

고객 구매행태의 지속적 변화 파악을 위한 재귀적 변화발견 방법

A Recursive Procedure for Mining Continuous Change of Customer Purchase Behavior

김 재 경 (Jae-Kyeong Kim)	경희대학교 경영대학
채 경 희 (Kyung-Hee Chae)	경희대학교 경영대학
최 주 철 (Ju-Cheol Choi)	경희대학교 경영대학
송 희 석 (Hee-Seok Song)	한남대학교 경영정보학과
조 영 빈 (Yeong-Bin Cho)	건국대학교 사회과학대학 경영학과

요 약

데이터 마이닝의 연관성규칙 분석 기법(Association Rule Mining)은 현실문제에의 많은 활용에도 불구하고 시간의 흐름에 대한 변화 파악 및 분석에서는 한계를 가지고 있다. 본 연구에서는 기존의 두 시점에서의 고객 행위 변화 파악 기법을 재귀적 방법을 통하여 다시점으로 확장하여 분석할 수 있는 방법론을 제시한다. 즉, 본 연구에서는 연관성규칙의 패턴 및 변화의 추세를 장기간에 걸쳐 지속적으로 관찰함으로써, 고객의 일시적인 변화보다는 지속적인 행위 변화를 관찰할 수 있도록 하는 방법론을 구성한다. 방법론을 검증하기 위해 L백화점의 4년간의 구매관련 데이터를 분석하여 그 결과를 제시하고 있다.

키워드 : 고객관계관리, 연관성규칙 분석, 재귀적 변화발견 방법

I. 서 론

1996년 1월부터 전면 개방된 우리나라의 유통 시장은 현재 많은 변화를 겪고 있으며 치열한 경쟁 환경에 직면해 있다. 그로 인해 유통시장에서의 소비자 파워는 더욱 증대되고 있으며 고객에 대한 깊이 있는 이해가 기업 생존의 필수적인 요소로 인식되고 있다. 즉, 기본적으로 우수한 제품을 판매해야 하는 것 이외에도 고객별로 개인화된 서비스를 제공하기 위한 고객관계관리(Customer Relationship Management: CRM)를 위

한 기법들이 기업 생존을 위해 필수적인 요소로 인식되고 있다. 그러나 최근 고객의 요구사항이나 주변 환경들이 급격히 변화하고 있다. 따라서 고객을 이해함으로써 개인화된 서비스를 제공하는 것이나, 고객의 성향을 예측하고 파악하는 것이 매우 어려워지고 있다. 이에 본 연구에서는 시간에 따른 고객 구매행태의 변화를 발견하기 위한 방법론을 제시하는 것을 그 목적으로 한다.

시간에 따른 고객 구매행태의 변화를 발견하기 위한 방법은 Song *et al.*(2001)의 연구에서 찾아볼 수 있다. 이 연구에서는 서로 다른 두 시점

의 데이터 집합으로부터 연관성 규칙을 생성하여 두 시점의 연관성 규칙을 비교함으로써 고객 행위 변화를 발견하고 있다. 그러나 이러한 분석은 두 시점간 비교일 뿐 장기적인 패턴의 변화나 특징은 발견하지 못한다는 단점이 있다. 따라서 본 논문에서는 다시점(Multi-Point)에서 고객 데이터를 비교 관찰하여 지속적인 시간 흐름에 따른 고객 구매 행태의 변화를 발견하고자 한다.

본 논문에서는 고객의 구매행태의 변화를 파악하기 위하여, 먼저 고객의 인구통계학적 특성과 구매 제품에 관련된 자료를 수집하여 연관성 규칙을 생성한 다음 처음 두 시점의 연관성 규칙을 비교하여 지속적으로 관찰할 대상이 되는 변화를 일차적으로 선별한다. 이후, 본 논문을 통해 제안하고자 하는 재귀적 변화발견 방법(Recursive Change Mining)을 통해 지속적 변화만을 필터링 한다. 과거 단편적인 시점에서만 관리되던 연관성 규칙의 변화들을 중장기적인 관점에서 지속적으로 살펴보고 관리하게 되면, 고객 구매행태에 대한 새로운 정보나 지식을 추출할 수 있을 것이다. 즉, 고객 구매행태의 변화를 과거부터 지속적으로 비교하여 살펴봄으로써 규칙적인 변화나 일정한 패턴을 파악할 수 있게 되면 구매행태에 대한 예측이 가능해져서 상품판매, 고객이탈 방지, 고객 활성화 전략 등의 캠페인 수립에 활용될 수 있을 것이다. 뿐만 아니라 본 연구에서 발견하고자 하는 다시점 부상패턴과 다시점 미세측변화에는 일시적으로 관찰되는 단편적인 변화는 제외되기 때문에 장기적 전략 수립에 유용한 정보를 제공할 수 있을 것으로 판단된다.

본 논문의 구성을 살펴보면 다음과 같다. 먼저 II절에서 고객의 구매행태를 파악하기 위해 활용된 연관성규칙 분석 방법에 대해 설명하고 고객의 구매나 행위 패턴의 변화, 고객 환경의 변화 등의 분석과 관련된 기존 연구들에 대해 살펴보고자 한다. III절에서는 본 논문에서 제안하고자 하는 방법론과 절차에 대해 설명하고, IV

절에서는 논문을 통해 제안한 방법론을 실제 기업 데이터에 적용시켜 봄으로써 방법론을 검증하고 결과에 대한 시사점을 도출해 보고자 한다. 마지막으로 V절에서는 본 연구의 내용을 요약 정리하고 한계점 및 향후 연구 방향을 제시하고자 한다.

II. 문헌 연구

2.1 연관성규칙 분석

연관성규칙 분석은 데이터마이닝 기법 중 하나로써 복잡한 대용량의 데이터를 이해하기 쉬운 규칙의 형태로 표현해 주며, $A \Rightarrow B$ 와 같이 조건부 A 와 결과부 B 의 연관성규칙으로 표현된다. 연관성규칙 분석은 분석의 방법이 간단하고 많은 양의 데이터에 대해서도 쉽게 적용할 수 있으므로 데이터마이닝에 유용하게 이용되고 있다. 대용량의 데이터를 분석하다 보면 수많은 연관성규칙이 생성되는데, 그 중에는 의미가 서로 상충되는 경우도 나타날 수 있다. 따라서 다수의 연관성규칙 중에서 유의한 것을 찾아내고 상충하는 규칙 중에서 어느 것을 선택할 것인지 결정해야 하는데, 이 때 주로 사용되는 기준이 지지도(Support)와 신뢰도(Confidence)이다. 연관성 규칙에서 지지도는 전체거래 건수에 대한 조건부와 결론부를 동시에 만족하는 거래건수로 계산되며 규칙의 유용성을 평가하는 기준으로 자주 활용된다. 신뢰도는 조건부를 만족하는 거래에서 결론부가 동시에 만족되는 조건부 확률로써 계산되며 규칙의 신뢰성을 평가하는데 사용될 수 있다. 본 연구에서는 다양한 연관성규칙 분석의 응용 방법 중 하나인 프로파일 연관성규칙 분석 방법을 활용하였다(Agrawal et al., 1998). 일반적인 연관성규칙 분석의 경우에는 조건부에 포함된 항목이 결과에 포함될 수도 있고, 결과의 항목이 조건부에 포함될 수도 있다. 그러나 프로파일 연관성규칙 분석에서는 조건부에 포함되는

항목과 결과에 포함되는 항목을 다르게 구성한다. 예를 들어, 조건부에 나이, 수입, 교육수준, 사회적 위치 등의 고객 프로파일 관련 항목을 포함하며 결과에는 구매 제품, 매출액, 시간 등과 같은 고객 행위 데이터 만을 포함하도록 하여 연관성규칙을 생성한다. 이 경우, 생성된 연관성 규칙은 “성별 = 남성 \Rightarrow 매출=high”와 같이 타겟 시장의 파악과 마케팅 실행에 용이한 의미 있는 규칙으로 나타나게 되고, 고객 관리자는 생성된 규칙을 통해 남성을 대상으로 타겟 광고를 늘리는 식의 캠페인을 수립할 수 있다. 본 연구에서 제시하고 있는 방법론은 실제 관리자의 의사결정을 돕는 자동화된 시스템을 지향하고 있으므로 프로파일 연관성규칙 분석방법을 활용하였다.

