
통신 산업의 고객 분류를 위한 예측 모델 설계

Design of a Forecasting Model for Customer Classification in the Telecommunication Industries

이병엽*, 조규하**, 송석일***, 유재수**

배재대학교 전자상거래학과*, 충주대학교 전기전자정보공학부**, 충북대학교 전기전자컴퓨터공학부**

Byoung-Yup Lee(bylee@pcu.ac.kr)*, Kyu-Ha Joh(khjoh@einssnc.co.kr)**,
Seok-Il Song(sisong@chungju.ac.kr)***, Jae-Soo Yoo(yjs@chungbuk.ac.kr)**

요약

최근 데이터 수집 및 저장기술의 발달, 데이터베이스 관리시스템과 데이터웨어하우스 기술의 광범위한 사용은 기업내부의 대량의 데이터를 축적할 수 있도록 하고 있으며, 축적된 데이터는 의사결정에 필요한 새롭고 가치 있는 정보와 지식을 획득할 수 있는 잠재적인 원천으로 인정되고 있다. 본 논문에서는 이동통신업체의 데이터를 가지고 데이터 마이닝 방법론을 이용하여 기존고객을 세분화하기 위한 예측모델을 설계한다. 이를 통해 고객 개개인의 특성에 맞는 마케팅 프로모션을 하게 하고 신규고객을 획득할 때는 신규 고객의 특성을 미리 예측하여 세분화함으로써 고객의 평생가치를 촉진하여 기업과 고객과의 관계성을 높여서 기업은 안정된 고객층으로부터 수익을 창출하고 고객들은 해당 기업으로부터 더 많은 혜택을 받게 하는데 목적이 있다.

■ 중심어 : | 데이터 마이닝 | 신경망 | 회귀분석 | 고객분류 | 예측모델 |

Abstract

Recently, according to the development of computer technology, a large amount of customer data have been stored in database. Using such data, decision makers extract the useful information to make a valuable plan with data mining. In this paper, we design a forecasting model that classifies the exiting customers in the telecommunication industries using the classification rule, one of the data mining technologies. In other words, this paper builds a model of customer loyalty detection and analyzes customer patterns in mobile communication service market with data mining using neural network and regression methods. This model improves the relationship of customers and enterprises. As a result, the enterprise creates the profits from many customers and the customer receives more benefits from the enterprise.

■ Keyword : | Data Mining | Neural Network | Regression | Customer Classification | Forecasting Model |

* 이 논문은 산업자원부의 지역혁신 인력양성사업의 연구결과로 수행되었습니다

I. 서론

기업들은 급변하는 경영환경에서 기업의 경쟁력을 강화하기 위해서 축적된 데이터를 분석하고 정보와 지식을 획득할 수 있는 능력과 정보기술을 보유해야 한다[8][16]. 이동 통신 업체가 늘어나고 기업간 보유고객의 수가 늘어남에 따라 고객 데이터와 정보의 양도 기하급수적인 증가를 보여왔으며 이러한 데이터와 정보의 증가는 데이터 활용에 대한 방법과 그 영역의 증대를 가져왔다[4][13]. 하지만 이동전화로 주로 사용하는 15-59세 연령층이 약 3,200만 명에 해당하고 이들 인구의 81%에 해당하는 가입자를 이미 확보했다는 사실을 고려하면 향후 신규 가입자의 추가적인 폭증은 현실적으로 어려운 실정이라 할 수 있다[17]. 따라서 이동 통신의 경쟁방식은 기존고객의 유지, 서비스의 차별화와 고도화를 통해 이루어져야 하며 신규고객 1명 유치에 드는 비용은 기존고객 1명을 유지하는데 드는 비용보다 4-6배 더 들게 되는 것을 감안한다면 양적 성장인 신규고객 유치보다는 보유고객에 대한 다각적인 분석에 따른 효율적인 마케팅 방법 적용으로 심화되는 이동 통신 업체간의 경쟁에서 우위를 차지할 수 있을 것이다[8][16].

본 논문은 국내 이동 통신 업체인 A사의 자료를 입수하여 실제 데이터를 중심으로 연구하였다. 또한 SAS를 이용하여 고객 데이터에서 변수 추출하여 정제작업을 실시하였다. 데이터 마이닝은 SAS의 데이터 마이닝 툴인 엔터프라이즈 마이너(Enterprise-Miner)를 이용하여 분류에 의한 고객그룹속성 및 고객 그룹에 대한 분류 예측 모형을 제시하였다.

본 논문의 구성은, I장에서는 연구 목적, 연구 범위와 방법을 제시하였고, II장에서는 데이터 마이닝의 개념과 기법 및 데이터 마이닝의 CRM 활용 분야 등에 대해 살펴보았다. III장에서는 통신사의 사례 연구 및 본 연구에서 사용한 데이터 마이닝 기법들에 대한 세부적인 내용을 알아봤으며 IV장에서는 연구모형의 실험결과를 살펴보고 의미를 분석하였다. V장에서는 연구결과를 종합하고 연구의 한계점을 밝히고 미래의 연구방향을 제시하였다.

II. 관련 연구

1. 데이터 마이닝

최근 수년동안 데이터 마이닝에 대한 많은 연구가 이루어졌다. 그 동안 제안된 다양한 데이터 마이닝 기법들은 어떤 형태의 지식을 발견하고자 하는가, 어떤 종류의 데이터베이스에 적용될 수 있는가, 어떤 분야의 기술에 바탕을 두고 있는가 등의 기준에 의하여 데이터 집합의 일반적 특성을 분석하고 있다[8][15]. 특히 대용량의 자료로부터 숨어 있는 정보와 유용한 패턴을 찾아내는 것을 의미하는 데이터 마이닝은 현재 그 원래의 의미에 데이터 분석이라는 광의의 의미가 좀더 첨가되어 사용되어 지고 있다.

