

IP기반 유선인터넷전화 가입요인 도출을 위한 분석적 연구: 통신상품결합서비스의 영향

하성호

(경북대학교 경영학부 교수, hsh@mail.knu.ac.kr)

양정원

(경북대학교 일반대학원 경영학과)

... Abstract ...

Recently, Internet Telephony has become increasingly popular in telecommunication industry. However, previous research on Internet Telephony has focused on analyzing specific Internet Telephony solutions, identifying the Internet Telephony movement itself. The research on prediction models about Internet Telephony adoption has been minimal. The main propose of this study is to develop models for predicting transition intention from using traditional telephones to using Internet Telephony. To do so, this study uses data mining methods to analyze demands in the IT communications market and to provide management strategies for Internet telephony providers. Especially this study uses discriminant analysis, logistic regression, classification tree, and neural nets to develop the prediction models for the Internet Telephony adoption. The models are compared with each other and a superior model is chosen.

Keywords: Internet telephony, data mining, transition intention

I. 서 론

정보통신기술이 빠른 속도로 발전함에 따라 통신서비스도 나날이 진보하고 있다. 차별화되고 진보된 통신서비스가 기존 서비스

를 대체하는 형태로 늘어나고 있는 것이다. 그 중 인터넷전화는 기존의 유선전화를 대체할 수 있는 가장 대표적인 서비스로 향후 발전과 이용 확산이 예견된다.

유선전화(Plain Old Telecommunication

Service: POTS) 란 가입자와 교환국 사이의 유선 선로(가입자망)를 통해 시내전화(통화권내), 시외전화(통화권간), 국제전화, 공중전화가 가능한 통신서비스이다. 인터넷 전화(Voice over Internet Protocol: VoIP)는 기존의 유선전화 기능을 제공하면서도 요금이 저렴하여 유선전화를 대체할 수 있는 잠재력을 보유한 서비스이다 [권오상 외 1인 2002]. 특히 최근 사업자들의 품질개선 노력과 인터넷전화로의 번호이동이 간편해지면서 인터넷전화에 대한 인지도가 향상되고 유선전화에 대한 대체가능성이 높아지고 있다. 또한, 인터넷전화는 물리적 네트워크에 종속되지 않고, 통화품질도 우수할 뿐 아니라 초고속 인터넷의 높은 보급률에 힘입어 전망이 아주 밝다고 할 수 있다 [함창용 외 4인 2007].

KISDI 조사에 따르면, 국내 유무선 및 인터넷전화 서비스 가입현황은 유선전화 72%, 이동전화 97%, 인터넷전화가 21%이며, 인터넷전화 서비스 가입자 중에서 유선전화를 해지한 고객은 81%에 이르는 것으로 조사되었다 [KISDI 2009]. 유선전화 해지 사유로는 ‘인터넷 전화의 통화요금이 저렴해서(75%)’ 또는 ‘결합요금제 때문에(18%)’라는 응답비중이 높아, 유선전화에서 인터넷전화로의 전환에 비용이 중요한 역할을 하는 것으로 분석되었다. 또한, 인터넷전화와 유선전화에 동시 가입한 사용자 중, 70%가 유선전화를 해지하겠다고 응답한 반면, 인터넷전화를 해지하겠다는 응답은 27%에 불과한 것으로 나타났다. 이것은 인터넷전화가 기존 유선전화를 대체할 수 있는 서비스로서의 가능성을 보여주며, 기존 유선전화에서 인터넷전화 서비스로 전환하도록 고객을 유인할 수 있는 서비스 개발이 필요함을 시사한다.

인터넷전화는 저렴한 요금뿐만 아니라 다양한 부가서비스와 영상통화를 통하여 새로운 시장을 창출할 것으로 기대되고 있다. 특히 와이브로를 비롯하여 무선 인터넷이 활성화되면, 유무선 융합의 비즈니스 모델을 통해 유비쿼터스 환경에서 다양한 인터넷전화 서비스를 제공할 수 있을 것으로 기대된다 [김문구 외 2인 2008]. 그러므로 유선 인터넷전화의 효율적인 확산을 위해서는 고객의 특성을 파악하고 적절한 비즈니스 전략을 수립하는 것이 필요하다. 인터넷전화에 대한 수요가 급증하고, 기존 유선전화를 대체할 수 있는 서비스로서 가능성이 커지고 있으나 현재 인터넷전화에 대한 문헌 연구나 실증적 분석은 미미한 실정이다. 인터넷전화에 대한 대부분의 기존 연구들은 기술동향 [강태규 외 2인 2004], 제도 및 정책연구 [권오상 외 1인 2002, 김도환 2009], 관련 기술에 대한 개선 방안 [송관호 2007] 등이다. 이에 본 연구에서는 기존의 유선전화에서 인터넷전화로의 전환을 유도할 수 있는 요인들을 도출하는 인터넷전화 사용의 향 예측모델을 개발하고자 한다. 또한, 정보통신 기술의 급속한 발달에 따라 통신 산업의 비용예측과 합리적인 통신기술 수요예측의 필요성이 제기되고 있다. 본 연구에서는 인터넷전화 도입에 관한 예측모형을 개발하기 위해 판별분석(Discriminant analysis), 로지스틱 회귀분석(Logistic regression), 분류나무(Classification tree), 신경망(Neural network) 알고리즘을 비교 분석하고 가장 우수한 예측모델을 제시한다.

