

다수의 결측치가 존재하는 가전업 고객 데이터 활용을 위한 고객분류기법의 개발

장영순^{1*} · 서종현²

¹명지대학교 경영학과 / ²중소기업진흥공단

Customer Classification Method for Household Appliances Industries with a Large Number of Incomplete Data

Young Soon Chang¹ · Jong Hyen Seo²

¹Department of Business Administration, Myongji University, Seoul, 120-728

²Small Business Corporation, Seoul, 150-718

Some customer data of manufacturing industries have a large number of incomplete data set due to the customer's infrequent purchasing behavior and the limitation of customer profile data gathered from sales representatives. So that, most sophisticated data analysis methods may not be applied directly. This paper proposes a heuristic data analysis method to classify customers in household appliances industries. The proposed PD (percent of difference) method can be used for the discriminant analysis of incomplete customer data with simple mathematical calculations. The method is composed of variable distribution estimation step, PD measure and cluster score evaluation steps, variable impact construction step, and segment assignment step. A real example is also presented.

Keywords: customer relationship management, incomplete data, household appliances industries, percent of difference (pd) method, classification, discriminant analysis

1. 서론

기업 간 경쟁이 심화되고 시장이 포화상태가 되어감에 따라 기업들은 자사 고객의 충성도(loyalty)를 높이기 위한 마케팅 전략을 수립하여 실행에 옮기고 있다. 이러한 전략의 기본 방향은 고객의 구매행동 패턴 등에 대한 통찰력(insight)을 확보하고 이를 캠페인(campaign) 등의 고객관리활동에 적극적으로 활용하는 것이다. 따라서, 고객의 기본정보(profile)와 제품에 대한 구매성향 등을 파악하여 고객의 life stage, life style 등을 분석/예측하기 위한 고객분석의 필요성이 대두되었으며, 이러한 목적으로 CRM(Customer Relationship Management)에 대한 관심이 지

속적으로 증대되고 있다. 특히 각종 데이터의 통합적인 분석을 통해 고객성향을 파악하기 위한 수단으로 analytical CRM 및 data mining에 대한 중요성이 최근 더욱 강조되고 있다.

CRM은 고객세분화를 바탕으로 일대일 마케팅(one-to-one marketing)의 실현을 통해 고객과의 장기적이고 지속적인 관계를 강화함으로써 궁극적으로 고객의 충성도를 제고하는 것으로 정의할 수 있다(Frederick, 2000). 기존에 유치한 고객들과의 거래를 지속적으로 유지함으로써 추가적인 수익을 창출하는데 비해 신규고객을 유치하는 것이 몇 배 이상의 비용이 소요된다는 것이 알려지면서 CRM의 중요성이 더욱 강조되고 있다. 따라서, 기업의 마케팅 전략이 신규고객을 유치하는 데서

*연락처 : 장영순 교수, 120-728, 서울특별시 서대문구 남가좌동 50-3 명지대학교 경영학과, Fax : 02-300-0734,

E-mail : yschang@mju.ac.kr

2005년 10월 접수, 2회 수정 후 2006년 2월 게재 확정.

벗어나 기존 고객들을 오랜 기간 자사의 충성고객으로 유지하기 위한 방안을 모색하는 것으로 초점이 옮겨가고 있다(Bart *et al.*, 2000).

CRM 전략을 수립하고 이를 성공적으로 실행에 옮기기 위해서는 객관적인 데이터를 바탕으로 고객의 니즈 및 구매 패턴을 정확히 파악하는 것이 매우 중요한 요소이다. 이를 위해 고객으로부터 각종 정보를 수집/분석하기 위한 기반을 마련하는 analytical CRM과 대용량 고객 데이터 분석기법인 data mining이 등장하였다. 전통적인 통계적 방법은 적은 수의 표본을 활용하여 모집단에 대한 추정 및 검정을 위한 기법들의 개발에 중점을 두고 있으나, data mining은 거대하고 복잡한 데이터를 효과적/효율적으로 탐색/분석하여 비즈니스 전략수립에 필요한 지식(knowledge)을 발견하고 고객에 대한 통찰력을 획득하기 위한 다양한 기법들의 개발에 중점을 두고 연구되고 있다. 즉, analytical CRM 및 data mining은 고객의 기본정보 및 상품자료와 캠페인 수행 이후 얻어진 고객 반응자료 등 기업이 보유하고 있는 모든 고객 관련 자료를 활용하여 고객의 특성을 파악하고 이를 실제 마케팅 활동에 적용할 수 있도록 가공하는 기법으로 정의할 수 있다(Addison, 1996).

그동안 analytical CRM에 관한 연구는 소매유통분야(Zeithaml *et al.*, 1996), 금융분야(Athanassopoulos, 2000; Van and Lariviere, 2004), 통신분야(Mozer *et al.*, 2000; Weerahandi and Morita, 1995) 등에서 활발히 연구되어 왔다. 이러한 산업분야에서는 고객과의 접촉 및 거래가 빈번히 발생하고 있으며 비교적 정확한 고객 데이터의 수집이 가능하므로, 다양한 분석방법을 활용하여 고객에 대한 통찰력을 확보할 수 있고 이를 바탕으로 마케팅들이 고객들에 대한 관리방안을 즉시적으로 변경할 수 있다. 그러나, 구매가 드물게 발생하거나 고객 데이터의 누락(missing)이 많은 경우에는 기존에 개발된 분석방법으로는 고객의 구매행태로부터 고객의 니즈를 파악하기가 용이하지 않아 고객이 원하는 시점에 합리적인 제품과 서비스에 대한 즉시적인 제공이 불가능할 수 있다. 이러한 문제점들로 인해 가전업, 자동차업, 가구업 등의 제조업에서는 CRM, 특히 data mining 및 analytical CRM의 적용이 미진한 상태이다.

이 논문에서는 제조업, 특히 가전업에서 발생하는 고객의 기본정보 및 구매행태 등에 관한 불완전한 정보들을 바탕으로 고객을 효율적으로 분류할 수 있는 발견적 방법(heuristic method)을 제안하고, 이를 활용하여 가전업에서 교차판매

(cross-selling)를 위해 고객분석을 수행한 사례를 제시한다.

이 논문에서 제안하는 고객분류방법은 다음과 같다. i) 주요 제품을 모두 구매한 일부 고객을 대상으로 군집화(clustering)를 실시하여 고객 cluster를 설정한다. ii) 임의의 고객을 설정된 cluster에 할당하기 위해 변수를 이산화하고 제안하는 PD(Percent of Difference) measure-cluster에 대한 고객의 평균적인 분포 대비 특정한 변수값을 갖는 경우의 각 cluster에 분포하는 정도를 측정하기 위한 척도 - 를 활용하여 군집점수(cluster score) - 임의의 고객이 각 cluster에 속할 가능성 - 및 변수별 중요도(variable impact)를 계산한다. iii) 고객에 대한 데이터가 주어지면 변수의 중요도 및 군집점수를 가중평균한 후, 가장 큰 값을 갖는 segment에 할당한다.

