

# 학생정보마이닝 시스템의 설계 및 구현

공헌선<sup>†</sup> · 김 명<sup>††</sup>

## 요 약

현재 초·중등학교에서는 교육행정시스템을 사용하여 학생들의 데이터를 관리한다. 이 시스템의 일부인 교무업무지원 시스템은 학사 및 교무 업무를 지원하는 단위 시스템으로써 학생들에 대한 각종 데이터를 저장하고 있으며, 데이터의 단순 검색 및 통계 데이터 계산을 허용하여 교육 정보 자원의 통합관리에 큰 도움을 준다. 그러나 현재의 교무업무지원 시스템은 데이터를 개별적으로 관리하고 있기 때문에 데이터 사이의 연관 관계로부터 도출될 수 있는 유용한 정보는 제공하지 못한다. 본 연구에서는 저장되어 있는 데이터의 연관 관계 분석을 통해 학생 교육에 유용한 데이터 추출이 가능하다는 것을 보이고, 그러한 분석을 손쉽게 할 수 있는 도구인 SIMS를 설계하고 구현하였다. SIMS는 데이터 상호 관계 분석에 연관규칙 기법을 사용한다. 유사한 기능을 제공하는 기존의 도구들을 사용하려면 전문 지식이 필요한 반면 SIMS는 일선 교사들이 수월하게 사용할 수 있으며 현재의 교육행정시스템에 연동되어 사용될 수 있다는 장점이 있다.

## Design and Implementation of A Student Information Mining System

Hyun-Seon Kong<sup>†</sup> · Myung Kim<sup>††</sup>

### ABSTRACT

Elementary schools and middle schools currently manage their student data by using the education administration system. One of its sub-systems called 'the academic affairs support system' is especially dedicated to handle school and academic affairs data. By allowing simple data search and statistical data calculations, it helps teachers easily integrate and manage education information resources. However, it is not easy for teachers to analyze the correlations among student data. In this paper, we showed by examples that a lot of meaningful information can be extracted by analyzing the relations among student data. Based on the results, we designed and implemented SIMS as a tool to provide teachers with such services. SIMS makes use of Association Rules for data correlation analyses. SIMS can be used in connection with the academic affairs support system, and is much easier to use than previously developed commercial products for similar services.

### 1. 서 론

초·중등학교에서는 교육정보화 사업의 일환으

로 1997년부터 추진된 초·중학교교종합정보관리 시스템을 교육행정업무처리에 사용하고 있다[2, 9]. 이 시스템은 학생의 기본 인적 사항, 가족 사항, 출결 사항, 성적, 행동 발달 사항, 취미 활동 등과 같은 기본 데이터를 교사들에게 제공하여 학생 관리를 편리하게 할 수 있도록 한다[5]. 그러나 데이터가 개별적인 화면에서 결과 위주로

<sup>†</sup> 준 회 원: 이화여자대학교 교육대학원 컴퓨터교육전공  
<sup>††</sup> 정 회 원: 이화여자대학교 컴퓨터학과 교수  
 논문접수: 2002년 12월 23일, 심사완료: 2003년 1월 23일  
 본 연구는 한국과학재단 목적기초연구 (R-04-2001-000-00191-0) 지원으로 수행되었음.

입력되며 데이터간의 연계 기능이 미흡하여 학생 데이터의 상호 연관 관계를 분석하기는 힘들다.

교사는 학생들의 생활 지도나 성적 지도, 학생의 성향 파악 등을 위해 학생들에 대한 여러 자료를 통합적인 관점에서 바라볼 필요가 있다. 실제로 가족 환경, 성적, 교우 관계 등의 데이터는 서로 유기적으로 연결되어 각각 다른 요인에 대해 원인으로 작용하기도 하고 결과로 나타나기도 한다. 학생 데이터로부터 숙련된 교사들의 실제적 경험과 지식을 바탕으로 유용한 패턴이나 데이터간의 상호 연관 관계를 추출하여, 점차 다양해지고 빠르게 변해 가는 학생들의 변화에 적응적으로 대처할 필요가 있다. 그러나 현재의 종합정보시스템으로부터는 이와 같은 부가가치정보를 추출하기가 쉽지 않다.

