

데이터마이닝을 이용한 eCRM

장형진*, 최성**, 한정란***, 이기민****

● 목 차 ●

- 1. 서론
- 2. 데이터베이스 마케팅관점에서의 고객관리
- 3. 데이터마이닝을 이용한 고객관리
- 4. 결론

요약

본 고에서는 인터넷 쇼핑물 기업들 중 신생기업들을 대상으로 이들의 기업환경에 맞는 데이터베이스 마케팅 방법론을 제시하고자 한다. 그러므로 데이터마이닝(Data Mining)을 이용하여 기존고객을 세분화한 다음 고객 개개인의 특성에 맞는 마케팅을 프로모션(Promotion)하고 신규고객을 획득할 때는 신규고객의 특성을 미리 예측하여 고객의 평생가치(LTV:Life Time Value)를 촉진하여 기업과 고객과의 관계성을 높이고, 기업은 안정된 고객층으로부터의 수익을 창출하고, 기업으로부터 더 많은 혜택을 받게 하는 것에 대하여 연구하였다.

1. 서론

1.1 데이터 마이닝의 정의

데이터마이닝이란 대량의 데이터로부터 쉽게 드러나지 않는 유용한 정보들을 추출하는 과정을 말

한다. 여기서 정보는 묵시적이고 잘 알려져 있지 않지만 잠재적으로 활용가치가 있는 정보를 말한다. 다시 말해 데이터마이닝이란 기업이 보유하고 있는 일일 거래자료, 고객자료, 상품자료, 마케팅 활동의 피드백 자료와 기타 외부자료를 포함하여 사용가능한 데이터를 기반으로 숨겨진 지식, 기대하지 못했던 패턴, 새로운 법칙과 관계를 발견하고 이를 실제 경영의 의사결정 등을 위한 정보로 활용하고자 하는 것이다.

1.2 데이터마이닝의 활용분야

데이터마이닝의 활용분야는 매우 다양하며, 현재에도 그 활용분야가 점차 늘어나고 있다. 그 중에서 특히 기업의 의사결정과 향후 대처에 대한 문제에서 많이 활용되고 있으며 주요 활용분야는 다음과 같다.

첫째 데이터베이스 마케팅으로서 데이터마이닝이 가장 성공적으로 적용되고 있는 분야로 소매, 통신판매, 금융서비스, 보험, 통신, 운송 등 다양한 분야에서 활용되고 있다.

둘째 신용평가로서 특정대상의 신용거래 대출한도를 결정하거나, 불량채권이나 기타의 금융문제에서 손실을 최소화하기 위한 것으로 활용되고 있다.

셋째 품질개선에서는 불량품을 찾고, 그 원인을

* (주)블루넷 대표이사

** 남서울대학교 컴퓨터학과 교수

*** 협성대학교 경영정보학부 교수

**** (주)아이뱅크코리아 대표이사

파악하여 궁극적으로 문제를 예방하는데 중점을 두고 있다.

넷째 부정행위적발의 목적은 고도의 사기행위를 발견할 수 있는 패턴을 알아내는 것이다.

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

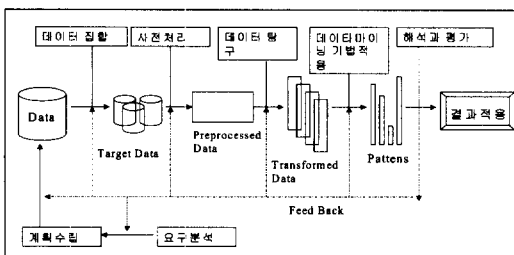
여섯째 과거로부터 현재에 이르는 판매관련 자료로부터 일정한 소비패턴이나 계절별 또는 월별 판매량의 변동 등을 파악, 모형화하여 미래 값을 예측하는 것을 목적으로 한다

1.3 데이터마이닝의 활용

인터넷 쇼핑몰 기업들 중 신생 기업을 대상으로 이들의 기업환경에 맞는 데이터베이스 마케팅 방법론을 제시하고자 한다. 신생 인터넷 쇼핑몰 기업들은 많은 양의 정보를 가지고 있지 못하기 때문에 많은 누적된 정보가 필요한 기존의 데이터베이스 마케팅 방법론들은 적절하지가 않다. 그러므로 누적된 정보가 1년정도 밖에 안되는 인터넷 쇼핑몰 기업의 경우 기존의 데이터베이스 마케팅 방법론에 비해 시간이 많이 단축되는 데이터마이닝을 이용하는 것이 좋다.

