

# 웹 데이터베이스 응용을 위한 액티브데이터마이닝 컴포넌트 개발

최 용 구\*

## Development of Active Data Mining Component for Web Database Applications

Yong-Goo Choi\*

### Abstract

The distinguished prosperity of information technologies from great progress of e-business during the last decade has unavoidably made software development for active data mining to discovery hidden predictive information regarding business trends and behavior from vary large databases. Therefore this paper develops an active mining object(ADMO) component, which provides real-time predictive information from web databases. The ADMO component is to extended ADO(ActiveX Data Object) component to active data mining component based on COM(Component Object Model) for application program interface(API). ADMO component development made use of window script component(WSC) based on XML(eXtensible Markup Language). For the purpose of investigating the application environments and the practical schemes of the ADMO component, experiments for diverse practical applications were performed in this paper. As a result, ADMO component confirmed that it could effectively extract the analytic information of classification and aggregation from vary large databases for Web services.

Keywords : ADMO, Active Data Mining, Component Object Model, XML Web Services

논문접수일 : 2007년 07월 23일

논문게재확정일 : 2008년 05월 31일

※ 이 논문은 동서울대학 산업기술 연구비지원 "웹 데이터베이스에서 데이터마이닝을 위한 컴포넌트 객체 모델"의 연구결과를 새롭게 수정 보완한 것임.

\* 동서울대학 컴퓨터정보과 부교수, (461-714) 경기도 성남시 수정구 복정동 432번지 동서울대학 컴퓨터정보과, Tel : 031-720-2155, e-mail : ygchoi@dsc.ac.kr

## 1. 서론

최근 인터넷 정보기술의 눈부신 발달로 인하여 데이터베이스에 저장되는 다양한 웹 데이터가 현저하게 증가하게 되었다. 맞춤형 웹서비스(adaptive web services)[Balke et al., 2003]를 위하여 이러한 웹 데이터베이스로부터 숨겨진 실시간 예측 정보를 발견하기 위한 데이터마이닝(data mining)[Brachman et al., 1996; Chen et al., 1996; Fayyad et al., 1996; Mehmed et al., 2002]기법의 연구가 활발히 진행 중에 있다. 웹 데이터베이스에서 맞춤형 웹서비스를 위한 데이터마이닝은 전자상거래, 금융, 유통, 보험, 의료 등 여러 분야에서 고객관리나 판매관리, 위험분석 등 다양한 업무에 그 필요성이 점점 더 증가하고 있는 추세이다. 특히 전자상거래 쇼핑몰에서 고객 세분화를 통한 목표 고객에게 차별화된 정보서비스 제공을 위한 고객 맞춤형 웹서비스(customer matching web services)[Berry et al., 2004]에 점점 더 관심이 높아지고 있다. 쇼핑몰에서 고객의 만족도를 높일 수 있는 고객 맞춤형 웹서비스를 위해서는 데이터베이스로부터 언제 어디서나 필요한 실시간 예측 정보를 추출하기 위한 액티브데이터마이닝이[Agrawal et al., 1995; Bailey et al., 2003] 필요하다. 하지만 현재 비교적 많이 사용하고 있는 통계분석 및 데이터마이닝 도구인 SAS의 *Enterprise Miner*, SPSS *Clementine*, Oracle *Data Mining*, IBM *Intelligent Miner*, 그리고 Silicon Graphics *MineSet*[Mehmed, 2002; Elder et al., 1998] 등은 정적인 데이터베이스로부터 데이터분류와 클러스터링, 회귀분석, 탐지모델링 등의 통계분석 정보를 제공하기 위한 목적으로 개발되었기 때문에 실시간 예측 정보를 추출하기 위한 액티브데이터마이닝에 적용하기 위해서는 여러 가지 한계가 있다.

액티브데이터마이닝을 위해서는 네트워크상에서 클라이언트 서버사이에 상호협력을 통하여 언제 어디서나 운영 데이터베이스로부터 실시간 정보를 용이하게 추출하기 위한 웹서비스가 필요하다. 웹서비스 환경에서 클라이언트 서버기술이 인터넷 기술과 객체지향 패러다임 정보기술 분야에 적용되면서 서버 객체 컴퓨팅 기법의 관심이 고조되고 있다. 이러한 서버 객체 컴퓨팅 기법으로 소프트웨어의 재사용을 위한 컴포넌트 기반의 소프트웨어 개발 기술이 정립되어 가고 있는 추세이다. 근래에 컴포넌트 객체 기술의 프레임워크를 제공하는 업체들은 분산 웹 서비스를 위하여 분산 컴퓨팅 웹 기술과 서버 객체 기술을 접목을 통한 컴포넌트의 표준화 방안을 마련하고 있다. 표준화 방안으로 개발된 대표적인 컴포넌트는 OMG(Object Management Group)의 CORBA(Common Object Request Broker Architecture), 마이크로소프트웨어사의 COM(Component Object Model), SUN사의 EJB(Enterprise Java Bean) 및 J2EE(Java 2 Enterprise Edition)[Heineman 2001; Snyperski, 1998]을 들 수 있는데 이들은 서로 보완적 관계를 유지하면서 발전하고 있다. 이러한 객체기반 컴포넌트는 여러 응용 시스템에 공동으로 사용할 수 있기 때문에 소프트웨어 생산성을 획기적으로 향상시킬 수 있다는 장점을 가지고 있다. 따라서 데이터마이닝도 이러한 환경을 현실적으로 가능하게 하는 다양한 액티브데이터마이닝 소프트웨어개발이 필요하게 되었다.

따라서 본 논문에서는 웹 데이터베이스 응용을 위한 액티브데이터마이닝 객체(active data mining object : ADMO) 컴포넌트를 개발한다. ADMO는 운영 데이터베이스로부터 실시간 데이터마이닝을 통하여 언제 어디서나 숨겨진 실시간 정보를 발견하기 위한 응용프로그램인터페이스(application program interface : API)이

다. ADMO 컴포넌트는 실시간 데이터마이닝이 가능하도록 COM기반의 데이터관리 컴포넌트인 ADO(ActiveX Data Objects) 확장한 것이다. ADMO 컴포넌트 구현은 XML(eXtensible Markup Language)[Choi et al., 2005; Tang et al., 2005]기반의 윈도우스크립트 컴포넌트(window script component: WSC)[Powers, 1999]를 이용한다. 이를 위해 제 2장에서는 이론적 배경으로 데이터마이닝 도구를 데이터 접근 방법에 따라 비교분석하고 컴포넌트 객체 모델을 분석한다. 제 3장에서는 ADMO 컴포넌트를 설계하고, 제 4장에서는 ADMO 컴포넌트를 구현한다. 제 5장에서는 구현된 ADMO 컴포넌트의 응용사례를 제시한 후, 제 6장에서는 본 연구결과의 의의와 향후 연구방향에 대하여 기술한다.

## 2. 이론적 배경

### 2.1 데이터마이닝 도구 비교분석

기존 데이터마이닝 도구들은 대량의 데이터로부터 새롭고 의미 있는 정보를 추출하기 위한 목적으로 개발되었지만 그들이 지원하는 응용 환경은 서로 다르다. 즉 데이터마이닝 도구들은 자신들만이 지원하는 플랫폼과 알고리즘, 데이터접근 방법, 그리고 시각화 능력 등에 따라 분류[Elder et al., 1998; Huh et al., 2000]할 수 있다. 오늘날 무수히 많은 데이터마이닝 도구들이 개발되었고, 앞으로도 계속해서 연구 개발되어 질 것이다. 따라서 본 논문에서는 개발된 데이터마이닝 도구들 중에서 비교적 널리 알려져 있는 SAS의 *Enterprise Miner*, SPSS의 *Clementine*과 *Answer Tree*, 그리고 Waikato 대학교에서 개발한 공개소프트웨어인 *Weka*들을 데이터 접근 방법에 대하여 살펴볼 것이다.