데이터의 연관관계를 추출하는 한 가지 방법은 데이터 마이닝 기법을 이용하는 것이다. 데이터 마이닝은 대용량의 데이터로부터 데이터 자체에 나타나 있지 않은 패턴, 부가가치 정보, 지식 등을 추출하는 기법이다[3]. 이러한 기법들 중에서 연관규칙(Association Rules)은 장바구니 분석(Market Basket Analysis)이라고 불리기도 하며, “전체 학생들 중에서 상위 10%로 학업 성적이 우수하고, 성격이 외향적인 학생들의 90%이상은 친구 관계가 좋다.”와 같은 규칙을 얻어내는데 쓰일 수 있는 방법이다[1, 3].

본 연구에서는 데이터 마이닝 기법을 사용하여 학생들의 정보로부터 학생 지도를 위해 유용한 정보를 추출할 수 있다는 것을 여러 예제를 통해 보였다. 또한 현재 초·중등학교 교육행정업무 위해 체계적으로 구축되어 있는 교육행정시스템과 연계하여 손쉽게 사용할 수 있는 학생정보마이닝 시스템(SIMS: Student Information Mining System)을 설계하고 구현하였다. 이 시스템은 전문지식을 요하는 Clementine(SPSS), Darwin(Oracle), Enterprise Miner(SAS), Intelligent Miner(IBM)[10]등과 같은 기존의 시스템들보다 사용이 간단하여 일선의 교사들이 유용한 정보를 손쉽게 추출할 수 있도록 설계되어 있다.

## 2. 관련 연구

교무업무 지원을 위해 데이터 마이닝 기법을 직접 응용하는 시스템은 아직 개발되어 사용되지 않는 것이 없다. 여기서는 본 연구의 기반이 되는 초·중등학교 종합정보관리시스템의 구조와 데이터 마이닝 기술에 대해 간단히 소개하고자 한다.

### 2.1. 초·중등교종합정보관리시스템

교육 정보화 사업은 정부가 선진적인 학교 교육 환경 구축을 위해 다각적으로 벌인 교육 환경 개선 사업이다. 교육 행정분야에서 교육 정보화 사업의 하나로 추진된 초·중등 종합정보관리시스템[9]은 클라이언트/서버(Client/Server), GUI(Graphic User Interface), 관계형 데이터베이스를 기반을 개발되었으며, 교무업무지원, 학교경영업무지원, 교육정보유통, 교육정보통합안내를 주요 기능으로 하고 있다.

그 중 교무업무지원 시스템은 학사, 교무 업무를 지원하는 단위 시스템으로, 학생의 기본 인적 사항, 가족 사항, 출결 사항, 성적, 행동 발달 사항, 취미 활동 등의 내용으로 구성되어 있다[5,9]. 주요 기능은 기본적인 학생 자료의 검색 및 저장, 반편성, 시간표 작성, 성적 처리, 입시 자료 생성 등이며, 월별, 년도별 출결 통계나 성적 자료의 경우 반별 과목별 점수평균, 표준편차 등을 통계 자료로 제공하고 있다.

이러한 기능을 가지는 초·중등학교종합정보관리시스템이 구축됨으로써 교육 행정의 투명성과 효율성이 향상되었으며, 교육 관련 정보 자원을 통합적으로 관리할 수 있는 기반과 환경이 조성되었다. 그러나 아직까지 이러한 자원들은 개별적으로 활용되고 있고, 서로 연계되어 복합적인 분석이 가능한 단계에는 이르지 못하고 있다.

### 2.2. 데이터마이닝과 연관규칙 기법

데이터 마이닝은 지식의 발굴(Knowledge Discovery in Databases, KDD)이라는 용어로 사용되기도 하는데, 이는 대용량 데이터베이스에

존재하지만 숨겨져 있는 상호 관련성과 패턴에 찾아 부가가치를 창출하는 IT기술이다. 데이터로부터 특정한 연관규칙을 찾아내는 것은 데이터 마이닝의 주요 작업으로 자율학습을 기반으로 하는 지식발견시스템의 핵심이라 할 수 있다[4]. 연관규칙 탐사는 시장 바꾸니 분석과 같이 동시에 발생하는 사건 그룹 내에서 존재하는 친화성(Affinity)이나 패턴을 발견하는 작업을 말한다.

