

## IRT와 데이터 마이닝을 이용한 효과적인 평가 및 추천시스템

김 천 식\*, 정 명 희\*

# Efficient Assessment and Recommendations System using IRT and Data Mining

Cheonshik Kim\*, Myunghee Jung\*

### 요 약

이러닝 교육 방법은 오프라인 교육의 단점을 보완하는 많은 장점을 갖고 있다. 이와 같은 이유로, 기존의 많은 오프라인 교육기관은 학습효과를 높일 목적으로 온라인 기술을 현장에 도입하였다. 최근의 일반 대학교에서도 온라인 학습을 부분적으로 도입하고 있다. 그 결과 기존의 오프라인 교육의 장점을 온라인 교육에 도입하여 교육의 효과를 높일 수 있는 방법에 대한 연구를 진행하고 있다. 그 결과 온라인에서 학습 효과를 높이기 위해서 학습자를 적절히 평가하고 적절한 피드백을 제공하는 것이 필요하게 되었다. 따라서, 본 논문에서는 오프라인의 장점을 온라인에 도입하여 학습 능력을 향상 시킬 수 있는 모델을 제안하고자 한다. 제안한 모델의 핵심은 올바른 평가에 있다. 그러므로, 학생들을 위한 정확한 평가를 위해서 문항반응검사를 실시하여 학습자를 평가하고 수준별 학습을 할 수 있도록 하였다. 또한, 온라인에서 학생의 학습 스타일을 알아내어 각 학생의 잘못된 학습 성향을 개선하게 하여 학습의 효과를 높도록 하는 방법을 제안하였다. 향후 본 논문에서 제안 방법이 현장에서 활용된다면 학습자의 학업능력 개선에 도움이 될 것으로 기대한다.

### Abstract

E-learning method has many advantages that supplement the shortfalls of offline education. For this reason, today's offline educational institutions adopted the online education technique to improve learning effectiveness. Recently, general universities have partially adopted online learning. As a result, a study is searching for ways to improve the effectiveness of education by copying the merits of the existing offline education onto the online education. Thus a proper evaluation of learners and a feedback provision are considered necessary to improve the effectiveness of online learning. This study aims to suggest a model that will improve learning efficiency by adapting the advantages of offline education to online learning. To evaluate properly, this study conducted Item Response Test to examine the learners and finally ensure them an adequate level of education. Also, this study suggested a way to enhance learning efficiency by finding out each learner's study habits and to address the weaknesses of online learning. It is expected that the suggested method would be helpful in bettering learner's ability to study in school environment.

▶ Keyword : e-Learning, online, IRT, Data mining

• 제1저자 : 김천식

• 접수일 : 2006.08.18, 심사일 : 2006.08.28, 심사완료일 : 2006.09.20

\* 안양대학교 디지털미디어공학 전공 교수

## I. 서론

전문가들은 “한국만큼 온라인 교육 환경이 좋은 곳이 없다”고 말한다. 전체 가구의 70%가량이 초고속 인터넷 서비스를 사용하고 있기 때문이다. 동영상 강의는 물론이고 쌍방향 대화식 강의도 이뤄지고 있다. 지금까지의 교육에 대한 견해는 초등 교육부터 대학교육까지의 정규 교육을 일반적으로 교육이라 통칭했다. 하지만, 산업분야의 지식은 이미 정규 교육으로는 전달하기에 한계가 있을 만큼 많은 지식이 생성되고 있다. 따라서, 이러닝은 이러한 교육 시장의 변화에 적합한 교육의 한 형태이다.

이러닝은 전통적 교육과 비교할 때 컴퓨터기반의 온라인 교육으로 다음과 같은 여러 가지 장점을 가지고 있다. 첫째, 인터넷을 통한 학습체제는 전 세계 모든 학습자에게 정보를 개방하고 공유하는 시스템으로 되어 있어 가치 있는 지식을 공유할 수 있으므로 지식경영시스템으로 활용할 수 있다. 둘째, 시간적, 공간적 제한이 없고 비교적 적은 비용으로 다양한 기법을 통하여 off-Line 학습체제의 한계를 극복하고 최신의 정보를 신속하게 얼마든지 학습할 수 있다. 인터넷을 통한 학습체제는 전 세계 모든 학습자에게 정보를 개방하고 창출된 정보를 공유하는 시스템으로 되어 있어 가치 있는 지식 보유가 가능하다.

셋째, 자기주도학습(Self-Directed Learning)으로 개별화 학습이 가능하고, 필요할 때 수시로 학습자가 자기 수준에 맞는 다양한 교육프로그램을 통하여 스스로의 능력을 향상시킬 수 있다. 조직 구성원들 개개인인 이러한 학습 환경을 바탕으로 자신들에게 필요한 사항을 중심으로 학습목표를 스스로 수립하고, 그 목표를 성취하기 위한 내용을 스스로 찾아가면서, 자신에게 익숙한 방법으로 언제 어디서든 학습할 수 있는 환경을 지향하고 있다. 이러한 자기 주도적 학습의 특징 때문에 이러닝의 자생적인 발전과 구성원의 능동적인 참여가 가능하게 된다[1].

