

데이터마이닝을 이용한 세분화된 고객집단의 프로모션 고객반응 예측

Predicting the Response of Segmented Customers for the Promotion Using Data Mining

홍 태 호 (Taeho Hong)

부산대학교 경영학부 부교수

김 은 미 (Eunmi Kim)

부산대학교 경영학과 박사과정, 교신저자

요 약

정보기술의 발전과 더불어 기업과 고객간의 대부분의 정보가 축적되면서 기업은 거래고객의 자세한 정보를 활용하여 차별화된 마케팅을 제공할 수 있다. 본 연구는 기업이 제공하는 마케팅 전략을 보다 효과적으로 실행하기 위해 고객을 세분화하고, 세분화된 고객집단별 마케팅 프로모션에 대한 반응을 예측하는 모형을 제시하였다. 고객세분화에는 데이터마이닝 기법 중 SOM(Self-organizing Map)을 적용하였으며, 세분화된 집단별 프로모션 반응예측에는 로짓모형, 신경망 등의 단일모형과 k-최근접이웃법을 이용한 단일모형들의 통합모형을 적용하였다. 제시된 방법론으로 기업은 프로모션에 대한 고객반응을 예측할 뿐만 아니라 프로모션에 대한 반응을 쉽게 예측할 수 있는 고객집단과 반응예측이 어려운 고객집단으로 구분하여 프로모션의 효과를 극대화하고 각 집단에 맞는 프로모션 전략을 수립할 수 있다.

키워드 : 캠페인 관리, 고객반응, 고객세분화, 데이터마이닝

I. 서 론

기업은 새로운 고객을 획득하는 것만큼 기존 고객과의 지속적인 관계유지가 중요하다는 것을 인식하고 있으며 기존고객과의 관계를 강화하기 위해 복잡하고 다양한 고객의 특성을 미리 파악하여 고객의 요구사항을 충족시키고자 한다. 정보기술의 발달로 고객과의 거래가 전자적으로 이루어지는 경우가 많아져 고객 데이터가 풍부해졌으며 고객 데이터베이스에 대한 접근이 용이해지고 활용도 쉬워졌다. 따라서 기업은 고객

의 구매 데이터뿐 아니라 인구통계학적 정보나 심리학적 정보를 포함한 다양한 데이터를 통해 고객의 특성을 파악할 수 있다. 기업들은 어떤 고객에게 어떤 마케팅 전략을 제시하는 것이 효과적인지 결정하여 차별화되고 맞춤형된 마케팅 전략을 통해 고객을 유지하기 위해 힘쓰고 있다. 고객은 기업이 제공하는 동일한 마케팅 전략에도 다르게 반응하기 때문에 고객의 특성에 맞는 적절한 고객관계관리를 위한 전략이 필요하며 고객의 특성을 반영하지 않은 고객관계관리 전략은 좋은 성과를 기대하기 어렵다. 효과적인 고

객관계관리를 위해서는 고객의 특성과 성향을 반영한 고객관리 전략이 수립되어야 한다.

기업은 상이한 성향을 가진 고객을 동일한 집단으로 분류하는 세분화 전략을 통해 고객의 특성을 반영하고자 한다(Rigby *et al.*, 2002). 일반적으로 고객세분화는 최신성, 구매빈도, 구매액을 이용한 RFM(Recency, Frequency, Monetary)분석을 통해 얼마나 최근에, 얼마나 자주, 그리고 얼마나 많은 구매를 했는가에 대한 정보를 기반으로 이루어진다(Cullinan, 1977). 모든 고객을 대상으로 동일한 마케팅 전략을 수립한다는 것은 더 이상의 의미가 없으며 세분화된 고객집단별 특성을 기반으로 한 차별화된 마케팅 전략이 더욱 강조되어야 한다. 마케팅 전략을 제공할 목표고객의 선정이 잘 이루어지면 기업은 수익을 증가시킬 수 있지만 목표고객을 잘 선정하지 못하면 기업은 마케팅 비용의 증가뿐 아니라 고객과의 관계도 악화될 수 있기 때문이다(Cönül *et al.*, 2000).

기업은 마케팅 전략에 반응하는 고객을 구별하기 위해 상당한 노력을 기울이고 있으며 데이터마이닝 기법을 사용하여 방대한 고객데이터 분석을 통해 기업의 마케팅 전략에 대한 고객반응 예측모형을 구축해 왔다(Kim and Street, 2004). 김진화 등(2008)은 고객의 구매 데이터를 사용하여 고객의 구매의도를 예측하기 위해 인공신경망, 로지스틱 회귀분석, 의사결정나무, 베이지안망, SVM(Support Vector Machines) 등의 다양한 데이터마이닝 기법을 적용하였으며, SVM이 가장 우수한 성과를 나타내었다. Yu and Cho(2006)는 반응모형 구축을 위한 입력변수의 선정을 통해 모형의 예측성과를 높이고자 하였다. Baesens *et al.*(2002)은 RFM 변수와 고객의 다른 프로파일링 변수를 사용하여 베이지안 신경망과 같은 데이터마이닝 기법을 적용하여 우수한 예측의 성과를 보여주었다. 예측모형은 일반적으로 예측성과가 가장 높은 모형을 선택하나, 적용된 기법에 따라 예측성과가 매년 조금씩 달라질 수 있기 때문에 데이터마이닝 기법 중 어느 한 기법이

가장 우수하다고 단정 짓기는 쉽지 않다. 이러한 문제점을 해결하고 예측모형의 예측성과를 높이기 위한 방안으로 두 가지 이상의 데이터마이닝 기법을 통합하는 연구가 진행되어 왔다. Lin and McClean(2001)은 부도기업을 예측하기 위해 일반적인 다변량 통계분석 기법과 인공지능기법을 결합하였으며 Carvalho and Freitas(2004)는 의사결정나무를 통해 데이터를 분류하고 유전자 알고리즘을 통해 최종 규칙을 찾아내는 통합모형을 제시하였다. Ahn and Kim(2008)은 온라인 쇼핑물 고객의 구매예측과 주가예측을 위해 로지스틱 회귀분석, 인공신경망, 사례기반추론 등의 단일모형과 유전자알고리즘과 사례기반추론을 통합한 통합모형을 비교하여 단일모형보다 통합모형에서 우수한 성과를 얻을 수 있음을 보여주었다. 홍태호, 박지영(2009)은 로짓, 인공신경망, SVM 등의 다양한 데이터마이닝 기법들을 사례기반추론으로 통합하여 단일모형보다 통합모형의 예측성과가 우수함을 보였으며 데이터마이닝 기법 선택에 대한 문제도 해결하였다.

