

e-Business에서의 BI지원 데이터마이닝 시스템

(A Data Mining System for Supporting of Business Intelligence in e-Business)

이 준 욱[†] 백 옥 현^{**} 류 근 호^{***}

(Jun Wook Lee) (Ok Hyun Baek) (Keun Ho Ryu)

요 약 비즈니스 인텔리전스에 대한 관심이 증대되면서 핵심 기술로써 데이터마이닝의 적용이 증대되고 있다. e-Business에서의 비즈니스 인텔리전스를 지원하기 위해 다양한 마이닝 연산을 통합적으로 제공하는 마이닝 시스템은 데이터베이스 시스템과 유연하게 통합될 수 있어야 하며, 또한 다양한 비즈니스 응용에서의 마케팅 프로세스를 쉽게 구현할 수 있는 인터페이스를 제공하여야 한다.

이 연구에서는 e-Business영역에서의 BI를 지원하기 위해 데이터마이닝 기법을 통합적으로 제공하는 시스템으로써 EC-DaMiner 시스템을 설계, 구현하였다. 데이터마이닝 시스템은 기존의 데이터베이스 시스템과의 표준적인 인터페이스를 통하여 연동될 수 있도록 하였다. 아울러 비즈니스 어플리케이션들은 마이닝 질의어인 MQL을 통하여 규칙을 탐사하고 탐사된 규칙을 기존의 마케팅 데이터베이스에 모델화하여 반영함으로써 마케팅 전략의 구현을 용이하게 하였다.

키워드 : 비즈니스 인텔리전스, 데이터마이닝 시스템

Abstract As the interest in business intelligence is increased, data mining is increasingly used in BI as the core technique. To support Business Intelligence in e-business environment, the integrated data mining system which included in various mining operations should be able to flexibly integrate with database system and also it must provide the easy and efficient interface to implement the marketing process in various business applications.

In this paper, we have implemented the EC-DaMiner system to support business intelligence in e-business area. The implemented system can be integrated with the conventional database system with the standard interface. Business applications can use MQL mining query language to discover the rules and mining result is modeled in marketing database, and the EC-DaMiner system make the implementation of business marketing process more easy.

Key words : e-Business, Business Intelligence, Data Mining System

1. 서 론

비즈니스 인텔리전스(business intelligence)는 다양한 비즈니스 데이터를 이용하여 좀 더 나은 의사결정을 지원할 수 있도록 하는 기법이다[1,2]. e-Business에서는

개인화 서비스(personalization) 및 고객 관계관리 등에 있어서의 비즈니스 인텔리전스(이하 BI)를 제공하기 위하여 데이터베이스나 데이터웨어하우스 등의 데이터 저장소와 OLAP, 통계적 분석 및 데이터마이닝과 같은 분석기술 등을 접목하고 있다[3]. 데이터마이닝은 방대한 양의 누적된 고객 및 트랜잭션 데이터로부터 알려지지 않은 묵시적이고 잠재적으로 유용한 지식을 추출함으로써 성공적인 의사결정 지원을 위한 BI 시스템의 핵심 기술이다[4,5]. 최근 데이터마이닝 기법들은 복잡해지고 다양해지는 기업 활동에 있어서 고객의 프로필, 고객의 구매 내역, 상품 특성 및 웹 접근 로그 분석 등을 위하여 시스템에 직접 내장되거나 밀결합(tightly coupled)되는 독립 시스템을 형성하고 있다[6,7]. 이러

· 이 연구는 한국과학재단 특정기초연구(R01-1999-00243) 및 ETRI 우정연구부의 지원으로 수행되었음.

† 비 회 원 : 충북대학교 컴퓨터학과

junux@dblab.chungbuk.ac.kr

** 비 회 원 : 국방과학연구소 연구원

ohpaek@add.re.kr

*** 종신회원 : 충북대학교 전기전자및컴퓨터공학부 교수

khryu@dblab.chungbuk.ac.kr

논문접수 : 2001년 8월 8일

심사완료 : 2002년 6월 18일

한 마이닝 시스템은 고객 세그멘테이션(segmentation) 및 분류(classification) 그리고 구매 상품간의 연관성 탐사 등의 지식을 제공함으로써 e-Business 시스템의 서비스 및 구매력 향상과 의사결정자의 다양한 구매 유발, 판매 및 캠페인 전략 수립 등을 가능케 한다[8,9].

현재 다양한 비즈니스 데이터의 분석을 위하여 데이터베이스 또는 데이터웨어하우스와 데이터마이닝의 통합적 사용이 증대되었다. e-Business에서의 데이터는 급격하게 증가하고 탐사된 규칙이나 패턴 역시 빠르게 변화되는 특성을 갖기 때문에 e-Business 시스템 내에서의 마이닝 규칙은 동적으로 탐사되어 마케팅 프로세스로 구현될 수 있는 구조를 가져야 한다. 그러나 기존의 마이닝 시스템들은 주로 의사결정자를 위한 레포팅 중심적이고 다차원 분석적인 결과를 제시하고 있으며 특정 시스템에 밀접합되는 구조를 갖고 있어 이를 기존의 e-Business 시스템에 통합하여 적용하기 힘들다[10, 11,12,13]. 따라서 독립형 구조를 갖는 마이닝 시스템은 다양한 데이터베이스와 통합 가능한 구조를 갖는 이식성이 높고 확장 가능한 특성이 요구된다[3].

기존의 연관규칙 탐사, 분류 및 클러스터링 기법 등을 적용한 마이닝 서브시스템이 e-Business를 위한 백엔드(back-end) 시스템으로 적용되고 있다. e-Business 영역에서의 백엔드 시스템으로서 데이터마이닝 서브시스템을 적용하기 위해서는 다음과 같은 사항이 고려되어야 한다. 우선 마이닝 시스템의 데이터마이닝 알고리즘 즉 마이닝 연산의 데이터 접근 구조, 마이닝 시스템의 아키텍처 및 데이터베이스 시스템과 결합구조 등이다. 또한 마이닝 시스템의 성능향상과 효율적인 적용을 위해 기존의 데이터베이스 기술을 고려한 성능향상 방안, 그리고 데이터베이스 또는 데이터웨어하우스 기반의 비즈니스 어플리케이션을 위한 배경지식 및 마이닝 지식 표현 스키마 등이 고려되어야 한다.

이 논문에서는 e-Business 데이터 분석을 위한 이식성과 확장성을 갖는 데이터마이닝 시스템으로서 EC-DaMiner를 설계, 구현한다. 이 시스템의 구현을 위하여 다음과 같은 설계 목표를 설정한다. 첫째, 시스템은 특정 시스템에 종속적이지 않은 데이터베이스 시스템들과의 통합을 지원한다. 둘째, 시스템은 다양한 플랫폼에서 수행 가능한 이식성을 갖는다. 셋째, 시스템은 Java Servlet이나 JSP와 같은 비즈니스 응용을 위한 인터페이스를 지원한다.

EC-DaMiner 시스템은 e-Business에서의 BI 지원을 위해 다양한 데이터마이닝 기법을 확장할 수 있으며 기존의 데이터베이스 시스템과는 표준적인 인터페이스를

통하여 연동될 수 있는 구조를 갖는다. 또한 이식성이 높은 Java언어를 통해 구현하였으며 Java RMI 기술에 기반한 비즈니스 응용 인터페이스를 제공하고 있다. 마지막으로 비즈니스 어플리케이션들은 마이닝 질의어인 MQL을 통하여 규칙을 탐사하고 탐사된 규칙을 기존의 마케팅 데이터베이스에 모델화하여 반영함으로써 마케팅 전략의 구현을 용이하게 한다.

이 논문의 구성은 다음과 같다. 2장에서는 시스템 설계 및 구현과 관련된 연구에 대하여 기술하고 3장에서는 비즈니스 인텔리전스 및 e-Business 마이닝의 프레임 워크를 제시한다. 4장에서는 EC-DaMiner 마이닝 시스템의 전체적인 구조 설계와 기능 그리고 마이닝 질의어 MQL을 제시한다. 5장에서는 EC-DaMiner의 구현과 그 수행을 기술한다. 6장에서는 가상우체국에서의 수행 사례 및 시스템의 구현에 대한 평가를 제시하고 마지막으로 7장에서는 결론 및 향후 연구를 말한다.