2.2 변화를 발견하기 위한 데이터마이닝 연구

본 절에서는 대용량의 데이터로부터 변화를 발견하거나 변화를 따라잡기 위한 기존 연구들을 소개한다. 이러한 연구들은 대부분 연관성규칙 분석을 기반으로 이루어지고 있는데 이는 연관성규칙 분석 기법이 다른 분석 기법에 비해 지식 발견의 누락을 줄일 수 있다는 장점 때문이다(Collectively Exhaustive). 즉, 연관성규칙 분석 기법은 데이터로부터 부분적인 지식을 발견하기보다는 가능한 모든 지식을 발견하는데 근접한 방법이기 때문이다. 변화를 따라잡기 위한 기존 연구로는 연관성규칙의 유지 보수와 관련된 연구를 들 수 있다(Cheung *et al.*, 1996; Feldman *et al.*, 1997; Thomas *et al.*, 1997). 이 연구들은 변화하는 데이터로부터 이미 발견된 연관성규칙을 어떻게 유지 보수함으로써 변화를 따라잡을 것인가에 초점을 맞추고 있다. 즉, 새로운 거래 데이터가 추가 되었을 때 효과적으로 규칙을 유지 보수하기 위한 방법을 제시하고 있다. 한편, 두 개의 서로 다른 데이터 집합으로부터 변화된 내용만을 발견하기 위한 변화탐지(Change Mining)

연구가 시도되어 왔다. Agrawal and Psaila(1995), Dong and Li(1999)의 연구에서는 지지도에 있어서 차이가 임계치 이상인 연관성규칙을 부상패턴(EP: Emerging Pattern)으로 정의하여 기준데이터 집합과 비교데이터 집합의 두 집합으로부터 EP를 발견하는 방법을 제시하고 있다. 이렇게 발견된 EP는 두 개의 데이터 집합 간에 차이를 설명하는데 필요한 지식이 된다. Liu *et al.*(2000)은 기존연구를 토대로 서로 다른 시점에 수집된 두 개의 데이터 집합으로부터 EP를 발견함으로써 두 시점에 있어서 변화된 내용이 무엇인지를 규명하고자 하였다. EP외에 중요한 변화로 인식될 수 있는 유형으로는 미예측변화(Unexpected Change)가 있다. Liu and Hsu(1996)는 중요하지 않은 연관성규칙을 필터링하는 기법을 제안하는 과정에서 기존의 전문가의 신념(belief)을 규칙으로 표현하고 이러한 신념에 위배되는 연관성규칙을 Unexpectedness로 정의하고 이러한 연관성 규칙을 발견하고자 하였다. Silberschatz and Tuzhilin(1996), Liu *et al.*(1997), Padmanabhan and Tuzhilin(1999), Suzuki(1997)의 연구는 Unexpectedness에 대한 이론적 배경과 프레임워크를 제공하고 있다. Unexpectedness에 대한 개념을 서로 다른 두 시점에 수집된 데이터 집합에 적용함으로써 미예측변화(Unexpected Change)를 찾고자 하는 연구는 Song *et al.*(2001)로부터 비롯되었다. Song *et al.*(2001)은 서로 다른 두 시점의 데이터 집합으로부터 연관성 규칙을 생성하여 두 시점의 연관성 규칙을 비교함으로써 부상패턴(Emerging Pattern), 미예측변화(Unexpected Change), 추가/삭제규칙(Added/Perished Rule) 등을 발견하는 고객 행위 변화 방법론을 제시하고 있다. 이 연구에서는 미예측변화(Unexpected Change)를 시간이 지남에 따라 규칙의 조건부는 유사하나 규칙의 결과부는 달라지는 형태의 변화로 정의하고 있다. 예를 들면, “성별=여성 \Rightarrow 매출= high”인 연관성 규칙이 “성별=여성 \Rightarrow 매출=low”와 같은 연관성 규칙으로 변화되는 것을 말한다. 그러나 이러한

분석은 두 시점간 비교일 뿐 장기적인 패턴의 변화나 특징은 발견하지 못한다는 단점이 있다. 따라서 본 논문에서는 다시점(Multi-Point)에서 고객 데이터를 비교 관찰하여 지속적인 시간 흐름에 따른 고객 구매 행태의 변화를 발견하고자 한다. 이외에도 Das *et al.*(1997, 1998)에서는 시계열과 관련된 규칙을 발견하는 것에 관심을 기울였으며, Han *et al.*(1999)은 규칙의 주기성을 발견함으로써 데이터마이닝의 효과성을 높이기 위한 몇 가지 알고리즘을 제시하였다.

III. 연구방법

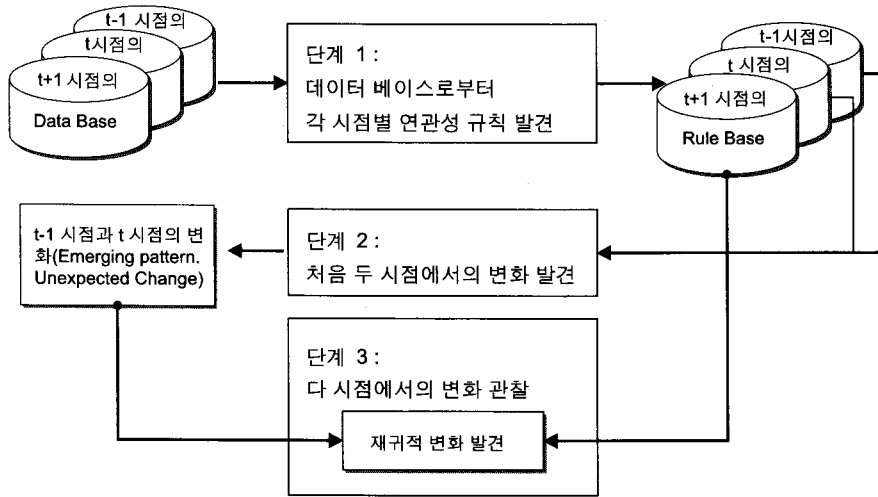
3.1 방법론 개요

본 연구에서는 기존의 단편적인 시점에서 관찰되던 연관성규칙 분석 방법을 확장하여, 시간의 흐름에 따라 지속적인 연관성규칙 기반의 변화를 발견하는 방법을 제시하고자 한다. 이를 위해 지속적인 변화의 유형으로 **다시점 부상패턴**과 **다시점 미예측변화**를 정의하고 이들을 발견하는 것을 목표로 한다. 다시점 부상패턴(다시점 미예측변화)은 부상패턴(미예측변화)이 다시점에 걸쳐 변화의 지속성여부가 확인된 것이다. 다시점 부상패턴은 시간에 따라 지지도가 지속적으로 임계치 이상 증가 또는 감소하거나 임계치 이상의 등락을 반복하는 변화를 보이는 부상패턴이다. 예를 들면 “기혼 여성이면 A브랜드 구매금액이 높다”라는 연관성규칙의 지지도가 2001년부터 2004년까지 꾸준히 5%씩 증가 되었다면, 변화량에 대한 임계치를 2%로 가정할 때 이 연관성 규칙은 다시점 부상패턴으로 분류될 수 있다. 다시점 미예측변화는 규칙의 내용이 변화되어 지속적으로 변화가 고착되는 형태이다. 예를 들면, “기혼 여성이면 A브랜드 구매금액이 높다”라는 연관성규칙이 2002년에는 “기혼 여성이면 A브랜드 구매금액이 낮다.”라는 형태로 변화 되어 이러한 변화가 2006년까지 지속적으로

고착화되는 경우이다.

본 연구에서 다시점 부상패턴과 다시점 미예측변화를 발견하는 것을 최종 목표로 하는 이유는 기업측면에서 미래 고객에 대한 행동 예측이 가능하다는 점 때문이다. 즉, 다시점 부상패턴이나 다시점 미예측변화는 일정 시점 후의 고객 행위 예측을 가능케 하는 신뢰할 수 있는 지식이기 때문이다. 예를 들어 유통업체에서 “기혼 여성이면 A브랜드 구매금액이 높다”라는 연관성규칙의 지지도가 3년 연속 지속적으로 상승하는 변화를 인지했다면 이는 다른 연관성규칙에 비해 매우 신뢰할 수 있는 변화이므로 높은 투자 비용에도 불구하고 A브랜드 특설코너를 설치하는 의사결정을 내릴 수 있다.

본 연구에서 제안하고자 하는 방법론은 크게 세 단계로 나누어 볼 수 있으며 전체적인 프로세스를 살펴보면 <그림 1>과 같다. 첫 번째 단계에서는 연관성규칙 분석(Agrawal *et al.*, 1993; Agrawal and Srikant, 1994)을 이용하여 데이터 베이스로부터 각각의 시점에 대한 연관성규칙을 생성한 후 규칙 베이스(Rule base)에 저장한다. 두 번째 단계에서는 처음 두 시점을 선택하여 두 시점간의 변화의 정도가 정해진 임계치 이상인 연관성 규칙들을 부상패턴(Emerging Pattern)과 미예측변화(Unexpected Change)로 구분하여 규칙베이스에 저장한다. 두 시점 간 변화를 발견하는 방법은 Song *et al.*(2001)의 방법을 적용한다. 마지막 단계에서는 다시점(multi-point)에서 연관성규칙의 변화를 지속적으로 관찰한다. 이를 위해 본 연구에서는 재귀적 변화발견 방법(Recursive Change Mining)을 제안한다. 재귀적 변화발견 방법은 처음 두 시점($t-1$, t)에서 일정 임계치 이상 변화된 연관성 규칙을 발견하고 이 규칙과 차기 시점($t+1$)의 연관성규칙을 직접 비교하여 다시 변화를 발견함으로써 지속적이지 못한 연관성규칙을 필터링하게 되고, 남아있는 연관성 규칙을 다시 차차기($t+2$)의 연관성규칙과 비교하여 지속적이지 못한 변화를 반복적으로



〈그림 1〉 지속적인 변화 탐지 방법론의 구성

필터링하는 재귀적 방식을 채택하고 있다. 이러한 재귀적 변화발견 방법은 본 연구에서 목표로 하고 있는 다시점 부상패턴과 다시점 미예측변화를 모두 발견하는 것을 보장한다.

