

고객ID 통합구조에 기반한 고객이탈방지 마이닝 모델

김혜정^o 임정연 성진동
KT 서비스개발연구소, KT 정보시스템본부
(elliis^o, horang21, jdsung)@kt.co.kr

Churn Management Minging Model based on Integrated Customer ID

Hye-Jung Kim^o, Jung-Yeon Lim, Jin-Dong Sung
Service Development Laboratory, KT

요 약

CRM이 기업의 핵심 경영전략으로 도입되면서 기업이 보유하고 있는 고객데이터를 수집, 통합, 가공, 분석하여 마케팅을 위해 활용하고자 하는 시도가 계속되고 있다. 특히, 기존고객의 유지 전략과 기존고객을 활용한 신상품 유도 전략이 중요한 이슈로 대두되면서 마이닝을 통한 CRM관점의 고객이탈방지는 각 통신사에서 지속적으로 추진하고 있는 분야이다. 본 연구에서는 KT의 고객이탈방지 모델 구축을 사례로 효율적인 마이닝 모델 구축을 위한 고객통합구조를 제안하고자 한다. 그리고, 고객이탈방지 모델 구축의 전처리 과정으로 고객통합구조를 적용하여 고객중심의 변수 도출, 이용형태 추적 등을 통해 의미 있는 해지변수를 찾아내는 방법과 그 효과에 대해 기술한다.

1. 서론

다른 산업분야보다 빠르게 변화가는 통신업계에서 신상품 출시와 더불어 시장 선점 및 우위 확보 경쟁이 날로 심화되고 있으며 기존고객의 유지 전략과 기존고객을 활용한 신상품 유도 전략이 중요한 이슈로 대두되고 있다. 특히 데이터 마이닝이 소개된 이래 마이닝을 통한 CRM관점의 고객이탈방지는 각 통신사에서 지속적으로 추진하고 있는 분야이다. 본 논문에서 제안한 고객통합구조는 운영계의 고객기본정보, 청구정보, 고객계약구조, 서비스별 계약정보 등을 통합하여 고객중심으로 재설계함으로써 마이닝 모델 구축시 고객중심의 변수 도출에 유용하게 활용하고자 하는 것이다. 본 논문은 다음과 같이 구성된다. 2장에서 데이터베이스 마케팅(DBM)과 데이터마이닝의 활용영역에 대해 살펴본다. 3장에서는 KT에서 실행한 고객 이탈방지 프로젝트를 사례로 데이터 마이닝을 위한 해지방어 업무 기반 구축 과정을 소개한다. 마이닝 모델 구축과정에서의 어려움과 해결방안을 통해 4장에서 전처리 과정으로 고객통합구조를 제안한다. 그리고 5장에서 각종 데이터를 통해 의미 있는 이탈변수를 찾기 위해 고객통합 구조에 기반한 탐색적 데이터 분석과 이에 따른 효과에 대해 기술한다.

2. DBM과 데이터마이닝의 활용영역

2.1 DBM

기업에서는 마케팅과제 해결을 위해 기업이 보유, 축적하고 있는 데이터를 활용한다. 기존의 통계 및 인공지능 기법으로 분석하여 예전에 알지 못했던 정보를 찾아내 의사결정에 도움을 주는 방법론이다.

이를 위한 기법과 활용영역은 다음과 같이 연관지을 수 있다.[1]

표 1. DBM기술과 활용영역

기술	활용영역
Decision Tree	Customer Acquisition
Logistic Regression	Churn Management, Customer Retention, Cross/Up Selling, DM, TM
Association Rule	Customer Segmentation & Profiling
Clustering & Segmentation	Winback, Forecasting
Neural Network	Market basket Analysis, Association

2.2 데이터마이닝의 활용영역

위와 같이 DBM을 위한 여러가지 기법들을 이용한 데이터마이닝이 고객이탈방지(Churn Management)에 활용되고 있다. 고객은 이탈을 고려 중일 때 일시적 혹은 지속적으로 이탈 징후를 보이는데 이것은 연관된 고객 데이터의 분석이나 의미 있는 이탈변수에 근거한 해지모형에 의해 감지할 수 있다. 연관된 고객데이터의 분석은 OLAP을 이용한 고객세분화에 의해서도 가능하다. 그러나 이는 마케터나 분석가의 감에 의존하는 경향을

