

## 신용카드업에서 데이터마이닝의 활용 -고객 행동기반의 고객세분화-

진서훈, 안상욱<sup>1)</sup>

### 요 약

기업들이 심화된 경쟁체제 속에서 고객에 대한 보다 심층적인 이해를 필요로 하고 정보기술의 발달로 각 요소활동내용의 데이터화가 가능해짐에 따라 CRM으로 대변되는 고객 정보의 전략적 활용이 매우 중요하게 되었다. 이를 위해 기업은 고객에 대한 이해를 바탕으로 고객관리 및 마케팅을 수행하기 위한 필수적인 도구인 고객세분화를 수행하고 있다. 본 연구에서는 신용카드고객의 카드사용행태에 근거하여 서로 유사한 사용행태를 보이는 고객군으로 세분화하는 과정을 소개한다. 고객이 실제로 카드를 사용하면서 발생시킨 거래정보에만 의존하여 고객세분화를 수행하였으며 이는 마케팅의 관점에서 상당히 의미있는 내용이라 볼 수 있다. 고객세분화를 위하여 데이터마이닝기법인 k-평균군집방법과 최장연결법에 의한 계보적 군집방법을 활용하였다.

주요용어 : 데이터마이닝, 클러스터링, 고객세분화, k-평균군집방법

### 1. 서론

현대적인 마케팅캠페인 프로세스는 마케팅기회를 탐지하여 마케팅을 기획하고 적절한 대상자를 선정하여 적합한 오퍼를 제공함으로써 캠페인을 수행하고 각 고객의 반응을 수집하고 분석하는 일련의 과정으로 이루어진다. 이때 캠페인분석, 캠페인대상자선정 및 오퍼매핑의 단계 등에서 데이터마이닝기법들이 많이 활용된다. 특히 대상자 선정을 위해서는 고객에 대한 정확한 이해가 필요한데 이를 위해 고객세분화는 필수적인 과정이다. 신용카드업은 시장의 급격한 팽창과 함께 서비스제공업체의 적극적 경쟁의 결과로 포화상태에 이르렀으며 각 업체에 있어 고객관리의 중요성은 매우 커져있다. 경쟁력있는 고객관리를 위해서는 고객에 대한 정보를 통합 분석하여 고객 하나하나의 특징을 파악하고 궁극적으로 원투원마케팅 및 관리를 실현하는 것이 필요하다. 이를 위한 과정으로 고객세분화를 통한 각 세분고객별로 차별적 맞춤관리를 수행하게 된다. 신용카드사용자의 행동은 고객이 신용카드사에서 제공하는 서비스를 사용함에 있어 "어떤 서비스를 어느 정도 이용하였는가", "신용카드의 사용처는 어떤 곳인가", "신용카드의 사용이 발생한 시간은 언제인가" 등에 따라 결정된다. 즉, 언제 어떤 곳에서 어떤 서비스를 얼마만큼 이용하였는가의 내용으로 고객의 행동을 정의할 수 있다. 따라서 이용서비스관련, 이용가맹점관련, 이용시간대관련의 3가지 영역별로 세부 변수들을 생성해 내고 생성된 변수에 근거하여 서로 유사한 행동패턴을 보이는 고객을 군집화하였다. 물론 고객이 신용카드회사와의 관계를 통해 경험하는 다른 것들, 예를 들면 콜센터 접촉, 지점방문, 홈페이지이용 등의 활동들도 있으나 이와 같은 경험은 발생빈도도 적을뿐더러 실제 신용카드 사용패턴에 미치는 영향이 상대적으로 적다고 판단하여 세분화 변수에서 제외하였다.

