

데이터마이닝에 의한 고객세분화 개발

진서훈¹⁾

요약

고객세분화는 기업이 관계하고 있는 고객을 이해하고 그 이해를 바탕으로 효과적인 고객관리를 수행하기 위해 필수적인 요소인데 데이터마이닝이 기업의 정보관리영역에 적극적으로 활용되면서 보다 과학적이고 최적화된 형태로 개발되고 있다. 본 연구에서는 신용카드고객의 카드사용행태에 근거하여 각 고객을 서로 유사한 사용행태를 보이는 고객군으로 세분화하는 과정을 소개하였다. 고객이 실제로 신용카드를 사용하면서 발생시킨 거래정보에만 의존하여 고객세분화를 개발하였으며 이는 마케팅의 관점에서 상당히 의미있는 내용이 될 수 있다. 고객세분화의 개발을 위하여 데이터마이닝기법인 k -평균 군집방법과 최장연결법에 의한 계보적 군집방법을 단계적으로 활용하는 이단계 군집방법을 이용하였다.

주요용어: 데이터마이닝, k -평균 군집방법, 계보적 군집방법, 고객세분화

1. 서론

기업들이 심화된 경쟁체제 속에서 고객에 대한 보다 심층적인 이해를 필요로 하고 정보 기술의 발달로 각 요소활동내용의 데이터화가 가능해짐에 따라 CRM(Customer Relationship Management)으로 대변되는 고객 정보의 전략적 활용이 매우 중요하게 되었다. 따라서 기업은 고객에 대한 이해를 바탕으로 고객관리 및 마케팅을 수행하기 위한 필수적인 도구인 고객세분화를 생성하여 활용하고 있다. 고객세분화 활용의 중요성은 마케팅프로세스에서도 잘 나타난다. 현대적인 마케팅캠페인 프로세스는 마케팅기회를 탐지하여 마케팅을 기획하고, 적절한 대상자를 선정하여 적합한 오퍼를 제공함으로써 캠페인을 수행하고, 각 고객의 반응을 수집하고 분석하는 일련의 과정으로 이루어진다(그림 1.1 참조). 이때 대상자 선정을 위해서는 고객에 대한 정확한 이해가 필요한데 이를 위해 고객세분화는 필수적인 도구이다.

국내의 신용카드업은 시장의 급격한 팽창과 함께 신용카드사의 적극적 경쟁의 결과로 상당히 빠른 속도로 포화상태에 이르렀다. 고객별 다사(多社)카드 소지가 일반적인 상황이 되었고 개별고객의 신용부실 문제가 겹치지면서 고객에 대한 이해가 어느 때보다 중요하게 되었다. 따라서 경쟁력있는 고객관리를 위해서는 고객에 대한 정보를 통합 분석하여 고객 하나하나의 특징을 파악하고 궁극적으로 원투원관리 및 마케팅을 실현하는 것이 필요하다. 이를 위한 과정으로 고객세분화를 통해 각 세분고객별로 차별적 맞춤관리를 수행하게 되는데 이는 고객과 기업간의 원투원관계의 실현을 위해 반드시 거쳐 가야할 단계이다.