2. 관련연구

이 장에서는 기존의 데이터마이닝 시스템의 특징과 통합방식 및 마이닝 연산의 구조 그리고 인터페이스 측면에서 기존 연구들을 분석하고 BI지원을 위해 이들 연구들이 갖는 문제점을 제시한다.

데이터베이스와 결합된 다양한 마이닝 태스크를 지원하는 데이터마이닝 시스템에 대한 연구는 Quest 프로젝트에서의 Quest시스템[10], DBMiner시스템[11,14] 등이 대표적이다.

Quest 프로젝트는 여러 응용에 걸쳐 적용되는 기본적인 마이닝 연산을 규정하였으며 Quest 마이닝 시스템의 기본 아키텍처를 제시하였다. 마이닝 연산은 크게 연관 규칙 탐사, 순차 패턴 탐사, 시계열 클러스터링, 분류 그리고 점진적 마이닝 등으로 규정하였으며 다수의 연관 규칙 탐사 알고리즘[15,16,17], 순서 패턴 탐사 알고리즘[18,19], 시계열 탐사 알고리즘[20] 그리고 분류 알고리즘[21,22] 및 점진적/능동 마이닝[23] 알고리즘 등이 개발되어 Quest 시스템에 적용되었다.

DBLearn시스템을 확장하여 다단계 개념 수준의 지식 마이닝을 위한 데이터베이스 통합 시스템으로써의 DBMiner는 기존의 일반화(generalization), 특성화(characterization), 연관규칙 및 분류(classification), 예측(prediction) 기법 외에 속성지향 추론, 통계분석 및 다중 레벨 규칙 탐사 기법 등을 포함하고 있다. GUI 또는 SQL과 유사한 DMQL[24] 마이닝 언어를 통한 대화적인 마이닝 지식 탐사를 지원하고 있으며 자동화된 개념 계층의 구성 및 ODBC를 통한 데이터베이스 통합

지원 등을 특징으로 한다.

ADaM 시스템[25]은 관계형 데이터베이스를 기반으로 하지만 트랜잭션 데이터 분석이 아닌 지구 과학 데이터를 분석하기 위한 확장 가능한 시스템이다. 이 시스템은 다양한 데이터 포맷의 파일을 표준 내부 형식으로 처리하는 입력모듈, 내부데이터에 대한 분석 모듈 및 다양한 외부 형식로서의 마이닝 지식을 출력하는 모듈 세 가지로 구성되고 있다. 이 시스템 역시 클라이언트-서버 환경을 지원하는 클라이언트 API를 지원하며 내부의 분석 모듈 역시 확장 가능한 플러그인 모듈로써 구성된다. 이러한 분석 모듈들은 입력 및 출력 형식에 독립적인 내부 표준 데이터포맷으로 데이터를 분석하는 특징을 갖고 있다.

시스템과의 통합 방식은 독립형 엔진 구조에서의 강결합 및 소결합 방식 그리고 마이닝 내장 방식[26]으로서 구분된다. 독립형 구조를 갖는 대부분의 마이닝 시스템[10,12,14]과 데이터베이스 시스템간의 표준적인 인터페이스를 통한 소결합 통합 방식은 마이닝 연산 기능이 데이터베이스 시스템의 기능에 대한 의존 정도에 따라 밀결합(tightly-coupled) 방식과 소결합(loosely coupled) 방식으로 다시 세분화된다[10,14]. Quest시스템을 상용화한 IBM Intelligent Miner[12]는 DB2와의 밀결합 방식을, DBMiner는 Sybase SQL Server와의 밀결합 방식을 취하고 있어 마이닝 연산에서의 데이터 접근을 DBMS의 내장 프로시저 및 인덱스를 사용함으로써 성능 향상의 이점을 갖는다.

데이터베이스 기반의 마이닝 연산들의 프로세싱 구조는 DBMS상의 데이터에 대한 ODBC 또는 SQL 커서 인터페이스 등을 통한 소결합 방식, 파일 시스템 캐싱을 통한 마이닝 방식, 마이닝 연산의 SQL 구현을 통한 방식 그리고 DBMS의 절차적인 내장 프로시저와 마이닝 알고리즘의 밀결합 방식 등이다[7]. [27]에서는 일반화된 연관규칙 탐사 및 순차 패턴 탐사 알고리즘의 SQL 구현을 통해 기존의 SQL 커서 방식에 비해 수행 성능의 향상을 가져왔음을 보였다. [7]에서는 마이닝 연산의 다양한 구현 기법에 대한 성능 비교를 통해 마이닝 연산 시 관계된 데이터에 대한 캐싱 구조가 효율적임을 보였다.

어플리케이션의 태스크 표현 및 처리를 위한 인터페이스 방식은 DBMiner와 같이 대부분의 시스템에서 제공되는 GUI를 이용한 대화식 처리와 DMQL[24]이나 M-SQL[28]과 같은 SQL 유형의 선언적 언어 제공 방식이다. 이러한 방식들은 사용자로부터 마이닝과 관련된 연관데이터 셋의 규정이나 마이닝의 다양한 임계치 표현 그리고 마이닝 결과에 대한 표현 구조 등의 명세를

가능케 한다.

기존의 이러한 연구들은 결합구조 측면에서는 독립적이며 소결합 구조를 갖는 시스템들도 대부분 마이닝 연산 구현에 있어서는 특정 시스템과의 밀결합 프로세싱 방식을 택하고 있어 확장성과 통합측면에 있어 제약이 따르며[12] 대부분의 마이닝 연산이 비절차적인 언어인 SQL로의 구현이 가능하지 못하다는 문제점[27]이 있다. 또한 성능상의 이점에도 불구하고 시스템마다 지원 방식의 차이점으로 인해 DBMS가 지원하는 사용자 정의 함수로써 마이닝 연산을 구현하는 방식은 일반화된 적용이 어렵고[6], 그리고 비즈니스 어플리케이션에서 요구되는 규칙의 모델화 및 이와 관련된 인터페이스를 제공하지 못하는 문제점이 있다[11,25].

이 논문에서는 기존 연구들이 BI지원 시스템으로서 갖는 이러한 문제점과 최근의 비즈니스 데이터 분석을 위한 시스템의 요구사항을 토대로 결합구조의 문제점을 극복하고 시스템 구조와 확장성을 위한 연산 구조로써 마이닝 연산을 시스템 독립적으로 구현한다. 이러한 DBMS 독립적인 구조로써 다양한 시스템과의 통합 및 이식성을 제공할 수 있으며 시스템은 추가로 구현된 마이닝 연산에 대하여 확장성을 갖는다. 소결합 시스템에서의 프로세싱은 밀결합 시스템에서의 프로세싱에 비하여 수행성능이 저하되는 문제점이 있지만 이러한 문제점은 마이닝 연산의 병렬적 수행 및 캐싱 메커니즘 방식을 통하여 해결 할 수 있다.

3. 비즈니스 인텔리전스

3.1 비즈니스 인텔리전스의 정의

비즈니스 인텔리전스는 데이터베이스나 데이터웨어하우스 등의 다양한 데이터 소스를 기초로 과거의 경험과 현재의 비즈니스 환경 변화를 정확히 이해하여 얻어진 정보 등을 통합적으로 이용하여 비즈니스 의사결정을 지원하는 개념과 방법 그리고 프로세스들의 집합이다. 즉, 효율적인 비용과 빠른 접근을 통해 비즈니스 질의에 응답하고 경향을 파악해서 적극적으로 경쟁력 있는 비즈니스를 지원하는 지식을 제공하는 것이다. [3]에서 제시된 바와 같이 BI개념은 새로운 것은 아니며 과거 10년간 1세대 호스트 기반 질의/레포팅 시스템과 EIS, 2세대 데이터 웨어하우스와 DSS 그리고 3세대 비즈니스 인텔리전스로 발전되었다.