본 연구에서 사용할 '연관규칙'은 항목의 집합으로 표현된 트랜잭션에서 각 항목간의 연관성을 반영하는 규칙이다. 즉,  $a \rightarrow b$ 의 형태를 갖는 패턴으로서, 이러한 규칙이 갖는 의미는 a 항목 집합이 나타날 때는 b 항목집합도 동반하여 나타나는 경향이 있다는 뜻이다.

연관규칙을 탐사하는 문제는 기본적으로 다음의 두 단계로 구성된다. 미리 결정된 최소 지지도 이상의 트랜잭션 지지도를 가지는 항목집합들의 모든 집합들인 빈발항목 집합들을 찾아내는 단계와 빈발항목 집합의 부분집합 A에 대하여, 만약 Support(A)에 대한 Support(L)의 비율이 적어도 최소 신뢰도 이상이면( $\text{support}(L)/\text{support}(A) \geq \text{MinConf}$ ),  $A \rightarrow (L-A)$ 의 형태의 규칙을 출력하는 단계이다. 이 때, 이 규칙의 지지도는 Support(L)이고 신뢰도는  $\text{Support}(L)/\text{Support}(A)$ 이다. 생성된 연관규칙이 트랜잭션들의 상황을 얼마나 잘 뒷받침해 주는가는, 지지도와 신뢰도의 두 가지의 척도로서 측정한다[7,8].

- 지지도(support degree) : 생성된 연관규칙이 전체 아이템에서 차지하는 비율을 말한다. 즉 데이터베이스에 속한 전체 트랜잭션의 개수 중 그 연관규칙을 지지하는 트랜잭션의 비율을 의미한다.
- 신뢰도(confidence degree) : 연관규칙의 강도를 의미하며 전체부를 만족하는 트랜잭션이 결론부까지를 만족하는 비율을 의미한다.

생성된 연관규칙은 데이터에 내재하는 유형 하나 하나를 매우 상세한 수준까지 보여준다. 또한 유형들을 전체적으로 광범위하게 보여준다고 할 수도 있다. 거시적 관점에서 보면 지지도가 높은 규칙들은 결국 데이터베이스를 요약하는 정보라

고 할 수 있으며, 데이터를 전반적으로 이해하는데 도움을 준다. 미시적 관점에서 보면 지지도가 낮더라도 신뢰도가 상당히 높은 규칙은 그것이 미치는 결과에 따라 충분히 가치를 가질 수 있다. 또한 과거에 비해 지지도가 현저히 낮아진 규칙이 발견되었다면 이는 새로운 변화를 의미하게 된다[4].

### 2.3. 교육분야에서의 자료분석시스템

다음은 교육 현장에서의 교육데이터 분석 현황에 대한 내용으로, 국내의 교육통계시스템과 미국의 교육통계시스템을 비교하여 설명하였다.

#### • 교육통계시스템

국내의 교육통계시스템으로는 국가적 차원에서 구축된 한국교육개발원의 교육통계시스템[11]이 있다. 통계 자료집 수준의 단순 조회 서비스에 머물렀던 기존의 수준에서 벗어나 2002년 10월 확대 개편된 교육통계시스템은 OLAP 기술을 도입하였다. 데이터의 요청과 열람 빈도가 높은 항목에 대해 사용자가 손쉽게 자료를 열람할 수 있도록 미리 설계해 놓은 상세 분석뿐 아니라 교육 통계에 관심이 있는 사람들이 실시간으로 맞춤형 분석을 할 수 있는 기능을 제공한다. 그러나 미리 정의한 항목에 대해서만 분석이 가능하며, 상호 연관 관계나 추론을 할 수 있는 통계 자료는 제시하지 못하고 있다.

#### • 미국국립교육통계센터(NCSE)

미국의 국립교육통계센터에서는 한국의 교육통계시스템에 비해 다양하고 많은 통계 정보를 제공한다. 또한 통계자료를 엑셀 파일, 로터스 파일, PDF 파일 등 다양한 형태로 자료를 내려 받을 수 있도록 하기 때문에 교사들은 이를 이용하여 별도의 자료분석을 할 수 있다. 그러나 학교별 통계나 학생 지도에 필요한 정보를 얻는 단계에는 이르지 못하고 있다[12].