이와 같은 장점이 있는 반면에 전통적인 학습 환경의 장점을 제공할 수 없는 단점이 있다. 즉, 학습자가 수업시간에 공부에 전념하는지는 오프라인 환경이라면 즉각적으로 알 수 있다. 그러나 온라인 환경에서는 학습자가 동영상 강의를 켜놓은 채 다른 볼일을 볼 수도 있다. 또한 오프라인 환경이라면 학습자의 능력을 평가하여 학습자의 능력에 따라서 반 편성을 다르게 하여 학습자가 학습에 도움이 되도록 하기도 한다.

따라서 본 논문에서는 이와 같은 오프라인 학습의 장점

을 이러닝에 도입하기 위해서 학습 평가와 피드백을 효율적으로 하기위한 방안을 제시하고자 한다. 또한 학습자의 학습 성향을 파악하고 적절한 조안을 통해서 학습 능력을 향상시킬 수 있는 모델을 제안하고자 한다.

본 논문의 2장에서 관련 연구를 통해서 기존의 이러닝 시스템과 데이터 마이닝에 관한 연구에 대해서 살펴보고 3장에서는 IRT를 이용한 학습 평가 방법에 대해서 설명한다. 4장에서는 이러닝 시스템에 대한 구조와 데이터 마이닝을 이용한 학습 스타일을 알아내는 방법에 관하여 설명하고 5장에서는 효율적인 학습을 위한 이러닝 모델에 관하여 설명하고 6장에서는 실험 및 분석을 통해서 제안한 방법에 대해서 평가한다. 7장에서 결론을 맺도록 한다.

## II. 관련 연구

이러닝(E-Learning)은 최신 기술에 가장 큰 영향을 받는 분야이다. 네트워크 기술의 비약적인 발전과 소프트웨어 기술의 발전은 오프라인 교육과 온라인 교육 분야에서 교육의 질 개선에 많은 도움을 주고 있다[2,3,4]. 전통적인 교육 과정을 이러닝 시스템에 도입하는 것은 여러 가지 고려사항이 필요하다. 전통적인 교육은 여러 학생이 한분의 선생님으로부터 강의를 일방적으로 받는 방식이다. 각 학생이 해당 수업내용을 충분히 이해하지 못한 경우라도 교사의 주도에 의해서 다음 단원으로 진행할 수밖에 없다.

이와 같은 형태의 학습은 학습능률 향상에 도움이 되지 않는다. 왜냐하면 개인마다 학습에 대한 이해력이 빠르고 느린 정도가 다르기 때문이다. 또 어떤 사람은 집중력이 좋은가 하면 그렇지 못한 사람도 있다. 따라서, 개인에게 적절한 평가와 평가 결과에 따라서 적절한 피드백을 주어야 학습능률이 향상될 것이다. 최근에 전자상거래 분야에서는 데이터 마이닝을 이용하여 어떤 사람이 홈페이지를 언제 방문하는지 어디를 둘러보는지 얼마동안 웹 서핑을 하는지 등의 정보를 알아낼 수 있다. 이와 같은 정보는 기업의 입장에서 비즈니스 모델을 개발하는데 이용이 된다. 마찬가지로 이러닝 분야에서도 어떤 학생이 홈페이지를 방문해서 얼마동안 공부를 했는지 그리고 어떤 콘텐츠를 사용했는지 등을 알 수 있다면 학습 스타일을 조언하고 학습 능률을 향상시킬 수 있도록 도울 수 있다. 이와 같은 개인화 된 웹은 전자상거래[5,6]에서 성공적으로 구현되었다. Blochl et al. [2]는 학습자에게 적응적인 학습을 제공하기 위한 심리적인 측면을 고려한 개인화 학습 시스템을 제공 하였다.

SPERO[6]는 IEEE LTSA(Learning Technology

Systems Architecture)에 기초한 개인화된 이러닝 시스템이다. 이 시스템은 외국어를 배우기 위한 방법으로 학습자의 흥미와 레벨에 맞는 학습 자료를 제공하는 특징이 있다. SPERO의 문제점은 학습자에게 질문을 하고 프로파일을 만들게 됨으로써 비용이 많이 드는 형태로서 너무 많은 작업량이 필요하다. Surfien(4)는 데이터 마이닝 기술을 사용한 추천 시스템이다.

본 논문에서 제안한 마이닝 기술은 마이닝을 이러닝에 도입한 논문 중에서 학생의 학습 성향을 분석하여 조언하는 시스템은 없었다. 따라서 본 논문에서는 학습자를 IRT를 이용하여 정확히 진단하고 적절한 난이도의 문제와 학습 콘텐츠를 제공하고 학습능률을 높일 뿐만 아니라, 학습자의 학습 성향을 데이터 마이닝의 연관성규칙 알고리즘을 이용하여 정확하게 진단하여 잘못된 학습 스타일을 가진 학생들에게 학습 스타일의 개선을 조언하여 학습 능률을 개선하도록 하는 것이 본 논문의 목적이다.

### III. IRT를 이용한 학습 평가 방법

문항반응이론(Item Response Theory)(8)은 검사 총점에 의하여 문항이 분석되는 것이 아니라, 문항은 문항 하나 하나의 불변하는 고유한 속성을 지니고 있으므로, 그 속성을 나타내는 문항특성곡선에 의하여 문항을 분석하는 검사 이론이다.