이 논문의 구성은 다음과 같다. 2장에서는 가전업에서 보유하고 있는 고객 관련 데이터의 특성에 관해 살펴보고, 3장에서는 고객을 분류하는 방법을 제안한다. 4장에서는 제안된 방법을 활용하여 가전업에서 고객분석을 수행한 예를 설명한다.

2. 데이터 특성

제조업에서 관리하고 있는 고객 관련 데이터들은 고객이 제품을 구입하거나 서비스를 받게 되는 경우 작성하게 되는 고객 카드와 고객을 방문하여 조사한 제품보유 현황을 바탕으로

- i) 고객 profile
- ii) 고객이 과거에 구매한 제품에 관한 구매이력
- iii) 제품의 설치 혹은 서비스를 위해 설치/서비스 기사가 고객에게 방문한 후 얻게 되는 자사 혹은 타사의 제품 보유 정보
- iv) 고객이 서비스를 받는 경우 얻게 되는 서비스 이력

등 4가지로 구분할 수 있다. <Table 1>은 A사에서 보유하고 있는 몇 가지 변수들을 위의 4가지 기준에 의해 구분한 것이다.

<Table 2>는 <Table 1>의 고객 profile 변수에 대한 누락률(missing rate)을 표시한 것이다. A사의 경우 고객의 유일성을 보장할 수 있는 주민등록번호를 key 값으로 하여 고객정보가 구성되어 이로부터 파생된 나이와 성별은 누락이 없는 정보를 얻을 수 있다. 그러나 이를 제외한 나머지 정보들 특히 직업이나 소득수준의 경우에는 고객에 따라 정보제공을 꺼리는 경우가 많으므로 40%에서 많게는 95% 이상의 데이터 누락률을 보이고 있다. 뿐만 아니라 누락률이 낮다고 하더라도 결론 여부,

Table 1. Customer variables

class	variables
customer profile	private information: resident registration number, address, phone number, name, scholarship, job, hobby, etc. income information: income level, dwelling type, living space, etc. family information: marriage, number of children, age of children, joint breadwinner, etc.
purchasing record	model, time, place, etc.
information of holdings	maker, model, number of holdings etc.
service record	model, failure type, failure time, service satisfaction, etc.

자녀수, 자녀나이 등과 같은 변수들은 정합성이 보장되지 않는 경우가 다수 존재하여 고객의 구매행태 파악이나 구매욕구가 큰 고객의 추출에 직접적으로 이용하기에는 한계가 있다.

고객의 구매성향을 파악하기 위한 중요변수인 제품구매이력에 대한 정확한 정보를 얻기 위해서는 판매처에서 판매한 제품과 고객을 빠짐없이 기록해 주어야 한다. 그러나 다양한 제조사의 제품을 판매하는 양판점의 경우에는 고객의 구매 데이터를 제조사에 제공해 주지 않는 경우가 많고, 소형 가전제품의 경우에는 하나의 제조업체 제품만을 취급하는 대리점에서 고객정보관리 시스템에 기록하지 않는 경우가 종종 발생한다. <Figure 1>은 구매한 제품의 종류에 따른 고객의 비율을 나타낸 것으로 한 제조업체에서 두 종류 이상의 제품을 구매한 고객의 비율이 상당히 낮은 것을 알 수 있다. 즉, 판매처에서의 미입력으로 인해 제품의 구매 데이터가 누락되고 있음을 추측할 수 있으며, 가전제품의 특성상 여러 개의 제품을 하나의 제조사에서 구매하기보다는 제품의 특성을 고려하여 다양한 제조업체를 선택하는 것을 알 수 있다. 따라서, 누락된 데이터가 많은 경우에도 손쉽게 사용할 수 있는 고객분류를 위한 모형이 필요하다.

3. Percent of Difference(PD)를 활용한 고객 Segmentation 방법

3.1 고객 Segmentation 방법의 기존 연구

고객 segmentation은 사전 정보 없이 데이터만을 활용하여 동질한 특성을 갖는 고객들을 찾아내어 그룹화하는 군집분석(cluster analysis)과 기설정된 집단에 고객을 분류하는 판별분석(discriminant analysis)으로 나누어 볼 수 있다. 군집분석은 개별 고객 간의 거리기준에 의하여 기준이 되는 대상들로부터 시작

하여 나무모양의 계층구조를 하위 상향식으로 형성해 나가는 single linkage cluster analysis, complete linkage cluster analysis, Ward's method 등의 hierarchical method와 전체 군집의 개수를 정해 둔 상태에서 설정된 군집의 중심에 가까운 개체를 군집에 하나씩 포함해 가는 K-means cluster analysis, hill-climbing method 등의 nonhierarchical method로 나누어 볼 수 있다. 판별분석을 위해서는 decision tree, logistic regression, neural network 등의 방법이 주로 사용된다(Dillon and Goldstein, 1984; Berry and Linoff, 1997). 특히, Punj and Stewart(1983)는 통계적 군집방법 중 Marketing Research 분야에서 활용되는 기법에 대해 자세히 정리하였다.

이러한 기법들은 대부분 데이터가 완전하여 결측값이 없는 경우에 활용할 수 있는 방법으로 결측값이 많은 제조업 고객 데이터에 직접 적용하기에 한계가 있다. 결측값이 존재하는 데이터에 기존의 방법을 적용하기 위해서는 결측값을 갖는 데이터를 분석에서 제외시키거나(Afifi and Elashoff, 1966), 적절한 모델을 수립하여 활용하고(Krishnamoorthy and Maruthy, 1998; Huang and Zhu, 2002), EM algorithm 등의 통계적 기법 혹은 인공신경망 등의 기계학습기법을 활용하여 결측값을 다른 값으로 대체하게 된다(Nijman and Kappen, 1997; Schafer, 1997; Granger et al., 2001; Lim et al., 2005). 그러나 이러한 방법은 결측값이 상당히 많이 존재하는 경우에는 분석의 정확도가 떨어지고 결측값들을 추정하는 데 많은 시간이 소요되어 사용하기가 어렵게 된다. 따라서, 결측값이 상당수 존재하는 경우에도 손쉽게 사용이 가능한 방법 및 절차의 개발이 필요하다.

