

빅데이터와 AI를 활용한 교육용 자료의 분석에 대한 조사

노영욱

신라대학교

A Survey on Deep Learning-based Analysis for Education Data

Young-uhg Lho

Silla University

E-mail : yulho@silla.ac.kr

요 약

최근에 빅 데이터와 AI 기술을 교육의 평가와 개별 학습에 적용하는 연구 성과가 있었다. 정보 기술의 혁신으로 소셜 미디어, MOOC, 지능형 개인지도 시스템, LMS, 센서 및 모바일 장치 등으로부터 학생들의 개인 기록, 생리학적 데이터, 학습 로그 및 활동, 학습 성과 및 결과를 포함하는 동적이고 복잡한 데이터를 수집 가능하였다. 또한 COVID-19 환경에서 e-러닝이 활성화 되어 많은 양의 학습 데이터가 생성되었다. 이 데이터로부터 학습 분석과 AI 기술을 적용하여 의미있는 패턴의 추출과 지식의 발견이 될 것으로 예상된다. 학습자 측면에서 학생의 학습 및 정서적 행동 패턴과 프로필을 식별하고, 평가 및 평가 방법을 개선하고, 개별 학생의 학습 성과 또는 중퇴를 예측하고, 개인화 된 지원을 위한 적응 시스템에 대한 연구는 필요하다. 본 연구에서는 교육용 데이터를 대상으로 이상탐지와 추천시스템에서 사용하는 기계학습 기술에 대한 조사와 분류를 하여 교육 분야의 연구에 기여하고자 한다.

ABSTRACT

Recently, there have been research results of applying Big data and AI technologies to the evaluation and individual learning for education. It is information technology innovations that collect dynamic and complex data, including student personal records, physiological data, learning logs and activities, learning outcomes and outcomes from social media, MOOCs, intelligent tutoring systems, LMSs, sensors, and mobile devices. In addition, e-learning was generated a large amount of learning data in the COVID-19 environment. It is expected that learning analysis and AI technology will be applied to extract meaningful patterns and discover knowledge from this data. On the learner's perspective, it is necessary to identify student learning and emotional behavior patterns and profiles, improve evaluation and evaluation methods, predict individual student learning outcomes or dropout, and research on adaptive systems for personalized support. This study aims to contribute to research in the field of education by researching and classifying machine learning technologies used in anomaly detection and recommendation systems for educational data.

키워드

Big Data, AI, Education COVID-19, Machine Learning

I. 서 론

서울교육정책연구소는 COVID-19 유행으로 원격수업이 장기화하면서 학력 저하와 양극화 문제에 대한 조사 결과를 발표하였다. 서울 시내 중학교 2·3학년 학생의 1학기 국어·영어·수학 내신 성적 등급 비율 변화를 3년에 걸쳐 비교한 결과에 의하면 COVID-19로 인한 원격수업은 상·하위권 학생의 학력 격차도 더욱 커지고, 상위권 학생들에

겐 거의 영향이 없었지만 중위권 학생들에게는 타격을 많이 준 것으로 나타났다. 또한 “원격 수업으로 중위권 아이들이 제대로 된 수업을 받지 못하면서 하위권으로 추락했고, 가정에서 돌봄을 잘 받는 중상위권 학생들은 사교육 등으로 보충하면서 격차가 벌어졌다”는 해석을 하고 있다[1].

지능형 교육정보통계시스템(EDS)은 교육부 및 시도교육청, 소속·산하기관 등 기관별로 산재해 있는 교육통계정보를 한곳에 모아 연계·저장하고, 정

제 표준화하여 교육부 시도교육청에서 다양한 형태의 대용량 데이터 및 분석결과를 활용하기 위한 환경이 마련되었다[2].

빅데이터와 인공지능 기술의 발전에 따라 학습자에 대한 학습과정 등의 분석으로 학생에게 최적화된 맞춤형 교육이 가능한 지능형 학습 시스템 구축이 가능하게 되었다.

II. 학습 데이터

LMS(Learning Management System)는 교수자와 학습자의 학습을 지원하고 관리하는 시스템이다. e-learning에서 학습자의 학습과정을 추적하고 학습 이력을 관리하여 학습자 개인에 대한 맞춤형 학습을 제공하게 된다. LMS의 주요 기능은 학급 편성 기능, 협동학습기능, 출결관리 기능, 게시판 기능 등이 있다. LMS의 기능이 고도화될수록 학생의 개별학습을 위한 맞춤형 학습 환경을 효과적으로 구성할 수 있다.

IMS의 Edu Graph 모델에서는 학습 분석을 위한 데이터 유형을 [그림 1]과 같이 다섯 개의 그룹으로 분류한다[3].