1) (150-757) 서울 영등포구 여의도동 15-22, 국민은행 카드마케팅팀, 과장
E-mail: shjin9@nate.com

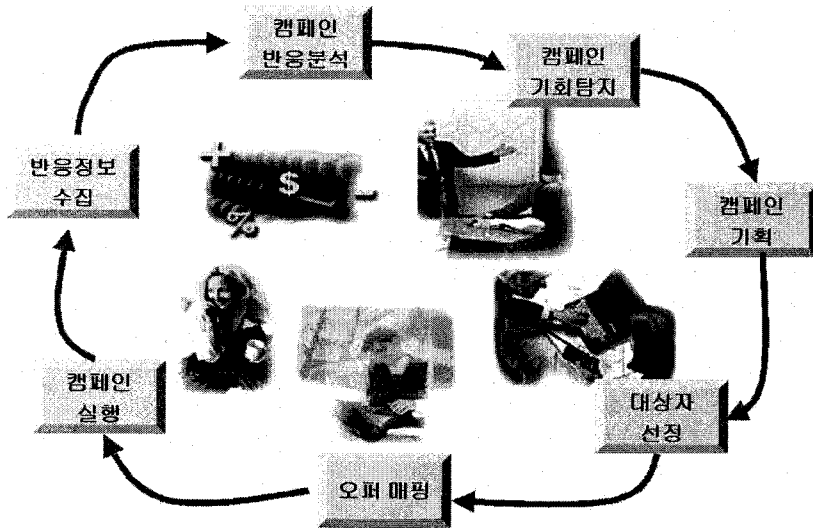


그림 1.1: 캠페인 프로세스 closed loop

본 연구는 신용카드고객의 카드사용행태에 근거하여 서로 유사한 사용행태를 보이는 고객군을 만들어냄으로써 고객관리 및 마케팅의 효율화를 위한 도구를 마련하는 내용을 담고 있다. 고객이 실제로 카드를 사용하면서 발생시킨 정보에만 의존하여 고객세분화를 개발하였으며 이는 고객의 실제 거래(transaction)에 의한 정보이므로 마케팅의 관점에서 상당히 의미있는 내용이라 볼 수 있다.

신용카드사용자의 행동은 고객이 신용카드사에서 제공하는 서비스를 사용함에 있어 언제, 어떤 곳에서, 어떤 서비스를 얼마만큼 이용하였는가의 내용으로 정의할 수 있다. 따라서 이용서비스관련, 이용가맹점업종관련, 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 적을 뿐만 아니라 실제 신용카드 사용패턴에 미치는 영향이 상대적으로 적다고 판단하여 세분화 변수에서 제외하였다. 고객세분화를 위하여 데이터마이닝기법인 k -평균 군집방법과 최장연결법에 의한 계보적 군집방법을 단계적으로 활용하는 이단계 군집방법을 이용하였다. 이단계 군집방법과 같은 군집방법의 단계적적용은 선행단계의 군집결과를 후행단계에서는 개체로 간주하여 군집분석을 수행함으로써 대규모의 자료에 적용하는 것을 수월하게 해준다. 단계적인 군집방법의 적용은 이재창 외(1997), Xiong 외(1994), El-Yaniv와 Souroujon(2001) 등에서 찾아볼 수 있다. 한편, Bazzi와 Glass(2002)는 k -평균 군집방법의 적용을 위한 초기값의 선정을 위해 계보적 군집방법을 활용하는 이단계 군집방법을 이용하기도 하였다.

수행단계	세부 단계	주요 업무
고객세분화 전략방향수립	행동기반의 고객세분화 전략방향도출	<ul style="list-style-type: none"> 고객 행동/니즈 정의 이케팅 관점의 전략방향성 도출
	고객세분화 속성요소 도출	<ul style="list-style-type: none"> 고객프로파일, 운영중인 고객 스코어모델, 전분기지식 등으로부터 고객세분화 속성요소 도출
분석대상 데이터선정	속성요소 별 행동변수 매핑(Mapping)	<ul style="list-style-type: none"> 각 행동변수의 원천 영역 및 대응 변수 존재 여부 파악 및 속성요소와의 매핑
	분석표본 추출	<ul style="list-style-type: none"> 고객세그먼트 모델링을 수행할 대표표본 추출
데이터품질파악 및 파생	분석 데이터마트 구성	<ul style="list-style-type: none"> 경외단 행동변수로 구성된 분석 데이터마트 생성
	데이터 기용성 및 품질 파악	<ul style="list-style-type: none"> 데이터 품질파악 / 변수별 분석기간 결정
	추가 파생변수생성	<ul style="list-style-type: none"> 고객세분화를 위한 파생변수 생성
고객세분화 알고리즘적용	고객세분화 알고리즘 적용	<ul style="list-style-type: none"> 군집분석 알고리즘의 적용을 통한 고객세분화의 수행
	전체 고객대상 확장 적용	<ul style="list-style-type: none"> 표본으로부터 만들어진 고객 군집을 전고객으로 확장
세분화 군집 프로파일링	세분화된 고객군의 특성파악	<ul style="list-style-type: none"> 세분화된 고객군별로 프로파일을 구함
	고객세분군별 프로파일 특성에 의한 명명	<ul style="list-style-type: none"> 세분화 군집별로 프로파일 특성을 파악하고 군집 이름을 부여함