3.2 PD를 활용한 고객분류방법

Percent of Difference(PD)를 활용한 고객분류방법은 구매가 빈번하게 발생하지 않고 정보획득의 어려움으로 인해 결측값이 많이 발생한 경우 간단한 수치계산을 통하여 고객을 판별할 수 있는 방법으로, i) loyalty가 높다고 판단되는 일부 고객들 중

Table 2. Missing rate of variables related to customer profile

private information	missing rate	income information	missing rate	family information	missing rate
resident registration number	0%	family income	96%	marriage	47%
scholarship	92%	dwelling type	94%	number of children	38%
job	59%	living space	39%	age of first child	95%
hobby	91%	price of house	84%	joint breadwinner	39%

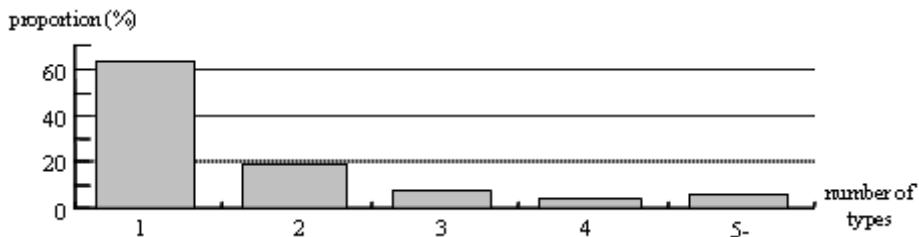


Figure 1. Customer proportion according to the number of purchasing product types.

특정한 변수로 구성된 데이터를 활용하여 군집분석을 실시한 이후 ii) 모든 변수들을 활용하여 전체 고객을 각 군집에 분배하는 발견적(heuristic) 방법이다. 이 논문에서 사용하는 PD는 기준값에 비해 관측된 값이 어느정도의 차이가 존재하는가를 알아보기 위해 널리 사용되는 척도로

$$\frac{\text{관측값} - \text{기준값}}{\text{기준값}} \times 100$$

으로 정의된다.

<Figure 2>는 가전업의 데이터 구조와 제안하는 방법을 활용하여 고객을 분류하는 방법을 도식화한 것으로, 가전업에서는 각 변수들에 대해 모든 고객들의 데이터를 얻을 수 없으나 주요변수에 대해 완전한 데이터를 갖고 있는 일부 고객이 존재함을 알 수 있다. 제안하는 방법은 i) 주요변수에 대한 모든 값을 보유하고 있는 고객 데이터를 활용하여 cluster를 설정하는 clustering 단계와 ii) clustering에 사용된 변수뿐 아니라, 그 이외의 모든 변수를 활용하여 설정된 cluster에 고객들을 배분하는 규칙을 생성하는 discriminant rule 설정단계, iii) 설정된 rule에 따라 고객을 각 segment에 할당하는 segment 할당단계로 나누어진다.

3.2.1 Clustering 단계

이 단계는 고객의 segments를 정의하는 단계로 잘 정제되고 누락이 없는 일부 고객 데이터를 추출하여 군집분석을 실시하는 단계이다. 즉, 고객의 특성을 파악할 수 있거나 향후 관측을 하고자 하는 주요 제품군을 모두 구매한 고객들을 선별하여 이를 새로운 모집단으로 정의하고 분류하는 작업을 수행한다. 이때 사용되는 data set에는 결측값이 존재하지 않으므로 기존에 개발된 clustering 기법들을 활용할 수 있다. 만일 고객관리 목적으로 사전에 고객에 대한 cluster를 설정해 놓았다면 이 단계를 생략할 수 있다.

예를 들어 가전제품을 생산, 판매하는 기업의 경우에는 대

부분의 소비자가 소유하고 있다고 판단되며 대표적인 가전제품이라 할 수 있는 TV, 냉장고, 세탁기를 모두 구매한 고객들을 선별하여 K-means cluster analysis를 수행할 수 있을 것이다.

3.2.2 Discriminant Rule 설정 단계

이 단계는 clustering 단계에서 얻어진 고객 cluster에 모든 고객들을 분배하기 위한 단계로 제조업의 특성상 불가결하게 발생하는 다수의 결측값을 효과적으로 처리하기 위한 것이다. 임의의 고객이 clustering 단계에서 설정된 cluster 중 어디에 속할 것인가를 결정하기 위한 기준을 수립하는 것으로 변수들의 관측치에 따라 각 cluster에 속할 점수를 계산하고 변수들의 중요도 - 변수들의 discriminant 능력 - 를 고려하여 가중평균하는 방법을 사용한다.

이 논문에서는 변수들이 이산형(discrete)인 경우에 적용할 수 있는 방안을 제시하며, 제안하는 방법은 i) 변수의 cluster별 분포계산, ii) PD measure 계산, iii) cluster score 계산 및 iv) variable impact 계산의 절차로 수행된다.

이 단계에서는 clustering 단계에서 사용된 변수뿐 아니라, 고객과 관련된 모든 변수를 활용하여 분석을 실시하게 된다. 즉, clustering 단계에서 사용된 변수(예를 들어 V_1, V_2) 이외에 사용되지 않은 변수(예를 들어 V_3, V_4, V_5) 모두를 활용하여 discriminant rule을 설정한다. 이렇게 설정된 rule은 고객이 V_1, \dots, V_5 중 하나의 변수값만을 갖더라도 고객을 기 설정된 cluster에 할당하는 데 사용할 수 있다. 이때, 결측치의 패턴이 랜덤하다면 기 설정된 cluster는 통계적으로 전체 고객을 대표하게 되므로 V_1 및 V_2 를 결측치로 갖고 있는 고객들에 대해서도 분류가 가능하게 된다.

변수의 Cluster별 분포계산

각 cluster별 변수의 분포는 n 개의 cluster (C_1, C_2, \dots, C_n)에 대한 각 변수 (V_1, V_2, \dots, V_k)의 분포를 구하는 것으로

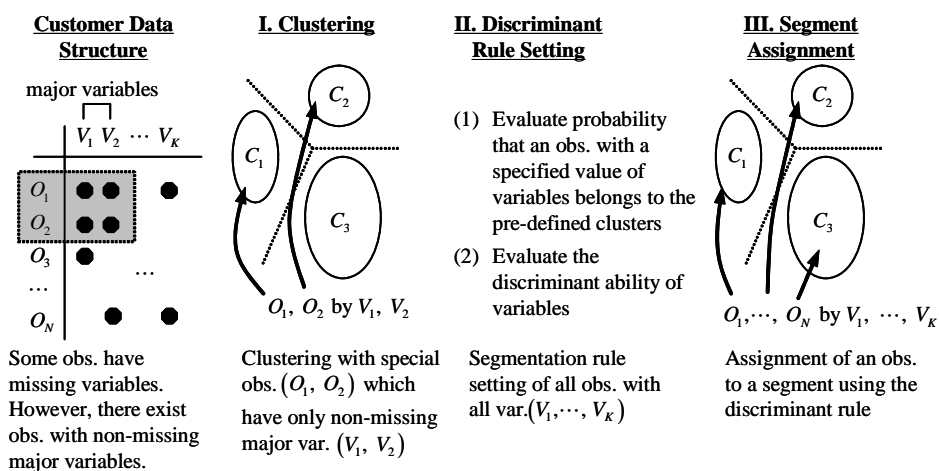


Figure 2. Proposed customer segmentation procedure.

$$P_{C_i}(V_k = v) = \frac{n(\{V_k = v\} \cap \{C_i\})}{n(\{V_k = v\})},$$

$$i = 1, \dots, n, \quad k = 1, \dots, K \quad (1)$$

로 추정할 수 있다. 여기서 $n(\{V_k = v\})$ 는 변수 V_k 가 v 의 값을 갖는 관측치의 개수이고, $n(\{V_k = v\} \cap \{C_i\})$ 는 cluster C_i 에 속한 관측치 중 $V_k = v$ 인 관측치의 수를 의미한다.