데이터 마이닝의 활용분야는 매우 다양하며 현재에도 그 활용 분야가 점차 늘어나고 있다. 그 중에서 특히 기업의 의사결정과 향후 대처에 대한 문제에서 많이 활용되고 있으며 주요 활용 분야 및 적용사례를 요약하면 다음과 같다[10][12][표 1].

• 데이터베이스 마케팅

데이터 마이닝이 가장 성공적으로 적용되고 있는 분야로 소매, 통신판매, 금융서비스, 보험, 통신, 운송 등 다양한 분야에서 활용되고 있다[3]. 분석방법으로는 목표 마케팅, 고객 세분화, 고객성향변동분석, 교차판매, 시장바구니분석 등이 있다.

• 신용평가

신용평가는 특정대상의 신용거래 대출한도를 결정하는 것으로 불량채권이나 기타의 금융문제에서 손실을 최소화하기 위한 것으로, 신용평가의 중요한 사안은 현재의 대출시스템을 그대로 유지하면서 불량채권이나 금융사고에 대한 최선의 대응책을 결정하는 것이다. 이러한 신용평가와 관리는 은행, 금융서비스, 보험 등의 분야에서 많이 적용되고 있다.

• 품질개선

품질개선에서는 불량품을 찾고, 그 원인을 파악하여 궁극적으로 문제를 예방하는데 중점을 두고 있다. 제조

업체에서는 불량품의 감소를 통한 이윤의 증가에 중점을 두고 있으며, 병원과 의료보험조합에서는 의료사고, 불필요한 장기입원, 의료비 과다청구 등에 초점을 맞추고 있다.

월별 판매량의 변동 등을 파악하고, 모형화하여 미래 값을 예측하는 것을 목적으로 한다. 이 외에도 위험관리, 망 관리 등 다양한 분야에서 활용되고 있다. 다음 [표 1]은 는 데이터 마이닝의 응용분야 및 적용사례를 나타내고 있다[5].

표 1. 데이터 마이닝의 응용분야 및 적용사례

분야	적용사례
소매/마케팅	고객의 구매패턴과 선호도 발견 DM(Direct Mail)응답 가능성 높은 고객 예측 판매 실적에 영향을 미치는 요소 고객 분류, 그룹별 특성 발견 광고, 프로모션, 이벤트의 효과측정
은행/카드	신용카드 도용 패턴 추적 이탈 예상고객 선정 및 특성분석 우수고객 선정 및 특성 분석 신용평가 모형 개발 주식 거래규칙 발견
보험	고객분류를 통한 보험료 정책 수립 보험료 청구 사기 패턴 추적 클레임 처리시간에 영향을 미치는 요소발견
통신	장거리 전화/무선전화의 부정합 이용패턴추적 이탈 예상고객 선정 및 특성 발견 서비스간의 연관관계 발견 우수고객 선정 및 특성 분석
제조	생산품 품질에 영향을 미치는 요인 발견 경쟁사의 입찰액 예측 제품의 수요 예측 대리점 여신 평가 모형 개발
유통	매장진열 전략 수립 상품 카탈로그 디자인 상품 교차판매
의료	질병 진단이나 질병의 예후 발견 환자의 특성에 따른 약약품의 부작용 분석

• 부정행위적발

부정행위적발의 목적은 고도의 사기행위를 발견할 수 있는 패턴을 알아내는 것이다. 은행에서는 발견된 패턴을 이용하여 신용카드 거래 사기 및 불량수표 등을 적발할 수 있으며, 통신회사에서는 전화카드거래 사기를 방지하며, 보험회사에서는 허위 및 과다청구를 예방할 수 있다.

• 이미지분석

이미지분석은 디지털화된 사진으로부터 패턴을 추출하는 기법으로 천문학, 문자인식, 의료진단, 방위산업 등 다양한 분야에서 활용되고 있다.

• 수요 및 판매예측

수요 및 판매예측에서는 과거로부터 현재에 이르는 판매관련 자료로부터 일정한 소비패턴이나 계절별 또는

2. 데이터 마이닝 기법

데이터 마이닝의 기법에는 일반적으로 통계학에서 사용되는 여러 분석기법들을 포함하여 연관규칙 (Association), 클러스터링(Clustering), 의사결정 트리 (Decision Trees), 그리고 신경망(Neural Network)과 같은 기법들이 있다. 본장에서는 분석모형에 사용된 인공신경망 기법과 로지스틱 회귀분석에 대해 기본적인 데이터 마이닝 기법을 요약 하였다[1].

(1) 인공신경망(Artificial Neural Networks)

인공 신경망에 관한 연구는 인간의 두뇌와 신경세포 모델에 관한 연구에서 시작되었으며 신경 시스템의 가장 기본적인 단위인 뉴런(neuron)의 정보전달과정을 모방하고 있다. 인공신경망은 연결가중치(weighted connection)를 사용하여 서로 상호 작용하는 뉴런 또는 노드라고 부르는 단순 처리요소들로 구성된 다이나믹 시스템이다. 일반적으로 인공신경망은 입력층, 출력층, 하나 이상의 은닉층의 계층구조를 갖는다 (Multilayer perceptron 구조). 은닉층의 수에 대한 이론적인 제한은 없으며 대부분의 문제를 푸는 데는 한 개의 은닉층이면 충분한 것으로 알려져 있다. 이들 각각의 층은 특정치의 연결강도 또는 가중치(weight)를 갖는 링크(link)로 연결되어 있고, 각각의 노드는 전 단계의 출력값을 입력값으로 받아 특정 전이함수(transfer function)에 의해 출력값을 생성하게 되는데 이 값은 다시 다음 단계의 입력값으로 적용하게 된다[10].

