

# 데이터마이닝기법을 이용한 인터넷교육 맞춤 시스템

이진호, 류준석, 김응모  
성균관대학교 컴퓨터공학과  
e-mail:gnori@skku.edu

## Internet Learning customized System for using Data Mining Techniques

Jin-Ho Lee, Joon suk Ryu, Ung mo Kim  
Dept of Computer Science, Sungkyunkwan University

### 요 약

정보통신 기술의 발전은 우리의 생활 전반에 걸쳐 빠르게 흡수되며 급속히 진행되고 있다. 특히 교육의 패러다임이 변화됨에 따라 오늘날 인터넷을 기반으로 한 가상교육의 형태는 학생들로 하여금 더 많은 지식 습득 기회를 제공한다.

본 논문에서는 인터넷상의 교육 시스템에서 개인의 정보를 수집하고, 개인별 교육성향을 분석하여 개인별로 적절한 서비스를 제공하기 위한 연구를 하였다. 데이터 마이닝 기법 중 연관규칙과 클러스터링 협업 필터링을 이용하여 학습자의 교육성향을 파악할 수 있다. 이를 마케팅에 적용한다면 학습자의 선호도를 상승시키고 해당 회사에 신뢰도가 높아져 이익을 증가시킬 수 있는 시스템으로 활용될 수 있다.

### 1. 서론

정보기술의 발달과 함께 인터넷 교육의 시장이 커지고 있다. 인터넷교육은 시간과 공간의 제약을 받지 않고 언제 어디서나 교육이 가능하고, 비용이 저렴하며, 쌍방향 커뮤니케이션과 데이터베이스에 저장된 고객별 구매 데이터와 이력 데이터 등의 활용을 통해 고객의 요구를 신속하게 파악함으로써 차별화된 서비스를 제공할 수 있다는 장점을 가지고 있다. 하지만, 기업 측면에서는 기업 간 경쟁이 심화되고 고객 욕구가 다양해짐에 따라 가격할인이나 판촉경쟁만으로는 한계를 느끼고 있으며, 완벽한 고객정보의 수집 및 분석을 통한 맞춤 서비스 제공, 글로벌 서비스 등 차별화전략을 통한 고객 만족도 제고가 기업 생존의 필수 조건이 되어 가고 있다.

그러나 기존 연구의 대부분이 교육의도와 이용만족, 재교육 의도 등에 대해 설문 조사 등의 방법을 채택하고 있으며, 수강 데이터 및 고객정보 데이터 등의 데이터를 기반으로 한 인터넷교육시스템에서의 고객에 대한 연구는 부족한 상황이다.

지금까지의 연구는 CRM(Customer Relationship Management)에서 e-비즈니스 시장에서의 수익 측면만 고려하거나 데이터베이스 구축과 같은 기술적 측면에 많이 치중되어 있었다. 인터넷 쇼핑물의 고객 충성도 제고나 수익률 향상을 위한 전략적 접근으로서의 데이터마이닝

같은 분석 기법들에 치중해야 할 때이다.

본 연구에서는 데이터베이스에 축적된 주문 데이터와 고객 데이터의 과학적 분석을 통해 개별 학습자의 구체적 성향과 요구를 파악하고, 각각의 학습자에게 차별화된 추천 서비스를 제공할 수 있는 시스템을 제안하고자 한다.

### 2. 기존 연구 및 이론적 배경

인터넷교육은 흔히들 E-learning으로 불려진다. e-Learning이란 Electronic Learning의 약자로 전기적 신호를 매개로 강의를 제공하는 원격 교육방법을 말한다. 처음 원격으로 제공된 교육은 Text 형태로 시작되어 현재는 동영상과 음성도 함께 제공되는 멀티미디어 형태로 발전하였다. 또한 다양한 콘텐츠가 등장하였다. 현재 교육자의 실력까지 고려되어 학습 내용과 학습 진도를 조절해주는 등의 더욱 다양한 콘텐츠의 서비스가 요구되고 있다.[2]

효율적인 인터넷교육시스템을 구축하기 위해서 학습자의 개인정보, 수강정보 등 다양한 형태의 정보가 데이터베이스에 기록된다. 이 정보를 데이터웨어하우스로 통합하여 향후 의사결정에 도움을 주는 분석결과를 제공하고, 데이터베이스 마케팅과 전략기획을 위한 자료로 활용한다. 또한, 데이터마이닝을 통해서 온라인상에서 제공하는 서비스의 교육 콘텐츠간의 연관관계를 밝혀내어, 적절한 콘텐츠가 적절한 서비스와 함께 학습자에게 제공되고 있는지의

여부를 밝혀내게 될 수도 있다.

### 3. 데이터마이닝을 이용한 시스템 설계

#### 3.1 처리 단계

전체 시스템 구성은 전처리, 패턴분석, 맞춤형 페이지 생성의 세 모듈로 구성한다. 개인별 정보수집 및 처리과정을 단계별로 정리하면 첫째, 학습자의 기본 신상정보를 수집. 둘째, 수강신청 정보 기록. 셋째, 학습자의 로그 데이터 추출. 넷째, 패턴 분석 및 성향 파악. 마지막으로 학습자 개별 맞춤 콘텐츠 제공이 있다.

