

---

# 데이터마이닝 시스템의 커스터마이징 방법론

김근형\* · 오경훈\*\* · 김민철\*

## Methodology for customizing data mining system

Keun-Hyung Kim\* · Kyung-Hoon Oh\*\* · Min-Cheol Kim\*

### 요 약

상용화된 데이터마이닝 제품들은 대부분 완성품의 형태로 이들 기능들의 일부 또는 전부를 지원한다. 완성품은 구입 즉시 사용할 수 있다는 장점은 있으나 사용자 및 기업의 요구사항이나 환경에 유연하게 적용할 수 없는 문제가 있다.

본 논문에서는 데이터마이닝시스템을 완성품이 아닌 커스터마이징의 형태로 도입 및 구축하기 위한 방법론을 제안하고, 실제 오라클 제품을 이용하여 커스터마이징을 하는 사례를 살펴본다. 그리고, 구축된 데이터마이닝 시스템을 이용하여 판매 데이터를 분석한 사례를 제시한다.

### Abstract

As a finished product, most commercialized products of Data mining are supposed to support some or all of these functions. These products have an advantage to be available as soon as they are purchased. But, they have some limit to meet users' needs and companies' requirements and to be adaptable to different settings.

This study is focused on how to introduce Data mining system and implement it as a form of customizing, instead of a form of finished product. Also, this study try to present several cases which actually used customizing with Oracle products and some other cases which used constructed data mining to analyse sales data.

### 키워드

데이터마이닝, 커스터마이징, 오라클

## 1. 서 론

데이터마이닝(Data Mining)은 기업경영 활동의 결과로서 수집된 대량의 데이터를 분석하여 기업의 이윤 추구에 도움이 될 수 있는 정보와 지식을 획득할 수 있는 기술이다[장남식, 1999]. 데이터마이닝은 이러한 비즈니스적 요구로 인해 등장한 정보 추출 방법론으로써 통신, 은행, 소매, 의료, 제조, 유통, 항공 등 다양한 산업 분야에서 널리 사용

되고 있다.

데이터마이닝의 기능 또는 기술에는 연관규칙탐사(association rule mining), 항목분류(classification), 클러스터링(clustering), 요약(summarization), 순차패턴탐사(sequential pattern) 등이 있다 [김정자, 1998]. 이러한 기술들은 다양한 응용영역들에서 활용될 수 있고 데이터마이닝의 사용자들에 따라 필요한 응용영역 및 데이터마이닝 기술이 다양하다[Berry&Linoff, 1997]. 즉, 사용자들에 따라

데이터마이닝에 대한 요구사항은 다르게 나타날 수 있다.

상용화된 데이터마이닝 제품들은 대부분 완성품의 형태로 데이터마이닝 기능들의 일부 또는 전부를 지원한다. 완성품은 구입 즉시 사용할 수 있다는 장점은 있으나 사용자의 요구사항이나 기업의 환경에 유연하게 적용할 수 없는 문제가 있다. 즉, 데이터마이닝의 기능들 중 일부만을 제공하는 완성품의 경우 다양한 응용영역의 다양한 사용자들을 만족시킬 수 없게 되고, 데이터마이닝의 모든 기능을 제공하는 완성품의 경우는 다양한 사용자들에 의하여 사용되어질 수 있지만 제품가격 대 효용 측면에서 비효율성이 생기게 된다. 따라서, 데이터마이닝 기능을 제공하는 패키지를 커스터마이징 가능한 형태로 제공하고 사용자들의 요구사항에 적합하게 커스터마이징하여 사용할 수 있으면 비용 대 효용 측면에서 효율화를 꾀할 수 있을 것이다.

본 논문에서는 데이터마이닝시스템을 완성품이 아닌 커스터마이징의 형태로 도입 및 구축하기 위한 방법론을 고찰해보고, 실제 오라클 제품을 이용하여 커스터마이징을 적용하는 사례를 살펴본다. 그리고, 구축된 데이터마이닝시스템을 이용하여 판매 데이터를 분석한 사례를 제시한다.

## II. 데이터마이닝

데이터마이닝이란 자동화되고 지능을 갖춘 데이터베이스 분석기법으로 90년대 초반부터 지식발견, 정보발견, 정보수확 등의 이름으로도 소개되어 왔는데 일반적으로 "대량의 데이터로부터 새롭고 의미있는 정보를 추출하여 의사결정에 활용하는 작업"이라 정의한다[장남식, 1999].