지금까지의 고객반응에 대한 예측모형은 반응 고객을 가능한 많이 구별하는데 중점을 두고 개발되어 왔기 때문에 다양한 기법들을 적용시켜 모형의 예측성과를 높이고자 하였다(김진화 등, 2008; Kim and Street, 2004; Yu and Cho, 2006). 한편 고객반응예측 연구와는 별도로 고객세분화에 대한 연구가 진행되어 다양한 고객세분화 방법을 제시하였다(이민수, 2007; 장남식, 2005; Chen *et al.*, 2007; Cheng *et al.*, 2009). 그러나 고객반응 예측모형이 고객의 특성을 반영하지 않고 모든 고객을 대상으로 구축되어 고객세분화 집단별로는 제시되지 않았다. 고객은 기업의 마케팅 전략에 모두 동일한 반응을 보이지 않기 때문에 고객의 특성에 따라 예측모형도 다르게 제시되어야 한다. 고객세분화 집단별 예측모형은 구매고객의 예측이 쉬운 집단과 어려운 집단을 구분할 수 있어 기업에서도 마케팅 전략을 제시할 목표고객군을 보다 효과적으로 선정할 수 있다. 따라

서 본 연구에서는 고객세분화 모형에 구매예측모형을 적용하여 고객집단의 특성을 기반으로 마케팅 전략을 제시하였다. 먼저 SOM(Self-organizing Map)을 이용하여 고객을 세분화하였다. 구매한 브랜드의 수, 구매 거래의 수, 구매 거래액에 대해 고객을 세분화하고 세분화된 고객 집단별로 예측모형을 개발하였다. 본 논문에서 제시한 고객세분화 집단의 반응고객 예측모형은 각 고객집단별 특성에 맞는 마케팅 전략을 제시할 수 있도록 하여 실제 기업에서 마케팅 전략에 대한 효과를 극대화할 수 있다.

II. 마케팅 캠페인과 데이터마이닝

마케팅 캠페인은 고객의 구매를 유도하거나 일정기간 동안 반복 구매가 보다 많이 일어나도록 하기 위해 카탈로그, 팸플렛, 견본, 전단 등의 배포나 사은품 전달 등의 다양한 활동으로 이루어질 수 있다(고용식, 2005). 마케팅 캠페인에 대한 고객반응은 고객이 기업의 마케팅 활동에 노출되면서 일어나는 고객의 선호도, 기대, 허락, 태도 등이 있으며 고객반응이 높을수록 기업은 더 많은 이익을 기대할 수 있다. 기업은 어느 고객에게 어떤 마케팅 캠페인을 제시하는 것이 효과적인지를 판단하여 적절한 마케팅 캠페인을 제시해야 하며 모든 고객에게 마케팅 전략을 제시하는 것 보다 목표고객을 결정하여 구매가능성이 높은 고객에게 집중적으로 제시함으로써 기업의 비용절감은 물론 이윤을 극대화 시킬 수 있다(Baesens *et al.*, 2002; Cönül *et al.*, 2000; Suh *et al.*, 1999).

기업의 마케팅 캠페인에 대한 긍정적인 고객반응은 구매에도 긍정적인 영향을 미치기 때문에 고객의 반응을 미리 예측하여 반응할 고객에게 보다 적극적인 마케팅 캠페인을 제시하여 마케팅 캠페인의 효과를 높일 수 있다. 고객반응 예측모형은 제공되는 마케팅 캠페인에 대한 반응고객을 가능한 많이 구별하는 것에 초점을 두며 고

객정보, 구매정보, 상품정보 등을 기반으로 이루어진다(Kim and Street, 2004). 방대한 고객데이터를 기반으로 반응고객을 구별하기 위해 주로 데이터마이닝 기법이 사용되며, 김진화 등(2008)은 고객이 구매한 거래내역 데이터를 통해 성질이 유사한 품목들을 카테고리화 하여 특정 상품에 대한 구매의도를 예측하였다. 인공지능경망, 로지스틱 회귀분석, 의사결정나무, 베이지안 망, SVM 등의 다양한 데이터마이닝 기법을 적용하였으며, SVM에서 가장 우수한 성과를 나타내었다. Yu and Cho(2006)는 반응모형 구축을 위한 입력변수를 의사결정나무, 유전자 알고리즘 등의 다양한 기법을 통해 적용시켜 모형의 예측성과를 높이고자 하였다. 안현철 등(2005)은 인터넷 쇼핑물의 구매데이터를 사용하여 로지스틱 회귀분석, 인공지능경망, SVM 등을 적용하여 고객구매예측을 하였으며 SVM에서 우수한 예측성과를 보였다. Baesens *et al.*(2002)은 RFM 변수와 고객의 다른 프로파일링 변수를 사용하여 베이지안 신경망을 적용하여 통신판매회사의 고객에 대한 구매를 예측하였다. 고객반응의 예측은 반응여부에 대한 이진분류 의사결정의 문제에 해당되며, 데이터마이닝을 이용한 이진분류 예측모형에 관한 연구들은 데이터마이닝 기법의 적용과 관련하여 다음과 같이 두 개로 분류될 수 있다. 고객반응과 같은 이진분류를 예측하기 위해 다수의 데이터마이닝 기법을 적용하여 예측성과가 높은 모형을 제시하는 연구들과, 특정 기법이 가장 우수한 기법이라고 결정하기가 쉽지 않다는 점을 고려하여 높은 예측성과를 얻기 위해 서로 다른 기법들을 통합하려는 연구들로 구분된다. Lin and McClean(2001)은 부도기업을 예측하기 위해 일반적인 다변량 통계분석 기법과 인공지능기법을 결합한 통합모형을 제시하였으며 Carvalho and Freitas(2004)은 의사결정나무를 통해 데이터를 분류하고 유전자 알고리즘을 통해 최종 규칙을 찾아내는 통합모형을 제시하였다. Kim and Shin(2004)은 신용등급의 예측 정확도를 높이