3세대로 구분되는 BI시스템은 2세대의 기술적인 해결책을 기반으로 비즈니스의 실질적인 문제를 해결할 수 있도록 하며 데이터 분석가 및 의사결정권자뿐만 아니라 일반 고객 등과 같이 광범위한 요청자들에게 비즈니스

스 정보의 효율적 접근 및 전달을 목적으로 하고 있다. 이러한 것들은 OLAP과 데이터마이닝 기술 등을 기반으로 다양한 어플리케이션의 제공을 통해 달성되고 있다. BI 시스템의 목적은 첫째, 수입을 증대하고, 비용을 줄이며 효과적으로 경쟁할 수 있는 지식을 제공한다. 둘째, 빠르게 변화하는 사용자 요구와 복잡한 비즈니스 환경에서의 요구를 모델링한다. 마지막으로 기존의 중복적이고 비일관적인 비즈니스 정보를 통합하고 경제적으로 경감하는 것이다.

3.2 비즈니스 인텔리전스 시스템 구조

Colin은 BI시스템은 그림 1과 같이 다양한 컴포넌트들로 구성된다고 보았다[3]. 이러한 구조는 이전의 데이터웨어하우스와 OLAP를 이용한 비즈니스 정보 시스템의 구조에서 발전된 구조라고 볼 수 있다. 최상위의 BI 어플리케이션은 BI시스템을 구성하는 최상부로서 특정한 기업 또는 응용 영역을 지원하는 솔루션 패키지를 말한다. 이러한 패키지들은 BI시스템 하부의 기능을 이용하여 개발된다. 비즈니스 인텔리전스 어플리케이션의 예로는 데이터마이닝 수행 결과에 대한 비주얼 툴 또는 OLAP 질의 도구 및 분석 툴 등이 있다. 의사결정 지원 툴(decision support tools)은 기본적인 질의 및 보고서 툴로부터 진보된 OLAP나 데이터마이닝 도구에 이른다. 기본적으로 의사결정 지원 툴들은 GUI기반의 클라이언트 툴이 대부분이며 최근 들어서는 Web기반 UI를 통합적으로 제공하기도 한다. 이러한 툴들은 파일 시스템과 다양한 유형의 데이터베이스 시스템에 의해 유지되는 구조적 또는 비구조적 정보를 다룰 수 있도록 개발된다. 메타데이터 관리 컴포넌트는 전체 비즈니스 BI시스템과 관련된 메타데이터를 관리한다. 이러한 메타데이터는 개발자나 관리자에 의해 사용되는 기술적인 데이터와 비

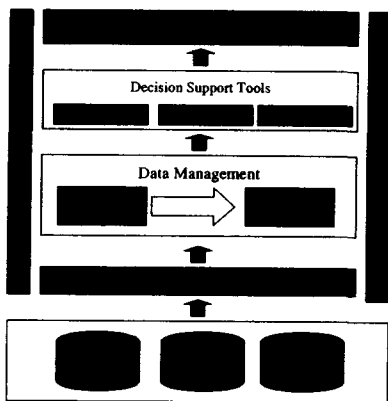


그림 1 전형적인 BI 시스템 구조

즈니스 사용자를 지원하는데 이용되는 비즈니스 메타데이터 등이 있다.

3.3 프로세스 모델

e-Business의 BI지원 시스템을 설계하기에 앞서 비즈니스 영역에서의 BI를 지원하기 위한 여러 프로세스 단계를 규정하고 이들 단계와 각 단계간의 관계성을 규정할 필요가 있다. 데이터에 기반한 의사결정 지원 모델에 대한 많은 연구 중에서도 [29]는 BI지원 프로세스들과 각 프로세스간의 관계를 비즈니스 인텔리전스 가치사슬(business intelligence value chain)로써 제시하고 있는데 이 연구에서는 그림 2와 같이 데이터마이닝을 이용한 백엔드 시스템으로써의 마이닝 시스템의 설계를 위해 BI 프로세스 모델로 구체화하였다. 모델에서 각 단계간에 의사결정 프로세스를 위한 가치는 증대된다.

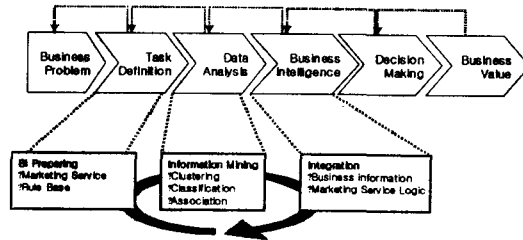


그림 2 BI 프로세스 모델

EC-DaMiner 시스템은 BI 프로세스 중 태스크 정의(task definition)단계에서 BI단계까지의 부분을 e-Business에 적용 가능한 부분으로 규정하고 이를 좀더 구체화하여 BI준비(BI preparing)단계, 정보마이닝(information mining) 단계 그리고 통합(integration)단계 지원을 중요하게 고려하였다. 따라서 BI 준비단계에서는 e-Business 마케팅 서비스를 주요 태스크로 규정하고 각 서비스를 위한 규칙베이스 구성이 결정된다. 두 번째 단계인 정보마이닝 단계에서는 주요 마이닝 연산을 통한 마케팅 태스크 지원 규칙을 생성한다. 그리고 마지막 통합단계에서는 백엔드의 비즈니스 정보와의 통합 및 마케팅 태스크를 처리하는 로직과의 통합을 지원한다.

3.4 데이터마이닝 기법

BI 시스템 프레임 워크에서는 의사결정 지원 툴로서는 데이터마이닝 기법이 적용됨을 보았고 또한 BI 프로세스 모델에서는 데이터 분석(data analysis) 단계에서 정보마이닝을 위해 데이터마이닝 기법이 적용됨을 제시하였다. 이 연구에서는 데이터마이닝을 수행함으로써 얻을 수 있는 지식에 따라 데이터마이닝을 분류(classifi-

cation)와 예측(prediction), 평가(estimation), 유연 그룹화(affinity grouping) 혹은 장바구니 분석(market basket analysis), 클러스터링(tering), 기술(description)의 다섯 가지의 태스크로 구분하였다. 각각의 태스크는 어느 한 가지의 데이터마이닝 툴이나 테크닉(또는 알고리즘)을 동일하게 사용하지는 않는다. 즉, 데이터마이닝을 통해 수행할 수 있는 태스크는 다양한 데이터마이닝 테크닉을 응용함으로써 이루어진다. 데이터마이닝 기법들에 대한 자세한 설명은 [4,5]를 참조한다.

4. EC-DaMiner 시스템의 설계

4.1 마이닝 시스템 구조

e-Business에서 비즈니스 인텔리전스를 적용하는 대표적인 예로는 고객관계관리(CRM)를 들 수 있다[30]. 일반적으로 CRM을 위한 시스템 구성은 고객과의 채널로서의 프론트 엔드시스템과 이를 지원하는 백엔드 시스템으로 구분된다. 그림 3은 CRM을 위한 백엔드 시스템에 BI 적용 모델을 보여준다. 이 구조에서의 데이터마이닝 시스템은 고객과의 상호작용을 하는 프론트엔드 인터페이스 채널로서 Web, Email, SFA 및 CTI 등을 통해 생성된 다양한 데이터로부터 지식을 추출하고 개인화 시스템 및 캠페인 관리 시스템과 같은 다른 어플리케이션에 적용 가능하게 규칙베이스화 한다. 전체적인 구조에서 보통 데이터마이닝 시스템은 고객 신상정보, 구매 이력, 상담이력 및 웹로그 정보 등을 통하여 구매 성향 및 이탈 가능 고객 판별 등과 같은 고객 관련 지식을 추출한다. 이 논문에서 제시하는 EC-DaMiner 시스템은 이러한 모델에서의 적용을 전제로 하며 다음과 같은 요구사항을 기반으로 설계하였다.

첫째, 다양한 데이터마이닝 기법을 제공할 수 있는 확장성 있는 시스템이어야 한다.