*Enterprise Miner*, *Clementine*, 그리고 *Answer Tree*는 ODBC(Open Database Connectivity)를

통하여 데이터베이스에 접근한다. 따라서 이러한 데이터마이닝 도구들은 ODBC 드라이버가 지원된다면 어떠한 지역 데이터베이스에도 접근이 가능하다. 하지만 이러한 데이터마이닝 도구들은 원격 데이터베이스에 접근하는데 있어서 각기 다른 제약을 가지고 있다. 먼저 *Enterprise Miner*와 *Clementine*은 별도의 공개 소프트웨어를 이용하여 원격 컴퓨터에 접속한 다음에 데이터 접근을 위해서는 ODBC를 통하여 원격 데이터베이스에 접근할 수 있도록 하고 있다. 즉 *Enterprise Miner*는 SAS에서 제공하는 *SAS Connect*기능을 이용하여 원격 컴퓨터에 접속하고, *Clementine*은 원격 컴퓨터와의 접속을 위해 이기종 컴퓨터에 접속하여 통합사용을 지원하는 *Exceed*를 사용한다. 반면에 *Answer Tree*는 원격 데이터베이스에 연결하는 기능을 제공하지 않고 있다.

자바로 구현된 *Weka*는 의사결정나무기법, 선형회귀, 모형트리 생성 등의 전통적인 데이터마이닝 통계 분석뿐만 아니라 다양한 데이터마이닝기법을 지원하고 있다. *Weka*는 지역 데이터베이스에 연결을 위하여 JDBC(Java Database Connectivity)를 사용하고, 원격 데이터베이스에 연결을 위하여 자바환경에서 컴퓨터나 프로그램, 객체들 사이에 통신할 수 있는 기능을 제공하는 RMI(Remote Method Invocation)을 사용한다. 따라서 *Weka*를 사용하면 플랫폼에 독립적으로 수행하는 자바의 특성을 기반으로 하기 때문에 서로 다른 기종의 정보자원간의 데이터의 교환이 가능하다. 하지만 *Weka*는 입력 파일을 ARFF(Attribute-Relation File Format)형식으로 변환해야 하는 단점을 가지고 있다.

### 2.2 컴포넌트 객체 모델 분석

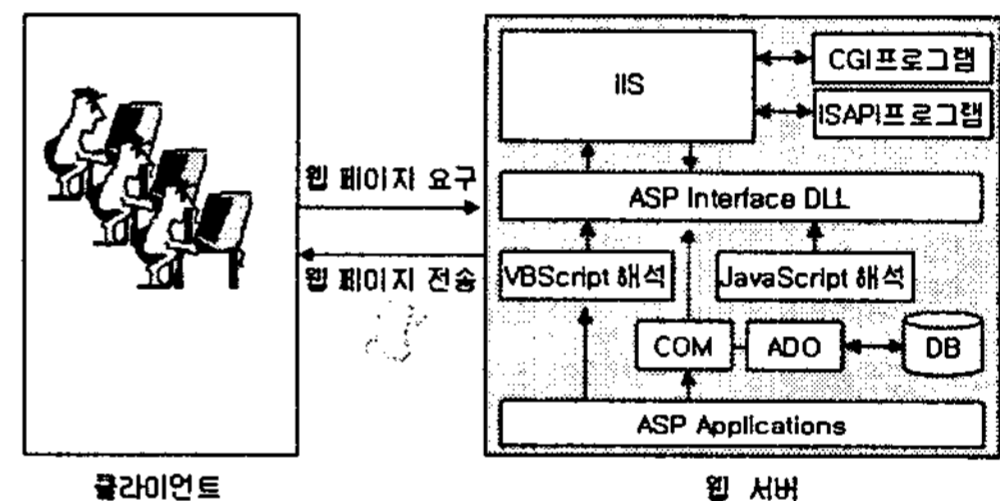
컴포넌트는 새로운 소프트웨어 개발과정에서 부품으로 사용될 수 있도록 미리 구현된 독립된

단위의 모듈 프로그램이다. 이러한 컴포넌트는 소프트웨어 개발과정에서 재사용함으로써 소프트웨어 개발 기간을 단축하고, 소프트웨어 생산성과 품질향상 및 유지보수 비용과 테스트 비용을 절감할 수 있는 장점을 가지고 있다[Philippe, 1998; Simons et al., 2001]. 컴포넌트는 다른 컴포넌트나 프레임워크와 상호작용할 수 있도록 하기 위해 미리 정의된 아키텍처에 맞게 개발되어야 한다. 컴포넌트 기반 개발은 소프트웨어 개발 방법론의 새로운 패러다임으로 여러 관련 기술이 서로 융합하여 이루어진다. 즉 컴포넌트 생성과 관련된 기술은 객체지향프로그래밍(object-oriented programming : OOP)[Sheets et al., 2001; Snyperski, 1998]에 기반의 재사용을 위한 설계 및 개발방법이 있다. 요즘은 OOP의 한계를 극복하고 소프트웨어의 재사용을 극대화할 수 있도록 OOP 패러다임을 이용한 컴포넌트 기반 개발방법이 일반적이다. 요즘은 이러한 컴포넌트 소프트웨어 재사용 기술과 분산 객체 컴퓨팅 기술을 통하여 서비스지향구조(service-oriented architecture : SOA)[Erl, 2005]로 급속하게 발전하고 있다.

마이크로소프트사는 객체지향프로그래밍 기반의 컴포넌트 개발을 위한 표준으로 컴포넌트 객체 모델(component object model : COM)[Box, 1997; Harkey et al., 1998]을 발표하였다. COM은 내부구현을 자세히 모르더라도 운영환경에 독립적으로 사용언어에 중립적인 방법으로 컴포넌트를 사용할 수 있도록 하였다. 이러한 COM이 서로 다른 컴포넌트들 사이에 동적인 상호작용을 위한 소프트웨어 프레임워크를 제공함으로써 컴포넌트 개발자로 하여금 이러한 컴포넌트를 이용하여 원하는 응용시스템을 용이하게 구축할 수 있도록 하였다. 실제로 윈도우시스템의 일부 기능과 마이크로소프트에서 제공하는 대부분의 어플리케이션은 COM으로 개발되어

있어서 이 기술을 활용하는 다양한 어플리케이션 출현의 기틀을 마련하게 되었다.

마이크로소프트사가 기업용 웹서버 IIS(Internet Information Server)에 수행할 수 있는 ASP(Active Server Page)를 처음 발표했을 때 단순히 비주얼베이직 스크립트(VisualBasic script)나 자바 스크립트(Java script) 등과 같은 서버스크립트를 처리하는 목적으로 개발하였다. 하지만 이제는 ASP가 <그림 1>과 같이 서버스크립트 처리뿐만 아니라 COM을 통하여 서로 다른 언어나 개발 틀로 만들어진 서버 컴포넌트를 사용할 수 있도록 발전을 하였다.