기 위해 사례기반추론과 퍼지 로짓을 통합하는 모형을 제안했으며 Ahn and Kim(2008)은 온라인 쇼핑물 고객의 구매예측과 주가에측을 위해 로지스틱 회귀분석, 인공신경망, 사례기반추론 등의 단일모형을 개발하고 사례기반추론에 사용된 k-최근접이웃법의 최적 k를 선택하기 위해 유전자 알고리즘을 적용한 통합모형을 제시하였으며 단일모형보다 우수한 예측성과를 보여주었다. 홍태호, 박지영(2009)의 연구에서는 통신판매 데이터인 DMEF04(<http://www.directworks.org>) 데이터 셋을 대상으로 로짓, 인공신경망, SVM 등 다양한 데이터마이닝 기법들을 사례기반추론으로 통합한 통합모형을 제시하였으며 통합모형에서 단일모형보다 우수한 예측성과를 보여주었다.

기업은 고객 반응예측으로 마케팅 전략을 제시할 목표고객을 선정하기 위해 상이한 특성을 가진 고객을 같은 집단으로 분류하는 고객세분화를 많이 활용하고 있다(Rigby *et al.*, 2002; Suh *et al.*, 1999). 고객세분화는 이질적인 고객들을 보다 정확하게 이해할 수 있어 효율적이고 효과적으로 마케팅전략을 개발하고 수행할 수 있게 한다(장남식, 2005). 고객세분화를 위해 k-means와 SOM의 군집화 기법이 주로 사용되고 있으며 코호넨이 제안한 SOM은 자율학습기법의 인공신경망 방법론으로 분석결과를 시각적으로 제시해주며 기존의 선형판별분석이나 전통적인 다계층 인공신경망과 견줄만한 결과를 보여주고 있다(Serrano-Cinca, 1996). 또한 k-means는 비계층적 군집화 기법으로 다양한 형태의 데이터에 적용이 가능하며 개체간의 거리만이 분석에 필요한 입력자료로 사용되기 때문에 다양한 분야에 적용이 용이하다. 홍태호, 서보밀(2004)은 인터넷 बैं킹의 고객세분화에서 통계적으로 유의한 고객의 심리통계학적 데이터를 사용하는 방안을 제시하기 위해 고객의 심리통계학적 데이터에 k-means를 적용하여 고객세분화를 한 후 판별분석과 인공신경망을 이용해 고객세분화 집단을 예측하는 모형을 제시하였다. 제안된 방법을 통한 모형은

고객세분화를 반복하지 않고서도 새로운 고객의 세분화 집단을 미리 예측하여 마케팅에 활용할 수 있게 하였다. Kuo *et al.*(2006)은 항공운송산업에서의 세분화를 위해 SOM으로 군집의 수를 정하고 최적의 솔루션을 찾기 위해 k-means를 적용하여 두 가지 기법을 모두 사용하였다. 장남식(2005)은 카드사, 이동통신사, 보험사의 고객 인적데이터 및 거래 데이터를 대상으로 k-means 기법을 적용하여 고객을 세분화하였으며, 선정된 고객군만을 대상으로 고객의 특성을 파악하고 관리하는 것이 전체 고객을 대상으로 하는 것보다 의미가 있다는 것을 검증하였다.

III. 연구 프레임워크

세분화된 고객집단의 구매를 예측하여 기업의 마케팅 전략을 제시하기 위해 고객세분화 모형과 고객반응 예측 모형을 구축하여 세분화된 고객집단별로 프로모션에 대한 고객반응 예측모형을 제시하였다. 기업이 제공하는 프로모션에 대한 고객의 반응은 모두 다르게 나타나기 때문에 모든 고객에게 동일한 프로모션을 제공하기 보다는 세분화된 고객집단별로 특성에 맞는 프로모션을 제공할 필요가 있다. 프로모션이 필요한 고객집단과 불필요한 고객집단을 선별하여 프로모션 전략을 제공한다면 기업은 비용절감은 물론 효과적인 프로모션을 통한 매출의 극대화를 얻을 수 있다.

고객세분화를 통한 마케팅 전략은 <그림 1>과 같은 단계로 나타낼 수 있다. 고객의 특성에 따라 고객을 세분화하여 고객의 구매를 예측하고 세분화된 집단의 특성에 맞는 마케팅 캠페인을 제시한 후 고객의 반응을 통해 마케팅 캠페인의 효과를 극대화한다. 본 연구에서는 세분화된 고객집단의 특성에 맞는 캠페인 전략을 수립하기 위해 일반적으로 많이 사용되는 RFM 변수를 사용하였다. 구매빈도와 구매 거래액에 대해 SOM으로 고객을 세분화하고 세분화된 고객집단별

특성과 마케팅 캠페인에 대한 반응률을 통해 고객집단별 특성에 적합한 마케팅 전략을 제공할 수 있다. 또한 마케팅 전략에 대해 반응이 높은 고객집단에게 반응이 낮은 고객집단보다 우선적으로 마케팅 전략을 제공할 수 있다.