둘째, 특정 시스템에 종속적이지 않은 데이터베이스

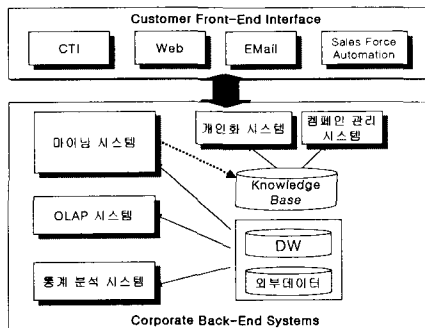


그림 3 CRM을 위한 Business Intelligence 적용 모델

시스템들과의 통합을 지원한다

셋째, 시스템은 다양한 플랫폼에서 수행 가능한 이식성을 갖는다.

넷째, 비즈니스 응용을 위한 인터페이스를 지원한다.

EC-DaMiner는 CRM 시스템과 같이 개별 고객의 지속적인 가치를 증대하기 위한 다양한 지식을 관리자 및 응용 어플리케이션 개발자에게 제공하도록 DBMS기반의 기존의 프론트 어플리케이션에 유연하게 통합될 수 있는 구조로 설계하였다.

현재 EC-DaMiner 시스템은 분류, 클러스터링 및 연관규칙 탐사 등의 대표적인 세 가지 범주의 마이닝 태스크를 수행한다. 데이터마이닝 태스크를 표현하고 처리하기 위한 마이닝 질의어로서 MQL(Mining Query Language)를 정의하였으며 엔진은 이 MQL 질의를 분석하고 내부의 연관규칙 탐사모듈 및 분류기, 클러스터링 모듈 등을 통하여 연산을 수행하고 이를 지식베이스에 저장한다. EC-DaMiner 마이닝 엔진의 구조는 그림 4와 같다.

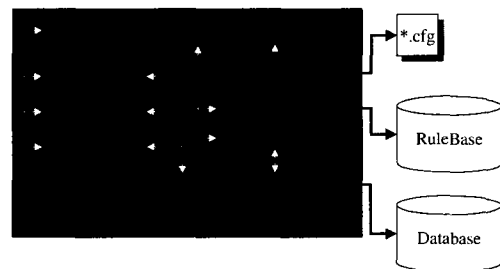


그림 4 EC-DaMiner 마이닝 엔진 구조

4.2 시스템 기능

데이터마이닝 시스템의 최상 레이어로 MQL processor는 사용자로부터의 마이닝 요구를 표현하는 마이닝 질의를 분석하고 처리하는 파서와 실행기를 말한다.

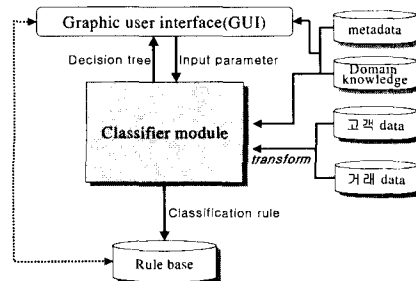


그림 5 분류기 모델

분석된 마이닝 질의는 시스템에 포함된 마이닝 연산 프리미티브 등을 통해 처리되게 된다. 질의처리는 이러한 연산 수준의 프리미티브들로 구성되는데 구현되는 마이닝 엔진은 연관규칙 탐사 모듈, 분류 모듈, 클러스터링 모듈 등을 포함한다.

위의 그림 5는 엔진의 분류기 모델로서 이 모델에서는 분류를 위한 배경지식과 분류 모델 생성을 위한 데이터 간의 흐름을 보여준다.

각 마이닝 연산을 수행하는 프리미티브는 표준적인 데이터베이스 인터페이스 레이어를 통하여 실제 마이닝 될 고객, 구매, 상품 등에 대한 정보를 담고 있는 데이터베이스에 접근하고 결과를 생성한다. 관계형 데이터베이스에 대하여는 SQL을 통하여 데이터에 접근한다.

마이닝 연산에 필요한 메타데이터 정보는 메타데이터 관리기를 통해 관리되며 각 연산 수행 시 이용된다. 메타데이터 관리기는 마이닝을 수행하는 데 필요한 정보 및 마이닝을 통해 획득된 정보 등을 관리한다. 주요 메타데이터는 데이터베이스에 대한 테이블의 구조 및 주요 컬럼에 대한 정보, 획득된 연관 규칙 관리를 위한 연관 규칙 테이블 등에 관한 정보 그리고 분류규칙 테이블에 관한 정보 등이다.

마지막으로 마이닝 질의의 수행결과로 규칙 베이스(Rule Base)를 생성한다. 규칙 베이스는 기반이 되는 여러 관계형 데이터베이스 시스템의 데이터베이스로부터 추출된 정보를 표현하는 부분이다. 이러한 규칙 베이스의 구성요소는 연관규칙, 분류규칙 그리고 클러스터링 규칙 등이 있다. 규칙 베이스는 시스템 내부 저장구조로써 관리될 수 있지만 관계형 데이터베이스 기반 비즈니스 어플리케이션과의 통합을 통한 규칙의 적용을 위해 접근되는 관계형 데이터베이스 시스템상의 테이블로 표현한다. 응용 프로그램은 마이닝 시스템에서 구성하는 규칙 스키마를 기초로 응용을 구현한다. 따라서 규칙 스키마 자체에 대한 종속성은 포함하고 있게 되지만 규칙 자체의 제약은 없으며 규칙의 유효성을 스키마내에 포함시킴으로써 응용에서는 유효성에 기반하여 규칙을 활용한다.

4.3 데이터마이닝 질의언어

EC-DaMiner시스템의 마이닝 질의어 MQL은 DMQL [24]과 유사한 SQL형식으로 다음과 같은 몇 가지 사항들을 고려하여 설계되었다[31].

첫째, 마이닝 질의어는 마이닝 기법들 중 e-Business 적용될 수 있는 기법들을 포괄할 수 있는 일반적인 형식을 취한다. 광범위한 영역의 연구 기법 등이 복잡하고 다양한 마이닝 패턴의 탐사에 적용될 수 있기 때문에

이를 다 포괄적으로 수용할 수 있는 마이닝 언어의 설계는 상당히 어려운 점이 많다.

둘째, 마이닝 질의어를 통해 마이닝 질 관련성 있는 데이터 세트를 명시할 수 있어야 한다. 마이닝 작업은 상당히 방대한 양의 데이터를 이용하기 때문에 연관 데이터 세트를 형성하는 것은 마이닝 작업의 초기단계에서 성능을 위한 중요한 부분이다.

셋째, 마이닝 질의는 그 결과로써 고유한 규칙이 생성되고 사용자에게 의해 식별되고 평가될 수 있는 형태로 저장 관리되어야 한다. 즉 마이닝 질의는 해당하는 질의 유형에 따라 규정된 지식의 표현 방법으로 결과를 생성한다.

넷째, 마이닝 결과에서 유용한 지식을 검출해 내기 위한 기준으로써 각 연산에 따른 임계치(threshold)의 표현이 가능하여야 한다.

EC-DaMiner 시스템은 관계형 데이터베이스에 대한 SQL 인터페이스를 기반으로 하였기 때문에 이 논문에서는 이를 고려하여 SQL과 유사한 형식의 MQL을 설계한다. 위에서 언급된 사항을 고려하여 MQL을 통해 탐사하는 지식의 유형은 다음과 같다.

- 연관규칙(association rule) : 하나의 연관규칙은 트랜잭션을 구성하는 데이터(항목) 간의 관련성을 기술한다.
- 분류규칙(classification rule) : 분류규칙은 일반적으로 클래스가 설정되어 있는 데이터 셋을 통해 연관된 데이터를 어떻게 분류할 것인지에 대한 모델(혹은 기술)을 획득한다.
- 클러스터링 규칙(clustering rule) : 클러스터링은 데이터를 유사성에 기반하여 몇 개의 그룹으로 구분한다.