3.2 두 시점에서의 변화 발견

본 방법론은 먼저 연관성규칙 분석(Agrawal *et al.*, 1993; Agrawal and Srikant, 1994)을 이용하여 데이터 베이스로부터 각각의 시점에 대한 연관성규칙을 생성한다. 본 연구는 먼저 두 시점에서의 규칙간 변화를 측정한 후, 다시점으로 확장하여 다시점 부상패턴을 관찰할 수 있도록 확장된 연관성규칙 분석 방법을 제시하고 있다. 문헌연구를 통해 두 시점에서 관찰할 수 있는 변화를 살펴보면 크게 부상패턴(Emerging Pattern)과 미예측변화(Unexpected Change)로 나누어 볼 수 있다(Song *et al.*, 2001).

부상패턴(Emerging Pattern)은 기준 시점에서 발견된 특정 연관성 규칙의 지지도가 비교 시점에서 임계치 이상 현저히 증가하거나 감소한 형태의 변화로써 다음과 같이 정의된다. 만일 다음의 두 가지 조건이 충족된다면, $t+k$ 시점에서의

j 번째 규칙인 r_j^{t+k} 는 t 시점에서의 i 번째 규칙인 r_i^t 의 관점에서 볼 때, 부상패턴(Emerging Pattern)으로 정의된다.

- (1) r_i^t, r_j^{t+k} 두 규칙의 조건부와 결과부가 같은 경우
- (2) 두 규칙들 간의 지지도 차이가 최소 지지도 변화량에 대한 임계치보다 큰 경우

예를 들어, t 시점에서 “주소 = 강북, 나이 = 50대 \Rightarrow 브랜드명 = 테스트니(지지도 = 0.1)”이라는 규칙 r_i^t 가 발견되었고, $t+k$ 시점에서 “주소 = 강북, 나이 = 50대 \Rightarrow 브랜드명 = 테스트니(지지도 = 0.13)”이라는 규칙 r_j^{t+k} 가 발견되었다고 가정하자. 만일 사용자가 최소 지지도 변화 비율을 20%로 설정하였다면, 이 예의 경우, r_i^t 에 비해 r_j^{t+k} 의 지지도가 30% 가량 증가하였으므로, r_j^{t+k} 는 r_i^t 에 대한 부상패턴으로 정의할 수 있다.

한편, 두 시점의 규칙이 조건부는 유사하고 결론부가 상이하면 미예측변화(Unexpected Change)로 정의한다. 미예측변화(Unexpected Change)를 발견하기 위해서는 Liu and Hsu(1996)의 특이도(difference) 지표가 이용된다. 특이도(difference)

를 계산하는 방법은 식 (3-1)과 같은데 만일 특이도의 수치가 0이 나왔다면, 이는 두 규칙의 조건부와 결과가 서로 다른(관련성이 없는) 규칙임을 의미한다. 그리고 1이 나왔다면, 이는 조건부는 동일하고 결과가 서로 다른 규칙임을 의미하며, -1이 나왔다면, 이는 결과는 서로 같고 조건부가 다른 규칙임을 의미한다. 따라서 $\delta_{ij} > 0$ 인 경우를 미예측변화(Unexpected Change)로 정의할 수 있다. 두 규칙 간의 특이도 δ_{ij} 는 두 규칙의 조건부 간의 특이도 ℓ_{ij} 와 조건부에 포함된 각 변수간의 유사도 x_{ijk} 의 합을 곱한 것에서, 다시 두 규칙간 결과부의 유사도 y_{ij} 를 차감한 것으로 나타낼 수 있다.

$$\delta_{ij} = \begin{cases} \frac{\ell_{ij} \times \sum_{k \in A_{ij}} x_{ijk}}{|A_{ij}|} - y_{ij}, & \text{if } |A_{ij}| \neq 0, c_{ij} = 1 \\ -y_{ij}, & \text{if } |A_{ij}| = 0, c_{ij} = 1 \end{cases} \quad (3-1)$$

특이도를 구하기 위해서는 먼저, ℓ_{ij} 와 x_{ijk} , y_{ij} 를 계산해야 하는데, ℓ_{ij} 를 계산하는 방법은 식 (3-2)와 같다. 다음의 식에서 $|A_{ij}|$ 는 두 규칙에 공통적으로 포함되어 있는 변수의 개수를 의미한다. 그리고 $|X_i^t|$ 는 t 시점 연관성규칙의 조건부에 포함된 변수의 개수를 의미하고, $|X_j^{t+k}|$ 는 $t+k$ 시점 연관성규칙의 조건부에 포함된 변수의 개수를 의미한다.

$$\ell_{ij} = \frac{|A_{ij}|}{\max(|X_i^t|, |X_j^{t+k}|)} \quad (3-2)$$

다음으로 x_{ijk} 는 두 연관성규칙 조건부의 공통 변수가 가지는 값에 대한 유사성을 나타내며, 변수의 값이 같으면 1, 다르면 0으로 계산된다(식 (3-3)).

$$x_{ijk} = \begin{cases} 1, & \text{변수의 값이 같은 경우} \\ 0, & \text{변수의 값이 다른 경우} \end{cases} \quad (3-3)$$

마지막으로 y_{ij} 는 결과에 포함된 변수가 가지

는 값의 유사성을 나타내며, 변수의 값이 같으면 1, 다르면 0으로 계산된다(식 (3-4)).

$$y_{ij} = \begin{cases} 1, & \text{변수의 값이 같은 경우} \\ 0, & \text{변수의 값이 다른 경우} \end{cases} \quad (3-4)$$

미예측변화(Unexpected Change)에 대한 예를 들면, t 시점에서 “주소 = 강북, 나이 = 50대 \Rightarrow 브랜드명 = 테스트니”라는 규칙 r_i^t 가 발견되었고 $t+k$ 시점에서 “주소 = 강북, 결혼여부 = 미혼 \Rightarrow 브랜드명 = 센존”라는 규칙 r_j^{t+k} 가 발견되었다고 가정하면, 특이도는 0.5가 된다. 만일 사용자가 임계치를 0.4로 정해 놓았다면 특이도가 임계치 이상이므로 r_j^{t+k} 는 r_i^t 의 관점에서 미예측변화(Unexpected Change)라 정의할 수 있다. 그러나 예를 통해 제시되는 두 규칙은 구조가 다른 규칙들이므로 부상패턴(Emerging Pattern)과 같이 지지도의 변화만으로 변화의 정도를 측정하는 것은 불가능하다. 따라서 Padmanabhan and Tuzhilin(1999)의 결합 규칙의 개념을 활용하여 미예측변화(Unexpected Change)의 정도를 판단한다. 예를 들면, “펭귄은 날지 못한다”라는 규칙이 기존 관념인 “새는 난다”라는 규칙에 대해 유의한 미예측변화(Unexpected Change)인지 여부를 파악하기 위해서는 결합규칙인 “새이면서 펭귄은 날지 못한다”의 지지도를 조사하여 지지도가 높다면 “새는 난다”라는 규칙에 대한 예외가 많이 발생한 것이므로 유의한 변화로 판단할 수 있다. r_j^{t+k} 가 미예측변화(Unexpected Change)로 정의되기 위해서는 t 시점에서의 연관성규칙 r_i^t 와 $t+k$ 시점에서의 연관성규칙 r_j^{t+k} 를 결합한 규칙 $r_{i \cap j}$ 의 지지도가 충분히 커야 한다. 따라서 결합 규칙의 지지도를 기반으로 미예측변화(Unexpected Change)의 정도를 평가하게 된다.

부상패턴(Emerging Pattern)과 미예측변화(Unexpected Change)에 대한 변화의 정도를 평가 방법은 두 변화의 특성이 서로 다르기 때문에, 다른 관점에서의 접근이 필요하다. 식 (3-5)는 부상

패턴의 변화의 정도를 평가하기 위한 척도로써 $\alpha_i^{t,t+k}$ 를 미예측변화의 정도를 평가하기 위한 척도로써 $\beta_{ij}^{t,t+k}$ 를 각각 제시하고 있다.

$$\alpha_i^{t,t+k} = \frac{|Sup^{t+k}(r_i) - Sup^t(r_i)|}{Sup^t(r_i)}$$

$$\beta_{ij}^{t,t+k} = \frac{Sup^{t+k}(r_{i \cap j})}{Sup^{t+k}(r_j)} \quad (3-5)$$

위의 식에서 $Sup^{t+k}(r_i)$ 는 $t+k$ 시점에서의 연관성규칙 r_i 에 대한 지지도를 의미한다. 그리고 $Sup^t(r_i)$ 는 t 시점에서의 연관성규칙 r_i 의 지지도를 의미한다. 이 두 가지 지지도를 이용하여 부상패턴(Emerging Pattern)의 변화의 정도에 대한 순위가 결정된다. 다음으로 미예측변화(Unexpected Change)의 정도에 대한 우선 순위는 $t+k$ 시점에서의 결합 규칙의 지지도 $Sup^{t+k}(r_{i \cap j})$ 와 $t+k$ 시점에서의 연관성규칙 r_j 의 지지도, $Sup^{t+k}(r_j)$ 에 의해 결정된다.