가지며 연속된 변수를 핸들링 하는데 한계가 있고 향후 고객 행동에 대한 예측요인에 의한 세분화가 불가능하다. 반면 데이터마이닝은 고객의 행동 예측 모델에 의해 고객을 세분화 한다. Clustering, Classification, Regression 등의 분석기법을 이용하여 고객 세분화를 실시하며, Decision Tree, Neural Network 등 다양한 모델링 기법들을 활용할 수 있다. 데이터 이해, 추출, 탐색 등 각 절차상의 작업은 이미 구축했거나 구축 예정인 데이터 웨어하우스의 정제된 데이터를 이용했을 때 제 효과를 발휘할 수 있다. 그뿐 아니라, 데이터웨어하우스를 관리하는 데 중요한 메타 데이터를 원활히 이용할 수 있는가의 여부가 성공적 작업의 결정적인 요인이 된다.[2]

3. KT의 메가패스 고객이탈방지 모델링 적용사례

3.1 배경

KT의 메가패스 고객은 2002년 7월 현재 가입자가 440만에 이르고 있으며 상품출시 시점인 1999년 초부터 지속적인 성장 추세를 유지하고 있다. 타 통신업체와의 경쟁에 따라 월 해지율이 2.2% 정도에 이르고 있으며 이들 해지고객은 실제로 다양하게 분류할 수 있다.[3] 따라서 해지고객에 대한 정확한 정의와 함께 고객분석을 통한 이탈변수를 도출하고 해지 예측모형을 구축하여 예상고객을 추출함으로써 해지방어 캠페인에 적용하고자 하는 것이 모델 구축의 목적이다.

3.2 분석환경 및 모델구축과정

고객분석에 사용된 데이터는 품질 관련된 변수인 민원(VOC) 정보, A/S요청신고 여부/횟수/유형, 5분내 재접속 비율, 사용시간(접속횟수/기간) 등이며 마케팅 변수인 약정기간 잔여일수, 이용정지/일시정지 여부 및 횟수, O14XY 이용여부 등이다. KT의 영업계, 시설계 등에 흩어져 있는 메가패스 고객 관련 정보들을 수집하여 고객중심의 분석작업을 하기 위해서 다음과 같은 분석 환경을 구축하였다.[3]

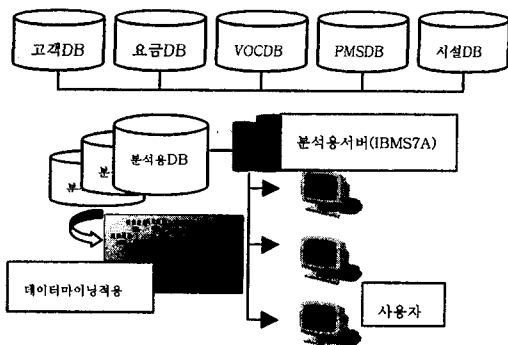


그림 1. 마이닝 모델 구축환경

그림1. 에서처럼 분석용 서버에 영업계, 시설계의 누적 데이터 1년분을 저장하고 고객중심의 분석작업을 위해

분석용 ID를 생성하는 작업이 선행되었다. 마이닝 모델을 구축하기 위해 고객 개개인의 행태분석을 위해서는 다음과 같은 절차가 필요하다.

- 고객을 파악하기 위한 기준정보를 활용하여 복수개의 상품을 사용하는 단일 고객의 가치를 산출하여야 한다.
- 영업관련 데이터 뿐 아니라 시설, 민원정보의 데이터를 통합하여 고객단위의 분석이 가능하여야 한다.
- 서비스 전환, 해지 후 재가입, 허수가입/해지 등 고객의 변동이력에 대한 추적이 가능하여야 한다.

이러한 절차를 통해서 고객단위의 각종 이력을 추적하고 행태를 분석함으로써 마이닝 모델에 필요한 의미있는 변수의 도출이 가능하다.