예를 들어 구매한 TV의 Level (V_1)을 4단계 (L_1, L_2, L_3, L_4)로 나눈 경우, clustering 단계에서 사용된 고객 중 L_1 의 TV를 구매한 고객이 1,000명이고 이 중 cluster 1에 속한 고객이 0명, cluster 4에 속한 고객이 150명이라고 하면,

$$P_{C_1}(V_1 = L_1) = \frac{0}{1,000} = 0.0,$$

$$P_{C_4}(V_1 = L_1) = \frac{150}{1,000} = 0.15$$

가 된다.

PD Measure 계산

cluster C_i 에 속한 고객의 비율을 P_{C_i} 라 할 때, P_{C_i} 값보다 $P_{C_i}(V_k = v)$ 의 값이 크다면 $V_k = v$ 인 고객은 C_i 에 속할 가능성이 높다고 할 수 있을 것이고, $P_{C_i}(V_k = v)$ 의 값이 작다면 C_i 에 속할 가능성이 낮다고 판단할 수 있다. 따라서, P_{C_i} 와 $P_{C_i}(V_k = v)$ 의 차이를 고객을 분류하는 기준으로 삼을 수 있다. PD measure는 이러한 차이를 나타낼 수 있는 기준으로 cluster에 속한 고객비율 대비 그 고객들이 특정한 변수값을 갖는 비율로 표현되며 다음의 식(2)와 같이 정의할 수 있다.

$$PD_{C_i}(V_k = v) = 100 \times \frac{P_{C_i}(V_k = v) - P_{C_i}}{P_{C_i}},$$

$$i = 1, \dots, n, \quad k = 1, \dots, K \quad (2)$$

이때, $P_{C_i}(V_k = v) \geq 0$ 이고 $P_{C_i} > 0$ 이 되어 $PD_{C_i}(V_k = v) \geq -100$ 이 된다. 이때 변수 V_k 의 값이 v 인 고객의 경우, $PD_{C_i}(V_k = v)$ 가 크면 cluster C_i 에 속할 가능성이 크다는 것을 의미하고, -100 에 가까울수록 cluster C_i 에 속할 가능성이 작다는 것을 의미한다.

예를 들어 clustering 단계에서 고객 cluster가 C_1, \dots, C_4 의 4개로 나누어졌고 구매한 TV의 Level (V_1)을 4단계로 나눈 경우 P_{C_i} 및 $P_{C_i}(V_1 = L_1)$ 이 <Table 3>과 같다고 하면,

$$PD_{C_1}(V_1 = L_1) = 100 \times \frac{0.0 - 0.2}{0.2} = -100$$

$$PD_{C_2}(V_1 = L_1) = 100 \times \frac{0.35 - 0.3}{0.3} = 16.67$$

$$PD_{C_3}(V_1 = L_1) = 100 \times \frac{0.0 - 0.1}{0.1} = -100$$

$$PD_{C_4}(V_1 = L_1) = 100 \times \frac{0.65 - 0.4}{0.4} = 62.5$$

이 된다. 즉, 구매한 TV의 Level이 L_1 이라면 고객들의 평균적인 성향에 비해 C_2 와 C_4 에 속할 가능성이 크고, C_1 과 C_3 에 속할 가능성은 거의 없다는 것을 의미한다.

Table 3. Evaluation of percent of difference

cluster	C_1	C_2	C_3	C_4
P_{C_i}	0.2	0.3	0.1	0.4
$P_{C_i}(V_1 = L_1)$	0.0	0.35	0.0	0.65
$PD_{C_i}(V_1 = L_1)$	-100	16.67	-100	62.5

Cluster Score 계산

PD measure 값을 특정 변수값을 갖는 고객이 각 cluster에 속할 정도를 나타내나, P_{C_i} 및 $P_{C_i}(V_k = v)$ 에 따라 다양한 값을 갖게 된다. 따라서 PD measure를 표준화하여 상대적인 값을 사용하는 것이 바람직하다. 이 논문에서는 PD 값이 0과 1 사이의 값이 되도록 표준화하여 사용하는 방안을 제안한다. 모든 i 에 대하여 $PD_{C_i}(V_k = v)$, $i = 1, \dots, n$ 값 중 가장 큰 값을 $MaxPD(V_k = v)$ 라 하고 가장 작은 값을 $MinPD(V_k = v)$ 라 하면 표준화된 변수별 cluster score는 다음의 식(3)과 같이 정의할 수 있다.

$$SPD_{C_i}(V_k = v) = \frac{PD_{C_i}(V_k = v) - MinPD(V_k = v)}{MaxPD(V_k = v) - MinPD(V_k = v)} \quad (3)$$

여기서 $SPD_{C_i}(V_k = v)$ 는 $0 \leq SPD_{C_i}(V_k = v) \leq 1$ 의 값을 갖으며 cluster C_i 에 속할 상대적인 가능성이 가장 낮은 경우는 0, 가장 큰 경우는 1이 된다. 이때, $SPD_{C_i}(V_k = v)$ 는 $V_k = v$ 인 값 내에서의 상대적인 크기만을 비교하는 것으로 $SPD_{C_i}(V_k = v)$ 와 $SPD_{C_i}(V_k = u)$, $v \neq u$ 혹은 $SPD_{C_i}(V_k = v)$ 와 $SPD_{C_l}(V_l = v)$, $k \neq l$ 등 서로 다른 변수값과 서로 다른 변수 간의 비교는 의미가 없다.

<Table 3>의 자료를 활용하여 cluster score를 구해 보면,

$$SPD_{C_1}(V_k = L_1) = \frac{-100 - (-100)}{62.5 - (-100)} = 0,$$

$$SPD_{C_2}(V_k = L_1) = 0.72,$$

$$SPD_{C_3}(V_k = L_1) = 0, \quad SPD_{C_4}(V_k = L_1) = 1$$

이 된다.