1) 국민은행 카드마케팅팀, 서울 영등포구 여의도동 15-22

## 2. 수행방법론

행동기반의 고객세분화는 데이터마이닝의 자율학습기법(unsupervised learning)인 클러스터링이 핵심 기술요소로 활용된다. 데이터마이닝을 수행하는 방법론은 여러 데이터마이닝 툴 벤더들이 각각 고유한 이름으로 제시하고 있으나 그 내용들은 대동소이하다. 예를 들면 SAS의 경우 SEMMA(sample, explore, modify, model, assess)(SAS, 2004), SPSS의 경우 CRISP-DM(cross industry standard process)(Chapman et al., 2000)이라는 데이터마이닝 방법론을 제시하고 있으나 그 내용을 살펴보면 실제 데이터마이닝 분석 단계에 있어 큰 차이를 찾기 어렵다. 본 연구는 A카드사에서 진행되었던 행동기반의 고객세분화의 내용으로서 수행방법론은 다음의 <그림 1>과 같은 과정을 따라 수행하였다. 자율학습기법에 의한 고객세분화작업이므로 지도학습의 경우처럼 데이터마이닝 모형의 옳고 그름을 따지는 검증의 단계가 강조되지는 않는다. 고객세그먼트의 도출을 위한 알고리즘은 k-평균 군집방법과 계보적군집방법을 순차적으로 수행하였다.

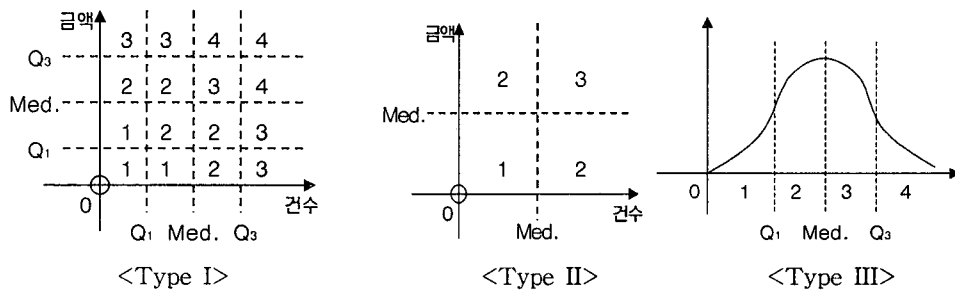
수행단계	세부 단계	주요 TASK
고객 Segmentation 전략방향수립	행동기반의 고객세분화 전략방향도출	<ul style="list-style-type: none"> <li>• 고객 행동/니즈 정의</li> <li>• 마케팅 관점의 전략방향성 도출</li> </ul>
	고객 Segmentation Attribute 도출	<ul style="list-style-type: none"> <li>• Customer profile, 운영중인 고객 스코어모델, Insight 등으로부터 고객세분화 Attribute 도출</li> </ul>
분석대상 데이터선정	Attribute 별 행동변수 Mapping	<ul style="list-style-type: none"> <li>• 각 행동변수의 source 영역 및 대응 변수 존재 여부 파악 및 Attribute과의 Mapping</li> </ul>
	분석 표본 추출	<ul style="list-style-type: none"> <li>• 고객세그먼트 모델링을 수행할 대표표본 추출</li> </ul>
데이터 품질 파악 및 파생	분석 Data Mart 구성	<ul style="list-style-type: none"> <li>• 정의된 행동변수로 구성된 분석 데이터마트 생성</li> </ul>
	Data Availability and Quality Check	<ul style="list-style-type: none"> <li>• Quality check / 변수별 분석기간 결정</li> </ul>
	추가 파생변수생성	<ul style="list-style-type: none"> <li>• 고객세분화를 위한 파생변수 생성</li> </ul>
	고객 Segmentation 알고리즘 적용	<ul style="list-style-type: none"> <li>• Clustering 알고리즘의 적용을 통한 고객 Segmentation의 수행</li> </ul>
segmentation 알고리즘적용	전체 고객대상 확장 적용	<ul style="list-style-type: none"> <li>• 표본으로부터 만들어진 고객 Segment를 전고객으로 확장</li> </ul>
	세분화된 고객군의 특성파악	<ul style="list-style-type: none"> <li>• 세분화된 고객군별로 Profile을 구함</li> </ul>
Segment Profiling	고객 Segment별 프로파일 특성에 의한 명명	<ul style="list-style-type: none"> <li>• Segment별로 Profile의 특성을 파악하고 세그먼트의 이름을 부여함</li> </ul>

