

데이터 마이닝을 이용한 시험 응답데이터 분석시스템 설계 및 구현

곽은영[†] · 김현철^{††}

본 논문은 데이터 마이닝 기법 중 연관규칙 추출 알고리즘을 이용하여 시험 응답 데이터에서 서로 연관된 문항들을 찾아내고, 그 원인을 규명함으로써 교육평가에서 사용되고 있는 기존의 검사이론 기반의 분석 결과와 함께 사용되면 문항의 질뿐만 아니라 피험자의 성취 수준을 심층적으로 분석하는데 도움을 줄 수 있는 시험 응답데이터 분석시스템을 개발하고 구현하는데 연구의 목적이 있다. 현재의 교육평가 분야에서 문항 분석에 사용되는 고전검사 이론과 문항반응 이론은 각 문항의 독립성을 전제로 하고, 피험자들이 각 개별 문항에 반응하여 나타나는 결과를 통계적 수치를 이용하여 설명하고 있다. 그러나 실제 학교 현장에서 실시한 시험의 결과를 보면, 피험자들의 반응에 의하여 문항간 연관성이 발생하게 되며 이러한 연관성은 각각의 문항들을 분석하고 피험자의 능력을 추정하는 데 의미 있는 영향을 미치게 된다. 제안된 시스템은 연관규칙 마이닝을 이용하여 흥미로운 문항간 연관성을 추출하고, 그 원인을 분석하여 사용자에게 제공함으로써 교수-학습 방법 개선이나 문제은행의 질을 향상시키는데 도움을 줄 수 있도록 하였다.

키워드 : 응답데이터 분석시스템, 문제은행, 문항분석, 평가시스템, 연관규칙 마이닝,

Design and Implementation of Analysis System for Answer Dataset with Data Mining

Eun-Young Kwak[†] · Hyeoncheol Kim^{††}

ABSTRACT

In this paper, we introduce an analysis system for answer dataset by using a data mining method. We analyze students' answer data collected from a test including multiple choice question items, and find associations between the items. Analysis of evaluation results based on our system will not only provide correct information on students' achievement levels but also provides a basis for modifying weaknesses of the evaluation procedures, question items, or teaching/learning procedures. Furthermore, it will enable us to improve the quality of question items for future use so that we can secure itemsets of high quality

Keywords : Item Analysis, Data Mining, Association Rule

1. 서 론

교육 현장에서 평가는 학습자들의 학습 성취도

를 분석하는 마지막 단계이면서 동시에 분석된 학습 성취도 결과를 다음 학습과 평가를 위해 활용함으로써 새로운 학습 과정의 시작 단계로서의 역할을 하게 된다. 학습 성취도의 평가는 일반적으로 시험이라는 검사 도구를 통해 이루어지며 학습자들의 시험 응답 데이터를 대상으로 다양한

[†] 정 회 원: 고려대학교 컴퓨터교육과 박사과정(교신저자)
^{††} 중 신 회 원: 고려대학교 컴퓨터교육과 교수
논문접수: 2007년 10월 23일, 심사완료: 2007년 11월 16일

측정치들을 구하고 결과를 분석하게 된다.

시험 응답 데이터를 분석하여 학습자의 성취도 뿐만 아니라 시험지를 구성하고 있는 각 문항에 대한 학습자의 반응을 통해 문항의 질도 평가하게 된다. 이를 위하여 교육학 분야에서 일반적으로 사용하고 있는 방법은 고전검사이론과 문항반응이론에 입각한 다양한 평가시스템을 이용하는 것이다. 고전검사이론은 총점에 의하여 분석되는 이론으로서 문항 난이도, 문항 변별도, 문항 추측도를 추정하게 된다. 고전검사이론에서는 측정오차가 피험자 집단의 성질에 관계없이 동일하다고 가정하고 있으나 실제로 측정오차가 피험자들의 능력수준에 반비례하기 때문에 그 가정이 충족되기 어렵다. 고전검사이론의 문제점을 극복하기 위해 제안된 검사이론이 바로 문항반응이론이다. 문항반응이론은 피험자의 검사 결과에 영향을 미치는 관찰할 수 없는 잠재적 특성이 있다고 가정하고 피험자의 검사 점수로부터 잠재적 특성을 추정하는 절차와 관련된 이론이다. 문항반응이론은 피험자의 잠재된 능력수준과 문항에 대한 반응의 관계를 수학적으로 나타내며, 피험자의 능력에 따른 문항의 답을 맞힐 확률을 나타내는 문항특성곡선에 기초한다. 이 두 이론은 어떤 능력을 가진 피험자의 하나의 문항에 대한 응답은 다른 문항의 응답에 전혀 영향을 주지 않는다는 지역독립성을 기본 가정으로 하고 있다. 지역독립성의 가정을 통계적으로 말하면, 피험자가 어떤 문항의 답을 맞힐 확률과 다른 문항의 답을 맞힐 확률이 상호 독립적이어야 함을 뜻하고, 이는 한 문항의 내용이 다른 문항의 정답의 단서가 되지 않아야 한다는 의미를 포함한다[2][6]. 그러나 현실적으로는 이것을 위배하고 한 문항에서의 수행이 다른 문항의 수행에 영향을 미치는 지역의존성이 발생하게 된다. 일반적으로 교육평가 측면에서 검사 도구 개발자들은 지역의존성을 제거하기 보다는 지역의존성을 발생시키는 요인들을 적절히 조절하여 지역의존성이 발생하지 않도록 하는데 초점을 두고 있다.

