

데이터 마이닝을 이용한 CRM

전남대학교 이도현*

1. 서 론

CRM(Customer Relationship Management)이란 말 그대로 고객과의 관계를 효과적으로 관리하는 기법을 의미한다. 굳이 CRM이라는 특별한 용어를 사용하지 않더라도 고객과의 관계를 효과적으로 관리하기 위한 노력은 거의 모든 고객 상대 업무 현장에서 이미 쉽게 찾을 수 있다. 소규모 소매점의 경우 과거의 판매 경험을 바탕으로 단골 고객이 선호하는 브랜드 혹은 상품을 기억하고 있기 때문에 해당 고객이 방문하면 자세히 얘기하지 않아도 원하는 상품을 추천하는 일은 이미 일상화되어 있다. 단골을 주고객으로 하는 이발소의 경우 머리 모양을 지정하지 않아도 의자에 앉기만 하면 이발사가 스스로 알아서 머리를 깎아준다. 만약 그러한 고객이 다른 이발소를 가게 되면 머리 모양도 지정해야 되고, 과연 자신이 원하는 바에 맞게 해 줄 것인가 확신도 서지 않기 때문에 특별한 문제가 없는 한 그 고객은 계속해서 그 이발소를 찾게 된다. 자동차 영업 사원들은 좀더 정교한 고객 관계 관리를 구사한다. 자신을 통해 자동차를 구입한 고객이 언제 어떤 부품을 교체해야 할 것인지 언제 어떤 점검을 받아야 할 것인지 미리 알아서 통보해 주고 그런 서비스를 통해 얻은 신임을 바탕으로 자동차를 바꿀 때 다시 자신을 통해 구입이 이루어지도록 유도한다. 또한 처음 만나는 고객의 경우에도 그 고객과 비슷한 직업, 연령대의 다른 고객들의 취향을 근거자료로 판단하여 그 고객이

관심을 가질만한 차종과 선택사양을 추천하기도 한다.

이러한 고객 관계 관리의 과정을 요구 인지(recognition), 요구 기억(memory), 전략 학습(learning)의 삼 단계로 나누어 볼 수 있다[1]. 먼저 요구 인지 단계는 고객과의 상호 작용을 통해 고객이 요구하는 것이 무엇인지를 알아내는 것을 의미한다. 둘째, 요구 기억 단계는 요구 인지 단계에서 획득한 정보를 기억하는 것을 의미한다. 마지막으로 전략 학습 단계는 기억한 정보를 바탕으로 보다 효과적인 영업 전략을 세우는 것을 의미한다. 앞서 예시한 소규모 비즈니스의 경우에는 처리해야할 정보의 양이 적기 때문에 수작업 혹은 간단한 소프트웨어를 이용하여 이러한 삼 단계 과정을 수행할 수 있다. 하지만 수만에서 수백만의 고객을 상대로 하는 대형 비즈니스의 경우에는 그리 간단한 일이 아니다.

하지만, 지난 수십년간 정착된 OLTP 시스템의 보급, 최근 급속히 보급되고 있는 데이터웨어하우스 시스템과 OLAP, 데이터 마이닝 솔루션에 힘입어 대형 비즈니스에서도 과거 소규모 비즈니스에서나 가능하던 정교한 고객 관계 관리를 하는 것이 가능해졌다. 비즈니스의 거의 모든 분야를 처리하고 있는 OLTP 시스템에 의해 시시각각 생성되는 고객의 요구를 인지하는 효과가 부수적으로 얻어지며, OLTP 데이터가 데이터웨어하우스에 누적됨에 따라 요구 기억 단계 역시 가능하게 되었다. 마지막으로 OLAP 및 데이터 마이닝 기술의 적용에 의해 전략 학습 역시 자동화할 수 있게 되었다.

이러한 배경에 의해 하나의 독립적인 기술 분

* 정회원

야로 떠오르게 된 CRM은 접근하는 관점에 따라 다양하게 정의된다. 아직 대표적인 정의로 공감대를 형성하고 있는 것은 없지만, 잠정적으로 다음과 같이 정의하자. CRM 시스템이란 고객에 대한 정확한 이해를 바탕으로 고객이 원하는 상품과 서비스를 지속적으로 제공하여 고객을 오래 유지시키고 결과적으로 고객의 평생 가치(life time value)를 극대화하여 수익성을 높일 수 있도록 하는 시스템이다.