1.4 데이터마이닝의 절차

본 논문에서의 진행절차는 데이터마이닝 프로세스의 각 단계별 과정에 따르고자하며 데이터마이닝 프로세스의 각 과정은 아래와 같다.



(그림 1) 데이터마이닝 프로세스

대개 6가지 작업단계 중 다섯 번째 작업인 데이터 마이닝 기법 선택 및 적용에 역량을 집중하고 나머지 단계에 대하여 간과하는 경우가 많이 있지만 사실은 그러하지 않다. 왜냐하면 데이터의 충실도가 만족되지 않는다면 데이터마이닝 작업은 별 의미가 없으며 기대하는 결과도 얻을 수 없다.

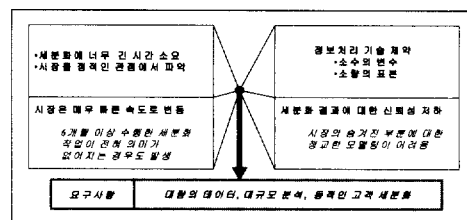
마케팅 목적으로 기존 고객에 대한 데이터베이스를 구축한 후에는 분할된 고객구성을 확실하게 이해하고 있어야 한다.

고객의 데이터베이스를 방대한 DB로만 이해하고 있으면 안된다. 이 고객들이 어떻게 구성되어 있는가? 수익성기여도 면에서는 어떤 구성을 갖고 있는가? 기타 구매 패턴 측면에서는 어떻게 세분화되는가? 등을 이해하고 있어야 활용방안을 생각할 수 있다. 현재의 인터넷 쇼핑몰은 국내의 벤처열기를 타고 급속도로 늘어나고 있는 추세이지만 아직까지는 제대로 된 고객관리가 이루어지지 않고 있는 경우가 많다. 거의 모든 기업이 매스 마케팅(Mass Marketing)에 의존하고 있는 형편인 것이다.

2. 데이터베이스 마케팅관점에서의 고객 관리

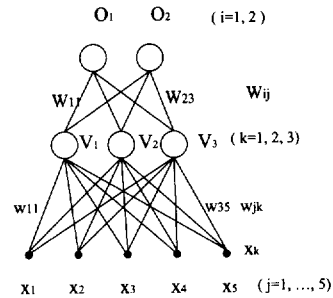
2.1 단일기준분석에 의한 고객관리

데이터베이스 마케팅(Database Marketing)이란 기업의 기존고객 혹은 가망 고객에 대한 데이터를 사내 전산 시스템에 축적하여 두고, 고객에 대한 개별적인 정보 속성을 고려하여 마케팅 활동에 적용하는 것을 의미한다.



(그림 2) 단일기준분석에 의한 고객 관리의 단점

기존의 데이터베이스 마케팅은 흔히 소비자행동 모델링을 만들어 이를 적용하여왔다. 가장 잘 알려진 고객분석방법은 R-F-M 분석방법으로 여기서는 거래의 최근성(recency), 거래빈도(frequency), 거래규모(monetary)등으로 단순화시킴으로써 고객이 얼마나 최근에 구입했는가, 고객이 얼마나 빈번하게 우리 상품을 구매했는가, 고객이 구입했던 총 금액은 어느 정도인가 등에 관한 정보를 축약하여 구입 가능성이 높은 고객들을 추려내기도 하고 상품, 연령, 지역, 구매액 등 단일기준에 대한 세분화를 할 수 있는 분석방법이다.



(그림 3) 다층신경망

2.1.1 단일기준분석에 의한 고객관리의 단점

R-F-M분석은 구축하는데 꽤 많은 시간이 걸리며, 많은 량의 누적 데이터가 있어야 좋은 R-F-M 모델을 구축할 수가 있다는 단점이 있다. 또한 모델링을 마치고 분석을 하였을 경우 신뢰성이 많이 저하된다. 그러므로 정교한 모델링이 어렵다는 것이 현실이다.

이러한 단점을 가지고 있기 때문에 신생 인터넷 쇼핑몰 기업환경에서 적절하지 못한 것이라 생각되어 대량의 데이터를 정교하게 모델링 할 수 있는 데이터마이닝 기법을 데이터베이스 마케팅에 적용하는 것이다.

