

온라인학습 중도이탈 예측시스템 개발

박희순¹, 구영현²

¹ 세종대학교 컴퓨터공학과 박사과정

² 세종대학교 컴퓨터공학과 교수

hiran@sjcu.ac.kr, yhgu@sejong.ac.kr

Development of a prediction system for online education

Hee-Sun Park¹, Yeong-Hyeon Gu²

¹Dept. of Computer Engineering, Sejong University

²Dept. of Computer Engineering, Sejong University

요 약

온라인 교육의 보급이 증가함에 따라 온라인 학습자의 중도이탈율을 줄이기 위한 효과적인 전략이 필요해 졌다. 본 논문은 인공지능 기술을 온라인 학습 분야에 적용하여 온라인 학습 중도 이탈 문제를 해결하고자 한다. 특히 온라인 교육기관에서 중도이탈 예측 시스템을 이미 운영 중인 온라인교육시스템(LMS)을 새롭게 구축하거나 크게 변경하지 않고도 개발 또는 도입이 가능한 방안을 제시한다. 운영 중인 온라인교육시스템(LMS)에 저장되고 있는 온라인 학습자들의 학습자정보 및 학습 이력 데이터를 활용하여 중도 이탈 위험이 있는 학생을 식별하고 맞춤형 관리가 가능하도록 온라인 교육 중도 이탈예측 시스템 개발하고 LMS와 연동하여 서비스는 방법을 제안한다. 이를 통해 온라인교육기관에서는 중도이탈이 있는 학생을 효과적으로 관리함으로써 온라인 교육의 질을 향상시킬 수 있고, 더 많은 수의 온라인 학습자가 성공적으로 과정을 이수할 수 있도록 하는 데 기여할 것이다.

1. 서론

최근 IT 기술의 발전과 COVID-19 팬데믹으로 인해 온라인 교육이 급속도로 성장하고 있다. 이러한 온라인 교육은 시간과 공간의 제약 없이 학습할 수 있는 유연성을 제공함으로써 많은 학습자들에게 매력적인 대안으로 자리 잡고 있다. 하지만, 이러한 온라인 교육은 오프라인 교육에 비해 상대적으로 높은 중도이탈율을 보이고 있다. 중도 이탈 문제는 온라인교육기관의 경쟁력을 낮추고 학습자들이 시간과 비용을 낭비하는 문제를 발생시키기 때문에 중도이탈율을 낮추기 위한 방안이 필요하다. 인공지능 기술의 발전으로 많은 분야에서 예측 서비스를 통해 만족할 만한 성과를 내고 있기에 온라인 교육 분야에서도 중도이탈에 대한 문제를 인공지능 기술을 통해 해결해 보려는 노력이 많이 이루어지고 있다. 그러나 이탈예측 서비스를 도입하여 서비스하는 교육기관은 아직 많지 않다. 이는 운영 중인 온라인교육 시스템(LMS, 이하 LMS)을 통해 데이터 수집 및 가공이 가능하더라도 수집된 데이터를 가공하여 예측 모델 생성, 서비스까지의 개

발이 어렵거나 비용이 많이 들것으로 판단하기 때문일 것이다.

본 연구에서는 이러한 교육기관들이 기존의 LMS와 학사관리시스템을 연동하여 중도이탈 예측시스템을 개발하고 서비스에 활용할 수 있는 방법을 실제 구현 결과를 바탕으로 제안한다는 점에서 의미가 있다.

무엇보다 예측시스템을 통해 교육 기관은 이탈가능성이 있는 개별 학생에 맞춤형 지원을 제공할 수 있으며 학생들의 학업 성공률을 높이고, 이탈율을 감소시킬 수 있다.

본 논문에서는 온라인교육의 높은 이탈 문제를 예측을 통해 해결할 수 있도록 사전에 정보를 제공하기 위한 중도이탈예측시스템 개발을 목표로 하고 있으며, 무엇보다 보유하고 있는 LMS에 중도이탈 예측시스템을 어떻게 연동할 수 있는지를 보여주고자 한다. 향후에는 이러한 시스템을 활용한 실제 결과를 바탕으로 예측과 실제 중도이탈 결과의 일치정도, 실제 예측서비스에서 발생한 문제를 제시하여 정교화 할 수 있는 대안을 제시함으로써 온라인 교육의 질을 향상

시키고 교육 기관의 지속 가능한 발전을 도모하는 방안을 모색하고자 한다.

