

Travel Agency에서 CRM을 위한 DataMining, OLAP 적용

김민정^o 박승수
이화여자대학교 컴퓨터학과
(mjkim,sspark@ewha.ac.kr)

Applying Datamining and OLAP for CRM to Travel Agency

Min-Jung Kim^o Seoung-Soo Park
Dept. of Computer Science & Engineering, Ewha Womans University

요 약

World Wide Web(WWW) 데이터가 폭발적으로 증가하고 있는 시점에서, WWW의 데이터로부터 유용한 정보를 찾아내고 분석하는 일이 필요해졌다. 또한 WWW의 데이터 만으로는 얻을 수 없는 기업의 의사결정을 위한 정보를 얻기 위해, 웹 페이지 접근 기록에서 얻어진 웹 로그 기록들과 기업의 판매 트랜잭션 데이터베이스, 광고 데이터베이스 그리고 고객 정보를 통합하여 데이터 웨어하우스를 구축한다. 이러한 과정은 기업활동의 결과로 축적된 데이터 자원과 WWW의 데이터를 통합하여 체계적인 정보기반을 구축하고, 이러한 자원을 전략적으로 재활용하는 것이 목적이다. 본 논문에서는 WWW의 데이터와 기업의 데이터베이스를 통합하여 웨어하우스를 설계하고 여기에 데이터마이닝, OLAP을 적용하여 CRM에 활용하는 방안을 제안하고자 한다.

1. 서론

WWW의 유용한 데이터가 폭발적으로 증가하고 있으며, WWW의 데이터로부터 유용한 정보를 찾아내고 분석하는 일이 실질적으로 필요해졌다. 또한 많은 기관들이 그들의 고객이나 비즈니스 파트너에 대한 풍부한 데이터를 가지고 있음에도 불구하고 이 데이터에 숨겨진, 지식이 될 수 있는 가치 있는 정보를 찾지 못하고 있다. 이러한 정보를 기업의 의사결정 활동에 활용하는 기술과 데이터마이닝의 가장 빠르게 성장하고 있는 응용분야가 CRM(Customer Relationship Management)이다.[4] CRM은 고객에 초점을 둔 접근으로서, 사업자들은 다양한 채널을 통해 고객의 데이터를 모으고, 이것을 데이터웨어하우스에 통합하고, 이렇게 축적된 데이터를 분석하고, 그 분석에 기반하여 활동하는 것이다.[3] 본 논문에서는 CRM을 위한 준비단계로서 기존 고객들의 패턴을 분석하기 위하여 웹 사이트를 방문하는 고객들의 행동패턴과 기업의 판매 데이터베이스 그리고 광고 데이터베이스를 결합하여 스타스카마를 가지는 데이터 웨어하우스를 구축한 후 다양한 데이터마이닝 기술과 OLAP 분석을 적용해 보고자 한다. 웹 페이지 접근 기록들을 분석하는 일은 데이터마이닝과 WWW 기술의 발달로 가능하다. WWW 사용자들의 행동은

웹 서버 로그 파일들에 저장되며, 이렇게 모아진 사용자들의 행동을 분석하고 규칙성을 찾는 일은 시스템의 성능을 높이고, 최종 사용자에게 전해지는 인터넷 정보 서비스의 질을 높일 수 있으며, 전자 상거래의 잠재적인 고객 집단을 추정할 수 있게 해 준다.[2] 또한 웹 로그 기록들을 분석하는 것 뿐만 아니라 이것을 기업의 광고활동 및 실질적인 판매량과 연관 지어 보는 것이 좀 더 의미 있는 분석이 될 것이다.

본 논문의 구성은 2 장에서는 웹 로그 기록들의 분석에 관련된 연구와 CRM 관련 연구에 대해 알아보고 3 장에서는 본 논문에서 데이터마이닝 기법으로 쓰이게 될 연관규칙과 클러스터링에 대해서 살펴보고 전체적인 시스템 구성을 제시한다. 마지막으로 5 장에서는 결론과 향후 과제를 논의한다.

