

통신시장에서 데이터 마이닝 분석

장일동^o, 위승민

한국해양대학교 산업공학과, 해사수송과학부

idchang1@dreamx.net, weesm@hanara.kmaritime.ac.kr

The Analysis Telecommunication Service Market with Data Mining

Il-Dong Jang^o, Seung-Min Wee

Dept. of Maritime transportation science, Korea Maritime University

요약

이 논문에서는 지식발견과 데이터 마이닝에 관한 전반적인 소개와 고객이탈에 관한 것이다. 데이터 마이닝이란 과거에 수집된 데이터로부터 반복적인 학습과정을 거쳐 데이터에 내재되어 있는 패턴을 찾아내는 모델링 기법이며 통신서비스시장에서 데이터 마이닝 활용으로 고객이탈방지 모델을 인공지능망을 통해 구축하였다.

통신서비스시장의 경쟁이 심화됨에 따라 통신서비스 제공 업체가 고통으로 겪는 어려움 중의 하나가 고객이탈률이다. 따라서 데이터베이스에서 보다 가치 있는 정보를 찾아내 고객 이탈고객 분류의 적중률에 관하여 논의하였다.

다.

1. 서론

통신시장 다변화 정책과 정보화 사회시대의 급속한 발전에 힘입어 IMT-2000 통신시장은 현재 730만여명의 이동전화와 초고속인터넷 가입자가 있으며 과거 한국통신에 독점되었던 것이 국제전화, 시외전화, 이동전화, 무선호출 등 통신서비스가 과점 내지 경쟁체제로 전환되면서 통신서비스 시장에도 무한 경쟁이 벌어지고 있다.

이와 같이 통신서비스를 제공하는 업체가 늘어남에 따라 기업간의 생존을 위한 치열한 경쟁이 심화되고 고객을 위한 서비스가 한층 확대되고 있는 이때 각 업체는 무엇보다도 고객이탈을 줄이기 위한 방안에 고심을 하고 있다.

고객이탈방지를 위한 전략으로 요금인하, 장려금 증대, 단말기 무료 교환 등 다양한 방법으로 시장점유율을 유지하려는 정책적 대안이 바로 고객차별화(differentiation) 전략이다. 이러한 고객차별화전략을 세우고 의사결정을 쉽게 할 수 있도록 본 원고에서는 인공지능망분석을 통한 데이터 마이닝 분석으로 이탈방지 모델을 구축하였

2. 데이터마이닝 기본적인 이해

2.1 데이터마이닝의 지식 발견

데이터마이닝은 지식발견(Knowledge Discovery in Database: KDD) 프로세스 중에서 데이터로부터 정보를 추출하기 위해서 기법을 적용하는 단계로 실제적인 지식 탐사 및 발견 단계이다. 즉 데이터 마이닝은 어떤 가설 검증(verification)을 위한 방법이라기 보다는 지식의 발견을 가능하게 하는 것이라 할 수 있다[2].

현대의 엄청나게 방대한 데이터로부터 새롭고 의미 있는 정보를 추출하여 지식으로 활용하기 위해서 데이터 마이닝은 필수적인 요소라고 말할 수 있다.

데이터 마이닝을 하기 위해서는 데이터웨어하우스를 구축이 필수적이며 최적의 시스템이다[3]. 일반적인 시스템에서 GIGO(Garbage In Garbage Out)규칙이 적용되었듯이 데이터 마이닝에서도 마찬가지이다. 이 말은 데이터 작업을 하기 이전에 반드시 점검해야 할 사항이 데이터의 충실도이다.

충실도란 데이터를 구성하는 레코드와 속성의 수가 총 분란가에 대한 양적 척도와 속성이 취하는 값들이 정확한가 또는 빠진 속성 값은 어느 정도인가에 대한 질적 척도, 이 두 가지를 함께 의미한다[4].