EC-DaMiner 시스템의 마이닝 질의어 MQL은 기존의 DMQL과 유사한 형식으로 서론에서 언급한 바와 같은 시스템의 설계 원칙 및 앞서 언급한 고려사항들을 기초로 하여 설계하였다. 마이닝 질의어는 여러 응용분야에서의 마이닝 수행 태스크 수행을 위한 필수 요소로써 다양한 설계 및 구현 연구가 진행되고 있다. 제시된 MQL의 다수의 다른 유형의 마이닝 태스크의 표현이 포함되지 못한 점과 또한 다른 그래픽적인 요소로의 규칙의 변환 등과 관련하여 제한점을 갖고 있어 이에 대한 추가적인 연구가 진행되어야 할 것이다.

4.4 마이닝 질의어의 질의 형식

마이닝 질의어 MQL의 구문구조는 그림 6과 같이 확장된 BNF형식으로써 표현하였다.

MQL은 크게 규칙 질의 구문과 내포된 SQL 질의 구

문으로 나누어 진다. 규칙질의 구문은 연관규칙, 분류규칙, 클러스터링 규칙을 표현하는 구문을 말하고 내포된 SQL질의 구문은 표준 데이터베이스 인터페이스를 통하여 SQL을 수행시키고 그 결과를 디스플레이 할 수 있는 구문으로 해당 데이터베이스에서 수행되는 질의어를 그대로 포함하게 된다.

```
MQL ::= use dataset <dataset >
      <rule_mining> SEMI
      | use dataset <dataset >
      <sql_query> SEMI

rule_mining ::= <rule_spec>
              related to <attribute_list>
              from <relation_list>
              [where {<condition_string>}]
              {with {<kind_of>} threshold = <threshold_value>}

sql_query ::= sql { <sql_string>
```

그림 6 마이닝 질의어 MQL 구문 구조

MQL의 시작인 *use*절은 데이터마이닝을 수행할 데이터세트를 지정하는 것으로 시스템상에서 정의된 이름이다. 이 이름을 통하여 데이터베이스 인터페이스를 위한 필요한 정보를 전달하게 된다. MQL의 <rule_mining> 비단말에서 앞서 언급한 세 가지 마이닝 규칙을 포함하는 비단말은 <rule_spec>이다. <rule_mining> 비단말에서 *related to*절을 통하여 연관된 속성을 규정하며 *from*절을 통해 관련된 릴레이션 리스트를 기술하게 된다. 그리고 선택적인 *where*절을 통해 튜플을 한정하는 조건을 기술하게 된다.

<rule_spec>에서 각 규칙에 대한 구문은 다음과 같다.

```
<rule_spec> ::=
  find association rule as <rule_name>
  | find classification rule as <rule_name>
  [ according to <attribute_list>]
  | find cluster
  for <cluster_name> of <#clusters>
```

탐사된 연관규칙은 정의된 <rule_name>을 통해 저장되고 관리되어진다. 분류규칙의 탐사 시 타겟 클래스와 관련된 속성은 *according to* 절을 통하여 기술하게 되며 탐사된 규칙은 <rule_name>을 통해 저장되고 관리된다. 클러스터의 탐사는 클러스터의 최대 개수를 한정하는 <#cluster>에서 정의된 수 만큼을 탐사하며 정의된 <cluster_name>을 통해 저장되고 관리되어진다.

탐사된 규칙에 따른 임계치를 명시하는 *with*구문에서 임계치의 종류를 나타내는 <kind_of> 비단말은 다음과 같다.

```
with {<kind_of>} threshold = <threshold_value>
<kind_of> ::= confidence | support | probability
```

각각의 규칙에 따라 적용되는 임계치는 다른데 연관규칙의 경우에는 confidence와 support 임계치가 적용되고 분류규칙의 경우에는 probability가 적용된다. 임계치의 값 <threshold_value>는 0 ~ 1내의 실수 값을 말한다.

5. EC-DaMiner 구현

e-Business를 수행하기 위한 시스템은 크게 쇼핑몰 시스템과 같은 프론트엔드 시스템과 BI지원을 위한 OLAP이나 데이터마이닝 모듈로 구성되는 백엔드 시스템으로 나눌 수 있다. 이 절에서는 BI를 지원하기 위한 데이터마이닝 시스템인 EC-DaMiner 엔진과 사용자 UI의 구현하고 사용자 UI를 통해 마이닝 연산을 수행한 결과를 기술한다.

5.1 시스템 구현

EC-DaMiner 시스템의 구현환경으로 엔진은 Spark Ultra 1s 워크스테이션의 UNIX 운영체제 하에서 Java 언어 1.2로 구현하였다. 사용자를 위한 UI는 Windows 98환경에서 Java언어와 JDBC 등을 사용하여 구현하였다. EC-DaMiner 마이닝 테스트를 위한 마케팅 데이터베이스 및 규칙 베이스는 Oracle DBMS 7.3엔진을 이용하였다.

5.2 EC-DaMiner

EC-DaMiner는 데이터 분석을 필요로 하는 사용자 및 관리자가 쉽게 데이터마이닝 엔진을 통하여 분류 및 연관규칙 등을 탐사하고 이를 가시화할 수 있도록 하기 위하여 개발된 마이닝 도구이다. 다양한 플랫폼으로의 이식성을 위하여 Java언어를 통하여 구현되었으며 지식관리를 담당하는 규칙 베이스 및 지식 탐사를 수행하는 데이터마이닝 엔진과의 인터페이스를 제공하고 있다. 효율적인 고객관리 및 마케팅 전략의 수립과 적용을 위하여 데이터마이닝 시스템은 기존의 쇼핑몰 시스템과 같은 전방시스템과 자연스럽게 통합될 수 있어야 한다. 앞장에서는 데이터마이닝 엔진의 구조와 기능을 설명하였고 이 엔진을 통하여 규칙을 생성하기 위한 기본 연산들 그리고 연산들을 통해 하나의 마이닝 작업을 표현하기 위한 마이닝 질의언어를 제시하였다. 데이터마이닝 엔진은 연관규칙 탐사 및 분류를 위한 연산 기능을 포함하고 있다. 이러한 연산을 통하여 규칙 탐사의 수행은 정의된 MQL 구문에 맞게 작성되어 마이닝 엔진에 의해 질의가 처리된다. 마이닝 엔진은 기존의 관계형 데

이타베이스와의 표준적인 인터페이스를 통하여 데이터 분석을 위한 접근을 시도하며 또한 생성된 규칙을 규칙 베이스 상에 저장하게 된다. 그림 7은 마이닝 엔진과 마이닝 UI 및 프론트 엔드 시스템, 데이터베이스 간의 연동구조를 보여준다.

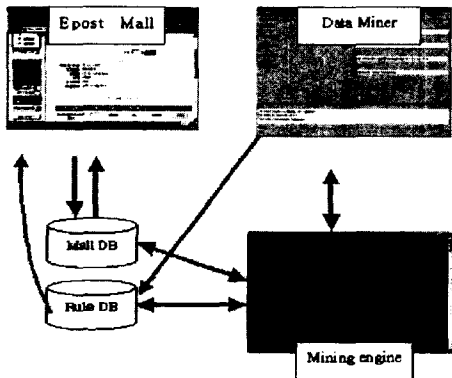


그림 7 BI 지원 마이닝 시스템 구성도

위의 그림 7과 같이 Data Miner UI는 데이터마이닝 엔진 그리고 규칙베이스에 접근을 위한 인터페이스를 제공한다. 마이닝 엔진과의 접근은 RMI기반의 메소드 호출을 통해 MQL 질의를 수행하며 규칙 베이스와는 표준적인 JDBC인터페이스를 통해 배경 지식 및 규칙 등을 접근한다.

