

데이터마이닝을 활용한 동적인 고객분석에 따른 고객관계관리 기법

하성호
경북대학교 경영학부 교수
(hsh@knu.ac.kr)

이재신
경북대학교 대학원 경영학과 석사과정
(fury@korea.com)

전통적인 고객관계관리 연구는 특정 시점에서 고객관계관리에 중점을 두어 연구되었다. 이러한 정적인 고객관계 관리와 고객 행동에 관한 지식은 마케팅 관리자가 제한된 마케팅 자원을 이익의 극대화를 위해 사용할 수 있게 해주었다. 그러나 시간이 경과하게 되면 이러한 정적인 지식은 쓸모가 없어지게 된다. 그러므로 고객관계관리는 고객의 동적 특성을 반영해야 한다. 과거 고객의 구매 행위를 관찰하여 현재 또는 미래 시장의 고객을 세분화하며 구분된 고객 군집에 대해 서로 다른 마케팅 전략을 사용할 수 있다. 고객의 구매행동을 근간으로 한 고객관계관리는 수십 년 전부터 연구되어왔지만 동적인 고객관계관리에 대한 연구는 최근에 들어와서야 활발하게 진행되고 있다.

본 논문은 인터넷 상점의 고객 데이터로부터 추출된 지식과 시간 경과에 따른 고객 행동 패턴의 분석을 위해 데이터마이닝과 모니터링 에이전트 시스템(MAS)을 이용하며, 이를 통한 동적인 고객관계관리 모델을 제시한다. 이 모델은 고객이력경로에 대한 예측과 고객에게 나타나는 집단이력경로의 분석, 그리고 시간 경과에 따른 고객 군집의 변화에 대한 분석과 그에 따른 마케팅 전략 도출을 포함한다. 이 모델의 제안은 많은 온라인 소매상이 직면하고 있는 경영상의 문제를 해결하는데 유용할 것이다.

논문접수일 : 2003년 3월

게재확정일 : 2003년 7월

교신저자 : 하성호

1. 서론

최근 기업과 소비자간 전자상거래(Business-to-Consumer electronic commerce)의 성장은 온라인 소매산업의 경쟁을 심화시켰다. 치열한 경쟁 환경 속에서 생존하고 나아가 경쟁력을 확보하기 위해서는 고객의 요구를 충족시킬 수 있는 개인화된 서비스를 제공해야 한다. 이런 필요성에서 인터넷에서의 고객관계관리(customer relationship management: CRM)에 대한 연구가 활발히 이루어지고 있다.

고객관계관리의 목적은 고객을 파악하고 고객

의 구매 패턴을 이해하고 예측하며 적절한 마케팅 도구를 결정한 뒤 고객에게 개인화된 방법을 통해 직접 전달하는 것이다. 따라서 고객관계관리는 차별적인 프로모션 활동의 전제인 고객 집단을 표적화하는 마케팅 행위의 전제 조건으로 여겨지고 있다(Rossi *et al.*, 1996).

전자상거래에서 고객관계관리의 전형적인 예는 온라인 소매상이 매출액을 향상시키기 위해 마케팅 프로모션을 시행하는 경우이다. 온라인 소매업자는 구매 경험이 있는 사람을 고객으로 정의하고 있으며 인터넷의 폭발적인 성장으로 인해 이러한 고객과 시장에 대한 데이터 또한 기하

급수적으로 증가하게 되었다.

고객에 대한 정보의 획득은 더욱 용이해졌고 이를 통해 개별 고객을 대상으로 직접 마케팅(direct marketing)을 하는 것이 가능해졌다(Russel and Kamakura, 1994). 즉 풍부한 고객 정보는 마케팅 담당자로 하여금 직접 마케팅과 표적 결정에 유용한 구매 모델을 수립하고 이용할 수 있도록 하고 있다.

고객의 구매 행위를 측정하기 위해서 여러 변수들이 있지만 그 중에서도 Recency, Frequency, Monetary(RFM) 변수들이 많이 활용된다. RFM 값은 고객의 행동학적 특성을 표현하는 정보를 제공한다. 구체적으로 Recency는 최종 구매 이래로 얼마나 시간이 경과하였는지를 이야기하고, Frequency는 얼마나 자주 구매하였는지를, 그리고 Monetary는 얼마나 많이 구매하였는지를 나타낸다(Roberts and Berger, 1989; Hughes, 1996).

고객관계관리는 단기적으로는 어떤 고객에게 마케팅 인센티브를 제공하는 것이 고객의 즉각적인 재구매를 유도할 것인지 결정하는 것과 중기적으로는 신규 고객 유치와 기존 고객 유지에 얼마나 많은 프로모션 비용이 투입되어야 할 것인지를 결정하는 것, 그리고 장기적으로 기존 고객의 미래 가치를 제고할 수 있는 방안을 결정하기 위해 이용된다.

본 논문에서는 소매 고객의 구매행위를 분석하는 연구 목적을 달성하기 위해 온라인과 오프라인에서 행해지는 고객의 구매내역을 담은 고객 구매 데이터마트를 구축하였다. 정보처리와 저장에 소요되는 비용이 갈수록 저렴해지고 있는 상황은 저렴한 비용으로 대용량의 고객 구매 데이터마트 구축을 가능하도록 하고 있다. 이 데이터마트에서 추출된 고객 군집 지식은 소매상이 개

별 고객의 행위를 이해하고 다른 고객의 행위를 예측할 수 있는 원천으로 활용된다(Peppers and Rogers, 1999). 데이터마트에서 지식을 추출하는 과정으로 이해되고 있는 데이터마이닝 혹은 데이터베이스에서 지식 발견(KDD)은 자동화된(혹은 반자동화된) 도구를 활용하여 대량의 데이터에서 의미있는 패턴이나 규칙을 발견하는 탐색적 분석을 의미한다(Fayyad *et al.*, 1996; Berry and Linoff, 2000).

실제 세계의 고객은 변화하고 따라서 그들의 행위 패턴도 시간에 따라 변화하기 마련이다. 특정 시점에 추출된 고객에 관한 규칙들은 단지 추출시점에서만 유효할 뿐이며 금방 쓸모가 없어진다. 이러한 정적인 규칙을 사용하는 정적인 고객관계관리는 다음과 같은 실제 문제를 해결하기 위해서는 미흡한 면이 있다.

- 새로운 데이터에서 새로운 규칙들을 생성하는 주기
- 마케팅 캠페인 또는 프로모션의 효과를 측정하는 방법
- 현재 우량 고객의 수와 증가 또는 감소 경향
- 시간에 따른 고객 구매 행위의 변화와 마케팅 전략 구현

상기의 문제를 해결하기 위해서, 고객관계관리 시스템은 시간의 변화에 따른 고객의 변화를 적극 수용할 수 있어야 한다. 따라서 본 논문에서는 데이터마이닝 기법을 활용하는 동적인 고객관계 관리모델을 구축하고 시간의 변화에 따른 고객 행위 패턴을 분석하여 동적 지식을 추출한다.

시간이 지남에 따라 고객은 군집간을 이동하게 되며 고객은 RFM의 분류의 일종인 우량고객 군집이나 이탈고객 군집으로 이동할 수 있다. 각

고객의 군집간 이동 행위를 관찰함으로써 고객의 군집간 이동 이력에 대한 정보를 생성할 수 있으며, 시간 경과에 따른 고객 군집의 변화를 예측하는 것이 확률적 과정을 통해 이루어질 수 있다. 마코프 속성(전이 확률의 불변성)을 지닌 고객으로 인해 군집이 변화한다면, 확률적 과정의 한 형태인 마코프 체인을 이용해 시간 변화에 따른 고객 군집의 변화를 분석할 수 있다. 본 논문은 고객 군집의 단기간에 걸친 변화 가능성과 장기간의 걸친 변화 가능성에 대한 기술과 함께 그에 상응하는 경영전략을 제시할 것이다.

본 논문은 다음과 같이 구성되어 있다. 2장에서는 전자상거래에서 고객관계관리 모형의 최근 경향에 대한 문헌연구의 내용을 담고 있으며, 동적인 고객 행동 데이터를 분석하기 위한 동적 고객관계관리 방법론은 3장에서 제시하고 있다. 4장은 본 논문에서 제시된 방법론의 특징을 보여주는 응용 시스템에 대한 적용 사례를 담고 있으며, 5장에서는 본 논문에 대한 견해와 토의를 담고 있다.

2. 문헌연구

고객관계관리의 핵심은 고객과의 관계를 더욱 증진하는데 있다. 고객관계관리는 ‘고객에 대해 단지 아는 것이 아니라, 고객이 말하는 것에 바탕을 두고 고객에 대한 행동을 기꺼이 변화시키는 것이다’(Peppers *et al.*, 1999).

기존 고객이 새로운 고객보다 더 유익하다는 전제는 기존의 고객에게 교차판매를 하는데 있어서 더 적은 비용이 소요되는 점과 고객의 기대에 더 부합하는 서비스 수준과 상품의 제공으로 고객 유지를 극대화 할 수 있다는 것과 일치한다.

이에 반해 새로운 고객을 만든다는 것은 기존의 고객을 유지하는 것에 비해 더 많은 비용을 요구한다. 그러므로 고객관계관리의 가장 중요한 목적은 기업에 있어서 고객의 존속 가치를 극대화하는 것이다(Peppard, 2000).

전통적인 정적 고객관계관리는 본질적으로 기업의 의사결정을 위해 고객 정보를 분석하는 것이다. 이러한 고객관계관리의 목적은 고객의 요구를 이해하고 시장 세분화를 통해 어떤 고객이 기업에게 더 유익한가에 따라 고객을 구분하는 것이며(Hughes, 1996), 형성된 고객 집단의 고객 충성도와 수익성, 채널의 효과성과 수익성을 분석하여 차별화된 프로모션을 시행하는 것이다.

일반적으로 온라인 소매업자는 마케팅 관점에서 자신의 사이트의 효과성을 분석하게 된다. 웹 마케팅은 온라인 상점에서 고객을 획득하고, 그 고객을 유지하는 활동으로 폭넓게 정의할 수 있다(Schafer *et al.*, 2001).

온라인 마케팅은 데이터베이스 마케팅, 일대일 마케팅 그리고 표적 광고(ad targeting) 등의 기법을 이용한다.

데이터베이스 마케팅은 더욱 개인화된 서비스를 고객에게 제공하려는 시도로써 고객의 인구통계적 특성을 근거로 하여 군집으로 나누고 각각의 군집에 대해 서로 다른 마케팅 정책을 사용하게 된다.

