한국컴퓨터정보학회지
제21권 제1호, 2013. 6.

빅데이터와 교육

• 최제영(영동대학교 임베디드소프트웨어학과)
• 박충식(영동대학교 스마트IT학부)

I. 서론

네트워크와 유선 또는 무선으로 연결된 온라인 기기의 증가와 디지털 공간상의 활동의 일상화로 발생되는 인터넷 홈페이지 접속 기록, 온라인 서비스 이용 기록, 인터넷 검색사이트의 검색 기록, 기업에서 축적되는 소비자 정보 및 거래정보, 인터넷에 연결가능한 자동차 · 산업시설 · 도로 · 건축물 등에 내장된 센서와 모바일 기기 등의 사물 대 사물 간의 사물통신(Machine to Machine 또는 Internet of Things)의 데이터, 스마트폰과 같이 인터넷에 연결 가능한 기기를 이용하여 블로그 · 트위터 · 페이스북 등과 같은 사회적 미디어(또는 사회 연결망 서비스, SNS)에서 생성되는 데이터 등의 디지털 데이터가 폭발적으로 증가하고 있다[1][2].

데이터의 저장 비용과 처리 비용이 저렴해져서 이전에는 버려져야 했을 데이터를 수집하여 처리할 수가 있게 되고 클라우드 컴퓨팅(cloud computing)의 확산이 비용을 낮추고 여러 기술적 장애를 제거하여 준다. 또한 데이터들을 원래의 저장장치에서 한 곳으로 이동하느라 시간과 비용을 낭비하지 않고 데이터를 분석 · 처리하는 기술이 개발되고 있다.

매킨지 세계 연구소[3]는 빅데이터를 전형적인 데이터베이스 소프트웨어 도구가 획득, 저장, 관리, 분석할 수 있는 한계를 초과하는 크기의 데이터 집합(dataset)으로 정의하였고 현재 많은 분야에서 수십 TB(테라 바이트) 내지 수 PB(페타 바이트)의 범위이지만 기술이 발전함에 따라 빅데이터로 간주되는 데이터 집합의 크기도 증가한다고 밝혔다.

Oracle[4]은 빅데이터의 3가지 유형으로 전통적인 기업 데이터, 기계의 센서가 생산하는 데이터, 사회 연결망 데이터를 들고 빅데이터를 규정짓는 특징으로 양(Volume), 속도(Velocity), 다양성(Variety)에 가치(Value)를 추가하였다.

스마트교육을 기술적 관점으로 바라보면 초고속 인터넷, 모바일 기기, 클라우드 컴퓨팅 환경을 이용하여 스마트 기술을 교육에 적용함으로써 개인의 취향과 적성에 따라 교육 프로그램을 선택 및 활용할 수 있는 교육이라고 할 수 있다. 이에 따라 스마트 교육은 21세기 학생들의 학습 역량을 강화하기 위해서 도입하게 된 지능형 맞춤 학습 체제로서 교육환경, 교육내용, 교육방법 및 평가 등의 교육체계를 스마트세대에 맞춰 새롭게 혁신하는 것을 말한다[5][6]

스마트교육에서 학생들과 교사들의 모든 교육활동은 디지털화되어 모니터링되고 축적될 수 있기 때문에 공공교육의 규모로 볼 때 스마트교육 활동정보는 전형적인 빅데이터라고 할 수 있다[7]. 이러한 교육 빅데이터는 개인정보보호를 전제로 하여 현재 빅데이터 처리 기술을 활용하면 각 학생에게 개별화학습을 제공하는 것은 물론이려니와 교육정책, 교육체제, 교육방법 등에서 최적의 전략을 구사할 수 있는 데이터로 활용될 수 있기 때문에 데이터에 기반을 둔 교육혁명을 이룰 수 있는 계기가 될 수 있다.

스마트교육의 빅데이터를 제대로 활용하기 위해서는 온톨로지 기술을 이용하여 학습콘텐츠를 통한 모든 학습활동이 정교하게 수집되고, 분석되어 활용될 수 있는 교육 빅데이터 프레임워크가 마련되어야 한다. 교육 빅데이터 프레임워크는 스마트교육을 위한 표준 제정을 통하여 모든 학습콘텐츠와 학습활동자료가 수집될 수 있도록 해야 하며, 이렇게 수집된 정보는 세심한 개인정보보호 원칙하에 대용량 처리 서버에 관리되어야 하고, 다양한 교육 빅데이터 활용 아이디어를 적

용할 수 있도록 교육 빅데이터 처리 환경을 많은 연구자들에게 오픈 API 형태로 제공할 필요가 있다.

또한 교육 빅데이터의 다양하고 성공적인 결과물들은 교육 일선현장과 교육정책기관에서 잘 활용될 수 있도록 교수자들에게는 사용하는 학습관리시스템에 플러그인 형태로 제공하고, 정책집행자들에게는 교육지표모니터링 시스템의 형태로 제공할 필요가 있다. 스마트교육의 시행과 발맞추어 잘 확립된 교육 빅데이터 활용 인프라를 구축할 수 있는 단계적인 정책의 시행이 필요할 것으로 보인다.

본고에서는 다음의 순서로 스마트교육 환경에서의 빅데이터에 대하여 논의한다. II장에서 스마트교육의 개념, 시도 사례, 구성을 살펴보고 III장에서 스마트교육 서비스에서 발생할 수 있는 교육 활동정보의 종류와 크기를 추정해 보고 이를 활용하기 위한 빅데이터 처리기술, 온톨로지 기술, 사생활과 윤리, 빅데이터 활용 프레임워크를 살펴본다. IV장에서 스마트교육 환경에서 발생하는 빅데이터의 활용방안을 생각하고 마지막으로 V장에서 마무리한다.

II. 스마트교육

1. 스마트교육의 개념

스마트교육은 기존의 IT를 기반으로 하는 이러닝과 오프라인 교육을 혼합하되 최근 급격히 부상된 스마트기술을 기반으로 새로운 교육서비스로 정의한다. 스마트기술은 모바일 기기들의 발달로 이루어진 정보기기의 이동성과 대규모 자료 처리능력을 기반으로 하는 지능화로 특징지을 수 있다. 이동성은 개인화된 모바일 기기와 고속 통신기술, 클라우드 컴퓨팅 기술을 기반으로 개인 정보처리 능력이 극대화되어 생긴 IT환경 변화이다. 지능화는 전례 없이 대규모 자료를 수집, 분석할 수 있는 빅데이터 기술을 기반으로 다양한 기계학습과 데이터 마이닝기술을 활용한다[6].

이러한 스마트기술로 인하여 학습자에게 교육의 최대 목표인 학습자 개개인에 적합한 개인화된 학습서비스를 제공할 수 있는 스마트 교육환경을 제공할 수 있게 되었다. 기존의 컴퓨터를 활용한 교육은 교실에서 풍부한 멀티미디어를 활용한 학습 콘텐츠를 제공하거나 원격에서 온라인으로 교육을 제공하는 형태로 이루어졌으나 스마트기술로 인하여 모든 장소에서 교육기기를 활용하여 단순한 멀티미디어 교육자료 제공을 넘어서 개인화된 피드백을 제공할 수 있게 되었다.

개인화된 정보기기를 통하여 장소에 무관하게 온라인으로 학습콘텐츠에 접근가능하게 되고, 세부적인 학습활동 정보를 대량으로 수집하고 분석할 수 있게 됨으로써 교실 현장교육과 원격 온라인 교육이 변화할 수 있는 계기가 마련되었다. 개인화된 모바일기기는 기존의 교실이나 가정에서의 PC환경과는 상상 이상으로 완연히 다른 변화를 초래한다. 충분한 처리속도와 통신기능을 가지며 휴대가능하고 충분한 화면 크기와 쉬운 사용자 인터페이스를 가진 개인화된 모바일기기는 책과 노트, 그리고 PC를 대신하여 언제 어디에서나 하나의 학습도구를 사용할 수 있게 함으로써 교실 수업과 온라인 교육의 구별을 없애고 완전히 새로운 학습자 교육환경을 제공한다.

이러한 학습자 교육환경은 다양한 학습콘텐츠를 제공하고 학습자의 활동정보를 수집하고 분석하여 피드백할 수 있는, 상상 가능한 다양한 교육방법의 아이디어를 실현할 수 있는 기본적인 바탕이다. 더불어 클라우드 컴퓨팅과 빅데이터 기술은 이러한 학습자 교육환경과 더불어 개인화된 학습을 더욱 용이하게 하는 스마트교육 환경을 구성한다.