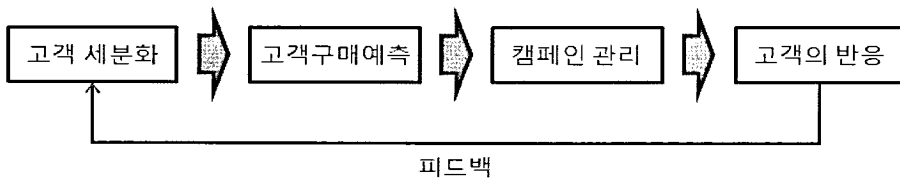
고객세분화를 통해 비슷한 특성을 가진 고객이 동일한 집단으로 분류가 되면 각 집단에 <그림 2>와 같이 고객반응 예측모형을 적용하여 고객세분화 집단별로 적절한 마케팅 캠페인 전략을 제시할 수 있다. 본 연구에서는 고객반응 예측모형을 개발하기 위해서 홍태호, 박지영(2009)이 제시한 다수의 데이터마이닝 기법을 사례기반추론으로 통합한 통합모형을 적용한다. 홍태호, 박지영(2009)이 제시한 통합모형은 k-최근접이웃법 알고리즘에 기반한 새로운 사례와 기존사례와의 유사도를 측정하여 가장 가까운 사례를 통해 결과값을 추론하는 방법이다. 본 연구에서는 로지스틱 회귀모형, 인공신경망, SVM의 고객반응 예측값들을 사례집합에 저장하고 새로운 사례에 k-최근접이웃법 알고리즘을 적용하여 가장 가까운 k이웃을 선정한 후 각 k 사례의 고객반응 결

과의 다수결로 프로모션에 대한 고객반응을 예측한다. 고객세분화 집단별로 마케팅 전략에 대한 반응률과 구매 예측모형의 성과를 통해 마케팅 전략을 제시할 목표고객의 선정과 각 집단에 적절한 캠페인 전략을 제시할 수 있도록 한다.

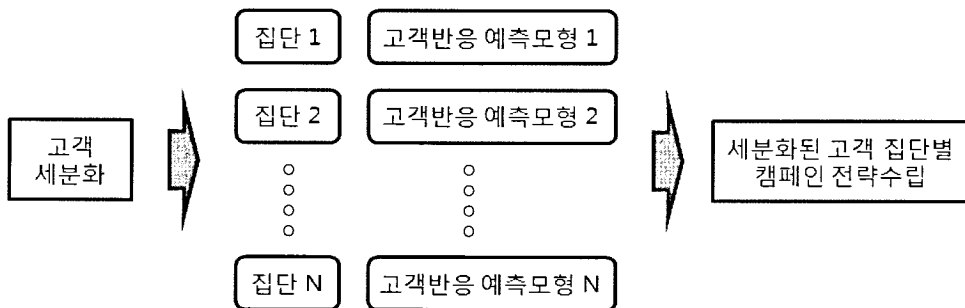
IV. 프로모션 전략 수립

4.1 데이터

본 연구에서는 목욕세제 고객에 대한 데이터를 사용하였다. 목욕세제 데이터는 공개 데이터로 소비재에 대한 고객의 구매행태를 추적하는 아시아의 시장조사기관에서 제공되었다(Shmueli et al., 2007). 데이터는 목욕세제에 대한 구매가 이루어진 600명 고객에 대한 인구통계학적 데이터, 구매거래 데이터, 프로모션 활동에 대한 데이터, 각 브랜드에 대한 구매 데이터 등 총 45개의 변수로 구성되어 있다. 인구통계학적 데이터는 매년 갱신되고 있으며 보유중인 TV, 자동차, 세탁기 등의 내구재에 대한 가중 평균값으로 고



<그림 1> 고객세분화를 통한 마케팅 캠페인



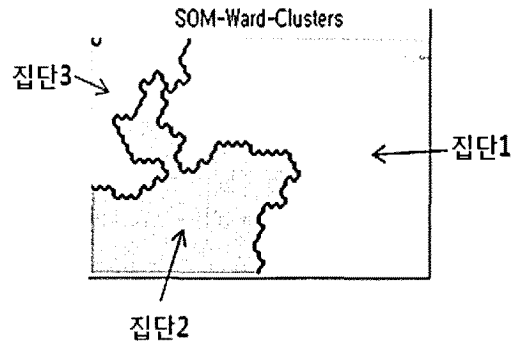
<그림 2> 세분화된 고객집단별 고객반응 예측결과를 이용한 캠페인 전략 수립

객의 부유지수를 나타내고 있다. 또한 구매거래 데이터와 브랜드에 대한 데이터는 매달 갱신된다. 기업은 600명의 고객을 위해 묶음판매, 가격 할인, 쿠폰제공, 사은품 제공 등과 같은 프로모션을 제공하였으며 그 중 묶음판매 프로모션에 대한 고객의 구매량을 제시하고 있다. 본 연구에서는 묶음판매 프로모션에 대한 반응여부를 종속변수로 하여 프로모션에 의해 구매량이 있는 고객을 프로모션에 대해 반응한 고객(299명)으로, 프로모션에 의해 구매량이 없는 고객을 반응하지 않은 고객(301명)으로 분류하였다.

4.2 고객세분화 모형

상이한 특성을 가진 고객을 동일한 집단으로 분류하기 위해 RFM 변수가 일반적으로 사용된다(Cullinan, 1977). 목욕세계 데이터의 구매 데이터는 매달 갱신이 이루어지기 때문에 최근성에 대한 변수를 제외하였다. 또한, 구매빈도를 구매한 브랜드의 수와 구매거래의 수로 나누어 단지 프로모션에 의한 구매인지 프로모션과 상관없이 브랜드 충성도에 의해 일어난 구매인지를 알아보았다. 구매한 브랜드의 수가 많은 고객은 프로모션이 이루어지는 다양한 브랜드의 제품들에

대한 구매가 이루어지며 구매한 브랜드의 수가 적은 고객은 프로모션과 상관없이 브랜드 충성도에 의해 구매가 많이 이루어진다. 따라서 구매한 브랜드의 수, 구매 거래의 수, 구매 거래액을 사용하여 Viscovery사의 SOMiner을 사용하여 3개의 집단으로 분류하였으며 <그림 3>과 같다.