그림 2.1: 고객세분화 수행방법론

2. 고객세분화를 위한 데이터마이닝 방법론

데이터마이닝은 일반적으로 대규모의 자료를 다루며 작업기간도 비교적 길기 때문에 효율적인 데이터마이닝작업을 위해서는 체계화된 방법론에 따라 수행하는 것이 필수적이다. 데이터마이닝을 수행하는 방법론은 여러 데이터마이닝 툴 업체들이 각각 고유한 이름으로 제시하고 있는데 그 내용들은 대동소이하다. 예를 들면 SAS의 경우 SEMMA(sample, explore, modify, model, assess, www.sas.com), SPSS의 경우 CRISP-DM(cross industry standard process, www.crisp-dm.org)이라는 데이터마이닝 방법론을 제시하고 있으나 그 내용을 살펴보면 실제 분석단계에서의 차이를 찾기 어렵다. 본 연구는 A카드사에서 진행되었던 고객세분화의 내용으로서 수행방법론은 그림 2.1과 같은 과정을 따랐다. 고객세분화는 데이터마이닝의 비지도학습기법(unsupervised learning)인 군집분석이 핵심 기술요소로 활용되는데 비지도학습기법의 적용은 지도학습의 경우처럼 데이터마이닝 모형의 옳고 그름을 따지는 검증의 단계가 강조되지는 않는다.

3. 데이터준비

고객세분화의 대상고객은 2003년 8월말 기준으로 가입기간이 1년이상이며 기준시점으로부터 최근 6개월 신용카드의 이용실적이 있는 고객을 대상으로 하였다. 최근 6개월 이내에 이용실적이 없어 고객행동정보가 충분히 축적되지 않은 고객은 고객의 행동을 파악하기 어려우므로 별도의 군집으로 구성하였다. 또한 가입기간이 1년 미만인 신규가입고객들도 별도의 군집으로 구성하였다. 신용카드사의 데이터는 고객이 신용카드를 사용하면서 받

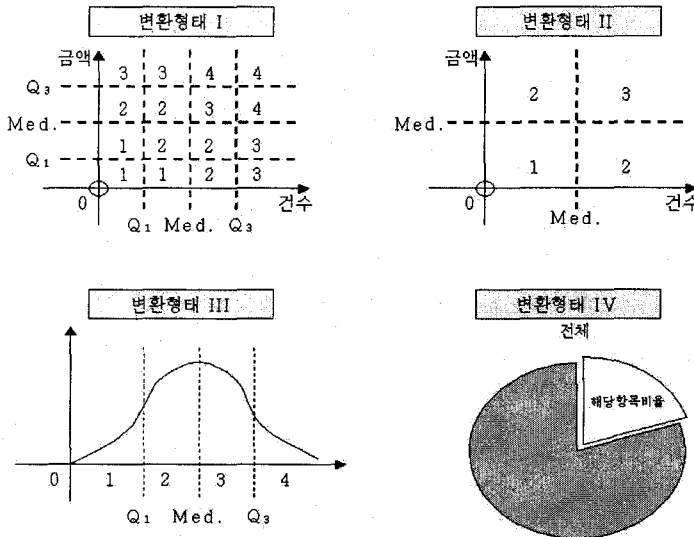


그림 3.1: 변수 변환형태들

생시키는 트랜잭션 데이터와 고객이 가입시에 제공한 각종 데모정보, 그리고 외부 신용평가기관으로부터 획득한 각 고객의 신용정보 등으로 구성되는데 본 연구에서는 고객의 행동정보만을 대상으로 고객을 세분화하도록 하였으며 기타정보는 고객세분화 이후에 세분화 집단별로 특성을 파악하는 데에 사용하였다. 고객의 행동정보는 언제, 어디서, 무엇을, 얼마만큼 구매/이용하였는지의 정보가 가장 기본적인 정보가 되며 이는 다른 산업에서는 얻기 어려운 상당히 풍부하고 상세한 정보이다. 이와 같은 행동정보를 기반으로 고객세분화의 생성을 위해 고객의 행동을 담고 있는 변수를 도출하였다. 변수생성은 산업의 이해가 해박한 전문가의 도움이 꼭 필요한 단계로서 고객행동을 설명하기에 충분한 형태로 도출된 변수로 만들어진 군집결과라야만 의미있는 고객세분화가 될 수 있는 것은 너무도 자명하다.