### 3. SIMS: 학생정보마이닝시스템

SIMS는 연관 규칙 기법을 사용하여 다양한 학

생 관련 데이터의 상호 연관 관계를 발견하는 시스템으로써, 교사들은 생성된 연관규칙을 통해 수업 준비나 학생 지도 및 학부모 상담에 필요한 유용한 정보를 얻을 수 있다. 여기서는 SIMS의 유용성과 주요 모듈에 대해 설명하기로 한다.

### 3.1. 유용한 학생 정보 추출 질문 예시

SIMS의 사용자는 교사로 한정하며, 대상 데이터는 현재 일선 학교에서 학사, 교무 업무를 위해 사용하고 있는 교육행정시스템에 등록되어 있는 자료를 사용하는 것으로 한다. SIMS를 사용하여 실제 학교 현장에서 교사들이 유용한 정보를 제공받을 수 있는 질문들을 대략적으로 열거해 보면 다음과 같다.

#### (1) 학생들의 성적 자료 관련 질문

- 전체 성적 상승률과 과목 성적과의 관계
- 인문계, 자연계 선택과 각 과목 점수와의 관계
- 시험문제 문항 분석을 통한 성별, 학년별, 반별 학업성취도 분석, 과목별 취약부분 분석
- 지도교사와 각 과목/단원별 학업성취도와의 관계

#### (2) 학생들의 생활, 진로 지도를 위한 질문

- 적성검사 결과와 성별, 각 과목 점수와의 관계
- 부모의 직업, 소득 수준과 학생의 성적, 생활과의 관계
- 성격, 가족사항, 성격 등과 교우관계와의 연관성
- 학생생활기록부의 각종 자료와 진로 선택과의 관계

위에서 제시한 질문 중에서 시험 문제 문항 분석을 통한 과목별 취약 부분에 대한 분석은 학생과 교사 모두에게 유용한 정보를 제공할 수 있다. 학생은 특정 과목에서 어떤 단원에 대한 이해가 부족하며, 현재 풀 수 있는 문제의 난이도는 어떤 정도인가에 대한 분석 자료를 제공받을 수 있고, 교사도 지도 내용 중 어느 부분에 보충이 필요할 것인지 예측할 수 있다. 시험 문제가 관련 단원, 난이도, 평가 영역 등의 항목들과 더불어 데이터베이스에 저장된다면, 수준별 학습을 제공하는 문제 은행의 역할도 할 수 있다.

교우 관계에 대한 분석은 집단 따돌림 현상이 사회 문제화된 현재의 시점에서 매우 유용한 질

문이라 할 수 있다. 학생의 성격, 성적, 가족 사항, 취미 활동, 학급 내 봉사 활동, 부모의 소득 수준 등 다양한 자료를 대상으로 데이터 마이닝을 수행한다면 유용한 결과를 얻을 수 있다. 이를 위해서는 학생의 교우 관계에 대한 정확한 정보가 사전에 입력되어 있어야 한다. 집단 따돌림 현상은 일부 학생이 당하는 문제이므로 교사들이 생성된 연관 규칙을 검토할 때 빈도수가 낮더라도 정확도가 높은 규칙들에 특히 관심을 가져야 한다.

진로에 관한 예측은 부모의 학력과 직업, 소득 수준, 거주 지역, 교사의 태도, 학생의 건강 상태, 신체 조건, 흥미도, 적성검사 결과, 장래 희망, 지능지수, 특기 및 취미 활동, 행동 발달 사항, 성적 등의 자료를 토대로 유추해 낼 수가 있다[1]. 졸업생의 직업에 대한 자료까지 저장되어 있다면 졸업생과 유사한 성향, 적성, 관심을 가진 재학생에게 졸업생의 진로, 직업, 진학, 취업 등의 정보를 분석하여 제공함으로써 효과적인 진로 지도를 할 수 있게 될 것이다.