<그림 3> 고객의 세분화 집단

분류된 3개의 고객집단이 서로 이질적인 집단으로 분류되었는지 확인하기 위해 각 요인별 집단간 차이를 분석하는 분산분석(ANOVA)과 이에 대한 사후 검정방법으로 Scheffe 방법을 이용하여 다중비교 분석을 실시하였다. <표 1>과 같이 3개의 고객집단이 서로 이질적인 집단으로 분

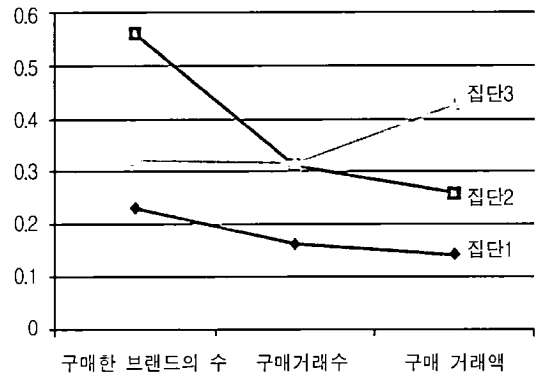
<표 1> 분산분석(ANOVA) 결과

		제곱합	자유도	제곱평균	F	유의확률
구매한 브랜드의 수	집단간	12.250	2	6.125	329.237	0.000**
	집단내	11.106	597	0.019		
	합계	23.356	599			
구매 거래 수	집단간	3.282	2	1.641	152.836	0.000**
	집단내	6.411	597	0.011		
	합계	9.693	599			
구매 거래액	집단간	5.489	2	2.744	268.995	0.000**
	집단내	6.091	597	0.010		
	합계	11.580	599			

주) **: p < 0.01, *: p < 0.05.

류되었음을 알 수 있으며 <표 2>의 사후검정 결과에서는 구매거래 수에서 집단 2와 집단 3이 유의하지 않은 것으로 나타났으나 구매한 브랜드 수와 구매 거래액의 다른 요인들은 집단들 간에 서로 유의한 차이가 존재하는 것으로 나타났다.

각 변수별로 집단의 특성을 살펴보면 <그림 4>와 같으며 집단 1에서는 구매한 브랜드의 수와 구매거래 수, 구매 거래액이 모두 적게 나타나는 집단으로 고객의 대부분을 차지하고 있는 일반 고객 집단이라 할 수 있다. 집단 2와 집단 3의 고객은 구매거래 수는 비슷하지만 구매한 브랜드의 수와 구매 거래액에서 큰 차이를 보인다. 즉 집단 2의 고객은 다양한 브랜드를 구입하지만 비교적 저렴한 제품을 구매하는 가격지향적 고객이라 할 수 있다. 선호하는 브랜드가 정해져 있지 않고 한 번의 구매에서 저렴한 여러 가지 브랜드를 구매하는 것으로 프로모션에 대한 반응률도 높게 나타난다. 집단 3의 고객은 집단 2의 고객에 비해 구매한 브랜드의 수는 적고 구매 거래액이 높게 나타나는 것으로 보아 구매하는 브랜드가 정해져 있거나 고액의 제품을 구매하는 품질지향적 고객인 것으로 보인다. 각 집단별 고객의 특성을 나타내면 <표 3>과 같이 정리할 수 있다.



<그림 4> 각 요인별 집단의 특성

4.3 고객반응 예측

고객집단별 구매 예측을 위해 본 연구에서는 로지스틱 회귀분석, 인공신경망, SVM을 이용하여 고객반응 예측을 위한 단일모형을 구축하고 이들 모형들을 사례기반추론을 이용하여 통합한 통합모형을 제시하였다. 먼저 변수선정을 위해 고객의 인구통계학적 데이터와 구매 데이터를 입력변수로 하여 로지스틱 회귀분석의 stepwise 방법을 사용하여 총 45개의 변수 중 <표 4>과 같이 6개의 변수가 묶음판매 프로모션에 유의한

<표 2> 사후검정 결과

독립 변수	집단 (I)	집단 (J)	평균차이 (I-J)	표준 오차	유의 수준
구매한 브랜드의 수	1	2	-0.331**	0.013	0.000
		3	-0.090**	0.017	0.000
	2	3	0.242**	0.019	0.000
구매 거래 수	1	2	-0.150**	0.010	0.000
		3	-0.156**	0.013	0.000
	2	3	-0.006	0.015	0.929
구매 거래액	1	2	-0.116**	0.010	0.000
		3	-0.285**	0.013	0.000
	2	3	-0.168**	0.014	0.000

주) **: p < 0.01, *: p < 0.05.

〈표 3〉 집단별 고객의 특성

집단	집단의 특성에 따른 구분	반응률	집단의 특성
집단 1 (367명)	일반고객	40.60%	가장 많은 고객이 속하는 집단으로 구매거래도 적게 일어나며 구매 거래액도 적은 집단
집단 2 (160명)	가격지향 고객	69.38%	구매하는 브랜드의 종류가 많으며 구매 거래액이 적게 나타난다. 또한 프로모션에 대한 반응률은 가장 높은 집단이다. 즉 브랜드와 상관없이 저렴한 제품의 구매를 선호하기 때문에 프로모션에 의한 구매가 많이 일어나는 집단
집단 3 (73명)	품질지향 고객	53.42%	구매하는 브랜드의 수에 비해 구매 거래액이 높게 나타난다. 프로모션에 의한 제품구매도 이루어지지만 저렴한 제품을 대량으로 구매하기 보다는 사용하던 브랜드나 고객의 제품을 소량으로 구매하는 집단