2.2 데이터 마이닝 알고리즘

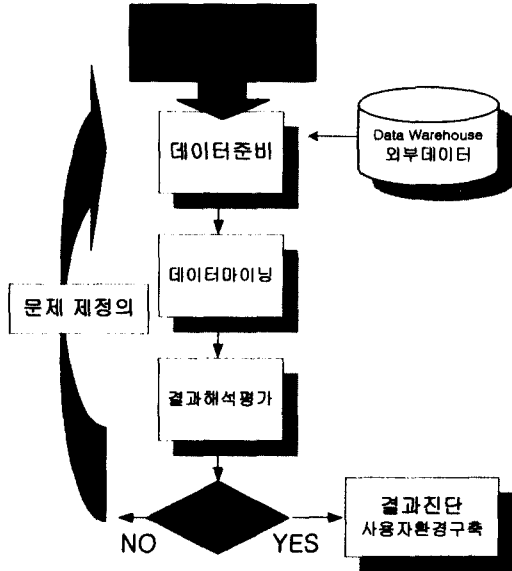


그림 1 데이터 마이닝 알고리즘

데이터 마이닝을 위한 알고리즘은 매우 다양하다. 분류(classification)을 위해서는 주로 결정트리[5]와 신경망이 사용된다. 추정(estimation)을 위해서는 주로 신경망이 사용되고, 예측(prediction)을 위해서는 결정트리, 신경망, 연관규칙탐사와 같은 방법이 사용될 수 있다[6].

분석 모형화(Modeling)에는 신경망분석(Multi-Layer Perceptrons), 로지스틱 회귀분석(Logistic Regression), CART(Classification And Regression Tree) 그리고 C5.0이 있다. 또한 군집화(clustering)을 위해서는 코호넨 네트워크(Kohonet Network), K-평균 군집분석 같은 알고리즘을 사용할 수 있다. 이들 중 어느 알고리즘이 우수하다고 말할 수는 없으며 데이터의 속성에 따라 보다 정확한 알고리즘 선택이 중요하다. 데이터 마이닝은 한번에 답이 도출되는 것이 아니며 반복적(iterative)인 작

업을 통해 여러 번 수행으로 수행결과에 근거하여 인수 값을 조절하거나 적용데이터를 보완하여 수 차례 반복적인 작업을 거쳐야만 만족한 결과를 얻을 수 있다는 것이다.

현재 신경망 분석은 비선형의 자료나 대용량의 자료 예측에 탁월한 능력을 보여, 현대 데이터 마이닝의 가장 대표적인 모형화 기법으로 각광을 받고 있다.

3. 신경망분석

신경망은 예측, 무리 짓기, 그리고 분류 과업에 사용되는 전형적인 알고리즘 가운데 하나이다. 신경망은 인간두뇌의 작용을 흉내낸 것으로서 단순하면서 수없이 많은 처리단위가 서로 연결되어 있는데 연결강도는 상황에 따라 변할 수 있다. 신경망은 실제 데이터로 구성된 '학습 데이터 (training set)'를 바탕으로 예측모형을 만든다[7].

신경망 모형은 모든 입력 변수와 목표변수가 0에서 1 사이 범위의 값을 취할 때 최적의 성능을 제공한다. 따라서 변수의 성격에 상관없이 모든 변수의 값을 0과 1사이의 값으로 변환시키는 것이 바람직하다.

은닉층을 포함한 망구조에서는 은닉층의 마디수를 증가시키면 분류능력이 향상되어 더 많은 패턴을 인식할 수 있지만 자칫 과잉 맞춤의 결과를 초래하기 쉽다. 따라서 적절한 은닉마디 수를 결정하는 것이 필요하며[8] 후진전파(Back Propagation)[9] 방법은 은닉층을 포함하는 구조이다.

신경망에서 은닉 노드의 수를 몇 개로 하는 것이 좋다는 규정은 없다. 다만, 적정 수준의 노드를 이용하여 모형을 만드는 것이므로 노드가 너무 많으면 overfitting의 문제 때문에 설명력이 있지만, 예측에서는 쓰레기가 되는 경우가 많다. 또한 몇 개의 층으로 할 것인가는 한 개의 Layer로 노드의 수와 연결함수, 활성화수를 잘 조절되지 필요이상의 Hidden Layer는 뉴럴 넷의 수행 속도에 저해만 가져다 줄 것이 분명하다.