<그림 1> 고객세분화 방법론

## 3. 데이터준비

고객세분화의 대상고객은 2003년 8월말 기준으로 가입기간이 1년이상이며 최근 6개월 A카드의 이용실적이 있는 고객을 대상으로 하였다. 최근 6개월 이내에 이용실적이 없어 고객행동정보가 충분히 축적되지 않은 고객은 고객의 행동을 파악하기 어려우므로 별도의 세그먼트로 구성하였다. 또한 가입기간이 1년 미만인 고객들은 신규가입고객으로 별도의 세그먼트를 구성하였다. 신용카드사의 데이터는 고객이 신용카드를 사용하면서 발생시키는 트랜잭션 데이터와 고객이 가입시에 제공한 각종 데모정보, 그리고 외부 신용평가기관으로부터 획득한 각 고객의 신용정보 등으로 구성된다. 본 연구에서는 고객의 행동정보만을 대상으로 고객을 세분화하도록 하였으며 기타정보는 고객세분화 이후에 세분화 집단별로 특성을 파악하는 데에 사용될 수 있다. 고객의 행동정보는 언제, 어디서, 무엇을, 얼마만큼 구매하였는지의 정보가 가장 기본적인 정보

가 될 수 있으며 이는 다른 산업에서 확보되는 정보에 비해 상당히 풍부하고 상세한 정보이다. 행동기반고객세분화를 위해 우선 고객의 행동을 담고 있는 변수를 생성하였다. 변수생성의 단계는 산업의 이해가 해박한 전문가의 도움이 꼭 필요한 단계로서 고객행동을 설명하기에 충분한 형태로 도출된 변수로 만들어진 세분화결과라야만 의미있는 세분화가 될 수 있는 것은 너무도 자명하다. 도출된 변수들을 클러스터링에 이용하게 되는데 이때 각 변수를 동일한 스케일로 변환해주어야 한다. k-평균군집분석은 일반적으로 유클리드거리를 사용하는데 변수가 서로 다른 척도를 갖게 되면 척도에 따라 거리에 영향을 미치는 정도가 다른 변수가 생길 수 있기 때문이다. 여기서는 중위수와 4분위수를 활용한 서열(Ordinal) 척도 및 전체 사용에 대한 상대적 개념인 비율(Ratio) 척도로의 변환을 시도하였다. 중위수와 4분위수를 활용한 변환방법은 변수의 중요도와 특징에 따라 다음의 <그림 2>과 같이 3가지로 나누어진다. 비율척도의 변환은 Type IV로 두었고 이때 변환값은 0에서 1사이의 값을 갖게 되므로 다른 변환과의 일치를 위해 4를 곱하였다. 물론 Type II의 경우는 0~3의 값을 갖게 되어 다른 Type의 변환에 비해 그 변이가 작을 수 있는데 여기서는 변수 중에 상대적으로 출현빈도가 적어 0~3사이의 변환으로 그 차이를 충분히 나타낼 수 있는 변수들을 택하였다. 각 Type별로 변환된 변수의 대표적인 것들을 살펴보면 다음의 <표 1>와 같다.



<그림 2> 변수 변환 Type들

Type I	의류, 일반음식, 연료(주유), 유통(할인점), 유통(소매점), 유통(EC)
Type II	신변잡화, 유흥업, 식품(일반), 내구재(전자), 내구재(정보통신), 유통(백화점), 유통(통신판매), 여행(숙박), 여행(운수/항공), 자동차(일반), 레저, 문화, 의료, 미용, 교육(학원), 교육(자재), 보험, 각종요금
Type III	신판건수합, 신판금액합, 신판건당금액, 신판비율, 할부건수합, 할부금액합, 할부건당금액, 할부비율, 현금서비스건수합, 현금서비스금액합, 현금서비스건당금액, 현금서비스비율, 월간이용건수분산, 반기교통카드사용금액 누적합
Type IV	시간대별건수비율, 시간대별금액비율, 주말건수비율, 주말금액비율, 보너스포인트소진율