온전한 의미의 CRM 시스템으로서 인정받을 수 있기 위해서는 다음과 같은 요소를 갖추어야 한다[2]. 첫째, 비즈니스 인텔리전스 및 분석 능력이다. 고객 신상 정보, 상품 구매 이력, 콜 센터 상담 이력 등을 분석하여 구매 성향, 이탈 가능성과 같은 정보를 추출할 수 있는 자동화된 도구를 갖추어야 한다. 둘째, 통합 채널을 통한 고객과의 상호 작용이다. 예를 들어 회사의 대표 콜 센터 혹은 웹 사이트를 통해 제품 상담, 주문, 유지보수 등 고객과의 모든 상호 작용이 처리될 수 있도록 해야 한다는 것을 의미한다. 세 번째, 중앙화된 고객 정보 저장소가 필요하다. 부서별 혹은 상품 및 서비스별로 별개의 고객 파일 혹은 데이터베이스를 구축하는 것이 기존 환경에서의 일반적인 상황이다. 하지만, 온전한 의미의 CRM 시스템을 구현하려면, 별개의 고객 파일 혹은 데이터베이스를 통합하여 중앙화된 고객 정보 저장소를 구축해야 한다. 네 번째, 통합된 워크플로우가 필요하다. 업무별로 별개의 워크플로우가 존재한다면, 일관성있고 중복이 없는 고객 관계 관리가 곤란하므로, 통합된 형태의 워크플로우를 구축해야 한다. 따라서, 부분적인 혹은 전사적인 비즈니스 재공학(reengineering)이 필요하게 된다. 마지막으로 전사적 자원 관리(Enterprise Resource Planning:ERP) 시스템과의 통합을 들 수 있다. 고객 정보 역시 중요한 회사의 자원이므로 ERP와 별도로 구분될 수 없다. 따라서, CRM 시스템과 ERP 시스템간의 정보 공유 및 유통이 효과적으로 이루어져야 한다.

CRM 시스템은 이제 선택이 아닌 기업 활동의 필수 요소로 자리잡고 있다. 심지어는 미국의 한 카지노 업체조차 다양한 고객의 패턴을 분석하여 차별화된 고객 마케팅을 하고 있는 실정이다. 한국 소프트웨어 산업협회에 따르면 2000년

1/4분기 국내 CRM 시장 규모는 약 800억원에 달하며, Metagroup 보고서에 따르면 전 세계 CRM 시장 규모는 올해 약 50% 성장세를 기록했다. 이러한 시장 확대에 힘입어 많은 국내의 업체들이 CRM 솔루션을 개발, 공급하고 있다. 현재 국내에서 활동하고 있는 CRM 솔루션 공급 업체를 크게 네 가지로 구분할 수 있다[2]. 첫째 그룹은 데이터 웨어하우스 확장형이라고 볼 수 있다. 데이터 마이닝, 데이터 웨어하우징, 데이터 마트, OLAP, 위험 관리(risk management) 기능 등을 제공하며 주로 금융권에 고객을 형성하고 있다. 대표적인 업체로는 인포믹스[3], 오라클 [4], 사이베이스, SAS를 들 수 있다. 두 번째 그룹은 ERP 확장형이라고 볼 수 있다. 이 그룹은 기존의 ERP 시스템에 고객 관계 관리 기능을 결합한 형태로서 SAP, Oracle, Vantive 등의 업체가 여기에 속한다. 세 번째 그룹은 E-CRM형으로서 E-Business에서 출발하여 고객관계 관리를 추구하는 그룹이다. 주로 웹 로그 분석을 통한 원투원 마케팅, 개인화된 웹 페이지등을 제공하는 인터넷 벤처기업이 여기에 속하며, 미국의 BroadVision 제품군이 현재 세계 시장을 주도하고 있다. 마지막 그룹은 CTI(Computer Telephony Integration) 확장형이라고 볼 수 있다. 기존의 콜센터 기능을 확장 개편하는 방식으로 고객 관계 관리를 추구한다. 따라서, 앞서 기술한 비즈니스 인텔리전스 측면 보다는 워크플로우 통합의 측면이 강한, 부분적인 CRM 솔루션이라고 할 수 있다.

2. CRM 시스템의 구성

CRM 시스템은 고객 관계 관리에 필요한 다양한 기능 모듈의 복합체라고 볼 수 있는데, 크게 전방 시스템(Front Office System)과 후방 시스템(Back Office System)으로 나누어 볼 수 있다[5]. 전방 시스템은 고객과 직접 상호 작용을 하는 모듈의 집합체를 의미하며, 후방 시스템은 고객 관계 관리에 관련된 데이터의 분석 및 가공을 담당하는 모듈의 집합체를 의미한다. 그림 1은 CRM 시스템을 전방 시스템과 후방 시스템으로 구분한 개념적인 구조를 도식하고 있다.