<그림 1> ASP 컴포넌트의 상호작용

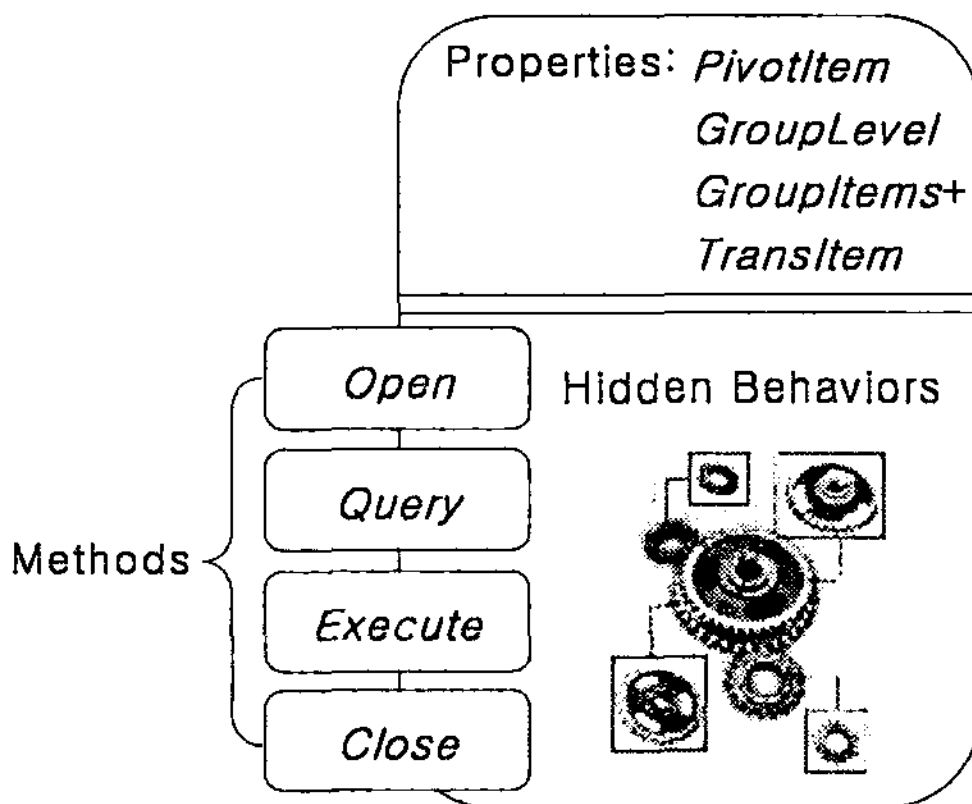
ASP에서 액티브 데이터베이스에 접근은 COM 컴포넌트를 이용한 ADO(ActiveX Data Object) 컴포넌트를 통하여 이루어진다. ADO는 OLEDB(Object Linking and Embedding Database)[Fronckowiak, 1997] 데이터 제공자에게 접근할 수 있는 응용프로그램을 개발하기 위한 응용 프로그램 인터페이스이다. ADO 컴포넌트는 ASP뿐만 아니라 비주얼 C++, 비주얼 J++, 비주얼 베이직 등 수많은 프로그래밍 언어에서 사용할 수 있다. 즉 OLEDB는 직접적으로 데이터 소스에 접근하기 위한 상호작용을 제공하는 반면에 ADO는 OLEDB와 상호작용을 통하여 간접적으로 데이터 소스에 접근한다. COM을 기반으로 하는 ADO 컴포넌트는 특별히 클라이언트 서버 응용시스템 개발을 위하여 고안되었으

므로 클라이언트와 동적으로 대화할 수 있는 데이터베이스 웹서비스 응용프로그램 개발에 적합하다고 할 수 있다.

### 3. ADMO 컴포넌트 설계

#### 3.1 ADMO 클래스 설계

ADMO의 주요한 역할은 운영 데이터베이스로부터 실시간적으로 데이터를 읽어서, 그 데이터를 분류집계(classification and aggregation) 하는 것이다. 이를 위하여 ADMO는 액티브데이터베이스에 저장된 데이터로부터 부류별 특성을 찾아내어 분류모형을 만들고 이를 토대로 집계통계 정보를 추출하는데 목적이 있다. ADMO 클래스는 액티브데이터마이닝 객체생성을 위한 프레임으로서 객체지향 모델링과 디자인을 통하여 다음 <그림 2>와 같이 표현한다.



<그림 2> ADMO 클래스 설계 명세

ADMO 클래스는 이 객체의 상태 값을 표현하기 위한 객체의 특성(property)들의 집합과 이러한 특성 값의 활성화를 위한 메서드(method)들의 집합으로 구성한다. ADMO를 위한 객체의 특성과 메서드의 종류와 그들의 역할은 다음 <표 1>과 같다.

<표 1> ADMO의 프로퍼티와 메서드

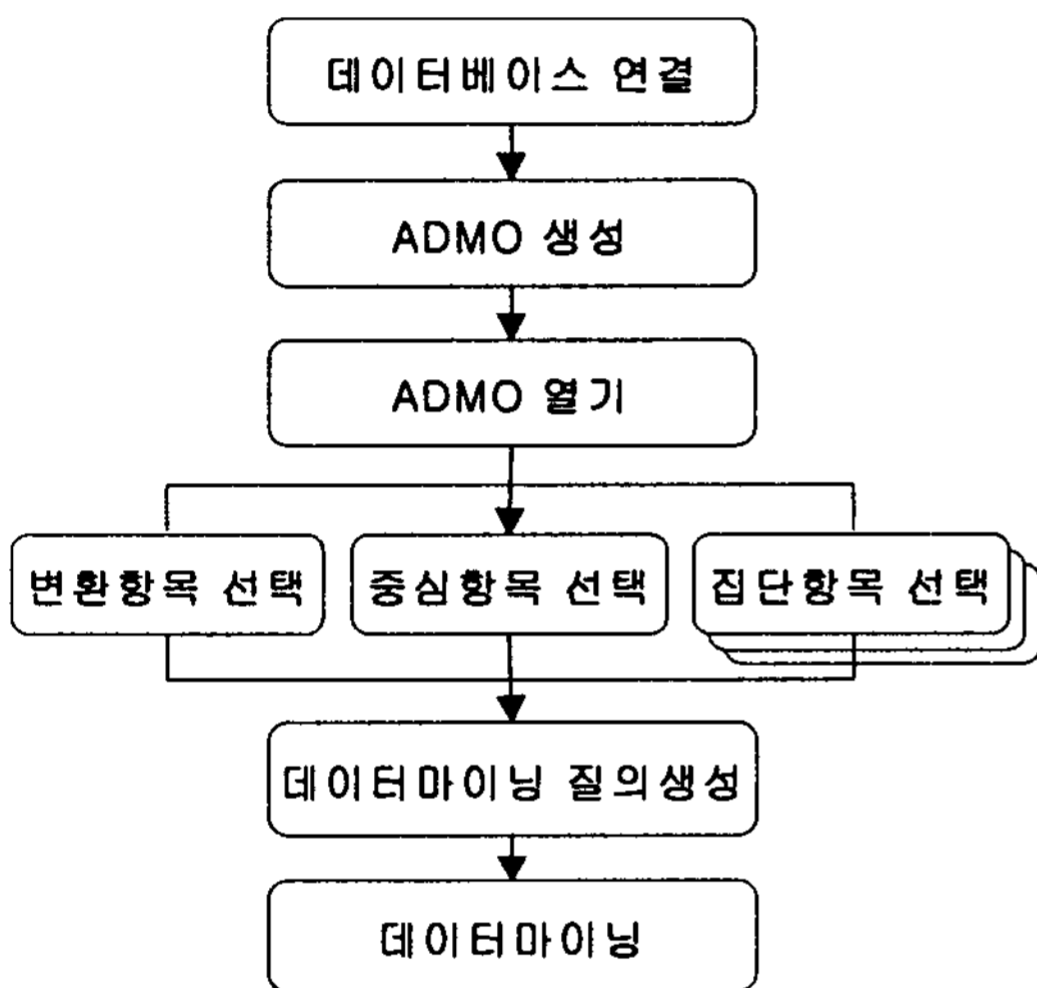
구분	이름	역할
프로퍼티	<i>PivotItem</i>	분류기준 중심항목
	<i>GroupLevel</i>	데이터분류 차원
	<i>GroupItems+</i>	분류 집단항목
	<i>TransItem</i>	변환기준 항목
메서드	<i>Open</i>	객체열기
	<i>Query</i>	질의생성
	<i>Execute</i>	질의처리 및 결과 표시
	<i>Close</i>	객체닫기

ADMO 클래스의 프로퍼티는 <표 1>과 같이 다양한 분류를 만들어 내기 위한 축과 축을 변경하기 사용할 메타데이터를 기술할 피벗항목인 'PivotItem'과 분류 단계를 지정하기 위한 'GroupLevel', 데이터 분류기준이 되는 항목을 지정하기 위한 'GroupItems+'으로 구성한다. 그리고 변환기준 항목으로 'TransItem'을 사용하도록 설계한다. 예를 들면, 회원의 직업과 학력이 신용등급과의 상관관계를 찾아내기 위하여 데이터 분류기준 중심항목을 신용등급이 되고 데이터 분류항목은 직업과 학력의 항목이 되며 분류단계는 2차원이 된다. 그리고 변환기준 항목은 회원의 직업과 학력에 따라 신용등급을 분류해야 함으로 변환기준 항목은 각각의 회원을 식별할 수 있는 회원식별자를 사용하면 된다. 한편, ADMO 클래스의 메서드는 <표 1>과 같이 액티브데이터마이닝 객체 열기를 위한 'Open'과 액티브데이터마이닝 질의 생성을 위한 'Query'로 구성한다. 그리고 액티브데이터마이닝 질의 처리 및 처리결과를 표시하기 위한 'Execute'와 열려진 액티브데이터마이닝 객체를 닫기 위한 'Close'로 이루어진다.