그러나 학교 현장에서 검사가 실시될 때, 교사가 의도하지 않았던 문항간 연관성이 발생할 수 있고, 어떤 문항들끼리 연관성이 발생하였는지, 그 연관성의 원인이 무엇인지를 파악함으로써 각

각의 문항들과 피험자의 능력을 보다 더 정확하게 측정할 수 있다. 이를 위하여 피험자의 응답 결과에 대한 정·오답 분석 시 문항과 문항간의 관계를 분석할 수 있는 도구가 필요한데, 이에 본 연구에서는 데이터 마이닝 기법중 연관규칙 마이닝을 사용하여 선다형 문항으로 이루어진 평가 검사에 대한 학습자들의 문항 응답 데이터를 분석하고, 문항간 연관성을 추출하고자 한다. 일반적으로 연관규칙은 지지도(support) 값을 기반으로 빈발 항목을 찾아내어 추출되어진다. 그러나 단순히 빈발 정도로만 발견된 연관규칙들이 모두 의미 있는 것이 아니기 때문에 흥미로운 연관규칙을 발견하는 것이 중요하다. 흥미도란 연관규칙의 의미성을 나타내는 것으로 강한 빈발 정도에 의하여 생성된 규칙이 존재한다 하더라도 그것이 모두 의미 있는 연관성을 의미하는 것은 아니기 때문에 규칙의 의미성을 판단하는 절차가 필요한 것이다. 특히 교육용 데이터인 문항 응답 데이터를 다루는 본 연구에서는 각 문항이 평가하고자 하는 내용 간의 연관성이 발생하기 때문에 단순히 빈발정도에 의하여 발생한 연관성은 더욱 무의미하게 될 것이다.

본 연구에서 제안하는 시스템은 한 개 시험지에 구성된 각 문항(item)에 대한 응답 데이터를 분석하기 위하여 연관규칙 알고리즘을 기반으로 하여 연관규칙을 생성하며, 이 때 무의미한 연관규칙의 발생을 감소시키기 위하여 우선적으로 불필요한 응답 데이터들을 제거하고 연관규칙을 추출하도록 하였다. 추출된 연관규칙은 정해진 기준에 의해 원인이 분석되고 사용자에게 제시되며, 사용자는 이러한 분석 결과를 문제은행에 반영할 수도 있고, 교수 방법을 개선하는데 활용할 수도 있다. 그러므로 기존의 검사이론 기반의 문항분석 시스템에서 제공하는 결과와 함께 사용하면 다각적인 분석을 하는데 도움이 될 것이다.

본 연구의 시험 응답데이터 분석시스템은 문항 분석모듈과 문제은행모듈로 구성되었으며, 문제은행을 구축하기 위하여 인문계 고등학교 정보사회와 컴퓨터 교과와 컴퓨터 교과의 내용을 문항으로 구성하였다. 시스템에서 제공하는 분석 결과의 타당성과 현장에서의 활용성은 컴퓨터 교과를 담당하고 있는 현직 교사들을 대상으로 설문조사를 통하여 검증

하였다.

2. 이론적 배경 및 관련연구

2.1 문항분석이론

학습 능력은 직접 측정이 불가능하므로 검사(test)라는 도구를 사용하여 측정하며, 검사를 실시한 결과를 평균, 총점, 표준편차, 상관분석 등의 통계적 방법을 이용하여 분석함으로써 학습자의 능력을 평가하게 된다. 따라서 평가가 타당한지의 여부가 중요하고, 평가의 타당성을 위하여 검사의 질이 양호한지를 우선적으로 분석하여야 하며 이를 위하여 검사를 구성하고 있는 문항을 분석하게 된다. 교육평가에서 문항을 작성할 때 개개 문항이 제대로 기능을 수행하고 있는지를 확인하는 작업을 문항분석이라고 한다[7].

문항의 질과 피험자들의 능력을 평가하기 위하여 시험 응답 자료를 대상으로 문항을 분석하는 수리적 통계 모형 기반의 검사이론으로서 고전검사이론과 문항반응이론이 있다.

고전검사이론은 총점에 의하여 분석되는 이론으로서 문항난이도, 문항변별도, 문항추측도를 계산, 추정하게 된다[6]. 고전검사이론은 검사점수의 해석과 검사의 양호도를 점검할 때 기본모형으로 이용되지만, 피험자의 능력에 따라 문항모수가 검사문항에 따라 피험자의 능력이 달리 추정되는 단점이 있다. 이 이론에 의하여 문항을 분석하였을 경우, 같은 문항이라도 능력이 높은 학생들이 응답한 자료를 가지고 분석하면 쉬운 문항으로, 능력이 낮은 학생들이 응답한 자료를 가지고 분석하면 어려운 문항으로 분석되므로 문항특성의 불변성 개념을 지지하지 못한다. 또한 검사의 총점으로 능력을 추정하므로 검사가 쉬우면 학생의 능력은 과대 추정되고, 검사가 어려우면 능력이 과소 추정되므로 피험자 능력 불변성 개념을 유지하지 못하는 단점을 지니고 있다[6].