CRM 전방 시스템에 속하는 모듈은 매우 다양하지만, 대표적으로 콜 센터(call center), 판매

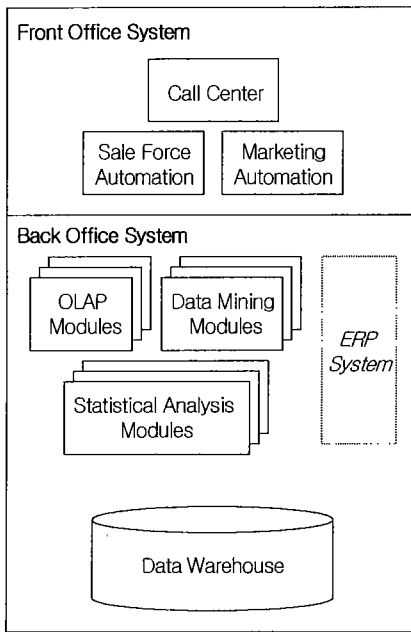


그림 1 CRM 시스템의 개념적인 구조도

자동화(Sale Force Automation:SFA) 서브 시스템, 마케팅 자동화(Marketing Automation:MA) 서브 시스템 등을 들 수 있다. 먼저, 콜 센터는 고객으로부터 걸려오는 전화 통신을 전담하는 채널이다. 단순한 상품 및 가격 문의로부터 시작하여, 주문, 결제, 유지보수 요청과 같은 다양한 업무를 처리할 수 있다. 고객의 편의성을 위해 단일 전화번호를 사용하면서 동시에 충분히 많은 수의 전화 접속을 처리할 수 있는 다중 접속 구조를 갖고 있는 것이 보편적이다. 고객으로부터 전화가 오면 일단 적절한 내용의 전화인지 판정한 후, 담당자에게 포워드(forward)하며, 전화 내역을 로그에 기록한다. 고전적인 의미의 콜 센터는 주로 우편 주문 판매 회사, TM(Tele-Marketing) 회사 등에서 운영하는 전화 접수 창구를 의미하지만, 최근에는 회사 웹사이트를 통한 고객 대응 창구도 콜 센터의 범주에 포함된다고 할 수 있다.

판매 자동화(Sale Force Automation:SFA) 서브 시스템은 계약 혹은 주문으로부터 회계, 배송까지 판매에 해당되는 전 과정을 자동화하는 시스템이며, 마케팅 자동화(Marketing Automation:MA) 서브 시스템은 마케팅 캠페인과 기

타 마케팅 관련 활동을 설계하고 시행 평가하는 시스템을 의미한다[2]. 이와 같이 CRM 전방 시스템은 주로 고객 관계 관리의 표면적인 운영을 담당하기 때문에 운영적(operational) CRM이라고 불리기도 한다.

CRM 후방 시스템을 구성하는 대표적인 요소는 데이터 웨어하우스, OLAP(Online Analytical Processing) 모듈, 데이터 마이닝 모듈, 통계적 분석 모듈이다. 데이터 웨어하우스에는 고객 관계 관리에 필요한 모든 정보가 축적, 관리된다. OLAP 모듈은 다양한 차원 및 차원 조합에 대하여 각종 정보를 집계, 가시화함으로써, 데이터에 대한 이해를 쉽게 해준다. 데이터 마이닝 모듈은 데이터에 내재된 패턴 예를 들어 고객의 구매 성향, 우량 고객의 대표적인 특성과 같은 정보를 추출한다[1, 6]. 통계적 분석 모듈은 기존의 통계적 분석 기법을 이용하여 가설 검증, 회귀 분석과 같은 작업을 처리한다. 이처럼 CRM 후방 시스템은 주로 자료의 분석에 주안점이 있기 때문에 분석적(analytical) CRM이라고 불리기도 한다. 본 논문에서는 주로 분석적 CRM에 초점을 맞춘다.

또한 전사적 자원 관리(Enterprise Resource Planning:ERP) 시스템을 CRM 후방 시스템에 포함시켜 생각할 수도 있다. ERP는 고객 정보는 물론 회사의 모든 자원에 대한 정보를 유지, 관리하기 때문에, 고객 관계 관리에 필요한 데이터를 추출하거나, 필요한 조치를 실현하는데 활용할 수 있다.

3. 고객 라이프 사이클과 분석적 CRM

CRM은 결국 고객과의 모든 상호 작용을 관리하는 것이기 때문에, 보다 효과적인 상호작용을 위해서는 고객 라이프 사이클의 모든 단계에서 현 고객 및 잠재 고객에 대한 각종 정보를 적절히 활용해야 한다. 고객 라이프 사이클은 크게 삼단계로 볼 수 있다. 첫 번째 단계는 신규 고객 유치이다. 이 단계를 통해 회사 입장에서 개별 고객의 라이프가 시작된다. 둘째, 고객 가치 증대이다. 고객 중 일부는 회사에 많은 이익을 주기도 하지만, 일부는 오히려 회사에 손해를 끼치기도 한다. 상위 20%에 해당하는 고객 1인의 매출이 나머지 80%에 해당하는 고객 16인의 매출과

비슷하다는 통계가 있다. 따라서, 회사 입장에서 다양한 홍보 및 서비스 제공을 통해 개별 고객을 우량 고객화 함으로써, 최대한 많은 이익을 얻을 수 있도록 노력하는 것이 중요하다. 셋째, 우량 고객 이탈 방지이다. 신규 고객을 확보하는데 드는 비용은 기존 고객에게 서비스하는데 드는 비용의 5배라고 알려져 있다. 따라서, 상당한 비용이 들더라도 기존의 우량 고객이 이탈하지 않도록 노력하는 것이 새로운 고객을 확보하는 것보다 비용면에서 유리하다. 이제 고객 라이프 사이클(customer life cycle)의 각 단계에서 어떻게 분석적 CRM이 적용되는지 살펴보자[11].