인공 신경망 모형은 통계학적 모형과는 달리 그 자체의 귀납적 특성으로 인해 모형을 도출하기 위한 이론수립 과정을 생략할 수 있으며, 통계학적 모형에서 요구하는 엄격한 가정(strict assumption)에 전혀 제한을 받지 않는다. 기존의 통계학적 방법론들은 엄격한 가정을 요구하기 때문에 그 사용범위에 제한을 받거나, 가정을 위

배한 경우에는 그 결과에 대한 신뢰성이나 사용기법의 적정성 등이 문제 시 된다[7]. 반면 인간의 뇌의 신경조직을 모방하여 수학적으로 모형화한 신경망(neural network)은 고도의 병렬분산처리를 하는 모형으로서, 이러한 통계학적 가정에 전혀 구애받지 않으며 통계학적 가정을 만족하지 못하는 자료의 분석에 적합한 비모수·비선형 문제이다.

(2) 로지스틱 회귀분석

본 논문에서 연구모형 성과를 비교하기 위해 사용할 통계적인 기법으로 단지 2개의 값(예: 우수, 비 우수)만을 가지는 종속변수와 독립변수간의 인과관계를 추정하는 하나의 통계기법이다[1]. X가 주어질 때 기대값 Y는 특별한 의미를 갖는다. 이것은 확률값으로 대출심사시 대출결정(성공)일 경우 1을 갖는다. 로지스틱 회귀식은 선형이 아닌 S 형태의 곡선을 이루고 있다. 즉 독립변수와 종속변수와의 관계는 비 선형적이다. 확률추정은 Z 값에 상관없이 항상 0과 1사이에 존재한다[9].

3. 통신 산업에서의 CRM

CRM은 고객전략을 정의하고, 경영자원의 분배를 최적화하며, 고객과 관련된 모든 부분에서 최상의 서비스를 제공하기 위해, 고객의 욕구와 행동 그리고 고객의 수익성을 규명해 가는, 지속적이고 자동화된 프로세스로서 CRM은 고객과 관련된 기업의 내외부 자료를 분석하고 통합하여 고객특성에 기초한 마케팅활동을 계획하고, 지원하며, 평가하는 과정이다. 또한 CRM은 데이터 마이닝을 통해 고객데이터의 세분화를 실시하여 신규고객획득, 우수고객 유지, 고객가치증진, 잠재고객 활성화, 평생 고객화와 같은 사이클을 통하여 고객을 적극적으로 관리하고 유도하며 고객의 가치를 극대화 시킬 수 있는 전략을 통하여 마케팅을 실시한다. 이러한 CRM의 정의로 볼 때 CRM에서의 데이터 마이닝은 필수 불가결한 도구라고 볼 수 있다[5][15].

2000년대에 접어들어 비즈니스 세계에서, 고객중심의 시각에서 모든 것을 바라보며, 그것에 맞추어 경영 전략을 수립하는 일들이 이제는 거의 상식의 관점에서 이루어지고 있다. 이러한 관점에서 비즈니스의 당위 목적들

이 제품 또는 서비스의 단순한 판매에만 있는 것이 아니라 고객과 관계를 맺어가는 과정임을 인식하는 것이라고 볼 수 있다.

통신 산업에서 데이터 마이닝과 CRM의 활용을 요약하면 다음과 같다. 첫째, 고객 세분화와 고객가치의 이해 증대로 더 많은 고객의 획득이 예상된다. 특정한 고객집단의 세분화와 가치의 극대화는 캠페인의 목표가 될 수 있다. 둘째, 고객 세분화와 상품 구매간의 관계에 대한 더 나은 이해로 교차 판매의 효율이 증대 될 수 있다. 셋째, 언제 그리고 왜 고객들이 이탈을 하는지, 그리고 어떤 경우에 긴급한 대처 방안은 무엇인지에 대한 개념 파악으로 우수고객의 유지율의 향상을 도모 할 수 있다. 이러한 요약된 개념들은 데이터 마이닝 또는 분석적 CRM의 기법에서 얻어진 결과물로서 운영 CRM에 의해 활용 될 수 있다[15].

데이터 마이닝은 각 산업의 특성에 따라 표준화된 분석방법과 적용사례가 있다. 본 논문에서 제시하는 고객 분류 모형은 데이터 마이닝 분석자들이 새로운 모형을 구축하는데 소요되는 노력을 최소화 할 수 있으며, 더 나아가 기존 고객 및 신규 고객의 관리측면에서 마케팅 자료로 활용될 수 있다.

III. 사례분석의 절차

1. 연구 절차와 방법

본 연구에서는 국내 이동 통신 업체인 A사의 자료를 입수하여 연구하였고 수행하는 연구의 진행절차는 데이터 마이닝 프로세스의 각 단계별 과정에 따라 진행을 하고자 하며 과정은 [그림 1]과 같다.