문항반응이론은 피험자의 검사 결과에 영향을 미치는 관찰할 수 없는 잠재적 특성이 있다고 가정하고, 피험자의 잠재된 능력수준과 문항에 대한 반응의 관계를 수학적으로 나타내며, 피험자의 능

력에 따른 문항의 답을 맞힐 확률을 나타내는 문항특성곡선에 기초한 검사이론이다. 문항반응이론에서 사용되는 문항모수는 문항난이도, 문항변별도, 문항추측도가 있고, 문항마다 각기 다른 고유한 곡선을 나타내는 문항특성곡선이 있다. 문항특성곡선(item characteristic curve)이란 문항의 고유한 속성을 나타내는 것으로, 학습자 능력에 따라 문항의 답을 맞힐 확률을 나타내는 곡선으로 일반적으로 S자 형태를 나타낸다[8].

문항반응이론에서는 어떤 능력을 가진 피험자의 하나의 문항에 대한 응답은 다른 문항의 응답에 전혀 영향을 주지 않는다는 지역독립성(Local Independence)을 기본 가정으로 하고 있고 고전검사이론에서도 이것을 잠재적으로 가정하는데, 이것을 위배하고 한 문항에서의 수행이 다른 문항의 수행에 영향을 미치는 지역의존성(Local Dependence)이 발생하게 된다[8][9]. 피험자가 어떤 문항의 답을 맞힐 확률과 다른 문항의 답을 맞힐 확률이 상호 독립적이어야 함을 뜻하고, 이는 한 문항의 내용이 다른 문항의 정답의 단서가 되지 않아야 한다는 의미를 포함한다. 반면, 지역의존성은 어떤 문항에 대한 피험자의 응답이 다음 문항의 수행에 영향을 끼칠 때 발생하는 것으로, 문항 점수의 조건적 상관(the conditional correlation of item scores)으로 말할 수 있다[9]. 문항의 지역독립성과 의존성은 상황에 따라 그 중요성이 달라진다. 교수-학습 상황에서는 한 문항에서의 수행이 다른 문항에 대한 수행의 안내 역할을 함으로써 긍정적인 효과로 작용할 수 있다. 반면, 평가를 실시할 때는 평가의 단위가 명확히 구분되어야 각 단위에 대한 평가를 정확히 실시할 수 있다[13][17].

지역의존성을 측정하는 일반적인 방법으로 문항간 상관계수 추정, 요인분석, Q₃ 치수 측정 등의 통계치를 구하는 방법이 있고, 다분문항반응이론에 근거한 측정모형을 사용하는 방법이 있다. 지역의존성의 원인과 측정방법에 대한 많은 연구가 계속해서 이루어지고 있는데, 대부분의 연구에서는 지역의존성이 발생하지 않도록 변인들을 조절하는 데 초점을 두고 있다[1][4][5][13][14][17][19].

2.2 연관규칙 마이닝

연관규칙 마이닝은 빈발항목집합을 생성하고 연관규칙을 발견하는 것으로서 빈발항목집합은 특정 트랜잭션(transaction)에서 자주 동시에 발생하는 두개의 항목들의 집합으로서 지지도 값으로 구하게 된다. 강한 빈발항목집합으로 추출된 연관규칙이라도 모두 흥미로운 것이 아니기 때문에 다양한 흥미도 측정 척도들을 사용하여 흥미로운 연관규칙을 발견하는 것이 중요하다[15]. 그러나 원본 데이터집합의 데이터들이 쓸데없는 데이터들로 구성되어 있다면 생성된 빈발항목집합은 더욱 무의미하고 연관규칙의 흥미도는 떨어지게 됨으로 결국 연관규칙의 질은 낮아지게 된다. 특히 트랜잭션과 어트리뷰트(attribute)가 서열화 되는 특정 속성으로 인해 발생하는 빈발항목집합으로부터 추출된 연관규칙의 흥미도는 더욱 그러하다. 예를 들면 본 연구의 대상인 시험 응답데이터의 경우, 각 트랜잭션에 해당하는 학생 데이터는 고득점에서부터 저득점까지 구분되고, 각 어트리뷰트에 해당하는 문항들은 어려운 문항부터 쉬운 문항으로 서열화 되는 속성이 있다. 여기서 고득점의 학생이 어려운 문항을 맞거나 저득점의 학생이 쉬운 문항을 맞히는 것은 당연한 일이다. 그러므로 추출된 연관규칙 $X \rightarrow Y$ 에서 X 는 어려운 문항, Y 는 쉬운 문항이라면, 어려운 문항을 맞힌 학생들의 대부분이 쉬운 문항을 맞힌다는 관계가 성립되며 이러한 사실은 너무나 당연하기 때문에 우리에게 흥미로운 정보가 될 수 없다. 그러나 고득점의 학생이 쉬운 문항을 틀리거나 저득점의 학생이 어려운 문항을 맞히는 사실은 매우 드문 일이며, 이러한 사실로부터 추출된 연관규칙은 우리에게 흥미로운 정보를 제공하게 된다.