2.2 신경망을 이용한 고객관리

신경망 모형은 신경생리학 분야에서 두뇌의 활동을 이해하고자 하는 목적 하에 신경의 작업을 설명하려는 시도에서 출발하여 생물학적인 프로세스를 컴퓨터를 이용하여 모형화 하려는 노력에서 비롯된 것이다.

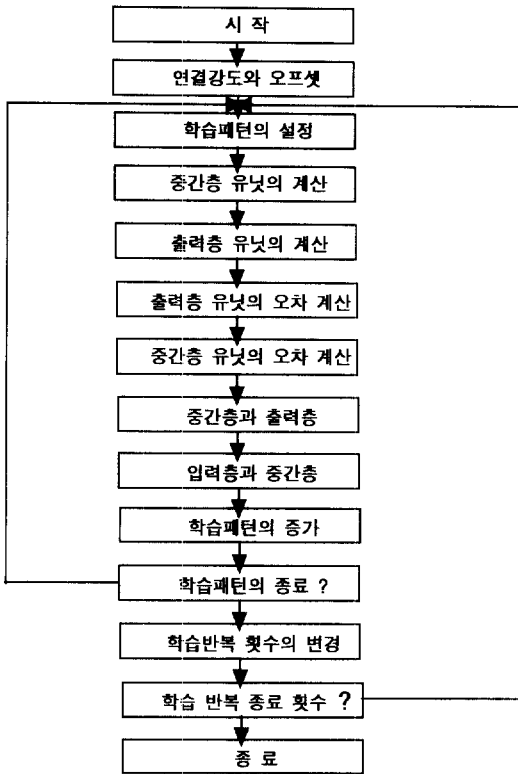
신경망은 은닉마디와 은닉계층을 몇 개로 하느냐에 따라 다양한 아키텍처를 구성할 수 있다. 특정문제나 데이터에 대해 최적의 아키텍처가 무엇인지 정해져 있지는 않다. 적용하고자 하는 문제 혹은 데이터가 얼마나 복잡한가에 따라 달라질 수

있다. 충분치 못한 아키텍처의 구성은 데이터에 내재된 복잡한 비선형적인 구조를 찾아내는데 불충분한 적합(Underfitting)이 될 수 있고, 반면 너무 많은 은닉계층이나 은닉마디는 과도적합(Overfitting)이 될 수 있다. 그러므로, 적절한 아키텍처를 구성하기 위해서는 어느 정도 시행착오가 수반되어진다.

많은 경우에 있어서 몇 개의 은닉계층을 두느냐 보다는 몇 개의 은닉노드를 두느냐 즉 가중치의 수를 몇 개로 할 것인가가 더 중요하게 고려되어야 할 사항이다. 따라서 이러한 신경망 기법 중 다층 퍼셉트론에서 가장 일반적인 방법인 오류역전파 알고리즘(Back Propagation)을 사용하기도 한다.

일단 설정한 아키텍처에 대해 트레이닝 데이터를 이용하여 트레이닝하고, 그 결과를 테스트 데이터에 평가해보고 적당한 아키텍처를 수정하여 다시 트레이닝, 그 결과를 평가하고 앞 단계와 비교하는 다소 반복적인 과정을 거치면서 만족할만한 아키텍처를 찾아나가게 된다. 접근방법으로는 우선 간단한 구성에서 시작하여 점차 복잡하게 구성해 나가는 방법과 반대로 크고 복잡한 구성에서부터 점차 줄여나가는 방법이 있다.

그리고 신경망을 지원하는 틀에서 기본적으로 제공되는 아키텍처를 중심으로 은닉노드(혹은 은닉계층)를 증가 혹은 감소시키면서 그 예측력을 비교해보고 적당한 아키텍처를 구성해 나갈 수가 있다.

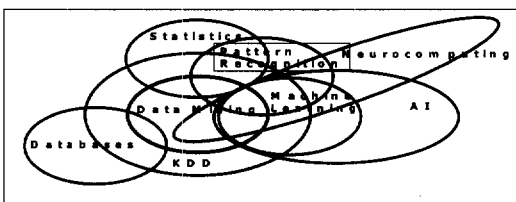


(그림 4) 오류역전파 알고리즘의 학습과정

3. 데이터마이닝을 이용한 고객관리

3.1 데이터마이닝의 구성 및 성격

데이터마이닝은 구성요소가 다양하다.