2. 관련연구

2.1 학습분석

학습분석이란 학습환경에 대해 여러 정보를 수집하여 학습환경과 학습활동 및 과정을 최적화하고 교육 관련 의사를 결정하기 위해 데이터를 실시간으로 수집, 모델링, 분석, 예측, 평가하는 것이며[1]. 학습환경에서 얻어지는 데이터 고유의 유형을 탐색하는 방법을 개발하고 이를 바탕으로 학습환경이나 학습자들을 보다 심층적으로 이해하는 데 목적이 있다[2]. 이러한 점에서 학습분석은 빅데이터와 유사하다고 볼 수 있으나 학습 분석에서 다루고자 하는 데이터는 교육 분야에서의 특화된 데이터만을 포함하며 분석 모델이나 알고리즘 결과를 시각화 하는 방식도 교육이라는 목표를 지향하며, 학습 분석에서는 학습 데이터를 어떻게 수집하고 분석하여 그 결과가 암시하는 바가 무엇인지 밝혀내고 이를 통해 얻은 결과를 LMS에 어떻게 적용시켜 개선할 것인지가 중요함을 설명하였다[4].

2.2 중도이탈 예측 모델

교육 분야에서 중도이탈 예측에 대한 연구는 중도탈락 예측 요인 분석[5], 머신러닝, 딥러닝 기술을 이용한 예측모델 개발[6] 등 다양하게 진행되어 왔다. 특히 교육에 적용하기 위한 학생들의 중도이탈 예측 모델에 대한 연구결과[7]로 기존의 머신러닝 기법인 DecisionTree, SVM(Support vector machine)보다 Random-Forest(RF) 또는 딥러닝 모델(DNN)이 좋은 성과를 내는 것으로 확인되었다.

3. 온라인학습 중도이탈 예측시스템

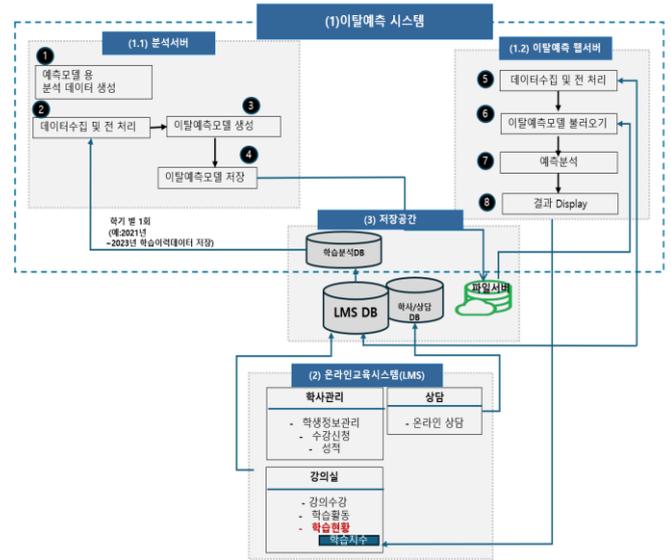
3.1 시스템 구성도

본 연구를 위해 구현한 온라인학습 중도이탈 예측 시스템은 그림 1 과 같다.

(1) 이탈예측시스템 내에 (1.1) 분석서버, (1.2) 이탈예측 웹서버가 있으며, 분석서버에서는 분석모델 생성을 위한 데이터 수집과 생성, 전 처리 작업을 수행한 후 이탈예측 모델을 생성하고 파일서버에 모델을 저장하게 된다. 이탈예측 웹서버에는 실제 분석용 데이터로 저장된 분석모델을 이용하여 분석 후 결과를 보여주는 기능을 한다.

(2) 온라인교육시스템은 기존에 운영 중인 LMS 및 학사관리를 위한 시스템이다. LMS 와 학사관리시스템에서 처리되는 데이터는 저장공간의 데이터베이스에 저장된다.

(3) 저장공간에는 기 운영 중인 데이터베이스와 파일서버가 있으면 본 연구를 위해 분석용 데이터를 별도 저장하는 학습분석 DB 를 논리적으로 추가하였다.



(그림 1) 중도이탈 예측 시스템 구성도

3.2 개발환경

본 논문의 개발환경은 <표 1>과 같다.

<표 1> 개발환경

환경		OS 및 플랫폼
이탈예측 시스템	분석서버	Windows2016, VSCode,python
	웹서버	Windows2016, Flask, python
온라인 교육시스템	LMS 서버	Linux, php
	학사관리	Windows2016, Spring, java
저장공간	DB 서버	Windows2016, MS-SQL2019
	파일서버	Windows2016

본 연구에서는 기존의 LMS 과 학사관리 시스템을 그대로 활용하였으며 LMS 는 moodle 기반의 LMS 로 linux 서버의 php 로 개발되었으며, 학사관리시스템은 윈도우 기반의 서버에 java 로 개발된 시스템이다.본 연구의 핵심인 이탈예측시스템은 윈도우 기반의 분석 서버에서 데이터수집, 전 처리 및 이탈예측 모델 생성을 위해 VScode 환경에서 python 으로 개발하였다. 이탈예측분석 결과를 보여주기 위한 이탈예측 웹서버는 윈도우 기반, python 으로 개발하여 Flask 웹서버로 구성하였다. LMS 학습이력 및 학사관리, 이탈예측모델의 데이터 저장을 위한 데이터베이스는 MS-SQL2019 를 사용하였다.