Variable Impact 계산

특정 변수값을 갖는 경우 각 cluster에 속할 가능성을 구한 이후에는 각 변수들이 고객을 구분하는 능력의 정도를 계산해야 한다. PD measure의 값이 크면 고객이 그 cluster에 속할 가능성

이 높고, 작으면 속할 가능성이 낮은 것이므로 PD measure의 산포(dispersion)가 크면 클수록 그 변수값은 고객을 잘 분류한다고 할 수 있다. PD measure의 산포는 분산(혹은 표준편차), 범위, 사분위 범위 등 다양한 방법으로 계산이 가능하나 이 논문에서는 계산의 편의를 위해 PD measure의 범위(range) - $RPD(V_k = v) = MaxPD(V_k = v) - MinPD(V_k = v)$ - 를 이용하여 변수의 판별능력을 계산하는 방법을 제안한다. 따라서, 다음의 식 (4)와 같이 $RPD(V_k = v)$ 의 평균값을 구하여 variable impact $VI(V_k)$ 를 계산할 수 있다.

$$VI(V_k) = \frac{1}{100} \cdot \sum_{all v} [RPD(V_k = v) \cdot Pr\{V_k = v\}] \quad (4)$$

이때, $VI(V_k)$ 의 값이 큰 변수들이 고객을 분류하는 능력이 뛰어나다고 할 수 있다.

예를 들어, TV 및 냉장고의 구매 Level (V_1, V_2)을 각각 4단계 (L_1, L_2, L_3, L_4)로 하고 각각의 $RPD(V_k = v)$ 와 V_k 의 분포가 <Table 4>와 같다고 한다면

$$VI(V_1) = \frac{1}{100} \times \{100 \cdot 0.5 + 200 \cdot 0.25 + 300 \cdot 0.2 + 700 \cdot 0.05\} = 1.95$$

$$VI(V_2) = \frac{1}{100} \times \{250 \cdot 0.3 + 200 \cdot 0.3 + 350 \cdot 0.25 + 600 \cdot 0.15\} = 3.13$$

Table 4. Evaluation of variable impact

variables	range of PD and distribution of variables	$v = L_1$	$v = L_2$	$v = L_3$	$v = L_4$	variable impact
TV	$RPD(V_1 = v)$	100	200	300	700	1.95
	$Pr\{V_1 = v\}$	0.5	0.25	0.2	0.05	
refrigerator	$RPD(V_2 = v)$	250	200	350	600	3.13
	$Pr\{V_2 = v\}$	0.3	0.3	0.25	0.15	

Table 5. Cluster score and variable impact for segment assignment

variables	variable impact	cluster	cluster score				$DC_{A,i}$	$DC_{B,i}$
			$v = L_1$	$v = L_2$	$v = L_3$	$v = L_4$		
TV level (V_1)	1.95	C_1	0.00	0.00	1.00	0.50	0.626	0.975
		C_2	0.72	1.00	0.50	0.10	4.065	0.195
		C_3	0.00	0.10	1.00	1.00	0.000	1.950
		C_4	1.00	0.90	0.00	0.00	5.080	0.000
refrigerator level (V_2)	3.13	C_1	0.00	0.20	0.70	1.00		
		C_2	1.00	0.85	0.10	0.00		
		C_3	0.00	0.00	1.00	0.90		
		C_4	0.70	1.00	0.00	0.00		

이 된다. 따라서, 냉장고의 구매 Level이 TV의 구매 Level에 비해 고객을 구분하는 능력이 뛰어나다고 할 수 있다.

3.2.3 Segment 할당단계

이 단계는 discriminant rule 설정단계에서 계산된 rule에 따라 기 설정된 고객 group에 모든 고객을 배분하는 단계이다. 고객 별로 얻어진 변수들에 대한 cluster score와 variable impact의 가중 평균을 취하여 그 값이 가장 큰 cluster에 할당한다. 즉, $V_k = v_k, k = 1, \dots, K$ 인 변수값을 갖는 고객 j 가 cluster C_i 에 포함될 정도를 나타내는 판별함수(discriminant function) DC_{ji} 는

$$DC_{ji} = \sum_{k=1}^K SPD_{C_i}(V_k = v_k) \cdot VI(V_k), i = 1, \dots, n \quad (5)$$

가 된다. 위의 식 (5)를 이용하여 계산된 판별함수 값이 가장 큰 cluster에 고객을 배분하면 되며, 만일 고객 j 의 V_k 값이 누락되었다면, 해당하는 변수의 cluster score를 0으로 하여 계산하면 된다. 따라서, 어느 고객이 하나의 변수에 대해서만이라도 값을 갖게 되면 기 설정된 cluster 중 한 곳에는 반드시 배분될 수 있다. 즉, 결측치가 많은 데이터라 할지라도 고객분류를 용이하게 할 수 있는 장점이 있다.

예를 들어 TV와 냉장고 구매 level을 활용하여 고객을 4개의 cluster에 분배하고자 하며, 이때 변수별 cluster score와 variable impact는 <Table 5>와 같다고 하자. 만일 고객 A 는 TV level 1,

냉장고 level 2를 구매하였고 고객 B는 TV level 4만을 구매하였다고 하면, cluster 1에 대한 판별함수는 $DC_{A,1} = 0.00 \cdot 1.95 + 0.20 \cdot 3.13 = 0.626$ 되고 고객 B는 냉장고를 구매하지 않았으므로 $SPDC_A(V_2 = v) = 0$ 이 되어 $DC_{B,1} = 0.50 \cdot 1.95 + 0.00 \cdot 3.13 = 0.975$ 가 된다. 같은 방법으로 모든 cluster에 대해 판별함수 값을 구하면 <Table 5>의 $DC_{A,i}$ 와 $DC_{B,i}$ 와 같다. <Table 5>에서 고객 A는 cluster 4에, 고객 B는 cluster 3에 배분된다.

3.3 성능평가

이 절에서는 관측치에 결측값이 존재하지 않는 경우와 존재하는 경우에 대해 제안된 방법을 최우추정(MLE)법과 C4.5 알고리즘을 사용하는 의사결정나무(decision tree) 방법 및 비교·평가한다. 많은 경우에 적용이 가능하고 MLE를 쉽게 얻을 수 있는 정규확률변수를 이용하며, 비교를 용이하게 하기 위해 2변량자료를 사용한다. 또한, 모집단을 3개의 서로 다른 cluster로 구성하고, 다음과 같은 2종류의 data set을 사용한다.