문항반응이론이 고전검사이론 보다 수리적으로 복잡함에도 불구하고 계속 관심의 대상이 되고 있는 이유는 다음과 같다. 첫째, 문항특성의 불변성 때문이다. 문항반응이론의 기본개념은 문항이란 하나의 특성을 측정하기 위해 만들어진 것이라는 점이다. 따라서 문항은 문항이 가지고 있는 고유한 특성이 있기 때문에 검정자 집단의 특성이 다르다고 하여 문항특성이 다르게 추정되지 않는다는 점이다. 둘째, 검정자 능력의 불변성 때문이다. 피험자는 각기 가지고 있는 고유한 능력수준이 있기 때문에 다른 검사 도구를 사용한다 할지라도 피험자의 능력이 다르게 추정되지 않는다는 점이다. 셋째, 피험자의 능력추정이 정밀하다는 이유 때문이다.

문항 난이도는 성취도 검사 (예, 수학시험, 지능검사)에서 문항 'I'를 얼마나 맞추기 어려운지를 나타내는 척도이다. 높은 문항 난이도를 지닌 문항은 맞추기 어렵고, 심리학에서 많이 쓰이는 태도검사나 임상진단용 검사에서 문항 난이도는 문항 'I'에 얼마나 동의하기 어려운지를 나타낸다.

피검사 능력점수는 성취도에서 능력점수가 높은 피검자는 어려운 문항을 맞출 가능성이 (낮은 피검자에 비해) 상대적으로 높다. 태도검사나 임상진단용 검사에서 능력점수가 높은 피검자는 동의하기 어려운 문항에 동의할 가능성이 상대적으로 높다.

(공식 1)은 문항의 난이도를 측정하기위한 식이다. 이 공식은 문항의 난이도를 알아내는데 사용된다. (공식 2)는 난이도 조절을 위한 식이다. 학생A가 문항 1의 문제를 풀었는데 오답처리가 되었고 이 문항이 난이도가 매우 높은 상황이라는 결론이 나온다면 A를 위해서 난이도를 조정하는 작업이 필요하다. 이때 (공식 2)를 활용한다.

(공식 3)은 피험자의 문항에 대한 능력을 평가하는 방법이다. 문항 난이도와 개인의 능력에 달려 있다. 문항 난이도와 능력의 차이값 ( $\beta - \delta$ )이 응답을 결정 한다. 즉,  $\beta$ 가  $\delta$ 보다 크면, 피검자가 문항을 맞출 확률이 1/2 이상,  $\beta$ 가  $\delta$ 보다 작으면, 피검자가 문항을 맞출 확률이 1/2 이하,  $\beta$ 가  $\delta$ 보다 같으면, 피검자가 문항을 맞출 확률이 1/2 이 된다.

$$\delta = \sum_{i=1}^i \frac{n}{N} \dots\dots\dots (1)$$

N : 총 피험자  
n : 답을 맞힌 피험자 수

$$\delta = w \times \delta (init) + (1 - w) \times \delta (used) \dots (2)$$

w : weight values,  $\delta (init)$ : 초기 난이도 값,  
 $\delta (used)$  : 계산된 난이도 값.

$$p(x = 1/\beta, \delta) = \frac{\exp(\beta - \delta)}{1 + \exp(\beta - \delta)} \dots\dots\dots (3)$$

$\beta$ : 맞춘 문항의 수,  $\delta$ : 문항의 난이도

### IV. 이러닝 시스템

이 장에서는 학습 시스템의 설계 구조와 학습자의 학습 스타일을 알아내는 기술을 소개한다.

#### 4.1 시스템 구조

본 논문의 이러닝 시스템 구조 B는 (그림 1)과 같다.

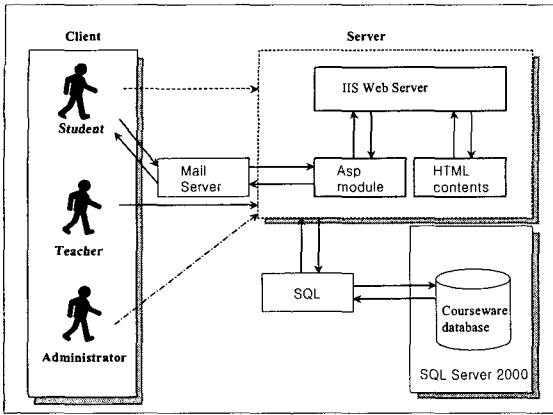


그림 1. 이러닝 시스템의 구조  
Fig 1. Structure of e-Learning System

이 시스템의 3-티어 클라이언트 서버로 구성될 수 있다. 이 시스템은 웹 서버, 데이터베이스와 메일 서버의 기본 컴포넌트로 구성된다. 웹 서버는 IIS 웹 서버와 스크립트 모듈을 사용하여 설계한다. HTML 내용은 정적인 문서를 처리하기 위한 HTML 언어와 동적인 문서 생성을 위해서 ASP를 사용하고 있다. 학습자는 웹 서버를 이용해서 코스웨어에 접속하여 자신에게 가장 적합한 학습을 진행한다. 교사(Teacher)는 웹 서버에 접속하여 학습자를 위한 문제를 생성하고 평가를 한다. 관리자는 웹 서버의 전반적인 보안 관리와 시스템의 안정적 운용과 데이터베이스 백업 등의 시스템의 운용에 필요한 모든 일을 수행한다.