〈표 4〉 로지스틱 회귀분석에 선정된 변수

변수명	설명	B ¹⁾	S.E. ²⁾	Wald ³⁾	유의확률
X1	연속해서 구매한 브랜드의 건수	0.151	0.017	80.957	0.000
X2	묶음판매가 아닌 기업의 다른 프로모션 활동에 의한 제품의 구매비율	3.038	1.369	4.921	0.027
X3	특정제품에 대한 구매비율	3.993	1.293	9.537	0.002
X4	가격대가 저렴한 제품에 대한 구매비율	2.205	0.550	16.078	0.000
X5	천연재료의 제품에 대한 구매비율	3.665	0.818	20.055	0.000
X6	글리세린 제품에 대한 구매비율	-5.494	2.209	6.188	0.013
상수		-2.880	0.305	89.156	0.000

주) 1) 모형의 계수, 2) 표준오차, 3) Wald 통계량(클수록 선택변수의 유의성이 높음).

영향을 미치는 변수로 선정되었다. 연속해서 구매한 브랜드의 건수(X1), 묶음판매가 아닌 기업의 다른 프로모션 활동에 의한 제품의 구매비율(X2), 특정제품에 대한 구매비율(X3), 가격대가 저렴한 제품에 대한 구매비율(X4), 천연재료의 제품에 대한 구매비율(X5), 글리세린 제품에 대한 구매비율(X6)이 유의한 영향을 미치는 변수로 선정되었다. 본 연구에서는 묶음판매라는 기업의 프로모션에 반응하는 고객과 반응하지 않는 고객을 예측하기 위해 전체 600개의 데이터를 훈련용(480개)과 검증용(120개)으로 나누어 SPSS 15.0을 사용하였다.

인공신경망 모형에서는 로지스틱 회귀분석을 통해 선택된 6개의 변수를 입력변수로 사용하여

은닉노드의 수는 1, 3, 5, 7, 9, 11, 13개로 변화시켜 시행착오(trial-and-error)방법으로 하고 학습률과 모멘텀은 각각 0.1로 하였다. 각 층에 대한 출력함수는 시그모이드 함수를 적용시켰으며, 데이터를 훈련용(360개), 평가용(120개), 검증용(120개)으로 나누어 neuroshell 2 4.0을 사용하였다. 인공신경망에 대한 자세한 내용은 Rumelhart and McClelland(1986)을 참고한다.

SVM 모형은 앞에서와 같이 6개의 변수를 이용하여 SVM에서 비선형함수로 많이 사용되고 있는 가우시안 RBF 커널함수를 사용하였다. SVM 모형을 구축하기 위해 상한값을 나타내는 C와 가우시안 RBF 커널함수의 모수 g가 중요한 역할을 하며(Tay and Cao, 2001), C = {1, 20, 40, 60,

80, 100}, $g = \{1, 0.7, 0.5, 0.2, 0.1, 0.05, 0.01\}$ 로 나누어 실험하였다. 이를 위해 데이터를 훈련용(360개), 평가용(120개), 검증용(120개)로 나누어 SVM 공개 소프트웨어인 LIBSVM(<http://www.csie.ntu.edu.tw/~cjlin/libsvm>)을 사용하였다. SVM에 대한 자세한 내용은 Vapnik(1995)을 참고한다.

단일모형을 통합한 통합모형은 로지스틱 회귀 분석, 인공신경망, SVM을 통해 얻은 결과값을 이용하여 사례기반추론을 적용하여 구축하였다(홍태호, 박지영, 2009). 로지스틱 회귀모형, 인공신경망, SVM을 통해 얻은 확률값을 학습용 데이터로 사용하였으며 각 기법들에 대한 확률값과 새로운 데이터와의 유사성을 측정하기 위해 데이터 사이의 유클리드(Euclidean) 거리를 계산하여 측정하였다. 또한 k를 1부터 12까지로 하여 최적의 이웃 수를 정하였으며, k-최근접이웃법은 Excel VBA를 이용해 개발하여 실험을 수행하였다.

4.4 실험결과

세분화된 고객집단별로 프로모션에 대한 반응률과 예측모형의 성과는 <표 5>와 같다. 고객 전체의 반응률을 예측하는 것보다 세분화된 집단별로 예측한 성과가 더 높게 나타났으며 이는 프로모션을 제시할 고객집단의 선정을 통해 고

객의 특성에 맞는 맞춤형 전략을 제시할 수 있다.

집단 1의 고객은 구매한 브랜드의 수, 구매거래 수, 구매 거래액이 모두 낮게 나타나는 집단이다. 평균 구매하는 브랜드의 수가 2.8개로 나타나며 구매거래 수도 23회로 적고 구매 거래액도 \$20 이하로 적은 집단으로 프로모션에 대한 고객반응 예측 정확도가 다른 집단보다 높게 나타난다. 즉, 집단 1의 고객에게는 모든 고객에게 프로모션을 제공하기보다 목표고객을 선정하여 선별적으로 제공할 때 효과적이며 고객의 반응 여부에 대한 예측이 높기 때문에 보다 안정적으로 프로모션 전략을 제공할 수 있다. 또한 프로모션으로 인한 구매가 이루어지고는 있지만 소량의 구매만 이루어지므로 좀 더 많은 구매가 이루어질 수 있도록 프로모션 전략을 제시할 필요가 있다. 집단 2는 프로모션에 대한 반응률이 가장 높은 집단으로 프로모션에 대한 고객반응 예측모형의 예측성과도 높게 나타난다. 집단 2의 고객은 가격을 중시하는 고객들로 5종류 이상의 다양한 브랜드를 구매 하지만 구매 거래액은 평균 \$35로 집단 3의 고객(평균 \$58)보다 낮게 나타난다. 이는 프로모션을 통한 낮은 가격을 선호하기 때문에 프로모션에 대한 반응률도 높으며 모형의 예측성과도 높게 나타난다. 선호하는 브랜드가 없고 저렴한 가격으로 인한 구매가