그림 8과 같이 Data Miner UI는 Connection, Association Rule탐사, Classification, MQL Query 등을 위한 5개의 패널로 구성되어있다. 그림 8은 Connection 패널을 나타낸다. 마이닝을 수행할 데이터 베이스와의 연결을 설정하는 Connection 패널에서는 연

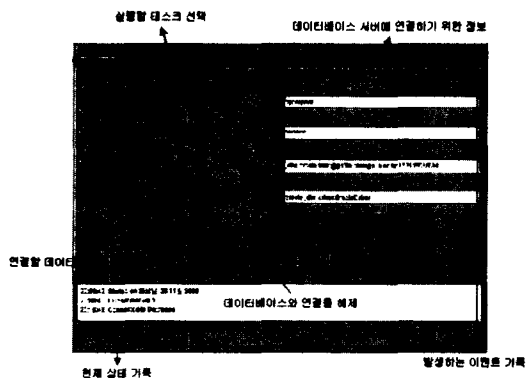


그림 8 EC-DaMiner의 Connection 패널

결합 데이터베이스 드라이버의 종류와, 서버에 대한 사용자 이름, 패스워드, URL, JDBC 드라이버 정보를 입력 받아 주어진 정보로 JDBC를 통해 데이터베이스와의 연결을 수행한다

EC-DaMiner에서는 Apriori알고리즘[16]을 이용한 연관규칙을 탐사 모듈을 구현하였으며 이 기법은 간략히 아래의 세 단계로 구분된다.

1. Large itemsets 발견(minimum support 고려)
2. 데이터베이스에 대해 연관규칙을 생성하기 위해 Large itemsets을 사용한다.
3. 생성된 주요 항목집합을 토대로 규칙을 생성하는 과정이다.

그림 9는 Association 패널을 보여준다. Association 패널은 연관규칙이 생성될 테이블을 선택하는 부분과 연관규칙과 관련된 지지도와 신뢰도를 지정하는 부분, 규칙 생성 후 규칙 베이스에 저장하도록 규칙 이름을 입력하는 부분으로 구성된다.

기계 학습 분야나 통계학 분야에서는 다수의 분류 기법의 소개되어 왔으며, 이 중 가장 많이 쓰이는 기법은 결정 트리를 이용한 분류 기법이다. 결정 트리에 기반한 분류기는 다른 분류 기법과 비교해 볼 때 상대적으로 빠르고 간단하며 이해하기 쉬운 분류 규칙으로 전환이 쉽다는 장점을 가진다. 이러한 결정 트리 알고리즘으로는 이전의 ID3나 CART와 같은 알고리즘이 제시되었다. 이들 기법은 상대적으로 적은 수의 데이터 셋을 갖는 것에 잘 적용되는 특성을 가지고 있다. 하지만 데이터 마이닝에서는 데이터가 대량이라는 특성을 가지고 있기 때문에 효율성과 확장성(scalability)에 대한 문제가 제기된다. 따라서 대용량의 트레이닝 셋을 처리하기 위한 알고리즘으로 SLIQ나 SPRINT와 같은 결정 트리

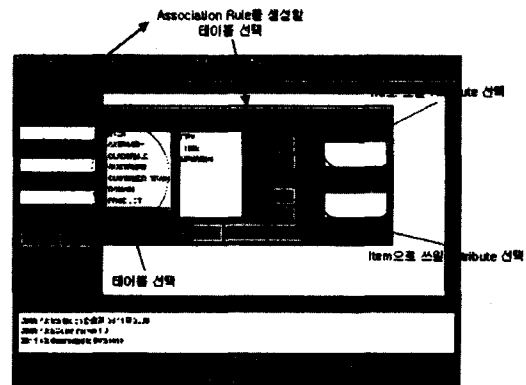


그림 9 연관규칙을 탐사할 테이블 선택

알고리즘이 제시되었다.

EC-DaMiner의 분류규칙 탐사 모듈은 이중 C4.5 알고리즘을 BI 시스템의 특성에 맞게 변형, 구현하였다[31]. C4.5가 결정 트리를 생성하는 과정은 다음과 같다.

- 단계 1: 한 속성을 트리의 루트노드로 선정하고 이 속성이 가질 수 있는 모든 값을 가지고 가지를 만든다.
- 단계 2: 어느 특정 리프노드에서 모든 레코드가 같은 클래스에 속하면 그 리프노드는 그 클래스의 이름이 붙는다. 모든 리프노드가 클래스의 이름이 붙으면 알고리즘은 끝난다.
- 단계 3: 그렇지 않으면 이 노드는 루트까지의 패스에 나타나지 않은 속성의 이름이 붙고, 속성의 가능한 값들에 대해 가지가 생성된다. 이 알고리즘은 2단계로 가서 모든 리프노드에 클래스의 이름이 붙을 때까지 계속 된다.

그림 10의 Classification 패널은 크게 분류규칙을 생성할 트레이닝 데이터로 쓰일 테이블을 선택하는 부분과 관련 인수 설정, 분류 규칙을 트리 형태로 가시화, 생성된 분류규칙을 데이터베이스에 적용하는 부분으로 나누어진다. 그림 10은 트레이닝 데이터로 쓰일 테이블을 선택하는 과정을 보여준다. 분류를 위해서는 분류를 수행할 테이블과 관련속성 선택 후, 분류기에 관련된 인수들이 설정되어야 한다. 결정할 인수들은 리프노드에 들어갈 최소 인스턴스의 수, 속성의 값에 대한 이진분할 여부, 트리를 가지치기 할 것인지 여부, sub tree raising을 할 것인지 여부와 트리를 가지치기 할 경우 사용되는 지지도가 있다.

그림 11은 분류기가 생성된 후에 나타나는 결과를 트리 형태로 가시화하는 화면이다. 결정 트리에서 각 인스턴스의 속성값에 대한 테스트를 수행하는 내부노드는 타원형으로 표시되고, 단말 노드는 사각형으로 표시된

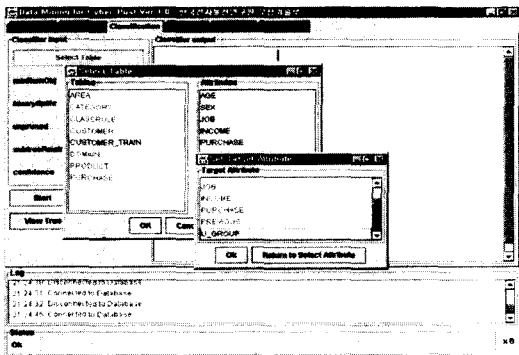


그림 10 분류가 수행될 테이블 선택

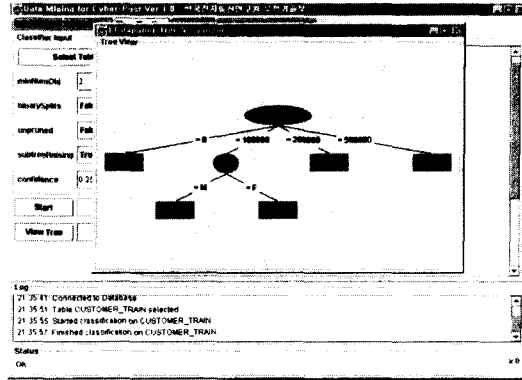


그림 11 분류기를 트리형태로 가시화

다. 루트 노드에서 시작해서 각각의 단말노드까지 가는 경로들은 각각 하나의 분류규칙이 된다.

6. 가상우체국 BI지원에의 적용 및 평가

구현된 EC-DaMiner 시스템의 BI 적용을 위하여 현재 구축되어 운영 중에 있는 우체국 전자상거래 쇼핑몰 사이트(<http://www.epost.go.kr>)와 유사한 형태로 구축하였다. 쇼핑몰은 Apache 웹서버를 사용하였으며 Oracle DBMS 7.3 상에 쇼핑몰 데이터베이스를 구축하였다. 상품검색 및 구매 등의 비즈니스 프로세스 구현은 자바 서버릿으로 구현하였다.