데이터베이스 서버는 코스워크를 저장하는데 사용된다. 이 시스템에 사용된 데이터베이스는 SQL 2000 서버로서 안정성과 편리한 인터페이스를 제공한다. 코스워크를 제공하기 위한 데이터베이스의 구조는 STUDENTS, TEACHER, COURSE, PROBLEM, MESSAGES 등으로 구성된다.

학생을 위한 스키마 구조는 (표 1)과 같다.

표 1. 학생 스키마  
Table 1. Student Schema

Name	Type	Description
id	int(4)	자동증가
name	varchar(50)	not null
address	varchar(200)	
phone	varchar(100)	not null
email	varchar(100)	not null
password	varchar(15)	not null
comment	varchar(300)	비고

과목을 위한 스키마 구조는 (표 2)와 같다.

표 2. 과목 스키마  
Table 2. SUBJECT Schema

Name	Type	Description
no	int(4)	
code	varchar(10)	강의코드
subject	varchar(200)	강좌명
fee	varchar(100)	수강료
name	varchar(100)	강사명
weekday	varchar(50)	강의요일
time	date	강의시간
room	varchar(100)	강의실

(표 3)은 코스워크를 위한 스키마 구조를 나타낸 것이다.

표 3. COURSEWORK 스키마  
Table 3. COURSEWORK Schema

Name	Type	Description
cid	int(4)	
sid	int(4)	과목 id
uid	int(4)	학생 id
file	varchar(100)	강의 파일
assessment	varchar(100)	학생평가
comment	varchar(350)	비고

(표 4)는 학습자(student)과 교사(teacher)간의 일대일 지도나 수업내용을 업로드 하는데 이용될 수이다.

표 4. MESSAGE 스키마  
Table 4. MESSAGE Schema

Name	Type	Description
mid	int(4)	
sid	int(4)	학생 id
name	varchar(100)	mail 이름
email	varchar(100)	mail 주소
subject	varchar(350)	mail 제목
body	varchar(500)	mail body

각 스키마간의 관계는 (그림 2)와 같은 관계를 갖고 있다.

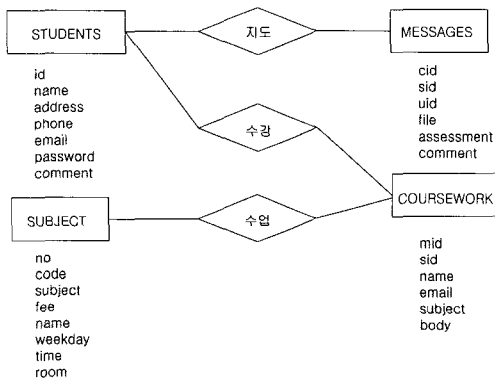


그림 2. 코스워크를 위한 E-R구조

Fig 2. E-R Structure for The Course Works

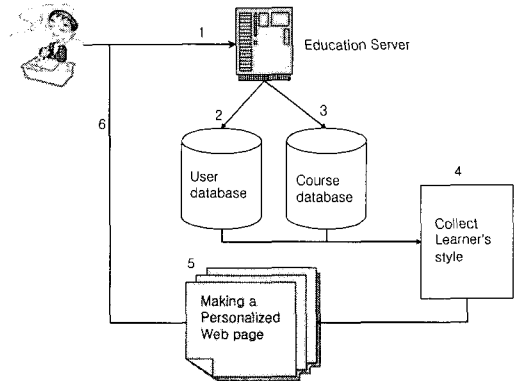


그림 3. 프로파일을 이용한 학습 스타일 추출

Fig 3. Extract Learning Style using The Profile

(그림 3)은 학습자가 학습시스템에 접속할 때 등록된 정보를 이용하여 학습자의 관심분야를 파악할 수 있도록 하기 위한 프로파일을 추출을 나타낸 것이다. (그림 3)의 과정을 단계적으로 설명하면 다음과 같다.

- ▶ STEP 1~2. 학습자가 시스템에 접속하면서 학습자의 개인정보를 등록하면 개인정보는 User 데이터베이스에 저장된다. 이후 시스템의 학습자의 개인정보를 이용하여 학습 성향을 파악하는데 이용한다.
- ▶ STEP 3. 학습자는 관심이 있는 여러 코스를 선택할 것이다. 이 단계에서는 코스를 운용하고 학습자를 평가하기 위한 여러 문항을 IRT 방법을 통해서 생성하고 평가를 수행하여 학습자의 능력을 정확히 판단할 수 있고 이 학습자에게 적합한 학습 자료와 수준을 결정할 수 있다.
- ▶ STEP 4. 1~3의 과정을 통해서 학습자의 스타일을 정확히 파악한다. 기초과정에 관심이 있는지 중급과정에 관심이 있는지 아니면 고급과정에 관심이 있는지를 파악할 수 있다. 또한, 자료가 동영상인 것을 주로 이용하는지 아니면 음성파일로 된 강의 자료에 관심을 보이는지 등을 파악할 수 있다. 이 때 어떤 학습자는 동영상 파일을 실행시키고 아마도 다른 작업을 하는 학생도 있을 것이다. 이와 같은 상황을 알아내어 학습 스타일 개선을 위한 요구 사항에 기록했다가 향후 학습자에게 학습 스타일개선을 요구하게 된다.
- ▶ STEP 5. 학습자의 스타일은 STEP 4에서 파악했다. 이 결과를 바탕으로 학습자에게 알맞은 학습 단계와 학습에 편한 학습 환경을 제공하고 추천 시스템을 이용하여 학습자의 진도나 학습과정을 지도할 수 있다.