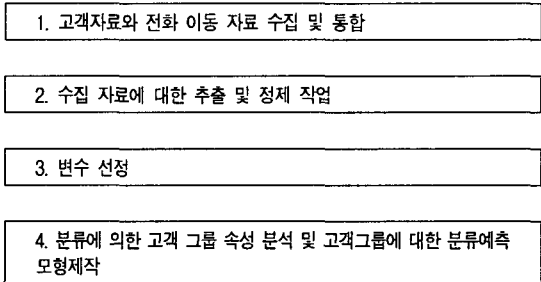


그림 1. 연구진행절차

2. 자료의 구성

본 연구에서 사용한 데이터는 국내 이동통신업체의 자료를 대상으로 한다. 데이터를 이루고 있는 고객의 수는 2000명으로 총 고객에서 샘플링을 하여 선정하였고, 기본 고객데이터와 통화관련 데이터로 나누어진다. 다음 [표 2]는 보유 데이터를 이용하여 산출한 변수를 설명한다.

3. 데이터 마이닝을 위한 변수 선정

데이터 마이닝 알고리즘을 적용하기 위해서는 데이터에 대한 사전처리가 필요하다. 표본의 각 사례의 특정 변수에 누락된 데이터가 있는지 확인하여 누락된 데이터에 대해서는 변수의 평균값, 중위수 혹은 중간값 등으로 결측치를 대체시킨다. 만약 결측치를 지닌 변수가 이산형이라면 해당 변수의 최빈값으로 대체하는 것이 일반적이다. 또한 중복된 레코드를 제거하고 변수의 값이 일관성이 없으면 이를 바로 잡아야 한다. 잠정적으로 선택된 많은 변수 중에서 원하는 모형에 가장 적합한 변수를 선정해야 하는데 선정방법에는 대표적으로 입력변수와 목표변수간의 결정계수를 이용하는 방법과 카이제곱(χ^2) 통계량을 이용하는 방법이 있다[1]. 본 논문에서는 카이제곱(χ^2) 통계량을 이용하였다.

표 2. 산출된 생성 변수

범주	변수	수
인적정보	나이, 성별, 신용등급, 연체횟수	4
이용정보	M to M 발신호수, S to S 발신호수, M to L 발신호수, 일반호수, Network Size, 총 통화시간, 주말 총 통화시간, 주중 총 통화시간, M to M 총 통화시간, S to S 총 통화시간, M to L 총 통화시간, 총 착신호수, 월 통화횟수	13
부가서비스	VMR 발신호수, Call Forward 호수, Ezi 호수	3
요금정보	각 월별 기본요금, 국내요금, 부가서비스료, 정보이용료, 요금합계	5
기타	마일리지 점수	1

3.1 카이제곱(χ^2) 통계량

카이제곱(χ^2) 통계량은 목표변수와 입력변수의 관측도수로 $r \times c$ 교차분석표 또는 상황표로부터 계산된다. 상황표상의 실제발생빈도는 두 변수가 아무 상관없이 있다는 가정 하에서 얻어진 기대빈도를 비교하여, 이 차이가 크면 두 변수간에 어떤 관계가 있음을 의미한다[9]. 본 연구에서 변수선정 문제는 총 26개의 변수에 대해 결정계수를 이용한 입력변수 선택 방법을 사용하여 결정계수의 값이 작은 S to S 발신호수, 6월과 7월의 요금 정보와 8월의 국내요금, 부가 서비스료, 정보이용료, 연체 횟수, 신용등급을 제외한 결정계수의 값이 큰 상위 20개의 변수를 입력변수로 선정하였다. [그림 2]는 SAS Enterprise -Miner를 이용하여 산출한 입력변수와 목표변수에 관한 결정계수 값을 나타낸다.

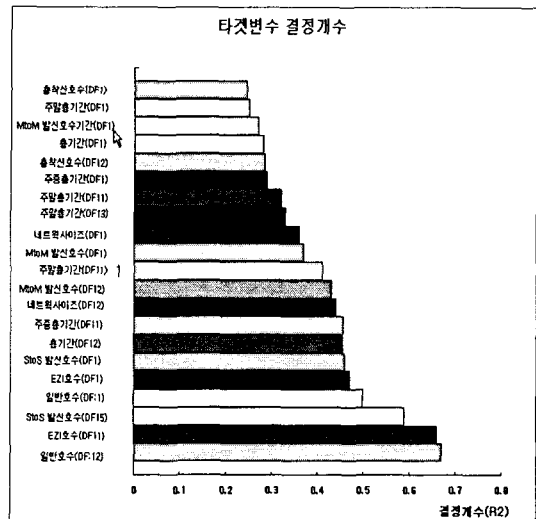


그림 2. 목표변수와 입력 변수간의 결정계수

3.2 선정된 입력변수의 특성

본 논문에서 사용된 입력 변수는 상기 결정계수의 값이 큰 상위 20개 데이터를 선정하였다[표 3]. 선정된 20개의 입력 변수중 주요 입력 변수들의 특성은 다음과 같다.

표 3. 선정된 입력 변수

범주	선정변수	변수설명
인적정보(2)	성별, 나이	-
이용정보(13)	네트워크 사이즈, 인터넷 서비스 사용횟수, M to M 발신횟수, M to L 발신횟수, 일반호수, 총기간, 주중총 기간, 주말총 기간, M to M 총기간, M to L 총기간, 총 통화시간, 총 착신횟수, 착신전환서비스 횟수	- M to M: Mobile 과 Mobile의 사용 - S to S: 동일 번호간 통화수 - M to L: Mobile 과 Local 통화수
요금정보(2)	기본요금, 월 DATA 요금	-
부가정보(2)	EZI 호수, VMR	- EZI: 인터넷서비스 등 부가 서비스 사용횟수 - VMR: 음성사서함
기타(1)	마일리지	-

(1) 성별

이동 통신 이용고객의 성별을 의미하는 변수로 [그림 3]와 같이 총 평균 통화수에 상위 30% 그룹의 평균 즉 월 150회 이상 통화하는 고객그룹(이하 “우수고객”이라 함)의 경우 남성의 비율이 60.57%이고 여성이 39.43% 이고 [그림 4]와 같이 월 150회 이하 통화하는 고객그룹(이하 “비우수고객”이라 함)의 경우 남성의 비율이 57.84%, 여성이 42.16%로 이동 통신 이용비율이 높아짐에 따라 남성의 비중이 커지고 있음을 알 수 있다.