2. 관련연구

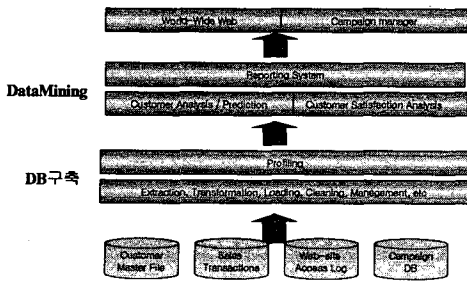
데이터마이닝은 KD(D Knowledge Discovery in Databases)의 한 단계로서, 대량의 실제 데이터로부터, 이전에 잘 알려지지 않은 것들이지만, 목시적이고, 잠재적으로 유용한 정보를 추출하는 작업이라 정의한다.[5] 데이터마이닝은 흥미 있는 패턴을 찾는 과정으로 여러 가지 기술들이 사용된다.[6] 웹 마이닝은 데이터마이닝 이러한 기술들을 거대한 웹 데이터 저장소에 적용한 어플리케이션이다.

마인딩은 Web content mining 과 Web Usage Mining으로 나누어진다. 후자는 웹 로그 마이닝으로도 불리며 웹 로그에서 흥미 있는 패턴을 추출하는 과정을 말한다. 패턴을 찾기 위해 마이닝의 여러 가지 기법들이 적용되는데, WAP[1]은 순차패턴과 연관규칙을 찾기 위해 WAP-tree 구조를 만들고, WAP-mine 알고리즘을 적용하였다.[1] 또한 다양한 마이닝 기법들을 포함하고 있는 웹 로그 마이닝 툴-Getstats, Analog, Microsoft Interse Market Focus, WebTrends, Weblogminer 등이 있다. 이 중 Weblogminer[2]는 마이닝의 여러 가지 기법들 중 time-series 분석에 초점을 두고 있다. 이를 위해 사용자와 데이터를 선택하고, 사이클을 찾고, 경향을 분석하는 과정을 두고, 이 과정에 OLAP 과 마이닝 기법들을 적절히 적용하고 있다.[2]

이외에도 CRM 연구의 예로 [3]은 telecomm-unications 고객을 귀납법(Rule induction)을 써서 분류하고, MBR(Memory-based reasoning)을 사용하여 점수를 매기는 방법으로 고객의 행동을 예측하기 위해 데이터마이닝의 기법을 적용한 것이 있다.[3]

형식으로 변환해야 하고, 구조 전처리에서는 페이지 사이의 하이퍼링크 관계 분석을 위해 변환해야 한다. 판매 트랜잭션의 경우에도 상품 내에 존재하는 상하 포함관계를 표현하는 등의 처리가 필요하다. 이러한 변환들의 경우 이와 관련된 지식 베이스를 만들어 이를 기반으로 처리해준 다.

3. 시스템 구성

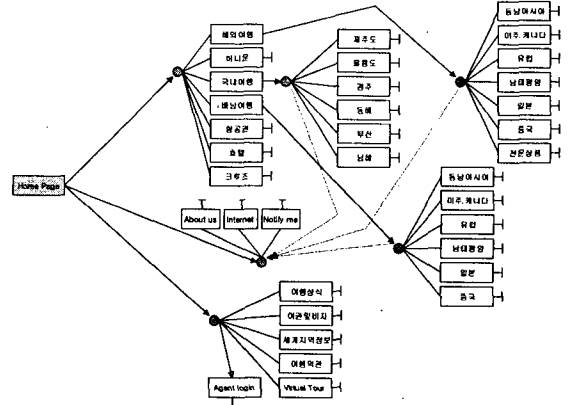


[그림1] 시스템 구성도

본 시스템의 데이터 소스는 웹 로그, 판매 트랜잭션, 고객, 광고 데이터베이스로 구성되어 있다. 전체 과정은 두 단계로 구성되는데, 데이터 소스들을 통합하기 위해 스타스키마를 가지는 데이터 웨어하우스를 구축하는 단계와 구축된 웨어하우스를 OLAP 연산들과 마이닝 기법들을 적용하여 여러 가지 측면에서 분석해 보는 단계로 나누어 진다.