2. 스마트교육의 시도 사례

2.1. 한국 정부의 스마트교육

2011년 6월 29일, 국가정보화전략위원회와 교육과학기술부는 청와대에서 이명박 대통령에게 「스마트교육 추진전략」을 보고하였다[8].

최근 사회연결망 서비스(SNS), 클라우드 컴퓨팅 등 IT기술의 발전과 스마트기기의 급속한 확산에 따라 개인 특성에 맞는 차별화되고 창의적인 학습 수요가 증가에 따라 스마트교육의 필요성이 제기되었다.

국가정보화전략위원회와 교육과학기술부는 공동으로 디지털교과서 개발 및 적용, 온라인 수업 활성화, 온라인을 통한 학습 진단 · 처방 체제 구축, 교육콘텐츠 자유 이용 및 안전한 이용 환경 조성, 교원의 스마트 교육 실천 역량 강화, 클라우드 교육 서비스 기반 조성 등의 「스마트교육 추진 전략」을 수립하고, 2015년까지 국가 교육경쟁력 세계 10위권, 2025년 세계 3위 진입을 목표로 설정하였다.

2.2. 아마존(Amazon)의 킨들 텍스트북 렌털(Kindle Textbook Rental) 서비스

전자책 시장 1위 사업자인 아마존은 2011년 7월부터 종이책 대비 최대 80% 저렴한 가격으로 전자책 형태의 대학 교재를 대여해 주는 킨들 텍스트북 렌털 서비스를 개시하였다.

킨들 전자책 단말은 물론 PC, 맥, 아이패드, 아이폰, 윈도우 폰7, 블랙베리, 안드로이드 단말 등 다양한 단말기에서 킨들 애플리케이션을 다운로드받아 사용할 수 있다.

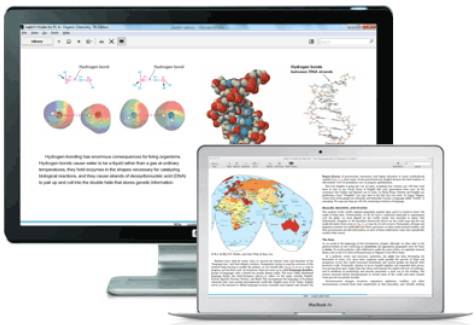

그림 1. 아마존의 킨들 텍스트북 렌털 서비스

2.3. 애플 스마트교육 사업

애플은 2012년 1월 19일 뉴욕 구겐하임 미술관에서 아이패드를 통해 디지털교과서를 구현할 수 있는 아이북스2(iBooks 2)를 공개하였다[9][10].

기존의 아이북스를 공공교육의 교과서로 사용하기 위하여, 애플은 아이북스를 확장한 아이북스2, 아이북스 오서, 아이튠즈유로 구성된 일련의 디지털교과서 서비스 체계를 확립하였다.

미국 교과서 시장의 90%를 점유하고 있는 피어슨, 맥그로-힐, 호톤 미플린하코트 등의 주요 출판 회사와 제휴해, 디지털교과서를 저렴한 가격(14.99달러)으로 제공한다.

애플 아이북스2는 애플의 디지털교과서 뷰어(viewer)로, 도표와 영상, 오디오, 애니메이션 등을 구현하는 등 학습에 유용한 쌍방향 디지털 교과서 플랫폼 기능을 갖추었다. 기존의 아이패드용 애플 아이북스를 확장하여 교과서로 활용할 수 있도록 하였다.

그림 2. 애플 아이북스2

아이북스2는 전자책 파일 형식으로 쓰이고 있는 EPUB3.0이나 아마존 전자책 파일 형식인 킨들 포맷8(KT8)처럼 상호작용이 가능하고 멀티미디어를 지원하는 형태로 제작된다. 애플 아이북스2는 애플 앱스토어를 통하여 학습콘텐츠를 구매할 수 있도록 고유한 유통 체계를 제공한다.

애플 아이북스 오서는 애플의 출판기술을 바탕으로 아이북스2 형태의 전자책을 쉽게 만들 수 있도록 개발된 아이북스2 전용 전자책 저작도구이다.

아이튠즈유는 기존의 대학교육을 위한 온라인 교육시스템으로, 아이북스2와 다양한 학습 보조교재를 결합할 수 있는 간단한 형태의 코스웨어 생성도구 및 콘텐츠 유통 플랫폼 역할을 맡는다. 그동안 교육 콘텐츠를 유통하는 장터였지만 별도로 분리돼 운영되지는 않았고, 애플이 교육 시장에 진출하며 아이튠스유는 애플의 교육 사업에서 허브 역할을 맡는다.

2.4. 미국 정부의 스마트교육

미국 오바마(Obama) 행정부가 2012년 2월에 향후 5년 내에 미국 전역에서 디지털 교과서 도입을 추진하겠다고 선언하고 정부 및 관련 업계 전문가들로 구성된 디지털교과서 협의체(The Digital Textbook Collaborative)를 통해 디지털교과서 도입 계획을 수립하였다.

디지털교과서 협의체를 통해, 디지털교과서 플레이북을 발표하여 교수자 및 관리자가 미국 전 지역의 초·중·고(K-12) 학생에게 풍부한 디지털학습경험을 제공해 줄 수 있도록 안내해 주는 가이드북으로 제공하고 학교나 교실, 가정에서 디지털학습환경을 위해 필요한 네트워크 인프라 정보와 단말기 정보를 제공한다.

그림 3. 디지털교과서 플레이북

디지털교과서 플레이북[11]에는 디지털학습환경으로 전환(Making the transition to digital learning), 학교 및 교실에서의 네트워크 환경(Connectivity at school), 학교 밖(가정 및 공동체)에서의 네트워크 환경(Connectivity beyond the school gates), 단말기 활용의 고려사항(Device perspectives)에 이르는 네 가지 사항에 대한 가이드가 언급되었다.

2.5. 일본 정부의 스마트교육

2012년 4월 5일에 일본 디지털교과서 교재 협의회(Digital Textbook and Teaching, 이하 DiTT)가 디지털교과서를 사용하여 학습하는 교육을 목표로 한 'DiTT 정책 제언 2012'를 발표하였다.

2015년까지 1,000만 명의 학생들에게 디지털 교과서 서비스가 가능하도록 '디지털교과서 실현을 위한 제도개정', '디지털교과서 보급을 위한 재정 조치', '교육정보화 종합 계획 수립 · 실행'의 3가지 제안 내용을 담고 있다.

2.6. 구글 세상보기(Google Currents)

구글은 안드로이드 기반의 모바일 기기와 아이폰 · 아이패드 등에서 손쉽게 웹 콘텐츠를 읽을 수 있는 애플리케이션인 구글 세상보기 서비스를 제공한다. 구글 세상보기는 구글 플레이와 애플 앱스토어에서 각각 다운받을 수 있다.

미국에서만 400개 출판사의 간행물과 1만4,000개 이상의 개인 제작 간행물이 등록되어, 사용자가 편한 언어를 선택할 수 있고, 구글 번역 기능이 추가되어 있어 외국어 콘텐츠도 읽는 것이 가능하다.

새로운 콘텐츠들이 항상 자동으로 업데이트되며, 오프라인 다운로드를 통해 기차나 비행기 등으로 이동할 때 잡지를 선택해 다운로드할 수 있다.

3. 스마트교육의 구성

3.1. 스마트교육 서비스 프레임워크

스마트기술의 이동성과 지능화를 기반으로 한 스마트교육의 원활한 서비스를 위한 참여자와 시스템의 구성은 스마트교육 서비스 프레임워크로 그려 볼 수 있다.

통합적 스마트교육 서비스 프레임워크에는 학습자, 강의자, 학습콘텐츠(이러닝 콘텐츠) 교육분석가, 행정가가 이해관계자로 참여한다. 통합적 스마트교육 서비스 프레임워크는 다양한 전자책 뷰어와 학습콘텐츠(이러닝 콘텐츠) 앱스토어, 전자책 앱스토어, 학습관리시스템(LMS), 교육 빅데이터 처리센터까지 포함한다.