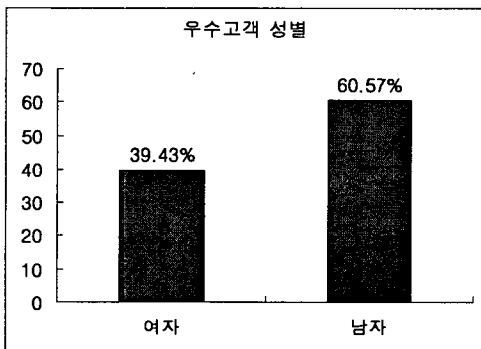


그림 3. 우수고객 그룹의 성별 비율

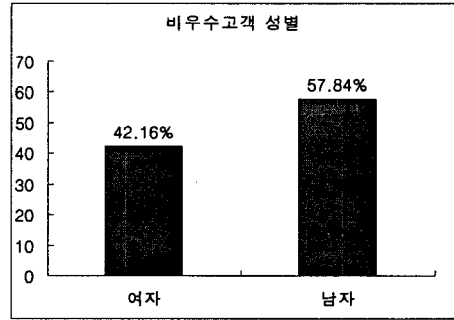


그림 4. 비 우수고객 그룹의 성별 비율

(2) 나이

이동 통신 이용고객의 연령을 나타내는 변수로 연령 그룹별 분류에서 10대 이용자 중 우수고객이 72.25%로 타 그룹보다 이용률이 월등히 높음을 알 수 있고 60대 이상의 고객은 비우수고객이 86.23%로 10대 그룹과는 상반되는 결과를 보여줬다. [표 4]는 이동통신 이용 고객의 연령별 사용비율을 나타낸다.

표 4. 연령별 이동 통신 사용 비율

고객그룹연령	150회 이하	150회 이상
10대	27.75	72.25
20대	64.55	35.45
30대	72.26	27.74
40대	75.86	24.14
50대	76.33	23.67
60대	86.23	13.77

(3) 네트워크 사이즈

이동 통신 사용자의 통화내역 중 동일하지 않은 번호로 통화 횟수를 나타내는 변수로 우수고객의 경우 통화내역 중 10개미만의 전화번호를 갖는 비율이 5%이고 비 우수고객인 경우 10개 미만의 전화번호를 갖는 비율이 62%로 통화량이 적은 고객일수록 고정번호에 대한 발신횟수 배율이 큰 것으로 나타났고 남성의 경우 10개 미만 비율이 63%이고 여성의 경우 20개미만 비율이 60%로 나타나 여성이 동일하지 않은 번호로 발신횟수가 많은 것으로 나타났다.

(4) 인터넷 서비스 사용 호수

이동 통신 이용 고객의 인터넷 서비스 사용횟수로 우수고객인 경우 인터넷 서비스를 이용하지 않은 비율이 95.7 %이고 비우수고객인 경우 인터넷 서비스를 사용하지 않은 비율이 98.76 %로 인터넷 서비스 이용횟수에 대한 두 그룹간의 차이점은 미비한 것으로 나타났다. 인터넷 서비스를 사용한 고객의 연령층은 20대, 30대, 10대 순으로 나타나 우수고객 증가를 위해 중,장년층에 대한 인터넷 서비스이용 활성화에 관한 대책이 필요하다는 것을 알 수 있다.

(5) Mobile to Mobile 발신호수

이동 통신 이용 고객의 핸드폰 대 핸드폰 통화횟수로 우수고객인 경우 월 평균 74통화, 최대 639통화의 Mobile to Mobile 통화량을 보였고 비우수고객의 경우 월 평균 18, 최대 90통화의 통화량을 보였다.

(6) 마일리지

이동 통신 이용 고객의 사용 요금에 따라 이동 통신 업체에서 제공하는 포인트 점수로 일정 점수에 이르게 되면 고객에게 편의를 주는 판촉 방법이다. 우수고객인 경우 요금할인이나 사은행사 등의 서비스 혜택으로 마

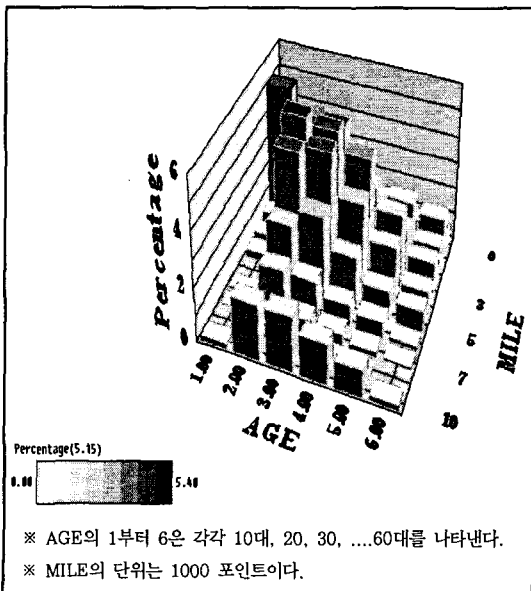


그림 5. 연령별 마일리지 비율

일리지 포인트적립 분포가 고르게 나타났고 비우수고객의 경우 4000포인트 이하의 사용자가 전체의 70%를 차지하여 고객의 우수성 판정을 위한 중요 변수임을 알 수 있다. [그림 5]은 이동 통신 이용 고객의 연령별 마일리지 비율을 나타낸다.