<표 5> 고객집단별 예측모형의 성과

구분	반응률	로지트	인공 신경망	SVM	통합 모형	프로모션 전략
전체	49.2% (59/120)	63.3%	66.7%	67.5%	68.3%	-
집단 1	48.6% (35/72)	68.1%	73.6%	72.2%	75.0%	선별적
집단 2	54.6% (18/33)	63.6%	60.6%	72.7%	69.7%	모든 고객
집단 3	40.0% (6/15)	40.0%	46.7%	26.7%	33.3%	불필요

많은 고객으로 집단 2의 고객에게는 적극적인 프로모션 전략을 펼칠수록 많은 구매가 예상되는 집단으로 모든 고객에게 제공할 수 있는 프로모션 전략이 필요하다고 하겠다. 집단 3의 고객은 가격보다는 품질을 우선으로 하는 고객들로 구매하는 브랜드의 수가 평균 3.5개로 많지 않으나 구매액은 \$58로 가장 높게 나타나는 집단이다. 집단 3의 고객은 프로모션과 상관없이 구매가 이루어지기 때문에 프로모션에 대한 고객의 반응률이 낮게 나타난다. 또한 프로모션에 대한 고객반응 예측모형을 적용시켰을 때에도 프로모션과 상관없이 구매가 일어나기 때문에 예측모형에 대한 성과도 낮게 나타난다. 즉, 집단 3의 고객은 품질과 브랜드에 대한 충성도가 높기 때문에 굳이 프로모션을 제공하지 않더라도 구매가 일어나는 고객으로 집단 3의 고객보다는 집단 1과 집단 2의 고객에게 프로모션 전략을 집중할 수 있다.

V. 결론 및 연구의 한계

본 연구에서는 비슷한 특성과 성향을 가진 고객집단별 고객반응 예측을 위한 모형을 구축하기 위해 구매한 브랜드의 수, 구매거래의 수, 그리고 구매거래액을 사용하여 고객을 3개의 집단으로 세분화하였다. 세분화된 고객의 집단에 로지스틱 회귀모형, 인공신경망 모형, SVM 모형과 단일 데이터마이닝 모형들을 사례기반추론으로 통합한 통합모형을 통해 어느 고객집단에 어떠한 프로모션 전략을 제시해야하는지를 판단할 수 있었다. 집단 1은 구매한 브랜드의 수, 구매거래 수, 구매거래액이 다른 집단보다 낮은 고객집단으로 대부분의 고객이 집단 1에 포함되었으며 집단 2의 고객은 저렴한 가격을 선호하기 때문에 프로모션에 대한 반응률이 높게 나타났다. 집단 3은 가격보다 품질을 중요시하는 고객으로 프로모션에 대한 반응률이 낮게 나타났으며 프로모션에 대한 반응예측모형의 성과도 가장 떨

어지는 것으로 나타났다.

이는 세분화된 고객집단의 특성에 따라 모든 고객에게 프로모션 전략을 제시할 것인지 선별적으로 목표고객을 선정하여 프로모션 전략을 제시할 것인지 프로모션 전략이 필요없는 집단인지를 알 수 있었다. 모든 고객에게 동일한 마케팅 전략을 제공하기 보다는 세분화된 고객집단별로 특성에 맞추어 마케팅 전략을 제공하면 기업의 비용절감은 물론이고 매출의 극대화도 얻을 수 있다. 프로모션에 대한 고객의 반응이 높은 집단에는 프로모션을 통해 고객의 구매를 촉진시킬 수 있을 것이며 프로모션에 대한 고객의 반응이 낮은 집단에는 비싼 비용을 들여서 프로모션을 할 필요가 없기 때문이다.

본 연구에서는 고객을 3개의 집단으로만 분류하여 다양한 고객의 특성을 다 반영하지 못하였으며 고객집단별 반응고객 예측모형에 같은 변수를 사용하여 모형을 구축하였다. 더 많은 데이터를 확보하여 향후에는 고객의 특성에 따라 더 많은 고객집단으로 세분화하고 고객의 특성을 좀 더 반영하며 세분화된 고객집단별로 상이한 변수를 선정하여 예측모형을 구축할 수 있을 것이다. 또한 단일모형을 통한 예측모형에서도 단일모형별로 서로 다른 변수를 적용하고, 사례기반추론을 통한 통합모형에서도 통합할 때 각 모형들에 대한 가중치를 각기 다르게 조절하여 반응고객의 예측성과를 더 향상시킬 수 있는 반응고객 예측모형이 구축되어야 할 것이다.

참 고 문 헌

- 고용식, “세일즈 프로모션 전략으로서의 VMD에 관한 연구”, 한국마케팅과학회 2005 춘계 학술대회 발표논문집, 2005.
- 김진화, 남기찬, 이상중, “Support Vector Machine 기법을 이용한 고객의 구매의도 예측”, *Information Systems Review*, 제10권, 제2호, 2008, pp. 137-158.