도출된 변수들은 군집분석에 이용되는데 이때 각 변수를 동일한 스케일로 변환해주어야 한다. k -평균 군집방법은 일반적으로 유클리드거리를 사용하므로 변수가 서로 다른 척도를 갖게 되면 척도에 따라 변수별로 거리에 영향을 미치는 정도가 다르게 되어 특정변수에 의존도가 높아지는 경우가 생길 수 있다. 여기서는 중위수와 4분위수를 활용한 서열(Ordinal) 척도 및 전체에 대한 상대적 개념인 비율(Ratio) 척도로의 변환을 시도하였다. 중위수와 4분위수를 활용한 변환방법은 변수의 중요도와 특징에 따라 그림 3.1과 같이 3가지로 나눈다. 한편 비율척도의 변환(변환형태 IV)은 0에서 1사이의 값을 갖게 되므로 다른 변환과의 일치를 위해 4를 곱하였다. 물론 변환형태 II의 경우는 0에서 3사이의 값을 갖게 되어 다른 변환형태의 변환에 비해 그 변이가 작을 수 있는데 여기서는 변수 중에 상대적으로 출현빈도가 적어 0에서 3사이의 변환으로 그 차이를 충분히 나타낼 수 있는 변수들을 택하였

표 3.1: 각 변환형태별 변수항목

변환형태	변수항목
변환형태 I	의류, 일반음식, 연료(주유), 유통(할인점), 유통(소매점), 유통(EC)
변환형태 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
변환형태 III	신판건수합, 신판금액합, 신판건당금액, 신판비율, 할부건수합, 할부금액합, 할부건당금액, 할부비율, 현금서비스건수합, 현금서비스금액합, 현금서비스건당금액, 현금서비스비율, 월간이용건수분산, 반기교통카드사용금액 누적합
변환형태 IV	시간대별별건수비율, 시간대별금액비율, 주말건수비율, 주말금액비율, 보너스포인트소진율

다. 각 변환형태별로 변환된 변수의 대표적인 것들을 살펴보면 표 3.1과 같다.

4. 군집분석

4.1. 단계 1 : k -평균 군집분석

척도를 통일한 고객행동변수를 이용하여 군집을 생성해 내기 위한 이단계 군집방법의 첫 번째 단계는 k -평균 군집분석의 단계이다. k -평균 군집방법은 대상개체를 상호배반적인 군집으로 묶어내는 데이터마이닝 기법인데 그 알고리즘은 다음과 같다(강현철 외, 2001).

- (i) 군집의 수 k 를 정한다.
- (ii) 초기 k 개의 군집의 중심을 선택한다.
- (iii) 각 관찰치를 그 중심과 가장 가까운 거리에 있는 군집에 할당한다.
- (iv) 각 군집별로 그에 속하는 관찰치를 이용해 새로운 중심(평균)을 계산한다.
- (v) 단계 (iii)과 단계 (iv)의 과정을 기존의 중심과 새로운 중심의 차이가 없을 때까지 반복한다.

k -평균 군집방법은 데이터마이닝의 활용분야에서 가장 인기있는 기법의 하나로 상대적으로 계산과정이 신속하여 대규모의 데이터를 다루는데 있어 좋은 군집방법이다. 신용카드업은 그 특성상 대규모의 자료를 다루는 것이 일반적이므로 k -평균 군집방법은 좋은 적용기법이 될 수 있다. 물론 k -평균 군집방법은 특이값에 영향받게 되는 단점을 지니고 있으나(진서훈, 1999) 본 연구에서는 그림 3.1과 같은 데이터의 사전처리를 통해 특이값의 영향으로부터 상당부분 벗어나 있으므로 k -평균 군집방법의 적용이 적절하다고 볼 수 있다.