II. 이론적 배경

인터넷전화에 고객관계관리를 적용한 연구로는 다음과 같은 연구가 있다. Ezawa et al. [1995]은 통신 서비스 고객의 연체 및 부정사용 유형 분류에 베이지안을 근거로 네트워크 모델을 개발하였다. 또한, 이홍재 외 2인 [2000]은 국내 통신 서비스 시장에 적합한 신규고객 예측방법론을 개발하였다. 이 연구에서 이동전화시장에서의 선택적 고객이전(selective migration)에 관한 모형을 제시하여 신규 통신서비스의 시장 잠재력에 대한 예측방법을 제시하였다. 또한, 신규 통신서비스의 최적도입시기에 관한 예측분석에 유사 서비스의 경험이나 경제이론에 기초한 베이지안 확산모형을 적용하였으며, 특정 시기별로 목표 소비자그룹을 찾아내고 각 시기에 환경변수(요금, 사업자수 등)가 평균적인 소비자의 가입률에 미치는 영향을 보여주었다. Lingfen et al. [2006]은 비선형 회귀모델을 개발하여 VoIP의 음성 품질을 측정하였으며 실제 인터넷 VoIP 추적 데이터를 활용하여 개발된 모델이 높은 예측 정확도를 나타낸을 보였다.

통신고객의 서비스에 대한 요구와 기대가 점점 다양해지고 세분화됨으로써 전화서비스 제공업자의 기존 유선전화 고객유지와 신규 인터넷전화 가입자 유치에 부단한 노력과 고도의 전략이 요구되고 있다 [Rinde 1999]. Kohli et al. [2001]의 연구에서 고객 관계관리(Customer relationship management: CRM)란 고객과의 장기적 관계를 유지하기 위해 제품 또는 서비스제공자가 고객의 요구에 맞춰 지속적인 변화를 추구하고 고객 기대에 부응하는 프로세스라고 정의하였다. Massey et al. [2001]의

연구에서는 CRM이 성공적인 고객관계의 개발과 유지를 포함하는 것으로 정의하였다. Peppers et al. [1999]은 고객에 대한 정보를 기반으로 개별고객의 기대에 맞춰 기업의 행동을 변화시키려는 의지와 능력이라고 정의하였다. VoIP 서비스제공업자들은 전통적인 전화이용 고객을 적극적으로 유치하기 위해 CRM을 도입/개발/활용한다. 이는 미래에 VoIP가 기존의 유선전화 시스템을 완전히 대체할 수 있을 것이라고 기대하기 때문이다 [Zhang et al. 2005].

VoIP와 CRM을 결합함으로써 운영비용 절감과 고객 서비스 강화라는 두 가지 목표를 성취할 수 있다. VoIP 서비스제공자들은 고객 경험을 향상시키고, 기업과 고객 간에 더 좋은 관계를 형성하기 위해 CRM을 고객 서비스에 포함시킨다 [Kabiraj 2003]. CRM이 VoIP 서비스제공자들에게 가져다주는 이점은 고객만족을 증대시킴으로써 고객 유지와 충성도를 향상시키는 것이다 [Zhang et al. 2005]. 이를 통해 기업은 장기적인 수익을 안정적으로 얻고 산업 내에서의 경쟁력을 강화시킬 수 있게 된다. Buyut et al. [2008]의 연구에서는 VoIP 서비스에서의 CRM 도전과제를 설명하고 CRM의 이점과 영향력을 강조하였다. 또한, Tseng [2005]의 연구에서는 고객 선호도 분석을 통해 제품, 가격 전략을 도출하였다. 인터넷전화의 주요 속성(음성전송품질, 통화료, 통화시간, 통화 중(line busy) 문제, 장치 속성)을 도출하였고 ‘음성전송품질’과 ‘통화 중 문제’를 우선 해결해야 한다는 제품전략을 제시하였다.

인터넷전화의 전략적 관점을 제시한 연구도 있었는데, 장범진 외 2인 [2006]의 연구에서는 수직적 차별화 및 기술경쟁차원에서 인터넷전화시장 발전을 위한 로드맵을 제시

하고 융합 환경에 적합한 규제체계 개선방향에 대하여 고찰하였으며, 인터넷전화와 기존유선전화 간의 차별화된 번호정책을 제시하였다. Jung [2004]은 신기술을 비용절감기술과 혁신기술의 두 가지 유형으로 분류하고 인터넷전화가 단기적으로는 비용절감기술의 특성을 보이지만 장기적으로는 혁신기술의 특성을 가진다고 주장하였다. 기존 인터넷 망에서 저렴한 음성서비스를 단순 제공하는 것이 초기 인터넷전화 서비스의 전략이지만, 장기적으로는 인터넷 망의 고도화, 다양한 부가서비스와의 결합 등을 통하여 인터넷전화의 서비스 품질을 향상시키고 차별화하기 위한 전략 수립이 필요하다는 지적이다.

또한, 박종현 외 2인 [2008]의 연구에서는 국내 유선 VoIP 이용자를 현재이용자, 잠재이용자, 비 이용자로 세분화하여 비교하였고 인터넷전화를 이용하는 이유와 이용하지 않는 이유, 이용조건, 지불의사수준을 파악하여 유선 VoIP가 활성화되기 위한 전략을 제시하였다.