#### 3.2 ADMO 프로세스 설계

ADMO 프로세스는 먼저 ADO의 연결객체를

이용한 액티브데이터베이스 연결 문자열과 ADMO 객체를 생성한 후 생성된 액티브데이터베이스 연결문자열을 이용하여 ADMO를 개방한다. 개방된 ADMO에 필요한 특성들을 선택한 후 데이터마이닝 질의를 생성한 후에 데이터마이닝 과정을 수행한다. 이런 ADMO 프로세스는 다음 <그림 3>에서 구체적으로 표현한다.



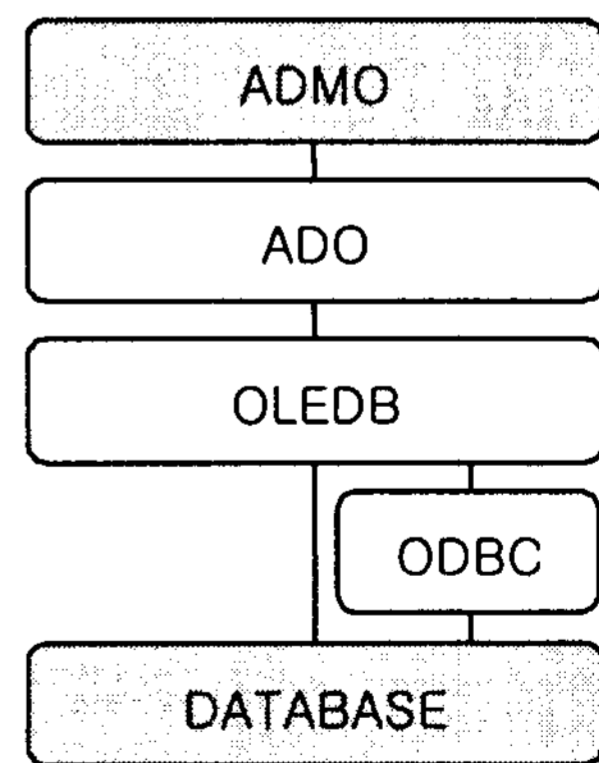
<그림 3> ADMO 프로세스 모델

데이터마이닝 질의를 생성하기 위해서는 변환항목, 중심항목, 그리고 집단항목을 선택하게 되는데 집단항목은 분류차원(classification dimension)에 따라서 한개 이상이 될 수 있다. 가령 신용등급별 직업과 학력에 대한 분류규칙을 발견하는 데이터마이닝 유형이 있다고 가정할 경우에 집단항목은 직업과 학력의 2개가 되고 이때 데이터 분류차원은 2가 된다. 데이터 분류차원은 의사결정 나무의 깊이를 결정하게 되는 변수이다. ADMO 컴포넌트는 데이터 분류차원을 3으로 제한하였다. 한편, 데이터마이닝 질의를 생성하는 것인데 이 단계에서 입력된 데이터마이닝 변수를 통하여 OLAP(On-Line analytic Processing)[Zaiane et al., 1998] 쿼리를

구성한다. 이렇게 생성된 질의는 데이터마이닝 단계에서 처리된다. 데이터마이닝 단계는 쿼리를 처리하는 일과 처리된 결과를 표시하는 역할을 수행하도록 한다.

### 3.3 ADMO 데이터접근 경로

ADMO 데이터접근 경로는 ADMO 컴포넌트가 액티브데이터베이스에 있는 데이터를 접근하기 위한 방법을 정의하는 것으로 각각의 컴포넌트들의 상호협력을 통해서 이루어진다. ADMO 컴포넌트는 기존에 웹 서버 프로그래밍에 많이 사용하고 있는 ADO를 확장한 것이다. ADO는 웹 응용 시스템에 사용하기 적합하도록 발달되어 왔기 때문에 웹 데이터마이닝 [Cooley et al., 1997; Srivastava et al., 2000] 객체 모델로 적합하기 때문이다. ADMO가 액티브데이터베이스에 있는 데이터를 접근하기 위한 표준경로는 <그림 4>와 같다.



<그림 4> ADMO 데이터 접근 표준경로

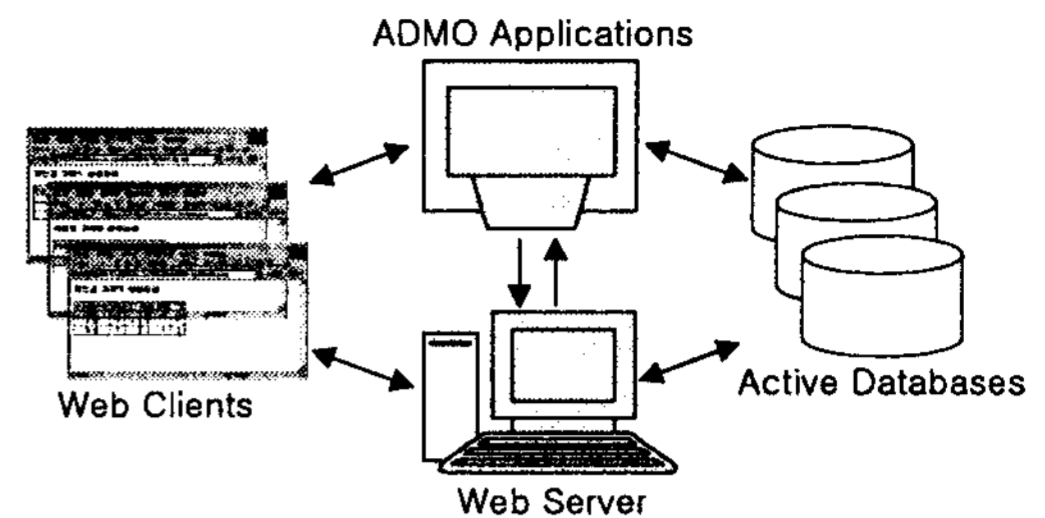
개방데이터베이스연결(open database connectivity : ODBC)은 데이터베이스의 데이터 소스에 대한 일관된 접근을 제공하는 것에 초점을 맞추고 있다. 이런 ODBC는 계속 발전하여 왔

고 현재 엑셀, 액세스 등과 같은 많은 데이터 처리 응용을 지원하고 있다. 그러나 ODBC에 데이터 소스 이름(data source name : DSN)은 운영환경에 종속적으로 존재하기 때문에 보다 폭넓은 목적으로 사용할 수 없다는 단점을 가지고 있다. 이러한 단점을 극복하기 위한 방안으로 데이터 소스에 관계없이 데이터 접근이 가능한 OLEDB(Object Linking and Embedding Database)가 개발되었다. OLEDB는 문서, 이메일, 파일, 스프레드시트, 그리고 데이터베이스에 접근할 수 있도록 하는 역할을 한다. 따라서 OLEDB는 ODBC와 비교해서 데이터 소스들에 대하여 보다 넓은 접근을 가능하게 해준다. 한편 ADO는 OLEDB를 기반으로 하는 응용프로그램 인터페이스(application program interface : API)이다. 액티브 컴포넌트들은 OLE 컴포넌트보다 훨씬 작고 빠르며, 특히 웹 응용프로그램에 포함될 수 있도록 고안되었다는 점을 제외하고 OLE 컴포넌트와 유사한 특징을 가지고 있다. 그리고 액티브 컴포넌트들은 특별히 인터넷 상의 안전한 동적 분배를 위하여 고안되어 왔으며 모든 윈도우 플랫폼과 매킨토시, 유닉스 플랫폼을 지원한다.