표 4.1: k -평균 군집결과

군집분석영역	군집	군집 특성
이용서비스영역	A	일시불 신용구매만 소액 사용하는 그룹
	B	현금서비스위주로 사용하는 그룹
	C	신용구매를 사용하되 할부구매도 겸해서 이용하는 그룹
	D	신용구매와 현금서비스를 모두 사용하는 그룹
	E	신용구매위주로 사용하되 사용액이 많은 그룹
이용업종영역	1	자동차 관련업종을 집중적으로 사용. 기타업종의 사용은 적음.
	2	전반적으로 사용이 적은 그룹으로 유흥업종, 일반음식, 숙박, 레저 등에 사용 다소 많음. 소매점 사용 매우 적음.
	3	할인점사용 매우 많음. 백화점, 통신판매 사용 다소 많음. 유흥업종 사용 매우 적음.
	4	전반적으로 카드사용이 많은 그룹으로 의류, 신변잡화, 백화점, 통관, EC, 항공, 문화, 미용, 요금, 레저 등에 사용이 매우 많음. 자동차관련업종 사용은 상대적으로 적음.
	5	전반적으로 카드사용이 적은 그룹으로 의류, 신변잡화, 미용, 백화점 등의 사용이 많음.
	6	식당, 유흥, 주유, 자동차, 숙박, 보험 등의 사용이 매우 많음. EC, 여행관련업종, 각종 요금 납부에 다소 이용.
	7	모든 업종에 있어 카드사용이 매우 적은 그룹.
	8	보통사용 그룹으로 소매점사용 매우 높음. 백화점, 할인점, 통관, 의류, 의료 등에 다소 이용.
	9	고사용 그룹으로 유흥업을 제외한 전업종을 두루 많이 씀.
	10	전반적으로 사용이 적은 그룹이나 EC사용 매우 높음.
이용시간대영역	가	주중사용 많음. 낮시간대 사용 많음.
	나	주말사용 상대적으로 많음. 밤시간대 사용 많음.

k -평균 군집방법의 적용에 있어 우선 데이터를 세 영역으로 나누었다. 첫 번째는 이용서비스 관련 변수 데이터이고 두 번째는 이용업종 관련 변수 데이터며 세 번째는 이용시간대에 관련된 변수로 구성된 데이터이다. 분할된 각각의 데이터영역에 대해 독립적으로 k -평균 군집방법을 적용하였다. 각각의 적용에 있어 군집의 개수를 다양하게 변경시켜가며 군집을 생성해 보고 군집간 구분이 용이하고 마케팅측면에서 타당한 군집결과를 택하였다. 우선 이용서비스정보에 의한 군집의 결과를 보면 일시불과 할부, 그리고 현금서비스의 이용정도에 따라 5개로 생성되었다. 다음으로 업종변수를 기준으로 한 군집은 총 10개로 구성되었는데 업종별 이용건수 및 금액정보를 이용하여 유사한 업종에서 신용카드를 이용하는 고객군을 찾아내었다. 업종변수를 기준으로 한 군집은 총 10개로 구성하였다. 끝으로 이용시간대 정보를 이용한 군집은 2개로 생성하였다(표 4.1 참조). 이와 같이 각각의 영역에서 생성된 군집들은 서로 교차시켜 상세 군집을 구성하게 된다. 즉, $5 \times 10 \times 2$ 의 총 100개

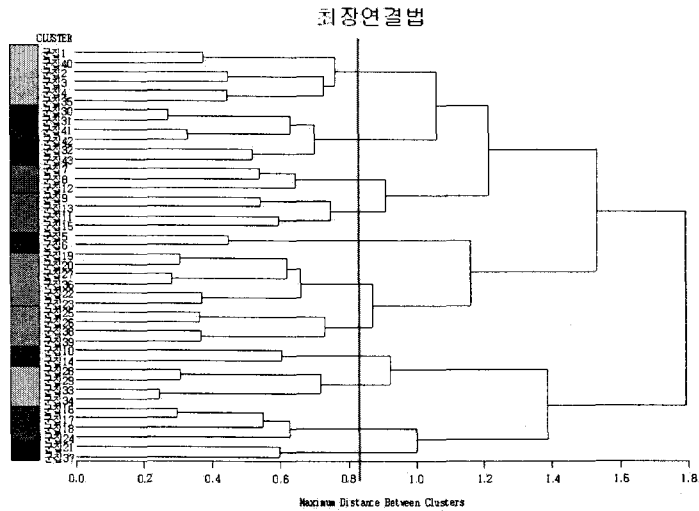


그림 4.1: 계보적 군집결과

의 상세 군집이 만들어 지게 되며 이들 군집을 논리적 존재성 및 군집의 크기를 기준으로 판단하여 43개의 군집으로 축소하였다.