Data Set 1:

$$Cluster 1 \sim BVN\left(\begin{bmatrix} -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

$$Cluster 2 \sim BVN\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

$$Cluster 3 \sim BVN\left(\begin{bmatrix} 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$

Data Set 2:

$$Cluster 1 \sim BVN\left(\begin{bmatrix} -3 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 2 \end{bmatrix}\right),$$

$$Cluster 2 \sim BVN\left(\begin{bmatrix} 3 \\ 3 \end{bmatrix}, \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right),$$

$$Cluster 3 \sim BVN\left(\begin{bmatrix} 0 \\ -3 \end{bmatrix}, \begin{bmatrix} 2 & 0.5 \\ 0.5 & 1 \end{bmatrix}\right)$$

각 group으로부터 MLE의 계산 및 제안된 discriminant rule을 설정하는 데 충분한 양인 100개씩의 관측치로 구성된 training set을 생성하였으며, 분류정확도를 평가하기 위해 50개씩의 관측치로 구성된 test set을 생성하여 사용하였다. 또한, 결측값의 비율에 따른 방법의 적정성을 평가하기 위해 각 변수에서

20%, 40%씩의 결측값을 임의로 생성하였으며, 두 변수 모두 결측값을 갖는 경우는 제외하였다.

<Table 6>은 결측치의 비율에 따른 정분류율을 보여주는 것으로, 이로부터 다음과 같은 결과를 얻을 수 있다.

- i) 결측치가 존재하지 않는 경우 data set 1에서는 제안된 방법의 성능이 우수하고, data set 2에서는 의사결정나무 방법이 우수하다. 즉, cluster가 정확하게 나누어진 경우, 제안하는 PD 방법은 결측치가 존재하지 않는 경우에도 우수한 성능을 나타낸다.
- ii) 결측치가 증가함에 따라 제안된 방법과 MLE를 사용하는 방법 모두 성능이 감소하는 경향을 보인다. 그러나 결측치가 많이 존재할수록 제안된 방법의 성능이 더 우수하게 됨을 알 수 있다.

4. 가전업에서 Cross-Selling을 위한 고객분류 사례

4.1 고객분류의 목적

A사는 대 고객 캠페인 활동을 수행하기 위해 주요 제품에 대한 구매확률 모델을 개발하여 사용하여 왔다. 그러나 각 제품에 대한 구매확률을 구하는 데 필요한 고객 profile 자료에 결측치가 상당수 존재하여 좋은 결과를 얻을 수가 없었다. 따라서, 고객의 구매성향을 중심으로 집단을 세분화하여 해당 집단의 특성을 파악한 후 집단별로 캠페인 전략을 세우기로 하였다.

4.2 Clustering 단계 및 변수선정

비교적 많은 고객이 구매한 대표적인 디지털 가전제품인 TV와 대표적인 생활가전제품인 냉장고 및 세탁기의 구매성향을 조합하면 전자제품에 대한 전체적인 구매성향의 파악이 가능할 것이고, 세 가지 제품을 구매한 고객은 A사에 대한 로열티가 높은 고객이라고 할 수 있으므로 TV, 냉장고, 세탁기를 모두 구매한 고객들을 활용하여 clustering을 실시하였다. 이 때, 사용된 고객의 수는 8만 명 정도로, 분석대상이 된 고객의 3.4% 정도가 되었다.

clustering을 위해 구매한 제품의 상대적인 가격을 변수화하여 활용하였으며 동일한 제품을 여러 대 구입한 경우 상대적인

Table 6. Accuracy of the proposed method and the decision tree method with MLE

data set	missing rate of each variable					
	0%		20%		40%	
	PD	DT(MLE)	PD	DT(MLE)	PD	DT(MLE)
data set 1	0.96	0.93	0.93	0.81	0.90	0.69
data set 2	0.93	0.96	0.89	0.81	0.79	0.67

PD: proposed method, DT(MLE): decision tree method with MLE

가격이 가장 높은 값을 사용하였다. 제품의 가격변동내역을 조사한 결과 <Figure 3>의 (a)와 같이 제품 모델에 따라 가격이 지수적으로 증가하는 경우(TV)와 (b)와 같이 선형적으로 증가하는 경우(냉장고, 세탁기)로 나누어지는 것을 알 수 있었다. 제품의 상대적인 가격을 변수로 하여 SAS E-miner의 K-means clustering 방법으로 임시적인 clustering을 실시하면 <Figure 4>와 같이 총 12개의 cluster로 나눌 수 있다. 이때, clustering 결과를 분석하면 TV, 냉장고, 세탁기의 상대적인 가격이 모두 4개 정도의 그룹으로 나누어질 수 있음을 알 수 있다. 따라서, 이후 discriminant 단계에서 제안하는 방법을 활용하기 위해 변수를 이산화하는 경우 지수적으로 가격이 증가하는 TV는 최고 가격부터 5%, 20%, 25%, 50%의 비율로 4개의 level(High Premium, Premium, General, Basic)로 이산화하여 사용하고 선형적으로 증가하는 냉장고 및 세탁기의 경우는 15%, 25%, 30%, 30%의 비율로 나누어 사용하는 것이 타당하다고 할 수 있다. 이러한 방

법으로 이산화된 변수를 활용하여 다시 clustering을 실시하면 <Table 7>의 clustering 결과를 얻을 수 있다. <Table 7>은 각 cluster에 대한 TV, 냉장고, 세탁기 level의 분포를 보여준다.

discriminant 분석에서 사용될 변수를 선별하기 위해, clustering에 사용한 고객들을 대상으로 <Table 1>의 변수와 고객의 cluster를 활용하여 탐색적 데이터 분석(exploratory data analysis)을 실시한 결과 <Table 8>과 같은 변수들이 분석을 위한 적당한 변수로 선정되었으며, 제품보유 및 서비스 이력은 활용이 불가능한 것으로 나타났다.

4.3 Discriminant Rule 설정단계

변수의 Cluster별 분포 및 Cluster Score 계산

9개의 각 cluster별로 분포하는 고객의 비율과 각 변수의 분포 및 cluster score를 계산한 결과, clustering에 사용되지 않은 변수

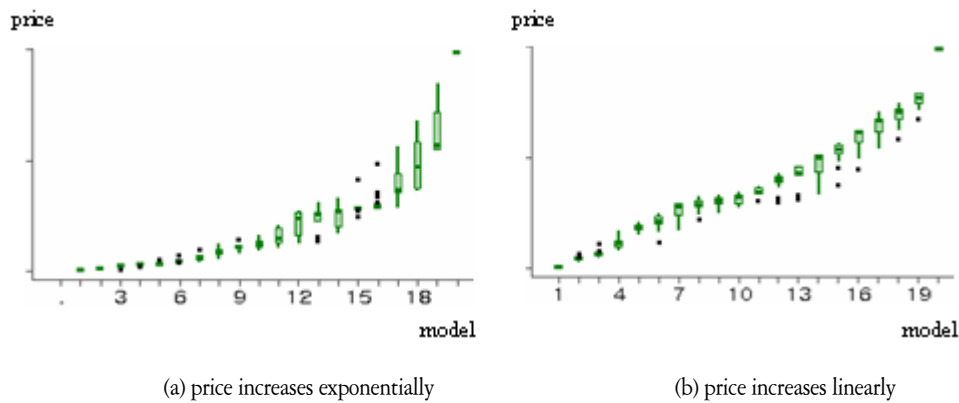


Figure 3. Product price pattern.