(그림 5) 데이터마이닝의 성격

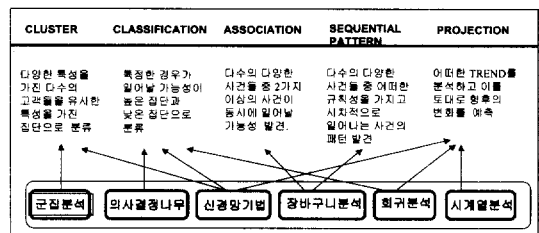
데이터마이닝은 알고자 하는 정보에 따라 작업 유형이 결정된다. 작업유형은 크게 연관규칙 (association), 연속규칙(sequence), 분류규칙(classification), 데이터 군집화(clustering) 등 4가지 유형으로 나누어진다. 그리고 이 네 가지 작업유형을 지

원하는 데이터마이닝 기법은 전통적인 통계기법, 의사결정나무(Decision Tree), 신경망(Neural Network), 동시발생 매트릭스(Co-Occurrence Matrix), K-평균 군집화(K-Means Clustering)기법 등이 있다.

<표 1> 데이터마이닝 작업 유형의 특성

이론적기법	마케팅 적용	마이닝 기법
시장바구니분석	연관규칙 탐사 연속패턴 탐사	통계, 확률
타겟마케팅 부정행위 적발 일탈관리	분류	의사결정수 신경망
주가예측	추정 혹은 회귀분석	통계, 신경망
시장세분화	군집분석	신경망 K-평균(Means)

다음 그림은 각각의 데이터마이닝 기법에 관한 특성을 설명한 그림이다. 본 연구에서 적용한 데이터마이닝 작업방법은 분류규칙이다. 분류규칙(classification)이라 하는 것은 데이터마이닝 에서 가장 많이 사용되는 작업으로 과거의 데이터로부터 부류별 특성을 찾아내어 분류모형을 만들고 이를 토대로 새로운 레코드의 부류값을 예측할 때 사용하게 된다



(그림 6) 데이터마이닝 기법의 특성

3.2 XML 유사성을 기반으로 하는 데이터마 이닝 (텍스트마이닝)

텍스트 마이닝 프로세스는 크게 두 단계로 텍스트 데이터로부터 특징을 추출하는 단계(Feature Extraction)와 흥미로운 패턴을 발견하기 위해 텍스

트 데이터를 분석하는 단계(Pattern Discovery)로 나뉜다.

첫 번째 단계에서는 자연어 처리기법이나 정보 검색의 여러 기법들이 혼용되어 사용될 수 있다. 이때 XML은 풍부한 의미적, 구조적 정보를 제공하는 특징을 가진 문서로서 문서 내에 사용되는 중요 태그를 명시적으로 표현할 수 있다. 이러한 풍부한 정보는 기존의 텍스트 마이닝 알고리즘을 변형하여 XML 문서를 위한 좀 더 정확한 알고리즘을 개발할 여지를 제공한다.

최근에 이르러 몇몇 텍스트 마이닝 기법을 개발 중에 있으나 의미와 구조정보를 그대로 살릴 수 있는 기법에 관한 연구는 이루어지고 있지 않다. 하지만 점점 더 많은 웹 문서들이 XML로 구성되고 있고 이를 자동으로 분석하고 분류하는 기술은 좀 더 XML의 문서가 가지는 정보를 최대한 반영할 수 있도록 해야 한다. 이를 위해 XML의 풍부한 의미정보를 그대로 추출하고 뿐만 아니라 태그의 의미를 최대한 확장한다. 이렇게 확장된 정보를 이용하여 구조적 유사성을 검사하여 자동으로 문서를 분류한다. 이는 문서 병합, 분류 및 관리 기술에 유용하게 사용될 수 있다

4. 결론

데이터마이닝에 의한 고객관리는 컴퓨터의 강력한 처리능력을 이용하여 실용화되고 있다. 또한 대다수의 Data Mining 기법들은 수학적으로 증명되고 발전된 것이 아니라 경험적으로 개발된 것이며, 기업의 다양한 의사결정 활동에 활용하기 위해서 사용된다.

Data Mining은 통계학, 전산과학, 인공지능, 공학 분야에서 개발되기 시작하였다. 그러나 실제로 이를 활용하는 전문가들은 경영, 경제, 정보기술 분야에서 배출되고 있다.