4.2. 단계 2 : 계보적 군집분석

고객세분화 결과는 전문 마케터들만이 사용하는 것이 아니라 전사적으로 활용될 수 있는 것이라야 한다. 따라서 군집의 규모나 개수가 쉽게 이해하고 활용할 수 있는 수준이어야 한다. k -평균 군집방법에 의해 생성된 43개의 군집은 서로 차별화된 행동패턴을 보이는 상호 배반적인 군집이라 할 수 있다. 그러나 43개의 군집은 비교적 많은 개수의 군집으로 각각의 대고객 채널에서의 고객응대에 있어 효과적으로 활용되기 어렵다. 따라서 활용이 용이한 수준의 구조로 군집을 재편하여야 한다. 이러한 목적으로 사용할 수 있는 데이터마이닝 기법이 계보적 군집방법이다. 계보적 군집방법은 각 개체를 독립된 군집으로 두고 시작하여 서로 가까운 군집을 묶어가는 과정으로 수행된다. 이때 군집간 유사성을 측정하는 방법에 따라 최단연결법(single linkage), 최장연결법(complete linkage), 평균연결법(average linkage), 중심연결법(centroid method), 워드방법(ward method) 등으로 구분된다(Anderberg, 1973). 계보적 군집방법은 군집이 형성되는 모습을 덴드로그램(dendrogram)에 의해 확인할 수 있으므로 적절한 군집의 형성을 연구자의 시각적 판단을 통해 이룰 수 있다는 장점이 있다. 더구나 이단계 군집방법에서는 선행단계에서 43개로 생성된 군집을 개체로 하여 군집방법을 적용하기 때문에 개체가 많을 경우 덴드로그램의 표현이 어렵다는 계보적 군집방법의 단점도 극복된다. 물론 단계 2에서도 k -평균 군집방법을 비롯한 최적분리군집방법의 적용을 고려할 수 있다. 그러나 최적분리군집방법은 적절한 군집의 개수선정을 위한 고민이 수반된다는 점과 43개의 단계 1의 군집결과정보가 단계 2의 군집결과에 일정부분 숨겨져 버

리는 단점을 갖게 된다. 실제 마케팅현장에서는 상위군집(단계 2의 군집결과)에 속하는 하위군집(단계 1의 군집결과)중 일부만을 대상으로 마케팅이 수행되는 경우가 생기는데 이때 하위에 구성되어있는 군집의 구조를 알고 있으면 대상을 선정하는데 도움이 될 수 있다. 계보적인 군집방법은 이와 같은 경우 적절한 정보를 줄 수 있다.

본 연구에서는 최장연결법에 의한 계보적 군집방법을 통해 군집을 생성하였다(그림 4.1 참조). 군집간 이질성을 크게 훼손하지 않는 범위에서 적절한 군집의 개수와 규모를 고려할 때 11개의 군집이 적당한 것으로 판단되어 11개의 군집을 최종 고객세분화 결과로 구성하였다.

5. 도출된 군집의 프로파일링

이단계 군집방법에 의해 고객세분화가 도출되고 나면 각각의 세분 군집의 특징을 파악하고 요약하는 단계가 필요한데 이것이 프로파일링의 단계이다. 각 군집의 프로파일링을 위해서 고려하는 변수들은 그 형태에 따라 연속형 변수와 이산형 변수로 구분할 수 있는데 연속형 변수의 경우 분산분석 및 다중비교를 통해 군집간 차이를 살펴볼 수 있으며 이산형 변수의 경우는 대응분석 등을 활용하여 군집의 특징을 파악할 수 있다. 그림 5.1과 그림 5.2는 군집의 프로파일링을 위한 시각화된 자료의 예이다. 프로파일링을 통해 각 군집의 특징이 파악되면 군집 특징을 요약하고 활용하게 된다(표 5.1 참조).

Seg.	일시불 신용구매				할부 신용구매				현금서비스			
	건수	금액	건당금액	건비율	건수	금액	건당금액	건비율	건수	금액	건당금액	건비율
1	12,1598	817,920	67,240	0.138	0,4649	158,798	100,829	0.3788	0,4605	200,341	76,391	0.0396
2	43,9454	2,371,527	54,002	0.4177	0,8619	189,316	115,804	0.0174	1,4648	944,298	202,169	0.0350
3	22,8964	1,388,901	60,715	0.2676	3,4308	1,479,856	540,718	0.3706	1,0281	261,840	144,568	0.0627
4	16,3051	831,177	51,000	0.2669	5,1297	249,843	380,318	0.1776	0,8059	213,023	117,525	0.0541
5	4,2166	158,723	29,113	0.2430	0,1114	16,577	14,289	0.0519	7,3493	2,724,220	371,146	0.0713
6	18,9916	878,880	51,209	0.4841	5,2672	1,139,993	329,568	0.1721	1,1241	849,158	257,146	0.0344
7	7,7688	355,484	50,550	0.3009	2,3258	1,087,840	470,422	0.1513	10,3478	30,942	319,116	0.0503
8	19,7002	906,310	51,064	0.2417	4,4441	2,411,451	326,627	0.1463	0,7047	138,970	87,689	0.0122
9	13,4839	617,091	45,809	0.3427	1,3676	329,689	185,859	0.0184	1,0797	613,356	130,994	0.0137
10	38,8211	1,884,093	51,076	0.5381	3,2256	1,555,920	471,041	0.3090	11,3322	1,013,120	253,310	0.0322
11	29,3213	1,234,795	49,987	0.1786	1,1477	0,443,877	320,288	0.1050	11,3322	9,521,011	100,722	0.1544