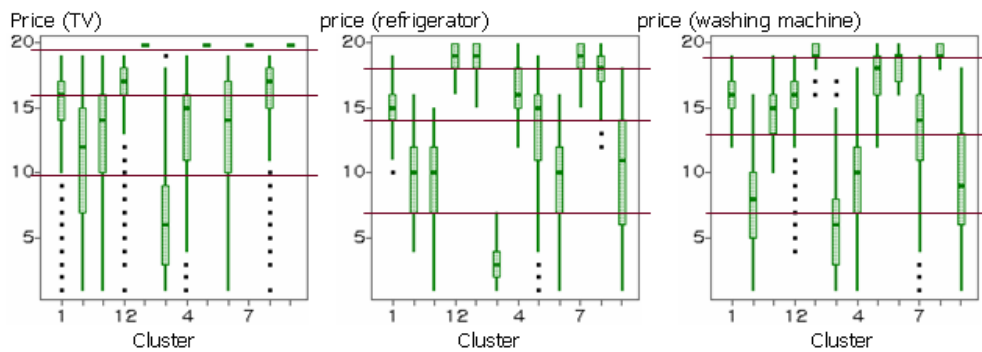


Figure 4. Result of preliminary clustering.

Table 7. Result of clustering and distribution of the level of TV, refrigerator and washing machine

cluster	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉
level of TV	HP/P	P	HP/P	G/B	G/B	HP/P	HP/P	G/B	G/B
level of refrigerator	HP/P	P	G/B	P/G	G/B	HP/P	G/B	HP/P	G/B
level of washing machine	HP/P	HP/P	G/B	G/B	HP/P	G/B	HP/P	HP/P	G/B

HP: High Premium, P: Premium, G: General, B: Basic

의 경우에도 cluster score가 상이하였으며 이를 활용하여 결측 값이 존재하는 고객도 각 cluster에 할당할 수 있을 것으로 분석되었다. <Table 9>는 TV 구매 level 및 소득 정도를 추정한 변수의 cluster score를 나타낸 것으로, 소득 정도가 높을수록 - 4의 값을 갖을수록 소득 정도가 높음 - cluster 1에 속할 가능성이 높아지며 소득 정도가 낮을수록 cluster 4, 5에 속할 가능성이 높아지는 것을 알 수 있다.

Variable Impact 계산

<Table 8>의 각 변수에 대한 variable impact를 구하면 <Table 10>과 같다. 여기서 고가의 대형제품에 대한 구매 level이 고객을 분류하는 주요 기준임을 알 수 있고, 이후의 기준으로는 비교적 저가의 소형가전, 제품의 구매이력 및 고객 profile임을 알 수 있다. 또한, 대형 가전의 경우에는 디지털 가전보다는 생활

가전의 구매수준이 고객을 나누는 데 더욱 중요한 역할을 하는 것을 알 수 있다.

4.4 Segment 할당단계

계산된 cluster score와 <Table 10>의 variable impact를 이용하여 230만여 명의 분석대상 고객을 9개의 segment에 각각 할당하였다. 할당한 결과, segment 1에는 11.7%, segment 2에는 12.0%, segment 3에는 6.3%, segment 4에는 12.5%, segment 5에는 9.0%, segment 6에는 4.4%, segment 7에는 9.7%, segment 8에는 6.18%, segment 9에는 28.2%의 고객이 할당되었다. 또한 각 segment의 특성을 분석한 결과 <Table 11>과 같은 이름의 부여가 가능하며, segment별로 상이한 특성을 갖고 있음이 확인되었다.

Table 8. Result of variable selection

class	variables
customer profile	age income level: a function of estimated income, price of house, and living space(principal component analysis is used.)
purchasing record	levels of refrigerator, washing machine, TV, gas range, vacuum cleaner, VTR, Kimchi refrigerator, air conditioner, rice-cooker, and microwave oven purchasing home theater or dishwasher the number of purchasing products time elapsed after last purchasing active time: time elapsed after first purchasing - time elapsed after last purchasing

Table 9. Cluster score according to TV level and income

cluster	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	
TV level	1	0.00	0.00	0.00	0.71	0.67	0.00	0.00	0.83	1.00
	2	0.00	0.72	0.00	0.95	1.00	0.00	0.00	0.81	0.61
	3	0.72	0.65	1.00	0.00	0.00	0.95	0.99	0.00	0.00
	4	1.00	0.00	0.24	0.00	0.00	0.38	0.28	0.00	0.00
income	1	0.00	0.28	0.52	1.00	0.95	0.55	0.64	0.86	0.88
	2	0.00	0.55	0.86	1.00	0.75	0.63	0.56	0.65	0.70
	3	1.00	0.50	0.25	0.00	0.20	0.44	0.50	0.33	0.10
	4	1.00	0.50	0.19	0.00	0.25	0.21	0.64	0.55	0.13

Table 10. Variable impact

variables (V_k)	$VI(V_k)$	variables (V_k)	$VI(V_k)$	variables (V_k)	$VI(V_k)$
refrigerator	3.33	dishwasher	3.24	home theater	3.00
washing machine	2.73	TV	2.40	gas range	1.23
vacuum cleaner	0.96	VTR	0.88	Kimchi refrigerator	0.87
air conditioner	0.86	rice cooker	0.84	microwave oven	0.69
time elapsed after last purchasing	0.63	age	0.53	active time	0.46
the number of purchasing products	0.38	income level	0.29		

제안된 방법의 분류정확도를 분석하기 위해 clustering 단계에서 사용된 고객을 활용하여 모의실험을 실시하였다. TV, 냉장고, 세탁기를 모두 구매한 고객을 8등분하여 TV/냉장고/세탁기를 모두 구매한 고객군, TV만 구매한 고객군, 냉장고만 구매한 고객군, 세탁기만 구매한 고객군, TV/냉장고 구매 고객군, TV/세탁기 구매 고객군, 냉장고/세탁기 구매 고객군, 모두 구매하지 않은 고객군으로 랜덤하게 나눈 후, 각 고객을 9개의 segment에 랜덤하게 배분하는 것에 비해 제안된 방법이 어느 정도의 성과를 보이는가를 분석한 결과 <Table 12>의 분류행렬을 얻을 수 있었다. <Table 12>의 각 행은 clustering 단계에서 얻어진 실제 cluster를 나타내는 것이고 각 열은 제안하는 방법을 통해 할당된 segment를 나타내며 (i, j) 의 행렬값은 ‘ C_j 로 할당된 고객수/ C_j 에 속한 고객수’를 의미한다. 따라서, $i=j$

인 경우 1에 가까운 값을 갖고 $i \neq j$ 인 경우 0에 가까운 값을 갖을수록 분류정확도가 높다고 할 수 있다. 제안된 방법의 정분류율은 약 50%로 고객을 cluster에 무작위하게 배분한 것에 비해 lift 4.4를 얻을 수 있다. 즉, 제안된 방법은 결측치가 많은 경우에도 만족할 만한 분류성적을 나타내는 것을 알 수 있다.