데이터마이닝은 '데이터베이스 내에서의 지식발견(Knowledge Discovery, KDD)'과 유사어로 언급되기도 하지만, 지식발견은 데이터로부터 유용한 정보를 발견하는 프로세스의 전 과정이며, 데이터마이닝은 지식발견 중에서도 데이터로부터 정보를 추출하기 위해서 기법을 적용하는 특정단계라고 정의할 수 있다[장남식, 1999].

KDD의 전체적인 프로세스는 연구자들마다 다소 표현에 차이는 있으나 이를 종합하여 보면, 1) 비즈니스 환경과 문제를 바탕으로 분석 목적 설정,

2)가용한 자료확인, 3)자료의 탐색적 분석, 4) 데이터마이닝을 통한 지식추출, 5) 결과에 대한 평가와 해석, 6)전략결정 및 실행, 7)실행결과 모니터링의 7단계로 나누어 볼 수 있다(Adrianns & Zantinge, 1996 ; Pyo, Uyal. & Chang, 2002).

이러한 KDD 과정에서 데이터마이닝을 통한 지식 추출과정이 차지하는 역할이 가장 결정적인 단계이기 때문에 결국 어떠한 비즈니스의 목적을 달성하기 위해 어떠한 데이터마이닝 기법을 적절히 적용하느냐가 중요한 문제가 되며, 이러한 이유로 KDD와 데이터마이닝을 유사어처럼 이용하는 것이다.

데이터마이닝 기법에는 연관규칙탐사(association rule mining), 연속규칙탐사(sequence rule mining), 분류(classification), 군집화(clustering) 등이 있다. 이들 데이터마이닝 기법은 특정 업종에만 국한된 것이 아니라 비즈니스의 환경과 목표, 사용 가능한 데이터의 속성에 따라 적용될 수 있는 적합한 기법들이 달라지게 된다[Berry & Linoff, 1997].

## III. 커스터마이징 방법론

이 절에서는 커스터마이징을 통하여 보다 효율적이고 효과적인 데이터마이닝시스템의 구현을 위한 방법론을 제안한다.

### 3.1 커스터마이징 관련 연구 고찰 및 분석

조직 내에서 정보시스템을 자체개발하는 경우 많은 시간과 비용을 소모하게 된다. 그런데 이렇게 개발된 시스템이 꼭 사용자의 모든 요구사항을 만족시키는 것도 아니다. 자체개발의 어려움에 대한 해결방안으로 패키지 소프트웨어의 사용이 제안되기도 하는데, 패키지 소프트웨어를 제대로 구축하여 성공으로 이끄는 데에는 여러 요인이 영향을 미치게 된다.

Lucas 등의 논문[1988]에서는 패키지 소프트웨어를 구축하는 과정을 연구하였는데, 우선 패키지 소프트웨어를 2가지 유형으로 구분하였다. 하나는 사용자가 문제를 해결하는데 직접적으로 활용할 수 있는 Lotus 1-2-3와 같은 일반목적의 패키지가고 다른 하나는 사용자의 요구사항에 맞게 커스터마이징(customizing)하여 활용할 수 있는 Dedicated 패키지 이다. 이 논문에서 커스터마이징

의 의미는 이미 개발된 패키지를 사용자의 요구에 맞게 시스템의 기능을 조정하고 수정하여 사용자의 욕구를 충족시켜주는 일련의 활동이라고 정의한다. 커스터마이징의 유형은 변경, 추가, 확장의 3가지 형태로 구분하여 살펴볼 수 있다. 첫째, 변경(Modify)이라 함은 패키지에서 제공하는 기능이 실제 업무처리 과정에서 필요한 기능과 상이하여 패키지의 프로그램 자체를 변경하는 활동을 말한다. 커스터마이징의 유형 중 가장 어렵다고 알려져 있다. 둘째, 추가(add-on)라 함은 패키지에서 제공되지 않는 별도의 기능을 개발하여 기존 기능들과 연계하여 사용할 수 있도록 하는 것을 말한다. 추가의 경우는 기존의 전통적인 개발방법과 거의 유사한 형태이다. 셋째, 확장(extension)의 경우는 변경과 추가의 중간 형태로서 기존의 패키지에 사용자가 원하는 기본적인 기능이나 모듈은 있으나 더 확장된 추가 기능이 필요하며 기본 기능이나 모듈에 부가적으로 프로세스나 기능을 첨가하여 개발하는 것을 말한다.