3.3. 구현방법

구현방법은 그림 1 시스템 구성도에 따라 다음 8 가지 단계로 이루어졌다.

Step1: 예측이탈모델의 feature 인 <표 2>의 데이터를 LMS DB 와 학사관리 DB 에서 지난 2020 년 1 학기에

서 2023 년 2 학기까지의 누적된 데이터 3,318,418 건을 학습분석 DB 에 저장하였다.

이탈예측의 주요 변인인 feature 는 알고리즘과 함께 선행연구[7]를 바탕으로 <표 2>와 같이 선택하였다. 이는 학습성과에 영향을 미치는 항목으로 온라인교육의 경우 주차, 차시 등으로 출석 및 학습활동을 평가하고 있기 때문에 분석단위를 주차로 지정하여 진행하였다. 본 연구는 과목별 이탈여부를 확인하기 위한 연구로 종속변인 이탈여부는 성적을 기준으로 해당 과목을 정상적으로 이수한 경우와 그렇지 못한 경우로 성적이 A~D 인 경우 이탈이 아님, F 또는 NULL(휴학,자퇴)인 경우 이탈로 설정하였다. 여기서 학습이력 및 학생정보 데이터를 조회하는 기능은 DB 서버에 View 와 프로시저로 개발하여 반복사용 및 수정이 용이하도록 하였다.

<표 2> 이탈예측모델의 feature

No	분석데이터	설명
1	분석단위	년도, 학기, 주차, 고유아이디, 과목코드,
2	학생정보	학과, 학년, 등록학기, 복수전공, 현수강 과목 수, 나이, 성별, 최종학력, 국적, 입학학년, 장애여부
3	학습이력	누적 접속 수, 평균접속횟수, 강의출석율, 강의수강시간(누적), 해당 주차 수강시간, 과제제출율, 시험참여율, 공지사항확인율, 기타학습활동참여율
4	이탈여부	성적(A~D): 아니요, 성적(F, NULL): 예

Step2: 학습분석 DB 로부터 이탈예측모델을 생성하기 위한 데이터를 조회하여 모델생성의 입력 항목에 맞도록 전 처리를 수행하였다.

전 처리는 모델 생성과 실제 이탈분석 용 데이터에 동일하게 사용하기 위해 공통 모듈로 개발하여 사용하였으며 데이터 정규화, 이상치 처리, 이탈 데이터에 대한 OverSampling 을 수행하였다.

Step3: 이탈예측 모델에 적합한 알고리즘으로 RandomForest 를 선택하고 수집한 데이터를 7:3 으로 분할하여 학습과 평가를 진행하여 모델 생성하고 평가를 진행, 성능을 확인하였다.

알고리즘은 의사결정트리(Decision Tree), 서포트 벡터머신(SVM), RandomForest, 딥러닝 알고리즘인 DNN 의 성과를 분석한 선행연구[7] 결과를 참고하였으며 성능은 정확도 0.96%, 재현율 0.76%, 정밀도 0.96%, F1-Score 0.84, AUC 는 0.95 이다

Step4: 생성한 모델은 이탈 예측 웹서버와 공유할 수 있도록 파일 서버에 저장하였다.

모델은 model_년도학기.pkl 로 버전을 관리할 수 있도록 저장하였다.

Step5: 이탈예측을 하기 위한 학습자의 데이터를 현재 주차 기준으로 LMS DB 및 학사관리DB에서 <표 2>에 대해서 조회 후 step2와 동일한 전처리를 수행하였다.

Step6: 파일 서버에 저장된 이탈예측 모델을 읽어온다.

Step7: step5에서 전처리를 거친 데이터를 step4에서 불러온 이탈예측모델에 입력하여 이탈예측분석을 수행하였다.

Step8: 분석된 결과는 JSON 형태로 웹페이지로 전달되어 화면에 출력되도록 하였다.

결과를 보여주기 위해 본 연구에서는 다음 두가지방법을 사용하였다.

(1) LMS의 학습현황에서 학번을 통해 이탈예측 웹서버에 ajax로 결과를 요청하여 LMS 상의 웹페이지로 결과를 표시하도록 구현하였다. 이 방법은 기존 LMS에 학습지수 및 대시보드를 구현한 웹페이지가 추가되어야 하지만 기존의 화면에서 바로 예측 결과를 조회할 수 있는 장점이 있지만 기존 시스템의 수정이 더 발생하는 단점이 있다.