- 안현철, 김경재, 한인구, "Support Vector Machine 을 이용한 고객구매 예측모형", 한국지능정보시스템학회논문지, 제11권, 제3호, 2005, pp. 69-81.
- 이민수, 최영찬, 유병준, "의사결정나무법을 이용한 귀납적 학습방법에 의한 정보시스템 사용자 세분화", *Information Systems Review*, 제9권, 제1호, 2007, pp. 67-84.
- 장남식, "사전 세분화를 통한 고객 분류모형의 효과성 제고에 관한 연구", *Information Systems Review*, 제7권, 제2호, 2005, pp. 23-40.
- 홍태호, 박지영, "사례기반추론을 이용한 다이렉트 마케팅의 고객반응 예측모형의 통합", *정보시스템연구*, 제18권, 제3호, 2009, pp. 375-399.
- 홍태호, 서보필, "인터넷 बैं킹에서 고객의 신념을 이용한 개인화모형을 위한 데이터마이닝", *인터넷전자상거래연구*, 제4권, 제2호, 2004, pp. 101-115.
- Ahn, H. and K. Kim, "Using genetic algorithms to optimize nearest neighbors for data mining", *Annals of Operations Research*, Vol.163, No.1, 2008, pp. 5-18.
- Baesens, B., S. Viaene, Van den D. Poel, J. Vanthienen, and G. Dedene, "Bayesian neural network learning for repeat purchase modelling in direct marketing", *European Journal of Operational Research*, Vol.138, No.1, 2002, pp. 191-211.
- Carvalho, D. R. and A. A. Freitas, "Hybrid Decision Tree/Genetic Algorithm Method for Data Mining", *Information Sciences*, Vol.163, 2004, pp. 13-35.
- Chen, Y., G. Zhang, D. Hu, and C. Fu, "Customer Segmentation based on Survival Character", *Journal of Intelligent Manufacturing*, Vol.18, No.4, 2007, pp. 513-517.
- Cheng, C. H. and Y. S. Chen, "Classifying the segmentation of customer value via RFM model and RS theory", *Expert Systems with Applications*, Vol.36, No.3, 2009, pp. 4176-4184.
- Cönül, F. F., B. D. Kim, and M. Shi, "Mailing smarter to catalog customer", *Journal of Interactive Marketing*, Vol.14, No.2, 2000, pp. 2-16.
- Cullinan, G. J., "Picking them by their batting averages' recency-frequency-monetary method of controlling circulation", Manual release 2103, Direct Mail/Marketing Association, NY, 1977.
- Kim, H. and Shin, "A Hybrid Approach Using Cased-Based Reasoning and Fuzzy Logic for Corporate Bond Rating", *Korea Intelligent Information System Society*, Vol.10, No.2, 2004, pp. 91-109.
- Kim, Y. S. and W. N. Street, "An intelligent system for customer targeting a data mining approach", *Decision Support Systems*, Vol.37, No.2, 2004, pp. 215-228.
- Kuo, R. J., Y. L. An, H. S. Wang, and W. J. Chung, "Integration of Self-organizing Feature Maps Neural Network and Genetic k-means Algorithm for Market Segmentation", *Expert Systems with Applications*, Vol.30, 2006, pp. 313-324.
- Lin, F. Y. and S. McClean, "A Data Mining Approach to the Prediction of Corporate Failure", *Knowledge-Based Systems*, Vol.14, 2001, pp. 189-195.
- Rigby, Darrell K., Frederick F. Reichheld, and Phil Schefer, "Avoid the Four Perils of CRM", *Harvard Business Review*, 2002, pp. 101-109.
- Rumelhart, D. E. and J. I. McClelland, "Parallel Distributing Processing: Exploration in the Microstructure of Cognition", Cambridge, MA: MIT Press, Vol.1, 1986.
- Serrano-Cinca, C., "Self Organizing Neural Net-

- works for Financial Diagnosis”, *Decision Support System*, Vol.17, 1996, pp. 227-238.
- Shmueli, G., N. R. Patel, and P. C. Bruce, *Data Mining for Business Intelligence*, Wiley, 2007, pp. 258-261.
- Suh, E. H., K. C. Noh, and C. K. Suh, “Customer list segmentation using the combined response model”, *Expert Systems with Applications*, Vol. 17, 1999, pp. 89-97.
- Vapnik, V., *The Nature of Statistical Learning Theory*, Springer, 1995.
- Tay, F. E. H. and L. Cao, “Application of support vector machines in financial time series forecasting”, *Omega*, Vol.29, 2001, pp. 309-317.

Information Systems Review

Volume 12 Number 2

August 2010

Predicting the Response of Segmented Customers for the Promotion Using Data Mining

Taeho Hong* · Eunmi Kim**

Abstract

This paper proposed a method that segmented customers utilizing SOM(Self-organizing Map) and predicted the customers' response of a marketing promotion for each customer's segments. Our proposed method focused on predicting the response of customers dividing into customers' segment whereas most studies have predicted the response of customers all at once. We deployed logistic regression, neural networks, and support vector machines to predict customers' response that is a kind of dichotomous classification while the integrated approach was utilized to improve the performance of the prediction model. Sample data including 45 variables regarding demographic data about 600 customers, transaction data, and promotion activities were applied to the proposed method presenting classification matrix and the comparative analyses of each data mining techniques. We could draw some significant promotion strategies for segmented customers applying our proposed method to sample data.

Keywords: *Campaign Management, Customer Response, Customer Segmentation, Data Mining*

* Associate Professor, Division of Business Administration, Pusan National University

** Doctoral Candidate, Division of Business Administration, Pusan National University

◎ 저 자 소 개 ◎



홍 태 호 (hongth@pusan.ac.kr)

현재 부산대학교 경영학부 부교수로 재직하고 있다. KAIST에서 산업공학사를 취득하였고 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 딜로이트 컨설팅에서 컨설턴트로 재직했으며, 주요 관심분야는 데이터마이닝, CRM, Business Intelligence 그리고 Social Networks 등이다. Expert Systems, Expert Systems with Applications, Asia Pacific Journal of Information Systems, 그리고 정보시스템연구 등을 비롯한 국내외 학술지에 논문을 발표하였다.



김 은 미 (keunmi100@pusan.ac.kr)

부산대학교 경영학과에서 석사학위를 취득하고, 현재 동대학원에서 박사과정 중에 있다. 주요 연구분야는 데이터마이닝, 고객관계관리, 지식경영 등이며 인터넷 전자상거래연구 등에 논문을 게재하였다.

논문접수일 : 2010년 06월 17일
1차 수정일 : 2010년 07월 31일

게재확정일 : 2010년 08월 06일