

고객 구매 행동 예측을 위한 새로운 고객 세분화 방안

A new Customer Segmentation Method for the Prediction of Customer Buying Behavior

이 장 희

한국기술교육대학교 산업경영학부

Abstract

This study presents a new customer segmentation method based on features that can predict the customer's buying behavior. In this method, we consider all variables that can affect the customer's buying behavior including demographics, psychographics, technographics, transaction pattern-related variables, etc. We define several features which are the combination of variables with the interaction effect by using C5.0, use SOM (Self-Organizing Map) neural networks in order to extract the feature's patterns and classify, and then make features' rules using C5.0 for the prediction of customer buying behavior .

1. 서론

고객세분화는 고객을 비슷한 욕구를 가지는 집단으로 나누는 작업으로 마케팅 및 CRM (Customer Relationship Management) 전략 수립에 중요한 역할을 한다. 일반적으로 마케팅 및 CRM 분야에서 고객 세분화에 지리, 인구통계, 심리, 관습, 이해관계, 고객 행동 같은 변수들을 고려하고 [1], 이중 인구통계 관련 변수나 RFM (Recency, Frequency, Monetary)[2]/FRAT(Frequency, Recency, Amount, Type) [3]과 같은 고객의 구매 행태적 척도를 활용하여 분류하여 고객의 구매 행동을 예측하고 있는데, 보다 고객 구매 행동을 정확하게 예측할 수 있는 새로운 변수 및 방법 개발에 대한 필요성을 인식하고 있다.

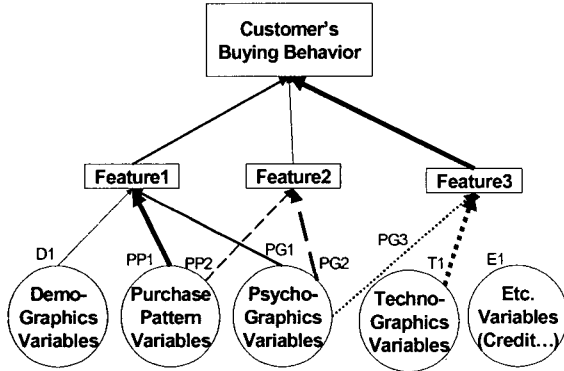
지금까지는 성별과 연령, 가족 생활주기, 소득 같은 단순한 기준으로 고객을 분류해도 무리가 없었지만 사회 변화가 빠르게 진행되면서 고객의 라이프 스타일이 다양하게 바뀌고, 이에 따라 고객의 욕구와 소비 패턴 역시 다양하게 변했기에 단순한 기준만으로는 고객을

효율적으로 분류하기 어렵다. 일례로, Forrester Research는 Technographics (기술 통계학) 라는 특정 기술 관련 제품 및 서비스의 구매 동기, 기술에 대한 고객의 기본적인 태도, 소득 수준이라는 세가지 축으로 고객을 분류하여 각 그룹별 특성 및 효과적인 마케팅 전략을 제시하기도 하였다.

본 연구는 새로운 예측 변수의 도입보다는 기존에 알려진 변수들의 상호 유기적인 작용에 의한 고객의 구매행동 변화를 예측하는 방법을 제시하고자 한다. 즉, 여러 변수들을 모두 고려하여 변수 간에 상호 유기적인 교호작용을 가지는 변수들을 찾아 그룹핑하여 이들의 변동으로 고객의 구매 행동을 설명하는 것이다.

본 연구에서는 고객 구매 행동에 영향을 줄 수 있는 모든 변수, 즉 인구통계학적 변수들, 과거 거래 이력 관련 변수들, 심리적 변수들, 기술통계학적 변수들, 고객 신용등 기타 변수들을 동시에 고려한 상황에서 상호 유기적인 교호작용을 가지는 변수들을 지능형 분석 도구인 C5.0[4]과 SOM (Self-Organizing Map) [5]의 단계적인 적용을 통해 추출하여

특성(Feature)이라 정의하고 다수의 특성의 변화에 근거하여 고객의 구매행동을 예측하는 새로운 방법을 제시한다(<그림 1> 참조).



<그림 1> 연구 모형

2. 고객 세분화

고객 세분화 분석은 얼마나 고객을 동질적인 그룹으로 잘 구분하느냐가 중요하다. 이를 위해 세분화에 사용된 척도, 세분화 그룹의 개수, 세분화에 사용된 분석기법등과 같은 요소를 잘 설계해야 한다.

