

전략중심의 CRM구조의 데이터마이닝

Data Mining for Strategy focused CRM Structure

윤 용 운¹⁾
Yong W. Yoon

(Abstract)

With the explosive growth of information sources available under various information technology and business environment, it has become increasingly necessary for determining effective marketing strategies and optimizing the logical structure of the CRM data mining system.

In this paper, we present an overview of the data mining for strategy focused CRM structure. This includes preprocessing, transaction identification and data integration components. We describe the main part of this paper to the discussion of processes and problems that characterize the mining tools and techniques, identify the CRM data mining, and provide a general architecture of a system to do focused CRM data mining that require further research and development.

[Keyword]

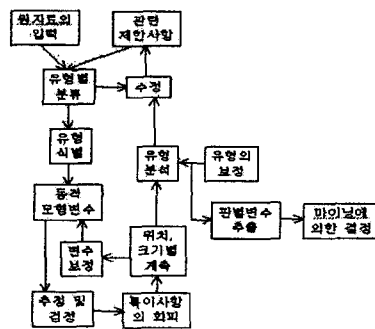
Data mining, CRM, Data mining architecture, e-Business, Knowledge discovery

1. 머리말

데이터베이스 중심의 CRM관리는 고객분석중심이나 고객접점 관리중심이던 간에 고객을 이해하려는 기업의 적극적인 CRM전략이다. CRM 자체가 고객에 대한 지식발견을 전제로 하기 때문에 고객데이터의 통합에 의한 데이터마이닝이 필수적이라 할 수 있다. 물론 CRM의 일반적인 정의에서 제시하듯[2,3] 고객과의 관계를 지속적으로 강화해 나가는 마케팅과 경영혁신과제에서 수반되는 고객정보관리 시스템은 객체지향의 환경에 목표를 두게 된다. CRM 관심영역의 확장에 따른 고객확보와 유지 및 개발의 목표에 대한 기대효과를 최대화하기 위해 기존사업의 성과와 신규사업진출의 플랫폼의 적극적인 활용을 전개하게 되며, 제품지식이 아닌 고객의 행동양식이나 고객 라이프사이클 등에 관한 지식과 관계를 통한 경영의 다각화를 모색한다. 또한 e-비즈니스 모형에 대한 유형화 작업과 특정 기업수준의 개별적인 e-비즈니스 모형에 적합한 CRM시스템은 다양해지거나 계속 변화과정을 거치게 되지만 대부분 기존의 사업방법에 부가적인 요소를 추가하거나 사업내용은 동일하지만 구현방법

1) 강원대학교 수리정보학부 정보통계학과 교수, (우) 200-701, 춘천
E-mail: yoonwoon@kangwon.ac.kr

을 차별화한다. 이러한 정보기술 요건이나 데이터처리기술의 변화는 고객정보의 배타성과 공유성의 전략적 활용체계에 따라 데이터탐색을 포함한 데이터마이닝 시스템을 전략적 CRM의 구축단계와 통합해야 하는 필요성이 대두된다. 그러므로 경영환경분석 및 고객분석을 바탕으로 CRM에 관한 전략 및 실행계획을 수립한 다음 이에 적합한 시스템을 구축하는 것이 바람직할 것이다. 따라서 다양한 경로로 수집된 고객정보의 데이터를 통합하여 (그림 1)과 같은 마이닝과정을 거쳐 제품시장 환경변화에 따른 데이터웨어하우스의 전략을 효율적으로 수행하게 되며, 특히 경영정보 환경에서 필연적으로 야기되는 CRM의 전략이 실제 다차원정보의 효율성을 제고시켜준다는 점에서 경영기술 환경의 중요한 변수가 된다.

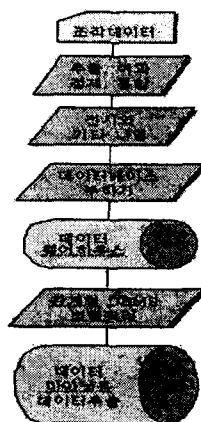


(그림 1) 데이터마이닝과정

이 논문은 다양한 정보기술과 경영환경의 상황에서 보다 전략적인 CRM구조의 데이터마이닝을 수행할 수 있는 시스템의 구축 및 운영에 관한 일반적인 절차를 언급하려는 것이다.