본 연구에서는 기존 유선전화서비스를 해지할 가능성이 높은 고객을 탐지하여 데이터마이닝 기법을 적용하고 이탈고객 또는 전환고객을 예측하여 VoIP 서비스제공자들의 인터넷전화서비스로 새로운 고객을 유치 할 수 있는 적절한 고객유치전략을 도출한다. 인터넷전화 사용의향을 예측하는 모델개발은 경쟁이 치열하고 고객의 해지가 빈번한 통신서비스 분야에서 고객 관리에 중요한 시사점을 제공한다. 본 연구에서는 인터넷전화 서비스를 제공하는 기업의 실제데이터를 이용하여 인터넷전화 사용의향 예측모델을 개발하고 유선전화에서 인터넷전화로의 전환고객에 대한 CRM 전략을 도출 할 수 있는 기반을 제공한다.

III. 연구 모델 개발

본 연구는 설문 자료를 기반으로 인터넷전화 사용의향 예측모델 개발을 목적으로 한다. 설문 자료의 특성상, 모든 자료는 숫자로 구성된 범주형이고 종속 변수는 인터넷전화 사용 의향이 있고 없음을 나타내는 이진형 범주로 나타낼 수 있다. 따라서 범주형 설명 변수와 이진형(Binary) 범주의 종속 변수를 가진 자료를 분석하는 알고리즘은 예측 목적용으로 로지스틱 회귀분석, 분류나무, 신경망 등이 있다. 본 연구에서 사용하게 될 또 하나의 알고리즘이 판별분석의 경우 연속형 설명 변수에 대한 분석 방법이지만 숫자로 구성된 범주형 설명 변수의 경우 적용이 가능하다 [Tabachnick et al. 2007, Shmueli et al. 2007].

로지스틱 회귀분석은 선형 모형의 단점을 극복하기 위해 종속 변수가 이진형 범주일 때 사용하는 특수 형태의 회귀 분석 알고리즘으로 회귀식을 통해 종속변수의 귀속 확률을 계산하여 예측을 하거나 통계적 추론을 하는 알고리즘이다 [Lee et al. 2002]. 판별분석 알고리즘은 관찰된 데이터를 근거로 판별식을 산출하여 주어진 상황에서 판별식의 산출 값에 따라 응답자들의 행동을 두 개 이상의 그룹에 각각 속하도록 예측하는 것으로 [정충영 외 1인 1998] 집단 간 공분산의 동일성 등의 조건을 전제로 하지만, 로지스틱 회귀분석과는 달리 두 개 이상의 그룹에 대해 한 개의 판별식으로 예측이 가능하기 때문에 신용 평가나 고객 예측 등의 많은 분야에서 이용되어 온 알고리즘이다.

분류나무 알고리즘은 수집된 데이터를 분석하여 데이터 사이에 존재하는 패턴의 분류 규칙을 나무의 형태로 만드는 것으로 신

용 예측이나 고객 평가에서 널리 사용되는 기법이다 [Breiman et al. 1984, 하성호 외 2인 2009]. 고객 분류나 신용 예측 규칙을 명시적인 규칙 셋(Rule set)으로 정의하고 있기 때문에 의사결정에 대한 근거를 설명하기 쉽다는 장점이 있다. 신경망 알고리즘의 경우는 통계적 가설 없이 비선형 회귀 모델을 설명하기에 적당해서 고객 신용 예측에서 뛰어난 결과를 보여 주고 있다 [Altman et al. 1994, Cheng et al. 1994, Desai et al. 1997]. 기존의 통계적인 알고리즘들과 데이터마이닝 알고리즘들을 혼합하여 구현한 연구들이 늘어나고 있는데 대표적으로 Lee et al. [2002]의 연구에서는 판별분석에서 선택된 변수를 신경망의 입력변수로 사용하는 혼합 모형을 구현할 경우 신경망만으로 구성된 단일 모형보다 예측율이 더 좋았다고 밝히고 있다. 따라서 본 연구에서는 판별분석, 로지스틱 회귀분석, 분류나무, 신경망(단일, 혼합) 모델을 이용하여 예측모델을 구현하고 성능을 비교 평가한다.

IV. 모델 적용 및 결과 해석

1. 데이터 분석

본 연구에서 분석한 자료는 국내 모 통신사가 실시한 기초조사(집 전화를 기반으로 한 추가 서비스 개발)에서 수집된 것으로 설문에 응한 고객의 인구 통계학적 정보와 고객들이 현재 이용하고 있는 통신 기술(유선전화, 초고속인터넷, 인터넷전화, 이동전화)에 대한 정보, 향후 이용할 계획이 있는 통신 기술에 대한 정보를 담고 있다. 설문

에 응한 고객의 수는 총 705명이며 모든 설문은 범주형 응답으로 구성되어 있다.

설문 자료를 기반으로 예측모델을 개발하기 위해 인터넷전화 향후 사용 의향을 나타내는 종속변수를 생성하였다. 향후 인터넷전화를 이용할 의향이 있는 고객이란 현재 유선전화를 사용하지 않고 인터넷전화를 사용하고 있는 고객이거나 향후 이용할 계획을 가지고 있는 고객, 현재 유선전화를 사용하고 있지만 추가로 인터넷전화를 사용할 계획을 가지고 있는 고객으로 정의하였다. 현재 유선전화를 사용하지 않고 인터넷전화를 사용하고 있는 고객 수는 36건, 그리고 현재 유선전화를 사용하고 있지만 추가로 인터넷전화를 사용할 계획을 가지고 있는 고객은 21건으로 조회되어 총 705건 중 57건이 인터넷전화 사용에 긍정적인 것으로 조회되었다. 인터넷전화 사용의도를 나타내기 위해 생성된 변수를 'D7'로 명명하고 인터넷전화 사용 의도를 가지면 '1'의 값을, 사용할 의도를 가지지 않으면 '0'의 값을 부여하였다.