마케팅의 환경이 일대일 환경으로 옮겨감에 따라 고객에 대한 일대일 마케팅은 필수적이 되었다. 일대일 마케팅은 각각의 고객에게 개별적으로 대응할 수 있도록 해주는 기술을 사용함으로써 마케팅의 몰개성적인 속성을 극복하려고 시도하고 있다(Peppers and Rogers, 1997). 일대일 마케팅 프로그램은 고객의 거래 행동, 구매가능성, 추가구매가능성 등의 고객정보와 데이터로

가능하다. 더욱이 기업에 있어서 고객의 수익성을 이해하고, 높은 수익성을 가진 고객을 재교육하고 획득하는데 필요한 자원의 중요성이 부각될수록 고객에 대한 정보는 성공적인 마케팅 전략 수행에 있어서 매우 중요한 요소가 된다고 할 수 있다(Peppers and Rogers, 1999).

추천 시스템은 소매업자가 일대일 마케팅 전략을 수행하는데 도움을 줄 수 있다. 많은 수의 웹 기반 개인 추천시스템(Web-based Personalized Recommendation Systems)은 최근에 들어 제안되었다(Konstan *et al.*, 1997; Resnick and Varian, 1997; Borchers *et al.*, 1998; Aggarwal *et al.*, 1999). 개인화(Personalization)는 개인의 프로필과 같은 약간의 정보를 통해 관련 상품이나 웹 페이지 같은 아이템 후보 집합을 필터링함으로써 그 역할을 한다. 필터링에는 크게 내용기반(content-based) 필터링과 협업(collaborative) 필터링 두 가지 방식이 있다.

내용기반 필터링 시스템은 고객이 과거에 선호했던 아이템으로부터 유사성을 근거로 하여 아이템을 추천하는 시스템이다. 일반적으로 아이템과 프로필은 공간에서 벡터로 나타낼 수 있으며 이들의 유사성은 유클리드 거리와 같은 표준거리 척도로 계산될 수 있다. 협업 필터링 시스템은 고객에 대한 개별적인 아이템 추천을 위해 그 고객이 속한 집단에 대한 정보를 이용하여 목표 고객과 비슷한 성향을 지닌 고객들이 선호하는 아이템을 추천하는 시스템이다.

표적 광고는 고객의 이전 행동을 기반으로 고객이 주문할 만한 아이템을 구분하여 고객에게 광고를 제공하는 것이다.

2.1 시장 세분화

시장을 세분화하는 것은 고객이 그 선호와 구매 행동에 있어서 이질적이라는 가정에 그 근간을 두고 있다(Green, 1977; Abell, 1980). 이러한 다양성은 일반적으로 제품과 사용자 특성의 차이로 설명된다(Kalwani and Morrison, 1977; Mahajan and Jain, 1978; Beane and Ennis, 1987). 그러므로 마케팅 담당자는 단 하나의 제품 또는 서비스로 시장의 모든 고객을 만족시킬 수 없다. 고객은 다수이고, 넓게 분산되어 있으며, 구매 욕구 또한 매우 다양하다(Kotler, 1991). 이러한 다양한 고객의 요구를 만족시킨다는 것은 상대적으로 비슷한 욕구를 지닌 고객 집단에 대해 각각 집중함으로써 더 효과적으로 달성될 수 있다(Choffray and Lilien, 1978). 그러므로 시장 세분화는 마케팅의 중심에 고객을 위치시키고자 하는 마케팅 개념의 논리적 발전이라 할 수 있다(Blattberg and Sen, 1976).

시장 세분화의 역할에 대한 연구논문과 응용 논문은 증가하는 추세이며, 이러한 문헌들은 두 가지 영역으로 구분된다. 그 첫 번째 영역은 고객 군집을 구분하고 분석하며 다른 세분화에 대한 기초변수의 적합도를 검증하기 위한 기술로써 고객 세분화를 고려하는 것이다. 이러한 의미에서 Bonoma와 Shapiro(1983)는 개별적인 고객을 군집으로 구성하는 과정만이 아니라 전체 시장을 군집으로 나누는 과정도 시장 세분화로 정의한다. 두 번째 영역은 Mahajan과 Jain(1978), Johnson과 Flodhammer(1980)가 정의한 것과 같이 군집화가 이루어진 시장에 대해 자원의 할당과 판별에 관한 측정과 분석의 방법으로써 시장 세분화를 정의한다.

2.2 시장 세분화 과정

Blattberg 등(1978)에 의하면, 세분화 분석의 경영적 유용성은 고객의 특징적인 구매 행동을 이용하여 동질적인 그룹으로 얼마나 잘 구분되는가에 의존한다. 때때로 규범적인 세분화로 여겨지는 그룹을 구분하는 과정이나 그 그룹에 대한 마케팅 자원의 할당을 결정하는 과정은 마케팅이나 전략경영 문헌에서 많은 주목을 받아왔다(Frank *et al.*, 1972; Mahajan and Jain, 1978). Johnson과 Flodhammer(1980)는 시장을 세분화하는데 있어서 폭넓게 수용되며 응용 가능한 과정을 다음과 같은 4단계로 제시한다.

- ① 세분화를 위한 원칙을 정하고 적절한 변수를 선택
- ② 그룹 전체 혹은 세분화된 시장에 대해 다변량 분석을 이용
- ③ 결과를 평가하고 확인
- ④ 경제적 관점에서 결과를 분석

2.3 시장 세분화 디자인을 위한 가이드라인

마케팅 담당자는 시장 세분화 디자인을 할 때 어떠한 분석이 행해질 수 있는지 또한 그 결과를 어떻게 이용할 수 있는지를 결정해야 한다. 일반적으로 디자인에 대해 선택할 수 있는 많은 대안이 있지만 세분화의 구조는 세분화에 사용된 척도, 군집의 수, 시간의 경과, 세분화에 사용된 기법, 선택된 고객의 수와 같은 요인에 많이 의존하게 된다.

2.3.1 세분화에 사용된 척도

세분화에 이용된 변수들은 일반적으로 지리,

인구통계, 심리, 관습, 이해관계 그리고 행동 같은 변수 가운데 하나 혹은 두 개 이상의 조합으로부터 도출할 수 있다(Weinstein, 1994). 이러한 변수들 가운데 인구통계적 특징은 비교적 쉽고, 매우 유용하며, 저렴한 세분화 변수로 여겨진다(Griffith and Pol, 1994; Dibb and Stern, 1995). 반면 RFM(Bult and Wansbeek, 1995; Hughes, 2000) 또는 FRAT(Frequency, Recency, Amount, Type)(Morwitz and Schmittlein, 1998; Stone *et al.*, 2001)과 같은 고객의 구매에 대한 행태적 척도는 인구통계적 정보의 어떠한 조합보다 미래의 고객 행동 예측에 있어서 더 정확함을 주장하고 있다. 인구통계적 정보는 단지 고객의 수입, 나이, 자녀의 수, 주택소유 여부 등과 같은 고객의 상태에 관한 정보를 제공하므로 고객의 행동 예측에 유용하지 않지만, RFM과 FRAT 같은 척도는 고객이 어떠한 형태의 제품과 서비스를 언제 구매하고, 얼마나 자주 구매하며, 얼마나 많이 구매하는지에 대한 정보를 제공하므로 고객 행동 예측에 있어서 유용함을 주장한다.

2.3.2 세분화된 군집의 수

최적군집의 수를 결정하는 노력은 계속되어왔다. Bonoma와 Shapiro(1983)는 세분화에 있어서 최적군집의 기준으로 조합수의 최소화와 세분화 기회비용의 최소화를 제시하였으며, Ladany와 Arbel(1991)은 최적화된 시장 구분을 결정하기 위해 같은 크기의 군집을 유도하는 방법을 제시하였다. 반면 몇몇의 연구자들은 해석가능성(interpretability)과 응용가능성(applicability)과 같은 관점에서 개별적 해법인 관리도구가 세분화 분석에 있어서 핵심이라고 생각한다(Doyle and Saunders, 1985). Chakravarti 등(1987)은 군집의

수를 최적화할 수 있을지라도 그것을 결정하는 것은 매우 어려운 일이므로 군집의 수는 식견있는 결정에 기반하여야 한다고 주장한다. 또한 Dibb과 Stern(1995)은 군집의 수는 해석가능성, 응용가능성, 중요도의 정도에 의해 결정된다고 주장하였다.

2.3.3 시간 경과에 대한 관점

시장 세분화 디자인에 있어서, 가장 간단하면서도 대중적인 가정은 시간이 경과해도 세분화의 대상은 변화하지 않는다는 것이다. 이러한 가정 하에서는 시간이 경과해도 군집은 변하지 않으며 그 군집에 속한 사람들 또한 변하지 않는다. 상대적으로 안정된 시장이라면 이러한 가정은 적절하다고 볼 수 있다.

불행하게도 시장이 불안정하다면 사람들이 변하지 않는다는 가정은 적절하지 않다. 이러한 단점을 극복하는 하나의 방법은 상황기반(occasion-based) 디자인을 이용하는 것이다(Wyner, 1992). 사람들은 제품을 사용하거나 구매하는 경우에 있어서 다양한 욕구를 가진다고 가정되며 특정한 시간이나 상태에 의해 여러 상황 중 하나가 선택된다. 사람을 이해하기 위해서는 그가 가질 수 있는 경우의 수에 관한 완전한 정보가 함축되어 있어야 한다.

불안정성을 극복하는 또 다른 방법은 같은 고객을 일련의 시간에 걸쳐 반복적으로 측정하여 시간을 기준으로 군집화한 고객에 관한 정보를 축적하는 것이다. 이것을 고객 패널 데이터(customer panel data)라 하며, 동적인 시장에서는 중요한 고객들의 군집간 이동을 쉽게 관찰할 수 있다. 이러한 이동은 고객의 삶의 변화나 부의 증대와 같은 외부적인 요인에서 기인하는 것

으로 볼 수 있다. 시간을 기준으로 군집화된 고객의 데이터를 이용할 수 있다면 각각의 고객을 안정성과 변화성에 따라 분석할 수 있으며 고객에 관한 더 자세한 정보를 얻음으로써 이러한 군집에 대한 정의는 변화할 수 있다.