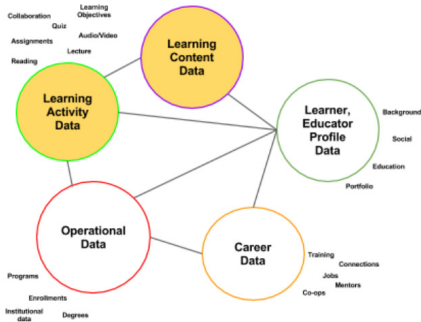


그림 1. Edu Graph 데이터 모델

Edu graph를 기반으로 제시되는 학습분석을 위한 데이터는 대부분 정형화된 데이터이다. 학습자의 학습활동에서 발생하는 학습활동 데이터(학습자의 게시글, 이메일, 다양한 형태의 로그 기록 등)는 대부분이 비정형 데이터이다. 비정형 학습 데이터는 학습 분석에서 학습자의 특성과 학습 환경을 나타낸다.

IMS Global은 [그림 2]와 같이 학습활동 중심의 데이터 매트릭스를 제안하였다. 대부분의 학습활동들은 읽기, 평가, 매체, 협동 등과 같은 하나의 학습활동의 의미를 기준으로 학습활동 분야로 그룹화 할 수 있다. IMS의 데이터 매트릭스는 학습활동에 세부적인 측정방법을 정의 및 표준화한다. 모든 학습활동에 일반적으로 적용할 수 있는 분류방법, 분류기준과 측정방법을 지원하는 학습참여자

성과평가와 같은 통합된 학습활동 분야에 대한 측정방법도 정의한다. 특정 학습활동은 하나 이상의 학습활동 측정 프로파일을 가지고 이와 관련된 학습 데이터를 수집하고 공유할 수 있도록 설계되어 있다.

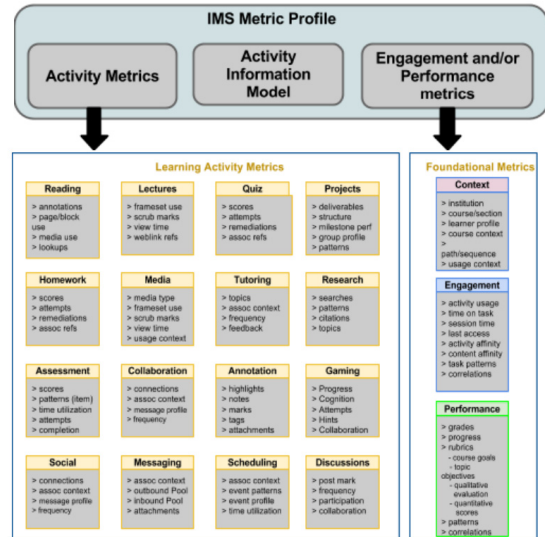


그림 2. 학습활동 중심의 데이터 매트릭스

III. 인공지능 적용 기술

인공지능 기술은 4차 산업혁명과 관련된 핵심기술이며 교육 분야에서도 영향을 미치고 있다. 교육용 데이터가 확보되고 이를 분석하여 교육 분야에 빅데이터와 인공지능을 활용하면 다음과 같이 학습이 편리해지고 개인별 맞춤화가 가능하여 교육 격차를 해소하고 학습자, 교육자 및 교육 기관에 도움이 된다[4-6].

- 개인화된 맞춤형 학습
- 가정교사 같은 개별지도, 피드백이 가능한 튜터링 학습
- 학습자가 어려워하는 개념 식별 후 맞춤 교육
- 시간/공간적 제약이 없는 글로벌 학습
- 실수를 두려워하지 않아도 되는 시행착오를 통한 학습
- 스마트 콘텐츠 기반의 맞춤형 자료
- AI 기반 자동채점
- 일상적인 교사 업무의 자동화 및 교사의 역할 변화

인공지능 기술을 교육에 활용하면 학습자의 상태를 진단하고 학습 순서를 조정하여 기존의 학습 방법보다 학습 속도가 빨라질 수 있다. 교육 분야에 인공지능을 활용한 대표적인 사례로는 MATHia 등이 있다[7]. MATHia는 Carnegie Learning는 학습

자의 개념 단위 학습 과정을 파악하고 적절한 피드백과 힌트를 제공해 주는 지능형 학습지원 시스템이다. 학습자 맞춤형 교육은 학습자의 학습 정도, 선호도 등에 맞게 문항 및 콘텐츠를 제공하는 방법이다.

이상탐지는 정상이 아닌 데이터 세트에서 패턴을 찾는 것으로 이러한 예상치 못한 동작을 이상치라고 한다. 이상탐지 단계는 [그림 3]과 같이 매개 변수화, 훈련, 모델, 탐지 단계로 구성된다[8].

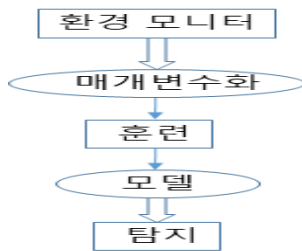


그림 3. 이상탐지 단계

교육 분야에 이상탐지 기술은 다양한 감성 변화에 대한 탐지에 사용될 수 있다. 학생들의 다양한 감정 빅데이터(글, 표정, 눈)를 딥러닝을 통해 컴퓨터가 스스로 학습하도록 할 수 있다. 컴퓨터는 이를 기반으로 학생의 표정 글 등을 통해 학생의 심리상태를 판단하고 이해할 수 있다.