현재 이용 중인 유선전화 서비스, 초고속 인터넷서비스, 인터넷전화서비스, 이동전화 서비스 중 같은 회사에 가입하여 이용하는 서비스가 무엇인지를 파악하기 위해 결합상품을 구분하는 변수 'D3'을 생성하였다. 이 변수는 결합통신서비스를 이용하는 경우를 묻는 변수로 인터넷전화 가입에 대한 고객의 의도를 파악하기 위해 설문 응답의 조합으로 만들어진 변수이다. '통신상품 결합서비스'란 2개 이상의 방송 또는 통신서비스 상품에 함께 가입해 이용하는 경우 방송통신 요금을 할인해 주는 상품이다. 방송통신 사업자들은 자사의 방송 또는 통신 서비스를 묶어서 제공하고 있으며 자사와 타사

의 방송 또는 통신 서비스를 묶어서 제공하는 경우도 있다.

변수 'D6'은 결합 통신서비스를 이용하는 경우와 그렇지 않는 경우를 나타내기 위해 생성된 변수이고, 변수 'D8'의 경우 유선전화에 대한 불만족 여부를 나타내기 위해 생성된 변수로 현재 유선전화를 사용하고 있는 응답자 중 유선전화를 해지할 의향이 있는 고객과 현재 유선전화를 사용하지 않고 있는 고객 중 사용하지 않는 이유를 비용이 비싸다거나 이용이 불편해서라고 응답한 경우를 유선전화에 대해 불만이 있는 경우로 해석하였다. 변수 'D9' ~ 'D12'는 결합 서비스 조합별로 이용하는 통신사를 표현하기 위해 생성된 변수들이다. 변수 'D14'는 통신요금을 나타내는 것으로 현재 사용 중인 유선전화 요금과 인터넷전화 요금을 이용하여 생성된 변수이다. 새로 생성된 변수들은 원래 자료와 마찬가지로 모두 범주형으로 구성되었다. 각 알고리즘에 사용된 설명변수와 종속변수들은 표1에 기술되어 있다.

표 1. 모델 구축에 사용된 설명변수와 종속변수.

변수 이름	정의
A12	성별
A13	나이
A15	주택 형태
A17	가구 유형
A22	유선 전화 서비스 회사
A23	초고속 인터넷 서비스 회사
A25	이동 전화 서비스 회사
C14	가구주 직업
C28	한 달 평균 수입
C32	맞벌이 여부
D3	사용하는 통신 결합 상품 유형
D4	사용하는 통신 결합 상품 통신사
D6	통신 결합 상품 이용 여부
D7	인터넷전화 사용 의향(종속변수)
D8	유선 전화 불만 여부

D9	통신사별 결합 유선-초고속-이동
D10	통신사별 결합 상품-유선/초고속인터넷
D11	통신사별 결합 상품-유선/이동전화
D12	통신사별결합-초고속/이동전화
D14	통신 요금

2. 인터넷전화 사용의향 예측모델

인터넷전화 사용의향 예측을 위한 모델을 개발하기 위해 범주형 독립변수를 이용하여 범주형 종속변수를 예측할 수 있는 데이터 마이닝 알고리즘을 이용하였다. 본 연구에서 채택한 알고리즘은 판별분석, 로지스틱 회귀분석, 분류나무, 신경망이다. 데이터마이닝 소프트웨어는 SPSS의 클레멘타인 12.0을 이용하였다.

표1에서 보이듯 인터넷전화 사용 의도를 가진 고객을 의미하는 성공 케이스가 전체 케이스의 8.1% 정도이기 때문에 성공 케이스와 실패 케이스의 균형을 맞추는 일이 필요하다. 개수가 작은 성공 케이스를 부풀려(Oversampling) 성공 케이스와 실패 케이스의 균형을 맞추고 학습(Training) 자료 셋을 구성하며, 검증(Validation) 자료 셋은 원래 자료 분포에 가깝게 성공 케이스를 작게 구성한다. 이와 같이 작은 수의 성공 케이스와 많은 수의 실패 케이스의 균형을 맞추는 일은 성공 케이스에 대한 데이터마이닝 알고리즘의 학습 능력을 극대화하기 위한 방법으로 알려져 있다 [허명희 외 1인 2003, Shmueli et al. 2007].

종속변수의 값으로 자료를 정렬하고 클레멘타인의 샘플링(Sampling) 노드에서 '1-in-n' 옵션을 사용하여 학습 자료 셋과 검증 자료 셋이 겹치지 않도록 자료를 분리한다. 선택된 학습 자료는 성공 케이스에 대한 학습을 극대화하기 위해 밸런싱

(Balancing) 노드를 사용하여 성공 케이스에 11,172의 균형 가중치를 주고 성공 케이스와 실패 케이스를 동수로 구성한다. 밸런싱 이전에 29개이던 성공 케이스는 밸런싱 이후 실패 케이스와 동수인 324건으로 부풀려졌다. 검증 자료 셋은 원래 자료의 분포와 유사하게 구성한다.