2.3.4 세분화 기법(방법론)

일반적으로 시장 세분화에 대한 방법론은 두가지 영역으로 구분된다. 첫 번째는 k -평균 알고리즘, 요인분석, 판별분석, 회귀분석과 같은 전통적인 통계적 세분화 기법을 사용하는 것이다. 두 번째는 비교적 최근에 들어 소개된 신경망(neural network)과 같은 기법을 이용하는 것이다.

시장 세분화에 대한 대안적인 방법론으로 신경망의 잠재력을 소개한 논문이 출간되어 있다(Balakrishnan *et al.*, 1996). 몇몇의 결과는 신경망이 판별분석이나 회귀분석보다 더 정확한 분류를 제공한다고 소개하고 있으며(Fish *et al.*, 1995), 마케팅 담당자는 신경망의 이용을 통해 해당 시장을 적절히 세분화할 수 있다고 주장하고 있다(West *et al.*, 1997). 특별히 신경망의 한 종류인 SOM(self-organizing map)은 세분화에 있어서 전통적인 통계적 방법에 비해 매우 좋은 결과를 제공할 수 있다(Kiang *et al.*, 1995; Mangiameli *et al.*, 1996).

2.3.5 고객의 수

시장 세분화는 모든 고객을 대상으로 하거나 고객 중 제한된 수를 대상으로 시행할 수 있다(Wyner, 1992). 시장 세분화가 제한된 수의 고객을 대상으로 이루어졌다면 마케팅 담당자는 각각의 군집에 전체 고객이 어떻게 분포하는지 예측해야 하며 샘플이 모집단을 잘 대변하는지 검증

해야 한다(Ha and Park, 1998). 즉, 통계적 추론을 통해 전체 고객의 분포를 예측해야 한다(Shepard, 1995; Morwitz and Schmittlein, 1998).

반면, 고객 세분화가 전체 고객을 대상으로 이루어졌다면 각각의 고객이 속한 군집에 대해 마케팅 담당자는 미리 알 수 있다. 마케팅 담당자는 각각의 군집에 적합하도록 개발된 메시지로 각 군집에 속한 고객과 직접 연락을 취할 수 있다. DM(direct mail)을 발송하는 경우를 가정한다면, 고객이 속한 군집에 대해 적절히 개발된 메시지를 각각의 고객에게 전송할 수 있다.

3. 동적 고객관계관리와 데이터마이닝을 활용하는 모니터링 에이전트 시스템

정적인 고객관계관리 시스템을 개발하는 과정은 소매상의 고객을 확인하고 특정 기간 동안의 구매 이력을 담은 구매 데이터마트를 구축하는 것으로부터 시작된다.

고객구매 데이터마트를 기업 인트라넷에 구축한 후, RFM 값과 같은 요약 정보를 데이터마트에서 추출하고 구매행위 패턴을 마이닝한다.

인공신경망의 일종인 SOM을 사용하여 소매상의 고객들을 유사한 RFM 값을 가지는 고객군집으로 나누고 각 고객을 결과군집에 할당한다(Kohonen, 1990). 그리고 의사결정나무 분류기인 C4.5를 활용, 고객군집을 클래스로 하고 RFM 값들을 속성으로 하여 고객을 우량고객과 이탈가능고객으로 분류한다(Quinlan, 1993). 나아가서, C4.5의 학습에 포함되지 않은 나머지 고객들을 분류하기 위해 분류규칙을 생성한다. 그리고 난

후, 고객군집의 전략적 포지셔닝을 통하여 마케팅 캠페인이나 프로모션의 대상이 될 고객을 선정한다(Ha and Park, 1998).

대부분의 전자상거래 사례에서 고객의 구매행위는 결코 정적이지 않으며 시간에 따라 크게 변화하는 모습을 보인다. 따라서 구축된 고객구매 데이터마트는 시간에 따라 변화를 거듭하고 그 데이터마트에서 추출한 규칙들도 특정 시간에만 유효하고 시간이 흐름에 따라 점차 쓸모가 없게 된다. 따라서 고객관계관리 시스템은 시간에 대해 적응 가능하여야 하며 이것이 전자상거래 환경에서 동적인 고객관계관리 시스템이 요구되는 이유가 된다.

본 논문에서는 동적인 고객관계관리가 용이하도록 <그림 1>에서 제시하는 모니터링 에이전트 시스템(Monitoring Agent System: MAS)을 활용한다.

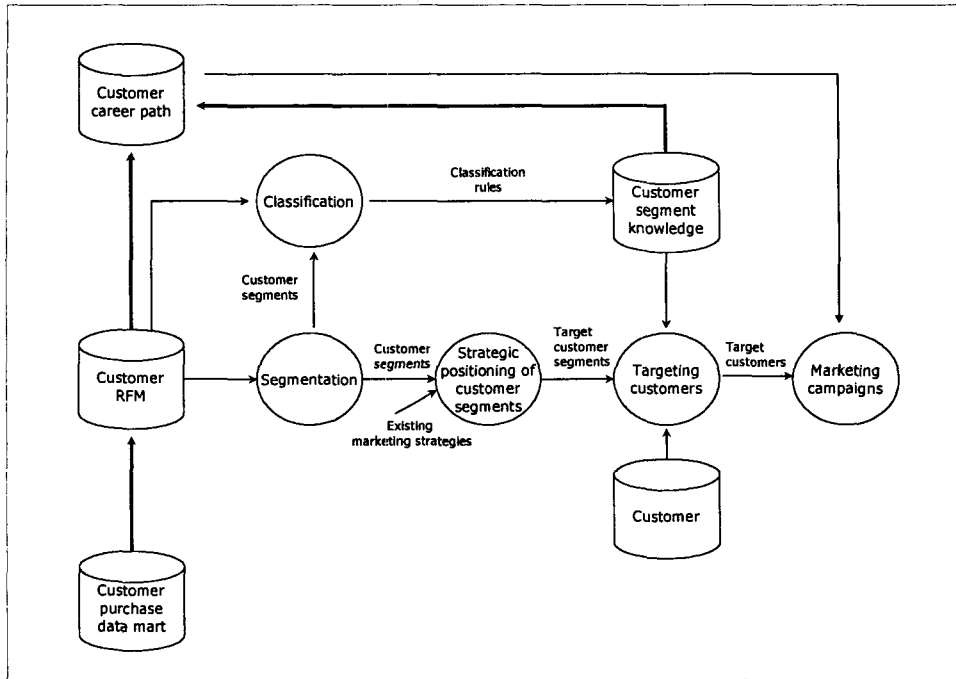
3.1 모니터링 에이전트 시스템

3.1.1 모니터링 에이전트 시스템의 역할

모니터링 에이전트 시스템은 동적인 고객관계관리에 있어서 다음과 같은 매우 중요한 역할을 수행한다.

■ 규칙 재생성의 시기

모니터링 에이전트 시스템은 새로운 구매 데이터가 데이터마트에 저장될 때마다 고객 구매 정보와 기존의 고객군집을 분류한 지식 사이의 차이를 항상 감시한다. 둘의 차이가 미리 정한 임계치를 상회하면 지식 관리자에게 고객에 대한 새로운 지식을 생성할 때임을 알리는 경고 메시지를 발송한다(Early Warning). 모니터링 에이전트에서 기존 규칙에 대해 반하는(Confliction) 규



<그림 1> 동적 고객관계관리를 구현하는 모니터링 에이전트 시스템 구조도

칙의 비율은 조기 경보를 발생시키는 임계치로서 이용한다. 그러기 위해서 모니터링 에이전트는 소규모의 샘플링과 고객 분류 규칙을 생성할 수 있는 데이터마이닝 능력을 보유해야 한다.

■ 표적 마케팅 캠페인 또는 프로모션의 효율성을 측정

고객의 구매는 광고, 가격할인과 같은 마케팅 캠페인 요인에 의해 영향을 받게 된다. 데이터마트에 저장되는 데이터가 증가함에 따라 마케팅 담당자는 앞으로의 마케팅 활동의 근거를 얻기 위해 기존 마케팅 활동의 효과를 측정할 수 있어야 한다. 마케팅 담당자의 입장에서는 새로운 마케팅 전략이나 인센티브 정책을 시행할 때 새로운 전략과 정책의 영향을 분석하기 위한 자료를

수집하고 분석할 수 있어야 한다. 일반적인 형태의 분석으로는 아래와 같은 것들이 있다.

- 마케팅 프로모션을 통해 실현된 이익
- 각 고객 군집별 이익
- 현재 프로모션 집행 비용
- 현재 프로모션과 이전 프로모션의 성과비교

표적 마케팅 캠페인의 효과를 측정하는 방법은 특정 마케팅 캠페인을 집행한 후 우량고객 군에 속하는 고객의 수가 증가하면 해당 캠페인은 효과적인 것으로 판단할 수 있다. 모니터링 에이전트는 마케팅 관리자가 고객 유지와 매출액 증가를 위해 더 나은 마케팅 기법을 선택할 수 있도록 도와준다.

■ 군집이동 경보(Proactive Prevention)

고객 구매 패턴을 분석하고 고객의 유형이 파악되었다면 모니터링 에이전트는 특정 유형의 고객이 예상과 달리 행동할 경우 경고 메시지를 마케팅 담당자에게 발송할 수 있다. 고객의 유형이 바뀌는 경우는 이탈과 같은 나쁜 표시도, 교차 판매와 같은 좋은 표시도 될 수 있으므로 고객의 일상적인 행위에 있어 변화가 있는가를 주시하면서 변화가 포착될 경우 마케팅 담당자가 적절하게 대응할 수 있도록 사전에 알려주는 것이 중요하다.

모니터링 에이전트 시스템은 고객이력경로 데이터베이스(customer career path database)를 포함하고 있다. 의사결정나무 분류기인 C4.5를 이용하여 추출한 고객 분류 규칙은 데이터마트 내 전체 고객을 분류하는데 활용되고 이 분류 정보는 고객이력경로 데이터베이스에 저장된다. 시간이 경과함에 따라 이력경로 데이터베이스 내 특정 고객의 군집이동 정보가 누적되고 각 고객

의 군집간 이동 양상을 모으면 각 고객의 이력경로가 만들어지게 된다.