일반목적의 패키지는 완성된 형태로 제공되기 때문에 즉각적인 활용이 가능하지만 그 기능들이 사용자의 요구사항과 완전하게 일치하지 않을 수 있으므로 사용자의 불편과 불만이 생길 수 있고, 또한 사용자가 패키지의 기능들을 익히고 적용해야 한다. Dedicated 패키지는 정제되지 않은 기본적인 기능들을 제공하지만 사용자의 필요에 따라 커스터마이징해서 사용할 수 있으므로 일반목적 패키지에 비하여 보다 융통성이 있다고 할 수 있다. 그러나, 커스터마이징 작업에 대한 비용이 필요하고 어떻게 커스터마이징했느냐에 따라서 패키지가 제공하는 기능에 대한 품질이 결정되므로 효율적이고 효과적인 커스터마이징 방법론이 적용되어야 한다.

Dedicated 패키지의 대표적인 예로써 ERP 패키지가 해당된다고 할 수 있다. ERP 패키지의 커스터마이징 방법론에 대해서는 몇몇 문헌에서 제시되고 있다. 김병곤[2000] 등의 연구에서는 ERP 패키지의 성공적인 커스터마이징을 위한 고려 요인으로 "ERP에 대한 정확한 이해, 목표의 명확한 설정, 철저한 도입준비, 최고 경영자의 강력한 의지, 커스터마이징의 최소화"를 들고 있다. 커스터마이징 작업 중 변경을 위해서는 원래의 패키지가 가지고 있는 기능이나 프로그램의 소스코드(source code)를 잘 알아야 하는데, 통상 일반 사용자나 개발자들에게는 패키지의 내용이 블랙박스화 되어 있기 때문

에 시스템 전체를 파악하는 것이 그리 쉬운 일이 아니다. Soh[2000] 등의 연구에서는 ERP시스템을 커스터마이징 해야 하는 경우, 추가는 하더라도 변경은 자제할 것을 제안한다. 일반적으로 패키지의 소스코드를 변경하는 것은 버전 상향 시 유지보수의 어려움 때문에 피하는 것이 좋다. 조직의 중요한 기능을 위해 커스터마이징이 필요한 경우에도 소스코드의 변경보다는 추가모듈의 개발을 권고한다. 정승민[2002] 등의 연구에서는 커스터마이징을 많이 할 수록 사용자 만족도 및 조직의 경쟁우위를 높일 수 있으며, 소스코드를 변경하는 것이 아니라 추가의 형태로 필요한 기능을 지원하여야 한다.

이상 문헌고찰의 결과로 dedicated 패키지의 커스터마이징은 사용자의 만족을 가져다 줄 수 있지만 패키지의 기능을 추가하는 방향으로 적용될 필요가 있음을 알 수 있다. 따라서, dedicated 패키지가 지원하는 기본적인 기능을 API(Application Programming Interface)의 형태로 제공하고 API를 이용하여 필요한 기능들을 추가할 수 있는 형태의 dedicated 패키지가 바람직할 수 있겠다. API는 프로그래밍언어의 함수 및 라이브러리와 비슷한 개념이라고 할 수 있다.

### 3.2 데이터마이닝의 커스터마이징 필요성

데이터마이닝 기능의 종류와 데이터마이닝 사용자의 유형은 다양하다. 즉, 데이터마이닝을 지원하는 제품은 다양한 기능을 제공해야 하고 다양한 사용자들에 의하여 편리하게 사용될 수 있어야 한다. 그러나, 실제 기업에서는 기업에서 필요로 하는 고급정보를 만들기 위하여 데이터마이닝의 모든 기능을 필요로 하지는 않을 것이다. 예를 들어, 판매매장에서는 판매되는 제품간의 연관성 정보는 필요할 수 있지만 제품들을 군집화하는 기능은 필요로 하지 않을 것이다. 또한, 데이터마이닝 제품은 데이터마이닝 전문가가 이용할 수 있어야 할 뿐만 아니라 판매점포의 점원도 편리하게 사용할 수 있어야 좋은 정보추출 도구가 될 수 있다. 결과적으로, 데이터마이닝 제품에 대한 요구사항은 사용자 또는 기업들에 따라 다양하다고 할 수 있으며, 이러한 다양한 요구사항을 하나의 일반목적 패키지의 형태로 제공하는 것은 비효율적일 수 있다. 패키지 용량이 커지면서 가격이 비싸지게 되고 모든 기능을 사용할 필요가 없는 사용자는 지불한 비용만큼 효용 회수가 되지 않을 것이기 때문이다.

따라서, 데이터마이닝 기능을 제공하는 패키지

의 형태도 dedicated 패키지가 바람직할 수 있다. 데이터마이닝 기능을 제공하는 dedicated 패키지는 데이터마이닝의 모든 기능을 API의 형태로만 제공하고, 사용자들은 필요한 기능들만을 API를 이용하여 추가 개발하는 방식으로 이용할 수 있다. 이러한 방식은 사용자들이 불필요한 기능들을 이용하지 않게 할 뿐만 아니라 사용자들의 수준에 적합하게 인터페이스(interface)를 구현할 수 있기 때문에 사용자들의 만족도를 높여줄 수 있다.