#### 4.2 데이터 마이닝을 이용한 학습 스타일

로그파일 분석은 사용자가 어떤 사이트를 방문한 경우 서버의 로그파일에 흔적을 남기게 되며 이러한 방문자의 정확한 데이터를 기반으로 고객 분석을 통하여 마케팅에서 피드백을 할 수 있는 고객 분석 방법이다[9].

웹 사이트 방문자는 웹 브라우저를 통해 해당 사이트를 방문하게 되는데, 이때 브라우저가 서버에 파일을 요청한 기록을 시간과 IP 등의 정보와 함께 남긴 것을 액세스 로그라고 한다. 서버로부터 브라우저에 파일이 전송된 기록으로 액세스 로그를 트랜스퍼 로그라고도 한다. 현재 로그 파일을 기록하는 표준은 NCSA의 "Common Log Format"을 따른다.

트랜스퍼 로그를 보면 사용자가 해당 사이트에서 어떤 페이지를 보았는지 알 수 있는 반면에 레퍼럴 로그에는 그 페이지를 보기 위해 어떤 경로를 사용했는지에 대한 기록이 남아있다. 이 로그를 살펴보면 사용자들의 웹 사이트를 찾아오기 위해 어떤 검색엔진을 이용하는지에 대한 정보를 알 수 있고, 사이트의 구조상 어떤 페이지를 Navigation을 도와주는 역할을 하는 페이지들인지 알 수 있다[10,11].

(그림 4)는 학습자가 컴퓨터를 이용해서 학습을 하기 위해 서버에 접속하면 서버 컴퓨터에 로그 정보가 기록되는 모습을 보여주고 있다.

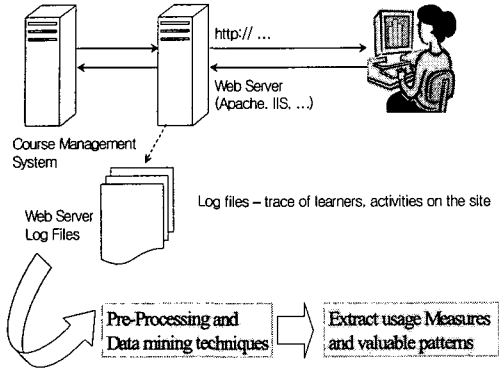


그림 4. 로그 마이닝 기술 적용과정  
Fig 4. Procedure Application of Log Mining

웹 로그 파일은 그 수가 많으나 실질적으로 분석 대상이 되는 정보를 가지고 있는 레코드 수는 소수에 불과하다. 또한, 분석목적에 따라 로그 파일 중에서 원하는 항목을 분석하여야 한다. 본 연구의 경우는 사용자의 IP, URL 그리고 참조한 웹 페이지 정보를 알면 된다. 따라서 나머지 정보는 전처리 과정에서 제거한다. 전 처리를 위해서 간단히 매크로 프로그램을 작성하여 제거하였다. 전처리 결과는 (표 5)와 같다.

표 5. 서버의 로그 데이터  
Table 5. Log Data of Server

Date	Time	IP	Page
03/Sep/2006	00:04:25	129.254.253.128	top.html
03/Sep/2006	00:04:25	129.254.253.128	main.html
03/Sep/2006	00:04:25	129.254.253.128	menu.html
03/Sep/2006	00:04:25	129.254.253.128	index.asp
03/Sep/2006	00:04:25	129.254.253.128	lecture-note.asp

(표 5)에서는 사용자에 대한 정보가 없다. 그러나 웹에서 사용자가 평가를 받기 전에 로그인 과정을 거치도록 함으로써 사용자를 구분할 수 있다. 사용자가 로그인 할 때 사용자가 이용하는 클라이언트 컴퓨터의 IP 정보를 얻을 수 있다. 따라서 로그인 때의 IP와 로그 데이터의 IP를 이용해서 사용자가 어떤 콘텐츠를 이용하는 지를 확인할 수 있고 이용시간도 알 수 있다.

또한, 우리는 학생들이 학습 사이트를 어떻게 이용하는지

에 대한 이용패턴을 분석하고자 한다. 이를 위해서, 연관성 규칙 알고리즘을 이용하고자 한다. 연관성을 평가하는 기준은 다음의 두 가지이다.

- ▶ 지지도(support) : 전체 데이터 집합에서 전체가 성립하는 레코드들의 비율이다. 이것은 규칙 적용이 얼마나 일반적인가를 말한다.
- ▶ 신뢰도(confidence) : 전체를 충족하는 레코드 중에서 지정된 결론을 갖춘 레코드들의 비율이다. 이것은 전체가 결론을 함축할 가능성을 나타낸다.