3.1 신규 고객 유치

CRM의 첫 번째 단계는 신규 고객으로 확보될 가능성이 높은 잠재 고객을 식별하는 것이다. 한 가상 신용카드 회사인 알파카드사를 예로 들어 살펴보자. 알파카드사는 100만명의 잠재고객을 대상으로 한 DM(Direct Mailing) 발송 홍보를 해마다 실시한다. 지금까지 통계를 살펴보면 DM에 의해 고무되어 신청서를 제출하고 신용평가를 거쳐 신규 고객으로 등록하는 잠재 고객 전체 DM 수신자의 1% 가량 된다. 다시 말해서, DM을 수신한 100만명의 잠재 고객 중 겨우 1만명이 신규 고객으로 획득된 것이다.

알파카드사가 DM 한 장을 발송하는데 소요되는 비용을 대략 1달러라고 가정하면 한 번의 DM 발송 홍보를 위해 100만달러가 소요된다. 신규 고객이 회사에 가져다주는 이익을 연평균 125달러라고 하면 당해연도에 획득한 신규고객 1만명으로부터 얻은 총 이익은 125만달러이다. 결국 100만달러를 들여서 125만달러의 이익을 얻었으므로 순이익은 25만달러라고 볼 수 있다. 만약 무작위로 100만명에게 DM을 보낼 것이 아니라, 실제로 신규 고객이 될만한 사람만 미리 선별해서 DM을 보낸다면 DM 발송 비용을 대폭 절약할 수 있으므로, 결국 훨씬 높은 순이익을 기대할 수 있을 것이다. 이러한 선별 작업을 위해 데이터 마이닝 기술을 활용할 수 있다.

우선 알파카드사는 100만명의 잠재 고객 중 5만명을 무작위로 추출하여 시험적인 DM을 발송한다. 이들의 DM에 대한 응답 여부와 최종 고객으로 선정된 비율과 같은 정보를 분석하여 어떤

특성을 가진 사람이 응답할 가능성이 높은지 예측하는 예측 모델(prediction model)과 개인 특성과 신용 수준과의 관계를 추정하는 추정 모델(estimation model)을 구축한다. 이 두 가지 모델을 바탕으로 응답할 가능성이 높으면서 신용 수준도 만족할 만한 수준인 대상에게만 DM을 발송하게 된다. 이러한 사전 선별 작업을 거친 결과, 시험 대상이었던 5만명을 제외한 95만명의 잠재 고객 중 70만명이 선정되었고, 이들에 대한 DM 홍보 결과 9천명이 최종적인 신규 고객으로 확보되었다고 가정하자. 결국 100만명을 상대로 한 무작위 DM 홍보 경우에 확보한 1만명에 비해 1000명의 신규 고객을 상실했지만, DM 발송 비용이 100만명 분에서 75만명(시험 고객 5만명 포함) 분으로 줄었으므로 결국 비용면에서 25만달러를 절감한 결과를 얻었다. 신규 고객이 9000명이므로 매출 신장은 112만 5천달러이다. 비록 신규 고객이 1만명인 경우에 비해 매출 신장 자체는 감소했지만 순이익은 오히려 37만 5천달러가 되므로 기존 상황의 25만달러에 비해 12만 5천달러가 증가했다. 물론, 데이터 마이닝 작업을 위해 구입한 하드웨어 및 소프트웨어, 인건비 등의 비용도 함께 고려해야 하므로 대략 4만달러가 투입되었다면, 결국 데이터 마이닝을 적용함으로써 8만 5천달러의 상대적인 이익을 얻을 셈이다. 또한 데이터 마이닝을 적용한 경우, 투자 비용은 4만달러인데, 그로 인해 얻은 이익이 8만 5천달러이므로, 투자 대 회수율(Return On Investment:ROI)은 약 213%가량 된다. 이러한 손익비교를 표로 정리하면 표 1과 같다.