위와 같은 분석을 통해 다양한 학생에 대한 성적과 학습 문제에 대한 문제 패턴을 적용하여 효율적인 수준별 학습 방법을 추론할 수 있으며, 범죄 및 폭력에 관한 예방, 학생의 성향을 토대로 한 생활 태도 관찰 및 집단 따돌림 예방, 진단, 건강 예측 및 진단, 학생의 장래 직업 및 진로에 대한 예측, 학부모의 관심 및 성향 등을 추론할 수 있다[1].

### 3.2. 연관규칙 추출

SIMS는 연관규칙 기법 적용을 위해 크게 다음과 같이 세 개의 모듈을 포함한다. 첫째 모듈은 서버에 있는 데이터베이스로부터 데이터를 가져오는 부분이다. 사용자가 분석하고자 하는 데이터와 조건을 편리한 인터페이스를 통해 제시하면 이에 해당하는 SQL문을 자동 생성하여 해당 데이터를 서버로부터 클라이언트측에 로드시킨다.

둘째, 클라이언트측에 저장된 데이터를 분석이 용이하도록 변환하는 부분이다. 0~100 사이의 모든 값을 가질 수 있는 성적과 같은 데이터는 적절히 그룹화하여 도메인 값을 축소시킨다. 연

관 규칙의 개수는 도메인 값의 개수에 비례하므로 사용자는 도메인 값의 범위를 조절하여 규칙의 개수를 조절해야 한다. SIMS의 두 번째 단계에서 이러한 작업이 이루어진다.

셋째, 연관규칙 기법이 적용되어 데이터 사이의 연관 관계를 추출하는 부분이다. 기본적인 연관규칙 알고리즘인 Apriori[7]를 사용한다. Apriori 알고리즘은 빈발항목집합을 찾을 때 빈발항목의 부분집합 역시 빈발항목에 속한다'의 개념을 바탕으로 하여 알고리즘의 성능을 높인다.

연관규칙 발견 과정을 설명하기 위해 <그림 1>과 <그림 2>를 살펴보기로 하자. <그림 1>은 분석할 대상 데이터베이스이며, 학생의 학년, 성별, 전체 성적, 성격, 교우관계로 이루어져 있다. 이로부터 <그림 2>의  $L_2$ 가 추출된다.  $L_2$ 가 나타내는 것은 성적이 좋은 학생의 성격이 외향적일 경우는 40% 발생하고, 성적이 좋으면 교우관계가 좋을 경우는 40% 발생하며, 외향적인 성격의 학생이 교우관계가 좋을 경우는 60% 발생한다는 연관관계를 나타낸다.

$L_2$ 의 추출과정은 <그림 2>에 자세히 나타나 있다. 각 항목에 대한 빈도 수를 계산하기 위해 <그림 1>의 데이터베이스를 검색하고 1-항목 집합들의 빈도 수를 계산하여 후보 1-항목 집합들의 집합  $C_1$ 이 얻어진다. 최소 지지도를 40%로 하면 최소 지지도를 넘는 후보들의 1-항목 집합들로 구성된 빈발 1-항목 집합들  $L_1$ 이 결정된다. 빈발 2-항목 집합들의 집합  $L_2$ 를 찾을 때 빈발 집합 항목들의 모든 부분 집합들의 모든 부분 집합들이 최소 지지도 이상을 가져야 한다는 사실에 기반하여  $C_2$ 를 생성하기 위해  $L_1 \times L_1$  ( $L_1 \times L_1 = \{(X, Y) | X \in L_1 \text{ and } Y \in L_1 \text{ and } X \neq Y\}$ )를 사용한다. 다음으로 데이터베이스 내의 3개의 트랜잭션을 읽고  $C_2$ 의 각 후보 항목 집합들의 지지도를 계산한다. 빈발 2-항목 집합들의 집합  $L_2$ 는  $C_2$ 의 후보 2-항목 집합의 지지도를 기반으로 결정된다.