3.3 다시점(multi-point)에서의 변화 관찰

기존의 단편적인 시점에서 관찰되던 연관성 규칙 분석 방법을 확장하여 시간의 흐름에 따라 지속적인 변화를 발견하는 방법을 제시하고자 한다. 이를 위해 다시점 부상패턴과 다시점 미예측변화에 대한 용어 정의와 충분조건을 정리하면 다음과 같다.

[정의 1] 시점 1로부터 시점 p 까지의 다시점 부상패턴

모든 $t, 1 \leq t \leq p-1$ 에 대해 $\alpha_i^{t,t+1} \geq T_1$ (T_1 는 최소 변화량에 대한 임계치)를 만족하는 부상패턴

[정의 1]에 의하면 다시점 부상패턴은 시점 1과 시점 2의 비교를 통해 발견된 부상패턴이 다시점에 걸쳐 변화의 지속성여부가 확인된 변화

로써, 시점 1에서 시점 p 까지의 임의의 인접한 두 시점에서 발견된 부상패턴의 변화의 정도 ($\alpha_i^{t,t+1}$)가 모두 일정 임계치 이상인 부상패턴으로 정의된다.

[정의 2] 시점 1로부터 시점 p 까지의 다시점 미예측변화

모든 $t, 1 \leq t \leq p-1$ 에 대해 $\beta_{ij}^{t,t+1} \geq T_2$ (T_2 는 최소 변화량에 대한 임계치)를 만족하는 미예측변화

[정의 2]에 의하면 다시점 미예측변화는 시점 1과 시점 2의 비교를 통해 발견된 미예측변화가 다시점에 걸쳐 변화의 지속성 여부가 확인된 형태로써, 시점 1에서 시점 p 까지의 임의의 인접한 두 시점에서 발견된 미예측변화의 정도 ($\beta_{ij}^{t,t+1}$)가 모두 일정 임계치 이상인 미예측변화로 정의된다. 다시점 미예측변화는 미예측변화가 일시적으로만 발견되는 것이 아니라 p 시점까지 지속적으로 고착화 되는 변화만을 의미한다. 이를 발견하는 방법은 미예측변화의 결합 규칙의 지지도가 p 시점까지 일정 임계치 이상 유지되는지를 검사하는 것이다.

본 연구에서 탐지 대상 목표로 삼고 있는 지속적인 변화의 발견은 재귀적 변화발견 방법을 통해 이루어진다. 재귀적 변화발견 방법은 초기 t 시점과 $t+1$ 시점에서 발생된 변화를 기반으로 향후 $t+k$ 시점까지도 이러한 변화가 지속되는지 여부를 관찰하는 것이며 각 단계별 프로세스는 <그림 2>와 같다. 즉, t 시점과 $t+1$ 시점에서 각각 연관성규칙을 생성하여 규칙베이스에 저장한다. 다음으로 저장된 두 시점 간의 규칙을 비교, 분석하여 부상패턴(Emerging Pattern)과 미예측변화(Unexpected Change)를 도출한 후 규칙베이스에 다시 저장한다. $t+2$ 시점의 연관성규칙을 생성하여 규칙베이스에 저장된 두 가지 종류의 변화와 각각 비교한 후 새롭게 부상패턴(Emerging Pattern)과 미예측변화(Unexpected

Change)을 갱신한다. 이와 같은 과정을 $t+k$ 시점까지 반복한다.

각 단계별로 예를 들어 설명하면 다음과 같다. 먼저 1단계에서 시작 시점이 결정되면, 그 시점을 1로 설정한다. 그리고 그 시점의 데이터를 이용하여, 연관성규칙을 생성하면, “주소 = 강북, 나이=50대 \Rightarrow 브랜드명=테스토니(지지도: 0.1)”와 같은 수 많은 규칙이 생성되므로 사용자는 최소 지지도를 임계치로 설정하여 최소 임계치를 만족하는 연관성규칙만을 선별하여 R^t 에 저장한다. 2단계에서는 두 번째 시점의 데이터집합에 대해 1단계와 동일하게 연관성규칙을 생성하고 R^{t+1} 에 저장한다. 이 때 생성된 연관성규칙 중에 “주소=강북, 나이=50대 \Rightarrow 브랜드명=테스토니(지지도: 0.13)”와 “주소=강북, 결혼여부=미혼 \Rightarrow 브랜드명=센존(지지도: 0.13)”, “주소=강북, 나이=50대, 결혼여부=미혼 \Rightarrow 브랜드명=센존(지지도: 0.1)”가 있다고 가정한다. 3단계에서는 1단계에서 생성된 R^t 와 R^{t+1} 을 비교하여 부상패턴(Emerging Pattern)과 미예측변화(Unexpected Change)을 찾아낸다. 위의 예를 이용하여 설명하면, 1단계에서 생성된 “주소=강북, 나이

=50대 \Rightarrow 브랜드명=테스토니(지지도: 0.1)”의 기준에서 볼 때, 2단계에서 생성된 규칙 중 “주소=강북, 나이=50대 \Rightarrow 브랜드명=테스토니(지지도: 0.13)”는 최소 임계치가 0.2일 때, 부상패턴(Emerging Pattern)으로 판단될 수 있으며, 또한 같은 기준에서 “주소=강북, 결혼여부=미혼 \Rightarrow 브랜드명=센존(지지도: 0.13)”는 “주소=강북, 나이=50대, 결혼여부=미혼 \Rightarrow 브랜드명=센존(지지도: 0.1)”라는 충분히 큰 지지도를 가지고 있는 결합규칙이 존재하므로 시점 1의 연관성 규칙에 대한 미예측변화(Unexpected Change)라 할 수 있다. 이후 5단계에서는 발견된 두 가지의 변화만을 이용하여 R^{t+1} 을 구성하고 다시 $t+2$ 기에 생성된 연관성 규칙의 집합인 R^{t+2} 와 비교함으로써 재귀적으로 변화의 지속성 여부를 판단하게 된다.

재귀적 변화발견 방법에서는 시점 t 부터 시점 $t+k$ 까지의 분석 대상 기간 중 특정 시점 $n(0 < n < k)$ 에서 지지도의 변화정도가 임계치 이상 되지 않으면 탈락되기 때문에 앞에서 정의된 다시점 부상패턴과 다시점 미예측변화의 탐지를 보장한다.

1단계 : Set $t=1$
 데이터집합 D^t 로부터 연관성규칙집합 R^t 를 생성

2단계 : 데이터집합 D^{t+1} 로부터 연관성규칙집합 R^{t+1} 생성

3단계 : t 시점의 규칙집합 R^t 와 $t+1$ 시점의 연관성규칙집합 R^{t+1} 를 이용하여 연관성규칙의 변화 분석 및 지속적 변화 선정

- 반복 1: 부상패턴(Emerging Pattern) 및 미예측변화(Unexpected Change) 발견
- 반복 2부터:
 - 3-1) 반복 1에서의 부상패턴에 대해, $\alpha_i^{t+1} \geq T_1$ 인 부상패턴만을 선정
 - 3-2) 반복1에서의 미예측변화에 대해 $\beta_{ij}^{t+1} \geq T_2$ 인 미예측변화만을 선정

4단계 : 만일 t 시점과 $t+1$ 시점에서 선정된 변화가 없다면, 종료.
 선정된 변화가 존재하면, 다음 단계로 이동

5단계 : $R^{t+1} \leftarrow \emptyset$
 t 시점과 $t+1$ 시점에서 선정된 변화(부상패턴 및 미예측변화)들을 R^{t+1} 에 저장
 Set $t=t+1$
 2단계부터 다시 반복

<그림 2> 재귀적 변화발견 방법의 프로세스

IV. 사례연구

4.1 분석 방법

본 연구에서 제시하는 방법론의 유용성을 검증하기 위해 유통업체인 L백화점의 실제 데이터를 수집하여 본 방법론을 적용하였다. 백화점의 온라인, 오프라인을 통해 보유하고 있는 다양한 고객 정보와 제품관련 정보, 구매정보를 이용함으로써 유용한 전략적 정보를 도출해 내는 것이 본 사례 연구의 최종 목표이다. 본 연구의 분석에 사용된 프로그래밍 도구는 Visual Basic 6.0이며, 연관성규칙 분석에는 클레멘타인 5.1을 활용하였다.