표 1 알파카드사의 데이터 마이닝 적용에 대한 손익 비교표

	적용전	적용후	차이
DM 건수	1,000,000	750,000	(250,000)
DM 비용	\$1,000,000	\$750,000	(\$250,000)
응답수	10,000	9,000	(1000)
응답당 매출	\$125	\$125	\$0
총매출	\$1,250,000	\$1,125,000	(\$125,000)
순이익	\$250,000	\$375,000	\$125,000
데이터 마이닝 비용	\$0	\$40,000	\$40,000
최종 순이익	\$250,000	\$335,000	\$85,000

이와 같이 고객 라이프 사이클의 첫 단계인 고객 확보 단계에 데이터 마이닝 기법을 적용함으로써 보다 수익이 높은 고객 확보 작업을 수행할 수 있다.

3.2 고객 가치 증대

일단 고객이 확보되면, 고객이 보다 많은 상품이나 서비스를 구입하도록 유도함으로써, 결국 기업 입장에서 개별 고객의 가치를 증대하는 것이 고객 라이프 사이클의 두 번째 단계에 해당한다. 고객 가치 증대를 위해서 다양한 기법이 있을 수 있지만, 대표적인 기법에 속하는 교차 판매(cross selling) 및 개인화된(personalized) 웹 페이지 기법을 살펴보자.

교차 판매란 상품 A를 구입하고자 하는 고객에게 관련된 상품 B도 함께 추천하여 결국 여러 개의 상품을 사도록 유도함으로써 매출을 증대하는 기법을 의미한다. 전화 주문 판매를 예로 들어 살펴보자. 먼저 고객이 상품 주문을 하기 위해 회사에 전화를 하면 송신자 번호 확인(Caller ID)기능을 이용하거나 직접 고객에게 전화 번호 혹은 회원 번호를 문의하여 고객의 신원을 인지한다. 일단, 신원이 인지되면 고객 데이터베이스에 접속하여 고객에 대한 상세한 정보를 확인한 후 응대를 시작한다. 이 때 주의해야 할 사항은 고객의 성향에 따라 교차판매 시도 자체를 불쾌하게 생각하고 즉시 전화를 끊는 경우가 의외로 많다는 것이다. 또한 고객의 관심과 의도를 정확히 알아내어 적절한 상품을 추천하지 못하고, 전혀 엉뚱한 상품을 추천할 경우, 매출로 연결되지 못한다. 특히 전자의 경우는 원래 구입하려고 했던 상품에 대한 판매까지 무산되기 때문에 오히려 교차 판매 전략이 역효과를 가져왔다고 볼 수 있고, 후자의 경우에는 교차 판매를 위한 홍보에 소요되는 인건비의 손실과 수신자 부담 전화 서비스를 채택한 경우 통신료의 손실까지 야기한다. 따라서, 과연 상담 중인 고객이 교차 판매 시도에 호의적으로 반응할 것인지, 또한 어떤 상품을 추천해야 고객이 관심을 가질 것인지를 알아내기 위해 데이터 마이닝 기법을 활용할 수 있다.

효과적인 교차 판매를 위해서는 먼저 일부 고객을 상대로 전화 혹은 우편 설문을 실시하여 교

차 판매에 대한 호의도를 조사한다. 이러한 조사를 바탕으로 호의적인 고객과 그렇지 않은 고객을 분류하는 분류 모델을 결정 트리(decision tree)의 형태로 구축할 수 있다. 또한 효과적인 상품 추천을 위해 추천 상품에 대한 관심도를 예측하는 예측 모델을 인공 신경망(artificial neural network)을 이용하여 구축할 수 있다.

웹을 통해 상품을 판매하는 환경에서도 개인화된 웹 페이지를 통해 상품의 추천이 가능하다 [7~10]. 우선 연령, 성별, 거주지와 같은 고객의 인구 통계적인 정보와 과거 상품 구매 기록등을 바탕으로 유사한 성향을 가진 고객을 군집화(clustering)한다. 특정 고객이 웹 사이트에 로그인하면 그 고객이 어떤 그룹에 속하는지를 판별하고 해당 그룹의 고객들이 주로 관심을 가진 상품을 추천함으로써 보다 적절한 상품 추천이 가능하다. 만약 이 고객이 추천 받은 상품을 구입하게 되면 바로 그 정보가 해당 그룹의 다른 고객을 위한 상품 추천의 근거 자료가 될 수 있다. 이처럼 여러 고객의 구매패턴을 종합적으로 이용하여 관심을 가질 만한 상품을 선별하기 때문에 이러한 기법을 협동 필터링(collaborative filtering)이라고도 부른다. 아울러, 적절한 상품을 추천 받은 고객은 그 웹사이트에 호감을 갖게 되는 것이 보편적이므로 매출 증대와 함께 회사의 이미지 제고에도 효과를 가져오게 된다.