(7) 기본요금

이동 통신 이용자의 기본요금을 나타내는 변수이다. 우수고객의 기본요금 평균은 15591원이고 비우수고객의 기본요금 평균은 14151원으로 통화량이 적은 그룹이 기본요금 부담이 적은 요금 상품을 선택하는 비율이 높은 것으로 나타났다. [그림 6]은 이동 통신 이용 고객의 연령별 요금을 나타낸다.

(8) 정보이용료

이동 통신 이용자의 정보이용료를 나타내는 변수이다. 우수고객의 기본요금 평균은 월 평균 150원이고 비우수고객의 기본요금 평균은 73원으로 통화량이 적은 그룹이 이동 통신을 이용한 정보 이용 빈도가 낮은 것으로 나타났다.

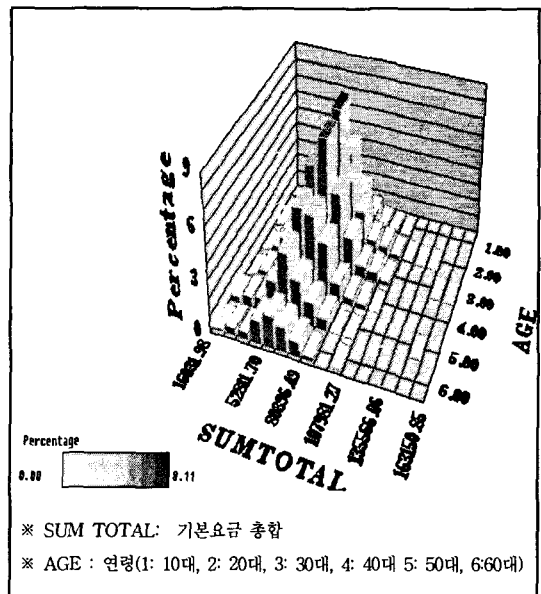


그림 6. 이동 고객의 연령별 요금

(9) 총 통화시간

이동 통신 이용자의 월간 통화시간을 나타내는 변수이다. 우수고객의 경우 월 통화시간이 123분 이하가 90.14 %이고 비우수고객의 경우 월 통화시간이 35분 이하가 80%로 나타나 그룹간의 통화시간이 4배 정도의 차이를 보이는 것으로 나타났다. [그림 7]은 이동 통신 이용 고객의 성별 월 통화시간을 나타낸다.

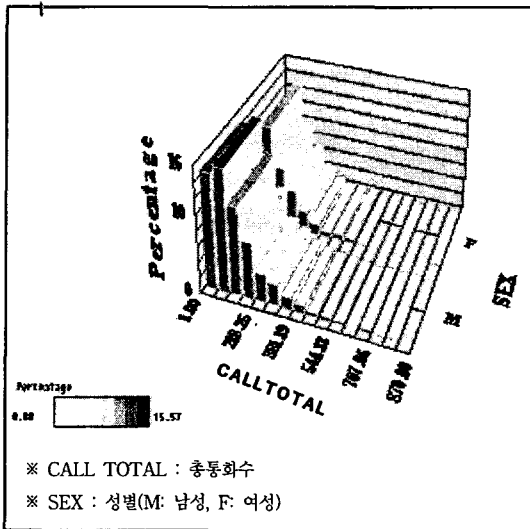


그림 7. 고객의 성별 월 통화시간

4. 모형의 구성

4.1 인공신경망 모형 구성

본 논문에서 사용한 인공신경망으로는 다층 퍼셉트론을 사용하였다. 다층퍼셉트론은 입력층과 은닉층 및 출력층의 3개의 계층으로 구성되어있다. 본 논문은 입력층은 20개의 노드로 구성하였고, 은닉층은 1개의 은닉층에 40개의 은닉노드로 구성하였다. 출력변수는 고객의 우수성 여부를 판단하는 것이 목적으로 1개의 층으로 구성하였다. SAS E-Miner는 인공 신경망의 학습방법으로 인공신경망 기법중의 하나인 오류 역전파알고리즘을 지원하며 이를 적용하여 분석 하였다.

4.2 로지스틱 회귀분석 모형 구성

본 연구의 연구모형 성과비교를 위해 사용할 통계적인 기법으로 입력변수와 종속변수는 신경망과 동일하게

선정하였다.

4.3 분석 절차

본 연구에서 설계된 인공신경망 모형은 2000개의 사례를 1000개의 훈련 집합과 1000개의 검증집합으로 구분하였다. 모형구축과 자료 분석을 위한 소프트웨어로는 SAS사에서 공급하는 통계용 툴인 SAS와 데이터 마이닝 툴인 엔터프라이즈 마이너를 이용하였다. 아래 [그림 8]은 엔터프라이즈 마이너를 이용한 인공신경망과 로지스틱 회귀분석 모형 구성도이다.