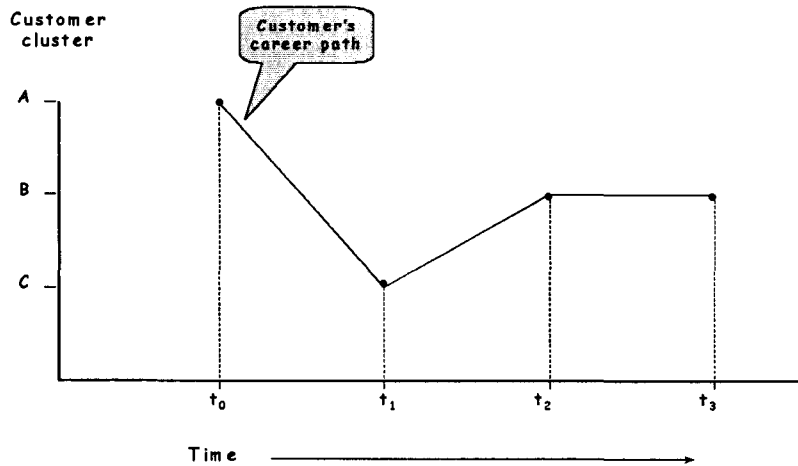
각 고객의 이력경로를 추출할 수 있도록 고객 이력경로 데이터베이스의 데이터는 다음 세가지 값들을 반드시 가져야 한다.

- 고객번호와 같은 고객식별 정보
- 고객군집과 분류지식
- 군집 이동시점

3.1.2 시간에 따른 고객 이력경로

■ 개인별 이력경로

<그림 2>는 샘플 고객의 이력경로를 보여준다. 간단한 예를 들기 위해 고객이 속할 수 있는 군집이 세 개가 있고 A, B, C라고 명명할 경우, 샘플 고객은 t_0 기에 A 군집, t_1 기에 C 군집, t_2 기에 군집 B, t_3 기에 군집 B에 각각 속한다. 이 경우, 고객의 이력경로는 A → C → B → B가 된다. 동일 고객의 군집간 이동은 순차적인 시계열



<그림 2> 고객 이력 경로

의 형태를 띠게 되며 이동 전후의 소속 군집을 보여주게 된다. 이 시계열 자료를 이용하면 시간의 경과에 따른 고객의 이력경로를 분석할 수 있고 다음 기에 어느 군집에 소속할 것인지를 예측할 수도 있다.

따라서 마케팅 담당자는 고객이 속할 것으로 예측된 군집의 특성에 따라 적절한 마케팅 전략을 집행할 수 있다. 예를 들어 이탈 가능 패턴을 보이는 군집에 속하는 것으로 판명된 고객에게는 고객 재활성 전략을 집행하고, 우량고객 군집에 속할 것으로 예측된 고객에 대해서는 고객 충성도를 유지하기 위한 고객유지 전략과 더욱 높은 충성도를 가질 수 있도록 고객 성숙 전략을 구사하면 좋다.

구체적으로 고객의 이력경로를 예측하는 방법으로는 유사한 이력을 가진 나머지 고객의 군집 이동 패턴을 이용하는 방법(협업 필터링)을 활용한다.

■ 집단 이력경로

특정 기간 동안에 소매상의 대부분 고객이 보이는 이력경로(집단 이력경로)를 추출하기 위해, 고객의 군집간 이동 가능한 형태와 실제 이동 회수를 계산하고 서로 비교해본다. 이동 회수가 많을수록 집단 이력경로가 될 소지가 많다. 두 가지 형태의 집단 이력경로가 관심을 끄는데, 하나는 우량고객 군집으로 이르는 집단 이력경로이고, 다른 하나는 이탈가능 군집으로 이르는 집단 이력경로이다. 우량고객 군집으로 이르는 집단 이력경로를 파악하면 이와 유사한 이력경로를 보이는 개별 고객을 식별하고 궁극적으로 우량 고객이 될 수 있도록 이력을 개발할 수 있다(성숙 전략). 이탈가능 군집으로 이르는 집단 이력경로

를 파악하는 것은 마케팅 담당자가 이력경로 상에 존재하는 고객이 경쟁사로 이탈하는 것을 막는 전략을 사전에 구사할 수 있도록 도와줄 수 있다(반이탈 전략).

3.1.3 고객 군집의 변화 예측

본 논문에서 제안하는 동적 고객관계관리 모델은 추계적 과정의 하나인 마코프 체인을 통해 시간의 경과에 따른 고객 군집의 변화를 미리 평가할 수 있도록 하는 분석적 방법을 포함한다. 구체적으로 고객 군집의 변화를 예측하면 다음과 같은 장점을 가질 수 있다.

- 마케팅 담당자의 입장에서는 우량 고객 군집이 향후 축소될 것인지 아니면 확대될 것인지를 예측하는 것이 중요하다. 우량 고객의 수가 줄어든다는 것은 미래 수익이 감소될 수 있다는 것을 의미하기 때문이다.
- 고객의 군집간 이동을 관찰하는 것은 마케팅 담당자가 마케팅 캠페인이나 프로모션의 형태로 전개하는 마케팅 전략의 효과를 측정하고 평가할 수 있는 단서를 제공한다.
- 소매상이 여러 대안 마케팅 전략을 고려하고 있을 때, 단기나 장기 수익 증대를 가져올 수 있는 마케팅 전략을 선택할 수 있도록 해준다.

고객의 군집간 이동은 고객이 과거에 어떤 군집에 속하고 있었는가에 상관없이 이루어지므로 마코비안 속성(무기억 속성: memory-less property)을 만족한다고 볼 수 있다. 따라서 본 논문에서는 고객 군집의 변화를 예측하는데 마코프 체인을 이용한다.

일반적으로 마코프 체인은 과거에 일어났던 이벤트와 독립적으로 현재의 상태만이 미래에 전개될 상태를 결정하는 특성을 가진다. 게다가, 마코프 체인은 유한개의 상태공간(state space)으로 구성된다. 마코프 체인의 특성을 조건부 확률로 묘사하면 다음과 같은 형태가 나타나게 된다 (Hillier and Lieberman, 1995).

$$\begin{aligned} & \text{모든 상태 } i_0, i_1, \dots, i_{n-1}, i, j \text{에 대하여,} \\ & P\{X_{n+1} = j / X_n = i, X_{n-1} = i_{n-1}, K, X_1 = i_1, X_0 = i_0\} \\ & = P\{X_{n+1} = j / X_n = i\} = P_{ij} \end{aligned}$$

고객 군집의 변화를 예측하기 위해 마코프 체인을 이용하는 것을 설명하기 위해 다음과 같은 고객 군집이 있다고 하자. 특정 0, 1, 2, ... 시점에서 각 고객은 유한 개수의 상호 배타적인 상태 (A, B, C, ...)에 속할 수 있다. 여기서 상태 A, B, C, ... 는 고객 군집을 의미한다. 모니터링 에이전트 시스템은 마코프 체인을 형성하면서 다음과 같은 전이 확률을 정의한다. 즉 n-1기에 고객 군집 A에 속하던 고객이 n기에 군집 B에 속하게 될 확률은 P_{AB}^n 로 정의한다. 마코프 체인의 확률은 전이 확률표(matrix of transition probabilities)의 형태로 만들어진다.

■ 전이 확률표의 안정도

고객 군집의 변화를 예측하기 위해 마코프 체

인을 이용하는 것을 설명하기 위해 A, B, C 세 개의 고객 군집만 있고 (3-상태 마코프체인) 각 군집 당 100명씩 총 300명의 고객이 있으며 <표 1>과 같이 관찰 기간 동안 군집간에 고객 이동이 이루어졌다고 가정해보자.

<표 1>은 군집 이동 전후의 각 군집별 고객 수를 표시하고 있다. 관찰 기간 동안 총 고객 수는 300명으로 변화가 없지만 군집 A와 B에 속하는 고객의 수는 감소하고 군집 C의 고객 수는 증가하고 있다.

<표 1>의 군집간 고객 이동 자료를 이용하면 전이 확률을 계산할 수 있는데 이 확률은 고객이력 경로 데이터베이스에 저장되어있는 고객별 이력 자료로부터 추출할 수 있다.

<표 2>의 전이 확률표에서 각 행은 군집별 고객 유지와 다른 군집으로의 유출을 보여주며, 각 열은 군집별 고객 유지와 다른 군집으로부터의 고객 유입을 보여준다. 최초 군집 전이 확률은 이번 기간에 고객이 군집 A에 있었다면 다음 기간에 군집 A에 머무를 확률이 0.5이고 군집 B로 유출될 확률이 0.1이며, 군집 C로 유출될 확률이 0.4임을 보여준다. P_{AA} 가 0.5로 계산된 것은 전이 이후에도 군집 A에 남아있을 고객의 수를 이전기에 군집 A에 있는 고객의 수로 나누었기 때문이다. 나머지 전이 확률도 동일한 계산법에 따라 얻을 수 있다.

<표 1> 고객 군집간 고객 이동

군집	이동 전고객수	군집간 이동						이동 후 고객수
		A로	B로	C로	A에서	B에서	C에서	
A	100	50	10	40	50	20	20	90
B	100	20	20	60	10	20	50	80
C	100	20	50	30	40	60	30	130
	300							300

<표 2> 전이 확률표

상태 (고객군집)	A	B	C	$\sum_j P_{ij}$
A	50/100=0.5	10/100=0.1	40/100=0.4	1.0
B	20/100=0.2	20/100=0.2	60/100=0.6	1.0
C	20/100=0.2	50/100=0.5	30/100=0.3	1.0

마코프 체인의 중요한 특징 중 하나는 전이 확률이 변화하지 않을 경우 장기에 일반 균형 상태에 도달하는 것이다. 마코프 체인의 장기 확률 분포의 존재는 다음과 같은 조건이 만족되었을 경우이다.

$$\sum_{j=0}^{\infty} \beta_j = 1$$

Irreducible aperiodic 마코프 체인의 경우, 모든 상태가 transient 하거나 null recurrent 하고, 모든 i 와 j 에 대해서 $n \rightarrow \infty$ 하고 $P_{ij}^n \rightarrow 0$ 수렴하면 균형 확률 분포는 존재하지 않는다.

상기 예제로부터 얻을 수 있는 균형 상태 확률은 각각 $\beta_A = 0.286$, $\beta_B = 0.297$, $\beta_C = 0.418$ 이다. 균형 상태에 이르는 전이 과정은 <표 3>으로 요약된다.

모든 상태가 상기의 조건을 만족하면,

$$\beta_j = \lim_{n \rightarrow \infty} P_{ij}^n, \quad j \geq 0$$

단, β_j 는 균형 상태의 확률 분포.

<표 3>에서 균형 단계에 도달하면 군집 A에 86명의 고객이 남아 있고, 89명은 군집 B, 125명은 군집 C에 남아있게 된다. 이 표를 이용하면 현재의 마케팅 전략 (마케팅 캠페인 혹은 프로모션)의 효율을 평가할 수 있다. 예를 들어, 군집 C가 수익성 있는 우량 군집이고 군집 A는 이탈가능 군집이라고 가정해 보자.