샘플링과 밸런싱 과정을 통해서 구성된 학습 자료 셋은 성공 케이스 324건, 실패 케이스 324건으로 구성되었고 검증 자료 셋은 성공 케이스 28건, 실패 케이스 324건으로 구성되었다. 검증 자료 셋에서 성공 케이스의 비율은 7.95%로 원 데이터 셋의 구성(8.1%)과 비슷하다. 자료 셋의 분할에 앞서 일부 널 값을 가진 독립 변수의 경우 미리 지정된 값으로 치환하는 과정을 밟았다. 널 값을 가진 설명 변수들은 주로 고객이 사용하고 있는 통신 서비스의 종류를 묻는 문항들이었다.

2.1 관별분석

관별분석 알고리즘을 이용한 모델 구축은 ‘Enter’ 옵션과 ‘Stepwise’ 방식을 각각 수행하여 검증 자료 셋에서 결과를 비교하였다. 관별분석 알고리즘에서 집단 간 예측 변수들의 분산-공분산의 일치성은 Box-M 테스트를 통해 검증되는데 두 가지 옵션을 사용한 관별분석 검정 결과 유의확률 0.000으로 유의수준 0.05에서 모집단 공분산 행렬이 동일하다는 귀무가설이 기각되었다. 이것은 관별분석의 기본 가정에 위배되는 사항이지만 공분산 행렬의 동일성이 극단적으로 위배되지 않는다면 관별식을 이용할 수 있다고 알려져 왔으므로 [정충영 외 1인 1998], 본 연구에서는 자료를 그대로 이용하였다.

관별분석 실행 옵션은 다음과 같이 설정

하였다. ‘Stepwise/Prior Probability (All group equal) / Use Covariance Matrix (Within-group covariance) / Method (Wilks' Lambda or mahalanobim distance) / Criteria (F value-Entry (3.84) - Removal (2.71)’. 모델 실행 결과, 고유값 (Eigenvalue) 0.745, Wilks 람다 값 0.573(유의확률 0.000), 정준 상관 계수 0.653을 보여주었다. 중요변수로 A22, D4, D10, D14, D3, D11, D6, A15, D12가 선택되었다. 관별분석 알고리즘에 의해 선택된 예측 변수의 상대적인 공헌 정도를 나타내는 표준화된 정준 관별 함수의 값(Standardized canonical coefficient)이 비교적 큰 변수는 A22(0.768), D14(0.603), D3(0.597), D12(-0.823)로, 인터넷전화 사용 의향은 유선전화 서비스 회사, 통신요금, 사용하는 통신 결합 상품 유형, 통신사별 결합(초고속-이동전화) 변수와 밀접한 관계가 있음을 알 수 있었다. ‘Enter’ 방식을 사용한 경우 분류오류율(Classification error rate)이 검증 자료 셋에서 14.2%를 나타내었고, ‘Stepwise’ 방식을 사용했을 때 분류오류율은 검증 자료 셋에서 11.36%로 향상되었다.

2.2 로지스틱 회귀분석

관별분석이 집단 간 공분산의 일치성과 정규성 등의 조건이 충족되어야 함을 전제로 한다면 로지스틱 회귀분석의 경우 이러한 전제 조건에서 좀 더 자유롭다고 알려져 왔다 [Tabachnick et al. 2007]. 본 연구는 클레멘타인의 로지스틱 회귀분석 알고리즘을 사용하여 ‘Enter’ 방식으로 전체 변수를 대상으로 했을 때와 ‘Stepwise’ 방식으로 변수를 일부 선택했을 때를 비교하여 검증 자료 셋에서의 분류오류율을 구하였다.

'Enter' 방식으로 전체 변수를 대상으로 하였을 때 검증 자료 셋에서 예측오류율은 11.93%로 나타났고, 'Stepwise (FIN (0.05), FOUT (0.01))' 방식으로 로지스틱 회귀분석을 실행했을 때 예측오류율은 11.65%로 향상되었다. 이때 선택된 변수들은 A22, D14, A15, A17, D3, A25, C14, A12, D4이다. 예측 변수들의 유형이 모두 셋(Set)으로 설정되어 로지스틱 회귀분석에서는 각 셋에 대한 분석이 이루어졌는데 0.05 유의수준에서 통계적 유의성을 가지는 변수들은 A12(1), A15(1, 2), A25(1), D3(3, 5), D14(1, 2)였다. 'Stepwise' 로지스틱 회귀분석을 수행했을 때, 학습 자료 셋에서의 분류오류는 2.3%, 검증 자료 셋에서는 11.65%를 보여주었다.

2.3 분류나무 모델

클레멘타인의 C5.0 모델을 사용하여 분류나무 알고리즘을 적용하였다. C5.0 실행 옵션은 다음과 같다. 'Simple / Accuracy / Expected Noise(0)'이며 알고리즘 수행 결과 생성된 나무의 깊이(Depth)는 6개였다. C5.0 알고리즘에 의해 선택된 중요 변수는 A22, D14, C14, D3, A12, A15, C28, C32, D11, A17, A13 순서였고 성공 케이스 (D7=1)를 분류하기 위해 형성된 규칙 셋 (Rule set)은 총 10개였다. C5.0 알고리즘은 학습 자료 셋에서 분류 오류 0.62%를 나타냈지만, 검증 자료 셋에서는 11.65%를 보여주었다.

