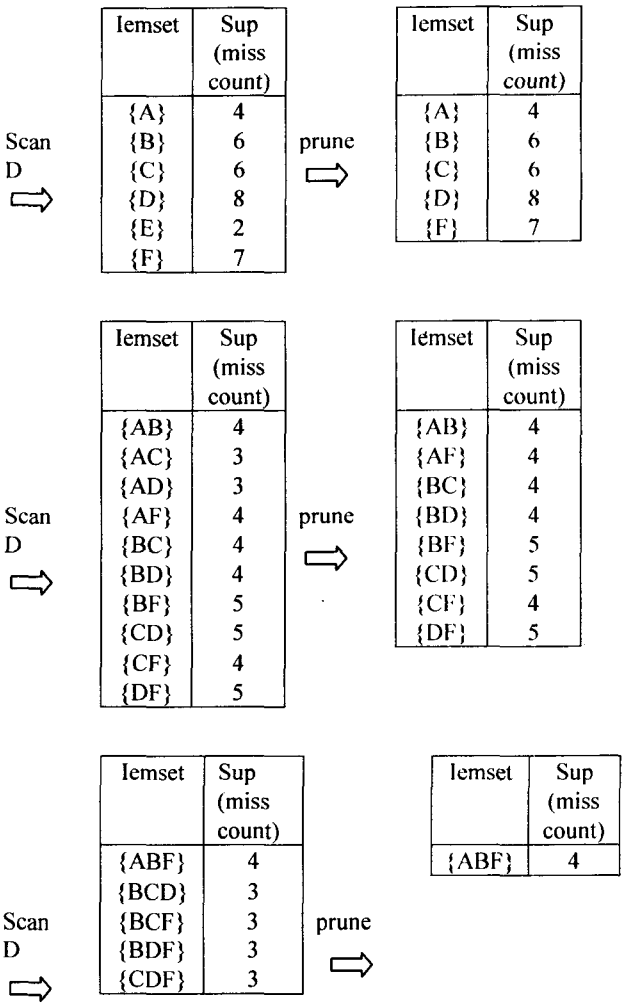
빈발항목집합과 역 빈발항목집합을 생성하기 위해서는 먼저 데이터베이스를 역 데이터베이스로 전환해야 한다. 4.2에서 설명한 첫번째 단계에 해당한다.

* 역 데이터베이스는 트랜잭션(TID)을 각 고객이 구매하지 않은 항목으로 나타낸 데이터베이스를 말한다.

먼저 역 데이터베이스에서 빈발항목집합을 생성한다. 4장에서 밝힌 바와 같이 각 항목의 지지도를 구하기 위해서는 역 데이터베이스에서 그 항목의 역 항목이 없는 트랜잭션을 센다. 이와 같이 세는 방법을 항목의 miss count를 센다고 정의한다. 예를 들면 항목 A의 지지도는 ~A가 없는 트랜잭션들 1, 2, 6, 10의 개수 4이다. 항목 집합 {CD}의 지지도는 ~C와 ~D가 모두 없는 트랜잭션들 2, 3, 4, 8, 10의 개수 5이다. [그림12]는 역 데이터베이스에서 고려하는 항목집합의 miss count를 세어 빈발항목집합을 구하는 과정을 요약한 그림이다. 빈발항목집합은 역 데이터베이스를 스캔하여 얻은 항목집합의 지지도가 최소지지도를 만족하면 생성된다. 최소지지도는 40%이다.

| TID | Itemset |
|---|---|
| 1 | ~D~E |
| 2 | |
| 3 | ~A~B~E |
| 4 | ~A~B~E~F |
| 5 | ~A~B~C~F |
| 6 | ~C~E |
| 7 | ~A~C~D~E |
| 8 | ~A~E~F |
| 9 | ~A~B~C~E |
| 10 | ~E |