β_j 확률은 유일하게 존재하고 최초 확률 분포에 대해서 독립적으로 존재한다. 이 경우 β_j 는 다음의 방정식 시스템으로부터 결정된다.

$$\beta_j = \sum_{i=0}^{\infty} \beta_i P_{ij}, \quad j \geq 0$$

군집 C에 고객들이 높은 확률로 남아 있고 군집 A가 낮은 확률로 없어져간다면 (즉, $\sum_j P_{ic} > \sum_j P_{ij}$ 부등호가 성립한다면) 현재의 마케팅 전략이 효과적이라고 주장할 수 있다. 그것은 우량 고객이 숫자적으로 증가하고 이탈가능 고객의 수

<표 3> 균형 상태에 도달하는 전이 과정

군집 \ 단계	고객 수					
	0	1	2	3	4	5
A	100	90	87	86	86	86
B	100	80	90	88	89	89
C	100	130	123	126	125	125

는 감소하기 때문이다.

군집 A, B, C 의 예상 수익률이 각각 15%, 25%, 40%이라면, 초기에 소매상이 올릴 수 있는 평균 수익률은 다음의 식으로 계산된다.

$$\frac{100}{300} \times 15\% + \frac{100}{300} \times 25\% + \frac{100}{300} \times 40\% = 26.67\%$$

한편, 균형 상태의 평균 수익률은 아래와 같이 계산된다.

$$\frac{86}{300} \times 15\% + \frac{89}{300} \times 25\% + \frac{125}{300} \times 40\% = 28.38\%$$

초기와 비교했을 때 장기 균형 상태에서 평균 수익률이 약간 증가하였음을 알 수 있다. 이것은 전이 확률이 변하지 않은 동안 현재의 마케팅 전략이 효과적이므로 소매상의 입장에서는 현재의 전략을 유지하는 것이 바람직하다는 것을 의미한다.

■ 대체 마케팅 전략 평가

소매상이 여러 대안이 될 수 있는 마케팅 전략을 고려하고 있을 때 마코프 분석은 대안을 평가할 수 있는 중요한 역할을 할 수 있다. 마케팅 전략을 결정하는데 있어 마코프 분석의 도움을 설명하기 위해 다음과 같은 전이 확률표와 소매상이 두 가지 새로운 마케팅 전략을 고려하고 있다고 생각해 보자.

	A	B	C
A	0.5	0.1	0.4
B	0.2	0.2	0.6
C	0.2	0.5	0.3

- 대체 전략 1: 대체 전략 1을 실행하면 군집 B 의 고객 유지율이 20%에서 60%로 증가하고

다른 군집으로의 고객 유출이 감소된다.

	A	B	C
A	0.4	0.4	0.2
B	0.2	0.6	0.2
C	0.1	0.6	0.3

- 대체 전략 2: 대체 전략 2는 군집 C로 고객을 유인하도록 설계되어 있다. 이 경우 전이 확률표는 아래와 같이 바뀐다.

	A	B	C
A	0.3	0.1	0.6
B	0.2	0.2	0.6
C	0.2	0.3	0.5

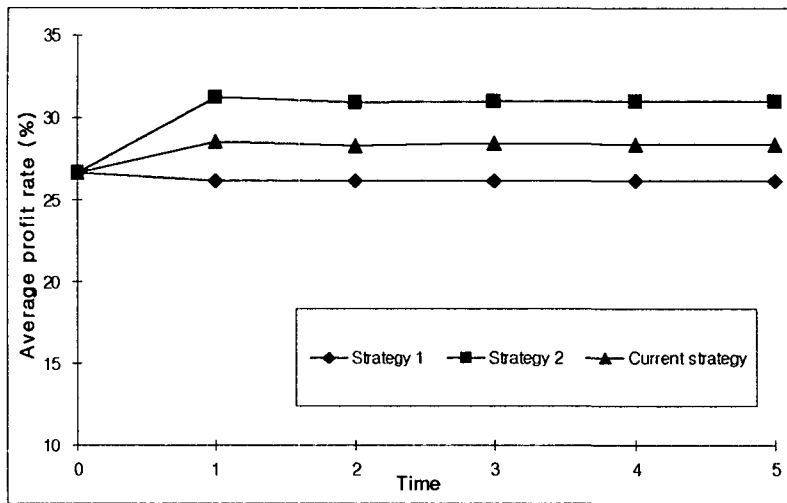
예제에서 볼 수 있듯이 현재의 전략을 포함하여 3개의 마케팅 프로그램을 실행하는 비용이 같다면 대체 전략 2가 최고의 전략이 된다. 전략 2는 장기에서뿐만 아니라 단기에서도 소매상에게 높은 수익 증대를 가져다 준다(<그림 3> 참조).

마케팅 전략 집행 비용이 상이하다면 2가지 관점을 추가적으로 고려해야 한다. 하나는 장기에 있어서 어떤 전략이 투자비에 대해 효율적인가라는 관점과 또 다른 하나는 각 전략별 손익분기점에 대한 관점이다.

4. 동적 고객관계관리의 적용

4.1 고객 군집

분석 대상이 된 소매상은 대한민국에서 오프라인과 온라인으로 면세점을 운영하고 있고 매출 규모 면에서 수위에 드는 업체 중의 하나를 선정해 회원권을 보유하고 있는 고객의 자료를 수집



<그림 3> 마케팅 전략별 평균 수익률 비교

하였으며, 1995년 7월 1일부터 1996년 12월 31일 까지 2,036명의 고객을 대상으로 분석을 하였다. 고객을 군집화하기에 앞서 고객의 RFM 자료를 추출하였는데 시간의 경과에 따른 동적인 분석이 가능하도록 다음 <표 4>와 같은 추출 방법을 이용하였다.

t 기간의 RFM 값은 다음과 같은 방식으로 계산된다. 먼저, t 기간 동안 신규 고객일 경우에는 이전에 구매한 기록이 없으므로 해당 고객의 RFM 값은 처음으로 계산된다.

- 기간 t 의 Recency 값 (R_t) - t 기간 동안 고

객이 최종 구매한 이래로 경과된 시간

- 기간 t 의 Frequency 값 (F_t) - t 기간 동안 고객의 구매 회수
- 기간 t 의 Monetary 값 (M_t) - t 기간 동안 고객의 구매 금액

고객이 기존에 구매 내력이 있을 경우에는 RFM 계산은 조금 달라질 수 있다. 이 경우에는 t 기간 동안 구매를 하였는지, 구매를 하지 않았는지에 따라 계산 방법이 달라진다. 먼저 t 기간 동안 구매를 하였다면 RFM은 다음의 방식으로 계산한다.

<표 4> 특정 고객 별 기간 t 의 RFM 값 계산 방법

	기존 고객			신규고객
	$t-1$ 기간	t 기간		t 기간
		구매함	구매하지 않음	
Recency값 (R)	R_{t-1}	R_t	$R_{t-1} + t$ 기간	R_t
Frequency 값 (F)	F_{t-1}	$F_{t-1} + F_t$	F_{t-1}	F_t
Monetary 값 (M)	M_{t-1}	$M_{t-1} + M_t$	M_{t-1}	M_t

- 기간 t 의 Recency 값 = R_t
- 기간 t 의 Frequency 값 = $t-1$ 기간 동안의 Frequency + t 기간 동안의 Frequency = $F_{t-1} + F_t$
- 기간 t 의 Monetary 값 = $t-1$ 기간 동안의 Monetary + t 기간 동안의 Monetary = $M_{t-1} + M_t$

여기서 주목할 것은 고객이 t 기간 동안 구매를 하였기 때문에 R_t 값이 기간 t 보다는 작다는 것이다. t 기간의 Frequency와 Monetary는 이전 기간의 ($t-1$) Frequency와 Monetary를 누적하여 계산한다.

반면에 기존 고객이 t 기간 동안 구매를 하지 않았을 경우에는 다음과 같은 방식으로 RFM 값을 계산한다.

- 기간 t 의 Recency 값 = $R_{t-1} + t$ 기간
- t 의 Frequency 값 = 기간 $t-1$ 의 Frequency 값 = F_{t-1}
- 기간 t 의 Monetary 값 = 기간 $t-1$ 의 Monetary 값 = M_{t-1}

여기서 주목할 것은 t 기간의 Recency 값은 고객이 t 기간 동안 구매를 하지 않았기 때문에 $t-1$ 기간의 값에 비해 기간 t 만큼 증가한다는 것이다. 기간 t 의 Frequency와 Monetary 값은 이전 기간인 ($t-1$)의 Frequency와 Monetary 값과 동일하다.

추출된 RFM 값은 3*3 SOM에 입력되는데 SOM을 이용한 고객 세분화는 학습 단계와 사용 단계의 두 단계를 거치게 된다. 학습 단계는 RFM 자료를 이용하여 고객 구매 행위 모델을 구축하는 단계로 종종 시간이 많이 소요되고 데이터마이닝 분석가의 도움이 요구되는 단계이다. 모델이 구축되면 사용 단계로 들어가는데 구축된 모델을 빠르고 쉽게 고객군집화에 적용하는 단계이다.

<표 5>는 기간 t 에 3 * 3 SOM을 이용하여 추출된 고객 군집의 특성과 상응하는 마케팅 전략을 보여주고 있다. 각 군집의 평균 RFM 값을 모든 고객의 평균 RFM 값과 비교한 후 군집의 평균값이 전체 평균값보다 크면 "h", 작으면 "l" 을 붙여 군집의 특성을 나타내었다.