연관규칙 탐색은 크게 두 단계로 나뉜다. 첫 번째 단계에서는 미리 정해진 최소지지도 이상을 만족하는 빈발항목집합을 구하고, 두 번째 단계에서는 빈발항목집합을 이용하여 미리 정해진 최소 신뢰도(confidence)를 만족하는 연관규칙을 추출하는 것이다. 연관규칙 마이닝의 성능은 주로 첫 번째 단계에서 결정되는데, 전체 데이터베이스에서 빈발항목집합을 얼마나 효율적으로 찾아내는

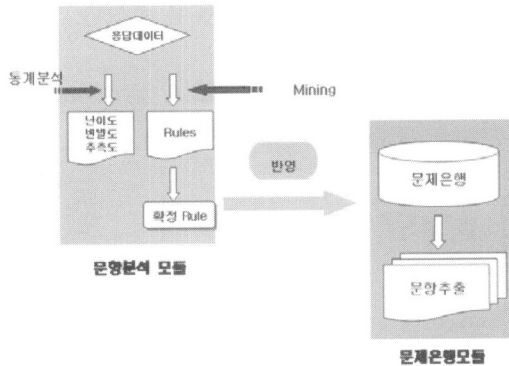
나가 이미 찾아진 목록을 통해 연관규칙을 추출하는 것보다 어렵기 때문이다. 대부분의 빈발항목집합 발견 알고리즘의 기본은 "Apriori"에 근간을 두고 있는데, 방법은 초기 빈발항목집합을 발견한 후, 그것을 바탕으로 다음 단계의 빈발항목집합을 찾는 과정을 반복 한다[11][12][18]. Apriori 알고리즘이 연관규칙 마이닝을 위한 가장 대표적인 알고리즘이지만 알고리즘 수행에 있어 빈발항목집합을 찾기 위해 생성해야할 후보항목집합의 수가 많고, 여러 번의 데이터베이스 스캔이 필요하다는 문제점이 있다. 이를 개선하기 위하여 DHP(Direct Hasing & Prunning), FP-tree, Common-Item Tree, CLOSET 등과 같은 알고리즘이 제시되었다[12][16][18].

이상으로 살펴본 내용을 정리하면, 교육 평가 분야에서는 기본적으로 문항간 연관성이 발생하지 않도록 검사 도구를 작성하는 것을 중요시 하지만, 실제 학교 현장에서 실시한 시험의 응답데이터를 분석하면 의도하지 않았던 문항간 연관성이 발생하게 되고, 이러한 결과는 피험자의 능력과 문항의 질을 평가하는 데 중요한 영향을 미치게 된다는 것을 알 수 있다. 그러므로 본 연구에서는 연관규칙 마이닝을 이용하여 응답 데이터를 분석하는 시스템을 제안하게 되었다.

3. 시스템 설계

본 연구에서 제안하는 시험 응답데이터 분석시스템은 <그림 1>과 같이 문항분석모듈과 문제은행모듈로 구성되어 있다. 문항분석모듈은 각 문항을 난이도, 변별도, 추측도의 통계치로 요약하고, 문항간 연관성과 연관성이 발생하는 원인을 제공하도록 하였다. 각 문항의 난이도, 변별도, 추측도 값은 고정되어 있는 것이 아니라 응답데이터의 정답률에 의하여 조정되도록 하였다. 문항간 연관성이 발생하는 원인은 두 문항이 정답에 대한 단서로 작용했는지 또는 해당 도메인에 대한 학습 성취가 잘 되었는지의 두가지 측면에서 분석되도록 하였다. 문제은행모듈은 텍스트 기반의 문제를 등록, 수정, 삭제하는 등의 기본 기능을 수행하며 문제은행의 문항들은 인문계 고등학교 정보사회와 컴퓨터 교과와 내용으로 구성하였다.

본 시스템은 현재 학교 현장에서 가장 많이 실시되는 평가 측정 형태인 지필고사를 기본 환경으로 전제하고 한 개의 검사 도구에 구성된 문항들을 분석할 수 있도록 하였다.



<그림 1> 문항분석시스템의 구성

3.1 문항분석 모듈

3.1.1 통계치 측정

문항에 대한 기본 특성인 난이도, 변별도, 추측도를 다음의 공식으로 구하였다.

- 난이도= 문항 정답자수/ 총피험자수
- 변별도 = $\frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\sum X)^2} \sqrt{N\sum Y^2 - (\sum Y)^2}}$
- 추측도=추측으로 맞힌 학생수/ 총피험자수

3.1.2 문항간 연관성 분석

Apriori 알고리즘을 이용하여 연관규칙을 추출하기 전에 응답 데이터 집합을 구성하고 있는 학생(transaction)과 문항(attribute) 데이터의 정보량 θ 값을 다음과 같이 구한다.[3].

$$\theta(S_i) = \frac{\sum_j \theta(B_{i,j})}{m}$$

N: 총피험자수 X: 각 피험자의 문항 점수 Y: 각피험자의 총점

$$\theta(I_j) = \frac{\sum_i \theta(B_{i,j})}{n}$$

학생 S_i 가 문항 I_j 에 정답 또는 오답한 것에 대한 $\theta(B_{i,j})$ 값은 다음과 같이 구할 수 있다.