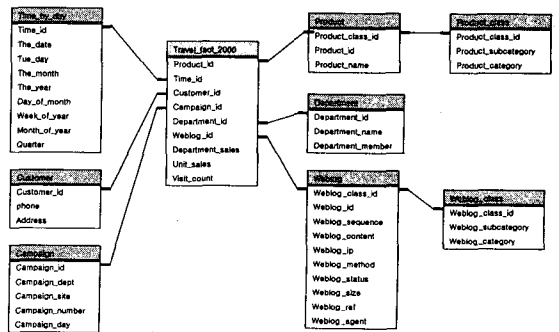
3.1 데이터웨어하우스 구축

데이터웨어하우스를 구축하는 단계에서 이루어지는 전처리(preprocessing)에서 웹 서버 로그의 경우 사용(Usage), 내용(Content), 구조(Structure) 전처리로 구성된다.[7] 사용 전처리에서는 사용자와 서버 세션들을 확인하고, 내용 전처리에서는 웹 문서에 있는 다양한 파일들을 분석하기 용이하도록 일정한



[그림2] 웹 사이트 구성도

웹사이트 지도는 고객들 사이에 링크의 집합체로 표현하는 것으로 전체적인 사이트 구성도이다. 이것은 사용자가 선택하는 임의의 객체를 중심으로 관계를 알아보기 편하다. 구조 전처리의 경우 위와 같은 웹 사이트의 계층 관계를 분석을 위한 표현으로 변환하는 것으로, 부모노드와 자식노드의 관계를 정하고 단말 노드들에 대한 처리를 말한다. 이것은 이후 패턴 분석단계에서 적용하게 될 데이터마이닝 기법 중 연관규칙에서 함께 참조되는 웹 페이지들을 웹 사이트의 계층관계에 기반해서 찾기 위한 것이다.



[그림3] 스타스키마

3.2. 패턴 발견

OLAP 연산들과 마이닝 기법들을 통해서 웹 트래픽 정보를 분석하여 시스템 성능향상이 가능하며, 고객 행태에 관한 분석을 통해서 웹 사이트의 구조 및 디자인에 관한 정보를 얻을 수 있고, 인터넷 비즈니스 기업에게 필수적인

고객의 웹 사이트 이용현황에 관한 정보 제공을 통해서 새로운 제품과 서비스 개발이 가능할 수 있고, 웹 사이트의 사용, 내용, 구조에 대한 특성이 가능하다.[7] 특히 고객관리의 측면에서는 고객을 세분화할 수 있으며, 신규고객과 기존고객의 행동을 예측하여 그들이 누구인지, 왜 그런 구매를 하는지, 어떤 상품이 함께 구매되는지, 어떤 상품이 촉진되기에 적절한지를 예측할 수 있다. 분석을 위해 본 논문에서 적용되는 마이닝 기법으로는 사용자들을 패턴에 따라 분류하기 위해 클러스터링 기법을 사용하고, 웹사이트 사용에서의 유사도를 알아보기 위해 연관규칙을 사용한다. OLAP 연산들로는 상품별 시간에 따른 판매량의 변화와 접근경로별 사이트별 시간별 사용자의 선호도를 알아볼 수 있다.

3.2.1 연관규칙 (Associations Rule)

본 연구에서 연관규칙은 단일 세션의 웹 페이지들을 각각의 아이টে็ม으로 하고, 단일 세션에서의 웹 페이지 탐색 순서를 트랜잭션으로 보았으며, 함께 참조되는 회수가 많은 페이지들을 관련이 있다고 보고 이를 찾아내는 것을 말한다. [그림3]과 같은 스타스키마에서 시간차원의 각 분석요소별로 참조되는 웹 페이지들을 찾을 수 있으며, 웹 페이지의 계층구조에 따라 동일 부모 노드 내에서의 참조 연관뿐만 아니라, 부모 노드가 다를 경우에 참조 연관도 분석이 가능하다. 또한 시간, 광고, 판매 트랜잭션, 웹 로그, 고객 관련등의 몇 개의 차원에 걸친 연관규칙 분석도 가능하다.