## 4. ADMO 컴포넌트 구현

### 4.1 액티브데이터마이닝 구조

ADMO 컴포넌트는 웹 응용프로그램에 일부 분으로 사용할 수 있도록 고안된 프로그램 모듈로 웹 서버에 의하여 수행된다. 웹 서버는 클라이언트에 의하여 요청한 내용을 처리하여 다시 요청한 클라이언트에게 처리된 정보를 제공해주는 역할을 한다. 이런 각 요소들 사이에 상호작용을 위한 액티브데이터마이닝 구조는 다음 <그림 5>와 같다.

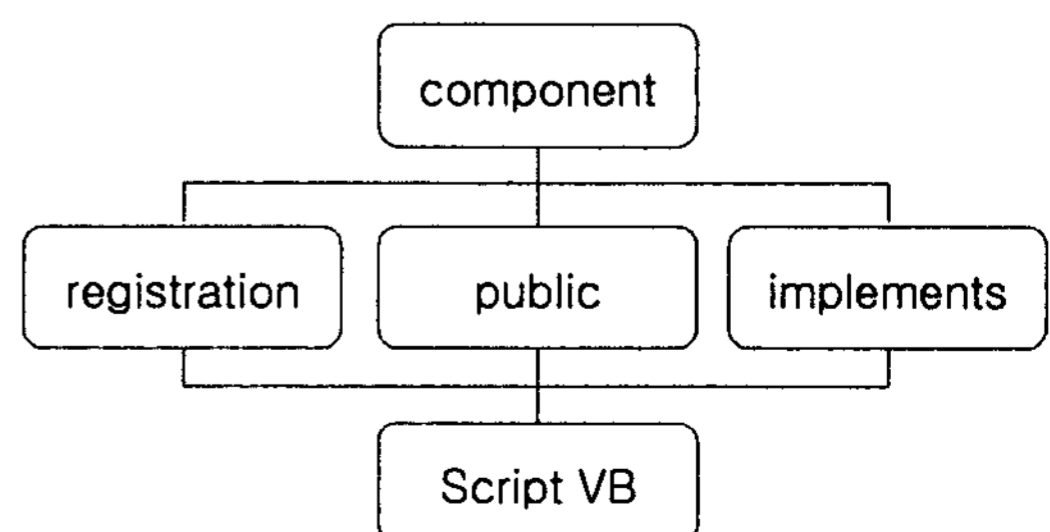


<그림 5> 액티브데이터마이닝 구조

<그림 5>에서 웹 클라이언트에 의하여 웹 서버에 있는 ADMO 응용프로그램의 수행을 요청하게 되면 ADMO는 액티브데이터베이스에 저장된 데이터로부터 부류별 특성을 분류모형을 만들고 이를 토대로 실시간 분류집계 정보를 생성한다. 이렇게 생성된 정보는 웹 서버에 의하여 이를 요청한 클라이언트에게 제공한다.

### 4.2 ADMO 컴포넌트 모듈

ADMO 컴포넌트 구현은 XML(eXtensible Markup Language) 기반의 윈도우스크립트컴포넌트(window script component : WSC)를 이용한다. XML은 컨텐츠 데이터를 표현하기 위한 데이터표기 언어로 메타데이터인 요소태그(element tag)와 속성(attribute), 그리고 텍스트 스트링(text string)으로 구성한다. ADMO 컴포넌트의 요소태그, 즉 구현모듈의 계층구조는 <그림 6>과 같다.



<그림 6> ADMO 컴포넌트 구현모듈 계층

ADMO 컴포넌트를 위한 XML 문서는 XML 문서임을 알려주기 위하여 `<?xml version="1.0" encoding="UTF-16"?>`으로 시작하고 <그림 6>과 같이 하나의 근 요소 태그(root element tag)인 `<component>`을 가진다. `<component>`는 구현 모듈로서 컴포넌트 등록 부, 공개 부, 구현 부로 각각 `<registration>`, `<public>`, `<implements>` 하위 요소 태그(subelement tags)를 포함하도록 하였다.

#### (1) 컴포넌트 등록 부

ADMO 컴포넌트의 등록 부 `<registration>`에서는 컴포넌트 설명 속성 *description*과 프로그램 식별자 속성 *progid*, 그리고 컴포넌트의 버전정보 속성 *version* 등을 기술한다.

```
<registration
  description="A data mining component for
              classification"
  progid="ADMODB.Classification"
  version="1.00">
</registration>
```

프로그램 식별자 속성 *progid*는 외부 응용 프로그램에서 컴포넌트 객체를 생성할 경우에 사용하는 문자열로 특별히 중요한 의미를 가진다. 예를 들면 ASP(ActiveX Server Page)[Francis et al., 2000]프로그램에서 *dmobj*이라는 이름의 컴포넌트 객체를 생성할 경우에 다음과 같은 구문형식의 문장을 기술하도록 하였다.

```
Set dmobj = Server.CreateObject
              ("ADMODB.Classification")
```

#### (2) 컴포넌트 공개 부

ADMO 컴포넌트의 공개 부는 외부 응용프로

그램과 본 컴포넌트 객체의 상호작용 수단을 기술한다. 이곳에서는 객체의 특성을 기술하는 프로퍼티와 객체의 행위의 메서드를 기술하였다. 객체의 프로퍼티는 테이블 1에서와 같이 *pivotitem*, *groupitem*, *grouplevel*, 그리고 *trasiitem*을 기술하였다. 그리고 객체의 행위를 기술하는 메서드는 *open*, *query*, *execute*, 그리고 *close*를 기술하였다.

#### (3) 컴포넌트 구현 부

ADMO 컴포넌트의 구현 부는 다음과 같이 공개부에서 정의한 각각의 프로퍼티와 메서드에 대한 함수를 정의하고, 메서드에 행위를 기술한다.

```
Algorithm ADMODB_Classification
Begin
  function get_PivotItem();
  function put_PivotItem(newValue);
  function get_TransItem();
  function put_TransItem(newValue);
  function get_GroupItem()+;
  function put_GroupItem(newValue)+;
  function get_GroupLevel();
  function put_GroupLevel(newValue);
  function Open(ActiveConnection);
  function Query(DataSoure, SourseType);
  function Execute(CommandType);
  function Close();
End
```

*Open* 함수에서 데이터베이스 액티브 연결 문자열을 받고 *Query* 함수에서 입력된 프로퍼티를 이용하여 부류별 특성을 찾아내어 분류모형을 만들고 이를 토대로 액티브데이터베이스로부터 집계통계 정보를 추출하기 하여 다음과 같은 형식의 OLAP(On-Line Analytic Processing) 쿼리를 생성하게 된다.



```

TRANSFORM Count(TransItem)
SELECT      GreoupItems, Count(TransItem)
FROM        ActiveDatabaseTable
GROUP BY    GroupItems
ORDER BY    GroupItems
PIVOT      PivitItem

```

Query 함수에서 생성된 쿼리는 *Execute* 함수에서 실행하여 레코드셀을 생성한 후 생성된 레코드셀을 읽어 집계통계 정보를 표시할 수 있도록 하였다. 끝으로 *Close* 개방된 액티브데이터베이스 데이터마이닝 객체를 해제하고, 사용했던 모든 컴퓨팅 자원을 반납하도록 하였다.

이와 같이 개발된 ADMO 컴포넌트는 운영체제의 하나의 소스파일로 존재하게 되는데 WSC나 SCT 파일의 속성으로 로 저장한다. 이렇게 존재하게 된 파일은 운영체제에 의하여 컴파일되어 동적 연결 라이브러리(dynamic linking library : DLL)에 등록하여 사용한다.