$$\theta(B_{i,j}) = -\log P(B_{i,j})$$

$P(B_{i,j})$ 는 학생 S_i 가 문항 I_j 를 맞힐 수 있는 확률로서 $\frac{P(S_i)}{P(M)} \cdot P(I_j)$ 로 구하게 된다. 여기서 $P(M)$ 는 모든 i, j 에 대한 $P(B_{i,j})$ 의 평균이며 $P(S_i)$ 와 $P(I_j)$ 는 다음과 같이 구한다.

$$P(S_i) = \frac{\sum_j P(B_{i,j})}{m}$$

$$P(I_j) = \frac{\sum_i P(B_{i,j})}{n}$$

θ 값이 기준 정보량보다 적은 데이터들을 제거함으로써 흥미롭지 않은 빈발항목집합이 생성되지 않도록 하고, 추출된 연관규칙은 신뢰도, 향상도(lift), 확신도(conviction)에 의하여 흥미도를 측정한다.

3.1.3 문항간 연관성 원인 분석

각 문항이 가지고 있는 특성 중 성취기준(평가 영역: 도메인), 지식의 종류를 정의하고 이들 간의 관계를 분석하여 연관성의 원인을 제시하도록 설계하였다. 성취기준은 각 문항이 평가하고자 하는 내용으로서 교과 내용을 분석함으로써 분류하였다[10]. 지식의 종류는 평가하고자하는 내용이 어떤 종류의 지식인지를 구분하는 것이다. 평가 내용이 동일하더라도 지식의 종류에 따라 난이도와 해결 방법이 다르기 때문에 피험자의 성취 정도를 단계적으로 접근할 수 있도록 한다. 본 시스템에서는 지식의 종류를 학교 현장에서 사용하고 있는 이원목적분석표에 제시된 지식, 이해, 응용

의 3가지로 구분하였다.

이 두 가지 특성에 의하여 연관성이 나타날 수 있는 경우를 다음과 같이 분류하였다. 성취기준과 지식의 종류가 모두 동일하고 두 문항이 정답-정답의 관계를 갖는 경우, 특정 문항이나 답지는 피험자가 다른 문항을 맞힐 수 있도록 단서를 제공할 수 있다. 성취기준이 동일하고 지식의 종류가 다른 경우, 정답-정답 관계인 두 문항은 한 문항이 다른 문항에 대한 정답 단서를 제공했거나, 해당 내용에 대해 학습 성취도가 매우 높다고 판단할 수 있다. 동일 경우에서 정답-오답 관계인 두 문항은 심화·응용 문제를 맞히지 못하는 상황으로 학습 성취도가 미흡하다고 판단할 수 있다. 본 시스템에서 제공하는 연관성의 원인은 정답 단서 제공과 성취도 평가라는 두 가지 측면에서 이루어지도록 하였다.

3.2 문제은행 모듈

문제은행 모듈은 문항작성, 문항추출의 기본적인 기능으로 구성되었다. 본 시스템의 특징은 문항분석 모듈에서 분석된 결과 중 두 문항이 정답에 대한 단서로 작용한다고 판단된 경우, 다음 검사지에 두 문항이 동시에 추출되지 못하도록 연관성을 데이터베이스에 저장하도록 설계한 것이다. 이를 위하여 문제은행은 다음과 같은 방향으로 설계하였다.

첫째, 성취기준은 교수-학습목표로서 데이터베이스에서는 도메인이란 용어로 사용되었다. 성취기준은 컴퓨터 교과의 내용을 분석한 후, 대영역, 중영역, 소영역으로 구분하였다[10].

둘째, 사용자가 입력한 텍스트 값은 일련의 숫자로 데이터베이스에 저장된다.

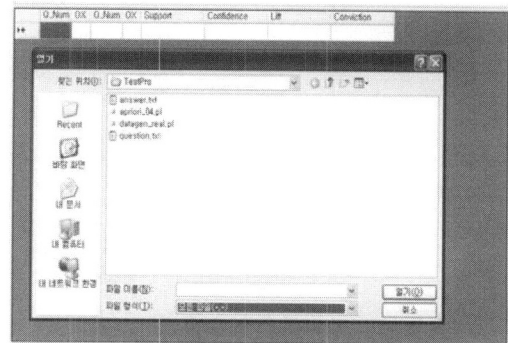
셋째, 각 성취기준에 대하여 지식, 이해, 응용의 3가지 종류로 구분하여 입력하도록 하였다[10].

4. 시스템 구현 및 검증

4.1 시스템 구현

시험 응답데이터 분석시스템은 .NET Frame

work를 기반으로 구현하였다. <그림 2>는 분석 시스템 초기화면에서 데이터집합의 데이터들에 대한 정보량 θ 값을 구하기 위하여 프로그램을 실행하는 화면이다. 최소 임계값 θ 이하의 정보량을 가진 트랜잭션 데이터가 필터링 되고 새로운 파일이 생성되면 사용자가 최소 임계값을 입력하여 어트리뷰트도 제거할 수 있도록 하였다.