TID	학년	성별	전체 성적	성격	교우관계
1	1	M	A	외향	상
2	1	M	A	외향	상
3	1	M	B	내향	중
4	1	M	B	내향	상
5	1	M	C	외향	상
6	1	M	A	외향	상
7	1	M	C	내향	상
8	1	M	B	외향	상
9	1	M	A	외향	상
10	1	M	C	외향	하

<그림 1> 대상 데이터베이스

<그림 2> 후보항목집합과 빈발항목집합의 생성

## 4. SIMS의 설계 및 구현

### 4.1. SIMS의 주요 기능

SIMS의 전체 흐름이 <그림 3>에 도식화되어 있다. 각 단계는 여러 모듈로 세분화 될 수 있으며, 상이한 단계들 사이를 반복적으로 수행하는 것이 일반적이다. 연관규칙 탐사를 수행할 요구사항이 정의되면 그에 적합한 데이터를 선정 단계와 변환, 데이터 마이닝 단계를 거쳐 생성된 연관 규칙을 조회하고 그 중에서 교사들끼리 공유할 가치가 있는 규칙들은 다시 서버에 저장되게 된다. 이렇게 저장된 유용한 연관규칙을 통해 교사들은 새로운 학생 관리에 필요한 정보를 얻게 된다.

자료 선정 단계는 연관 규칙 탐사를 위한 기초 데이터를 선택하는 부분으로써 동적으로 SQL문을 생성하여 실행한다. 전산 지식이 없는 교사들을 위하여 SQL문을 직접 생성하고, 자료 테이블명의 한글화, 컬럼명의 한글화, WHERE 조건을

정의하기 위한 도움 기능이 포함되어 있다. 이는 보안을 필요로 하므로 관리자용 프로그램에 포함되어 있다.

써 교사들이 정보를 공유할 수 있도록 하였다. <그림 4>에는 자료 선정 단계에서부터 데이터마이닝 단계까지의 모듈별 입출력을 명시하였다.

### <그림 3> SIMS에서의 데이터 마이닝 과정

교사들은 한글화된 테이블명, 쿼리명을 통해 자기가 분석하기 원하는 대상 자료를 선택할 수 있으며, 전산 담당 부서에서 미리 정의해놓은 조건식을 참고로 하여 자료 선택을 위한 조건을 입력하게 된다. 해당 SQL문이 자동으로 생성되어 서버에 있는 교육행정시스템의 데이터베이스로부터 자료를 가져오게 된다.

서버에서 가져온 자료가 클라이언트의 임시 테이블에 저장되고 나면 유용한 분석 결과를 추출하기 위해 변환 단계를 거쳐야 한다. SIMS에서는 자료의 분포를 파악하여 자동으로 구간을 설정해 주는 자동 분류와 일정 간격 또는 자료의 건수 별로 구간을 정의하는 수동 분류 방식을 제공한다.

변환 단계가 끝나면 연관규칙 기법 적용을 위해 지지도와 신뢰도를 정의하게 된다. 연관 규칙에 포함될 최대 조건 수를 정의하여 지나치게 복잡한 연관 관계가 추출되는 것을 피하도록 필터링 기능을 포함시켰다.

마지막으로 유용하다고 판단되는 규칙들을 서버에 저장하여 교사들이 공유할 수 있도록 하는 기능을 제공한다. 예를 들어, 어떤 학급이 담임 교사가 집단 따돌림 현상에 대한 마이닝 결과 중 정확도가 높은 규칙을 발견했다면, 이러한 정보는 모든 교사들이 공유하는 것이 바람직하다. SIMS는 데이터 마이닝 결과에 대한 제목, 상세 설명 등 추가 정보를 입력하여 웹에 게시함으로써

### <그림 4> SIMS에서의 각 단계별 입출력 내용

#### 4.2. 교사용 프로그램

교사용 프로그램은 관리자용 프로그램을 통해 테이블명 등록, 쿼리명 등록, 조건식 예시 등록 작업이 선행된 상태에서 사용되며, 유용한 마이닝 규칙을 학교의 홈페이지를 통해 공유하도록 하였다. 관리자용 프로그램 중 테이블명 등록과 쿼리명 등록은 각각 테이블명과 쿼리명을 한글화하여 교사들의 이해를 돕기 위한 프로그램이며, 조건식 입력 프로그램은 자주 사용하는 조건식을 미리 입력해 놓음으로써 교사들이 조건식을 입력할 때 참고로 하기 위한 것을 목적으로 하는 프로그램이다. 다음에서는 창의성 검사와 학년, 성별, 과목별 성적과의 연관 관계를 찾는 과정을 예로 교사용 프로그램을 설명하기로 한다.