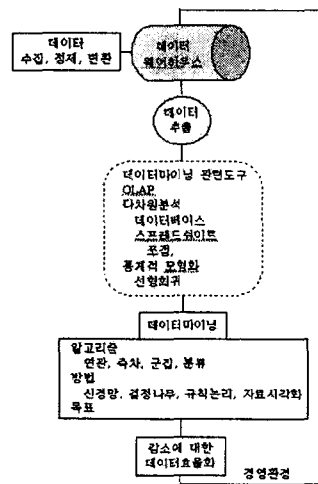
2. CRM 데이터마이닝시스템의 구조

초기 SMP, MPP, 클러스터링 등 상업용 병렬컴퓨터와 대용량 병렬컴퓨팅을 지원하는 RDBMS의 출현이 실제 데이터마이닝을 활용하는 배경이 되어 (그림 2)와 같은 절차를 응용한 관계형 데이터베이스 혹은 다차원 데이터베이스 모형이 적절하였다.



(그림 2) 관계형 데이터베이스의 절차

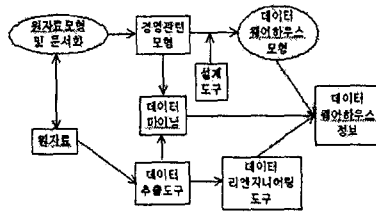
또한 고객데이터에 대한 DBMS는 잠재고객의 데이터부재와 부정확성의 문제가 발생되고, CRM 마이닝 시스템의 전략에 대한 이해부족과 기업 상황에 적합한 CRM의 명확한 인식의 부재는 CRM 분석 및 실행시스템이나 기술상의 문제로 나타나게 된다. 따라서 CRM 데이터 마이닝 시스템의 구조는 고객중심의 관리 형태를 전제로 하여 핵심적인 요소는 DBMS가 되며, Back-end에서 Front-end까지 통합적으로 접근할 수 있고, 가능한 한 모든 솔루션을 갖는 일관성의 형태를 취해야 할 것이다. 특히 ERP, PPM등 같은 기간 업무별 응용패키지와 결합하여 기능 중심적이고 변화 가능한 형태를 띄우는 유연성과 적응성이 있어야 한다. 대개 CRM은 운영시스템 형태와 정보시스템 형태의 성격 모두를 내포하지만 기업정보기술 환경의 변화에 따라 데이터처리기술은 탐색과 마이닝의 단계로 전환된다. 이런 데이터마이닝은 (그림 3)과 같이 데이터추출, 마이닝의 방법론과 분석도구 및 알고리즘을 적용하므로써 마이닝의 효율성을 제고하는 일반적인 절차과정을 거쳐 경영환경에 적합한 구조로 설정된다.



(그림 3) 데이터마이닝 절차

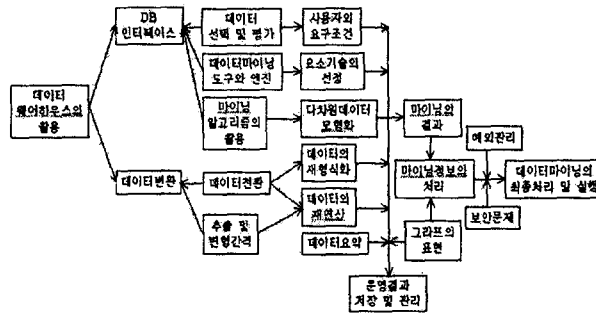
마이닝에 관련된 도구로써 OLAP이나 다차원 분석을 통해 통계적인 모형화를 설정하게 되고[4,5,10], 실제 마이닝 방법이나 알고리즘을 이용하여 전략중심의 CRM 데이터내의 지식 발견의 유형을 식별하므로써 고객데이터의 효율성을 높이는 것이다.

한편 장기적인 관점에서 CRM 데이터마이닝을 수행할 수 있도록 고객유지율, 고객성장률, 브랜드 선호도와 같은 지표를 파악하여 우량고객에 대해 정확하게 이해하고, 고객의 특성에 적합한 정보기술의 리엔지니어링을 수행한다. (그림 4)의 데이터마이닝 절차를 통해서 CRM 구조의 데이터마이닝 운영에 필요한 모든 소스데이터를 찾아내어 옮겨오고 전환시키는 과정을 거치게 한다. 반면에 다차원모형인 데이터웨어하우스의 모형으로부터 생성되는 데이터베이스 객체들을 위한 요구사항과 목표에 대한 적절한 절차를 수립하여야 할 것이다[1].