4.2 도메인

2001년 이후 고급 명품에 대한 수요가 지속적으로 증가하여 명품관을 별도로 마련한 유통점이 증가하고 있다. 이에 L백화점도 별도의 명품관을 운영해 왔으며 고가 명품 구매의 특성 상 고객에 대한 깊이 있는 이해와 관계관리가 매우 중요함을 인식하게 되어 전산실과 마케팅부서 주도로 고객데이터에 대한 광범위한 분석을 수행하게 되었다. L백화점의 경우, 2001년 이후 명품관을 운영하면서 축적된 데이터가 있었으며 이러한 시계열적 데이터를 통해 고객에 대한 이해를 높이고 니치 마케팅 전략을 수립하는 방법을 고민하였고, 이를 위한 방법론으로 본 연구에서 제시한 재귀적 변화 발견 방법론이 적합한 것으로 판명되었다. L백화점의 명품구매 행위 데이터를 이용한 데이터 마이닝의 목표는 다음 두 가지로 정리 될 수 있다. 첫째, 명품 구매 고객들의 선호(행위)가 어떻게 바뀌고 있는지의 경향을 미리 파악하여 명품관 구성 등의 장기적인 명품관 운영전략 수립에 필요한 정보를 얻는 것이다. 둘째, 고객 행위 변화를 토대로 특정 세분 고객집단을 선정하고 집단별로 차별화된 마케팅

캠페인을 실시하는 것이다. 이를 위해 L 백화점의 명품 데이터베이스에 저장되어 있는 1,000명의 고객에 대한 31,990건의 거래 데이터와 고객 데이터를 4개 연도별로 수집하였으며 데이터는 전처리 과정을 통해 고객유형, 성별, 연령, 띠, 결혼여부, 고객등급, 가입기간, 타사카드 고객여부, 문화센터 회원여부, 장기거래 고객여부, MVG여부, 주소, 판매수량, 판매금액, 브랜드 명 등을 추출하였다.

4.3 연관성규칙 분석 및 결과

분석 단계의 첫 번째로써 전처리 된 데이터를 이용하여 각 연도별 고객과 구매 행태 간의 연관성규칙을 도출한다. 도출 된 연관성규칙은 각 규칙 베이스에 저장됨으로써, 다음 단계의 부상 패턴(Emerging Pattern)이나 미예측변화(Unexpected Change)를 도출하기 위한 기반이 된다. 본 연구에서는 연관성규칙의 변화를 측정하는 것이 주된 목적이므로 초기 규칙의 발견에 있어 최소 지지도는 1%로 매우 낮게 설정하였다. 이는 기준시점에 비록 매우 낮은 지지도를 보이고 있는 규칙이라 하더라도 비교시점에 지지도에 있어서 큰 변화가 있다면 매우 의미 있는 규칙이 될 수 있기 때문이다. 연관성규칙 분석은 브랜드명, 판매금액, 판매량 등 각기 다른 결과 변수를 기준으로 실시되었는데 분석에 포함되는 결과나 조건부에 포함될 변수는 고객 관리자의 의사결정 목표에 따라 선택되고 변형될 수 있다. 다음의 <표 1>은 각 연도별로 결과 변수를 각각 다르게 하였을 때 생성된 연관성규칙의 개수를 나타내고 있다.

<표 1>을 보면 각 연도별 브랜드를 기준으로 생성된 연관성규칙은 상대적으로 소수이며, 판매금액, 판매량 순으로 연관성규칙의 수가 증가됨을 알 수 있다. 이는 브랜드의 경우 특정 계층이나 특정 고객군에 대해 공통적으로 구매되는 브랜드가 적으며, 반면에 판매금액이나 판매량

의 경우에는 특정 양을 구매하는 다양한 계층과 고객군이 존재함을 의미한다. 특히 판매량의 경우에는 대부분의 고객들이 1년에 1회 가량 명품을 구매하기 때문에 다양한 고객층에 대한 표현이 연관성규칙들을 통해 나타난 결과라고 판단된다. 그러나 브랜드의 경우, 구매하는 종류가 다양하고 같은 브랜드에 대한 공통적인 고객층에 대해 표현할 수 있는 방법이 매우 제한되어 있음을 알 수 있다.

<표 1> 각 연도별 및 결과 변수별 연관성규칙의 개수

	2001	2002	2003	2004
브랜드명	2,989	2,648	3,004	2,876
판매금액	5,397	5,028	6,047	5,825
판매량	13,907	12,825	14,452	14,786

4.4 변화의 발견

다시점 변화를 관찰하기 위해서는 관찰이 시작되는 두 시점간 변화의 분석이 우선적으로 수행되어야 한다. 본 사례연구에서는 전진적 재귀 방법을 통해 지속적인 변화를 발견하고 있다. 즉, 고객 관리자가 과거 4년 전부터 지금까지의 4년간 변화를 지속적으로 관찰하고자 한다면 4년 전의 연관성규칙과 3년 전의 연관성규칙을 비교하여 두 시점간 변화를 발견하고 이를 통해 재귀적으로 변화의 지속성 여부를 판단하게 된다.

<표 2> 처음 두 시점에서의 변화 발견 결과

	부상패턴 (Emerging Pattern)	미예측변화 (Unexpected Change)
브랜드명	153(10%)	716(70%)
판매금액	150(40%)	134(50%)
판매량	172(100%)	-

우선 처음 두 시점에서의 발견된 변화는 <표

2>와 같다. 그러나 이러한 변화가 일시적인 변화인지 혹은 지속적 변화인지는 추가적인 분석을 통해 파악되어야 한다. 임계치를 살펴보면, 부상패턴(Emerging Pattern)의 경우 브랜드 임계치는 매우 낮게 설정되고 판매량 임계치는 매우 높게 설정되었음을 알 수 있다. 이는 브랜드의 경우 공통된 브랜드를 구매하는 특정 계층이 드문 반면, 판매량은 주로 한 두 개를 구매한 고객들이므로 특정 계층에 대한 다양한 연관성규칙이 생성됨에 따라 나타난 현상이다. 반면에 미예측변화(Unexpected Change)의 경우 부상패턴(Emerging Pattern)과는 다르게 초기 브랜드 임계치가 판매금액의 임계치 보다 높게 설정되었는데, 이는 브랜드가 유행이나 흐름에 따라 한 쪽으로 편중되는 현상이 반영된 결과라고 판단된다. 미예측변화(Unexpected Change)의 경우 판매량은 분석에서 제외되었는데, 고객들이 한번에 하나의 브랜드에 대해 한 개의 제품만을 구매하는 경우가 대부분으로써, 결과값의 변형은 기대하기 어렵기 때문이다. 그러나, 변화량이 큰 고객의 경우 지속적으로 구매하는 특정 계층을 표현해 낼 수 있으므로 분석에 포함하였다.

4.5 다시점 부상패턴의 발견

재귀적 변화발견 기법을 이용하면 지금까지 지속적으로 변화되고 있는 연관성규칙들을 선별하여 보여줌으로써 고객 관리자가 장기적인 관점에서 중요하게 고려해야 하는 연관성규칙이 무엇인지 가시적으로 나타내 준다.

<표 3> 재귀적 변화발견 기법을 이용한 분석 결과의 개수

	2001~2002 (임계치)	2001~2003 (임계치)	2001~2004 (임계치)
브랜드명	153(10%)	1(10%)	0(10%)
판매금액	150(40%)	54(30%)	22(20%)
판매량	172(100%)	56(50%)	31(25%)

<표 3>을 살펴보면 변화의 개수가 두 시점간 비교 시 보다 현저히 줄어드는 것을 알 수 있다. 브랜드명을 결과로 한 연관성규칙의 경우 지속적으로 나타나는 변화는 없었다. 반면 판매금액, 판매량의 경우 지속적으로 나타나는 규칙이 있었으며 실제 결과 값은 <표 4>, <표 5>과 같다. <표 4>는 판매 금액을 결과로 생성된 연관성규칙의 변화를 4년 동안 지속적으로 살펴 본 결과이다. <표 4>에 나타난 다시점 부상패턴 중 가장 현저한 변화를 보인 내용은 첫 번째 레코드에 표시된 “판매금액=5 ← 성별코드=여자 & 장기거래=

예 & 브랜드명=센존”이다. 이는 ‘센존(St.John)’이라는 브랜드를 구매하고 장기거래 고객이면서 여성 고객인 그룹이 고객의 구매를 한다는 연관성 규칙의 지지도가 매년 임계치 이상 증가하여 과거 1.1%에 비해 2.5배 가량 증가한 2.8%를 차지한 것을 보여주고 있다. 해외명품 영역 전문가에 의하면, ‘센존’은 미국을 대표하는 여성 패션 브랜드 중 하나로 사회적으로 성공한 여유 있는 30~50대의 여성을 목표 고객으로 설정하고 있다고 한다. 이러한 맥락에서 볼 때, 상기의 다시점 부상패턴은 백화점 입장에서는 판매 목표에

<표 4> 2001~2004 다시점 부상패턴(판매금액)