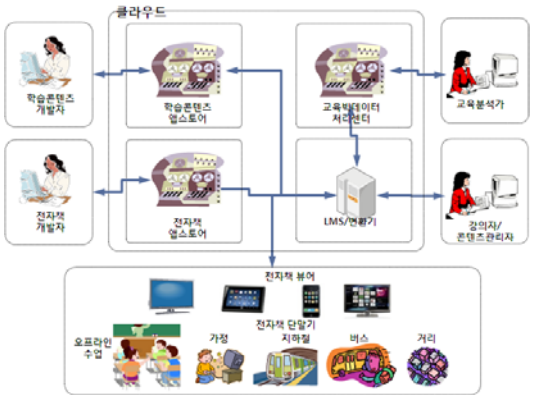

그림 4. 스마트교육 서비스 프레임워크

강의자(교사)는 학습관리시스템을 활용하여 해당 수업의 학습자에게 필요한 교과서(전자책)와 학습콘텐츠를 구성하여 수업을 위한 교재를 구성한다. 전자책은 전자책 앱스토어를 통하여, 학습콘텐츠는 학습콘텐츠 앱스토어를 통하여 유료 · 무료로 활용하고, 개인적으로 구성한 학습콘텐츠도 활용할 수 있도록 한다.

학습(이러닝) 콘텐츠와 전자책 콘텐츠는 학습활동을 위하여 상호보완적이고 스마트교육 서비스에서 관련 산업의 활성화를 위하여 공히 스마트교육에서의 역할이 필요하다. 전자책 뷰어는 학습자의 수준과 취향에 따라 선택될 수 있도록 학습콘텐츠 개발자나 전자책 개발자들에 의하여 개발되도록 해

야 한다.

Moodle[12], OpenClass[13], Blackboard[14] 등과 같은 학습관리시스템[15]은 조직적이고 체계적으로 교재를 구성하고, 학습자의 학습활동 정보를 수집하고 관리할 수 있기 위하여 필요하다. 학습자의 학습활동정보는 강의자의 선택에 따라 교육 빅데이터 센터에서 분석되어 학습관리시스템은 교사와 학습자에게, 직접 또는 간접으로 피드백을 제공해야 한다.

3.2. 스마트교육 서비스 프레임워크의 표준화

스마트교육 서비스가 체계적인 구조를 갖추고 관련 산업의 적극적 참여가 가능하도록 하기 위해서는 표준화 작업이 선행되어야 한다. 스마트교육을 위한 표준화는 전자책 표준, 이러닝 표준, 저작권 보호 표준, 개인정보 보호 표준, 유통 표준, 통신 표준, 빅데이터 표준 등에 관련된다.

스마트교육의 표준화를 위해서는 관련 표준들이 다수 존재하고 있으므로 관련 표준을 적극 수용하고 전체 서비스에 필요하지만 부족한 표준을 확장하거나 새로운 표준을 개발하는 것이 효율적인 서비스 표준화 전략이다. 표준화와 더불어 표준화를 수용하고 있는 관련 오픈 소스 소프트웨어를 활용하는 것이 서비스 시장을 빠르게 성장시키고 기존의 시장에 쉽게 적응할 수 있는 전략이 될 것이다.

III. 스마트교육에서의 빅데이터와 처리 기술

1. 스마트교육에서의 빅데이터

1.1. 교육 빅데이터의 종류

전통적인 교육현장에서는 학습활동과 관련된 대부분의 데이터가 시험성적이나 학생생활기록부에 기록된 내용을 제외하면 교육이 끝난 뒤에 모두 사라진다.

인터넷, 모바일 기기, 클라우드 환경을 활용하여 스마트기술을 교육에 적용한 스마트교과서를 이용하는 학습활동에서는 모든 것이 디지털 데이터로 저장된다.

학생이 스마트교과서를 활용하는 과정에서 생성되는 빅데이터와 관련된 활동의 예는 다음과 같다:

- 학생이 패드형 단말기를 사용하여 온라인 접속한 뒤 온라인 학습을 진행함
- 과제를 제출하거나 온라인 퀴즈를 풀고 온라인 설문조사에 응함
- 필요한 학습 콘텐츠를 다운로드하여 이용하고 다른 기기에 콘텐츠를 동기화하거나 다른 학생과 콘텐츠를 공유함
- 학습과 관련된 검색을 하거나 메모를 남김
- 대화방에서 다른 학생과 대화를 하거나 사회 연결망 서비스(SNS)를 이용하여 글을 게시함
- 각 학생의 학습목표, 학습 취향, 기술 및 능력, 성취도 등의 학습자정보가 갱신됨

1.2. 교육 빅데이터의 양

한 학생이 하루에 수행하는 각 활동의 평균 개수를 보수적으로 추정하여 표 1에 나타내었다[7].

2011년 국가통계포털[16]을 참조하여 전체 초중고교 학생수는 7×10^6명으로, 학교당 학생수는 600명으로 어림하여 하루에 각 학교와 전체 초중고교에서 발생하는 데이터 생성 건수를 표 1에 함께 표시하였다.

표 1. 하루당 데이터 생성 건수

활동	학생당	학교 (600명)당	전국 (7×106명)
로그인	10	6,000	70,000,000
온라인 학습	20	12,000	140,000,000
과제 제출	4	2,400	28,000,000
온라인 퀴즈	10	6,000	70,000,000
온라인 설문조사	2	1,200	14,000,000
콘텐츠 다운로드	10	6,000	70,000,000
콘텐츠 동기화	8	4,800	56,000,000
콘텐츠 공유	2	1,200	14,000,000
메모 기록	4	2,400	28,000,000
대화방 대화	10	6,000	70,000,000
검색	10	6,000	70,000,000
SNS 글 게시	10	6,000	70,000,000
합계	90	54,600	630,000,007

표 1에 따라 학생 1명이 하루에 100건의 활동을 한다면 전국 초중고 700만 명의 학생이 하루에 생성하는 데이터는

모두 7억 건이고, 1년에 생성하는 데이터는 2,555억 건에 달하게 된다. 학생 1명이 하루에 1MB의 데이터를 생성한다면 전국 초중고 700만 명의 학생이 하루에 생성하는 데이터는 모두 7.0TB이고, 1년에 생성하는 데이터는 2.56PB에 달하게 된다.

2. 교육 빅데이터 처리기술

인터넷 상에서 점점 더 많이 이루어지는 여러 분야의 활동에서 발생하는 빅데이터를 활용하려는 노력과 마찬가지로 온라인 교육 분야에서 발생하는 빅데이터를 활용하여 학생들의 교육 성과와 초중등 교육 시스템의 생산성을 향상시키기 위한 시도로 학습 애널리틱스(learning analytics)와 교육 데이터마이닝(educational data mining)이 연구되고 있다[17][18].

교육 데이터마이닝은 데이터에서 새로운 패턴을 찾아 새로운 알고리즘과 모델을 개발하려 하고 학습 애널리틱스는 교육 체제에서 알려진 예측 모델을 적용하려 하는 면에서 차이가 있다.

표 2. 학습 애널리틱스와 교육 데이터마이닝의 비교

학습 애널리틱스	교육 데이터마이닝
정보과학, 사회학, 심리학, 통계학, 컴퓨터과학, 교육학 전체를 이해하려 함 기존의 방법과 모델을 적용하여 학습과 관련된 문제를 분석 자동화된 처리 대신에 교사의 결정을 지원	통계학, 기계학습, 데이터마이닝 학습을 작은 부분들로 분해 데이터를 분류하고 관계를 찾아 예측모델을 개발 학생에 대한 자동화된 처리 (적응 학습 시스템)

2.1. 기술의 목표

교육 빅데이터를 처리하는 기술의 목표는 다음과 같다.

- 학생, 교사 및 관리자에게 의미 있고 활용 가능한 피드백을 제공함
- 각 학생에게 적절한 개별화된 학습 활동을 추천하는데 사용함
- 새로 제시된 학습 콘텐츠에 대하여 학생의 성과를 예측함
- 각 학생은 스스로 학습 계획을 구성하고 수행함
- 많은 학생들의 학습 활동의 결과로부터 각 콘텐츠의 난이도, 유용성, 사용 편리성 등에 대한 평가가 이루어짐
- 학생과 교사의 사회연결망을 분석하고 활용함
- 교육 데이터를 이용하여 교육 이론을 검증하고 개선하는데 활용함

2.2. 교육 데이터마이닝

교육 데이터마이닝은 통계학, 기계학습, 데이터마이닝 등의 기술을 사용하여 학생의 데이터를 수집하고 분석하여 데이터 속의 패턴과 경향을 발견하고 새로운 발견을 하거나 학생들이 학습하는 방식에 대한 이론을 검증한다. 학생에 개별화된 소프트웨어로 분석하고 처리할 수 있는 작은 요소들로 학습을 분해하는 것을 일반적으로 중시한다.