<표3> 역 데이터베이스




| Itemset | Sup (miss count) |
|---|---|
| {A} | 4 |
| {B} | 6 |
| {C} | 6 |
| {D} | 8 |
| {E} | 2 |
| {F} | 7 |

| Itemset | Sup (miss count) |
|---|---|
| {A} | 4 |
| {B} | 6 |
| {C} | 6 |
| {D} | 8 |
| {F} | 7 |

| Itemset | Sup (miss count) |
|---|---|
| {AB} | 4 |
| {AC} | 3 |
| {AD} | 3 |
| {AF} | 4 |
| {BC} | 4 |
| {BD} | 4 |
| {BF} | 5 |
| {CD} | 5 |
| {CF} | 4 |
| {DF} | 5 |

| Itemset | Sup (miss count) |
|---|---|
| {AB} | 4 |
| {AF} | 4 |
| {BC} | 4 |
| {BD} | 4 |
| {BF} | 5 |
| {CD} | 5 |
| {CF} | 4 |
| {DF} | 5 |

| Itemset | Sup (miss count) |
|---|---|
| {ABF} | 4 |
| {BCD} | 3 |
| {BCF} | 3 |
| {BDF} | 3 |
| {CDF} | 3 |

| Itemset | Sup (miss count) |
|---|---|
| {ABF} | 4 |

[그림12] 빈발항목집합 생성과정

생성된 빈발항목집합은 다음과 같다.

| | |
|---|---|
| 1빈발 항목 집합 | A(4), B(6), C(6), D(8), F(7) |
| 2빈발 항목 집합 | AB(4), AC(3), AD(3), AF(4), BC(4), BD(4), BF(5), CD(5), CF(4), DF(5) |
| 3빈발 항목 집합 | ABF(4) |

B를 타겟으로 하는 마케팅을 한다고 가정하면 B항목에 대한 최소신뢰도를 만족하는 연관규칙은 다음과 같이 생성된다. 최소신뢰도는 60%이다.

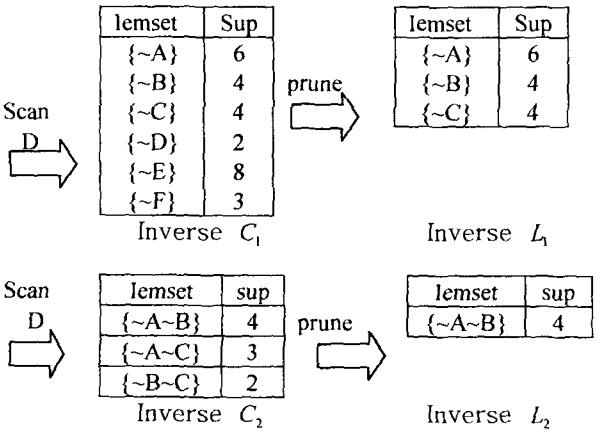
| |
|---|
| A ⇒ B |
| C ⇒ B |
| F ⇒ B |
| AF ⇒ B |

5.1의 절차는 데이터베이스를 역 데이터베이스

로 전환시켜 사용한다는 것과 지지도를 세는 방법이 miss count로 센다는 것을 제외하면 나머지는 apriori 알고리즘과 방법이 같다.

## 5.2 역 데이터베이스에서 역 빈발항목집합과 역 연관규칙 생성

역 데이터베이스에서 역 빈발항목집합을 생성한다. 4장에서 밝힌 바와 같이 각 역 항목의 지지도(support)를 구하기 위해서는 역 데이터베이스에서 역 항목이 있는 트랜잭션을 센다. 예를 들면 항목 ~A의 지지도는 ~A가 있는 트랜잭션들 3, 4, 5, 7, 8, 9의 개수 6이다. 항목 집합 {~A ~C}의 지지도는 ~A와 ~C가 모두 있는 트랜잭션들 5, 7, 9의 개수 3이다. 아래의 그림은 역 데이터베이스에서 역 빈발항목집합을 구하는 과정을 요약한 그림이다




[그림13] 역 데이터베이스에서
역 빈발항목 구하는 과정

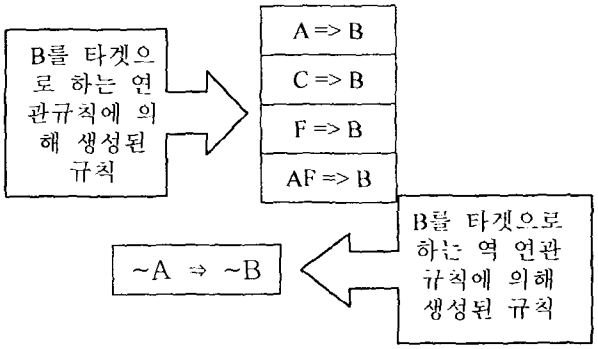
생성된 역 빈발항목집합은 다음과 같다. 최소지지도는 40%이다.

| | |
|---|---|
| 1 역 빈발항목집합 | ~A(6), ~B(4), ~C(4) |
| 2 역 빈발항목집합 | ~A~B(4) |

B를 타겟으로 최소신뢰도를 만족하는 역 연관규칙은 다음과 같이 생성된다. 최소신뢰도는 60%이다.