<표 1> Type 별 변환된 변수들

#### 4. Clustering 수행

클러스터링을 위해 데이터를 세부분으로 나누었다. 첫 번째는 이용서비스데이터이고 두 번째는 이용업종데이터며 세 번째는 이용시간대에 관련된 데이터이다. 각각의 데이터영역에 대해

신용카드업에서 데이터마이닝의 활용

독립적으로 군집분석을 수행하고 각각의 영역에서 생성된 군집을 서로 교차시켜 세부군집을 생성해 내었다. 각각의 영역별로 이용한 군집방법은 k-평균군집방법이며 군집의 개수를 다양하게 변경시켜가며 군집을 생성해 보고 가장 군집간 구분이 용이하고 마케팅측면에서 타당한 군집결과를 택하였다. 이용서비스정보에 의한 군집은 일시불과 할부와 현금서비스의 이용정도에 따라 5개로 생성되었다. 다음으로 이용업종별 군집을 위해 업종의 중분류수준의 이용건수 및 금액정보를 이용하여 유사한 업종에서 이용하는 고객군을 찾아내었다. 업종변수를 기준으로 한 군집은 총 10개로 구성하였다. 끝으로 이용시간대 정보를 이용한 군집은 2개로 생성하였다. 5X10X2의 총 100개의 군집을 논리적 존재성 및 군집의 크기를 기준으로 판단하여 43개의 군집으로 축소하였다. 43개의 군집은 다시 최장연결법에 의한 계보적 군집방법을 통해 11개로 군집화하였다<표 2>.

Segment	행동특성
01 Frugal Citizen	일시불 신용구매위주의 소량사용그룹으로서 할부구매 및 현금서비스사용 거의 없음. 교통카드 이용 많음.
02 Joyful Daddy	일시불 신용구매위주로 사용하며 할부구매 거의 없고 현금서비스사용 미미함. 일시불구매 건당금액 높음. 밤시간대와 주말사용많음. 유희, 자동차, 숙박, 레저 사용높음.
03 Reliable Rider	일시불과 할부의 신용구매위주로 사용. 사용건당금액 높음. 밤시간대사용많음. 주말사용많음. 주유, 자동차관련 사용많음.
04 Installment Payer	신용구매위주로 사용하며 타 세그먼트에 비해 할부구매가 가장 많음. 현금서비스사용은 거의 없음. 오후시간대 사용많음. 주말사용많음. 의류, 전자, 출판 사용 많음.
10 Amusement Seeker	일시불 신용구매와 현금서비스 위주로 사용금액이 높은 그룹. 유희, 자동차, 정보이용관련 사용량 많음.
11 Credit Card Advocate	신용구매와 현금서비스 고루 많이 사용함. 수익성 매우 높은 그룹으로 월간 신용카드 이용 금액 가장 많음. 직업중에 걸쳐 사용이 많으나 유희은 상대적으로 적음.

<표 2> 도출된 군집(일부)

5. 고객세분화결과의 활용

행동기반의 고객세분화결과는 고객과의 커뮤니케이션에 적극 활용될 수 있다. 각 채널에서 고객접촉 시에 고객이 속한 세그먼트별로 대응전략을 차별화하여 고객의 성향에 기반하는 응대를 가능케 한다. 또한 캠페인 시에 각 캠페인의 목적에 맞는 고객군을 타겟팅함에 있어 고객세분화 결과를 활용할 수 있다. 세분화 그룹별로 데모정보를 보면 각 군집의 특성파악에 도움이 되며 아직 고객행동을 파악하기엔 가입기간이 짧은 고객의 경우도 데모정보와 그간의 사용실적에 근거하여 유사한 군집에 해당하는 고객응대전략을 수립할 수 있다. 데이터마이닝은 정보의 축적과 함께 그 효용이 날로 증가하고 있으며 신용카드업 외에도 고객의 세분화를 통한 고객관리 및 마케팅의 활용을 위해 다양하게 사용될 수 있을 것이다.

참고문헌

SAS (2004), <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>  
 Chapman et al. (2000), CRISP-DM 1.0: Step by Step Data Mining Guide, SPSS