데이터를 분류하고 관계를 찾는 교육 데이터마이닝 기법을 온라인 교육에서 수집된 학습 데이터에 적용하여 예측 모델을 개발한다. 예측 모델은 개별화 학습 시스템(adaptive learning system)을 구성하는데 중요한 역할을 한다.

개별화 학습 시스템에서는 예측 모델의 예측에 근거를 둔 적응과 조정으로 학생이 앞으로 경험하는 것을 변경하고 학습을 지원하기 위하여 학교 외적인 분야의 활동까지 학생에게 추천한다.

교육 데이터마이닝 연구의 목표

- 학생의 지식, 동기, 초인지, 성향 등의 상세한 정보를 고려한 학생 모델을 고안함으로써 학생의 미래 학습 활동을 예측하기
- 학습할 콘텐츠와 최적의 학습 순서를 특징짓는 도메인 모델을 구축하고 개선하기
- 학습 소프트웨어가 제공할 수 있는 여러 종류의 교육적 지원 효과를 연구하기
- 학생, 도메인(과목), 소프트웨어 교육법의 모델을 구성함으로써 학습과 학습자에 대한 과학적 지식을 발전시키기

교육 데이터마이닝 연구에서 사용되는 5종의 기법은 다음과 같다.

- 예측하기 : 데이터의 여러 상황으로부터 데이터의 한 상황에 대하여 추론하는 모델을 개발하기로서 학생의 행동(시스템 갖고 놀기, 아는데도 질문에 틀린 답하기 등)을 탐지하는 것이 예임
- 군집화(clustering) : 자연스럽게 무리 지어져 종류별로

나눌 수 있는 데이터 집합을 찾는 것으로서 학습 패턴과 상호작용 패턴에 근거를 두고 학생을 묶기와 같은 것

- 관계 발굴하기 : 데이터 집합 안의 변수들 사이의 관계를 찾아 후에 활용할 수 있도록 규칙화하기
- 연관규칙 발굴하기 : 학생이 동시에 범하는 실수를 찾는데 쓰일 수 있음
- 순서패턴 발굴하기 : 순차적으로 일어나는 사건들 사이의 관계를 포착하는 규칙을 세우는 것으로서 학생의 수행과 교과과정 사이의 연관관계를 발견하고 어떤 교육 전략이 더 효과적이고 견고한 학습을 낳는지를 발견하는데 응용됨
- 인간의 판단을 돕기 위하여 추출하기 : 데이터의 특징을 사람이 빨리 구분하고 분류할 수 있도록 데이터를 나타내기로서 시각 데이터 애널리틱스(visual data analytics)와 겹침
- 모델로 발견하기 : 현상들의 (예측하기, 분류하기 등을 통하여 개발된) 유효한 모델을 추가 분석의 요소로 사용하기로서 학습 전략의 모델 가운데 각 전략이 다양한 상황에서 어떻게 변해야 하는가를 연구하기가 한 예임

2.3. 학습 애널리틱스

애널리틱스(analytics)는 컴퓨터과학, 수학, 통계학 등의 기술을 응용하여 대량의 데이터에서 유용한 정보를 추출하는 것을 말한다. 학습 애널리틱스(learning analytics)는 정보과학, 사회학, 심리학, 통계학, 컴퓨터과학, 교육학 등의 개념과 기술을 사용한다.

학습을 요소로 분해하는 교육 데이터마이닝과 달리 전체 시스템을 이해하려 하고 자동 시스템을 추구하는 대신에 인간의 결정을 지원하고자 한다. 교사와 학교가 학생의 학습을 예측하고 능력과 수준에 맞는 교육을 할 수 있도록 지원하는 것을 목적으로 한다.

새로운 데이터 분석 방법을 개발하는 대신에 기존의 방법과 모델을 사용하여 학생의 학습과 교육기관의 교육 시스템에 영향을 주는 중요한 질문에 대답하려 하고 학생에 대한 자동화된 응답을 강조하는 교육 데이터마이닝과 달리 학습 콘텐츠를 적응시키고, 위기의 학생을 중재하고, 피드백을 제공하는 방식을 취한다.

교육 데이터마이닝의 기술 이외에 사회연결망 분석(학생과 학생, 학생과 교사 관계의 분석)도 사용한다.

교육 데이터마이닝처럼 실행가능한 분석결과를 생성하는데에 시각화가 중요하며 대개 쉽게 요약할 수 있는 형태로 데이터를 표시하는 현황판(dashboard)에 나타낸다.

2.4. 시각 데이터 애널리틱스

시각 데이터 애널리틱스(visual data analytics)는 고급의 계산 기법과 그래픽 엔진을 결합하여 데이터의 패턴과 구조를 시각적으로 표현하는 기술이다. 복잡한 시스템에서 수집된 비균질이고 동적인 데이터 집합의 패턴, 경향, 예외값 등을 드러내도록 도와준다[19].

시각 데이터 애널리틱스의 예로 Gapminder[20]는 시간에 따라 변화하는 데이터 집합을 동적인 그래픽으로 보여주고, IBM의 ManyEyes[21]는 데이터 집합을 다양한 형식으로 시각화하는 기능을 제공하는 누리집이다.

Tableau[22]는 대화식으로 데이터를 시각화해 주는 소프트웨어이며 FlowingData[23]는 막대 그래프, 선 그래프, 파이 차트 같은 단순한 그래프는 물론, 시계열 그래프, 지도 그래프, 트리맵과 대화식의 네트워크 그래프까지 그릴 수 있게 한다.

2.5. 개별화 학습 시스템의 데이터 흐름

온라인 학습 시스템은 학습자의 세세한 활동을 포착하여야 하고 데이터에 여러 가지 기법을 사용하여 학습, 교육, 교육 정책 결정 등을 개선하는데 쓰일 피드백을 제공해야 한다.

온라인 학습 시스템의 6 요소는 다음과 같다.

- 콘텐츠(content) 관리, 유지, 제공 요소는 학생에게 개별화된 콘텐츠와 평가를 제공함
- 학생 학습 데이터베이스(student learning database)는 학생이 활동함에 따라 생기는 입력과 거동을 시간과 함께 저장함
- 예측 모델(predictive mode)은 학습자 정보 시스템으로부터의 데이터(학습자 정보)와 학생 학습 데이터베이스를 결합하여 학생의 발전을 추적하고, 미래의 교과목 성적 또는 낙제와 같은 미래의 수행을 예측함
- 보고 서버(reporting server)는 예측 모델의 결과를 사용하여 여러 사용자를 위하여 현황판(dashboard)에 피드백을 제공함

- 적응 엔진(adaptation engine)은 예측 모델의 출력에 근거하여 콘텐츠 제공 요소를 조정하여 학생의 수행 수준에 따라 학습 자료를 제공함으로써 학습을 연속적으로 개선시킴
- 중재 엔진(intervention engine)은 교사, 관리자, 시스템 개발자가 학생의 학습을 더 잘 지원하기 위하여 자동 시스템을 중재하거나 우선하도록 허용함

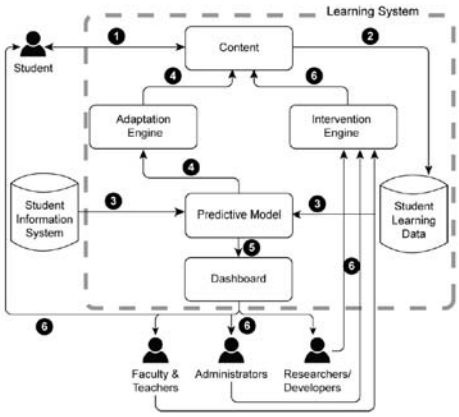

그림 5. 개별화 학습 시스템의 요소와 데이터 흐름

6 요소와 더불어 개별화 학습 시스템에는 학교나 기관에서 유지하는 학습자 정보 시스템(student information system)이 사용된다. 학습자 정보 시스템의 학습자 정보는 자주 변하지 않으므로 수작업으로 내려 받아서 학생 학습 데이터베이스의 수행 데이터와 연결되며 학생들을 범주로 구분하거나 특정 학생이 위기 상황인가를 판단하는 변수를 제공하는데 쓰일 수 있는 배경 정보를 포함함

학생은 개별화 학습 시스템으로 학습하는 콘텐츠와 이루어지는 상호작용에 대한 피드백을 제공받는다. 피드백에는 평가에서 맞춘 비율, 습득한 것으로 증명된 개념들의 목록, (힌트를 요청하거나 문제 풀이 시도 등의) 상세한 학습 활동 정보 등이 포함된다. 한 학생의 상세한 학습 정보는 높은 점수를 받은 학생의 것과 비교하여 그 결과에 따라 학습을 조정하는데 쓰일 수 있다.