Seg.	가립장차대신 연체원금		이동전화수분산 점수소진률		지역전력	
	연체율	연체액	연체율	연체액	지역전력	지역전력
1	5.7771	35,475	8.1609	5,1533	0.4112	313,146
2	22.2252	24,943	0.0448	16,5557	0.3225	842,794
3	25.7275	48,851	8.1485	8,4389	0.3748	1,128,158
4	32.1735	34,753	8.1120	0.0585	0.4146	905,776
5	24.1111	312,738	0.1111	3.5139	0.2925	887,391
6	29.5740	245,556	0.2472	11,2839	0.2950	330,022
7	40.7729	560,515	0.1117	7,9493	0.4291	1,100,075
8	12.8476	11,152	0.0259	22,7094	0.3833	1,722,673
9	10.8828	8,709	0.0148	1,1039	0.3038	1,152,567
10	30.8739	374,076	0.2626	17,2501	0.4917	621,947
11	16.2770	143,058	0.0793	12,416	0.3916	947,777

그림 5.1: 연속형 변수의 프로파일링 예시

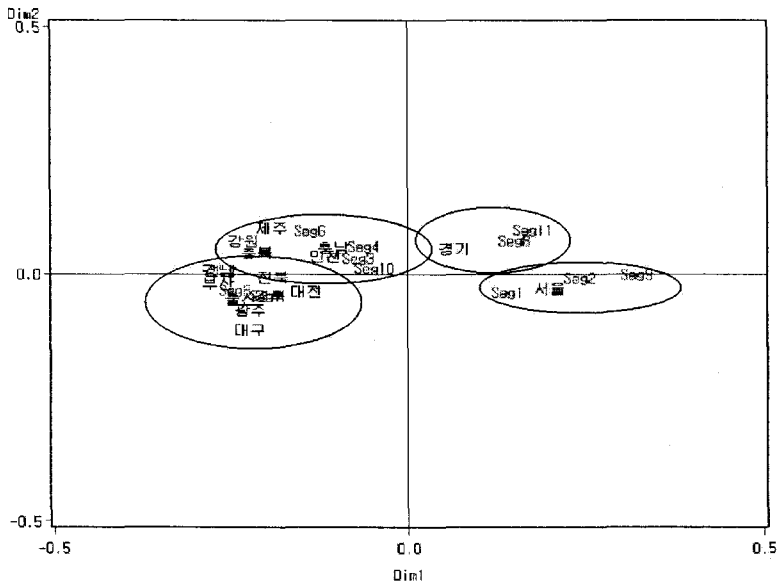


그림 5.2: 이산형 변수의 프로파일링을 위한 대응분석 예시

표 5.1: 도출된 군집의 특징요약

군집	군집 특성
군집 1	일시불 신용구매위주의 소량사용그룹으로서 할부구매 및 현금서비스사용 거의 없음. 교통카드 이용 많음.
군집 2	일시불 신용구매위주로 사용하며 할부구매 거의 없고 현금서비스사용 미미함. 일시불 구매건당 금액 높음. 밤시간대와 주말사용 많음. 유흥, 자동차, 숙박, 레저사용 높음.
군집 3	일시불과 할부 신용구매위주로 사용. 사용건당 금액 높음. 밤시간대사용 많음. 주말사용 많음. 주유, 자동차관련 사용 많음.
군집 4	신용구매위주로 사용되 타 군집에 비해 할부구매가 가장 많음. 현금서비스사용은 거의 없음. 오후시간대 사용 많음. 주말사용 많음. 의류, 전자, 통신판매 사용 많음.
...	... 중략 ...
군집 10	일시불 신용구매와 현금서비스 위주로 사용하며 이용금액이 높은 그룹. 유흥, 자동차, 정보이용관련 사용량 많음.
군집 11	신용구매와 현금서비스 고루 많이 사용함. 수익성 매우 높은 그룹으로 월간 신용카드 이용금액 가장 많음. 전업종에 걸쳐 사용이 많으나 유흥업종은 상대적으로 이용 적음.