### 3.3 데이터마이닝시스템의 커스터마이징 방법론 제안

데이터마이닝시스템을 구현하기 위하여 데이터마이닝 기능이 내장된 dedicated 패키지를 효율적이고 효과적으로 커스터마이징하기 위한 방법론을 제안한다.

선행연구에서 살펴보았듯이, 커스터마이징은 추가 형태로 적용될 때 그 성공 가능성이 높다고 할 수 있다. 추가유형의 커스터마이징은 전통적인 개발방법인 SDLC(System Development Life Cycle)와 유사하면서도 커스터마이징이라는 특성이 반영되어야 하므로 단계별 실행지침들에서 차이점들이 존재한다. 추가유형의 커스터마이징을 성공적으로 수행하기 위한 전략을 다음과 같이 제안한다.

**전략 :** 구축단계별로 자체개발과 커스터마이징의 차이점을 비교분석한다.

일반적으로 기존의 개발방법은 업무분석, 설계, 구축, 테스트, 이행, 유지보수의 개발 사이클을 가진다. 그러나, dedicated 패키지의 구축에 있어서는 기존의 방법과는 다르게 보다 강화되어야 할 단계가 프로젝트의 준비단계이다. 프로젝트의 시작시점에 준비단계를 신설하여 이에 적합한 활동을 우선 실시하여야 한다. 일반적인 개발방법에 있어서는 앞에서 열거한 단계들을 진행하면서 하나의 상품화된 시스템이 탄생하게 된다. 그러나, dedicated 패키지는 상품화된 시스템이 이미 존재하고 있으며, 중요한 점은 자사의 요구사항을 정확하게 지원해줄 수 있는 기능이 제공되게 커스터마이징하기 위하여 dedicated 패키지의 전반적인 이해를 위한 교육 및 연구 과정이 필요하다는 것이다. 또한, SDLC의 각 단계별로 이행되어야 할 세부 시행지침들도 기존의 개발방법과 다른 관점에서 파악되어야 한다. <표1>은 기존의 전통적인 개발방

법과 커스터마이징 방법의 구축단계 별 차이점을 비교 분석하고 있다.

<표 1> 기존 개발 방법론과 커스터마이징의 단계별 비교분석표

구분	기존개발방법	커스터마이징
프로젝트 준비단계	<ul style="list-style-type: none"> <li>별도로 정의된 사항이 없음</li> </ul>	<ul style="list-style-type: none"> <li>dedicated 패키지의 기능 이해를 위한 전반적인 교육 또는 연구</li> </ul>
업무분석 단계	<ul style="list-style-type: none"> <li>현업 사용자와 면담 후 업무분석</li> </ul>	<ul style="list-style-type: none"> <li>업무분석을 통한 데이터마이닝의 응용영역 이해</li> <li>사용자의 요구사항 분석</li> </ul>
설계 단계	<ul style="list-style-type: none"> <li>분석 산출물을 중심으로 업무설계와 기술설계</li> </ul>	<ul style="list-style-type: none"> <li>사용자의 요구사항을 해결하기 위한 기능설계</li> <li>dedicated 패키지에서 제공하는 기능을 이용하여 설계</li> </ul>
구현 단계	<ul style="list-style-type: none"> <li>개발툴과 DB관리툴을 이용하여 설계내용에 따라 엔진기능을 개발</li> </ul>	<ul style="list-style-type: none"> <li>dedicated 패키지와 호환될 수 있는 개발툴을 이용하여 설계내용을 구현</li> </ul>
테스트 단계	<ul style="list-style-type: none"> <li>자동화된 테스트툴을 이용하거나 테스트 계획서를 작성하여 실시</li> </ul>	<ul style="list-style-type: none"> <li>테스트 시나리오를 작성하여 실시</li> </ul>
운영 및 유지보수 단계	<ul style="list-style-type: none"> <li>사용자의 요구사항에 맞도록 관리</li> </ul>	<ul style="list-style-type: none"> <li>추가개발 요청 시 dedicated 패키지의 기능을 고려하여 결정</li> </ul>

각 단계의 구체적인 차이점들과 이행지침들은 다음과 같다.