## 5. ADMO 컴포넌트 응용

본 논문에서 개발한 ADMO 컴포넌트의 응용 사례를 시험하기 위한 컴퓨터 시스템 플랫폼은 윈도우즈2000 운영체제의 IIS5.0(Internet Information Server Version 5.0) 웹 서버에 웹 응용프로그램 ASP3.0을 이용한다. 모의시험 데이터는 데이터마이닝을 위하여 전자상거래 쇼핑몰의 회원 정보를 가정한다. 그리고 ADMO 컴포넌트 응용 사례는 전자상거래 쇼핑몰에서 고객들의 회원등급을 통한 고객 세분화를 통한 목표고객을 찾기 위한 데이터마이닝 사례를 보일 것이다.

### 5.1 ADMO 컴포넌트 응용 환경

#### (1) 모의시험 데이터

현재 많은 전자상거래 쇼핑몰 중에서 고객관

계관리(customer relationship management : CRM)[Berry et al., 2004]를 위하여 회원들의 거래실적, 직위, 나이 등의 자료를 분류하고 집계하여 고객 각자의 특성에 적합한 일대일 마케팅 활동을 하고 있다. 고객관계관리를 효과적으로 하기 위한 방법으로 전자상거래 쇼핑몰은 회원을 여러 가지 등급으로 분류하고 있다. 본 논문에서 ADMO 객체의 응용 사례를 보이기 위하여 다음 <표 2>와 같이 어떤 쇼핑몰의 고객 데이터를 가정한다.

<표 2> ADMO 응용을 위한 모의실험 데이터

	id	name	gender	age	scholarship	occupation	rank
	7711	홍길동	남	52	고등	지역	보통
	7712	이수일	남	25	대원	직장	우수
	7713	심순애	여	44	고등	무직	보통
	7714	한송미	여	32	대학	직장	보통
	7715	심현섭	남	33	대원	지역	대우
	7720	김대희	남	30	고등	지역	우수
	7721	이상해	남	50	중등	직장	보통
	7724	미사도	여	25	고등	무직	보통
	7725	최지우	여	32	고등	무직	우수
	7726	배용준	남	55	중등	무직	불량
	7788	미소라	여	40	대원	지역	우수
	7789	차인표	남	53	고등	무직	불량
	7790	최시라	여	38	대학	직장	우수
	7791	김소은	여	47	중등	무직	불량
	7792	이진수	남	28	대원	지역	우수
	7793	박수진	여	30	대학	직장	불량
	7977	미유진	여	26	대학	직장	보통
	7978	김봉수	남	44	대학	지역	우수
	7979	백애란	여	37	고등	지역	우수
	7999	최성립	남	30	대학	직장	우수

모의실험 데이터는 20명의 회원을 대상으로 하였고, 각 회원의 메타데이터는 회원 아이디(id), 성명(name), 성별(gender), 나이(age), 학벌(scholarship), 직업(occupation), 그리고 신용등급(rank)으로 구성하였다.

#### (2) 응용시험 매개변수

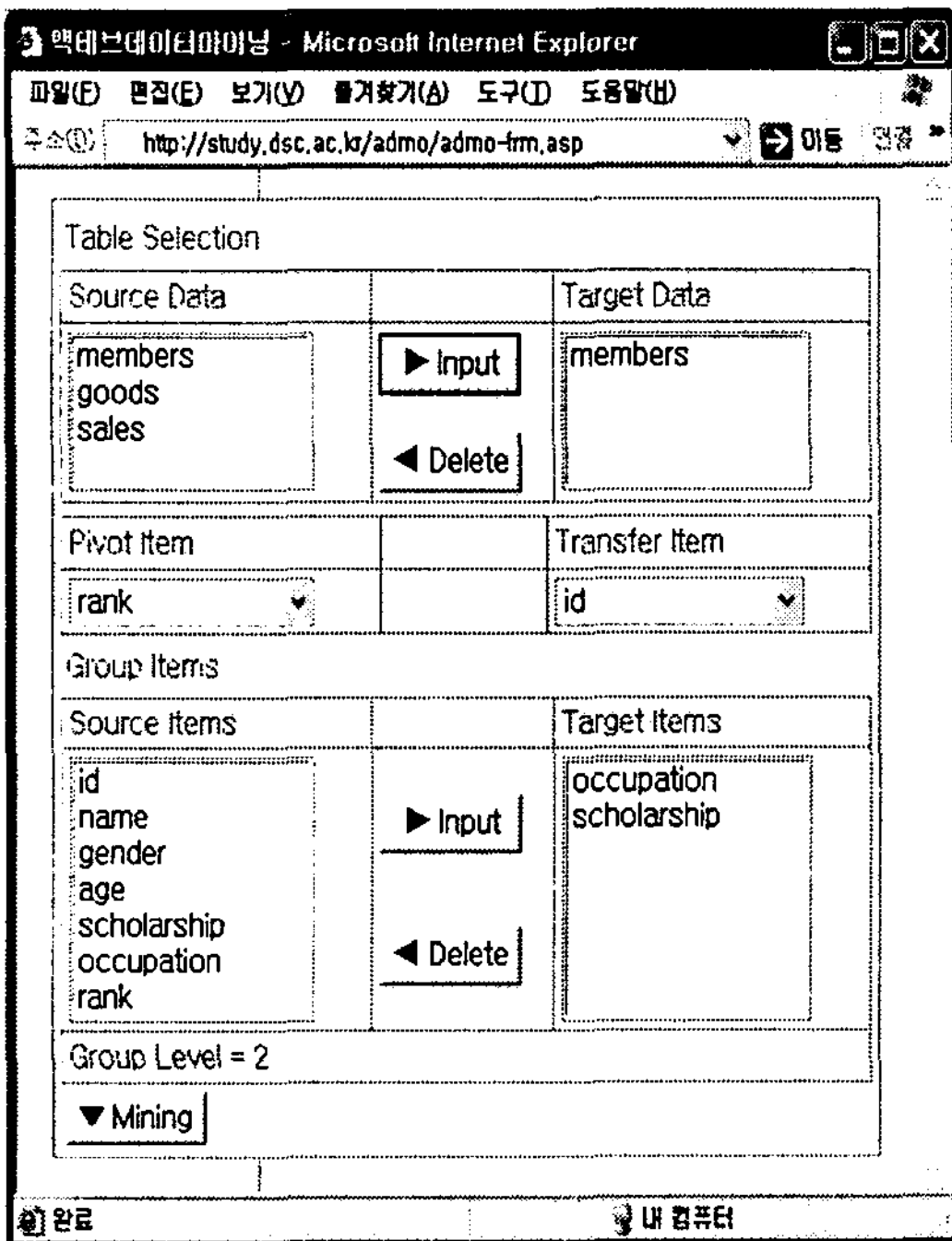
ADMO 컴포넌트 수행에 필요한 매개변수들은 데이터소스의 선택과 데이터마이닝 분류방법에 적합한 데이터 속성들이 있다. 즉 데이터소스는 분류할 데이터 속성이 포함된 데이터 테이블을 선택하게 되는데 2개 이상의 테이블이 선택되면 선택된 데이터 테이블은 서로 조인되

어 하나의 데이터 레코드 셀을 생성하게 된다. 한편 데이터마이닝 분류방법에 따른 데이터 속성으로는 ADMO 객체의 프로퍼티를 결정하는 것으로 피벗항목, 변환항목, 변환그룹항목을 선택하여야 한다. 예를 들어 회원의 직업과 학력에 대한 신용등급의 분류집계 정보를 추출하기 위해서 <그림 7>과 같이 ADMO 응용시험 매개변수를 선택하여야 한다.