<표 5> 고객 군집과 상응하는 마케팅 전략

고객 군집	특 성			마케팅 전략
	Recency	Frequency	Monetary	
RhFIMI	크다	작다	작다	고객 재활성
RhFIMh	크다	작다	크다	
RhFhMI	크다	크다	작다	
RhFhMh	크다	크다	크다	
RIFIMI	작다	작다	작다	고객 성숙
RIFIMh	작다	작다	크다	
RIFhMI	작다	크다	작다	
RIFhMh	작다	크다	크다	

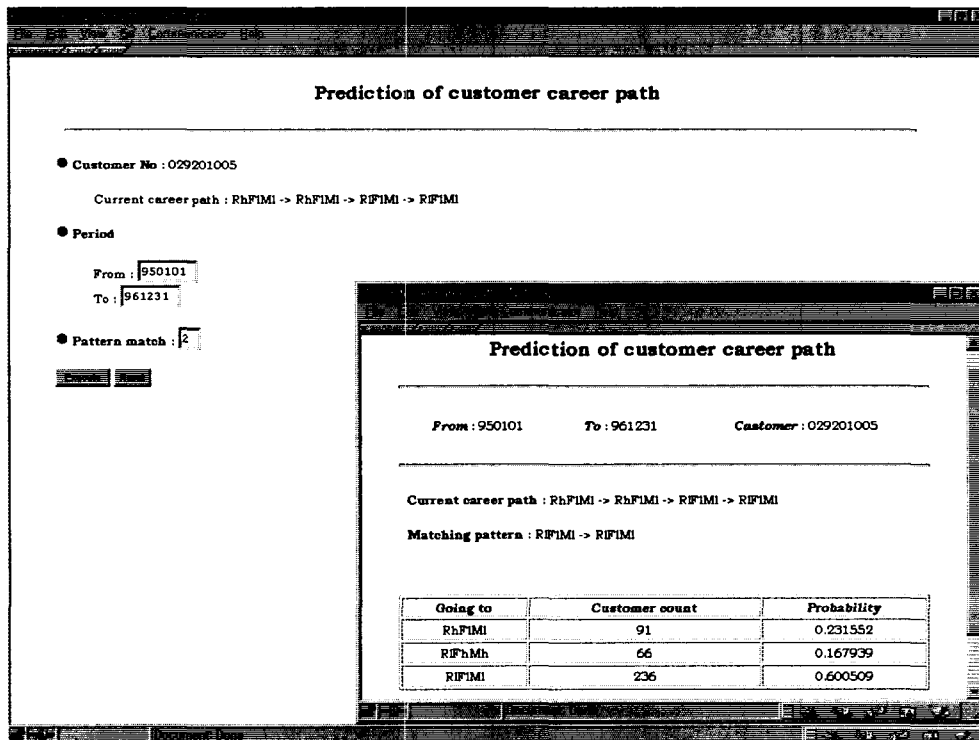
RhFIMI, RhFIMh, RhFhMI, RhFhMh 군집에 속하는 고객은 Recency 값이 전체 평균값보다 크므로 이탈가능고객일 가능성이 높다. RIFIMI, RIFIMh, RIFhMI 군집은 신규 고객을 포함하고 있어 Recency, Frequency, Monetary 값에서 전체 평균 이하 값을 보이고 있다. RIFhMh 군집에 속하는 고객은 최근에 자주, 많이 구매한 우량 고객으로 간주될 수 있다.

4.1.1 개별 고객의 이력경로

개별 고객의 과거 군집간 이동 이력과 유사한 이력을 가진 고객의 군집간 이동 이력을 협업 필터링하면 해당 고객의 향후 군집간 이동을 예측하는 것이 가능하다.

<그림 4>는 특정 고객의 이력 경로를 예측하는 모습을 보여주고 있다. 회원 번호 “029201005” 고객은 RhFIMI => RhFIMI => RIFIMI => RIFIMI 형태의 과거 이력경로를 가지고 있는데 다음 기간에 해당 고객이 보일 군집간 이동의 모습을 유사한 이력을 가진 고객이 보인 군집간 이동 패턴을 참조하여 예측할 수 있다.

<그림 4>에 의하면 RhFIMI => RhFIMI => RIFIMI => RIFIMI의 이력 경로를 보이는 고객은 다음 기간에 23.16%의 확률로 RhFIMI, 16.79%의 확률로 RIFhMI 군집으로 이동하거나, 60.05%의 확률로 RIFIMI 군집에 남는 것으로 예측되었다. 해당 고객이 RIFIMI 군집에 남아 있을 확률이 월등하게 크므로 이 고객과의 일대일 관계 마케팅



<그림 4> 협업 필터링에 의한 개별 고객의 이력 경로 예측

을 위해서 고객의 충성심을 고양하는 고객 성숙 전략을 구사하는 것이 좋겠다.

4.1.2 집단 이력경로

특정 기간 동안에 대부분의 고객이 보이는 집단 이력경로를 파악한 결과 중 <그림 5>은 우량 고객 군집인 RIFhMh 군집으로 이르는 집단 이력경로를 보여주고 있다.

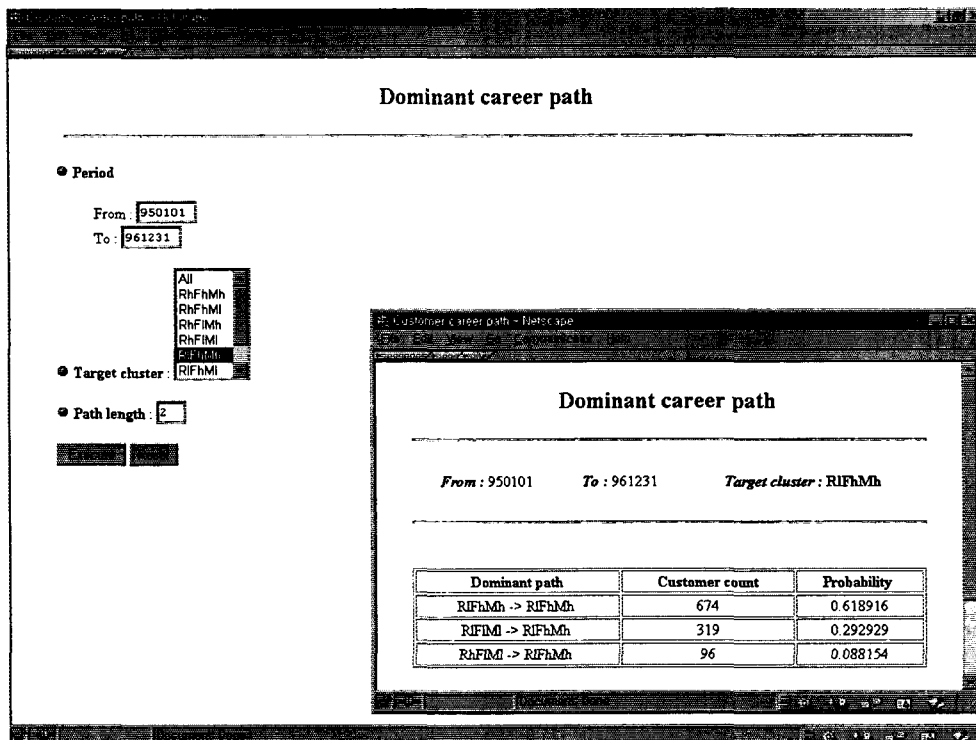
<그림 5>에 의하면 RIFhMh에 속할 것으로 예측된 고객 중, RIFhMh 군집에서 오는 (RIFhMh 군집에 남아있는) 고객의 수가 가장 많다. RIFhMh 군집은 우량 고객을 포함하는 군집이므로 우량 고객이 우량 고객 군집에 많이 남는 매우 고무적인 상황이라고 할 수 있다. 따라서,

마케팅 담당자는 RIFhMh 군집 소속 고객에 대해서는 “고객 유지 전략”을 구현하는 것이 타당하다.

RIFIMI 군집에서 RIFhMh 군집으로 이르는 확률이 그 다음으로 높는데 RIFIMI 군집에 속한 고객이 우량 고객이 될 가능성이 RIFhMh 군집을 제외하고 가장 높다는 것을 의미한다. 따라서 마케팅 담당자는 RIFIMI 군집 소속 고객에 대해서 “고객 성숙 전략”을 펴는 것이 적절하다.

<표 6>은 우량 고객 군집인 RIFhMh 에 이르는 집단 이력경로를 보여주는데 이번에는 경로 길이를 3으로 늘렸을 때의 경로를 보여준다.

<그림 5>에 의하면 RIFIMI 군집에서 바로 다음 기간 동안에 RIFhMh 군집으로 이동할 확률



<그림 5> RIFhMh 군집으로 이르는 집단 이력경로

<표 6> 경로 길이가 3일 때 RIFhMh 군집에 이르는 집단 이력 경로

집단 이력 경로	고객 수	확률 (%)
RIFhMh => RIFhMh => RIFhMh	331	48.32
RIFIMI => RIFhMh => RIFhMh	94	13.72
RIFIMI => RIFIMI => RIFhMh	86	12.55
RIFhMh => RIFIMI => RIFhMh	71	10.37
RhFIMI => RIFhMh => RIFhMh	41	5.99
RhFIMI => RhFIMI => RIFhMh	28	4.09
RhFIMI => RIFIMI => RIFhMh	24	3.50
RIFIMI => RhFIMI => RIFhMh	10	1.46

은 29.3 % 이었다. <표 6>의 결과와 비교해서 보면 바로 다음 기간 동안에 이동하지 않으면 RIFIMI 군집 고객이 RIFhMh 군집으로 이동할 확률은 27.73%(= 13.72 + 12.55 + 1.46)으로 떨어 진다는 것을 알 수 있다. 이것은 적절한 시기에 적절한 고객을 대상으로 적절한 마케팅 전략을 펼쳤다면 더 많은 고객을 우량 고객으로 확보할 수 있었음을 암시하고 있다.

이번에는 이탈 가능 군집으로 분류된 RhFIMI 군집에 대해서 유사한 분석을 해보았다. <표 7> 은 경로 길이가 3일 때 고객 이탈로 이어지는 집단 이력 경로를 보여주고 있다.

<표 7>의 첫번째와 네번째 행을 보면 RhFIMI 에 속했던 고객은 좀처럼 RhFIMI 군집을 벗어나 지 않는다. 따라서 RhFIMI 군집에 속하기 전에

적절한 마케팅 전략을 구사해야 함을 의미한다. 두번째와 세번째 행은 신규 고객을 포함하는 RIFIMI 군집에서 이탈 가능 군집으로 편입되는 고객이 많고 신규 고객에 대한 적절한 성숙 전략 을 구사하지 않는다면 첫 구매 이후 이탈 가능 고객이 속출할 수 있고 한번 이탈한 고객은 다시 돌아올 가능성이 낮음을 지적하고 있다.