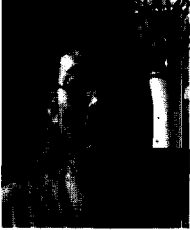
본 연구에서는 신생 인터넷 쇼핑몰의 환경을 고

려하여, 기존의 데이터베이스 마케팅과 달리 고객 관리 문제를 데이터마이닝 기법을 통해 기술하였다. 대용량의 데이터만 계속 쌓이고 있는 요즘상황을 개선해 나가기 위해서 데이터마이닝에 의한 쇼핑몰에서의 고객관리는 필수 불가결하다. 그러나 데이터마이닝은 오프라인 형태의 쇼핑몰과 달리 접속횟수(방문횟수)가 가장 중요한 변수가 되기 때문에 앞으로 이러한 자료가 계속 누적되고 웹페이지의 Hit수가 효율적으로 관리될 수 있는 설계가 이루어져야 한다.

참고문헌

- [1] 장남식, 홍성완, 장재호, "데이터마이닝", 대청, 1999
- [2] 조재희, 박성진, "OLAP 테크놀로지", 시그마 컨설팅 그룹, 1999
- [3] 노형진, "한글 SPSSWIN에 의한 조사 방법 및 통계분석", 형성출판사, 1999
- [4] 원태연, 정성원, "한글SPSS 통계조사 분석", SPSS 아카데미, 1999
- [5] 박난희, 장기화, 박희창, 우용태, "WWW 접근패턴 분석을 위한 데이터마이닝모델 설계", 99 한국 데이터베이스 학술대회 논문집
- [6] 정인환, "데이터베이스 마케팅의 전략에 관한 연구", 숭실대학교 학위논문, 1998
- [7] 조용준, 최인규, "Neural Connection을 이용한 데이터마이닝 신경망 분석", SPSS 아카데미, 1999
- [8] 송현수, "통합DBMS", 새로운 제안, 1998
- [9] 정재호, "인공 신경망을 이용한 신용카드 부정 사용 색출", 한국과학기술원, 1998
- [10] 피터아드리안, "데이터마이닝", 그린출판사, 1998
- [11] 김정수, "통합 데이터베이스 마케팅 시스템", 범우사, 1997

저자 약력



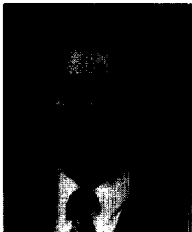
장형진

서울대학교 졸업
 서울대학교 물리학과 박사과정 수료
 한국정보처리전문가협회 정회원
 정보처리학회 편집위원
 도산 뉴벤처포럼 부회장
 홍사단 서울지부 회원
 현 블루넷 대표이사



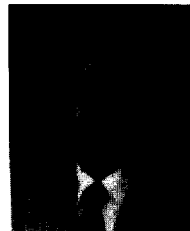
한정란

1985년 이화여대 전산과 졸업
 1987년 동대학원 전산과 프로그래밍 언어론 전공 이
 학석사
 1999년 동대학원 컴퓨터학과 프로그래밍 언어론 전
 공 공학박사
 1999년-현재 협성대학교 경영정보학부 교수
 관심분야: 전자상거래, e-Business, Java 응용, 인터넷
 프로그래밍



최성

1975년-1980년 기업은행 전산개발부
 1983년-1985년 제주은행 전산실장
 1986년-1994년 한국생산성본부 OA추진 사무국장
 1999년 강원대학교 대학원 컴퓨터학과 이학박사
 1983년 연세대학교 산업대학원 전자계산학과 공학석사
 1994년-현재 남서울대학교 컴퓨터학과 교수
 현재 한국정보처리전문가협회 이사, 한국정보통신 기
 술사회 이사
 관심분야: EC, ERP, S/W Engineering, VR영상게임 등



이기민

1993년 광주대학교 전자계산학과(공학사)
 1996년 경희대학교 산업정보대학원 전자계산학과(공
 학 석사)
 2001년 동국대학교 전산통계학과
 - 현재 박사 과정
 1981년-1998년 금융결제원 정보시스템부
 1998년-2001년 (주) 티엘정보통신 기술개발 연구소장
 2000년-현재 (주) 아이뱅크코리아 대표이사
 관심분야: 프로그래밍 언어론, 컴파일러, 시뮬레이션,
 소프트웨어 개발방법론