6.1 가상 우체국 전자상거래 머천트 시스템과의 연동

전자상거래 시스템에서의 DB마케팅 역할은 머천트시스템 전반에 걸쳐 있으며 고객관리를 통해 판매를 높이는 것을 핵심으로 한다. 전자상거래 시스템에 마이닝 규칙을 적용하기 위해서는 마이닝 시스템의 관리자가 쇼핑몰 DB의 데이터를 이용하여 마이닝 규칙을 생성해야 한다. 관리자는 적용 도메인에 적합한 테이블과 속성을 선택하고 적절한 임계치를 반복적으로 수행하면서 의사결정에 적용할 수 있는 규칙을 생성하게 된다. 비즈니스 프로세스를 구현한 각 서버릿은 Java RMI 인터페이스를 통해 데이터마이닝 엔진으로부터 마이닝 요구를 요청하거나 JDBC를 통하여 데이터베이스로부터 탐색된 규칙을 검색하여 적용하게 된다.

6.2 마이닝 기법의 적용

마이닝 결과로 생성되어질 연관규칙, 분류규칙, 클러스터링 규칙 등은 현재 운영중인 전자상거래 시스템의 프로세스에 적용할 수 있다. 일반적인 전자상거래 쇼핑몰은 상품을 구매하기 전에 상품에 대한 정보를 브라우저를 통해 검색하며 장바구니에 구매할 상품을 추가하

고 마지막으로 결제하는 일반적인 전자상거래 프로세스 내에 다음과 같은 방법으로 마케팅을 적용할 수 있다.

- **크로스 상품 제안:** 쇼핑물에서 특정 상품을 구매할 경우 이와 연관성이 높은 다른 상품을 연계하여 판매하는 것을 크로스 세일링 마케팅 기법이라고 하는데 이를 구현하기 위해서 연관 규칙 탐사 테이블을 이용할 수 있다.
- **업세일링 상품 제안:** 고객이 어느 시점에 특정 상품을 구매하면 해당 구매 기록을 바탕으로 이후에 새로운 제품이 출시되었을 경우에 이를 연계시켜 판매하는 것을 업세일 기법이라고 하는데 이를 구현하는 데 있어 연관 규칙 테이블을 이용할 수 있다.
- **타켓 상품 제안:** 고객의 구매 성향 등을 파악하여 그룹단위 또는 특정 대상에 대하여 상품 구매를 유도하는 것을 타켓 마케팅이라 하는데 분류규칙 테이블과 클러스터링 규칙 테이블을 이용할 수 있다.

6.3 수행

탐사된 마이닝 규칙은 다양한 비즈니스 프로세스에 적용될 수 있다. 여기서는 고객의 특성이 고려되지 않은 상품간의 연관성 탐사규칙의 적용과 고객의 특성이 고려된 분류규칙과 연계된 수행을 보인다.

그림 12에서는 일반적으로 쇼핑물에서 각 사용자의 특성이 고려되지 않은 상품들간의 연관성 규칙 탐사결과를 기반으로 특정 상품에 대한 정보를 디스플레이 할 때 해당 상품과 연관성이 높은 상품을 추천하는 추천 마케팅 전략에 사용되고 있는 사례를 보여준다.

그림 13에서는 특정 고객의 장바구니 내의 구매 상품간의 연관성을 기존에 탐색과 마이닝 결과를 바탕으로 하여 연계한 추천상품 소개를 보여준다. 이외에도 패키지 상품내의 상품과 사용자의 관심이 있는 상품간의 관계성에 기반하여 추천하거나 고객 프로필을 기반으로 분류된 우량고객과 장바구니 연관성을 연계한 추천을

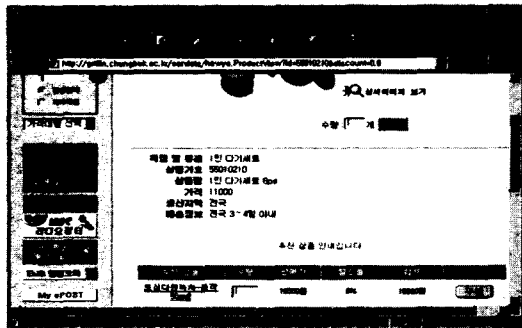


그림 12 상품 자세히 보기에서의 연관상품 추천

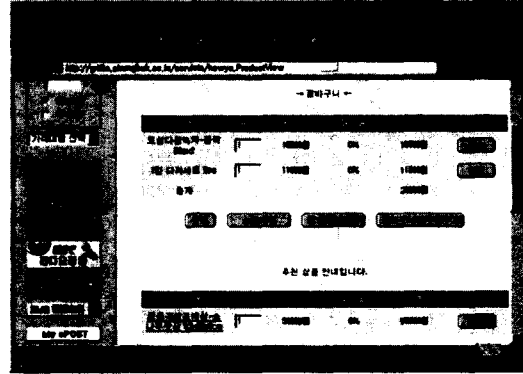


그림 13 장바구니 분석을 통한 연관상품 추천

수행할 수 있다.

6.4 평가

수행 사례에서 제시된 바와 같이 BI를 지원하는 데이터마이닝 시스템은 시스템 독립적인 통합성과 이식성 그리고 응용인터페이스 지원이라는 세 가지 요구사항을 만족한다.

표 1과 같이 EC-DaMiner는 기존의 데이터베이스 시스템과 JDBC를 이용한 표준적인 인터페이스 방식을 통하여 소결합 구조의 독립 시스템으로써 구현되었고 시스템 내부의 주요 마이닝 연산인 분류, 클러스터링, 연관 규칙 탐사 모듈 역시 특정 시스템과 밀접한 형태로 구현되지 않고 순수한 Java클래스로써 구현되었다. 따라서 EC-DaMiner 시스템은 다양한 이기종 플랫폼으로 이식될 수 있으며 기존의 DBMS와 유연하게 통합될 수 있는 구조적 특성을 갖는다. 또한 EC-DaMiner의 마이

표 1 EC-DaMiner의 특징 및 문제점

설계 목표	구현 방식	주요 특징	문제점
BI 통합성	-독립 마이닝 시스템 -연산의 소결합 구조	-JDBC를 이용한 데이터 접근 -마이닝 연산자의 독립적 연산구조 -분류, 클러스터링, 연관 규칙 탐사 지원	-DW 접근 구조 확장 -병렬수행 및 캐싱 연산 구조 지원 필요
플랫폼 이식성 및 확장성	-Java 기반 구현 -마이닝 질의어 명세	-마이닝 엔진의 이식성 -마이닝 연산의 확장성 -MQL을 이용한 태스크 표현	-마이닝 연산 확장 -MQL의 Rule 처리 명세 강화
BI 응용 인터페이스	-원격 메소드 호출 -Rule Base 구현	-RMI 기반 C/S 구조 지원 -JSP 및 Servlet 지원 -Rule 스키마를 통한 응용에의 적용	-Rule 변환 및 가시화 지원

닝 수행 결과는 마케팅 데이터베이스 내에 구현되기 때문에 앞의 수행 사례에서와 같이 상품 추천 및 구매 등의 비즈니스 프로세스에서는 각 규칙의 스키마에 대한 접근을 통하여 쉽게 규칙을 적용할 수 있다.

현재 구현된 EC-DaMiner는 표 1에서 제시되고 있는 바와 같이 연산 성능 향상을 위한 구조적인 지원과 마이닝 연산의 확장 및 Rule 처리 그리고 가시화 지원을 위한 인터페이스 설계 등이 추가 적으로 보강되어야 하는 문제점을 갖고 있다.

7. 결론

경쟁적인 E-Business에서 다양한 고객과의 지속적인 마케팅 전략을 구사하기 위하여 마이닝 시스템은 BI의 핵심이다. 특히 데이터에 대한 판매에 대한 예측과 분석 그리고 고객에 대한 개인화 서비스 및 판매 향상을 위해 다양한 데이터의 분석과 그 적용이 빠른 시간 간격으로 구현되어야 하기 때문에 기존의 비즈니스 시스템과 통합되어 효율적인 지식을 추출하고 이를 적용할 수 있게 하는 데이터마이닝 시스템의 중요성이 커지고 있다.