여러 변수를 통해 해지징후를 분석한 결과 다음과 같은 해지예측모형에 투입할 변수리스트를 도출하였다.[3]

표2. 해지모형 투입변수 리스트

구분	활용영역
고객/요금 정보	납부형태, 체납여부/횟수, 요금조정여부/횟수, 이용정지횟수
네트워크정보	5분내 재접속 비율, 주평균 로그온 횟수, 주평균 사용시간
민원정보	불만여부/횟수, VOC처리기간
고장정보	고장신고횟수, 고장처리기간, 동일고장 재신고 여부

그리고 위와 같은 변수들을 통해 지속적이고 주기적인 데이터 갱신과 함께 고객의 행태변화를 모니터링 함으로써 안정적인 해지모델 구축과 해지방어를 위한 Outbound 캠페인을 기획할 수 있었다.

4. 고객ID 통합 구조와 데이터 마이닝

4.1 고객통합구조의 필요성

어느 회사나 CRM전략수립과 실행 초기 단계로서 고객DB를 통합하고 DW와 OLAP 시스템을 구축한다. 그리고 마케팅부서에서는 우수고객관리, 잠재고객 추출, 경쟁시장 방어 등을 위해 세분화된 고객정보를 요구한다. 실제로 고객 세분화는 고객을 세분화 할 수 있는 기준, 그리고 그에 따른 정확한 데이터 없이는 불가능하며 어떤 표준화된 형식으로 주기적인 세분화 정보를 제공하기에는 어려움이 많다. CRM을 추진하고 있는 많은 회사와 마찬가지로 KT에서도 고객에 대한 세분화 기준이 되는 데이터 확보와 데이터 클린징, 데이터 갱신 등의 문제로 고민하고 있다. 고객을 분류할 수 있는 명확한 기준은 무엇인가? 하는 문제로부터 시작하여 고객의 여기저기 흩어져 있는 데이터들을 묶어 하나의 의미있는 정보로서

다듬어 나가는 것만이 CRM과 마이닝 성공의 시발점이 될 것이다. 이탈방지 모델 구축을 통한 캠페인 실행과 같은 운영계, 분석계 CRM에 있어서 고객통합을 통한 고객의 세분화는 필수적인 작업이다. 더욱이 데이터 마이닝 모델의 성공적인 구현은 고객중심의 행태분석과 이를 반영하는 변수의 도출, 그리고 적합한 모형개발에 있다. 따라서 영업과 시설의 각 시스템에서 각각 운영되고 있는 고객을 통합하여 단일 고객으로 볼 수 있는 ID체계를 갖는 것이 고객 세분화, 등급화를 통한 고객가치 산출을 용이하게 하며 마이닝을 통한 효과를 기대할 수 있게 하는 부분이다. 고객ID의 통합은 선결되어야 하는 작업이나 반드시 운영계에서 이루어져야 하는 것은 아니다. 기존 운영계 시스템에 미칠 영향과 교체, 유지 비용 등을 고려한다면 정보계에서 운영계의 각 데이터를 통합하고 통합된 고객정보를 운영계에서 활용할 수 있도록 하는 방법도 효과적이라 할 수 있다. 기존의 상품중심으로 흩어져 있던 가입정보, 이용정보, 민원정보, 시설정보를 통합하여 정보계에서 고객중심으로 재구조화 하는 것이다.

4.2 고객ID 통합구조

정보계의 ID통합은 운영계에서 각 서비스별로 부여하고, 동일고객이라도 고객명에 따라 달리 부여하던 고객ID를 식별번호(주민번호, 법인번호) 기준으로 통합하여 고객관점의 정보관리를 가능하게 하는 것이다. 그림2.에서 보는 바와 같이 운영계의 ID와 연결고리를 가지면서 정보계에서 고객 세분화나 고객 타겟팅 등 CRM관점의 마케팅 정보로 활용할 수 있도록 한다.