연관성 분석을 통해서 우리는 학생들이 학습을 할 때 주로 이용하는 학습 자료를 알 수 있다. 이 결과는 최종적으로 학생이 성적이 좋지 않을 때 학습 성향이 한쪽 부분만 주로 이용하는 경우학습 성향을 개선하도록 요청함으로써 학습 능력을 높이도록 하고자한다. 연관분석을 위해서 논문에서는 (표 5)의 데이터를 (표 6)과 같이 변형하였다. 즉, 어떤 사용자가 동영상 자료를 이용하였다면 T로 표시하고 이용하지 않았다면 F로 표시하였다. 이 실험은 Weka 3.5를 사용해서 연관분석을 할 것이다.

표 6. 연관성 분석을 위한 데이터  
Table 6. Data for The Association Analysis

IP	강의자료			
	동영상	GVA	게시판	채팅
125.178.215.150	T	F	T	F
125.178.215.150	F	T	F	F
125.178.215.150	F	F	T	F

## V. 효율적인 학습을 위한 이러닝 모델

이 장에서는 학습자의 학습 능력을 향상시키기 위해 학습자를 평가하고 평가의 결과 학습자에게 적합한 레벨을 학습하게 한다. 컴퓨터를 이용한 학습에서 학습자의 학습 수준에 따라서 문제의 난이도를 조정한다. 이것은 학습자의 능력에 맞는 개인화된 평가를 도와준다.

다음의 (그림 5)는 피드백을 제공하기 위한 러닝시스템의 구조를 보인 것이다.

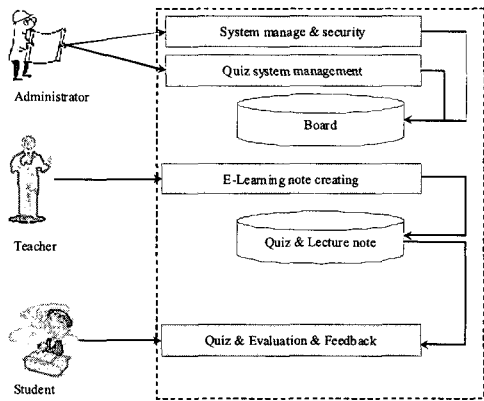


그림 5 피드백을 위한 러닝 시스템의 구조  
Fig 5. The Structure of Learning System for Feedback

**Evaluations Management**

평가는 학습자의 능력을 평가하는데 사용된다. 이 부분은 학습자에게 제공된 문제의 수준에 대한 복잡도의 적합성 여부를 평가하는데 이용된다. 학습자에 대한 분석과 평가의 신뢰성에 대한 끊임없는 개선을 통해서 학습자에게 도움이 되는 평가가 이루어지도록 되어야 할 것이다.

**Adaptive Feedback**

피드백 장치는 평가 관리를 위한 구조가 될 수 있다. 피드백의 또 다른 타입은 "힌트"로 알려져 있다. 학습자에게 힌트를 제공하기 위해서는 학습자의 프로파일(Profile)과 학습자의 평가 데이터 등을 알고 있어야 하며 그 결과로서 학습자를 위한 충고, 예상, 경고 등을 보여 줄 수 있다. 이 시스템은 토픽과 문제의 레벨을 선택할 수 있다. 개인화된 피드백은 개별적인 평가의 가장 중요한 부분이다. 피드백 시스템은 학습자에게 조언을 함으로써 학습자가 자신의 학습 진도나 수준을 선택하는데 도움을 주고자한다.

**Feedback System Engines**

(그림 6)은 학습자가 코스를 수강하는 동안 학습자를 평가하기 위한 처리과정을 나타낸 것이다. 학습자는 각 장에 대한 공부를 하고 퀴즈를 선택해서 시험을 본다. 학습자의 시험에 대한 답안 기록은 데이터베이스에 저장하여 보관되고 학습자의 능력 평가와 학습자의 학업 성취도 측정 등에 이용한다. 본 논문에서 피드백 시스템은 시험 성적을 이용하여 학습자의 능력을 분석하는 처리를 수행한다.

처음으로 학습자가 코스를 선택해서 퀴즈 시험을 본다면 학습자의 레벨을 알 수 없으므로 기초 단계부터 시작한다.

즉 레벨 1부터 시작한다. 해당하는 레벨의 문제를 60점 이상 맞추면 다음 레벨로 단계를 높여서 문제를 풀도록 한다. 만일 그렇지 못한 경우 해당 장에 대한 학습을 다시 할 수 있도록 지도 메시지를 보내고 수업을 마친 후 해당 레벨의 다른 문제를 제공하여 학습의 평가를 실시한다.