4. 신경망의 실행 결과

4.1 코호넨 네트워크



그림 2 코호넨을 통한 군집화의 결과

그림 2에서 크게 00, 02, 20, 22집단에서 가장 많은 군집군을 이루고 있음을 알 수 있다. 즉 군집분석은 일종의 비슷한 성격을 가진 것끼리 모으는 작업을 의미하지만 많은 데이터를 한꺼번에 군집으로 묶을 수는 있는 것이 결코 올바른 군집의 형태로 보기는 어렵다. 동일한 성향을 나타내는 성격을 묶지만 몇 개 안되는 경우에는 군집을 묶을 필요가 없으며 아주 세세하지는 못하더라도 수많은 고객을 몇 개의 집단으로 묶고, 묶여진 집단의 공통 특성을 파악할 수 있다면, 미래의 예측이나 기획 전략을 수립하는데 상당한 이익을 가져다 줄 것이다.

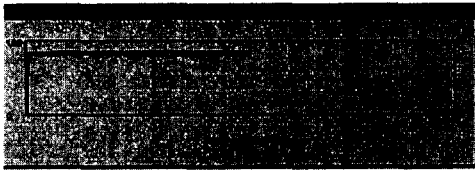


그림 3 신경망 모형의 Feedback Graph

많은 변수 중에서 고객의 특정 속성을 인공신경망에 적용함으로써 이탈한 고객의 유형을 추출하였으며 입력 변수가 단순화됨으로써 모델을 쉽게 이해 할 수 있다. 그림 3에서 볼 수 있듯이 인공신경망에 의한 이탈고객 분류 적중률은 84.57% 수준이며 상당히 높은 적중율을 보이고 있는 것으로 나타났다.

5. 결 론

통신서비스 시장에서 경쟁 치열할수록 대중마케팅이 지양되고 세분화마케팅과 틈새마케팅이 중요해졌다. 더 나아가 개별고객을 하나의 시장으로 보는 일대일 마케팅일 필요하게 되었으며 일대일 마케팅 방식으로서의 전환은 이러한 의미를 잘 반영하고 있다.

본 원고에서는 데이터 마이닝에 대한 개념과 신경망에 대해 살펴보았으며 일반적으로 마케팅 전략에 있어 시장 점유율(market share)도 중요하지만 현재 이탈고객을 예상하여 어떻게 하며 계속적으로 유지고객으로 확보하는가에 대한 마케팅 전략 개발에 주력해야 할 것이다.

참고문헌

- [1] Adriaans, P. and Zantinge, D., (유환승 역, 「데이터마이닝」, 그린, 1998.), Data Mining, Syllogic, 1996.
- [2] Syllogic B. V., "An introduction to th Syllogic Datat Mining Tool", white paper, 1996.
- [3] 유영일, Michael Park, 이재관, 최연준, 김병진 공역, "데이터웨어하우스 구축방법론", 홍릉과학출판사, 1997.
- [4] 장남식, 홍성완, 장재호 공저 "데이터 마이닝", 대청출판사, 2000, 43.
- [5] Quinlan, J. R., C4.5 , Programs for machine Learning, Morgan Kaufmann, San Mateo, California, 1993.
- [6] Michael J. A. Berry and Gordon S. Linoff, "Data Mining Technique", CIP, 1997.
- [7] Joseph P. Bigus, Data Mining with Neural Networks, McGraw-Hill, 1996.
- [8] Danie E. O' Leary, "Using Neural Networks to Predict Corporate Failure," International Journal of Intelligent System in Accounting, Finance & Management.7, P187-197, 1988.
- [9] Rumelhart et al. "Parallel Distributed Processing", MIT Press, Cambridge, 1986.
- [10] Eyden, R. J., "The Application of Neural Networks in the Forecasting of Share Price, "Finance & Technology, 1996.