서론에서도 언급한 바 있지만, 기존 연구에서는 인구통계, 심리, 지리, 관습, 이해관계와 같은 변수중 하나 또는 두개 이상의 조합을 기준으로 세분화를 수행하거나 RFM 또는 FRAT과 같은 구매 행태적 척도를 기준으로 세분화를 수행하였다.

인구통계적 변수의 어떠한 조합보다 RFM/FRAT 척도에 의한 미래의 고객 행동 예측이 더 정확하다고 주장되고 있다. 그러나 본 연구에서는 인구통계적 변수와 RFM/FRAT 척도를 별도로 활용하지 않고 고객의 구매 행동에 영향을 줄 수 있는 모든 변수들을 동시 고려하여 이들 간의 교호작용에 의한 구매 행동 변화에 초점을 두어 교호작용을 가지는 변수들을 규명, 이를 특성으로 정의함으로써 특성 변화로 구매 행동을 설명하고자 하는 것이다.

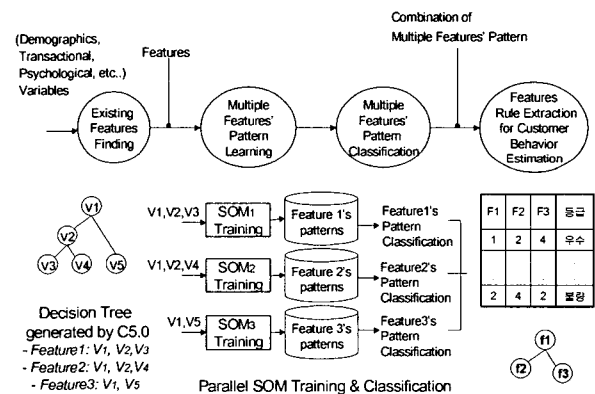
세분화되는 그룹의 수는 세분화 비용에 대한 최소화, 그룹의 해석 가능성과 응용 가능성 관점에서 최적의 그룹 수를 결정하는 연구가 수행되었다. 본 연구에서는 특성별로 각 특성

에 속하는 변수들의 동시 움직임의 패턴 수만큼 그룹 수를 결정하는 방안을 추천하였고, 전체 고객 그룹의 수는 각 특성의 그룹수를 곱한 수만큼 존재하는 것이다.

세분화에 사용되는 기법은 통계적인 방법과 신경망에 의한 방법이 많이 활용되고 있다. 즉, K-means Clustering, Principal Component Analysis, Factor Analysis, Discriminant Analysis와 같은 다변량 분석기법의 활용이 두드러지고, 신경망의 경우는 SOM의 활용이 활발히 사용되고 있다. 본 연구에서는 보다 정확한 고객 세분화를 위해 C5.0와 SOM의 연속적인 적용을 제안하고 있다. 즉, C5.0 분석을 통해 특성을 정의하고 SOM을 통해 특성의 패턴을 추출하고 분류하여 세분화 하는 것이다.

3. 특성 기반 고객 구매 행동 예측 방안

연구의 간편성을 위해, 고객별 과거 구매 실적을 기준으로 VIP, 우수, 일반 고객과 같이 고객의 등급을 결정하는 상황을 고려하여 특성을 추출하여 향후 고객의 등급을 예측하는 방법에 한정하여 설명하고자 한다. (<그림 2> 참조).



<그림 2> 고객 행동 예측을 위한 특성 규칙 추출 방안

3.1 존재하는 특성 정의 단계

본 연구에서 제안하는 방법의 1단계는 고객의 모든 데이터, 예를 들면 인구통계학 데이터, 고객 만족도 조사와 같은 각종 서베이로부터 얻을 수 있는 고객의 기호 및 심리적 경향 데이터, 기술 통계학 (Technographics) 관련 데이터, 거래 데이터 (RFM/FRAT) 및 외부에서 얻을 수 있는 개인 신

용 또는 타기업 제품의 구매 거래 데이터 등과 고객 등급 결과 (VIP/우수/일반 고객) 값으로 구성된 분석 대상 데이터 셋을 C5.0 분석을 통해 상호 교호작용이 있는 변수들을 선별하여 특성을 구성하는 것이다.