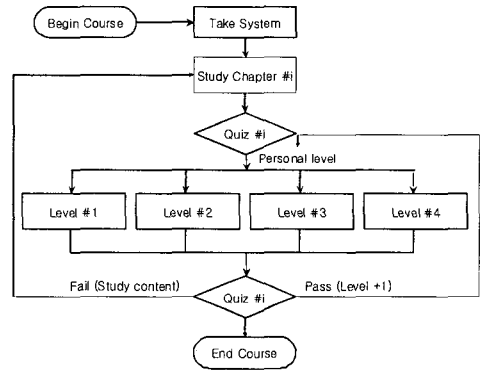


그림 6 피드백 엔진의 흐름도  
Fig 6. The Flowchart of Feedback Engines

**VI. 실험 및 분석**

본 연구의 분석을 위해 대학교 2학년 자료구조 단원에 관해 평가 문항을 만들어 실험을 수행하였다. 학습 평가에 참여한 인원수는 50명이다. 고전 검사이론에 의하여 문항과 검사의 특성을 TestAn(성태제)의 툴을 사용하여 (그림 7)와 같이 문항을 분석하였다.

문항 번호	문항 난이도	문항 상관계수	문항 변별도	문항 점수	문항 평균	문항 표준편차	문항 점수	문항 순위	문항 순위	문항 순위	문항 순위
1	.72	.63	.77	.08	.00	.23					
2	.12	.00	.31								
3	.04	.00	.15								
4	.04	.00	.08								
5	.72	1.00	.23								
기타	.00	.00	.00								

그림 7. 고전 방법을 사용한 문항 분석 결과  
Fig 7. The Result of Item Analysis using Traditional Methods

문항난이도는 총 피험자 중 문항의 답을 맞힌 피험자의 비율을 나타낸다. (그림 7)의 문항 번호 1의 문항 난이도는 .72로서 쉬운 문항으로 분석된다. 상관계수에 의한 문항 변별도는 문항점수와 총점과의 상관계수로서 1번 문항의 상관계수에 의한 문항 변별도는 0.63이고, 정답비율 차에 의한 문항 변별도는 상위 27% 피험자 집단의 정답비율 차이에 의한 문항 변별도는 77%이다. 1번 문항은 변별력이 없는 문항으로 최소한 수정되거나 제외되어야 할 문항이다. '응답비율 전체'는 전체 피험자들이 각 답지에 응답한 비율을 말한다. 1번 문항의 경우 1번 답지에 전체 피험자의 8%, 2번 답지에 12%, 3번 답지에 4%, 4번 답지에 4%, 5번 답지에 72%가 응답하였다. 문항분석 결과에 의하면 50명 피험자 점수의 평균은 13.040, 표준편차는 4.133 분산의 17.078, 최저 점수는 5점 최고 점수는 30점이다.

전통적인 시험의 평가는 맞은 문항수를 이용하여 학생을 평가하고 평가 결과에 피드백을 주어 학습 능력을 향상시키게 된다. (표 7)는 50명 학생에 대한 평가 결과로서 정답문항 수, 즉 총점과 2모스 로지스틱 모형에 의한 능력추정치와 추정오차 및 진점수를 나타내었다. 피험자 3, 10, 16은 같은 15의 문항의 정답 문항을 기록했다. IRT를 이용하여 진점수로 바꾼 결과 각각 15.719, 14.886, 16.733을 다른 결과를 얻었다. 따라서 문항에는 문항의 난이도가 있고 그 결과 어떤 문항에 정답을 알고 있는냐에 따라서 그 사람에 대한 평가를 달라져야 한다.

표 7. 피험자 능력 추정결과

Table 7. Inference Result of Testee Ability

피험자ID	맞은 문항수	능력	능력 모수 추정오차	진점수
1	14	-0.366	0.081	11.55
2	12	0.295	0.144	14.907
3	15	0.472	0.111	15.719
4	11	-0.54	0.09	10.602
5	8	-1.022	0.188	8.029
6	18	0.988	0.221	17.886
7	14	-0.37	0.078	11.527
8	7	-1.503	0.147	5.902
9	6	-1.509	0.15	5.878
10	16	0.5	0.116	15.843
11	15	0.29	0.146	14.886
12	8	-1.143	0.151	7.439
13	8	-1.115	0.161	7.574
14	8	-0.855	0.2	8.891
15	14	0.02	0.204	13.574
16	15	0.706	0.185	16.733
17	14	0.407	0.112	15.424
18	12	-0.314	0.109	11.833
19	17	0.957	0.222	17.763
20	16	0.551	0.131	16.069

우리는 학습자의 학습 스타일을 확인하기 위해서 서버에 접속하는 학습자들의 서버로그 정보를 이용하였다. 데이터 마이닝의 연관성 알고리즘을 적용하기 위해서 IP 별로 데이터를 모은 다음 (표 6)과 같이 학습자별로 데이터를 정리했다. 준비된 데이터에 연관성 알고리즘을 적용한 결과 (표 8)의 결과를 얻었다.

표 8. 로그 데이터에 연관성 알고리즘 적용

Table 8. The Result that Apply Association Algorithm

학생	연관성 분석 결과
1	animation=T 3 ==> GVA=F BOARD=F 3 acc:(0.9484)
2	animation=T 4 ==> GVA=F 4 acc:(0.94146)
3	chat=T 4 ==> BOARD=F 4 acc:(0.93571)
4	animation=T 4 ==> chat=F 4 acc:(0.92927)
5	GVA=T 4 ==> BOARD=F 4 acc:(0.91123)

첫 번째 학생은 동영상상을 주로 이용하고 동영상상을 이용할 경우 게시판과 GVA 강좌를 활용하는 것을 알 수 있다. 두 번째 학생은 동영상상을 이용하면 주로 GVA를 같이 이용한다. 세 번째 학생은 채팅과 게시판 이용률이 높은 학생이다. 네 번째 학생은 동영상상을 활용하면서 채팅도 함께 사용하는 성향을 갖고 있다.