규 칙 명	파일1 SUP	파일2 SUP	변화도
판매금액=5←성별코드=여자&장기거래=예&브랜드명=센존	1.10%	2.80%	154.55
판매금액=5←판매구분=정상판매&장기거래=예&브랜드명=센존	1.10%	2.70%	145.45
판매금액=5←고객유형=개인고객&판매구분=정상판매&장기거래=예&브랜드명=센존	1.10%	2.60%	136.36
판매금액=5←고객유형=개인고객&장기거래=예&브랜드명=센존	1.20%	2.80%	133.33
판매금액=5←장기거래=예&브랜드명=센존	1.30%	3.00%	130.77
판매금액=5←고객유형=개인고객&판매수량=1&장기거래=예&브랜드명=센존	1.00%	2.20%	120.00
판매금액=5←판매수량=1&장기거래=예&브랜드명=센존	1.10%	2.30%	109.09
판매금액=5←고객유형=개인고객&성별코드=여자&결혼여부=미혼&브랜드명=센존	1.10%	2.30%	109.09
판매금액=5←성별코드=여자&결혼여부=미혼&브랜드명=센존	1.20%	2.40%	100.00
판매금액=5←성별코드=여자&판매수량=1&결혼여부=미혼&브랜드명=센존	1.00%	2.00%	100.00
판매금액=5←판매구분=정상판매&성별코드=여자&결혼여부=미혼&브랜드명=센존	1.00%	2.00%	100.00
판매금액=5←고객유형=개인고객&판매구분=정상판매&결혼여부=미혼&브랜드명=센존	1.10%	2.10%	90.91
판매금액=5←고객유형=개인고객&결혼여부=미혼&브랜드명=센존	1.30%	2.40%	84.62
판매금액=5←판매구분=정상판매&성별코드=여자&판매수량=1&브랜드명=센존	1.40%	2.40%	71.43
판매금액=5←성별코드=여자&판매수량=1&브랜드명=센존	1.60%	2.70%	68.75
판매금액=5←고객유형=개인고객&판매수량=1&결혼여부=미혼&브랜드명=센존	1.20%	2.00%	66.67
판매금액=5←고객유형=개인고객&성별코드=여자&판매수량=1&브랜드명=센존	1.60%	2.60%	62.50
판매금액=5←고객유형=개인고객&판매구분=정상판매&판매수량=1&브랜드명=센존	1.60%	2.50%	56.25
판매금액=5←고객유형=개인고객&판매수량=1&브랜드명=센존	1.80%	2.80%	55.56
판매금액=5←판매수량=1&결혼여부=미혼&브랜드명=센존	1.50%	2.10%	40.00
판매금액=5←판매구분=정상판매&판매수량=1&결혼여부=미혼&브랜드명=센존	1.30%	1.80%	38.46
판매금액=5←결혼여부=미혼&장기거래=예&가입기간=2-5년&판매구분=행사판매	3.40%	2.70%	20.59

〈표 5〉 2001~2004 다시점 부상패턴(판매량)

규 칙 명	파일1 SUP	파일2 SUP	변화도
판매수량=1←고객유형=개인고객&판매금액=5&띠분류=개띠	1.20%	2.80%	91.67
판매수량=1←고객유형=개인고객&판매구분=정상판매&띠분류=개띠	1.20%	2.80%	91.67
판매수량=1←판매구분=정상판매&결혼여부=기혼&띠분류=개띠	1.80%	3.10%	72.22
판매수량=1←성별코드=여자&고객등급=최우수&결혼여부=기혼&가입기간=1-5년	1.40%	2.40%	71.43
판매수량=1←판매구분=정상판매&고객등급=최우수&결혼여부=기혼&가입기간=1-5년	1.30%	2.20%	69.23
판매수량=1←결혼여부=기혼&띠분류=개띠	2.10%	3.50%	66.67
판매수량=1←판매구분=정상판매&장기거래=예&가입기간=11-15년&연령코드=70대이상	1.40%	2.30%	64.29
판매수량=1←판매구분=정상판매&성별코드=여자&결혼여부=기혼&띠분류=개띠	1.60%	2.60%	62.50
판매수량=1←장기거래=예&고객등급=최우수&결혼여부=기혼&가입기간=11-15년	1.30%	2.10%	61.54
판매수량=1←성별코드=여자&결혼여부=기혼&띠분류=개띠	1.80%	2.90%	61.11
판매수량=1←성별코드=여자&장기거래=예&고객등급=VIP&주소=서대문구	4.80%	2.10%	56.25
판매수량=1←판매구분=정상판매&성별코드=여자&가입기간=11-15년&연령코드=70대이상	1.50%	2.30%	53.33
판매수량=1←성별코드=여자&고객등급=VIP&주소=서대문구	4.80%	2.30%	52.08
판매수량=1←고객유형=개인고객&성별코드=여자&고객등급=VIP&주소=서대문구	4.80%	2.30%	52.08
판매수량=1←성별코드=여자&결혼여부=미혼&고객등급=VIP&띠분류=닭띠	5.00%	2.50%	50.00
판매수량=1←고객유형=개인고객&판매구분=정상판매&결혼여부=기혼&띠분류=개띠	1.80%	2.70%	50.00
판매수량=1←판매구분=정상판매&가입기간=11-15년&연령코드=70이상	1.60%	2.40%	50.00
판매수량=1←고객유형=개인고객&판매구분=정상판매&가입기간=11-15년&연령코드=70대이상	1.60%	2.40%	50.00
판매수량=1←결혼여부=미혼&장기거래=예&고객등급=VIP&주소=서대문구	4.40%	2.30%	47.73
판매수량=1←성별코드=여자&장기거래=예&가입기간=11-15년&연령코드=70대이상	1.90%	2.80%	47.37
판매수량=1←결혼여부=미혼&고객등급=VIP&주소=서대문구	4.50%	2.50%	44.44
판매수량=1←고객유형=개인고객&결혼여부=미혼&고객등급=VIP&주소=서대문구	4.50%	2.50%	44.44
판매수량=1←장기거래=예&고객등급=최우수&가입기간=11-15년&연령코드=70대이상	1.20%	1.70%	41.67
판매수량=1←고객유형=개인고객&성별코드=여자&결혼여부=기혼&띠분류=개띠	1.80%	2.50%	38.89
판매수량=1←고객등급=최우수&결혼여부=기혼&가입기간=11-15년	1.90%	2.60%	36.84
판매수량=1←고객유형=개인고객&고객등급=최우수&결혼여부=기혼&가입기간=11-15년	1.90%	2.60%	36.84
판매수량=1←판매구분=정상판매&성별코드=여자&가입기간=11-15년&띠분류=개띠	1.20%	1.60%	33.33
판매수량=1←성별코드=여자&가입기간=11-15년&연령코드=70대이상	2.10%	2.80%	33.33
판매수량=1←고객유형=개인고객&성별코드=여자&가입기간=11-15년&연령코드=70대이상	2.10%	2.80%	33.33
판매수량=1←고객등급=최우수&가입기간=11-15년&연령코드=70대이상	1.30%	1.70%	30.77
판매수량=1←고객유형=개인고객&고객등급=최우수&가입기간=11-15년&연령코드=70대이상	1.30%	1.70%	30.77

위배되지 않는 바람직한 변화로 파악될 수 있을 뿐 아니라 향후에도 지속적인 비율의 증대가 예상되므로 “판매금액=5 & 성별코드=여자 & 장기거래=예 & 브랜드명=센존”인 타겟 집단에 대해 관계를 강화할 수 있는 캠페인을 마련하는 것이 요구 된다.

<표 5>는 판매량을 결과로 한 다시점 부상패턴들을 표시하고 있다. 판매량은 대체로 한번 방문했을 때, 고객들이 한 개의 제품을 주로 구매하기 때문에 결과 값이 주로 “판매수량=1”을 나타내고 있다. 그러나 이와 같은 결과에서도 금액이나 브랜드에 상관없이 지속적으로 구매하고 있는 고객층은 어떠한 특성을 지니고 있는지 알 수 있기 때문에 매우 유용한 결과라고 판단된다. <표 5>에 나타난 다시점 부상패턴 중 지지도에 있어서 가장 현저한 변화를 보인 내용은 첫 번째 레코드에 표시된 “판매수량=1 ← 고객유형=개인고객 & 판매금액=5 & 띠분류=개띠”이다. 이는 개인고객이면서 개띠의 고객 구매자가 1개의 명품을 구입하는 경향이 시간이 지날수록 더욱 뚜렷해지고 있음을 의미한다. 따라서, “개인고객 & 판매금액=5 & 띠분류=개띠”인 타겟 집단에 대해서 판매수량을 확대할 수 있도록 교차판매 캠페인을 수립하는 것도 고려해볼 만 하다.

위에 언급한 <표 4>, <표 5>에 대해 전반적으로 살펴본 결과, 판매량과 판매금액의 경우에는 2001~2004년까지 지속적으로 관찰되는 연관성 규칙이 있었으나, 브랜드 명의 경우에는 2003년, 즉 3년간 관찰된 연관성 규칙이 매우 소수 관찰되었을 뿐이다. 이는 브랜드명의 경우 유행이나 신제품에 따른 변화가 매우 다양하기 때문인 것으로 추정된다.

4.6 다시점 미예측 변화의 발견

본 절에서는 다시점 미예측변화에 대한 분석 결과를 제시한다.