교사는 개별 학생과 학급 전체에 대한 피드백을 받아 자신의 교육 방식을 조정한다. 피드백 데이터를 분석하여 권고나 도움이 필요한 학생을 선정하여 학습 속도를 조절하거나 습득하지 못한 내용을 이해시키기 위하여 추가 자료를 제시할 수 있고, 피드백 데이터로부터 학급 전체 학생이 어려워하는 영역을 구분하여 학생의 오개념에 적절히 대처할 수 있다.

관리자는 여러 학급의 상세한 데이터를 분석하여 모든 학생이 향상되는지를 판단할 수 있다. 시스템이 제공하는 상세한 학습 데이터를 (예를 들어, 선수과목을 이수하지 않은 학생이 어떤 성과를 얻는가를 알기 위하여 또는 남녀별 향상 정도를 비교하기 위하여) 학생 집단별로 또는 교사별, 연도별로 분류할 수 있다. 학습 데이터를 이용하여 교육정책을 정하거나 다양한 교육 프로그램을 시행하고, 교육과 학습을 개선하기 위하여 정책과 프로그램을 조정할 수 있다.

연구자는 상세 데이터를 사용하여 학습 이론을 실험하고 여러 가지 교육 방침과 교육과정 설계의 효과를 분석할 수 있다. 학습 시스템 개발자는 학생, 교사, 관리자에게 더 나은 온라인 학습 시스템을 제공하기 위하여 다수의 사용자를 대상으로 검사를 수행할 수 있다. 온라인 학습 시스템을 사용하는 연구자는 많은 학생에게 여러 가지 학습 방식 중에서 무작위적으로 제공되는 실험을 수행할 수 있다.

2.6. 개발되어 있는 학습분석 도구의 예

A. LOCO-Analyst

Loco-Analyst[24][25]는 학생이 학습 콘텐츠를 가지고 하는 활동들(읽기, 퀴즈 풀기 등)에 대한 학습 콘텐츠 상황(learning object context)에 근거를 두고 학생의 활동을 추적하여 분석하여 학생의 학습활동에 대한 피드백을 교사에게 제공하고 교육 내용과 구성을 개선하도록 도와준다.

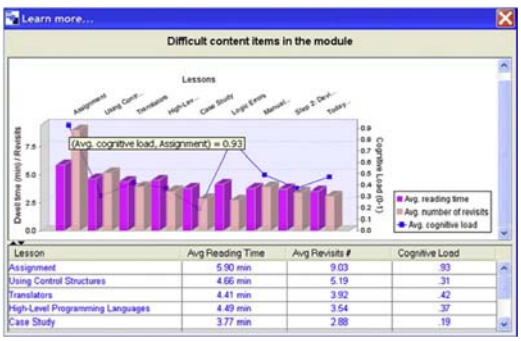

그림 6. 각 단원의 난이도를 나타내는 LOCO-Analyst의 화면. 인지 부담(cognitive load)은 학생들이 각 단원에서 소비한 평균 학습 시간과 단원 내용의 양(글자, 그래프, 그림 등의 개수)을 비교하여 계산됨

이를 달성하기 위하여 학습 콘텐츠 상황 온톨로지(Learning Object Context Ontology), 도메인 온톨로지, 학습 설계 온톨로지, 학습 콘텐츠 구조 온톨로지 등을 이용한다. (온톨로지에 대해서는 아래의 2.8절을 참조)

B. Moodog

Moodog[26]는 California 대학교 Santa Barbara 분교에서 학습관리시스템 Moodle에 학습 활동 추적 기능을 보완한 시스템으로서 Moodle Watchdog의 약어이다.

학생의 온라인 학습 활동(읽은 횟수, 콘텐츠의 개수, 총 접속시간, 올린 글의 개수 등)을 추적하고 학습 활동 데이터를 쉽게 이해할 수 있도록 시각화한다.

그림 7. Moodog에서 제공되는 학생들의 학습활동 보고서

전체 학생이나 개별 학생이 학습 콘텐츠를 어떻게 사용했는가에 대한 정보를 교사에게 제공하고 각 학생이 자신의 학습 진도를 다른 학생과 비교할 수 있게 해준다. 또한 각 학생에게 아직 학습하지 않은 콘텐츠를 학습하도록 격려하는 전자우편을 발송하기도 한다.

C. SAM

온라인 교육이 활성화됨에 따라 학생들은 같은 학급의 학생들 속에서의 자신의 위치를 못 잡게 되고 교사는 학급 전체에 대하여 조망할 기회를 잃게 된다. SAM(Student Activity Monitor)[27]은 학생들 학습활동의 활성상태를 분석해서 시각화하는 학습분석 도구이다.

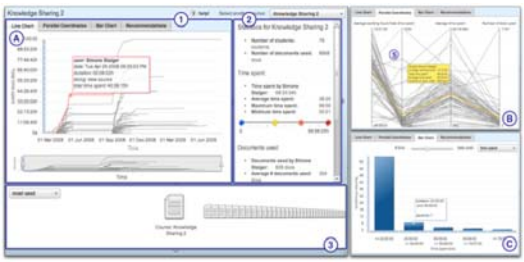

그림 8. SAM의 사용자 인터페이스에서 각 학생의 학습활동 상태를 시각화해서 보여줌

시각화 결과를 보고 학생들은 자기 자신에 대한 피드백을 얻고 학습 동기가 생길 수 있다.

D. SNAPP

SNAPP(Social Networks Adapting Pedagogical Practice)[28]은 대화방에 올린 글과 답글로 생겨나는 사회연결망을 시각화하는 도구이며 여러 학습관리시스템에서 데이터를 추출할 수 있도록 개발되었다.

교사가 학생들 행동의 패턴을 쉽게 찾을 기회를 제공하여 고립된 (즉 위기의) 학생을 분간하고 학급 내에서 연결망의 중심인 학생을 분간하게 해 준다. 또한 교사가 학생들의 연결망을 중재(예로 공동학습을 시키기)하기 전후의 상호작용의 변화를 알려준다.

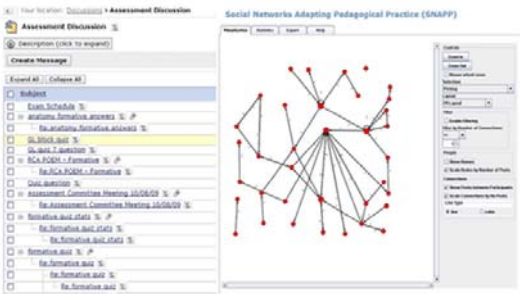

그림 9. SNAPP에서 대화방에 올려진 글과 답글들의 목록과 이들이 시각화된 네트워크 도표

2.7. 온톨로지

스마트 교육의 시나리오가 교사의 개입을 최소화하여 자동적으로 또는 반자동적으로 진행되려면 모든 콘텐츠의 메타데이터(metadata)가 도메인 온톨로지(ontology)에 근거를 두고 작성되어야 하고 콘텐츠들을 구조화하는 근거가 되는 학습설계 온톨로지가 필요하다.

온톨로지는 철학의 한 분야로서 최근 인공지능 및 정보과학의 분야에서 널리 쓰이고 있으며 전자적으로 정보를 공유하는 참여자들 사이에 개념의 혼동을 완화하거나 제거하기 위하여 도입되었다[29][30].

온톨로지는 사람과 컴퓨터 사이에 개념을 공유하기 위한 형식적(formal)이고 명시적(explicit)인 규약을 의미한다. 여기에서 “공유”는 사람과 사람, 사람과 컴퓨터, 컴퓨터와 컴퓨터가 합의된 표현 체계에 따라 개념을 공유함을 의미하고 “형식적”은 컴퓨터가 읽을 수 있다는 의미이며 “명시적”은 개념의 종류와 그들의 관계, 개념들의 제약 조건들이 명백하게 기술되어 있음을 나타낸다. 온톨로지의 구성 요소는 개념(concept, class), 인스턴스(instance), 관계, 공리(axiom) 등이다.