#### 3.2 전처리기

전처리기는 로그파일에서 필요한 정보를 추출, 사용자별로 로그를 분류, 개인별 기록을 세션으로 분류 그리고 패턴분석기에 사용될 트랜잭션 분류의 4가지 서브모듈로 구분할 수 있다.

#### 3.3 패턴 분석기

인터넷교육시스템의 개인별 맞춤서비스 콘텐츠 제공의 핵심인 패턴분석기는 실질적인 학습자의 성향을 파악하는 데이터마이닝 모듈로써 전처리 시스템의 처리결과로 구축된 데이터웨어하우스를 기반으로 처리 한다.

##### 3.3.1 연관규칙

데이터 마이닝 모듈인 패턴분석기에는 전처리기의 트랜잭션에 포함된 접근한 웹페이지가 항목으로 입력된다. 이 웹 페이지에 연관규칙(Association Rule)을 적용하여 접근한 웹 페이지들간의 패턴지식을 찾아내고 패턴 데이터베이스에 저장한다.

전처리과정을 통해 나온 데이터는 다음과 같은 방법으로 분석하여 페이지 항목 간 연관관계를 계산한다. 지지도와 신뢰도를 수식으로 표현하면 다음과 같다.

$$\text{supp}(R) = \frac{\text{subt}(XUY)}{\text{전체트랜잭션개수}}$$

$$\text{conf}(R) = \frac{\text{supp}(XUY)}{\text{supp}(X)}$$

인터넷교육 시스템 환경은 학습자의 학습 니즈를 반영하고 효율적으로 학습을 수행할 수 있게 방향을 제시해 줄 수 있어야 한다. 시스템은 개인의 기본 정보와 학습 정보를 가지고 교육 콘텐츠간의 지지도와 신뢰도를 구한다. 학습자가 새롭게 수강신청 및 학습 평가를 할 경우 해당 교육 콘텐츠와 가장 신뢰도가 높은 교육 콘텐츠를 같이 프 로세싱하여 개인에게 시간 및 비용을 줄여주고, 올바른 판단을 할 수 있도록 콘텐츠를 추천해준다.

##### 3.3.2 클러스터링에 의한 학습자 유형별 협업 필터링

Clustering은 학습자가 자발적으로 제공한 정보를 이용하여 사용자를 비슷한 선호도를 가진 집단으로 나누어 그 집단 내에서 서로에게 추천하는 방식을 사용한다.

이 프로세스는 학습자 성별, 성적, 학년, 난이도에 따른 협업 필터링을 이용해 구성된다. 학습자 입력의 데이터를 이용하여 개인화 된 패턴을 형성한다.

#### 3.4 맞춤형 페이지 생성기

학습자가 e-learning에 접근하였을 때 패턴 데이터베이스에 저장된 규칙을 적용하여 학습자의 성향에 맞는 교육 콘텐츠를 웹 페이지에 동적으로 구성하여 제공한다.

개인별 콘텐츠 서비스 시스템은 패턴분석시스템의 결과를 유도하고, 이를 어디에 어떻게 적용하며, 또한 이러한 패턴분석의 결과로 나온 데이터를 학습하여 새로운 가중치를 적용하는 기능을 한다.

### 4. 결론 및 향후 연구과제

인터넷의 발전에 따라 교육 연관규칙을 이용한 데이터 마이닝 기술을 학습자의 웹 페이지 접근 데이터 분석에 적용하기 위한 데이터 처리 시스템과 생성된 규칙을 이용한 활용법을 소개하였다.

제안한 데이터 마이닝 기법은 e-learning 시스템을 운영하는 사업자의 입장에서 지속적으로 동적인 학습자의 성향을 파악할 수 있는 기술이다. 점차 e-learning의 요구가 증가하고 있으며 이러한 시스템 기술의 방법이 다양해지고 있다. 이에 따라 데이터마이닝 기술 적용의 필요성이 강조되고, 여기에 맞는 학습자 교육 콘텐츠 서비스 기술의 지속적인 확장 연구가 필요하다.

#### 참고문헌

- [1] 강제정, 양은정. "인터넷 쇼핑몰 시스템 만족과 재구매 영향 요인에 관한 연구". 한국경영경제연구. 제1권 제2호. pp.19~43. 2001.
- [2] 고완기, 고석용. "고객관리를 위한 데이터베이스 마케팅 정보시스템에 관한 연구". 한국OA학회 논문집. 제4권 제2호. pp.88~94. 1999.
- [3] Bamshad Modasher, Robert Cooley and Jaideep Srivastava, "Automatic Personalization based on Web Usage Mining", Communications of the ACM, 43(8), pp. 142-151, 2000.
- [4] Hughes, M.. "Strategic Database Marketing. 2nd Ed". New York: McGraw-Hill, 2000.

이 논문은 2009년도 정부(교육과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임 (No. 2009-0075771).