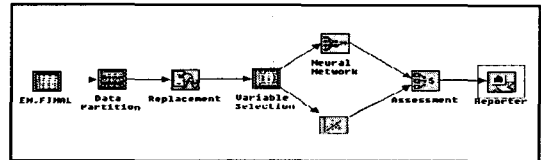


그림 8. 인공신경망과 로지스틱 회귀분석 모형 구성도

4.4 모형의 성과 측정 방법

인공신경망과 로지스틱 회귀분석의 성과측정을 위해 검증집합의 사례적중률을 이용하였다. 이는 실제 값과 모형에서 예측된 값을 비교하여 실제값과 예측값이 일치한 사례가 전체집합에서 차지하는 비율이다[4]. 이 결과는 오 분류표로 나타내어질 수 있다. 오 분류표란 목표변수의 실제 범주와 모형에 의해 예측된 분류범주 사이의 관계를 나타내는 표이다. 즉 목표변수의 범주별로 이를 제대로 분류한 빈도와 그렇지 못한 빈도를 함께 제시한 표로 목표변수의 범주가 s개인 경우 s * s개의 셀로 이루어진 표 형식을 취한다. 이때 분류표의 대각에 존재하는 셀은 목표변수의 원래 범주가 제대로 분류된 부분이고 비대각에 존재하는 셀은 분류가 잘못된 부분으로 구분되어 각각의 빈도가 제시된다. [표 5]는 오 분류표의 구성을 나타낸다[11].

표 5. 오 분류표의 구성

		예측된 변수	
		0	1
원래 목표변수	0	실제 0 예측 0	실제 0 예측 1
	1	실제 1 예측 0	실제 1 예측 1

IV. 실험 결과의 분석

1. 인공신경망 모형

인공신경망 모형에서의 고객 분류 적중률은 테스트집합에서 우수고객을 우수하다고 예측하고 비 우수 고객을 비 우수고객이라고 예측한 확률이 93.1%로 나타났고 우수고객을 비 우수고객으로, 비 우수고객을 우수고객으로 예측한 확률은 6.9%로 나타났다. 검증집합에서의 예측률은 각각 92.3% 와 7.7%로 나타났다. 따라서 고객그룹을 대상으로 마케팅을 하는 것을 전제로 할 때 시장 세분화를 위하여 같은 유형의 고객군을 분류할 때 잘못된 분류로 인하여 마케팅 목표를 선정함에 있어 그릇된 결정을 내릴 수 있는 소지가 매우 적다는 의미가 있다. [그림 9]은 인공신경망 모형의 수행결과를 나타낸다.

테스트 집합					검증 집합				
TABLE OF TARGET BY OUTPUT					TABLE OF TARGET BY OUTPUT				
TARGET	OUTPUT			Total	TARGET	OUTPUT			Total
Frequency/Percent					Frequency/Percent				
Row Pct					Row Pct				
Col Pct	0	1			Col Pct	0	1		
0	672	17		689	0	659	23		682
	67.20	1.70		68.90		65.90	2.30		68.20
	97.53	2.47				96.63	3.37		
	92.82	6.16				92.43	8.01		
1	52	259		311	1	54	264		318
	5.20	25.90		31.10		5.40	26.40		31.80
	15.72	83.28				15.98	84.02		

그림 9. 인공신경망 수행결과

2. 로지스틱 회귀분석

로지스틱 회귀분석 모형에서의 고객 분류 적중률은 테스트집합에서 우수고객을 우수하다고 예측하고 비 우수 고객을 비 우수고객이라고 예측한 확률이 88.4%로 나타났고 우수고객을 비 우수고객으로, 비 우수고객을 우수고객으로 예측한 확률은 11.6%로 나타났다. 검증집합에서의 예측률은 각각 87.6% 와 12.4%로 나타났다. [그림 10]은 로지스틱 회귀분석 모형의 수행결과를 나타낸다.

테스트 집합					검증 집합				
TABLE OF TARGET BY OUTPUT					TABLE OF TARGET BY OUTPUT				
TARGET	OUTPUT			Total	TARGET	OUTPUT			Total
Frequency/Percent					Frequency/Percent				
Row Pct					Row Pct				
Col Pct	0	1			Col Pct	0	1		
0	652	30		682	0	652	30		682
	65.20	3.00		68.20		65.20	3.00		68.20
	95.60	4.40				95.60	4.40		
	88.35	11.45				87.40	11.81		
1	66	232		318	1	94	224		318
	6.60	23.20		31.80		9.40	22.40		31.80
	27.14	72.86				29.56	70.44		

그림 10. 로지스틱 회귀분석 수행결과

3. 분석 모형의 결과 분석

위 실험 결과에서는 인공신경망에 의한 고객분류 적중률이 최고 93.1% 수준이었고 로지스틱 회귀분석은 최고 88.4%수준으로 인공신경망 모형이 로지스틱 회귀분석 모형보다 우수한 것으로 나타났다[표 6].

표 6. 인공신경망 모형과 로지스틱 회귀분석모형 결과

적중률	(단위:%)	
	인공신경망	로지스틱 회귀분석
훈련집합의 적중률	93.1	88.4
검증집합의 적중률	92.4	87.6

V. 결 론

이동통신서비스 시장에서 기업간 경쟁이 더욱 심화되고 최근 신규 가입자의 증가율이 급격한 감소 추세를 나타냄에 따라 기업이 마케팅 활동을 수행함에 있어 기존 고객의 유지가 중요한 과제가 되고 있다[9]. 본 논문에서는 국내의 이동통신업체의 고객을 대상으로 데이터 마이닝을 위한 기법 중 인공신경망과 로지스틱 회귀분석을 이용하여 기존 고객을 분석하였고 그 결과는 다음과 같다.