3.3 우량 고객의 이탈 방지

마케팅에 있어서 새로운 고객을 유치하는 비용은 우량 고객의 이탈을 방지하는 비용에 비해 훨씬 크다는 것이 통계적으로 나타나고 있다. 100만명의 고객을 확보하고 있는 대형 인터넷 서비스 공급업체(Internet Service Provider:ISP)의 경우 매달 8% 정도의 고객이 이탈한다고 가정하자. 한 명의 신규 고객을 유치하는데 약 200달러가 든다면 매달 1600만달러의 비용이 고객 유지를 위해 소요되는 셈이다. 따라서, 데이터 마이닝 기법을 이용하여, 이탈 가능성이 있는 고객을 사전에 판별하여 적절한 보너스 혹은 서비스를 제공함으로써 이탈율을 낮추는 것이 필요하다.

첫 번째 단계는 이탈 가능성이 있는 고객을 판별하기 위해 필요한 데이터를 수집하는 것이다. 인터넷 서비스 공급업체의 경우에는 특정 고객의

매 인터넷 접속당 평균적인 접속 유지 시간, 평균 접속 횟수, 평균 파일 전송 횟수 및 크기, 전자 우편의 송수신 빈도 및 크기, 고객 서비스 센터로 보낸 전자 우편의 횟수 및 성격, 보유하고 있는 전자 우편 주소의 개수, 이용료 납부 기록과 같은 데이터들이 필요하다. 두 번째 단계는 이탈을 방지하기 위해 노력해야 할만큼 회사에 이익을 주는 우량 고객인가를 판별하는 작업이 필요하다. 간단한 손익 계산을 통해 판별할 수도 있고, 비록 현재는 회사에 별 이익을 주지 못하지만 가까운 미래에 이익을 줄 가능성이 높은 고객을 선별하기 위해서는 좀더 정교한 결정 트리 혹은 인공 신경망을 구축할 필요가 있다.

세 번째 단계는 이탈 가능성을 예측하는 작업이다. 주로 사용 빈도의 증감, 고객 서비스 센터로의 문의 빈도 증감과 같은 정보가 활용되며 역시 결정 트리 혹은 인공 신경망과 같은 예측 모델을 채택할 수 있다. 마지막 단계는 이탈 가능성이 보이는 우량 고객에게 어떠한 보너스 혹은 서비스를 제공해야만 이탈을 방지할 수 있을지 판단하는 것이다. 예를 들어 어떤 고객의 경우는 이용료의 할인이 효과적인지 있지만, 어떤 고객의 경우는 사용할 수 있는 디스크 공간의 추가 배정이 효과적일 수 있다.

4. CRM을 위한 데이터 마이닝의 적용 절차

이 절에서는 CRM을 위해 데이터 마이닝을 적용하는 전체적인 절차에 대해서 기술한다. 편의상, 일곱 가지 단계를 순차적으로 기술하지만, 실제 환경에서는 각 단계가 순차적으로만 이루어지는 것이 아니고 필요에 따라 부분적으로 선행 단계를 반복하는 상황이 발생할 수 있다.

4.1 비즈니스 문제의 정의

가장 우선적으로 해야 할 일은 데이터 마이닝을 통해 해결하고자 하는 비즈니스 문제를 정확하고 구체적으로 정의하는 것이다. 예를 들어 “DM 응답율을 높인다”라든지 “각 고객의 가치를 높인다”와 같이 문제를 정의할 수 있는데, 문제가 정의되어야만 필요한 데이터를 수집하고 데이터 마이닝을 적용하는 후속 단계들을 수행할

수 있다. 아울러, 정확하게 기술된 비즈니스 문제 정의는 데이터 마이닝 적용의 결과를 평가할 때 직접적인 기준이 된다.

4.2 관련 데이터 수집 및 정화

일단 비즈니스 문제가 정의되면 필요한 데이터를 수집하여 하나의 데이터베이스로 통합해야 한다. 필요한 데이터의 일부는 관계형 데이터베이스에 저장되어 있고, 일부는 일반 파일로 저장되어 있을 수 있다. 경우에 따라서는 문서상으로만 존재할 수도 있다. 이처럼 데이터들이 여러 가지 형태로 산재해있기 때문에 이들을 식별하고 통합하는 일은 매우 많은 시간과 노력을 요하는 작업이다. 특히 데이터를 외부로부터 입수했거나 아웃소싱한 소프트웨어에 의해 생성된 데이터인 경우, 상당한 시간이 지나고 나면 비록 소프트웨어의 운영은 매뉴얼에 따라 하고 있지만 실제 데이터의 구성과 의미를 파악하기 매우 어려운 경우가 많다.