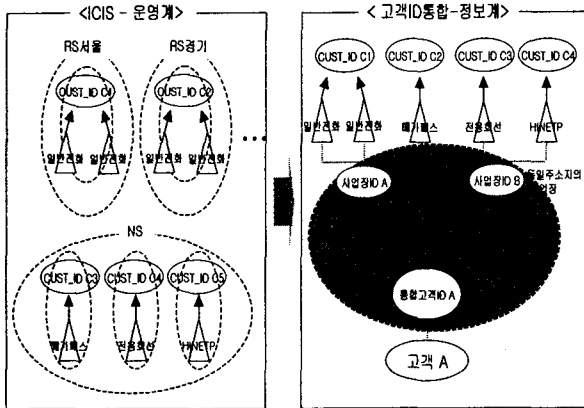


그림2. 고객ID통합 구조

그러나 통합고객ID는 시설계에 있는 데이터에도 동일한 기준으로 적용시키기에는 한계가 있다. 시설관점의 고객은 가계고객(예) 유선전화)의 성향을 가지고 있으며 고객의 이력을 추적해 나가거나 실사용 고객 관점에서 고객의 성향을 반영하기에는 어려움이 많다. 따라서 운영계와 시설계의 고객을 하나의 VIEW로 통일시켜 볼 수 있는 기준값을 설정하는 것이 진정한 고객통합의 과제이다.

4.3 전처리를 포함한 마이닝 절차

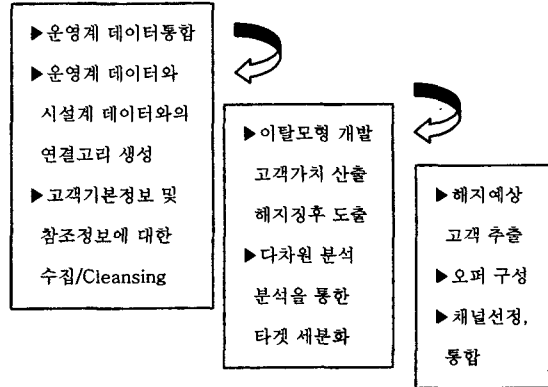


그림 3. 마이닝 절차

위와 같은 전처리 과정을 통해 고객의 행태이력 추적이 가능해짐에 따라 마이닝 모델 구축이 용이하고 보다 정확해진다. 즉, 데이터마이닝을 위한 데이터 탐색은 곧 데이터 통합의 문제로 귀결될 수 있다. 마이닝을 통한 고객세분화나 마이닝을 통한 모델 구축으로 캠페인을 실행할 경우 일반적인 프로세스 전처리 단계에 데이터 통합과정이 요구된다.

5. 결론

수많은 기업에서 CRM시스템 도입을 결정하고 캠페인을 위해 마이닝 TOOL을 활용하여 고객 세분화를 실행하고자 할 때 간과해서는 안될 것이 있다. 그 회사의 고객 데이터가 과연 마이닝을 적용시켰을 때 정확한 예측값을 뽑아낼 수 있을 정도로 정제되어 있는가? 정제할 수 있는가? 그리고 고객관점의 일관된 VIEW를 적용시킬 수 있을 만큼 통합되어 있는가? 하는 것이다. 무작정 마이닝 모델 구축을 시도하다가 데이터 탐색과정만을 반복하는 것으로 프로젝트를 종결 시키는 경우가 종종 있다. 이러한 실패를 예방하기 위해서, 그리고 성공적인 마이닝 프로젝트를 위해서는 통합된 고객구조가 필수적이다. 고객 통합구조는 운영계, 시설계의 데이터 뿐 아니라 민원정보, 건물정보, 자회사의 고객 데이터를 포함시킬 수 있을 만큼 확장성을 갖고 있어야 하며 이러한 고객 통합구조를 통해 마이닝은 효과를 발휘할 수 있다.

<참고문헌>

[1] 장남식, 홍성완, 장재호, "데이터마이닝", 대청, 1999
 [2] S.C Hui, G. Jhi, Data mining for Customer Service Support, Information&Management 2000 1-13, 1999
 [3] KT 경영연구소, "메가패스 해지방지 전략수립 최종보고", 2001
 [4] 피터 아드리안, "데이터마이닝", 그린출판사, 1998