기계학습 기반 이상탐지 기법은 [그림 4]와 같이 분류기반, NN(Nearest neighbor) 기반, 군집화(clustering) 기반, 통계적 기법, 정보이론 기반, 스펙트럴 기반, 맥락적 기반 방법 등이 있다.

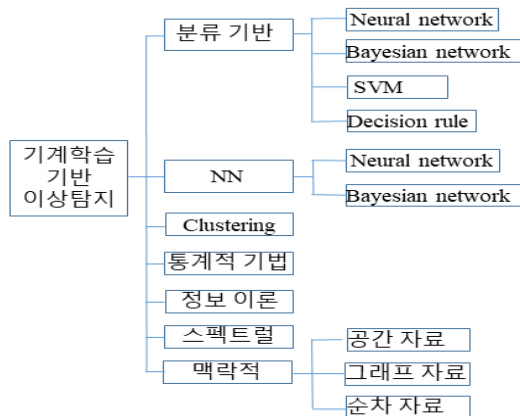


그림 4. 기계학습 기반 이상탐지

추천시스템에서 추천목록은 사용자 선호도, 항목 기능, 사용자 항목 과거 상호 작용, 시간 및 공간 데이터와 같은 일부 추가 정보를 기반으로 생성된

다. 이는 학생의 학습 패턴을 스스로 학습하여 개별화 교육과 적절한 교육 자료를 추천하는데 적용할 수 있다. 추천 모델은 주로 입력 데이터의 유형에 따라 협업, 콘텐츠 기반 추천 시스템 및 하이브리드 추천 시스템으로 분류된다. 최근 딥 러닝 기반 추천시스템은 [그림 5]와 같이 두 가지 범주로 분류할 수 있다[9].

- 뉴럴 블록형 추천시스템 : MLP는 사용자와 항목 간의 비선형 상호 작용을 쉽게 모델링할 수 있으며, CNN은 텍스트 및 시각적 정보와 같은 이기종 데이터 소스에서 로컬 및 글로벌 표현을 추출할 수 있으며, RNN은 콘텐츠 정보의 시간적 역학 및 순차적 진화를 모델링할 수 있다.
- 하이브리드형 추천시스템 : 하나 이상의 딥러닝 기술을 사용하며, 심층 신경망의 유연성 덕분에 여러 신경 빌딩 블록을 결합하여 서로를 보완하고 보다 강력한 모델 구성 가능하다.

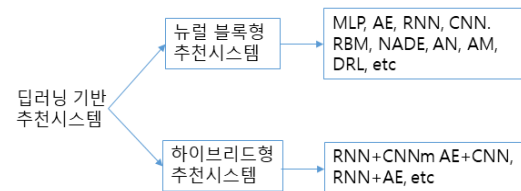


그림 5. 딥러닝 기반 추천시스템 기술 분류

IV. 결론

COVID-19로 인한 비대면 수업의 활성화와 e-learning 기술, 지능형 교육정보통계시스템 구축 등으로 빅데이터 분석이 가능한 학생들의 학습과 활동에 데이터를 수집 저장 가능하게 되었다. 이에 따라 최근 기계학습 기술의 이상탐지와 추천시스템 기술이 교육분야에 활용될 수 있는 기술을 분석하였다. 향후 연구 방향으로 구체적인 데이터를 대상으로 학습 모델을 개발하는 것이 필요하다.

References

- [1] 서울교육정책연구소, 코로나19 전후, 중학교 학업성취 등급 분포를 통해 살펴본 학교 내 학력 격차 실태 분석, 2021-1 현안분석 보고서, 2021.04.
- [2] <https://if-blog.tistory.com/11416> [교육부 공식 블로그]
- [3] IMS, Learning Measurement for Analytics Whitepaper, 2013 IMS Global Learning Consortium.

- [4] wire19.com, “9 ways Artificial Intelligence (AI) is impacting education,” 2020. 9. 17.
- [5] KAKAO AI REPORT vol. 13.
- [6] 전형배, 정훈, 강병욱, 이윤경, AI 기반 교육 현황과 기술 동향, 자동신동향분석 제36권 제1호 2021.2.
- [7] <https://www.carnegielearning.com>
- [8] Varun Chandola, Arindam Banerjee, and Vipin Kumar, “Anomaly Detection: A Survey,” ACM Computing Surveys, Vol.41, No.3, July 2009.
- [9] Shuai Zhang, Lina Yao, Aixin Sun, Yi Tay, “Deep Learning based Recommender System: A Survey and New Perspectives,” ACM Computing Surveys, Vol.1, No.1, July 2018.