C5.0 분석 후 얻어지는 의사결정 나무의 노드 (즉, 변수) 구조에 기반하여 특성을 구성하는 것이다. <그림 2>의 예에서 C5.0 분석을 수행하여 얻어진 의사결정 나무에는 5개의 노드 (V1,...,V5) 가 있는데 노드간 구조에 기반하여 특성 1은 V1,V2, V3의 3가지 변수로 구성하고 특성2는 V1,V2,V4의 3가지 변수로, 특성3은 V1,V5의 2가지 변수로 구성한다.

3.2 특성의 패턴 추출 단계

2단계는 1단계에서 얻어진 각각의 특성들에 대해 기존 데이터에 존재하는 모든 패턴을 추출한다. 이를 위해 각 특성별로, 특성에 속하는 변수들로 구성된 데이터 셋을 구성하고 특성에 속하는 변수들의 동시적 움직임의 패턴을 SOM (Self-Organizing Map) 신경망을 이용하여 추출한다.

<그림 2>의 예에서, 특성 1에 속하는 3가지 변수인 (V1,V2, V3), 특성 2에 속하는 3가지 변수인 (V1,V2, V4), 특성 3에 속하는 2가지 변수인 (V1,V5)에 대해 각각 SOM 신경망을 구성하여 각 특성들의 패턴을 추출한다.

SOM 학습을 통한 패턴 추출시 그룹의 수는 가능한 많은 수를 지정하고, 학습후 얻어지는 각 그룹의 분포 및 평균과 같은 그룹의 특징을 각 그룹간 비교를 동일성 검정과 같은 통계적 방법을 수행하여 그룹간 동일 그룹인지 여부를 확인하여 그룹수를 줄이는 방향으로 수행한다.

3.3 기존 변수 데이터의 특성 패턴 값으로의 변환 단계

3단계는 기존 변수들의 데이터를 2단계에서 추출한 각 특성의 패턴을 기준으로 분류하여 특성 값으로 변환하는 것이다. 특성 1에 대해 9개의 패턴을 추출하였다고 가정하면, 기존 (V1,V2,V3)의 데이터를 특성1의 9개 패턴 중 어느 패턴에 속하는지 분류하는 것이다.

3.4 고객등급 예측용 특성 값 규칙 추출

마지막 4단계는 3단계에서 분류한 각 특성의

패턴 값과 고객 등급 값을 C5.0 분석을 실시하여 특성 값에 따른 고객 등급에 대한 규칙을 추출하고 향후 이를 예측에 활용하는 것이다. <그림 2>에서 보인 것과 같이 3개의 특성별로 패턴 값과 그때의 고객 등급 값을 데이터 셋으로 구성하여 C5.0 분석을 수행하면 고객등급에 대한 특성들의 규칙을 얻을 수 있다.

4. 결과 및 향후 연구

본 연구는 기존의 고객 세분화 연구에서 별도로 또는 일부 변수의 조합으로 활용되는 것을 동시에 고려하고 변수간의 상호 교호관계가 있는 변수들의 조합을 선별하여 특성으로 정의하고 이들의 변동을 관리함으로써 고객 구매 행동을 효과적으로 예측할 수 있는 방안을 제시하였다.

본 연구에서는 특성 구성 및 예측을 위한 특성 변동 규칙 추출을 위해 C5.0 분석을 수행하였고, 특성별 패턴 추출 및 패턴 규명을 위해 SOM 신경망을 활용하였다.

향후 본 논문에서 제시한 방법의 타당성 검증을 위해 현장 데이터를 확보하여 실제 적용하여 보고 문제점 및 개선방안을 도출하여 연구의 최종 결론을 보고하고자 한다.

참고문헌

- [1] Weinstein, A, Market Segmentation, Revised Ed., Probus Publishing Company, Chicago, IL, USA, 1994.
- [2] Huges, A. M., The Complete Database Marketer, 2nd Ed., McGraw-Hill, New York, 1994.
- [3] Stone, B., R. Jacobs and H. R. Wientzen, Successful Direct Marketing Methods, 7th Ed., McGRAW-HILL, New York, 2001
- [4] Quinlan, J. Ross, "C4.5: Programs for Machine Learning", Morgan Kaufmann Publishers(San Mateo, California), 1993
- [5] Kohonen, T. ,Self-Organization & Associative Memory. 3rd ed. (Springer-Verlag, Berlin), 1989.