아울러, 잘못된 데이터, 일관성이 없는 데이터, 혹은 누락된 데이터를 보완하는 데이터 정화(data cleansing) 작업이 필요하다. 경우에 따라서는 데이터 정화가 간단하기도 하지만 매우 복잡하고 사실상 불가능한 경우도 있다. 예를 들어 일관성이 없는 데이터의 자료형이 나타나는 경우에는 대표적인 한가지 자료형으로 통일해야 하는데, 이것은 간단한 자료변환 프로그램으로 가능하다. 하지만, 고객 주소와 같은 자료처럼 동일한 주소지를 여러 가지로 표현한 경우에는 주소 데이터베이스와 같은 보조 자료가 필요하게 된다. 더욱 어려운 경우는 아예 실제와는 다른 데이터 값이 잘못 저장된 경우이다. 이런 경우는 실질적으로 검출자체가 어렵기 때문에 정화 작업이 불가능하다고 봐야 한다. 이처럼 이질적인 대량의 데이터를 취급해야 하고, 사용자가 직접 개입해서 판단해야 하는 일이 많기 때문에 전체 시간 비용 중 이 단계에서 소비되는 비용이 적게는 50%에서 많게는 90%까지 차차하기도 한다.

4.3 사전 데이터 분석

데이터 마이닝을 적용하기 위해서는 사전에 데이터의 분포 및 특성을 분석해야 한다. 예를 들어 데이터의 분류를 위해 데이터 마이닝을 적용

할 계획이라고 하더라도, 데이터가 주로 수치형으로 구성되어 있는 경우에는 인공 신경망을 사용하는 것이 보편적이고 범주형 데이터가 많은 경우에는 결정트리를 사용하는 것이 보편적이다. 또한 데이터 분포를 미리 분석함으로써 데이터 마이닝 작업의 인수값을 결정할 수 있다. 이러한 사전 데이터 분석을 위해 데이터 가시화 도구를 사용할 수 있다. 모형, 색상, 크기, 위치 등을 최대한 활용하여 데이터의 분포를 직관적으로 이해하기 쉬운 형태로 표현해 주는 데이터 가시화 도구를 활용함으로써 데이터 자체를 직접 관찰해서는 얻어내기 힘든 직관적인 통찰력을 얻을 수 있다.

4.4 데이터 마이닝을 위한 데이터 준비

특정 데이터 마이닝 알고리즘을 적용하기 위해서는 필요한 인수값을 결정하고, 독립변수와 종속 변수를 정하는 등 준비 작업이 필요하다. 경우에 따라서는 전체 데이터를 대상으로 데이터 마이닝 알고리즘을 적용하는 것보다 표본 데이터를 추출하여 적용하는 편이 유리할 때도 있다. 표본 데이터를 이용하는 경우 표본의 추출 방식에 따라 데이터 마이닝 결과의 정확도가 크게 영향을 받기 때문에 적절한 표본 방식의 선택이 중요한 관건이 된다. 또한 적용하고자 하는 데이터 마이닝 알고리즘의 특성에 따라 데이터를 변환(transformation)해야 하는 경우도 있다. 예를 들어 결정 트리를 이용하여 데이터 분류(classification)를 수행하고자 하는데 데이터의 자료형은 연속값인 경우 적절한 구간 분할을 통해 연속값을 범주값으로 변환하는 작업이 선행되어야 한다.

4.5 데이터 마이닝 알고리즘 수행을 통한 모델 도출

데이터 마이닝을 위한 알고리즘은 매우 다양하다[6]. 분류(classification)를 위해서는 주로 결정트리와 인공신경망이 사용된다. 추정(estimation)을 위해서는 주로 인공 신경망이 사용되고, 예측(prediction)을 위해서는 결정 트리, 인공 신경망, 연관 규칙 탐사와 같은 방법이 사용될 수 있다. 또한 군집화(clustering)을 위해서는 KNN과 같

은 알고리즘을 사용할 수 있다. 물론, 결정트리 중에도 CART와 같은 이항 결정 트리가 있는가 하면, C4.5와 같은 다항 결정 트리가 있고, 인공 신경망 모델 역시 다중 계층 퍼셉트론(Multi-Layer Perceptron:MLP)와 같은 대표적인 모형 이외에도 다양한 변형들이 제안되어 있다.

적합한 알고리즘의 선택과 함께 중요한 점은 데이터 마이닝 작업이 반복적(iterative)인 작업이라는 것이다. 즉, 한 번의 알고리즘 수행으로 원하는 결과를 얻어내는 것이 아니라, 수행 결과에 근거하여 인수값을 조절하거나 적용 데이터를 보완하여 수차례 반복적인 작업을 거쳐야만 만족할 만한 결과를 얻을 수 있다는 것이다. 또한 데이터 마이닝의 결과로 얻어진 패턴 혹은 지식 중에는 이미 알려진 상식적인 사항이나 비즈니스와 관련성이 없는 사항도 있으므로 결과 중 쓸모있는 부분만 선별해내는 것이 필요하다.

이와 같이 데이터 마이닝 알고리즘의 적용을 통해 얻은 결과를 모델이라고 부른다. 예를 들어 분류를 위한 결정 트리를 얻었다면 그 결정 트리는 현장에 내재한 분류기준을 모델링한 것이므로 분류 모델이라고 할 수 있다.