#### 가. 프로젝트준비단계

**기존개발방법 :** 기존 개발 방법에서는 정의된 사항이 없다.

**커스터마이징 :** 프로젝트 참여자 전원에게 dedicated 패키지에 대한 교육을 실시한다. dedicated 패키지의 기능과 사용법, 호환 가능한 개발툴 등에 대하여 교육을 실시한다. 교육 여건이 어려운 중소기업들의 경우는 프로젝트 참여자 스스로 매뉴얼등을 이용하여 연구함으로써 관련 지식들을 습득하도록 한다.

나. 분석단계

기존개발방법 : 현업업무 위주로 실사용자와 인터뷰 후 업무내용을 상세하게 기록한다.

커스터마이징 : 데이터마이닝의 응용 영역에 대한 이해를 정확히 함으로서 사용자의 요구사항을 최종 시스템에 정확히 반영할 수 있으므로, 데이터마이닝이 활용될 업무분석을 수행한다. 업무에 대한 지식을 바탕으로 현업의 실 사용자와 인터뷰를 한 후, 사용자의 요구사항을 파악하고 필요한 데이터마이닝의 유형을 결정한다. 사용자가 요구하는 데이터마이닝 기능이 dedicated 패키지에서 제공되고 있는 것인지 아니면 새롭게 개발되어야 할 부분인지를 파악한다. 또한, 데이터마이닝 사용자의 유형을 결정한다. 즉, 데이터분석 전문가인지 또는 일반 사용자인지를 파악한다.

다. 설계단계

기존개발방법 : 분석시의 산출물을 근거로 업무 설계와 기술설계를 수행해 나간다.

커스터마이징 : 분석단계에서 요구되어진 데이터마이닝 기능들을 수행할 수 있는 모듈들의 논리(logic)를 dedicated 패키지의 API 또는 라이브러리 등을 이용하여 설계한다. 필요한 경우 ERD(Entity Relationship Diagram)와 DFD(Data Flow Diagram) 등을 이용한다. 또한, 사용자의 요구사항과 사용자의 유형이 고려된 입출력 화면을 설계한다.

라. 구축단계

기존개발방법 : 선정된 개발틀과 데이터베이스 관리 틀을 이용하여 설계된 내용대로 구축작업을 진행한다. 본격적인 기능의 개발에 앞서서 각종 라이브러리나 API 등 엔진기능을 먼저 개발하여야 한다.

커스터마이징 : 설계된 내용대로 구축작업을 진행한다. 개발틀은 dedicated 패키지와 호환될 수 있는 것을 사용한다. 엔진 기능은 대부분 dedicated 패키지에서 API나 라이브러리 형태로 제공되므로 그대로 사용하면 된다.

마. 테스트단계

기존개발방법 : 자동화된 테스트 틀이 있을 경우 이를 이용하고, 테스트틀이 없을 경우 테스트 계획서를 작성하고 이에 따른 테스트 시나리오와 테스트 데이터를 개발하여 이를 근거로 테스트를 실시하고 결과를 기록한다.

커스터마이징 : 자동화된 테스트틀을 이용하는

경우 dedicated 패키지에서 제공하는 API에 대한 트레이스(trace)는 불가능하므로 테스트 시나리오를 만들 때 이러한 사항을 고려한다.

바. 운영 및 유지보수단계

기존개발방법 : 현업의 요구대로 개발된 내용이 원래의 의도나 시나리오대로 가동이 되는가를 면밀히 모니터링하고 실 사용자의 요청에 의거, 지속적으로 시스템을 개선해 나간다.

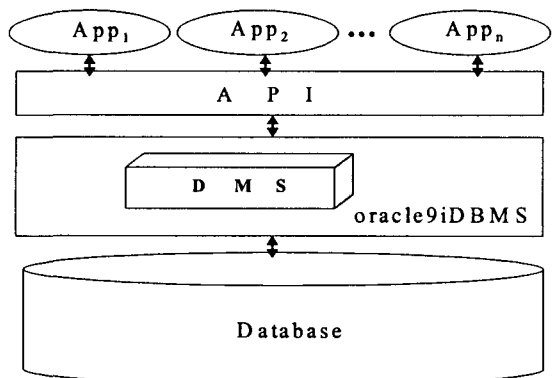
커스터마이징 : 사용자의 추가개발 요구 시 우선적으로 dedicated 패키지 내에서 해결책을 찾아야 한다. 그러기 위해서는 패키지의 기능에 대해서 전 유지보수 요원이 숙달되어 있어야 한다.

IV. 커스터마이징 사례

본 절에서는 dedicated 패키지의 일종인 오라클 데이터마이닝을 이용하여 앞에서 살펴 본 커스터마이징 방법론에 따라 데이터마이닝시스템을 구축하는 사례를 제시한다.