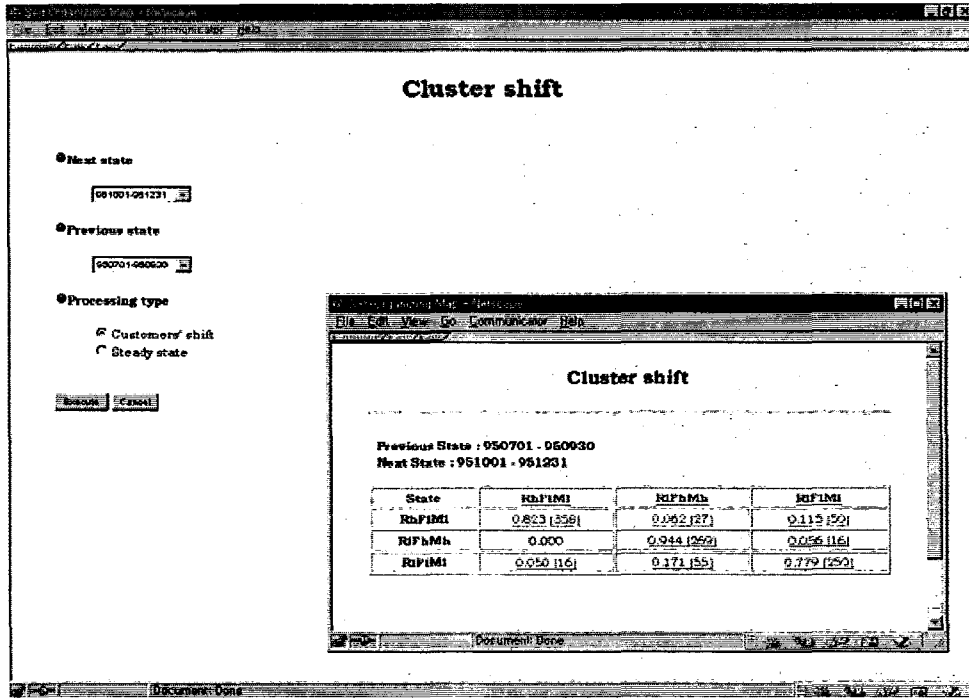
4.2 고객 군집의 변화 예측

<그림 6>은 1995년 7월 1일부터 1995년 12월 31일까지 기간 동안 고객들의 군집간 이동 확률 을 보여주고 있다.

<그림 6>에서 특이한 것은 해당 분석 기간 동안에는 3개의 고객 군집만 나타난다는 것이다.

<표 7> 경로 길이가 3일 때, RhFIMI 군집으로 이르는 집단 이력 경로

집단 이력 경로	고객 수	확률 (%)
RhFIMI=> RhFIMI => RhFIMI	476	65.20
RIFIMI=> RhFIMI => RhFIMI	145	19.86
RIFIMI=> RIFIMI => RhFIMI	91	12.47
RhFIMI=> RIFIMI => RhFIMI	18	2.47



<그림 6> 고객의 군집간 이동 확률

대체적으로, RIFIMI 군집은 신규 고객을 포함하고, RhFIMI 군집은 이탈 가능 고객 군집을, RIFhMh 군집은 우량 고객 군집을 나타낸다.

4.2.1 고객의 군집간 전이 확률

아래에 표시된 전이 확률 표에는 각 군집간 고객 유지, 유입, 유출 확률을 보여준다. 각 행은 고객의 유지와 유출을, 각 열은 유지와 유입을 나타낸다.

	RhFIMI	RIFhMh	RIFIMI
RhFIMI	0.823	0.062	0.115
RIFhMh	0	0.944	0.056
RIFIMI	0.050	0.171	0.779

표에 의하면 RIFhMh 군집은 소속 고객의 94.4% (269명)을 유지하고 RIFIMI 군집에만 5.6% (16명)의 고객이 유출되는 반면, RhFIMI 군집으로부터 6.2% (27명), RIFIMI 군집으로부터 17.1% (55명)의 고객이 유입되어 고객 유입이 유출보다 많고 시간이 흐름에 따라 소매상의 거의 대부분 고객을 흡수할 것으로 보인다.

4.2.2 장기적 경영 관점

소매상의 고객들이 안정적인 구매 패턴을 보인다면 전이 확률표는 변함이 없을 것이고, 마코프 체인을 이용하여 장기에 걸친 예측을 하는 것이 가능하다.

장기 경영의 관점에서 마케팅 담당자는 마코프 체인이 균형상태에 수렴하는지 아니면 수렴하

<표 8> 각 군집별 고객 수의 변화와 균형 상태

군집 \ 단계	0	1	2	3	4	5	...	44	45
RhFIMI	435	374	281	193	129	91	...	63	63
RIFhMh	285	351	458	570	658	712	...	754	754
RIFIMI	321	316	302	278	254	238	...	224	224

지 않는지에 관심이 있다. <표 8>은 최종 균형 상태에서의 안정 확률에 이르는 과정을 정리하고 있다.

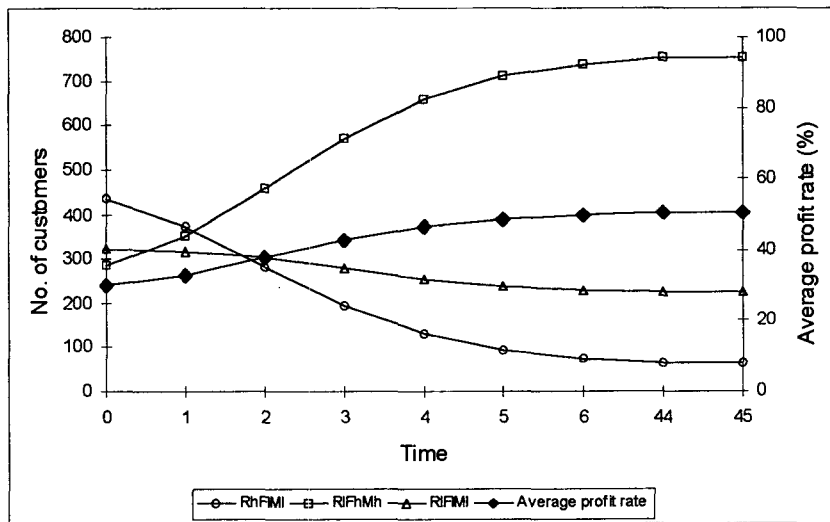
<표 8>에 따르면 44번의 전이 후에 세 개 군집은 궁극적으로 균형 상태에 도달하게 된다. 균형상태 확률은 각각 $\beta_{RhFIMI} = 0.061$, $\beta_{RIFhMh} = 0.724$, $\beta_{RIFIMI} = 0.215$ 이다.

세 개의 군집이 현재 소매상에 가져오는 수익률이 각각 RhFIMI = 10%, RIFhMh = 60%, RIFIMI = 30% 이므로 <그림 7>과 같은 시간의 경과에 따른 수익률의 변화를 가져오고 평균 수익률은 장기적으로 상승한다는 것을 알 수 있다.

4.2.3 단기적 경영 관점

마케팅 담당자의 입장에서는 장기에서의 균형 상태 못지 않게 단기적인 수익 향상에 더 많은 관심을 가질 수 있다.

경영 활동의 영향으로 실제 고객 군집의 분포가 전이 확률과 다르게 전개될 수 있는 상황이 발생할 수 있는데 단기적으로 마코프 체인을 활용할 수 있다. <그림 7>에서는 고객의 군집은 각 단계별로 상이한 변화의 모습을 보인다. RIFhMh 군집의 고객의 수는 지속적으로 증가하는 반면 RIFIMI, RhFIMI 군집의 고객의 수는 지속적으로 감소하는 방향으로 움직이고 있다. 마케팅 담당



<그림 7> 각 군집별 장기 수익률

자는 마코프 체인이 균형상태에 도달하는 장기에 있어 최고의 수익률을 얻기 위해 각 군집의 변화를 <그림 7>과 같은 모습이 유지될 수 있도록 단기적으로 관리하여야 한다.

예를 들어 마케팅 담당자가 단기간(첫 번째 전이 후)에 더 많은 수익을 올리길 원한다면 어떤 조치를 취해야 하는가? 마케팅 담당자가 취할 수 있는 행동은 고객의 군집간 이동을 관찰하고 통제하는 것이다. 그 방법으로는 ① 우량 고객 군집인 RIFhMh에 더 많은 고객이 유입되도록 하는 방법과 ② RIFIMI, RhFIMI 군집에서 고객을 더 많이 유출시키는 방법, ③ ①과 ②를 복합적으로 사용하는 방법이 있을 수 있다.

5. 결론 및 토의

고객의 구매 행동에 기초를 둔 고객관계관리의 개념은 오래 전에 나왔지만 동적인 고객관계관리를 구현하고 적용한 시스템은 발견하기 어려운 실정이다. 전통적인 고객관계관리는 주로 특정 기간에 적용된 고객관계관리라고 볼 수 있다. 일반적으로 이런 정적인 고객관계관리와 고객지식은 주어진 기간 동안 소매상의 이익을 증가시키기 위해 마케팅 자원을 집중하는 것에 도움을 준 바 있다. 하지만, 시간이 지나면 정적인 고객지식은 쓸모가 없어지므로 전자상거래에서 고객관계관리는 시간에 대해 동적으로 이루어져야 하며 본 논문의 목적과 같이 동적인 고객관계관리가 효과적으로 온라인 소매상이 직면한 경영 문제들을 해결할 수 있는 것을 보여주고자 한다.

그러나 동적 고객관계관리가 소매상에게 상당한 가치를 가져다 주는 것은 분명하지만 시간의 흐름에 따른 고객 구매 행위를 분석하고 고객

지식을 획득하는 실용적이고 비용 면에서 우수한 방법이 필요한데 본 논문에서는 데이터마이닝 기법과 마코프 체인을 활용하는 방법을 제안하였다.

데이터마이닝 기법을 활용하는 방법에는 ① 소매상의 고객을 RFM 분석의 관점에서 세분화 하는데 SOM을 이용하였고, ② 특정 고객을 우량 고객 혹은 이탈 가능 고객으로 분류하는데 있어 의사결정나무 분류기를 이용하였다.

개인의 이력경로와 군집의 이동 경로를 관찰, 예측함으로써 각 고객과 고객 군집에 대해 적절한 마케팅 프로모션을 시행할 수 있게 하였다. 마코프 체인을 이용하여 시간의 경과에 따른 고객 군집의 변화를 예측하는 확률 모형을 구하고 마케팅 담당자의 의사결정을 지원할 수 있는 여러 문제를 해결하였다. 또한 단기의 전이 확률과 장기의 균형 확률을 소개하였고 관련된 경영 전략의 수립과 평가 방법을 도출할 수 있었다. 마코프 체인을 이용하여 단기에 더 큰 이익을 가져오는 방법을 제안하였고 장기적인 마케팅 전략을 수립하는데 이용하였다. 마코프 분석의 전제 조건인 전이 확률의 불변성이 어떤 면에서는 비현실적인 가정이지만, 고객의 구매 행위의 변화에 대한 전체적인 실마리를 제공하고 마케팅 전략의 효과를 평가하는 방법을 제공할 수 있었다.