<그림 2> θ 값으로 데이터 정제 실행

원본 응답 데이터 집합으로부터 무의미한 데이터를 제거한 후, Apriori 알고리즘을 실행하면 <그림 3>과 같이 연관규칙이 추출된다. 사용자는 흥미도 특정 척도인 신뢰도, 향상도, 확신도의 값을 임의로 조절하여 연관성의 정도를 분석할 수 있다.

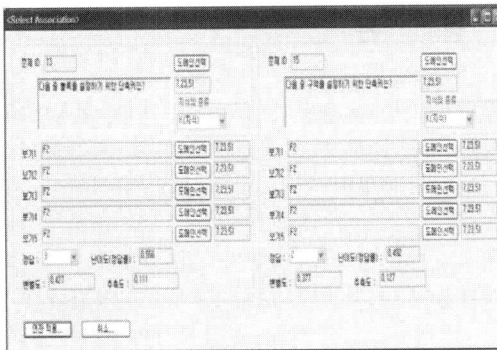
Association Rule							
List Option							
Confidence >=		LR >=		Conviction >=			
0.8		1.0		1.0			
Rule A							
G.Num	O	G.Num	O	Support	Confidence	LR	Conviction
13	0	19	0	0.095	1.004	1.403	
15	1	13	1	0.07	1.564	3.403	
0	1	5	1	0.00	1.112	1.704	
Rule B							
G.Num	O	G.Num	O	Support	Confidence	LR	Conviction
12	0	20	0	0.04	1.16	1.765	
20	1	12	1	0.04	1.022	1.765	

<그림 3> 문항간 연관규칙 추출

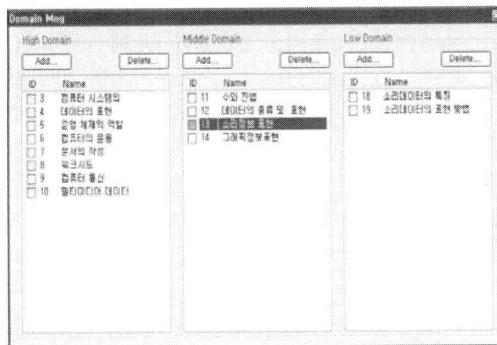
발견된 규칙들을 Rule A와 Rule B로 분류하였다. Rule A는 두 문항이 정답 단서를 제공하는 관계이고, Rule B는 학습 내용의 성취도 수준을 제고해야 하는 경우로서 사용자는 <그림 4>와 같

이 시스템에서 제공한 연관성과 원인을 구체적으로 확인할 수 있다. 두 문항의 특성을 분석하여 시스템이 분류한 Rule A에 해당되는 연관성을 인정하면, 두 문항이 다음 검사지의 문항으로 동시에 출제되지 않도록 문제은행 데이터베이스에 저장할 수 있도록 하였다.

문제은행모듈에서 데이터를 처리하는 역할을 하는 데이터베이스 서버는 오라클(Oracle) 10.0을 사용하였다. 문제은행에 문항을 추가하기 위하여 <그림 5>와 같이 대영역, 중영역, 소영역 순으로 도메인을 추가하고, 역순으로 제거할 수 있도록 하였다. 도메인이 구성되면 해당 문항을 작성한다.



<그림 4> 연관규칙을 데이터베이스에 적용



<그림 5> 계층적 구조의 도메인 구성

추가되는 문항은 정답률에 의해 정해진 난이도 수준과 도메인에 따라 검색할 수 있도록 하였다. 검사지를 구성하기 위해 해당하는 도메인, 지식의 종류, 문항 수를 입력하여 데이터베이스로부터 문항을 추출할 수 있다.

4.2 검증

본 연구에서 구현한 시험 응답데이터 분석시스템에서 제공하는 문항간 연관성과 원인 분석에 대한 타당성, 현장에서의 활용도를 검증하기 위하여 서울에 P고등학교에서 실시한 정보사회와컴퓨터 교과와 2개 시험 응답데이터를 대상으로 실험하고, 그 결과에 대한 평가를 컴퓨터 담당 교사 10명에게 설문지로 조사하였다.

4.2.1 실험결과

컴퓨터1은 2005학년도 1학기 기말고사의 응답 데이터로서 평가내용은 데이터 표현 영역 38%와 워드프로세서(한글 2002) 영역 62%로 구성되었다. 컴퓨터2는 2007학년도 1학기 중간고사의 응답 데이터로서 평가내용은 데이터 표현 영역 68%와 워드프로세서(한글 2002) 영역 32%로 구성되었다.

• 컴퓨터1 분석 결과

시스템으로 분석된 문항간 연관성은 $I11 \Rightarrow I12$, $I12 \Rightarrow I11$, $I18 \Rightarrow I17$ 이다. 문항 11번(정답률 80%)은 기능키 F3을 사용하여 문서의 특정 부분을 블록 설정할 수 있는지를 평가하고, 문항 12번(정답률 77%)은 기능키 F4를 사용하여 문서의 특정 부분을 구역 설정 할 수 있는지를 평가하는 것으로 난이도의 차이는 없고 성취기준은 모든 영역이 일치하므로 F3, F4 기능키 중 한 개의 기능만 정확히 알고 있다면 다른 하나는 추측으로 맞힐 수 있다는 가능성이 발견되었다. 문항 18번(정답률 57%, 응용)은 3월 3행의 표 안의 특정 셀을 선택하고 CTRL, SHIFT, ALT 키를 이용하여 크기를 조절하는 방법을 평가하고, 문항 17(정답률 81%, 지식)번은 4월 3행의 표 안에서 특정 셀의 위치를 열과 행으로 표현하도록 하여 열과 행의 개념을 평가하는 것이다. 두 문항은 워드프로세서 영역의 성취기준에서 대영역과 중영역이 동일하고 난이도의 차이도 있기 때문에 두 문항간의 정답 관계가 당연하다고 볼 수도 있겠지만, 18번 문항의 답지 1번에 열과 행에 대한 내용이 언급되어 있어 문항 17번을 맞히는데 단서가 될 수 있다는 가능