4.1 오라클 데이터마이닝 시스템 구성도

오라클 데이터마이닝은 오라클 9i DBMS와 연동되어서 데이터마이닝 기능을 제공할 수 있는 dedicated 패키지이다. 오라클 데이터마이닝에는 연관규칙탐사 기능과 분류기능을 지원하기 위한 API들이 포함된다. 오라클 9i DBMS와 오라클 데이터마이닝이 결합되어져서 구성된 오라클 데이터마이닝시스템의 구성도는 <그림1>과 같다.



<그림1> 오라클 데이터마이닝시스템의 구성도

<그림1>에서 API(Application Programming Interface)부분은 오라클 데이터마이닝을 설치하면

이용할 수 있는 라이브러리 함수이다. API의 각 기능들은 DMS(Data Mining Server)의 서비스를 통하여 수행된다. DMS는 oracle9i DBMS내에 포함되어 있고 데이터마이닝 기능을 수행하는 엔진이라고 할 수 있다. Appi는 데이터마이닝 기능을 포함하는 응용프로그램들인데 사용자의 요구사항에 따라 API를 이용하여 개별적으로 만들어질 수 있다. 데이터마이닝 프로젝트의 결과물은 사용자가 최종적으로 이용할 수 있는 Appi들이라 할 수 있다.

**4.2 오라클 데이터마이닝시스템을 구축하기 위한 커스터마이징 단계**

커스터마이징 방법론을 이용하여 오라클 데이터마이닝시스템을 구축할 때, 각 단계에서 실행되어야 할 사항들을 사례를 통하여 구체적으로 살펴본다. 본 사례는 앞에서 제시한 커스터마이징 방법론의 타당성을 검토를 위한 수단으로서 프로토타입(prototype)형태로 구축되었다.

**가. 프로젝트준비단계**

오라클 데이터마이닝의 매뉴얼을 분석하여 오라클 데이터마이닝에서 제공하는 기능, 특징들을 고찰한다. 오라클 데이터마이닝에서 제공하는 API들의 종류들과 개발도구에 대하여 학습한다. API는 자바 클래스 및 메소드의 형태로 제공되고 오라클 데이터마이닝과 연동될 수 있는 개발도구는 자바언어이므로 자바언어에 대하여 학습한다.

**나. 분석단계**

사용자의 요구사항은 판매매장에서 POS시스템에 의하여 생성된 판매데이터를 분석하려는 것이고, 사용자의 유형은 데이터마이닝에 대한 전문지식이 있는 데이터 분석 전문가로 가정하였다. 입력력화면은 편의상 커맨드방식으로 구현할 것이다.

**다. 설계단계**

판매 데이터들 사이의 연관성을 분석할 수 있는 모듈을 API의 기능을 고려하여 설계한다. 모듈의 로직은 다음과 같다.

/\* 연관규칙 탐사 프로그램의 절차 \*/

- 파일입력을 통하여 환경변수 값을 설정한다.
- DMS와 연결한다.
- 분석용 데이터화일의 형태를 변환한다.
- 연관규칙을 탐사한다.
- 탐사된 규칙들을 파일을 통하여 출력한다.

**라. 구현단계**

설계단계의 모듈을 자바언어를 이용하여 코딩한다. edit plus나 기타 자바 개발 툴을 이용한다.

**마. 테스트 단계**

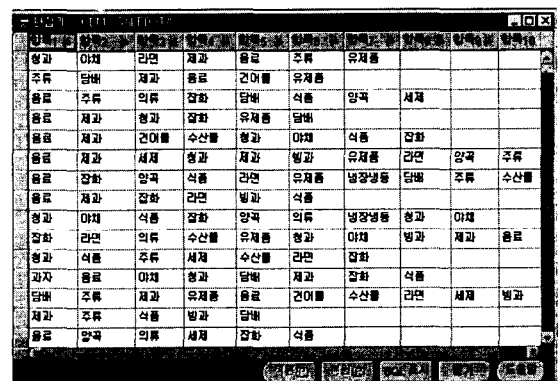
API의 입출력 데이터가 포함된 테스트시나리오를 작성하여 시스템의 오작동 여부, 사용자 요구사항의 만족여부를 테스트한다.

**바. 운영 및 유지보수 단계**

판매데이터를 이용하여 데이터마이닝 프로세스에 따라 연관규칙탐사를 한다. 5절에서 실제의 판매데이터를 분석한 사례를 제시한다.