2.4 신경망 모델

Lee et al. [2002]의 방법을 따라서 1) 판별분석에서 선정된 변수를 입력 변수로 사용한 혼합모형, 2) 로지스틱 회귀분석에서

선정된 변수를 입력 변수로 사용한 혼합모형, 3) 판별분석과 로지스틱 회귀분석에서 공통적으로 나타난 변수를 입력 변수로 사용하는 혼합모형, 4) 전체 변수를 모두 사용한 단일모형을 상호 비교하여 예측율이 높은 모형으로 신경망 모델의 성능으로 정하기로 하였다. 신경망 모델 수행 파라미터는 은닉층 1개, 은닉노드 20개, 학습률 0.9, 초기 모멘텀 0.3, 최종 모멘텀 0.01로 설정하였다.

네 개의 후보 모델을 실행한 결과 검증 자료 셋에서의 예측오류율이 전체 변수를 사용한 모형(4)이 6.82%, 로지스틱 회귀분석의 변수를 사용한 모형(2)이 10.23%, 판별분석의 변수를 사용한 모형(1)이 13.92%, 판별분석 모델과 로지스틱 회귀분석 모델의 공통 변수를 사용한 모형(3)의 예측오류율이 15.06%였다. 전체 변수를 사용한 모형이 가장 좋은 성능을 보였고, D14, D4, A17, D10, D9, A25, D6, A12, D11, D8 순으로 변수의 중요도가 매겨졌다. 이때 신경망 구조는 입력계층에 19개의 뉴론, 은닉계층에 20개의 뉴론, 출력계층에 1개의 뉴론이 사용되었다.

3. 모델 비교

표2는 구축된 다섯 가지 모델의 성과를 비교하기 위해 검증 자료 셋에 대한 예측오류율, Type I 오류율, Type II 오류율을 각각 나타내었다. 전체 예측오류율을 비교해보았을 때 단일신경망(전체변수 사용)을 사용했을 경우가 가장 성과가 좋은 것으로 나타났다. 그러나 Type I 오류율을 고려하면 판별분석이 더 좋은 성과를 보이는 것으로 나타났다. Type I 오류는 실제로 인터넷

전화를 사용할 의향이 있는 고객인데 그렇지 않은 고객으로 예측하여 이익을 낼 수 있는 고객을 잃어버리는 경우에 해당한다. 반면에 Type II 오류란 실제로는 인터넷전화를 사용할 의향이 없는 고객임에도 의향이 있는 것으로 예측하여 불필요한 마케팅 비용을 지불해야 하는 경우를 말한다. Type I 오류에 대한 비용과 Type II 오류에 대한 비용의 차이가 크지 않다면 인터넷전화 사용 의향을 예측하기 위한 최적의 모델은 전체 예측오류율에서 최고의 성과를 보인 단일신경망(전체변수 사용) 모델이 된다. 그러나 Type I 오류에 대한 비용이 Type II 오류에 비해 월등하게 크다면, 즉 고객을 잃는 비용이 마케팅 비용을 불필요하게 지불하는 것보다 더 높다면 인터넷전화 사용 예측을 위한 최적의 모델은 Type I 오류가 가장 작은 판별분석 모델이 된다.

표 2. 모델 별 예측오류율 비교

모델	예측 오류율	Type I 오류율	Type II 오류율
판별분석	11.36%	21.43%	10.49%
로지스틱 회귀분석	11.65%	35.71%	9.26%
분류나무	11.65%	32.14%	9.88%
단일신경망 (전체변수)	6.82%	39.29%	4.01%
혼합신경망 (혼합모형)	13.92%	28.57%	12.65%

그림1은 다섯 가지 모델의 게인즈 차트를 나타낸 것으로 Type I 오류 비용을 상대적으로 크게 했을 경우 가장 위쪽에 나타난 판별분석이 가장 성과가 좋은 모델이다.

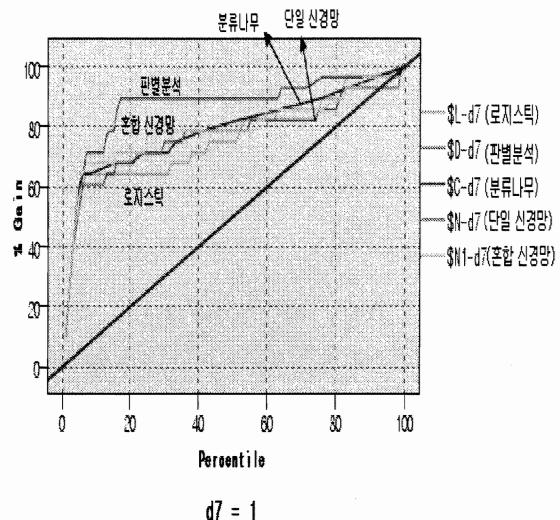


그림 1. 전체 모델에 대한 게인즈 차트

V. 예측 모델의 적용

인터넷전화를 사용할 의향을 가진 고객을 그렇지 않은 고객으로 잘못 예측하는 경우의 손해가 더 크다고 전제한다면(즉, Type I 오류가 더 크다고 할 경우), 본 연구에서 최고의 성과를 보인 모델은 판별분석 모델이 된다. 판별분석 모델에서 선정된 변수를 중요도에 따라 열거하면 A22(유선전화 서비스 회사), D4(사용하는 통신 결합 상품/통신사별), D10(통신사별 결합상품-유선/초고속인터넷), D14(통신요금), D3(사용하는 통신 결합 상품 유형), D11(통신사별 결합상품-유선/이동전화), D6(통신 결합 상품 이용 여부), A15(주택형태), D12(통신사별 결합-초고속/이동전화)이다.