(표 9)는 학습자별로 문항반응 이론에 의한 학습결과에 의한 평가와 각 학습자를 위한 학습 스타일 개선 사항을 제시하였다.

표 9. 이러한 평가 결과 및 조언

Table 9. Result of Assessment and Recommendation

학생	속성 (문항반응 이론)	학습 스타일 개선사항
학습자 1	중	게시판을 활용
학습자 2	상	없음.
학습자 3	상	없음.
학습자 4	중	GVA 및 게시판 사용을 권장
학습자 5	하	동영상 자료를 활용하기 바람.

#### IV. 결론

본 논문에서는 우리는 문항분석 이론과 웹 마이닝 기술을 이용하여 이러닝 학습자를 위한 개인별 학습 평가를 제공하고 학습 성향을 분석하여 학습 개선하도록 하는 시스템



을 제안하였다. 제안한 방법은 시험문제의 공정성과 적정성을 IRT를 이용하여 제공할 수 있었다. 또한, 웹 마이닝 기술을 사용함으로써 학습자가 어떤 분야에 관심이 있는지 그리고 어떤 타입의 학습을 선호하는지 등을 분석할 수 있었다.

IRT를 이용한 학습자에 대한 분석결과를 가지고 피드백을 통해서 해당 학습자가 취약한 단원을 찾아서 집중적으로 피드백을 함으로써 학습능력을 개선하였다. 또한, 학습자가 온라인으로 강의를 청취하거나 자료를 다운 받아 공부하는 등의 학습과정을 로그 마이닝을 이용하여 학생별로 학습 성향을 분석하였다. 이 결과를 학습 능력이 좋지 않은 학생에게 적용하여 학습 패턴을 바꾸는 추천시스템을 통해서 학습의 능력을 향상시킬 수 있었다. 향후 초중고 대학교에 제안한 학습 모델을 도입하면 학습 능률 향상에 도움이 될 것으로 기대 한다.

J.H. Suh, "A personalized recommendation procedure for Internet shopping support" Electronic Commerce Research and Applications Vol.1(2002), pp.301-311.

- [7] SPERO: Tele-Informatics System for Continuous Collection, Processing, Diffusion of Material for Teacher Training in Special Education. <http://www.image.natua.gr/spero>.
- [8] 성태제, "문항제작 및 분석의 이론과 실제", 학지사, 1996.
- [9] 박성준, 김주연, 김영국, "분산 이기종 인터넷 쇼핑물 환경에서의 벡터 모델 기반 개인화 서비스 시스템", 한국정보과학회 논문집, 제8권 제2호, pp.206-218, 2002.
- [10] Pieter Adriaans, Dolf Zatinge, "Data Mining", Addison-Wesley Professional, 1996.
- [11] 김석훈, 김은수, "웹로그 마이닝을 이용한 개인화 광고 서비스 기법", 컴퓨터교육학회논문지, 2005

### 참고문헌

- [1] 김정숙, 이희영, "동적인 학습 내용 구성과 실시간 과제물 평가 기능을 가진 e-Learning 시스템의 설계 및 구현", 한국컴퓨터 정보학회 논문지, 2005.11.27.
- [2] M. Blochl, H. Rumershofer, and W. Wob. Individualized e-learning systems enabled by a semantically determined adaptation of learning fragments. In Proceeding of the 14th international workshop on Database and Expert Systems Applications, pages 640-645, 2003.
- [3] S.K.Chang. A chronobot for time and knowledge exchange. In Tenth International Conference on Distributed Multimedia Systems, pages 3-6, San Francisco Bay, CA, Sep. 2004.
- [4] X. Fu, J. Budzik, and K.J. Hammond. Mining navigation history for recommendation. In Proceedings of the 5th international conference on Intelligent user interfaces, pages 106-112, 2000.
- [5] J.B. Schafer, J.A Konstan and J. Rieer, "E-mining and Knowledge Discovery, Vol. 5 (2001), pp.115-153
- [6] J.K. Kim, Y.H.Cho, W.J.Kim, J.R. Kim. and

### 저자 소개



김천식

1995년 안양대학교 전자계산학과 (공학사)  
 1997년 한국외국어대학교 컴퓨터 및 정보통신공학과 (공학석사)  
 2003년 한국외국어대학교 컴퓨터 및 정보통신공학과 (공학박사)  
 2000년~2003년 경동대학교 정보통신공학부 교수  
 2004년~현재 안양대학교 디지털미디어학부 교수  
 <관심분야> 데이터베이스, 데이터마이닝, 유비쿼터스, 텔레매틱스, TPEG, DMB, 홈네트워킹, e-Learning



정명희(정회원)

1989년 서울대학교 계산통계학과 졸업.  
 1991년 U. of Texas, Austin  
 1997년 U. of Texas, Austin 산업공학과 박사학위  
 2006년 현재 안양대학교 디지털미디어공학과 교수.  
 <주관심분야 : e-learning, 영상, 멀티미디어>