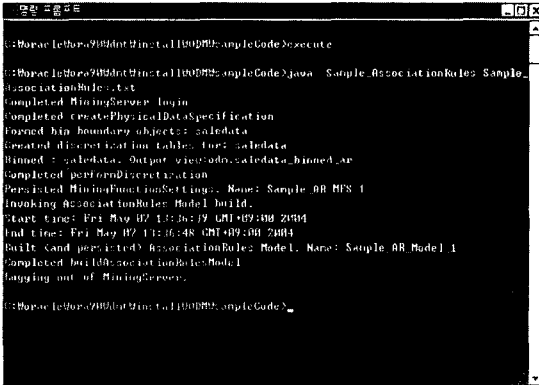
**4.3 오라클 데이터마이닝을 이용한 데이터분석 사례**

본 절에서는 구축된 오라클 데이터마이닝시스템을 이용하여 대형 할인매점의 판매데이터들 사이의 연관성을 분석한다. 판매데이터의 트랜잭션(transaction) 수는 500건으로 비교적 충분하지 않지만 시험용 분석이라는 측면에서 의미가 있다고 하겠다. <표2>는 494건의 판매 트랜잭션 데이터를 담고 있는 관계형 파일이다. 판매 물품들의 종류 즉, 항목들의 종류는 17가지이고 한 트랜잭션의 최대 항목 수는 10개로 설정하였다.



행과	아채	리면	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채
주류	담배	제과	음료	건어물	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료
음료	주류	의류	향과	담배	식물	양육	계란	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품
음료	제과	향과	유제품	담배	계란	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류
음료	제과	건어물	수산물	향과	아채	식물	양육	계란	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류
음료	제과	세제	향과	제과	빙과	유제품	리면	양육	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채
음료	제과	향과	리면	빙과	식물	양육	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류
향과	아채	식물	향과	양육	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품
향과	리면	의류	수산물	유제품	향과	아채	빙과	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품
향과	식물	주류	세제	수산물	리면	양육	제과	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류
과자	음료	아채	담배	제과	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류
담배	주류	제과	유제품	음료	건어물	수산물	리면	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품
제과	주류	식물	담배	제과	향과	아채	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품	계란	의류
음료	양육	의류	세제	향과	식물	양육	계란	제과	음료	주류	유제품	계란	의류	냉장생선	향과	아채	음료	주류	유제품

<그림 2>는 오라클 데이터마이닝을 이용하여 판매데이터 분석을 실행하는 화면이다. 실행시간은 9 초 걸리는 것으로 나타나고 있다.



<그림 2> 오라클 데이터마이닝의 실행화면

<그림3>은 연관규칙탐사를 위한 환경변수 값들에 따라 도출된 규칙들을 정리한 것이다.

<표2> 도출된 규칙들

환경변수	규칙	지지도	신뢰도
최소지지도 : 0.01 최소신뢰도 : 0.5 규칙의 길이 : 2	rule : 잡화 → 식품	0.06	0.52
	rule : 병과 → 제과	0.03	0.68
	rule : 담배 → 제과	0.02	0.53
	rule : 수산물 → 야채	0.01	0.58
	rule : 야채 → 청과	0.01	0.58
최소지지도 : 0.01 최소신뢰도 : 0.5 규칙의 길이 : 3	rule : 수산물 → 식품	0.01	0.50
	rule : 식품, 제과 → 잡화	0.02	0.60
	rule : 병과, 음료 → 제과	0.02	1.00
	rule : 잡화, 음료 → 제과	0.02	0.53
최소지지도 : 0.01 최소신뢰도 : 0.5 규칙의 길이 : 4	rule : 잡화, 음료 → 식품	0.02	0.53
	rule : 음료, 제과, 식품 → 잡화	0.01	0.86

최소지지도는 1% 로하고 최소신뢰도는 50% 로 설정하였다. 도출된 규칙들은 우리가 일반적으로 기대했던 결과를 반영하고 있다고 생각된다. 예를

들면, [rule : 잡화 → 식품]의 의미를 해석해 보면, 잡화와 식품을 포함하는 트랜잭션 건수는 전체 트랜잭션 494건 중 30건(즉,  $0728744 = 30/494$ )이고, 잡화를 포함하는 트랜잭션이 식품도 포함할 확률은 0.51724136, 즉 약 50% 라는 말이다. 달리 해석해 보면, 잡화와 식품은 많이 팔리는 제품이고 서로 연관성도 높다는 의미이다. 특히, 주목할 규칙은 [rule : 병과, 음료 → 제과]로써 병과와 음료를 구매하면 제과도 함께 구매될 확률이 100%로 나타나고 있다. 전반적으로 볼 때, 기대하지 않았던 특이한 규칙은 도출되지 않았음을 알 수 있다. 이는 데이터마이닝이 의미있는 의외의 결과를 항상 제공하지는 않음을 알 수 있다.