성이 분석되었다.

• 컴퓨터2 분석 결과

연관된 문항 13, 14, 15, 16 은 정답률이 각각 65%, 68%, 70%, 60%로서 난이도에 의해 발생한 연관성으로 판단되기 보다는 진법변환이라는 동일한 성취기준에 해당되기 때문인 것으로 분석된다. 특히 문항 16번은 16진수를 8진수로 변환하는 내용으로서 풀이 과정상 16진수를 2진수로 변환한 후 8진수로 변환하기 때문에 2진수를 8진수로 변환하는 문항 14번, 2진수를 16진수로 변환하는 문항 15번과는 매우 높은 연관성이 있는 것으로 분석되었다.

4.2.2 타당성 및 활용성 설문 결과

컴퓨터 교과 담당 교사 10명에게 시험 응답데이터 분석시스템에서 분석된 문항 연관성의 흥미도와 현장에서의 활용성에 대하여 5점 척도로 평가하도록 하였다. 문항간 연관성의 타당도와 본 시스템의 활용도에 대한 평가 결과는 <표 1>과 같다. 시스템에서 제공하는 문항간 연관성은 흥미롭고 의미가 있는 것으로 평가되었으므로 본 연구에서 제안한 마이닝 방법이 빈발항목 집합의 생성을 감소시키고 시험 응답데이터의 특성을 잘 반영하여 흥미로운 연관규칙을 추출하는 데 유용한 것으로 검증되었다.

시험 응답데이터 분석시스템의 활용성에 대한 설문 결과에서는 교육 현장에서의 활용도가 높게 평가되었지만, 문항 연관성에 대하여 시스템에서 분류하는 원인이 적합한지에 대해서는 평균 2.9의 낮은 평가를 받았다. 이러한 결과는 학습 환경, 교사의 주관적 판단 등에 따라 문항 연관성이 발생한 원인이 다양하게 분석될 수 있기 때문인 것으로 해석된다.

<표 1> 문항간 연관성 타당도 평가

설문내용	평가
추출된 연관규칙의 문항들이 연관성이 있는가?	4.0
문항간 연관성의 원인은 도메인과 관련되어 있는가?	4.1
시스템에서 제공하는 연관성에 대한 원인 분석이 적합한가?	2.9
연관성이 발견된 문항들을 문제은행에 반영하는 기능이 유용한가?	4.5
기존 분석시스템과 함께 사용되면 다차원적인 평가를 할 수 있는가?	4.7
교육 현장에서의 활용도가 있는가?	4.9

5. 결 론

일반적으로 교육평가 분야나 학교 현장에서는 검사이론과 통계적 도구를 기반으로 학생들의 성취 능력과 문항의 질을 분석한다. 이러한 검사이론은 특정 문항에 대한 반응은 다른 문항에 대한 반응에 전혀 영향을 미치지 않으며, 학생 능력과 문항특성에 의해서만 문항반응이 결정된다는 것이다. 그러나 실제 학교 현장에서의 검사 결과를 분석하면, 한 문항에 답을 하거나 맞히는 것이 다른 문항에 답하는 것에 영향을 미치는 경우가 발생한다. 이러한 상황에서 추정된 각 결과들은 문항과 피험자들의 능력을 평가하는 데 잘못된 판단을 초래 할 수 있다. 각 문항에 대한 피험자들의 반응도 중요하지만, 피험자들의 반응에 의하여 발생된 문항들 간의 연관성을 분석하는 것도 피험자의 능력을 추정하는 데 의미 있는 영향을 미치게 된다. 그러므로 본 연구에서는 연관규칙 마이닝을 기반으로 문항 간의 관계를 추출하여 피험자 반응에 의한 잠재적인 문항간 연관성을 찾아냄으로써 기존의 문항 분석 방법에 새로운 분석 방안을 제시하였다. 흥미로운 연관규칙을 추출하기 위하여 정보성이 약한 데이터들을 먼저 제거한 후, 연관규칙을 추출하고 그 원인을 분석할 수 있도록 하였다.

본 시험 응답데이터에서 추출된 문항간 연관성을 분석하여 다음과 같이 활용할 수 있을 것으로 기대된다.