이 논문에서는 e-Business환경에서의 BI에 대한 정의와 BI지원 시스템의 전반적인 구조 및 프로세스 모델을 분석하였다. 또한 e-Business에서의 BI 지원을 위해 기존의 연구들이 갖는 문제점을 제시하고 이를 해결하기 위해 시스템 독립적인 통합성과 이식성 그리고 응용 인터페이스 지원 등의 세 가지 요구사항을 만족하는 데이터마이닝 시스템인 EC-DaMiner를 설계/구현하였다. 또한 BI의 핵심으로서 다양한 데이터마이닝 태스크를 수행하는 소결합 방식의 독립 데이터마이닝 시스템을 가상우체국과 같은 기존의 시스템과 통합하여 다양한 비즈니스 마케팅 구현에 적용될 수 있음을 수행 사례를 통하여 보였다. EC-DaMiner시스템은 기존의 마이닝 시스템과 같이 다양한 응용 영역에 적용이 가능하며 특히 BI 지원을 위한 백엔드 시스템으로써 비즈니스 데이터 분석을 위한 특성을 포함하고 있다. 시스템은 다양한 데이터마이닝 연산을 수행하는 마이닝 연산모듈 들을 핵심으로 하고 있으며 기존의 데이터베이스와의 표준적인 데이터베이스 인터페이스로써 마이닝을 수행한다. 데이터마이닝 태스크 연산을 위한 인터페이스로써 마이닝 질의어인 MQL을 이용한다. MQL의 수행은 마이닝 태스크에 따라 규칙 베이스를 구성하며 다양한 비즈니스 프로세스 구현에 이용될 수 있다

향후에는 MQL에 포함되지 않은 다양한 마이닝 연산 및 e-Business 영역에서의 시간 지식 탐사를 위해 시간 데이터마이닝 모델 기반의 마이닝 연산 등을 포함하도

록 마이닝 질의어를 확장해야 한다. 또한 시간규칙 탐사 연산과 시간 규칙의 효과적인 생성 및 규칙의 관리 등에 관한 연구가 필요하다.

참고 문헌

- [1] Q. Chen, U. Dayal, and M. Hsu, "OLAP-Based Data Mining for Business Intelligence Applications in Telecommunications and E-commerce," In Proc. of the Int'l Workshop on Databases in Networked Information Systems, pp.1-19, 2000.
- [2] J. Mrazek, "Data Mining for Robust Business Intelligence Solutions," PKDD 1999.
- [3] W. Colin, "The IBM Business Intelligence Software Solution," Database Associates, 2000.
- [4] U. Fayyad, G. Piatetsky-Shaprio, P. Smyth, and Uthurusamy, *Advanced in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.
- [5] G. Piatetsky-Shaprio, J. Frawley, *Knowledge discovery in databases*, AAAI/MIT Press. 1991.
- [6] IBM Research Center, "Developing Tightly-Coupled Applications on IBM DB2/CS Relational Database System: Methodology and Experience," IBM Research Report RJ 10005(89094), 1995.
- [7] S. Sarawagi, S. Thomas, and R. Agrawal, "Integrating Mining with Relational Database Systems: Alternatives and Implications," In Proc. of the ACM SIGMOD Int'l Conference on Management of Data, pp.343-353, 1998.
- [8] H. T. Kim, *E-Commerce Marketing Strategy*, SamGakHyung Press, 1999.
- [9] 이용준, 서성보, 류근호, 김혜규, "시간 간격을 고려한 시간 관계 규칙 탐사", 정보과학회논문지, 제28권 제3호, 2001.
- [10] R. Agrawal, et al., "The Quest Data Mining System," In Proc. Of the Int'l Conference on Knowledge Discovery in Databases and Data Mining, pp.244-249, 1996.
- [11] J. Han, et al., "DBMiner: A System for Data Mining in Relational Databases and Data Warehouses," In Proc. of CASCON, 1997.
- [12] IBM DB2 Intelligent Miner for Data, <http://www-4.ibm.com/software/data/iminer/fordata/>
- [13] M. King, et al., "Evaluation of Fourteen Desktop Data Mining Tools," IEEE International Conference on Systems, Man, and Cybernetics, 1998.
- [14] J. Han, et al., "DBMiner: A System for Mining Knowledge in Large Relational Databases," In Proc. of the Int'l Conference on Data Mining and Knowledge Discovery, pp.250-255, 1996.
- [15] R. Agrawal, T. Imielinski, and A. Swami, "Mining

- association rules between sets of items in large databases," In Proc. Of the ACM SIGMOD Conference on Management of Data, pp.207-216, 1993.
- [16] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Data," In Proc. Of Int'l Conference on Very Large Databases, pp.487-499, 1994.
- [17] R. Srikant, and R. Agrawal, "Mining Generalized Association Rules," In Proc. Of Int'l Conference on Very Large Databases, pp.407-419, 1995.
- [18] R. Agrawal, et al., "Mining Sequential Patterns," In Proc. of the Int'l Conference on Data Engineering, pp.3-14, 1995.
- [19] R. Srikant and R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements," In Proc. of the Int'l Conference on Extending Database Technology, pp.3-17, 1996.
- [20] R. Agrawal, K. Lin, H. Sawhney, and K. Shim, "Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases," In Proc. of the Int'l Conference on Very Large Databases, pp.490-501, 1995.
- [21] M. Mehta, R. Agrawal, and J. Rissanen, "SLIQ: A fast scalable classifier for data mining," In proc. Of Int'l Conference Extending Databases Technology, 1996.
- [22] J. Shafer, R. Agrawal, and M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining," In Proc. of the Int'l Conference on Very Large Databases, pp.544-555, 1996.
- [23] R. Agrawal and G. Psaila, "Active data mining," In Proc. Of the Int'l conference on Knowledge Discovery in Databases and Data Mining, pp.3-8, 1995.
- [24] J. Han, Y. Fu, K. Koperski, W. Wang, and O. Zaiane, "DMQL: A Data Mining Query Language for Relational Databases," SIGMOD'96 Workshop. on Research Issues on Data Mining and Knowledge Discovery, 1996.
- [25] K. Keiser, J. Rushing, H. Conover, and S. Graves, "Data Mining System Toolkit for Earth Science Data," Earth Observation (EO) & Geo-Spatial (GEO) Web and Internet Workshop '99, 1999.
- [26] 이정무, Introduction to Data Mining with SQL Server 2000, Microsoft Tech-Ed 2000, 2000.
- [27] S. Thomas and S. Sarawagi, "Mining Generalized Association Rules and Sequential Patterns Using SQL Queries," In Proc. of the Int'l Conference on Knowledge Discovery in Databases and Data Mining, pp.344-348, 1998.
- [28] T. Imielinski, A. Virmani, and A. Abdulghani, "DataMine: Application Programming Interface and Query Language for Database Mining," In Proc. of the International Conference on Knowledge Discovery and Data Mining, pp.256-262, 1996.
- [29] M. Brohman, M. Parent, M. Pearce, and M. Wade, "The Business Intelligence Value Chain: Data Driven Decision Support in a Data Warehouse Environment: An Exploratory," In Proc. of the Int'l Conference on System Sciences, 2000.
- [30] 류근호, 이준욱, 이용준, "eCRM을 위한 시간 데이터마이닝", 한국정보과학회 데이터베이스연구회지, 제17권 제1호, 2001.
- [31] 류근호, "CyberPost BI 기술 개발", 한국 전자통신 연구원 위탁과제 최종 보고서, 2000.



이 준 욱

1997년 충북대 컴퓨터과학과 졸업. 1999년 충북대 대학원 전자계산학과(이학석사). 1998년 ~ 1999년 한국전자통신연구원 위촉연구원 근무. 2002년 현재 충북대학교 대학원 컴퓨터과학과 박사과정. 관심 분야는 시간 데이터베이스, 시공간 데이터 마이닝, CRM, 시간 데이터 마이닝.



백 욱 현

2000년 충북대학교 정치외교학과(학사). 2002년 충북대 대학원 전자계산학과(이학석사). 2002년 현재 국방과학연구소 연구원. 관심분야는 시간데이터베이스, 시간 데이터 마이닝, LBS,

류 근 호

정보과학회논문지: 컴퓨팅의 실제 제 8 권 제 1 호 참조