4.6 모델 적용의 효과 평가

데이터 마이닝을 통해 얻은 모델을 현장에 적용한 후 효과를 평가하기 위해서는 다양한 척도가 사용될 수 있다. 먼저 모델의 정확도(accuracy)를 평가할 수 있다. 예를 들어 결정 트리를 사용한 경우, 시험 데이터를 얼마나 정확히 분류해내는지 정량화할 수 있다. 하지만 정확도만으로는 실제 비즈니스 이득을 평가하기 어렵기 때문에 리프트(lift)를 평가하기도 한다. 리프트란 모델을 적용하기 전과 적용한 후의 매출액 차이라고 볼 수 있다. 최근에는 투자 회수율(Return On Investment:ROI)을 평가하기도 하는데, 리프트를 데이터 마이닝 적용에 소요된 비용으로 나눈 값이라고 볼 수 있다.

4.7 모델의 채택 및 CRM 시스템과의 통합

다양한 효과 평가 결과 적절하다고 판단된 모델을 채택하여 전체 CRM 시스템과 통합하는 작업이 필요하다. 앞서 논의한 바와 같이 CRM은

다양한 전방 시스템 모듈과 다양한 후방 시스템 모듈로 구성되는 복잡한 시스템이다. 따라서 데이터 마이닝 적용을 통해 얻은 모델을 CRM 시스템 요소로 통합시켜서 보다 효과적인 고객 관계 관리가 이루어질 수 있도록 하는 것이 필요하게 된다.

5. 결론

본 논문에서는 CRM의 정의와 요소 기술을 살펴보고, 전방 시스템과 후방 시스템으로 구분하여 필요한 성분을 설명하였다. 특히, 후방 시스템의 핵심 요소인 데이터 마이닝을 중심으로 분석적 CRM이 고객 라이프 사이클의 각 단계에서 어떻게 활용되는지, 데이터 마이닝을 적용하기 위해서는 어떠한 절차가 필요한지 논의하였다.

일반적인 마케팅 전략을 논할 때 시장 점유율(market share)이 중요한 척도로 사용되고 있다. 그와 대응되는 개념으로 CRM을 논할 때, 라이프 점유율(life share)이라는 척도를 사용할 수 있다. 즉, 개별 고객의 인생 전체 기간을 1로 보았을 때, 특정 회사 제품 및 서비스를 선호하는 기간의 비율이 얼마나 되는가를 고려하는 것이다. 결국, CRM 시스템은 시장 점유율은 물론, 라이프 점유율까지 극대화시킴으로써, 회사의 이윤을 최대화하기 위한 마케팅 인프라라고 할 수 있다.

참고문헌

[1] M. J. A. Berry and G. Linoff, "Data Mining Techniques: For Marketing, Sales, and Customer Support", John Wiley and Sons, Inc., 1997.

[2] CRM(Customer Relationship Management, 한국선소식, pp. 12-21, August 2000.

[3] 인포믹스의 CRM 전략과 솔루션, Informix Magazine, pp. 16-29, Summer 2000.

[4] 한국오라클, "Oracle CRM : Release 11i Statement of Direction", 오라클 백서, 2000

[5] R. Forsyth, "Components of a CRM System and a Value-Driven Approach to

Implementation", CRM Forum Report, http://www.crm-forum.com/crm_forum_presentations/cva/ppr.htm, 1999.

[6] 이도현, "데이터마이닝 : 개념 및 연구동향", 한국정보과학회 데이터베이스 연구회지, 13(4), pp. 122-137, 1998.

[7] Doug Riecken, "Introduction: Personalized Views of Personalization", Communications of the ACM, 43(8), pp. 26-28, 2000.

[8] Udi Manber, Ash Patel and John Robison, "The Business of Personalization: Experience with Personalization of Yahoo!", Communications of the ACM, 43(8), pp. 35-39, 2000.

[9] Maurice D. Mulvenna, Sarabjot S. Anand and Alex G. Bhner, "Personalization on the Net using Web mining: Introduction", Communications of the ACM, 43(8), pp. 122-125, 2000.

[10] Bamshad Mobasher, Robert Cooley and Jaideep Srivastava, "Automatic Personalization based on Web Usage Mining", Communications of the ACM, 43(8), pp. 142-151, 2000.

[11] H. Edelstein, "Building Profitable Customer Relationships with Data Mining", SPSS White Paper, 2000.

이도현



1990 한국과학기술원 전산학과(공학사)
 1992 한국과학기술원 전산학과(공학석사)
 1995 한국과학기술원 전산학과(공학박사)
 1995.9~1995.12 인공지능연구센터 연구원
 1999.3~2000.2 미국 텍사스 주립대 방문교수

1996~현재 전남대학교 전산학과 조교수
 관심분야: 데이터 마이닝, 워크플로우, 데이터베이스
 E-mail: dhlee@dbcore.chonnam.ac.kr