<표 6> 분석 결과의 개수

	2001~2002 (임계치)	2001~2003 (임계치)	2001~2004 (임계치)
브랜드명	716(70%)	50(50%)	15(30%)
판매금액	134(50%)	24(40%)	4(30%)

<표 6>을 살펴보면 변화가 측정된 연관성 규칙의 개수가 시간이 지남에 따라 줄어드는 것을 확인할 수 있다. 또한 부상패턴(Emerging Pattern)의 결과와는 달리 브랜드를 결론부로 하는 연관성 규칙에서 다시점 미예측변화가 많음을 알 수 있다. 이는 브랜드 구매가 유행에 민감하고 다양한 형태의 구매 행동 변화가 있었기 때문인 것으로 해석될 수 있다.

다음의 <표 7>은 다시점에 걸쳐 발견된 변화를 나타낸 것이다. <표 7>에 나타난 다시점 미예측변화 중 가장 현저한 변화를 보인 내용은 첫 번째 레코드에 표시된 “브랜드명=에프로 ← 고객유형=개인고객 & 판매수량=1 & 고객등급=VIP & 주소=동작구”로 부터 “브랜드명=T.S.E ← 판매수량=1 & 고객등급=VIP & 띠분류=개띠 & 주소=동작구” 로의 변화 이다. 이를 해석하면 두 연관성 규칙의 조건부의 결합으로 만들어지는 “고객유형=개인고객 & 판매수량=1 & 고객등급=VIP & 주소=동작구 & 띠분류=개띠” 타겟집단의 브랜드 선택이 2001년 “브랜드명=에프로”에서 2002년 “브랜드명=T.S.E”로 변화되었고 이후에도 지속적으로 변화가 유지되고 있음을 의미한다. 따라서 변화된 브랜드에 대한 정보제공을 강화하는 등 선호 변화에 적극 대처해야 할 것이다.

<표 8>은 고객들의 구매금액의 변화를 살펴본 것이다. <표 8>에 나타난 다시점 미예측변화 중 가장 현저한 변화를 보인 내용은 첫 번째 레코드에 표시된 “판매금액=1 ← 성별코드=여자 & 판매수량=1 & 고객등급=VIP & 주소=용산구”로 부터 “판매금액=5 ← 성별코드=여자 & 판매수량=1 & 고객등급=VIP & 주소=용산구”로의 변화이다. 이를 해석하면 “여자 & 판매수량=1 &

〈표 7〉 2001~2004 다시점 미예측변화(브랜드)

파일 1 규칙	파일 2 규칙	변화도
브랜드명=에프프로←고객유형=개인고객&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&띠분류=개띠&주소=동작구	72.78
브랜드명=에프프로←성별코드=여자&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&띠분류=개띠&주소=동작구	72.78
브랜드명=에프프로←판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&띠분류=개띠&주소=동작구	72.78
브랜드명=에프프로←판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&띠분류=개띠&주소=동작구	68.74
브랜드명=에프프로←판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←고객유형=개인고객&판매수량=1&띠분류=개띠&주소=동작구	68.74
브랜드명=에프프로←성별코드=여자&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←성별코드=여자&판매수량=1&MVG여부=예&주소=동작구	65.12
브랜드명=에프프로←판매수량=1&장기거래=예&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&연령코드=60대&주소=동작구	65.12
브랜드명=에프프로←고객유형=개인고객&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&가입기간=6-10년&주소=동작구	62.48
브랜드명=에프프로←성별코드=여자&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&고객등급=VIP&가입기간=6-10년&주소=동작구	62.48
브랜드명=에프프로←판매수량=1&결혼여부=미혼&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←판매수량=1&결혼여부=미혼&MVG여부=예&주소=동작구	58.92
브랜드명=에프프로←판매수량=1&장기거래=예&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←성별코드=여자&MVG여부=예&띠분류=개띠&주소=동작구	47.59
브랜드명=에프프로←판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←성별코드=여자&연령코드=60대&주소=동작구	32.56
브랜드명=에프프로←판매수량=1&장기거래=예&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←장기거래=예&MVG여부=예&띠분류=개띠	31.73
브랜드명=에프프로←성별코드=여자&판매수량=1&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←성별코드=여자&결혼여부=미혼&장기거래=예&주소=동작구	30.27
브랜드명=에프프로←판매수량=1&장기거래=예&고객등급=VIP&주소=동작구	브랜드명=T.S.E ←장기거래=예&고객등급=VIP&MVG여부=예&주소=동작구	30.18

〈표 8〉 2001~2004 다시점 미예측변화(판매금액)

파일 1 규칙	파일 2 규칙	변화도
판매금액=1 ←성별코드=여자&판매수량=1&고객등급=VIP&주소=용산구	판매금액=5 ←성별코드=여자&판매수량=1&고객등급=VIP&주소=용산구	37.64
판매금액=1 ←성별코드=여자&결혼여부=미혼&고객등급=VIP&주소=용산구	판매금액=5 ←성별코드=여자&결혼여부=미혼&고객등급=VIP&주소=용산구	37.12
판매금액=1 ←성별코드=여자&결혼여부=미혼&고객등급=VIP&주소=용산구	판매금액=5 ←판매수량=1&결혼여부=미혼&고객등급=VIP&주소=용산구	33.35
판매금액=1 ←성별코드=여자&결혼여부=미혼&고객등급=VIP&주소=용산구	판매금액=5 ←성별코드=여자&판매수량=1&고객등급=VIP&주소=용산구	31.96

고객등급=VIP & 주소=용산구” 타겟집단의 구매 금액이 고액구매 경향으로 변화되고 있음을 의미한다. 따라서 이들 집단 중 여전히 구매금액이 낮은 고객들을 대상으로 지속적인 고액구매를 유도하기 위해 업셀링을 위한 캠페인을 수립하는 것도 유용한 정책이 될 것이다.

V. 결 론

본 연구에서는 기존의 연관성규칙 분석 기법을 이용한 고객의 변화 분석 방법을 다차원적인 시점으로 확장하여 분석할 수 있는 방법론을 제시함으로써, 기존 연구들의 한계점을 극복하고 고객 관리자에게 유용한 정보를 제공 하고자 하였다. 고객 관리자는 제안된 방법론을 통해 변화의 트렌드를 파악할 수 있으며 이에 대응하는 적절한 마케팅 캠페인을 수립할 수 있다. 예를 들어, 나이가 30대이고 강북구에 거주하는 고객들의 ‘테스토니’ 구매 빈도가 최근 증가했으며, 과거에도 다시점 부상패턴이 있었던 연관성규칙이라면, 백화점에서는 ‘테스토니’의 주 목표고객을 30대 강북구 거주 주민으로 결정할 수 있을 것이다. 본 연구에서 제시하고 있는 재귀적 변화 발견 방법론은 지속적인 변화들을 선별하여 가시적으로 제시해 줌으로써 고객 관리자의 정보 검색 노력과 시간을 감소시켜 준다.

본 연구는 다음과 같은 측면에서 실무적인 시사점을 줄 수 있다. 첫째, 니치 마케팅 캠페인을 위한 타겟 집단의 선정 시 본 방법론을 활용할 수 있다. 특히, 명품이나 고가의 상품일수록 고객집단을 매우 작게 세분화 하여 해당고객 집단의 행태를 깊이 이해하여 대응할 필요가 있는데 이를 위해서 본 방법론이 활용될 수 있다. 이는 본 방법론이 다시점 부상패턴 및 다시점 미세측 변화라는 관점에서 변화의 양상이 현저한 고객집단을 발견하는 것을 지원하기 때문이다. 둘째, 장기적인 고객 행위변화의 경향을 예측하고 이에 대비할 수 있다는 것이다. 다시점 부상패턴과

다시점 미세측변화는 장기적인 고객행위변화를 예측하는데 매우 중요한 요소가 될 수 있으며, 장기적인 예측이 가능하다면 단기적인 마케팅 캠페인 만이 아니라, 신상품 개발, 매장 확대 등의 장기적 전략 구상을 가능케 하기 때문이다.

본 연구에서는 분석 단위를 1년 단위로 하여 분석하였으나 분기 및 계절 등의 보다 미시적인 관점에서 데이터를 전처리하여 분석한다면 계절별 특성을 감안한 트렌드도 살펴볼 수 있을 것이라 판단된다. 따라서 향후 연구로써 다양한 분석 단위를 정하여 실험하고 그 결과를 해석해 볼 필요가 있다. 본 연구의 또 다른 한계는 탐지된 변화들이 상호배타적이지 못하고 유사한 변화들이 동시에 많이 발견되기 때문에 특정 고객 집단의 경우 중복적으로 캠페인 대상 고객에 선정될 수 있는 문제가 존재한다. <표 4>의 경우에도, 많은 유형의 변화들이 “판매금액=5 ← 장기거래=예 & 브랜드=센존”의 변형된 형태의 변화들이다. 이는 연관성 규칙분석을 이용한 변화탐지 방법의 특성 때문에 기인한 것으로서, 이 방법이 모든 변화를 누락 없이 발견할 수 있다는 장점은 있지만 유사한 변화들이 동시에 발견되어 해석을 어렵게 하는 단점을 동시에 내포하고 있다. 따라서, 의사결정 나무기법을 통해 생성된 분류규칙을 이용하여 변화를 발견하는 방법도 고려해 볼 만 하다

참 고 문 헌

- Agrawal, C. C., Z. Sun, and P. S. Yu, “Online algorithms for finding profile association rules”, *Proceedings of the 1998 ACM CIKM International Conference on Information and Knowledge Management*, 1998, pp. 86-95.
- Agrawal, R., T. Imielinski, and A. Swami, “Database Mining : A performance perspective”, *IEEE Trans. Knowledge and Data Engineering*, Vol.5, No.6, 1993, pp. 914-925.