도메인 온톨로지(domain ontology)는 특정 도메인(분야 또는 교과목)을 특징짓는 개념들을 구분하고 정의하며 그들 사이의 관계를 명시하며 한 교과목에서 교육되는 내용들 사이의 계층적 구조나 학습 순서를 구조화하는데 필요하다.

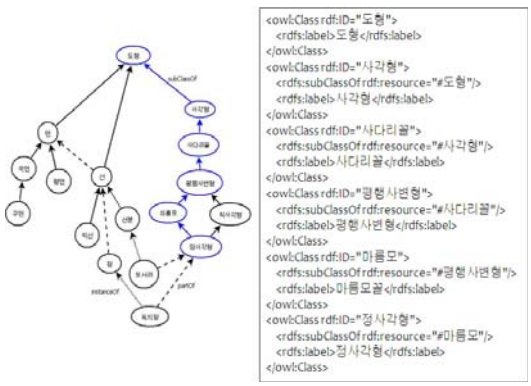

그림 10. 도메인 온톨로지의 일부분. 오른쪽은 왼쪽 그림으로 나타낸 개념과 관계를 온톨로지 언어인 OWL로 기술한 것임

학습설계 온톨로지(learning design ontology)는 도메인 온톨로지와 학생의 학습자정보를 결합하여 특정학생에게 최적화된 교육을 제공하기 위하여 콘텐츠들을 구조화하는 근거가 되는 온톨로지이다.

온톨로지를 구축하려면 여러 분야의 전문가들이 특정 도메인을 철저히 분석해야 하고 시간과 경비가 많이 들게 된다. 온톨로지 학습 또는 온톨로지 생성(ontology generation)은 [31][32] 기계학습과 자연언어처리를 활용하여 데이터베이스와 문서의 정형 데이터와 비정형 데이터에서 개념과 관계를 추출하여 반자동으로 온톨로지 구축을 돕는 방법이다.

2.8. 사생활과 윤리

빅데이터의 활용이 확산되어 감에 따라 사생활 보호가 큰 문제로 대두될 수 있다. 소프트웨어를 개별화시키기 위하여 사용자의 개인 정보를 수집할 필요가 있기 때문에 개별화된 상호작용과 사용자 모델링은 사생활에 상당한 영향을 끼친다.

온라인 사용자는 자신에게 개별화된 콘텐츠에 가치를 두지만 또한 인터넷 상의 사생활에도 관심을 가지며 사생활과 개별화는 간단히 균형지을 수 있는 것이 아니다.

개인과 사회의 사고 방식, 공개된 정보가 평균으로부터 벗어난 정도, 공개된 정보를 아는 사용자의 수, 정보를 통제할 수 있는 정도 등이 판단 기준이 된다. 법률의 제정을 통한 공식적인 통제와 데이터를 이용하는 기관이 자발적으로 채택한 윤리기준에 의한 비공식적인 통제로 분류할 수 있다.

교육 기관은 학생과 교사의 사생활을 보호하도록 요구받고 있다. 교육 데이터마이닝과 학습 애널리틱스가 학생의 활동을 더 많이 관찰하여 예측하고 바람직한 활동을 추천하려는 것이 사회윤리적인 문제를 야기한다.

데이터마이닝을 통하여 학생에 대하여 얻어진 정보에 대한 윤리적 의무사항으로서 혜택을 줄 수 있는 사람들과만 정보를 공유해야 하고, 피해를 주기 보다는 혜택을 주는 방식으로 공유되어야 한다.

교육 데이터 분석가가 “한 학생이 과거의 성과에 근거하여 교과목을 잘 학습할 확률이 20%에 불과하다는 정보를 학생에게 알려주는 것이 유용할까?”, “학급에서 또는 교사에게 이러한 발견은 어떤 의미일까?”, “이러한 결과가 공개되었을 때에 교사-학생간의 관계는 어떻게 될까?”와 같은 질문은 스스로에게 던져야 한다.

정책 결정자는 학생에 대한 결정을 하는데 쓰일 예측 모델의 타당성을 분석할 윤리적 책임을 가지며 학습 애널리틱스

에 근거를 두고 컴퓨터 시스템이 행하는 예측과 조처의 증거를 설명할 수 있어야 한다.

개인별 맞춤 서비스를 제공하려면 정밀한 고객 데이터가 필요하나 개인정보보호법에 의하여 개인정보를 수집, 이용하거나 제3자에게 제공할 경우에는 정보주체의 동의를 얻어야 한다.

사생활보호 데이터마이닝(Privacy Preserving Data Mining)은 개인정보 유출을 방지하는 방법으로서 개별 이용자의 프라이버시가 노출되지 않지만, 전체 동향에 대해서는 공유할 수 있도록 하는 것이다. 개인 정보가 여러 저장 장치에 분산 저장되어 전체 정보가 한꺼번에 유출될 수가 없고 분석을 분산 환경에서 수행한다.

3. 교육 빅데이터 활용 프레임워크

스마트교육 환경에서는 학습자의 세부적인 학습활동 정보들이 대규모로 축적될 것이므로 이를 체계적으로 수집하고 관리하여 이로부터 많은 강의자(교사), 교육전문가, 교육 행정가들이 다양한 학습전략과 분석이 가능한 환경을 제공하여야 한다.

통합적 스마트교육 서비스 프레임워크에서는 여러 학습관리시스템으로부터의 학습자 학습활동정보가 교육 빅데이터 처리센터로 모여질 수 있도록 구성한다. 교육 빅데이터를 분석하여 유용한 학습자 피드백 정보를 생성하고 교육행정 정보로 활용할 수 있는 방안은, 매우 다양한 아이디어가 도출될 수 있으므로 특정한 몇 가지 방식만의 처리방법을 고수하는 것은 매우 비효율이다.

그러므로 교육 빅데이터 처리센터는 학습관리시스템으로부터의 정보를 체계적으로 구축하고 이를 처리할 수 있는 물리적인 컴퓨팅 환경을 제공하고, 이를 활용하여 다양한 분석을 수행할 수 있는 교육 빅데이터 처리 API를 제공하고, 이를 통하여 얻어진 결과물들을 학습자, 교사, 교육행정가에게 제공할 수 있는 환경을 제공하는 것이 중요하다.

더불어 스마트교육 환경에서 학습자의 세부적인 학습활동 정보들을 원활히 활용하기 위해서는 학습자 개인정보를 보호할 수 있는 방안과 표준이 매우 필요하다. 학습자 학습활동 정보들을 효율적으로 활용하기 위해서는 스마트 교육 관련 온톨로지 모델의 확립이 필요하다.

그림 11은 스마트교육에서 데이터의 흐름을 모식적으로 나타낸 것이다. 학생이 콘텐츠를 사용하면 학습활동 이력데이터가 추출 · 저장된다.

예측모델이 학생의 학습자정보와 학습이력 데이터를 분석하여 학생의 학습활동을 추적하고 이후의 행동이나 성과에 대한 추측을 수행한다.

학습자 정보, 도메인 온톨로지, 학습 설계 온톨로지에 근거하여 학생의 성취 수준 및 관심에 따라 개별화학습을 위한 적절한 콘텐츠가 적절한 순서로 학생에게 제공된다.

시각화 기술을 사용하여 종합된 정보는 현황판(dashboard)에 게시되고 교사 또는 콘텐츠제작자가 활용하게 된다.

이와 같이 학습 애널리틱스, 빅데이터 처리 기술, 온톨로지 등을 스마트교육에 접목하면 학생에 적응하는 개별화학습 및 데이터에 근거를 둔 교육체제의 혁신이 가능해질 것이다.

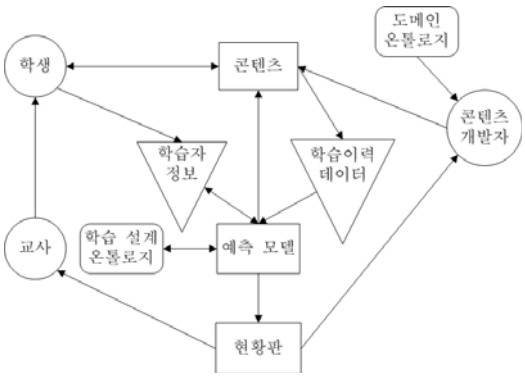

그림 11. 스마트교육의 데이터 흐름도

IV. 교육 빅데이터 활용방안

기존의 오프라인 교육에서는 교육의 결과인 시험 결과와 최종 평가만 수작업으로 기록되었고 (최근 OCR을 이용한 답안 작성으로 시험은 디지털화되었으며) 교육이 진행되는 과정(학습 활동)과 평가 과정이 오프라인 상에서 존재하다가 사라진다.