##### (1) 분석 대상 테이블 선택

연관규칙 탐사를 위한 첫 단계는 <그림 5>와 같이 마이닝 대상이 되는 분석 대상 테이블 선택이다. 테이블을 미리 선택하는 이유는 다음 단계인 분석 대상 쿼리 선택에서 지나치게 많은 쿼리 자료가 나오지 않도록 하기 위한 것이다.

하게 된다. 성적과 같이 여러 값이 나올 수 있는 컬럼은 일정구간으로 분할하여 마이닝을 수행하게 된다. <그림 8>에 예제가 있다. '자동분류'를 선택하면 시스템에서 분포를 고려해서 일정 구간을 정하게 되고, '수동분류'를 선택하면 '일정간격으로 분류'나 '개수로 분류'를 선택할 수 있다. '일정간격으로 분류'는 점수의 경우 5점 간격, 10점 간격과 같은 식으로 분류가 되는 것을 말하며, '개수로 분류'는 10건씩 또는 20건씩과 같은 식으로 묶여지게 되는 것을 말한다.

<그림 5> 분석대상테이블 선택

(2) 분석 대상 컬럼 선택 화면

분석 대상 컬럼 선택 프로그램이 실행되면 앞의 분석 대상 테이블 선택 프로그램에서 선택된 각 테이블의 컬럼들이 모두 조회된다. <그림 6>과 같은 분석대상 컬럼 선택 화면에서는 분석대상자료, 분석결과자료를 선택하게 된다. 아래의 예시는 창의성 검사의 각 항목별 점수가 학년, 성별, 과목별 성적과의 연관관계를 알아보기 위한 화면이다. 이때 검사항목과 검사지수가 분석결과가 되며, 학년, 성별, 각 과목 성적이 분석대상이 된다.

<그림 7> 조건 입력

<그림 6> 분석대상컬럼 선택

(3) 조건 입력

<그림 7>과 같이 조건 입력 프로그램을 실행하면 이미 관리자가 등록한 조건식 예시가 보이며, 추가, 수정, 삭제를 통해 조건식 입력을 하게 된다.

<그림 8> 자료분류

(4) 자료 분류

분석대상자료 선택과 조건 입력이 끝나면 마이닝 작업의 효율을 높이기 위해 코딩 작업을 수행

(5) 데이터 마이닝 수행

<그림 9>는 연관규칙 탐사를 위한 파라미터 설정 화면이다. 신뢰도와 지지도는 구간으로 입력할 수 있으며, 규칙에 포함되는 조건 수를 지정함으로써 지나치게 복잡한 규칙의 생성을 제한하고 시스템의 효율을 높이고자 하였다.

문에 대해서는 통계 자료로 쌓아서 정보의 변화를 볼 수도 있게 된다.

<그림 9> 데이터마이닝 수행

(6) 마이닝 결과 조회 및 저장

연관 규칙 탐사 결과는 신뢰도, 지지도와 함께 <그림 10>과 같이 리스트 형태로 제공된다. 규칙이 많을 경우에는 스크롤바를 통해서 조회해야 하는데, 이때 특정 분석대상별로 필터링을 할 수 있다. 이 때에는 하단에 있는 분석대상 데이터 분포에 그래프로 볼 수 있다.

<그림 11> 추가정보 등록

<그림 12> 웹에서의 마이닝 결과 조회

### 5. 결 론

SIMS는 데이터 마이닝 기법을 적용하여 대량의 학생 데이터에서 자료간의 상호 관련성과 유용한 패턴을 추출하여 교육에 활용할 수 있는 정보를 제공하기 위한 도구이며, 조작하기 쉬운 인터페이스 설계로 컴퓨터에 익숙하지 못한 교사들이 쉽게 접근할 수 있도록 하였다.

본 논문의 의의는 다음과 같이 정리할 수 있다. 첫째, 데이터 마이닝 기술의 새로운 시도이다. 대부분의 데이터 마이닝은 은행, 기업 등에서 고객 관리나 마케팅을 위해 사용되어지고 있는데, 이러한 데이터 마이닝 기법을 교육 정보를 처리하는데 이용해보므로써 데이터 마이닝 적용영역을 확대시키는 계기가 될 것으로 기대한다.