- 도메인과 지식의 종류가 동일한 문항의 경우, 출제자가 두 문항을 의도적으로 동시에 출제하지 않았다면 이 두 문항은 정답의 단서로 상호 작용한 결과일 가능성이 높기 때문에 다음 시험지에 함께 출제되지 않도록 할 수 있다.
- 도메인과 지식의 종류가 동일한 문항들을 $I_X \rightarrow I_Y$ 의 관계를 예상하고 출제자가 의도적으로 동시에 출제하는 경우가 있다. 즉, 문항 X번을 맞으면 문항 Y번도 맞혀야 하는 것이다. 그러나 실제 응답 반응에 의한 결과는 $I_X \rightarrow I_Y$ 의 관계가 발생되었다면 특정 도메인에 대한 학습 성취도가 미흡한 것이기 때문이다. 그러므로 교수-학습 과정이나 학생들의 이해 수준을 재점검하여 적절한 피드백을 제공하여 해결하도록 할 수 있다.
- 일반적으로 기대하거나 예상 할 수 없는 문항간 연관성이 발생하는 경우가 있다. 예를 들면, 도메인이 동일하거나 연관성이 없어도 두 문항간 연관성이 여러 번의 응답데이터 분석 때마다 자주 발견된다면 지금까지 알려지지 않았던 학생들의 학습·이해 과정의 구조를 재조명하고 새로운 방법으로 이해하는 계기를 마련할 수 있을 것이다.

참 고 문 헌

- [1] 강성진, 이명애(2002), 비선형 다층모형에 의한 검사 문항반응 분석방법, 교육평가연구, 15(1), pp. 273-294.
- [2] 강승호(1996), 현대 교육평가의 이론과 실제, 양서원,
- [3] 광은영, 김현철(2007), 문항 응답 데이터에서 문항간 연관규칙의 질적 향상을 위한 도구 개발, 컴퓨터교육학회논문지, 10(3), pp. 1-8.
- [4] 민정석, 김종필(2006), 다차원 문항반응모형의 척도연계방법, 교육평가연구, 19(1), pp. 1-19.
- [5] 박정(2001), 다분 문항반응이론 모형, 교육과 학사.
- [6] 성태제(2003), 고전검사이론과 문항반응이론에 의한 문항과 검사 분석, [http:// home.ewha. ac.kr/ ~tjseong/](http://home.ewha.ac.kr/~tjseong/).
- [7] 성태제(1998), 교육측정, 평가의 새 지평, 교육과학사.
- [8] 성태제(1991), 문항반응이론 입문, 양서원.
- [9] 성태제(2002), 현대교육평가, 학지사.
- [10] 이승현, 광은영, 김현철(2004), 7차 교육과정 에 따른 '정보사회와컴퓨터' 교과 의 평가도 구 개발, 컴퓨터교육학회논문지, 7(1), pp. 15-25.
- [11] 이형봉(2000), 연관성 규칙탐사 알고리즘 비 교분석, 호남대학교 산업기술연구소 논문집, 8, pp. 185-201.
- [12] 임영문, 최영두(2000), 연관규칙을 이용한 데 이터 분석에 관한 연구, 산업경영시스템학회 지, 23(61), pp. 115-126.
- [13] Ferrar, S., Huynh, H., Baghi, H.(1997), Contextual characteristics of locally dependent open-ended item cluster in a large-scale performance assessment, Applied Measurement in Education, 10(2), pp. 123-144.
- [14] Ferrar, S., Huynh, H., Michaels, H.(1999), Contextual explanation of local dependent in item clusters in a large scale hands-on science performance assessment, Journal of Education Measurement, 36(2), pp. 119-140.
- [15] Jiawei. H., Michaeline, K.(2000), Data Mining: Concepts and Techniques, Morgan Kaufmann Publisher.
- [16] Knorr, E. M., NG, R. T., Tucakov, V.(2000), Distance based outliers: Algorithms and applications. Int. J. Very Large Databases 8, pp. 237-253.
- [17] Yen, W. M.(1993), Scaling performance assessment: strategies for managing local item dependence, Journal of Education Measurement, 30(3), pp. 187-213.
- [18] Zaki, M. J. and K. Gouda(2003), Fast Vertical Mining using Diffsets, Proc. of the Int. Conf. on Knowledge Discovery and Data Mining, pp. 326-335.
- [19] Zenisky, A. L., Hambleton, R. K., Sireci, S. G.(2002), Identification and evaluation of local item dependences in the Medical College Admissions Test, Journal of Education Measurement, 39(4), pp. 291-309.

곽 은 영



1992 고려대학교 사범대학
가정교육과(교육학학사)

2001 고려대학교 교육대학원
컴퓨터교육전공(교육학석사)

1993~ 서울 백암고등학교 교사

2002~현재 고려대학교 대학원 컴퓨터교육학과
박사과정

관심분야: 컴퓨터교육, 데이터마이닝, 평가시스템

E-Mail: key@comedu.korea.ac.kr

김 현 철



1988 고려대학교 전산과학과
학사

1990 미조리 주립대학 (Rolla)
(전산학석사)

1998 플로리다 대학 (전산학박사)

1998 GTE Data Services, Inc. 시스템 분석가

1998~1999 삼성 SDS 책임컨설턴트

1999~현재 고려대학교 컴퓨터교육과 교수

관심분야: 컴퓨터교육, 데이터마이닝,
기계학습알고리즘, 바이오인포매틱스

E-Mail: hkim@comedu.korea.ac.kr