### 5.2 응용사례

(1) 응용사례 1 : 직업과 학력이 신용등급에 미치는 영향

모의시험 데이터 (1)에서 직업과 학력이 신용등급에 미치는 영향을 출력하기 위해서는 데이터 소스 테이블은 *members*를 선택 하고 분류 집계를 위한 변환 중심항목은 신용등급 속성명인 *rank*로 설정한다. 그리고 회원 직업과 학력에 대한 신용등급 회원 수를 추출하기 위하여 변환항목은 회원 아이디 속성인 *id*를 선택한다. 직업과 학력에 대한 분류집계를 위한 그룹 항목은 직업 속성 *occupation*과 학력속성 *scholarship*의 2개를 선택하여 2차원 분석을 한다. 그리고 데이터 분류 분류차원은 2가 된다. 이러한 응용시험 매개변수는 <그림 7>의 ADMO 응용 매개변수 입력 화면을 통해서 쉽게 결정할 수 있다. <그림 7>의 ADMO 응용 매개변수 입력 화면을 통해서 선택한 후 데이터 마이닝을 수행하면 다음과 같은 응용 프로그램이 자동으로 생성되어 실행된다. 하지만 <그림 7>의 ADMO 응용 매개변수 선택 화면을 이용하지 않고도 ADMO 응용 프로그램에서 직접 다음과 같은 소스코드를 직접 입력해도 같은 결과가 된다.



<그림 7> ADMO 응용을 위한 매개변수 입력 화면

즉 회원의 직업과 학력에 대한 신용등급의 분류집계 정보를 추출하기 위해 <그림 7>에서 회원 테이블인 *members*를 목표 데이터(target data) 테이블로 선택하고 피벗항목(pivot items)은 *rank* 속성을, 변환항목(transfer items)은 *id*, 그리고 그룹항목(group items)은 *occupations*, *scholarship*의 두 개를 선택한다. 따라서 데이터마이닝 분류 그룹 차원(group levels)은 그룹 항목의 개수인 2로 자동으로 결정된다.

```
<%
1. Set DBConn=Server.CreateObject("ADODB.Connection")
2. DBConn.Open "memdb"
3. Set DM = Server.CreateObject("ADMODB.Classification")
4. DM.Open(DBConn)
5. DMPivotItem = "rank"
6. DM.TransItem = "id"
7. DM.GroupItem1 = "occupation"
8. DM.GroupItem2 = "scholarship"
9. DM.GroupLevel = 2
10. CommandText = DM.Query("members", 2)
11. DM.Execute(CommandText)
12. DM.Close
13. Set DM = Nothing
%>
```

즉 ADMO 응용 프로그램은 DBConn이라는 데이터베이스 연결객체를 생성(행 번호 1)하여 memdb라는 데이터 소스이름을 통하여 액티브 데이터베이스에 접속(행 번호 2)을 한다. 또한 DM이라는 이름의 ADMO 객체를 생성(행 번호 3)한 후 액티브 연결객체를 통하여 DM객체를 개방(행 번호 4)한다. 그리고 DM객체의 분류기준 중심항목 지정(행 번호 5), 변환기준 항목 지정(행 번호 6), 분류 집단 항목 지정(행 번호 7~8), 데이터 분류 차원 지정(행 번호 9)한다. 한편, 데이터마이닝 질의 생성(행 번호 10)과 질의처리 및 결과출력(행 번호 11), 그리고 DM객체를 닫기(행 번호 13)를 한다. 이와 같은 ADMO 컴포넌트 응용 프로그램의 실행결과는 다음 <그림 8>과 같다.

occupation	scholarship	total	대우	보통	불량	우수
무직	고등	4		2	1	1
	중등	2			2	
자영	고등	3		1		2
	대원	3		1		2
직장	대학	1				1
	대원	1				1
	대학	5		2	1	2
	중등	1		1		

<그림 8> 직업별 학력별 신용등급 현황

<그림 8>과 같이 직업과 학력이 신용등급에 미치는 영향을 분석한 결과를 살펴볼 때 일반적으로 직업이 무직이며 학력이 중졸인 사람의 신용등급은 불량일 확률이 높다고 말할 수 있다. 한편 직업이 자영이고 학력이 대졸이거나 직업이 직장인이고 학력이 대원이면 우수회원일 확률이 높다는 것을 알 수 있다. 또한, 직업이 무직이고 학력이 고등이며 직업이 직장이고 학력

이 대학인 경우는 신용등급을 정확히 예측하기가 어렵다. 이러한 결과로 볼 때 전자상거래 쇼핑몰 관리자는 직업이 무직이고 학력이 중졸인 회원들에게는 제한적인 마케팅을 직업이 자영이고 학력이 대졸이거나 직업이 직장이고 학력이 대원인 회원들에게는 적극적인 마케팅을 통하여 구매를 촉진할 수 있다.

(2) 응용사례 2 : 직업과 학력, 성별이 신용등급에 미치는 영향

회원의 직업과 학력, 성별이 신용등급에 미치는 영향을 분석하기 위해서는 데이터 소스 테이블은 members를 선택하고 분류집계를 위한 변환 중심항목은 신용등급 속성명인 rank로 설정한다. 그리고 회원 직업과 학력에 대한 신용등급 회원 수를 추출하기 위하여 변환항목은 회원 아이디 속성인 id를 선택한다. 직업과 학력에 대한 분류집계를 위한 그룹 항목은 직업 속성 occupation과 학력속성 scholarship에 성별속성 gender을 추가하여 3차원 분석을 한다. <그림 7>의 ADMO 응용을 위한 매개변수 입력 화면을 이용하여 수행한 결과는 다음 <그림 9>와 같다.

occupation	scholarship	gender	total	대우	보통	불량	우수
무직	고등	남	1			1	
		여	3		2		1
자영	중등	남	1			1	
		여	1			1	
	고등	남	2		1		1
		여	1				1
직장	대학	남	2	1			1
		여	1				1
	대원	남	1				1
		여	1				1
	대학	남	1				1
		여	4		2	1	1
	중등	남	1		1		

<그림 9> 직업별 학력별 성별 신용등급 현황

<그림 9>와 같이 직업과 학력, 성별이 신용등급에 미치는 영향을 살펴보면 일반적으로 직업이 무직이고 학력이 중졸인 남자와 여자, 직업이 무직이고 고졸인 남자의 신용등급은 불량일 확률이 높다는 사실을 알 수 있다. 한편, 직업이 자영이고 학력이 대졸이거나 직업이 직장인이고 학력이 대원인 남자와 여자 모두는 우수일 확률이 높을 수 있다. 그리고 작업이 무직이고 학력이 고등과 직업이 직장이고 학력이 대학인 경우는 신용등급을 정확히 예측할 수 없다는 것을 알 수 있다. 이러한 결과로 볼 때 전자상거래 쇼핑몰 관리자는 직업이 무직이고 학력이 중졸인 남녀 모두에게는 제한적인 마케팅 정책을 수립하고 직업이 자영이고 학력이 대원의 여자이거나 직업이 자영이고 학력이 대학의 남자는 신용등급이 보다 높기 때문에 더욱 더 적극적인 마케팅 정책을 수립할 필요가 있다.

## 6. 결론 및 향후 연구

본 논문에서 개발한 액티브데이터마이닝 객체 ADMO 컴포넌트는 맞춤형 웹 서비스를 위하여 액티브데이터베이스에서 실시간 데이터마이닝을 통하여 언제 어디서나 필요한 통계분석 예측정보를 제공할 수 있도록 하는데 목적을 두었다. ADMO 컴포넌트는 다른 응용프로그램과 상호협력을 통하여 수행하는 프로그램 모듈이다. 이러한 개발결과는 기존에 데이터마이닝 도구들이 액티브데이터마이닝을 통한 통계분석 예측정보를 제공하는데 한계를 극복하기 위한 시도이다. 이러한 시도는 문헌적인 액티브데이터마이닝 개념을 실체화한 것으로 실용학문 연구에 중요한 의의를 가질 수 있을 뿐만 아니라 인터넷 웹 서비스를 위한 실시간 데이터마이닝 및 지식탐색의 실체적 모습을 제시한 것이 본 논문의 중요한 의의이다. 본 연구결과는 액티브데이

터마이닝 응용모델을 연구하는데 중요한 자료로 활용할 수 있을 뿐만 아니라 데이터마이닝 컴포넌트 개발에 중요한 명세서가 될 수 있을 것이다.