5. 결론

이 논문에서는 구매가 빈번하게 발생하지 않으며 고객 기본정보에 대한 결측치가 상당수 존재하는 가전업의 데이터를 분석하여 고객을 분류하는 발견적 방법을 제안하고 이를 적용한 실제 사례를 소개하였다. 제안한 방법은 PD(percent of difference)

Table 11. Segment naming and characteristics

segment	naming	segment characteristics
1	High premium customers	Customers in the 40th and 50th with high income level purchasing expensive goods
2	Household electrical appliances focused high premium customers	Relatively frequent purchasers with a preference for high standard household electrical appliances compared to other electrical equipments
3	Digital electrical equipment focused customers	Relatively frequent purchasers with a preference for high standard TVs compared to other electrical equipments
4	Occasional buyer(1)	Sporadic purchasers of relative large and expensive electrical equipment
5	Occasional buyer(2)	Sporadic purchasers buying expensive washing machines
6	Loyal small electrical equipment buyer	Frequent purchasers of small electrical equipment such as rice cooker, vacuum cleaner, microwave oven, etc
7	Joint breadwinning pattern customers	Mostly customers in the 30th purchasing standard electrical equipment with a preference for comfortable and convenient goods like TV and washing machine
8	Household electrical equipment focused housewifely customers	Customers in the 40th and 50th living mainly in provincial areas
9	Basics focused customers	Purchasing mainly inexpensive goods with average frequency

Table 12. Simulation result

		segments assigned with the proposed method								
		C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9
pre-determined segments from clustering step	C_1	.65	.15	.02	.00	.01	.05	.09	.03	.00
	C_2	.07	.69	.02	.02	.04	.04	.10	.02	.00
	C_3	.03	.04	.36	.06	.01	.17	.22	.02	.10
	C_4	.01	.20	.05	.36	.07	.11	.06	.06	.09
	C_5	.03	.13	.02	.04	.38	.01	.20	.07	.13
	C_6	.13	.21	.07	.06	.00	.43	.06	.02	.03
	C_7	.10	.10	.06	.00	.05	.03	.61	.02	.03
	C_8	.20	.24	.01	.04	.08	.06	.02	.30	.06
	C_9	.01	.04	.10	.14	.07	.06	.05	.03	.50

를 이용하여 간단한 수치적인 계산으로 판별분석이 가능하도록 고안된 것으로, 직관적으로 이해하기가 용이하고 계산이 간단하여 현장에서 복잡한 통계적 분석방법에 대한 지식 없이도 결측치가 상당수 존재하는 데이터의 처리에 사용될 수 있을 것으로 판단된다. 이 논문에서는 변수가 이산형인 경우에 대해서만 다루었으나, 향후 연속형 변수의 값을 이산화시키지 않고 사용할 수 있는 방법에 대한 연구도 필요할 것으로 여겨진다.

참고문헌

- Addison, W.(1996), *Data Mining*, Syllogic, USA.
- Affi, A. and Elashoff, R.(1966), Missing Observations in Multivariate Statistics I: Review of the Literature, *Journal of American Statistical Association*, 61, 595-604.
- Athanassopoulos, A. D.(2000), Customer Satisfaction Cues to Support Market Segmentation and Explain Switching Behavior, *Journal of Business Research*, 47(3), 191-207.
- Bart, B., Verstraeten, G., Poel, D.V.D., Egmont-Petersen, M., Kenhove, P.V. and Vanthienen, J.(2004), Bayesian Network Classifiers for Identifying the Slope of the Customer Lifecycle of Long-lift Customers, *European Journal of Operational Research*, 156(2), 508-523.
- Berry, M. J. A. and Linoff, G.(1997), *Data Mining Techniques*, John Wiley & Sons, NY, USA.
- Dillon, W. R. and Goldstein, M.(1984), *Multivariate Analysis*, John Wiley & Sons, NY, USA.
- Frederick, N.(2000), *Loyalty.com*, McGraw-Hill, USA.
- Granger, E., Rubin, M. A., Grossberg, S. and Lavoie, P.(2001). Classification of Incomplete Data Using the Fuzzy ARTMAP Neural Network, *Proc. IEEE International Joint Conference of Neural Networks*, 6, 35-40.
- Huang, X. and Zhu, Q.(2002), A Pseudo-nearest-neighbor Approach for Missing Data Recovery on Gaussian Random Data Sets, *Pattern Recognition Letters*, 23, 1613-1622.
- Krishnamoorthy, K. and Maruthy, K.(1998), Some Simple Test Procedures for Normal Mean Vector with Incomplete Data, *Annals of the Institute of Statistical Mathematics*, 50(3), 531-542.
- Lim, C. P., Leong, J. H. and Kuan, M. M.(2005), A Hybrid Neural Network System for Pattern Classification Tasks with Missing Features, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), 648-653.
- Mozer, M. C., Wolniewicz, R., Grimes, D. B., Johnson, E. and Kaushansky, H.(2000), Predicting Subscriber Dissatisfaction and Improving Retention in the Wireless Telecommunications Industry, *IEEE Transactions on Neural Networks*, 11(3), 690-696.
- Nijman, M. J. and Kappen, H. J.(1997), Symmetry Breaking and Training from Incomplete Data with Radial Basis Boltzmann Machines, *International Journal of Neural Systems*, 8, 301-306.
- Punj, G. and Stewart, D. W.(1983), Cluster analysis in Marketing Research: Review and Suggestions for Application, *Journal of Marketing Research*, 20(2), 134-148.
- Schafer, J. L.(1997), *Analysis of Incomplete Multivariate Data*, Chapman & Hall, London, UK.
- Van den Poel, D. and Lariviere, B.(2004), Customer Attrition Analysis for Financial Services Using Proportional Hazard Models, *European Journal of Operational Research*, 157(1), 196-217.
- Weerahandi, S. and Morita, S.(1995), Using Survey Data to Predict Adoption Switching for Services, *Journal of Marketing Research*, 32(1), 85-96.
- Zeithaml, V. A., Berry, L. L. and Parasuraman, A.(1996), The Behavioral Consequences of Service Quality, *Journal of Marketing*, 60(2), 2105-2111.



장영순

연세대학교 응용통계학 학사
한국과학기술원 산업공학 석사
한국과학기술원 산업공학 박사
현재: 명지대학교 경영학과 조교수
관심분야: 품질경영, 생산관리, CRM



서종현

한국과학기술원 산업공학 학사
한국과학기술원 산업공학 석사
한국과학기술원 산업공학 박사
현재: 중소기업진흥공단 전문위원
관심분야: 신뢰성공학, CRM