현재 도입되고 있는 온라인 스마트 교육에서는 모든 학습활동이 온라인 상에서 진행되므로 학습 활동과 관련되는 모든 데이터를 추출하기가 용이하고 개별화학습을 통하여 학생의 학업성취도를 향상시키고 교육 정책의 결정 과정에서 데이터의 활용가능성을 탐색해 볼 기회가 많아질 것이다.

교육 빅데이터를 최대로 활용하기 위하여 필요한 것은 방대한 양의 데이터를 처리할 수 있는 컴퓨터 하드웨어와 소프트웨어만이 아니라 방대한 데이터 속에서 의미 있는 요소를 찾아내기 위하여 사람과 교육이론이 함께 협력하여야 한다.

구체적으로 다음의 4가지 종류의 여러 가지 시나리오를 생각해 볼 수 있다

- 학생의 개별화학습
- 콘텐츠 개발
- 교육과정 및 교육정책 개선
- 학생 지도

1. 개별화학습

시스템 갖고 놀기(시스템이 정답을 제공할 때까지 클릭하고 소프트웨어가 피드백과 도움을 주는 규칙을 체계적으로 이용하여 다음 단계로 넘어가는 것) 등의 활동을 하는 학생들을 감지하고 추가로 연습문제를 제공하여 학습을 향상시킬 수 있다.

학생 반응의 정확성, 질문에 답하기 전에 연습에서 소비한 시간, 요청한 힌트의 횟수와 특성, 잘못된 답의 반복, 실수 등과 같은 상호작용으로부터 학생의 지식 수준을 추론한다.

학습관리시스템은 학생의 학습활동에 근거하여 학생, 교사 및 관리자에게 의미 있고 활용 가능한 피드백을 제공해 준다.

피드백 데이터를 분석하여 권고나 도움이 필요한 학생을 선정하여 학습 속도를 조절하거나, 습득하지 못한 내용을 이해시키기 위하여 추가 자료를 제시할 수 있다. 피드백 데이터로부터 학급 전체 학생이 어려워하는 영역을 구분하여 학생의 오개념에 적절히 대처할 수 있다.

각 학생의 학습자정보 및 학습활동 이력데이터를 종합하여 학습 능력, 집중력 등을 알아내어 다음 주제로 진행해도 되는지, 진도가 늦지 않은지를 판단하여 각 학생에게 알맞은 학습자료를 제공함으로써 동기를 부여하고 이들 데이터는 학생의 학습자정보를 갱신하는데 쓰인다.

학생마다 취향, 흥미, 배경, 학습 목표 등이 다르기 때문에 학습자 정보를 수집하여 학생 개인이나 학생 집단의 학습 효과와 효율을 최대화하기 위한 개별화된 학습 환경을 제공할 수 있다.

학습자 정보는 학습에 해로운 행동이 검출될 때 개별화된 상호작용을 제공하는 장기 목표를 가지고 새로운 학생을 분류하는 데에 쓰인다. 이해도, 학습 경로, 동기, 문화적 배경 등에 따라 학생을 분류하여 교사가 지도 방식을 변화시키는 것을 도울 수 있다. 각 학생은 자신의 학습진도와 학습성과를 다른 학생과 비교함으로써 스스로 학습 계획을 구성하고 학습을 수행한다.

여러 학생들이 콘텐츠를 이용한 뒤에 밝힌 콘텐츠의 평가와 개별 학생의 학습자 정보에 근거하여 그에게 적절한 콘텐츠를 추천할 수 있다.

새로 제시된 학습 콘텐츠에 대하여 학생의 성과를 예측할 수 있게 한다. 학생의 학습 성과를 관찰 · 예측하고 잠재적인 문제를 일찍 찾아내어 교과목에 낙제할 위험에 빠진 학생을 확인하는 데에 응용될 수 있다.

더 나아가 학생들의 방대한 자료를 분석하여 학업 지체요인을 밝혀내고 조건별로 각 학생들의 학업성취도를 향상시킬 수 있는 최적화 방안을 마련할 수가 있게 된다.

2. 콘텐츠 개발

학생의 이해도를 평가하기 위하여 치르는 퀴즈 및 시험문제의 난이도, 변별도, 추측도, 오답지 매력도 등의 문항분석[33]을 통하여 답지나 문제를 재조정하는데 쓸 수가 있다.

많은 학생들의 학습 활동의 결과로부터 각 콘텐츠의 난이도, 유용성, 사용 편리성 등에 대한 평가가 이루어질 수 있다. 각 학년의 모든 학생들의 학습활동을 수집할 수가 있으므로 제공하고 있는 학습 콘텐츠의 적정성을 분석할 수가 있다.

많은 학생들이 특정 콘텐츠에 시간을 오래 소비하거나 재방문을 한다면 그 내용이 학생들에게 어렵다는 방증이므로 콘텐츠를 변경하거나 쉽게 풀어쓸 필요가 있음을 알 수 있다.

학생들은 콘텐츠에 대하여 난이도, 사용의 편리성과 교육적 가치에 대하여 평가한다. 각 콘텐츠의 평가와 활용 상황으로부터 콘텐츠의 선호도를 취합하고 활용 방식에 대한 데이터를 획득하여 콘텐츠 개발자에게 콘텐츠를 개선할 수 있는 정보를 제공할 수가 있다.

3. 교육과정 및 교육정책 개선

시스템이 제공하는 상세한 학습 데이터를 (예를 들어, 선수과목을 이수하지 않은 학생이 어떤 성과를 얻는가를 알기 위하여 또는 남녀별 향상 정도를 비교하기 위하여) 학생 집단별로 또는 교사별, 연도별로 분류할 수 있다. 여러 학급의 상

세한 데이터를 분석하여 모든 학생이 향상되는지를 판단할 수 있다.

학생을 잘 가르치는 교사의 교육적 패턴을 발견하여 다른 교사에게 동일한 기술을 훈련시켜 학습 시스템에서 어떤 변화가 생기는가를 연구하고 효과가 있다면 이 기술을 확산시킬 수 있다.

학습 시스템 개발자는 학생, 교사, 관리자에게 더 나은 온라인 학습 시스템을 제공하기 위하여 다수의 사용자를 대상으로 검사를 수행할 수 있다. 온라인 학습 시스템을 사용하는 연구자는 많은 학생에게 여러 가지 학습 방식 중에서 무작위적으로 제공되는 실험을 수행할 수 있다.

상세 데이터를 사용하여 학습 이론을 실험하고 여러 가지 교육 방침과 교육과정 설계의 효과를 분석할 수 있다.

학습 데이터를 이용하여 정책을 정하거나 다양한 프로그램을 시행하고, 교육과 학습을 개선하기 위하여 교육정책과 교육 프로그램을 조정할 수 있다.

학생의 수행과 교과과정 사이의 연관관계를 발견하고 어떤 교육 전략이 더 효과적이고 견고한 학습을 낳는지를 발견하는데 응용된다. 새로 개발된 교과과정이 이전의 교육과정보다 더 나은 학습을 가능하게 하는가를 객관적으로 분석할 수 있는 기회를 제공하여 준다.

중간고사, 기말고사, 일제고사 등 정기적인 시험을 시행하지 않고 온라인 학습 및 온라인 퀴즈만으로 대체할 수 있는 가능성을 탐구해 볼 수 있다.

교육 데이터를 이용하여 교육 이론을 검증하고 개선하는데 활용할 수 있고 이를 이용하여 교육자료의 수준을 조정하고 학년별 수준에 맞추어 교육과정의 개편작업 및 교육정책 수립에도 활용할 수가 있다.

4. 학생 지도

온라인 학습, 콘텐츠 활용, 검색하는 내용을 종합하여 학생이 어떤 주제에 관심을 가지고 적성이 무엇인지를 판단하고 진로지도에 도움을 줄 수 있을 것이다.

온라인 학습에서 대화방의 참여, 전자우편 발송 횟수 등 학생의 행동 변수를 분석하여 학생의 성적과의 관계를 분석한다.