(그림 4) 데이터마이닝 절차

특히 비구조적이고 비규격화된 데이터의 중요성이 증가하고 있는 현재의 기업정보 기술 환경에서 고도의 정보서비스라는 CRM 데이터마이닝의 본래의 목적을 충족시키기 위해서는 데이터마이닝 시스템에 직접 접속하여 시각적인 분석과 통합된 기술 및 다양한 데이터마이닝 툴과 엔진을 처리한다[8]. 또한 데이터베이스에 수록된 개별 고객차원의 자료에 통계적 분석과 모델링기법등을 적용하여 잠재고객 및 현재고객이 장기간에 걸쳐 제품이나 서비스를 반복적으로 구매하기 위해서 고객과 기업 간의 쌍방연계 형태를 취한다. 따라서 지식수준의 CRM 데이터마이닝 구조를 구축하기 위해서는 고객의 평가에 근거한 경쟁업자와 경쟁전략의 유사성에 근거한 경쟁업자를 파악하여 비교우위전략을 선택하고 이 경쟁전략과 수익성제고를 동시에 고려하여 고객세분화문제를 해결한다. 이런 과정을 요약하면 (그림 5)와 같다.



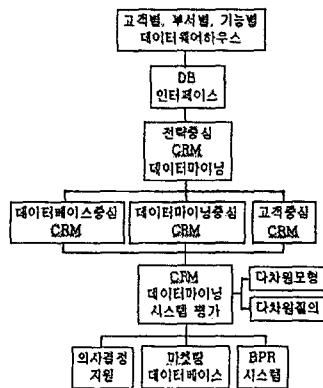
(그림 5) 데이터마이닝시스템의 과정

3. CRM 데이터마이닝 시스템의 운영

데이터마이닝 중심의 CRM 시스템은 마이닝된 데이터소스의 형태, 유도된 데이터모형, 발견된 지식의 유형이나 사용된 마이닝기법등에 따라 분류한다. 데이터소스의 형태에 따른 분류는 공간데이터, 멀티미디어 데이터, 시계열데이터, 텍스트데이터나 WWW등과 같은 조정된 데이터의 형태에 따라 데이터마이닝 시스템을 운영한다[6,7]. 또한 유도된 데이터모형은 관계형 데이터베이스, 객체지향형 데이터베이스, 데이터웨어하우스 등을 포함한 모형에 의한 것이다. 발견된 지식의 유형은 고객별의 거래특징, 고객데이터의 유형의 판별, 거래제품의 연관 및 분류, 고객별 구매유형의 그룹화등과 같은 데이터마이닝의 기능성이나 발견된 지식을 파악하게 된다. 반면에 CRM 데이터마이닝에 사용된 기법은 기계학습, 신경망, 유전자 알고리즘, 통계분석, 시각화, 객체지향 데이터베이스나 데이터웨어하우스 등 사용된 데이터분석기법에 따라 데이터마이닝 시스템을 운영하게 된다[9].

CRM 데이터마이닝의 데이터구조에서 고려하게 되는 특징은 구조적 데이터의 중요도와 관계들인데 중요도는 대개 수치화되거나 혹은 각 레코드마다의 필드와 테이블마다의 칼럼들이 표현되는 각 객체들과 여러 관계가 존재하므로 텍스트데이터와 비구조적 데이터를 동질성 아키텍처로 접근하는 것이 효율적이다. 또한 실시간 비구조적 데이터의 범위는 상세한 레벨과 요약에 의해 텍스트추출 데이터와 오디오 데이터로 구분되지만 두 레벨간의 관계는 일반적으로 알고리즘에 의해 파생되지 않기 때문에 비구조적 데이터는 시간에 대한 민감성이 작다. 따라서 데이터웨어하우스 규모에서의 비구조적 데이터의 규모를 신중하게 고려해야 한다.

한편 CRM 데이터마이닝 시스템의 평가는 의사결정지원 컴포넌트에 의해서 고객의 평가 기준에 근거하여 합리적인 의사결정을 할 수 있도록 한다. 적절한 인터페이스를 통해 평가 기준을 정의하고 평가기준 각각의 중요도를 지정한 다음 평가기준에 따라 대안들을 평가하고 대안순서를 결정한다. 그런데 실제 발생하게 되는 비즈니스 데이터는 사용자 애플리케이션이나 계층적인 데이터베이스와 관계형 데이터베이스를 이용하는 내부적으로 일관성을 갖도록 설계되는 과정인 것이다. 또한 클라이언트와 서버 모두에게 데이터의 구조와 내용을 분석하고 규칙들을 강화하기 위한 프로세스의 기능을 포함하게 되는데 각 속성이나 필드의 연결 및 조작, 필드 내용의 일관적인 형태로의 전환이나 극단적인 상황에서의 오류수정의 다양한 유형 등이다. (그림 5)는 CRM 데이터마이닝 운영절차를 나타낸 것이다.