첫째, 데이터 마이닝에 의한 고객분석 결과 인공지능경망을 이용할 경우 최대 93.1%의 적중률을 보였으며 로지스틱 회귀분석의 경우 최대 88.4%의 적중률을 보였다 둘째, 인적정보와 요금정보, 이용정보를 이용하여 기존 고객에 대한 우수, 비 우수성을 예측할 수 있다. 단순 통계량에 의한 방법의 경우 개별고객에 대한 점수를 부여하기 어렵고 실제로 적용하기에 한계가 있다. 또한 설문조사의 경우도 고객의 응답내용이 신뢰성이 결여되기 때문에 설문 결과를 개별 고객에게 적용하기에는 문제점을 가지고 있다. 셋째, 기존 고객유지를 위한 마케팅 방법을 제시한다. 고객에 영향을 미치는 변수를 분석하여 영향을 끼치는 변수에 대한 대응책을 강구함으로써 기업의 수익을 증가시킬 수 있다.

넷째, 다수의 변수 중에서 고객의 특정 속성을 인공지능망에 적용함으로써 우수, 비 우수 고객의 유형을 추출할 수 있다. 입력변수가 단순화됨으로써 모형을 쉽게 파악할 수 있고 우수고객과 비 우수고객을 구분함으로써 마케팅 표적을 세분화할 수 있고 효율적인 마케팅 전략수립이 가능하다. 추후 연구과제로는 우수, 비 우수 고객의 분석뿐 아니라 부정사용자, 고객이탈분석, 고객 신용도 평가, 트래픽 분석 등 다각적인 고객분석이 필요하다.

참고 문헌

- [1] 강현철, 한상태, 최종후, *데이터 마이닝 방법론 및 활용*, 자유아카데미, 1999.
- [2] 박찬욱, *데이터베이스 마케팅*, 연암사, 1996.
- [3] 조재희, 박성진, *데이터 웨어하우징과 OLAP*, 대청, 1996.
- [4] 최종후, 한상태, *데이터 마이닝 기능과 사용법*, 자유아카데미, 2000.
- [5] 허준, *통신시장에서의 데이터 마이닝 Telco-CAT*, SPSS아카데미, 2000.
- [6] C. Apte and S. M. Weiss, "Data mining with decision trees and decision rules." *Future Generation Computer Systems*, Vol.13, pp.197-210, 2003.
- [7] M. Berry and G. Linoff, *Data Mining Techniques For Marketing, Sales and Customer Support*, John Willy & Sons, 1997.
- [8] J. P. Bigus, *Data Mining with Neural Networks*, McGraw-Hill, 1996.
- [9] E. Carl and E. Katz, "Customer spending behavior and telecommunications services," *Telecomm-unications Policy*, Vol.22, pp.23-46, 2004.
- [10] A. Famili and W. M. Shen, "Data Preprocessing and Intelligent Data Analysis," *Intelligent Data Analysis*, Vol.1, pp.3-23, 1997.
- [11] U. M. Fayyad and G. P. Shapiro, *Advanced in Knowledge Discovery and Data Mining*, AAAAI Press/MIT Press, 2003.
- [12] A. Feelders, H. Daniels, and M. Holsheimer, "Methodological and practical aspects of data mining," *Information & Management*, Vol.37, pp.271-281, 2000.
- [13] B. A. Jain and B. N. Nag, "Performance Evaluation of Neural Network Decision Models," *Journal of Management Information Systems*, Vol.14, pp.201-216, 1997.
- [14] H. Lan, *Data Mining*, Morgan Kaufmann Publishers, 1999.
- [15] D. Pyle, *Data Preparation for Data Mining*, Morgan Kaufmann Publishers, 1999.
- [16] M. Rob, *Data Warehousing and Data Mining for Telecommunications*, Artech House, 1987.
- [17] S. Zhang, "Aggregation and maintenance for database mining," *Intelligent Data Analysis*, Vol.3, pp.475-490, 2004.

저자 소개

이 병 엽(Byoung-Yup Lee)

정회원



- 1991년 2월 : 한국과학기술원 전산학과(공학사)
- 1993년 2월 : 한국과학기술원 전산학과(공학석사)
- 1997년 2월 : 한국과학기술원 경영정보공학(공학박사)

- 1993년 1월~2003년 2월 : 대우정보시스템 차장
- 2003년 3월~현재 : 배재대학교 전자상거래학과 조교수
<관심분야> : XML, 지능정보시스템, 데이터베이스 시스템, 전자상거래학

유 재 수(Jae-Soo Yoo)

증신회원



- 1989년 : 전북대학교 컴퓨터공학과(공학사)
- 1991년 : 한국과학기술원 전산학과(공학석사)
- 1995년 : 한국과학기술원 전산학과(공학박사)

- 1995년~1996년 8월 : 목포대학교 전산통계학과 전임 강사
- 1996년 8월~현재 : 충북대학교 전기전자컴퓨터공학부 부교수
<관심분야> : 데이터베이스 시스템, XML, 멀티미디어 데이터베이스, 분산 객체 컴퓨팅 등

조 규 하(Kyu-Ha Joh)

준회원



- 1999년 2월 : 동국대학교 컴퓨터공학과(공학사)
- 1999년 1월~2003년 10월 : 대우정보시스템 대리
- 2003년 11월~현재 : (주)아인스 에스엔씨 부장

- 2004년 8월~현재 : 충북대학교 전기전자컴퓨터공학부(석사과정)
<관심분야> : XML, 데이터베이스 시스템, 분산 객체 컴퓨팅, 알고리즘, EAI 등

송 석 일(Seok-II Song)

정회원



- 1998년 : 충북대학교 정보통신공학과(공학사)
- 2000년 : 충북대학교 정보통신공학과(공학석사)
- 2003년 : 충북대학교 정보통신공학과(공학박사)

- 2003년~현재 : 충주대학교 전기전자정보공학부 조교수
<관심분야> : 데이터베이스 시스템, XML 색인, 다차원 색인 등