~A ⇒ ~B

연관규칙으로 생성된 4개의 규칙에 의하여 A, C, F, AF를 구입한 사람들은 B를 구입하는 경향이 많다는 것을 알 수 있다. 그러므로 기존 방식으로 타겟 마케팅에 연관규칙을 적용할 경우에는 A, C, F, AF를 구입한 고객 중 아직 B를 구입하지 않은 고객에게 B를 마케팅 하는 전략을 쓴다. 역 연관규칙 ~A ⇒ ~B도 생성되는데 이 규칙과 '이' 관계에 있는 A ⇒ B가 같이 생성되기 때문에 타겟 마케팅에 또 다른 전략이 첨가된다. 그 전략은 B를 판매하기 위해서는 우선 A를 판매해야 한다는 것이다. 즉 B를 판매하기 위해서 B를 마케팅 하는 전략 외에 B와 종속관계에 있는 다른 항목 A를 마케팅 할 필요가 있다는 정보가 생성된다. 결국 A의 마케팅이 B의 판매를 촉진시키는 역할을 한다. 아래의 그림은 이 설명을 뒷받침한다.




[그림14] 타겟 마케팅에 적용할 규칙들

# 6. 결 론

데이터마이닝의 목표는 알려지지 않은 유용한 정보를 얻어 경영에 활용하는 것이다. 이러한 정보를 얻으려는 목적으로 데이터베이스를 역 데이터베이스로 전환하여 새로운 규칙인 역 연관규칙을 생성한다. 본 논문은 연관규칙에 역 연관규칙을 첨가 시켜 타겟 마케팅에 새로운 정보를 생성시킨다. 연관규칙에서 생성된 규칙과 그 규칙에 나와 있는 항목 간의 역 연관규칙이 생성될 경우 항목집합 간의 종속관계를 알 수 있다. 그 종속관계에 의하여 타겟 항목을 판매하기 위하여 그 항목을 직접 마케팅 하는 전략뿐만 아니라 타겟 항목이 아닌 그 항목과 종속관계에 있는 다른 항목을 마케팅 하는 전략도 가능하게 된다. 다른 항목의 마케팅이 결국은 타겟 항목의 판매에 영향을 주는 경우가 있기 때문이다. 연관규칙과 역 연관규칙을 본 논문이 제시한 대로 타겟 마케팅에 적용하여 해석하면 이러한 경우를 알아낼 수 있으며 결국 좀 더 다양한 마케팅 전략을 세울 수 있다.

# 참고문헌


Agrawal R., R. Srikant: "Fast Algorithms for Mining Association Rules", *Proc. of the 20th Int'l Conference on Very Large Databases*, Santiago, Chile, (1994). Expanded version available as IBM Research Report RJ9839, (1994).

Han Jiawei, Micheline Kamber, *Data Mining: concepts and Techniques*, Morgan Kaufmann publishers, 2000.

Michael J. A. Berry, Gordon Linoff, *Data Mining Technigues*, John Wiley & Sons, NY, 1997

Mohammed J. Zaki, Ching-Jui Hsiao, "CHARM: An Efficient Algorithm for Closed Association Rule Mining,"

RPI Technical Report, (1999), 99-10.

Nicolas Pasquier, Yves Bastide, Rafik Taouil, Lotfi Lakhal: "Discovering Frequent Closed Itemsets for Association Rules." ICDT (1999): 398-416

Pei J. , J. Han, and R. Mao, "CLOSET: An Efficient Algorithm for Mining Frequent Closed Itemsets ", *Proc. 2000 ACM-SIGMOD Int. Workshop on Data Mining and Knowledge Discovery (DMKD'00)}*, Dallas, TX, (2000).

Roberto J. Bayardo Jr.: "Efficiently Mining Long Patterns from Databases." SIGMOD Conference (1998), 85-93