- Agrawal, R. and G. Psaila, "Active data mining", *Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95)*, 1995, pp. 3-8.
- Agrawal, R. and R. Srikant, "Fast algorithm for mining association rules", *Proceedings of the International Conference on Very Large Databases (VLDB-94)*, 1994, pp. 487-499.
- Cheung, D. W., V. T. Ng, and B. W. Tam, "Maintenance of Discovered Knowledge: A Case in Multi-Level Association Rules", *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 1996, pp. 307-310.
- Das, G., D. Gunopulous, and H. Mannila, "Finding Similar Time Series", *Proceedings of the First European Symposium: Principles of Data Mining and Knowledge Discovery (PKDD '97)*, 1997, pp. 88-100.
- Das, G., K. Lin, H. Mannila, G. Renganathan, and P. Smyth, "Rule discovery from time series", *Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (KDD-98)*, 1998, pp. 16-22.
- Dong, G., and J. Li, "Efficient Mining of Emerging Patterns: Discovering Trends and Differences", *Proceedings of the Fifth International Conference on Knowledge Discovery and Data Mining (KDD-99)*, 1999, pp. 43-52.
- Feldman, R., Y. Aumann, A. Amir, and H. Manila, "Efficient algorithms for discovering frequent sets in incremental databases", *SIGMOD'97 Workshop on Research Issues in Data Mining and Knowledge Discovery (DMKD'97)*, 1997, pp. 59-66.
- Han, J., G. Dong, and Y. Yin, "Efficient Mining of Partial Periodic Patterns in Time Series Database", *Proceedings of the Fifteenth International Conference on Data Engineering (ICDE 1999)*, 1999, pp. 106-115.
- Hussain, F., H. Liu, E. Suzuki, and H. Lu, "Exception Rule Mining with a Relative Interestingness Measure", *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 2000, pp. 86-97.
- Li, J., G. Dong, and K. Ramamohanarao, "Making Use of the Most Expressive Jumping Emerging Pattern(s) for Classification", *Proceedings of Pacific Asia Conference on Knowledge Discovery in Databases (PAKDD)*, 2001, pp. 131-145.
- Liu, B. and W. Hsu, "Post-Analysis of Learned Rules", *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, 1996, pp. 828-834.
- Liu, B., W. Hsu, and S. Chen, "Using general impressions to analyze discovered classification rules", *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 1997, pp. 31-36.
- Liu, B., W. Hsu, H. S. Han, and Y. Xia, "Mining Changes for Real-Life Applications", *Second International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2000)*, 2000, pp. 337-346.
- Padmanabhan, B. and A. Tuzhilin, "Unexpectedness as a measure of interestingness in knowledge discovery", *Decision Support Systems*, Vol.27, 1999, pp. 303-318.
- Silberschatz, A. and A. Tuzhilin, "What makes patterns interesting in knowledge discovery systems", *IEEE Trans. On Knowledge and Data Engineering*, Vol.8, No.6, 1996, pp. 970-974.
- Song, H. S., J. K. KIM, and S. H. KIM, "Mining the change of customer behavior in an Internet shopping mall", *Expert systems with applica-*

tions, Vol.21, No.3, 2001, pp. 157-168.

Suzuki, E., "Autonomous discovery of reliable exception rules", *Proceedings of the Third International Conference on Knowledge Discovery and Data Mining (KDD-97)*, 1997, pp. 259-262.

Thomas, S., S. Bodagala, K. Alsabti, and S. Ranka,

"An efficient algorithm for the incremental updation of association rules in large databases", *Proceedings of the Third International Conference on Knowledge Discovery in Databases*, 1997, pp. 263-266.

A Recursive Procedure for Mining Continuous Change of Customer Purchase Behavior

Jae-Kyeong Kim* · Kyung-Hee Chae* · Ju-Cheol Choi*
Hee-Seok Song** · Yeong-Bin Cho***

Abstract

Association Rule Mining has been successfully used for mining knowledge in static environment but it provides limited features to discovery time-dependent knowledge from multi-point data set. The aim of this paper is to develop a methodology which detects changes of customer behavior automatically from customer profiles and sales data at different multi-point snapshots. This paper proposes a procedure named 'Recursive Change Mining' for detecting continuous change of customer purchase behavior. The Recursive Change Mining Procedure is basically extended association rule mining and it assures to discover continuous and repetitive changes from data sets which collected at multi-periods. A case study on L department store is also provided.

Keywords: CRM, Association Rule Mining, Recursive Change Mining

* School of Business, Kyung Hee University

** Department of MIS, Hannam University (Corresponding Author)

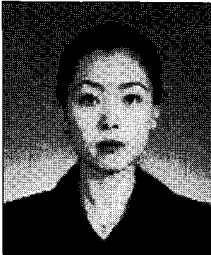
*** Department of Business Administration, Konkuk University

◎ 저 자 소개 ◎



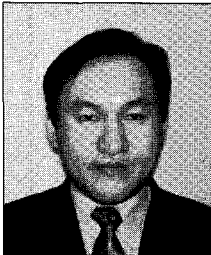
김 재 경 (jaek@khu.ac.kr)

서울대학교에서 산업공학 학사, 한국과학기술원에서 경영정보시스템 전공으로 산업공학 석사 및 박사학위를 취득하였으며 현재 경희대학교 경영대학 교수로 재직하고 있다. 미국 미네소타 주립대학교 경영정보학과에서 교환교수를 역임하였으며, 경희대학교 경영대학 정보센터 소장을 맡고 있다. 주요 연구분야는 비즈니스 인텔리전스, 상품추천시스템, e비즈니스 모형 및 전략 등이며, 주요저서로, e비즈니스 시대의 경영정보시스템(한경사, 2004), e비즈니스 시스템(울곡출판사, 2004), 비즈니스 인텔리전스(시그마 인사이트, 2004), e-Business(McGraw-Hill Korea, 2002), 의사결정분석 및 응용(영지문화사, 1999) 등이다.



채 경 희 (khchae@khu.ac.kr)

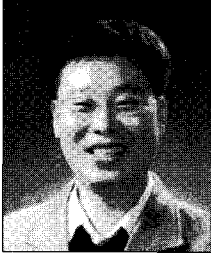
경희대학교에서 경영학 학사, 동 대학원에서 경영정보시스템 전공으로 경영학 석사를 취득하였으며, 현재 동 대학원 e비즈니스 전공 박사과정을 수료하였다. 주요 연구분야는 CRM, 데이터마이닝, e비즈니스 모형 및 전략 등이며, 연구 실적으로는 경영과학지(2004), ICEB(2004), 한국경영정보학회(2005) 등이 있다.



최 주 철 (kff122@ldcc.co.kr)

경희대학교에서 기계공학 학사, 동 대학원에서 경영정보시스템 전공으로 경영학 석사를 취득하였으며, 현재 동 대학원 e비즈니스 전공 박사연구과정에 재학 중이다. 롯데그룹 주요 계열사에서 16년간 근무하였으며 현재 롯데백화점 전산 정보팀장으로 재직 중이다. 주요관심분야는 CRM과 데이터마이닝, 모바일 전자상거래, 유통 정보화전략 등이다. 연구분야는 CRM, 데이터마이닝, e비즈니스 모형 등이며, 주요 연구 실적으로는 고객 구매행태의 지속적 변화 파악을 위한 확장된 연관성분석 방법(한국경영정보학회, 2005) 등이 있다.

○ 저자 소개 ○



송희석 (hssong@hannam.ac.kr)

고려대학교에서 경영학(경영학사)을 전공하였으며, KAIST에서 경영과학 전공으로 석사학위를 경영공학 전공으로 박사학위를 취득하였다. 대우정보시스템(주)에서 15년간 근무하였으며 현재 한남대학교 경영정보학과 조교수로 재직 중이다. *Information System Review* 편집위원을 맡고 있으며 주요 관심분야는 CRM과 데이터마이닝, 시맨틱웹, 모바일 전자상거래, 경영혁신과 정보화전략 등이다.



조영빈 (ybcho111@kku.ac.kr)

고려대학교에서 산업공학을 전공하였으며, KAIST에서 산업공학 전공으로 석사학위, 경영공학 전공으로 박사학위를 취득하였다. 한국생산성본부, 삼성경제연구소 등에서 15년간 근무하였으며 현재 건국대학교 사회과학대학 경영학과 조교수로 재직 중이다. 주요 관심분야는 CRM과 데이터마이닝, 사회연결망 분석 등이다.

논문접수일 : 2006년 1월 18일

1차 수정일 : 2006년 3월 27일

게재확정일 : 2006년 6월 19일

2차 수정일 : 2006년 5월 17일