(그림 5) CRM 데이터마이닝 운영절차

CRM 데이터마이닝 시스템의 평가는 CRM 시스템지원연쇄관리, e-비즈니스와 그 외의 전략적인 응용에 대한 효율성의 극대화를 전제로 하게 되며, 제품기술혁신에 따른 기업의 경쟁적인 요소를 그대로 유지하도록 CRM 마이닝 시스템의 핵심적인 차별화에 의해서 지원된 개인별 서비스를 지원하는 능력을 향상시킨다. 또한 새로운 고객을 획득하기 위한 고비용은 CRM 실용성과 시스템을 수행하는 중요성의 요인이 되므로 CRM 프로젝트는 다양한 데이터형식과 정의와 함께 다차원 소스의 데이터를 사용하여 식별되지 않거나 혹은 다시 해결되지 않은 데이터 품질문제를 CRM 시스템의 유효성으로 해결한다. 특히 데이터 품질문제는 데이터변환이나 여러 데이터 소스중 최선의 데이터를 선택하는 문제나 데이터의 요약, 데이터의 재형식화와 재연산, 데이터의 인코딩과 디코딩, 데이터에 대한 내부적 표현이 필요하다[4]. 따라서 구축프로젝트의 기술적 기초를 설정하고, 다차원모형인 데이터웨어하우스의 모형으로부터 생성되는 데이터베이스 객체들을 위한 요구사항과 목표에 대한 관리를 한다.

또한 최종 마이닝의 종료 또는 데이터웨어하우스의 실제 운용까지 필요로 하는 과제를 정의하고, 실행하는 과정과 운용중인 시스템에 대해 평가하고 앞으로의 데이터와 시스템의 향상을 위한 계획을 수립하는 것이다. 따라서 마케팅 과제의 관점에서 환경분석 및 고객분석을 바탕으로 CRM에 관한 전략을 수립하여 고객가치나 반응성, 고객속성기준 등 고객의 총체적인 특성을 파악하게 된다.

4. 결론 및 과제

데이터웨어하우스가 프로세스의 가능요인으로서의 통합된 정보기반이기 때문에 데이터마이닝 프로세스 리엔지니어링 시스템을 근본적으로 구축할 필요가 있다. 전략중심의 CRM 데이터마이닝 시스템의 구축은 정보기술의 의사결정을 최적화하려는 목적이며, 운영 프로세스를 최적화하는 것과 다르다. 다차원 질의와 모델링, 다양한 연산에 의한 데이터마이닝 시스템을 전략중심의 CRM구조에 활용한다는 것은 효과적인 정보처리와 정보환경을 이해하는데 필수불가결하다. 따라서 대용량의 고객별 데이터와 다양한 요소기술이 결부된 통합 환경을 구성하는 전략중심의 CRM 데이터마이닝 시스템을 구축하여 관리해야 할 것이다.

[참고문헌]

- [1] M.S. Chen, J. Han, and P.S. Yu, Data Mining: An Overview from a Database Perspective, *IEEE Trans. Knowledge and Data Engineering*, 8:866-883, 1996
- [2] W.J. Fraley, G. Piatetsky-Shapiro and C.J. Matheus, Knowledge Discovery in Databases: An Overview, In G. Piatetsky-Shapiro et al. (eds.), *Knowledge Discovery in Databases*, AAAI/MIT Press, 1991
- [3] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000
- [4] W.H. Inmon, *Building the Data Warehouse*, 2nd. ed. John Wiley & Sons, Wiley & 1996
- [5] I. Khosla, B. Kuhn, and N. Soparkar, Database search using information mining, In *Proc. of 1996 ACM-SIGMOD Int. Conf. on Management of Data*, 1996
- [6] R. King and M. Novak, Supporting information infrastructure for distributed, heterogeneous knowledge discovery, In *Proc. SIGMOD 96 Workshop on Research Issues on Data Mining and Knowledge Discovery*, Montreal, Canada, 1996
- [7] B. Mobasher, N. Jain, E. Han, and J. Srivastava, *Web Mining: Pattern discovery from world wide web transactions*, Technical Report TR96-050, University of Minnesota, Dept. of Computer Science, Minneapolis, 1996
- [8] R. Ng. and J. Han, Efficient and Effective Clustering Method for Spatial Data Mining, In *Proc. of the 20th VLDB Conference*, pp.144-155, Santiago, Chile, 1994
- [9] G. Piatetsky-Shapiro and W.J. Frawley, *Knowledge Discovery in Databases*,

AAAI/MIT Press, 1991

- [10] S.M. Weiss and C.A. Kulikowski, *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert Systems*, Morgan Kaufmann, San Mateo, CA, 1991
- [11] O.R. Zaiane and J. Han, Resource and knowledge discovery in global information systems: A preliminary design and experiment, In *Proc. of the First Int'l Conference on Knowledge Discovery and Data Mining*, pp.331-336, Montreal, Quebec, 1995