이들 변수 중 D4(사용하는 통신 결합 상품-통신사별)와 인터넷전화 사용의향을 가진 고객과의 관계를 교차표를 통해 살펴보면 표3에서 'C 통신 회사'의 결합상품을 쓰

고 있는 고객들의 인터넷전화 가입의향이 다른 통신 회사의 고객보다 높은 것을 알 수 있다. 따라서 C 통신사의 결합상품을 사용하고 있는 고객을 표적 마케팅할 경우 성공적으로 인터넷전화에 대한 사용을 늘릴 수 있을 것이다.

표 3. 통신사에 따른 인터넷전화 가입의향

		인터넷전화 가입의향		전체
		없음	있음	
D4	없음	237	6	243
	A 통신사	280	17	297
	B 통신사	110	7	117
	C 통신사	21	27	48
	전체	648	57	705

표4는 D14(통신요금)과 인터넷전화 사용의향을 비교하고 있는데 통신요금이 비싼 범주에 (범주 3) 들어가는 고객이 비록 수는 적지만 인터넷전화 사용의향을 매우 많이 가지고 있음을 알 수 있다. 따라서 현재 유선전화나 인터넷전화 요금을 포함한 통신요금이 비교적 높은 고객을 인터넷전화 마케팅의 대상으로 하는 것이 더 효과적인 방법임을 암시한다고 볼 수 있다.

표 4. 통신요금에 따른 인터넷전화
가입의향

		인터넷전화 가입의향		전체
		없음	있음	
D14	모름	134	5	139
	1	226	19	245
	2	238	16	254
	3	50	17	67
	전체	648	57	705

표5는 D3(통신 결합 상품의 이용 여부)와 인터넷전화 사용의향간의 관계를 나타낸 것

이다. ‘유선전화, 초고속인터넷, 인터넷전화, 이동전화’ 중 세 개 이상의 서비스를 같은 통신사의 상품으로 이용하는 경우가 가장 높은 인터넷전화 가입 의향을 보여주고 있어서 통신 결합 상품을 이용하는 고객의 경우 인터넷전화에 대한 잠재적 고객이 될 가능성이 높음을 암시한다.

표 5. 통신상품 결합 형태에 따른
인터넷전화 가입 의향

		인터넷전화 가입의향		전체
		없음	있음	
D3	사용 안 함	237	19	256
	유선전화-초고속인터넷 결합	174	3	177
	유선전화-인터넷전화 결합	0	1	1
	유선전화-이동전화 결합	74	5	79
	초고속인터넷-인터넷전화 결합	0	1	1
	초고속인터넷-이동전화 결합	76	8	84
	인터넷전화-이동전화 결합	0	3	3
	세 개 이상 결합	87	17	104
	전체	648	57	705

VI. 결 론

본 연구는 인터넷전화 사용의향에 대한 예측 모델을 개발하기 위한 것으로 기존 유선전화 서비스를 사용하고 있는 고객을 대상으로 실시한 설문에서 자료를 수집, 분석하였다. 응답자의 인구통계학적 정보와 통

신서비스 이용 정보를 분석하여 인터넷전화 이용의향에 관한 예측을 하였다. 이를 위해 새로운 변수를 추가 생성하였고, 예측 모델 개발을 위해 판별분석, 로지스틱 회귀분석, 분류나무, 신경망을 활용하여 각 모델의 성과를 비교하였다. 단일신경망이 전체 예측 오류율에서 6.82%로 최고의 성과를 보여주었고 판별분석(11.36%), 로지스틱 회귀분석(11.65%), 분류나무(11.65%), 혼합신경망(13.92%) 순으로 성과가 나타났다. 신경망 모델에서는 판별분석과 로지스틱 회귀모델에서 공통으로 선정된 변수들을 입력 변수로 사용할 경우가 (혼합모형) 성공 케이스에 대한 예측성과가 향상되는 것을 확인하였다(28.57%). 그러나 Type I 오류에 대한 비용이 클 경우 판별분석 모델이 최적의 모델로 평가되었고, Type II 오류에 대한 비용이 클 경우 단일신경망 모델이 최적의 모델로 평가되었다. 판별분석과 로지스틱 회귀, 분류나무, 단일신경망 모델에서 공통적으로 나타난 중요 변수는 D14와 D3로 통신요금과 통신결합 상품유형을 나타내는 변수들이었다.

향후 연구 과제는 다음과 같다. 첫째, 본 연구에서 이용한 자료는 설문의 특성상 응답하지 않은 문항에 대해서는 자료 값이 없었기 때문에 알고리즘의 실행을 위해서 기존 자료 값과 겹치지 않는 임의의 값을 입력하였다. 둘째, 모든 알고리즘에서 Type I 오류가 Type II 오류보다 월등하게 높게 나타나고 있다. 향후 연구에서는 Type I 오류를 줄일 수 있는 방법이 논의되어야 할 것이다. 특히 성공 케이스가 작을 경우 예측율을 높일 수 있는 방법들에 대한 논의가 필요하다.