분석의 기초가 된 RFM 값들이 고객 세분화의 중요한 변수로서 역할을 하고 있지만 더 나은 예측력을 가진 변수들을 개발할 필요성은 여전히 남아 있다. 그런 변수들을 개발할 수 있다면 이런 종류의 분석을 기반으로 하는 앞으로의 연구에 커다란 진전을 가져올 것으로 보인다.

참고문헌

- Abell, D. F., *Defining the Business : The Starting Point of Strategic Planning*, Prentice-Hall, Englewood Cliffs, New Jersey, 1980.
- Aggarwal, C. C., J. L. Wolf, K-L. Wu and P. S. Yu, "Horting Hatches an Egg : A New Graph-Theoretic Approach to Collaborative Filtering," In : S. Chaudhuri and D. Madigan, *Proceedings of the 1999 International Conference on Knowledge Discovery and Data Mining*, ACM, New York, (1999), 201~212.
- Balakrishnan, P. V., M. C. Cooper, V. S. Jacob and P. A. Lewis, "Comparative Performance of the FSCL Neural Net and K-means Algorithm for Market Segmentation," *European Journal of Operational Research*, 93(1996), 346-357.
- Beane, T. P. and D. M. Ennis, "Market Segmentation : A Review," *European Journal of Marketing*, 21(1987), 20-42.
- Berry, M. J. A. and G. Linoff, *Mastering Data Mining : The Art and Science of Customer Relationship Management*, John Wiley & Sons Inc, New York, 2000.
- Blattberg, R. C. and S. K. Sen, "Market Segments and Stochastic Brand Choice Model," *Journal of Marketing Research*, 13(1976), 34-45.
- Blattberg, R. C., T. Buesing, P. Peacock and S. Sen, "Identifying the Deal-Prone Segment," *Journal of Marketing Research*, 15(1978), 369-377.
- Bonoma, T. V. and B. P. Shapiro, *Segmentation the Industrial Market*, Lexington Books, Lexington, Mass, 1983.
- Borchers, A., J. Herlocker, J. Konstan and J. Riedl, "Ganging up on Information Overload," *Computer*, Vol.31, No.4(1998), 106-108.
- Bult, J. R. and T. Wansbeek, "Optimal Selection for Direct Mail," *Marketing Science*, Vol.14, No.4(1995), 378-394.
- Chakravarti, D., P. E. Hendrix and W. L. Wilkie, *Market Segmentation Research*, Vol.1 and 2, Electric Power Research Institute, Palo Alto, California, 1987.
- Choffray, J. M. and G. L. Lilien, "A New Approach to Industrial Market Segmentation," *Sloan Management Review*, Spring(1978), 17-29.
- Dibb, S. and P. Stern, "Questioning the Reliability of Market Segmentation Techniques," *Omega, International Journal of Management Science*, Vol.23, No.6(1995), 625-636.
- Doyle, P. and J. Saunders, "Market Segmentation and Positioning in Specialized Industrial Market," *Journal of Marketing*, 49, No.2(1985), 24-32.
- Fayyad, U. M., G. Piatetsky-Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," In : Fayyad, U. M., G. Piatetsky-Shapiro, P. Smyth and R. Uthurusamy(Ed.), *Advances in Knowledge Discovery and Data Mining*, AAAI Press/The MIT Press, Menlo Park, (1996), 1-34.
- Fish, K. E., J. H. Barnes and M. W. Aiken, "Artificial Neural Networks - A New Methodology for Industrial Market Segmentation," *Industrial Marketing Management*, 24(1995), 431-438.
- Frank, R. E., W. F. Massy and Y. Wind, *Market Segmentation*, Prentice-Hall, Englewood Cliffs, New Jersey, 1972.
- Green, P. E., "A New Approach to Market Segmentation," *Business Horizons*, Vol.20(1977), 61-73.
- Griffith, R. L. and L. G. Pol, "Segmenting Industrial Markets," *Industrial Marketing Management*, 23(1994), 39-46.
- Ha, S. H. and S. C. Park, "Application of Data

- Mining Tools to Hotel Data Mart on the Intranet for Database Marketing," *Expert Systems With Applications*, Vol.15, No.4 (1998), 1-31.
- Hillier, F. S. and G. J. Lieberman, *Introduction to Operations Research*, 6th Ed., McGraw-Hill, Singapore, 1995.
- Hughes, A. M., *The Complete Database Marketer*, 2nd Ed., McGraw-Hill, New York, 1994.
- Hughes, A. M., *Strategic Database Marketing*, 2nd Ed., McGraw-Hill, New York, 2000.
- Johnson, H. and A. Flodhammer, "Some Factors in Industrial Market Segmentation," *Industrial Marketing Management*, 9(1980), 201-205.
- Kalwani, M. U. and D. G. Morrison, "A Parsimonious Description of the Hendry System," *Management Science*, 23(1997), 467-477.
- Kiang, M. Y., R. T. Chi and K. Y. Tam, "A Decision Support System for Market Segmentation - A Neural Networks Approach," *Proceedings of Americas Conference on Information System*(1995).
- Kohonen, T., "Self-Organized Formation of Topologically Correct Feature Maps," In : Shavlik, J. W. and T. G. Dietterich, *Readings in Machine Learning*, Morgan Kaufmann Publishers Inc, California, (1990), 326-336.
- Konstan, J., B. Miller, D. Maltz, J. Herlocker, L. R. Gordon and J. Riedl, "GroupLens : Applying Collaborative Filtering to Usenet News," *Communications of the ACM*, Vol.40, No.3(1997), 77-87.
- Kotler, P., *Marketing Management : Analysis, Planning, Implementation and Control*, 7th Ed., Prentice-Hall, Englewood Cliffs, NJ, 1991.
- Ladany, S. P. and A. Arbel, "Optimal Cruise-Liner Passenger Cabin Pricing Policy," *European Journal of Operational Research*, 55(1991), 136-147.
- Mahajan, V. and A. K. Jain, "An Approach to Normative Segmentation," *Journal of Marketing Research*, 15(1978), 338-345.
- Mangiameli, P., S. K. Chen and D. West, "A Comparison of SOM Neural Network and Hierarchical Clustering Methods," *European Journal of Operational Research*, Vol.93, Iss.2(1996), 402-417.
- Morwitz, V. G. and D. C. Schmittlein, "Testing New Direct Marketing Offerings : The Interplay of Management Judgment and Statistical Models," *Management Science*, Vol.44, No.5, May(1998), 610-628.
- Peppard, J., "Customer Relationship Management (CRM) in Financial Services," *European Management Journal*, Vol.18, No.3(2000), 312-327.
- Peppers, D. and M. Rogers, *Enterprise One to One : Tools for Competing in the Interactive Age*, Doubleday, New York, 1999.
- Peppers, D. and M. Rogers, *The One to One Future : Building Relationships One Customer at a Time*, Bantam Doubleday Dell Publishing, New York, 1997.
- Peppers, D., M. Rogers and R. Dorf, "Is Your Company Ready for One-to-One Marketing," *Harvard Business Review*, Jan/Feb(1999), 151-160.
- Quinlan, J. R., *C4.5 : Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, 1993.
- Resnick, P. and H. R. Varian, "Recommender Systems," *Communications of the ACM*, Vol.40, No.3(1997), 56-58.
- Roberts, M. L. and P. D. Berger, *Direct Marketing Management*, Prentice Hall, Englewood Cliffs, NJ, 1989.
- Rossi, P. E., R. E. McCulloch and G. M. Allenby, "The Value of Purchase History Data in

- Target Marketing," *Marketing Science*, Vol.15, No.4(1996), 321-340.
- Russel, G. J. and W. A. Kamakura, "Understanding brand Competition Using Micro and Macro Scanner Data," *Journal of Marketing Research*, Vol.31, May(1994), 289-303.
- Schafer, J. B., J. A. Konstan and J. Riedl, "E-commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Vol.5, No.1-2(2001), 115-153.
- Shepard, D., *The New Direct Marketing : How to Implement a Profit-Driven Database Marketing Strategy*, 2nd Ed., Irwin, Homewood, IL, 1995.
- Stone, B., R. Jacobs and H. R. Wientzen, *Successful Direct Marketing Methods*, 7th Ed., McGraw-Hill, New York, 2001.
- Weinstein, A., *Market Segmentation*, Revised Ed., Probus Publishing Company, Chicago, IL, USA, 1994.
- West, P. M., P. L. Brockett and L. L. Golden, "A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice," *Marketing Science*, Vol.16, No.4(1997), 370-391.
- Wyner, G. A., "Segmentation Design," *Marketing Research*, December(1992), 38-41.

Abstract

Customer Relationship Management Techniques Based on Dynamic Customer Analysis Utilizing Data Mining

Sung Ho Ha* · Jae Shin Yi*

Traditional studies for customer relationship management (CRM) generally focus on static CRM in a specific time frame. The static CRM and customer behavior knowledge derived could help marketers to redirect marketing resources for profit gain at that given point in time. However, as time goes, the static knowledge becomes obsolete. Therefore, application of CRM to an online retailer should be done dynamically in time. Customer-based analysis should observe the past purchase behavior of customers to understand their current and likely future purchase patterns in consumer markets, and to divide a market into distinct subsets of customers, any of which may conceivably be selected as a market target to be reached with a distinct marketing mix. Though the concept of buying-behavior-based CRM was advanced several decades ago, virtually little application of the dynamic CRM has been reported to date.

In this paper, we propose a dynamic CRM model utilizing data mining and a Monitoring Agent System (MAS) to extract longitudinal knowledge from the customer data and to analyze customer behavior patterns over time for the Internet retailer. The proposed model includes an extensive analysis about a customer career path that observes behaviors of segment shifts of each customer: prediction of customer careers, identification of dominant career paths that most customers show and their managerial implications, and about the evolution of customer segments over time. Furthermore, we show that dynamic CRM could be useful for solving several managerial problems which any retailers may face.

Key words : Electronic Commerce, Business-to-Consumer, Customer Relationship Management, Data Mining, Marketing Strategy, Agent Model.

* School of Business Administration, Kyungpook National University