## V. 결 론

데이터마이닝은 축적된 대규모의 데이터들을 분석하여 기업의 경영활동에 도움이 될 수 있는 정보와 지식을 획득할 수 있는 기술이다. 데이터마이닝 기술에 의하여 다양한 정보분석 및 지식추출 기능을 활용할 수 있는데, 사용자들은 이러한 다양한 기능들을 전부 필요로 하지 않고 업무특성에 따라 제공 기능의 일부분을 이용한다. 또한, 데이터마이닝 기술은 전문가뿐만 아니라 비전문가도 편리하게 활용할 수 있어야 보다 폭넓은 공헌을 할 수 있다. 결과적으로, 데이터마이닝은 사용자에 따라서 다양한 요구사항이 존재할 수 있으므로 커스터마이징에 의한 데이터마이닝시스템 구축의 필요성이 생기게 된다.

본 논문에서는 데이터마이닝시스템을 커스터마이징에 의하여 효율적이고 효과적으로 구축하기 위한 방법론을 제시하였다. 기존의 개발방법론과의 비교분석 전략을 통하여 체계적인 단계별 실행 지침을 제시하였다. 커스터마이징 방법론은 기존의 개발방법에 비하여 프로젝트 준비단계의 비중을 중요시할 필요가 있고 미리 제공되는 라이브러리 기능을 정확히 파악하여 설계, 구현단계에서 효과적으로 이용하여야 한다. 본 논문에서 제안한 커스터마이징 방법론에 대한 타당성 검토를 위하여 커스터마이징 방법론의 각 단계별 실행지침에 따라서 오라클 데이터마이닝시스템을 구축해보았다. 또한, 대형 할인매장의 판매분석 데이터를 확보하여 구축된 데이터마이닝시스템을 이용하여 연관규칙 탐사를 실행하였다.

커스터마이징 방법론을 도출하는데 있어서 객관성을 높이기 위하여 커스터마이징 전문가들의 의

견을 수렴하여 반영할 필요성이 있지만, 이는 본 연구를 시발점으로 추가적으로 연구되어야 할 과제가 될 수 있을 것이다.

ledge discovery in database for tourist destinations", Journal of Travel Research, 40, May.

참고문헌

저자소개

- [1] 김병곤, 오재인(2000), "ERP 패키지의 성공적인 커스터마이징 전략", 경영정보학연구, 10권 3호.
- [2] 김정자, 이도현(1998), "데이터마이닝 기술 및 연구동향", 정보과학회지, 제16권 9호.
- [3] 장남식(1999), 데이터마이닝, 대청미디어, 1999.
- [4] 정승민, 김준석(2002), "ERP 시스템 도입 시 커스터마이징 정도가 사용자 만족도와 조직의 경쟁우위에 미치는 영향", 한국경영정보학회 2002 추계 학술대회, pp.529-540.
- [5] Soh, C., Kien, S.S., and Joanne Tay-Yap.(2000), "Cultural Fits and Misfits: Is ERP a Universal Solution?", Communications of ACM, pp.47-51.
- [6] Adrianns, P.&Zantinge, d.(1996), Data mining, New Yor NY: Addison-Wesley.
- [7] Berry.M.J.A. & Linoff.G.(1997), Data mining techniques: For marketing sales and customer support, New York, NY : John Wiley & Sons.
- [8] Lucas, H.C., Walton, E.J., and Ginzberg, M.J.(1988), "Implementing Packaged Software", MIS Quarterly, pp.537-549.
- [9] Pyo,S.P.,Uysal. M.&Chang,H.S(2002), "Know-

김근형(Keun-Hyung Kim)

1990년 서강대학교 컴퓨터학과(학사)  
1992년 서강대학교 대학원 컴퓨터학과(석사)  
2001년 서강대학교 대학원 컴퓨터학과(박사)  
2001년~ 현재 : 제주대학교 경영정보학과 조교수  
※ 관심분야 : 데이터베이스, 데이터마이닝

오경훈(Kyung-Hoon Oh)

2002년~현재 제주대학교 경영대학원 경영정보학과  
※ 관심분야 : 데이터마이닝, MIS

김민철(Min-Cheol Kim)

1991년 중앙대학교 경영학과(학사)  
1995년 고려대학교 대학원 경영학과(석사)  
2000년 고려대학교 대학원 경영학과(박사)  
2001년~ 현재 : 제주대학교 경영정보학과 조교수  
※ 관심분야 : MIS, 의료정보, 지식경영