참고 문헌

- 장태규, 김도영, 김영선, “BcN 인터넷전화(VoIP) 기술 동향,” 전자통신동향분석, 제 19권, 제 6호, 2004, 66-73.
- 권오상, 안재홍, “인터넷전화-시장, 요금, 규제,” 정보통신정책 ISSUE, 제 14권, 제 1호, 2002, 1-107.
- 김도환, “인터넷 전화의 품질보장제도 및 번호이동제도의 게임이론적 효과분석,” 경영학연구, 제 38권, 제 1호, 2009, 35-49.
- 김문구, 권수천, 박종현, “모바일 2.0 촉진을 위한 핵심 성공요인과 모바일 브로드밴드 전개를 수용 특성에 관한 연구,” 전자통신동향분석, 제 23권, 제 6호, 2008, 112-123.
- 박종현, 박희진, 백종현, “국내 유선 VoIP 이용 특성과 수용 영향요인 분석,” 전자통신동향분석, 제 23권, 제 3호, 2008, 163-174.
- 송관호, “인터넷전화 상호접속 개선을 위한 ENUM 도입방안과 전망-사업자 ENUM 시범사업 결과를 중심으로,” 한국통신학회지, 제 24권, 제 1호, 2007, 55-64.
- 이홍재, 김용규, 유제국, 통신서비스 수요예측 방법론, 정보통신정책연구원, 2000.
- 장범진, 나성현, 이은곤, 인터넷 전략 시장에서의 상품차별화 전략연구, 정보통신정책연구원, 2006.
- 정충영, 최이규, SPSSWIN을 이용한 통계분석, 무역경영사, 1998.
- 하성호, 양정원, 민지홍, “코호넨 네트워크와 생존분석을 활용한 신용 예측,” 한국경영과학회지, 제 24권, 제 2호, 2009,

- 35-54.
- 합창용, 광정호, 맹승찬, 나상우, 천병준, “VoIP 시장의 국내외 현황 및 시사점”, KISDI Issue Report, 정보통신정책연구원, 2007.
- 허명희, 이용구, 데이터마이닝 모델링과 사례, SPSS 아카데미, 2003.
- KISDI, 09년 상반기 통신시장 경쟁상황평가(소비자) 보고서, 정보통신정책연구원, 한국리서치, 2009.
- Altman, E.I., G. Marco, and F. Varetton, "Corporate Distress Diagnosis: Comparisons Using Linear Discriminant Analysis and Neural Networks", *Journal of Banking and Finance*, Vol. 18, 1994, 505-520.
- Breiman, L., J. Friedman, R. Olshen, and C. Stone, *Classification and Regression Trees*, Chapman & Hall/CRC, 1984.
- Buyut, V.C., S.H. Siadat, and W.Z. Abidin, "Electronic Customer Relationship Management for VoIP Service," *ICCEE: International Conference on Computer and Electrical Engineering*, 2008, 419-423.
- Cheng, B. and D.M. Titterington, "Neural Networks: A review from a Statistical Perspective," *Statistical Science*, Vol. 9, No. 1, 1994, 2-30.
- Desai, C.S., D.F. Convay, J.N. Crook, and G.A. Overstreet, "Credit Scoring Models in the Credit Union Environment Using Neural Networks and Genetic Algorithms," *IMA Journal of Mathematics Applied in Business Industry*, Vol. 8, No. 4, 1997, 323-346.
- Ezawa, K. and S. Norton, "Knowledge Discovery in Telecommunication Services Data Using Bayesian Network Models," In Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), 1995, 101-105.
- Jung, R., *Korea IP Telephony Service Forecast And Analysis, 2004~2008: 2003 Year End Review*, IDC, 2004.
- Kabiraj, S., "Electronic Customer Relationship Management: Origin and Opportunities," *IEMC: Engineering Management Conference*, 2003, 484-488.
- Kohli, R., F. Piontek, T. Ellington, T. VanOsdon, M. Shepard, and G. Brazel, "Managing Customer Relationships through e-Business Decision Support Application: A Case of Hospital-Physician Collaboration," *Decision Support Systems*, Vol. 32, No. 2, 2001, 171-197.
- Lee, T.S., C.C. Chiu, C.C. Lu, and I.F. Chen, "Credit Scoring Using the Hybrid Neural Discriminant Technique," *Expert Systems with Applications*, Vol. 23, No. 3, 2002, 245-254.
- Lingfen, S., E.C. Ifeachor, "Voice Quality Prediction Models and Their Application in VoIP Networks," *IEEE Transactions on Multimedia*, Vol. 8, No. 4, 2006, 809-820.
- Massey, A.P., M.M. Montoya-Weiss, and K. Holcom, "Re-engineering the Customer Relationship: Leveraging Knowledge Assets at IBM," *Decision Support Systems*, Vol. 32, No. 2, 2001,

155–170.

Peppers, D., M. Rogers, and B. Dorf, *The One to One Fieldbook*, Garden City, NY, Doubleday, 1999.

Rinde, J., "Telephony in the year 2005," *Computer Networks*, Vol. 31, No. 3, 1999, 157–168.

Shmueli, G., R.N. Patel, and C.P. Bruce, *Data Mining for Business Intelligence – Concepts, Techniques, and Applications in Microsoft Office Excel with XLMiner*, A John Wiley & Sons, 2007.

Tabachnick, C.B. and S.L. Fidell, *Using Multivariate Statistics* 5th, Pearson Educations, 2007.

Tseng, F.M., "Forecasting the Taiwan Customer Market for Internet Telephony," *Journal of the Chinese Institute of Industrial Engineers*, Vol. 22, No. 2, 2005, 93–105.

Zhang, G., M. Hillenbrand, P. Muller, "Facilitating the Interoperability among Different VoIP Protocols with VoIP Web Services," *First International Conference on Distributed Frameworks for Multimedia Applications*, 2005, 39–44.