둘째, 다양한 연관규칙 탐사 결과는 교사들의 교과 수업 연구나 학생 생활지도 및 학부모 상담

<그림 10> 마이닝 결과 조회

생성된 연관규칙 중 중 교사가 적절하다고 생각하는 규칙은 <그림 11>과 같이 설명을 추가하여 다시 저장할 수 있다. 이 때에는 자료가 다시 학교 서버에 저장되며, 학교 홈페이지나 별도의 프로그램을 통해서 교사들끼리 마이닝 결과를 공유할 수 있게 된다.

(7) 웹에서의 마이닝 결과 조회

각 교사들이 저장한 연관 규칙을 가지고 학교의 홈페이지나 별도의 프로그램을 통해 조회할 수 있다. 개별적인 데이터 마이닝 결과를 공유하는 차원에서 벗어나 지속적인 관심을 가지는 질



에 도움을 줄 수 있다. 교과 담당 교사는 해당 교과별로 자료를 분석함으로써 학생들의 학업 이해도, 수업 형태, 교과 난이도, 타과목과의 상호 관련성 등을 파악하여 교사의 수업 계획이나 연구에 도움을 얻을 수 있다. 교사뿐 아니라 학교 행정가도 전체 학생에 대한 분석을 통해 7차 교육과정의 특징 중 하나인 선택과목 선정 등 각종 정책 수립에 데이터 마이닝 결과를 활용할 수 있으며, 더 나아가 국가적 차원의 교육 정책 수립에도 기여할 수 있다.

셋째, 교사들이 분석한 유용한 연관규칙 결과를 학교의 통합 데이터베이스에 저장함으로써 교사들끼리 정보를 공유할 수 있으며, 의견 교환을 통해 공통의 경험과 지식을 축적할 수 있다.

마지막으로 쉽게 조작할 수 있는 인터페이스의 구현으로 SIMS는 학교 현장에서 사용하고 있는 교육행정시스템과 연계되어 충분히 활용될 수 있으리라 기대한다.

### 참 고 문 헌

[1] 이철환, 한선관, “교육에서의 효율적인 정보 활용을 위한 데이터 마이닝 기법”, 한국정보처리학회 논문지 3권 1호, 1999.

[2] 강신천, “ICT교육 활성화를 위한 기반 구축 방향”, 새교육, 2002년 통권 574호.

[3] 용환승 역, “데이터마이닝”, 그린, 1998.

[4] A. Berson, et. al., “CRM을 위한 데이터마이닝”, 2000.

[5] 한국교육학술정보원, 교무업무지원시스템 고등학교 사용자 지침서, 2000

[6] 김민정, “데이터마이닝 기법을 이용한 여행사 고객관계관리”, 이화여자대학교 과학기술대학원, 2001

[7] R. Agrawal and R. Srikant, “Fast algorithms for mining association rules”, In Proceedings of the 20th VLDB Conference, Santiago, Chile, Sept, 1994.

[8] Groth, R., “Data Mining:Building Compleitive Advantage”, Prentice Hall PTR, 1999

[9] <http://sims.keris.or.kr/>  
(초중등학교종합정보관리시스템 홈페이지)

[10] <http://kdnuggets.com/software/>

[11] <http://www.kedi.re.kr/>(한국교육개발원)

[12] <http://www.nces.ed.gov>  
(미국국립교육통계센터)

### 공 현 선

1995 이화여자대학교  
전자계산학과 (이학사)  
1999~현재 이화여자대학교  
교육대학원 컴퓨터교육전공  
1995~2001 동양 시스템즈  
2001~현재 이화여자대학교 정보통신처  
사무직원  
관심분야: 컴퓨터교육, 데이터마이닝  
E-Mail: khappy@ewha.ac.kr

### 김 명

1981 이화여자대학교 수학과  
(이학사)  
1983 서울대학교 계산통계학과  
(석사)  
1993 미국 캘리포니아 주립대학교  
(산타바바라)  
컴퓨터학 (박사)  
1993~1994 미국 캘리포니아 주립대학교  
(산타바바라) Postdoc.  
1995~현재 이화여자대학교 부교수  
관심분야: 지식공학, OLAP, 컴퓨터 교육,  
고성능처리  
E-Mail: mkim@ewha.ac.kr