ADMO 컴포넌트는 운영 데이터베이스로부터 실시간적으로 데이터를 읽어서, 그 데이터를 분류 집계하는 데에 개발범위를 두었기 때문에 다양한 예측정보를 추출하기 위한 액티브데이터마이닝에 적용하는데 그 한계를 가지고 있다. 데이터마이닝 기법은 의사결정 나무, 신경망, 연관규칙, 그리고 군집규칙 등이 있지만 비즈니스 환경에서 일반적으로 많이 사용하는 분류규칙을 발견하기 위한 컴포넌트 개발만을 본 연구의 범위에 포함하였다. 그리고 ADMO 컴포넌트 응용결과는 연관되는 항목들 사이에 값을 분류하는 결과를 도출하였지만 도출된 결과를 통한 가설의 결과를 이용한 지식을 도출하는 방안은 본 논문에서 다루지 않았다. 이러한 것은 의사결정을 위한 전략정보를 찾아내는데 중요한 역할을 담당하지만 본 논문에서는 그의 중간 결과를 도출하는 컴포넌트를 개발하는데 목적을 두었다. 한편 데이터마이닝은 정확한 예측 정보와 그 정보를 인지하고 올바른 의사결정을 지원하기 위해서는 데이터마이닝 결과를 시각화하는 것이 중요하다. 하지만 본 논문에서는 그의 결과를 테이블로 표시하는데 제한을 두었고 그를 도표화하는 데에 필요한 컴포넌트 개발은 다루지 않았다.

효과적인 액티브데이터마이닝을 위해서는 XML 데이터 웹 서비스 모델과 같은 새로운 형태의 유연한 데이터 모델의 연구가 필요하다. 본 논문에서 개발한 ADMO 컴포넌트의 응용결과를 나무구조의 XML 데이터로 변환하여 생성하고, XML 데이터를 통하여 의사결정에 필요한 숨겨진 다양한 예측정보를 발굴할 수 있을 것이다. 한편, 본 논문에서 제시한 분류규칙 이외에 다른 데이

터마이닝기법의 컴포넌트 객체를 개발하는 것도 향후 과제이다. 실제 비즈니스 환경에 필요한 정보는 데이터 소스에 따라 다양하며 이러한 다양한 정보를 제공하기 위해서는 다양한 유형의 데이터마이닝 기법이 필요하다. 이러한 데이터마이닝을 표준적으로 접근할 수 있는 통합된 표준 액티브데이터마이닝 컴포넌트와 데이터마이닝 결과를 검증하고 이들의 결과를 시각화할 수 있는 컴포넌트의 개발이 필요하다. 따라서 본 논문에서 다루지 못한 내용을 보완하고 발전 방안을 도출하여 향후 계속되는 연구에서 이를 수행할 것이다.

## 참 고 문 헌

- [1] Agrawal, R. and Psaila, G., "Active Data Mining", *In Proceedings of First International Conference on Knowledge Discovery and Data Mining(KDD-95)*, 1995, pp. 3-8.
- [2] Bailey-Kellogg, C. and Ramakrishnan, N., "Active Data Mining of Correspondence for Qualitative Assessment of Scientific Computation", *The 17th International Workshop on Qualitative Reasoning Hosted by University Brasilia, Brasil*, Aug. 2003, pp. 23-30.
- [3] Balke, W. and Wagner, M., "Towards Personal of Web Services", *In Proceeding of the International World Wide Web Conference(WWW)*, Budapest, Hungary, 2003.
- [4] Berry, J. A. and Linoff, G., *Data Mining Technigues : For Marketing, Sales, and Comstomer Relationship Management*, Second Edition, John and Sons, 2004.
- [5] Box D., *Essential COM*, Addison-Wesley International, 1997.
- [6] Brachman, R. and Anand, T., "The Process of knowledge Discovery in Database : A Human-Centered Approach", *In Advances in Knowledge Discovery and Data Mining*, 1996, pp. 37-58.
- [7] Chen, M. S., Han, J., and Yu, P. S., "Data Mining : Overview from a Database Perspective", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, December 1996, pp. 866-883.
- [8] Choi, Y. and Moon, S., "Lightweight Multi-granularity Locking for Transaction Management in XML Database Systems", *The Journal of Systems and Software*, Vol. 78, 2005, pp. 37-46.
- [9] Cooley, R., Mobasher, B., and Strivasta, J., "Web Mining : Information Discovery and Pattern Discovery on the World Wide Web", *In Proceedings of the 9th International Conference on Tools with Artificial Intelligence*, IEEE, 1997.
- [10] Elder IV, J. F. and Abbott, D. W., "A Comparison of Leading Data Mining Tools", *The Fourth International Conference on Knowledge Discovery and Data Mining*, New York, Friday, August 28, 1998.
- [11] Erl, T., *Service-Oriented Architecture (SOA) : Concept, Technology, and Design*, Prentice Hall, 2005.
- [12] Fayyad, U., Piatetsky-Shapito, G., Smyth, P., and Uthurusamy, R., "From Data Mining to Knowledge Discovery : An Overview", *Advance in Knowledge Discovery and Data*

- Mining*, AAAI/MIT Press, CA, 1996.
- [13] Fronckowiak, H. W., *OLE DB and ADO*, SAMS Pub., 1997.
- [14] Harkey, O. and Eddon, H., *Inside Distributed COM*, Microsoft Press, 1998.
- [15] Mehmed K., "Data Mining Concepts, Models, Methods, and Algorithms", *IEEE Press*, US, 2002, pp. 120-143.
- [16] Heineman, G. T. and Councill, W. T., *Component-Based Software Engineering*, Addison Wesley, 2001.
- [17] Huh, M. Y. and Song, K. R., "The Prospect of the Structure of Data Mining Solution in the Future", *International Conference on Data Mining, Visualization and Statistical System*, KSS, 2000.
- [18] Philippe K., "Modeling Component System with the Unified Modeling Language", *International Workshop on Component-Based Software Engineering*, 1998.
- [19] Powers, S., *Developing ASP Components*, O'Reilly, 1999.
- [20] Sheets, S. D. and Tegarden, D. P., "Illustrating the Cognitive Consequences of Object-Oriented System Development", *The Journal of Systems and Software*, Vol. 59, 2001, pp. 163-179.
- [21] Simons, J., Graham, K., Owen, A., Patterson, K., and Hodges, J., "Perceptual and Semantic Components of Memory for Objects and Faces : A PET Study", *Journal of Cognitive Neuro-Science*, Vol. 13, No. 4, 2001, pp. 430-443.
- [22] Snyperski C., *Component Software : Beyond Object-Oriented Programming*, Addison-Wesley International, 1998.
- [23] Srivastava, J., Cooly, R., Deshpande, M., and Tan, P.-N., "Web Usage Mining : Discovery and Applications of Usage Pattern From Web Data", *ACM SIGKDD Explorations*, ACM SIGKDD, Jan 2000.
- [24] Tang, Z., MacLennan, J., and Kim, P. P., "Building Data Mining Solutions with OLE DB for DM and XML for Analysis", *ACM SIGMOD Record*, Vol. 34, No. 2, June 2005, pp. 80-85.
- [25] Zaiane, O. R., Xin, M. J., and Han J., "Discovering Web Access Patterns and Trends by Applying OLAP and Data Mining Technology on Web Logs", *Proceeding Advances in Digital Libraries Conference*, 1998, pp. 19-29.

#### ■ 저자소개



#### 최 용 구

동서울대학 컴퓨터정보과 부교수로 재직 중이고, 한국과학기술원에서 경영공학전공 데이터베이스분야 박사학위를 취득하였다. 관심분야는 데이터 마이닝, XML 데이터 웹서비스, DBMS 등이다.