6. 결론

본 연구에서는 신용카드고객의 신용카드 이용행태에 근거하여 고객세분화를 생성해내는 과정을 소개하였다. k -평균 군집방법과 최장연결법에 의한 계보적 군집방법을 단계적으로 이용하는 이단계 군집방법이 적용되었으며 이는 신용카드업과 같이 군집화 대상의 수가 많고 자료의 규모가 매우 큰 경우 적절한 군집분석 방법이 될 수 있다.

본 연구와 같은 이단계 군집방법에 의한 고객세분화는 기업의 입장에서 고객과의 커뮤니케이션에 적극 활용될 수 있다. 각 채널에서는 고객접촉시에 고객이 속한 군집별로 대응 전략을 차별화하여 고객의 성향에 기반하는 응대를 가능케 한다. 또한 캠페인시에 각 캠페인의 목적에 맞는 고객군을 타겟팅함에 있어 고객세분화 결과를 활용할 수 있다. 세분화 그룹별로 데모정보를 보면 각 군집의 특성파악에 도움이 되며 가입기간이 짧아 고객행동을 파악하기 부적합한 고객의 경우도 데모정보와 그간의 사용실적에 근거하여 유사한 군집에 해당하는 고객응대전략을 수립할 수 있다.

군집분석을 비롯한 데이터마이닝기법들은 정보의 축적과 함께 그 효용이 날로 증가하고 있으며 신용카드업 외에도 고객세분화를 통한 고객관리 및 마케팅에서의 활용을 위해 다양하게 사용될 수 있을 것이다.

참고문헌

- 강현철, 한상태, 최종후, 김은석, 김미경 (2001). SAS Enterprise Miner를 이용한 데이터마이닝 - 방법론 및 활용 -, 서울 : 자유아카데미.
- 이재창, 전명식, 방윤식 (1997). 인구주택총조사자료에 근거한 군집화와 응용가능성, <한국분류학회논문집>, 1, 1-13.
- 진서훈 (1999). *A Study On The Partitioning Method For Cluster Analysis*, 박사학위논문. 고려대학교.
- Anderberg, Michael R. (1973). *Cluster Analysis For Applications*, Academic press.
- Bazzi, I. and Glass, J. (2002). A multi-class approach for modeling out-of-vocabulary words, *Proceedings of the 7th International Conference on Spoken Language Processing*, 1613-1616.
- El-Yaniv, R. and Souroujon, O. (2001). Iterative double clustering for unsupervised and semi-supervised learning, *Advances in Neural Information Processing Systems*, 14, 1025-1032.
- Xiong, X., Rosenstiel, W. and Barros, E. (1994). A method for partitioning UNITY language in hardware and software, *Proceedings of the conference on European design automation*, 220 - 225.
- <http://www.sas.com/technologies/analytics/datamining/miner/semma.html>.
- <http://www.crisp-dm.org/Process/index.html>.

[2005년 3월 접수, 2005년 6월 채택]

A Development of Customer Segmentation by Using Data Mining Technique

Seohoon Jin¹⁾

ABSTRACT

To Know customers is very important for the company to survive in its cut-throat competition among competitors. Companies need to manage the relationship with each and every customer, and make each of customers as profitable as possible. CRM (Customer relationship management) has emerged as a key solution for managing the profitable relationship. In order to achieve successful CRM customer segmentation is a essential component. Clustering as a data mining technique is very useful to build data-driven segmentation. This paper is concerned with building proper customer segmentation with introducing a credit card company case. Customer segmentation was built based only on transaction data which came from customer's activities. Two-step clustering approach which consists of k -means clustering and agglomerative clustering was applied for building a customer segmentation.

Keywords: Data mining, k -means clustering, Hierarchical clustering, Customer segmentation

1) Manager, Credit Card Marketing Team, Kookmin Bank, Yoido-Dong, Youngdeungpo-Gu, Seoul, 150-757
E-mail: shjin9@nate.com