학생들이 사회연결망 서비스(SNS) 및 대화방에 게시하는 글을 분석함으로써 학생들 사이의 사회연결망과 학생과 교사 사이의 사회연결망을 파악하여 사회연결망이 학습에 미치는 영향을 연구하고 활용할 수가 있다. 학생들 사이의 호 · 불호 관계에 대한 정보를 얻을 수가 있어서 왕따 및 폭력 단체의 존재에 대한 추정을 가능하게 할 것이다.

V. 시사점 및 제언

가장 바람직한 교육의 형태는 해당학생에게 적합하도록 모든 교육자원을 제공하여 스스로 학습할 수 있도록 하는 것이다.

스마트교육에서 손쉽게 다양한 교육자료를 제공할 수 있지만 해당 학습자에게 적합한 교육을 시행하기 위해서 교수자의 세심한 관찰도 필요하고 다양한 교육분석 정보도 필요하다.

스마트교육에서는 학생의 모든 학습활동 자료가 쉽게 수집될 수 있을 뿐만 아니라 모든 학생의 자료 또한 수집될 수 있다. 스마트교육에서 수집되는 모든 개별학생의 모든 학습활동자료는 전형적인 빅데이터이며 이를 통하여 개별학생에 적합한 피드백뿐만 아니라 전체 학생을 위한 교육콘텐츠 개발, 교육정책 발전에 획기적인 분석과 발전을 도모할 수 있는 바탕이 된다.

스마트교육은 단순히 손쉽게 다양한 멀티미디어 학습자료를 활용할 수 있도록 하는 것을 넘어 학습자에게 적합한 학습 피드백을 제공할 수 있어야 진정한 스마트교육이 될 수 있다.

스마트교육의 빅데이터를 제대로 활용하기 위해서는 온톨로지 기술을 이용하여 학습콘텐츠를 통한 모든 학습활동이 정교하게 수집되고, 분석되어 활용될 수 있는 교육 빅데이터 프레임워크가 마련되어야 한다.

교육 빅데이터 프레임워크는 스마트교육을 위한 표준 제정을 통하여 모든 학습콘텐츠와 학습활동자료가 수집될 수 있도록 해야 하며, 이렇게 수집된 정보는 세심한 개인정보보호 원칙하에 대용량 처리 서버에 관리되어야 하고, 다양한 교육 빅데이터 활용 아이디어를 적용할 수 있도록 교육 빅데이터 처리 환경이 많은 연구자들에게 오픈 API형태로 제공할 필요가 있다.

또한 교육 빅데이터의 다양하고 성공적인 결과물들은 교육 일선현장과 교육정책기관에서 잘 활용될 수 있도록 교수자들에게는 사용하는 학습관리시스템에 플러그인 형태로 제공하고, 정책집행자들에게는 교육지표모니터링 시스템의 형태로 제공될 필요가 있다.

정부가 주도적으로 진행하고 있는 우리나라의 스마트교육

은 상대적으로 해외 여러 나라보다 빨리 진행되고 있지만 스마트교육의 진정한 효과는 교육빅데이터를 통한 개별화 학습과 합리적인 교육정책으로 달성할 수 있을 것으로 여겨진다.

스마트교육의 시행과 발맞추어 잘 확립된 교육 빅데이터 활용 인프라를 구축할 수 있는 단계적인 정책의 시행이 필요할 것으로 사료된다.

참고문헌

[1] 시로타 마코토, *빅데이터의 충격*, 한빛미디어, 2013년.

[2] 한국정보화진흥원, "특집 빅데이터 시대의 국가발전전략", *2012 국가정보화백서*, 2012년

[3] L. Douglas, "The Importance of 'Big Data': A Definition", Gartner, 2012.

[4] J.-P. Dijcks, "Oracle: Big Data for the Enterprise", *An Oracle White Paper*, 2012.

[5] 정의석, 한승철, 박충식, "디지털교과서수업설계저작도구", *해양정보통신학회 춘계학술대회*, 2011년.

[6] 박충식 등, "디지털교과서의 적응적 서비스를 위한 표준화 방안 연구(부제: 미래의 교과서 - 스마트북)", *KERIS 연구보고 KR 2011-11*, 2011년.

[7] 최제영, 박충식, 최광선, 정의석, 김성진, 유인식, "스마트교육에서 발생하는 교육 빅데이터 활용방안", *2012 한국지능정보시스템학회 춘계학술대회*, (2012)

[8] 교육과학기술부, "인재대국으로 가는 길 스마트교육 추진전략 실행계획", 2011.10.

[9] 박충식, "미래교과서상상하기: 스마트북 - 애플아이북스를 넘어서", 한국교육학술정보원 교육정보화 수요포럼 73회, http://www.kocw.net/home/search/kemView.do?kemId=336257, 2012.2.22. 2012년

[10] 양재봉, 아이북스 오서 iBooks Author : 세상에서 제일 쉬운 전자책 만들기, 한빛미디어, 2012년 06월

[11] 디지털플레이북, http://www.fcc.gov/encyclopedia/digital-textbook-playbook

[12] Moodle, https://moodle.org/

[13] OpenClass, http://www.openclass.com/

[14] Blackboard, http://www.blackboard.com/

[15] 임진혁 등, "고등교육 범용 학습관리시스템(LMS) 기준모형개발 기획 연구", *KERIS 연구보고 CR 2010-10*, 2010년.

[16] 국가통계포털, http://kosis.kr/

[17] M. Bienkowski, M. Feng, and B. Means, "US Department of Education Featured Report: Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics [DRAFT]", 2012.

[18] T. Elias, "Learning Analytics: Definitions, Processes and Potential", http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotentia.pdf, 2011.

[19] N. Yau, *비주얼라이즈 디스 : 빅데이터 시대의 데이터 시각화+ 인포그래픽 기법*, 에이콘, 2012년.

[20] Gapminder, http://www.gapminder.org

[21] IBM의 ManyEyes, http://www-958.ibm.com/software/data/cognos/manyeyes/

[22] Tableau, http://www.tableausoftware.com/

[23] FlowingData, http://flowingdata.com/

[24] LOCO-Analyst, http://jelenajovanovic.net/LOCO-Analyst/

[25] J. Jovanovic, D. Gasevic, C. Knight, and G. Richards, "Ontologies for Effective Use of Context in e-Learning Settings," *Educational Technology & Society*, Vol. 10, No. 3, pp. 47-59, 2007.

[26] H. Zhang, et al., "Moodog: Tracking Student Activity in Online Course Management Systems", *Journal of Interactive Learning Research*, Vol. 21, No. 3, pp. 407-429, 2010.

[27] SAM, http://www.role-showcase.eu/role-tool/student-activity-monitor

[28] SNAPP, http://research.uow.edu.au/learningnetworks/seeing/snapp/index.html

[29] R. Mizoguchi, *(차세대 웹과 지식 처리의 핵심 기술) 온톨로지 공학*, 두양사, 2009년.

[30] 황금하, 신지애, 최기선, "개념 및 관계 분류를 통한 분야 온톨로지 구축", *정보과학회논문지 : 소프트웨어 및 응용*,

Vol. 35, No. 9, pp. 562-571, 2008년.
[31] R. Navigli, P. Velardi, and A. Gangemi, “Ontology Learning and Its Application to Automated Terminology Translation”, *IEEE Intelligent Systems*, Vol.. 18, No. 1, pp. 22-31, 2003.
[32] 최기선, 류법모, “온톨로지의 구축과 학습: 상하위 관계”, *정보과학회지*, Vol. 24, No. 4, pp. 24-30, 2006년.
[33] 황정규 등, *교육평가의 이해*, 학지사, 2011년.

저 자 소 개

최 제 영

1984: 서울대학교
물리학과 이학사.
1987: 서울대학교
물리학과 이학석사.
1990: 서울대학교
물리학과 이학박사
1990-1991: 미국 Temple Univ.
연구원
1991-1993: 서울대학교 CTP
연구원
현 재: 영동대학교
임베디드소프트웨어학과
교수
관심분야: 복잡계,
임베디드시스템,
스마트교육

박 충 식

1985: 한양대학교
전자공학과 공학사.
1987: 연세대학교
전자공학과 공학석사.
1992: 연세대학교
전자공학과 공학박사
현 재: 영동대학교
스마트IT학부 교수
관심분야: 지식기반시스템,
컴퓨터비전,
스마트교육