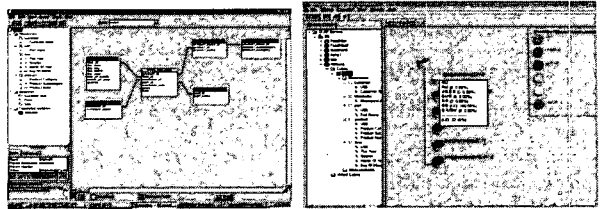
3.2.2 클러스터링

클러스터링이란 주어진 데이터 집합에서 서로 유사성을 가지는 몇 개의 클러스터로 분할해 나가는 과정으로, 하나의 클러스터에 속하는 데이터 점들 간에는 서로 다른 클러스터 내의 점들과는 구분되는 유사성을 갖게 된다.[8] 본 논문에서 클러스터링 기법으로 가능한 분석으로는 유사한 브라우징 패턴을 가지는 사용자들의 클러스터를 찾을 수 있고 연관된 콘텐츠를 가진 웹 페이지들의 클러스터를 찾을 수 있다. 클러스터링 또한 연관규칙과 마찬가지로 몇 개의 차원에 걸친 분석이 가능하다.

4. 실험

본 논문에서 제시한 시스템 구축을 위해서는 여러 가지 소스의 데이터를 모아 통합적인 데이터웨어하우스 구축이 필요하다. 이를 위해 각 데이터 소스별로 스타스키마를 만들어 차원을 공유하는 방식으로 통합할 계획이다. 우선, 판매 트랜잭션 데이터에 대한 스타스키마와 OLAP, 마이닝 분석 결과이다. 시스템 환경은 Windows NT server 상에서 데이터베이스로는 Access2000 을 사용하였고, 큐브 생성은 Microsoft OLAP Service 를 사용하고, 마이닝 툴은 DBMiner2.0을 사용하였다. 사용될 데이터는 실제 Travel agency 의 2000년 1월에서 6월 판매 데이터, 웹 서버 로그파일, 광고 데이터베이스 그리고 고객관련 데이터이다.

[그림 4] 는 판매 데이터베이스만으로 스타스키마를 만들고 DBMiner2.0을 사용한 화면이다



[그림 4] 판매 데이터베이스 분석

5. 결론 및 향후연구

본 논문에서는 판매 데이터, 웹 서버 로그파일, 광고 데이터베이스 그리고 고객관련 데이터들을 통합하고 OLAP, 데이터마이닝 기술을 적용하여 CRM 에 사용할 것을 제안하였다. 테스트를 위하여 Travel Agency 데이터를 사용할 것이며, 현재 데이터들을 통합하기위해 정제 및 변환 과정을 프로그램내에서 처리해주는 작업을 진행중이다.

본 연구에서는 고객데이터를 중심으로 분석하여 고객층을 세분화하고 개인화 된 서비스를 제공하고자 하는 취지보다는, 고객의 활동과 관련되어 고객과의 접점에 존재하는 데이터를 이용하여 고객에게 접근하고자 하였다. 향후 연구에서는 이를 바탕으로 실제 고객의 개인정보가 얻어질 경우 이를 분석하여 기업의 활동에 적용하는 CRM 시스템을 구축할 계획이다.

6. 참고 문헌

- [1] J. Pei, J. Han, B. Mortazavi -Asl, and H. Zhu , " Mining Access Pattern efficiently from Web logs ", Proc. 2000 Pacific -Asia Conf. on Knowledge Discovery and Data Mining (PAKDD'00), Kyoto, Japan, April 2000.
- [2] O. R. Zaiane, M. Xin, J. Han, " Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs ", Proc. Advances in Digital Libraries Conf. (ADL'98), Santa Barbara, CA, April 1998, pp. 19-29.
- [3] Toyohisa Morita, Yoshinori Satou, Erika Ayukawa, and Akira Maeda, " Customer Relationship Management Through Data Mining ", INFORMS & KORMS □ Seoul 2000(Korea)
- [4] Alex Berson, Stephen Smith, Kurt Thearling, " Building Data Mining Applications for CRM ", McGraw- Hill, 2000/8/24
- [5] Usama M. Fayyad, Gregory Piatetsky -Shapiro, et al., eds.(1996), " Advances in Knowledge Discovery and Data Mining ", McGraw-Hill, 1996.
- [6] Michael J. A. Berry and Gordon Linoff(1997), " Data Mining Techniques For Marketing, Sales, and Customer Support ", John Wiley & Sons
- [7] J. Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan, " Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data ", 2000 SIGKDD