(2) 학번과 인증정보를 이탈예측 웹서비스 전달하여 이탈예측 웹서비스 내 대시보드 화면을 개발하여 팝업창으로 출력하도록 하였다.이러한 방법은 기존 시스템에 최대한 수정을 가하지 않고 구현이 가능한 장점이 있다.

4. 결과

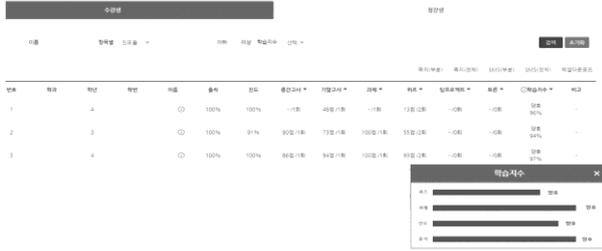
그림 3 과 같이 학습자 화면에서는 기존 나의 학습현황페이지내에 학습지수 항목을 추가하여 ajax 를 통해 이탈예측결과를 조회한 후 양호, 주의, 위험으로 표기하고 학습지수를 클릭 시 레이어 팝업으로 주요 예측 변인인 출석, 과제, 진도, 시험에 대한 상세한 현황을 막대 그래프로 시각화 하여 제공하였다.



(그림 3) 학습자 학습지수 대쉬보드.

교수자의 경우 그림 4 와 같이 학생관리 기능에 학

습지수를 양호, 주의, 위험으로 표기하고 클릭 시 학생의 상세 학습지수 데이터를 그래프로 결과를 확인할 수 있도록 하였다.



(그림4) 교수자 학습지수 대시보드.

5. 결론

본 논문에서 제안된 시스템은 과목별 학습 이력 및 학생정보를 수집하여 하나의 과목에 대한 중도이탈 위험성 정보를 학습자 본인 및 교수자, 관리자에게 제공하여 하였으며 이를 통해 학습자에게는 과목별 학습을 완료하기 위한 활동에 대한 우선순위 및 동기 부여가 가능하도록 하였으며, 과목 운영 교수자는 학생 별 맞춤형 피드백을 제시하거나 학습부진자들 만을 위한 활동을 추가적으로 진행할 수 있게 하였다. 이로써 전반적인 강의운영에서 보완할 부분을 파악하고 개선할 수 있도록 도움을 줄수 있음을 확인하였다. 향후 학생특성정보와 과목별이 아닌 학생 별 학습이력 데이터를 활용하여 다음학기 재등록 여부에 대한 예측정보를 제공하는 학생 별 온라인학습 중도이탈 예측 시스템개발에 대한 연구로 확대하여 과목별 , 학생별 중도이탈 예측시스템으로 확장할 수 있을 것이다.

참고문헌

[1] Tanya Elias, “Learning Analytics: Definitions, Processes and Potential”, Tanya Elias January, 2011

[2] 조용상, “학습 분석 기술 활용 가능성 및 전망”, 한국교육학술정보원(KERIS), 2013 KERIS 이슈리포트 , 2014

[3] Siemens, G., & Long, P, “Penetrating the Fog: Analytics in Learning and Education”, EDUCAUSE , 31-40, 2011

[4] 최재원, “학습과정 관리를 위한 데이터사이언스 활용 : 위험그룹 예측을 중심으로”, 아주대학교 박사학위 청구논문, 1-102, 2016

[5] 이정은, “대학 신입생 중도탈락 예측 요인 분석: S 대학 사례를 중심으로”, 2021

[6] 박혜진, “온라인 학습자의 중도탈락 예측 요인 탐색: 랜덤 포레스트를 적용하여”, Journal of Educational Technology 2022, Vol 38, No 1, 297-332 , 2022

[7] HS Park, “Early dropout prediction in online learning of university using machine learning”, JOIV , 347-353 ,

2021

[8] 서진선, “LMS 로그 데이터 분석을 통한 온라인 학업 성취 예측모형에 관한 사례 연구”, Journal of Field-based Lesson Studies, 2020, 제 1 권 제 2 호, 121 ~ 154, 2020

[9] 계보경, “지능형 학습 분석을 위한 데이터 수집·분석 API 고도와 연구”, 한국교육학술정보원 (KERIS), 연구보고 RR 2018-9, 2018

[10] 정제영, “빅데이터를 활용한 학업중단 학생 대응 모델 연구”, 한국교육학술정보원(KERIS), 연구보고 RR 2018-10, 2018

[11] 이혜윤, 대학 이러닝 환경에서 학습자 행동 로그에 기반한 군집별 학업성취 예측모형 비교, Journal of Educational Studies, Vol.49 No.1, 127-150, 2018

[12] 김경태, 딥 러닝과 Boosted Decision Tree 를 활용한 고객 이탈 예측 모델, Journal of Educational Studies , Vol.49, No1, 127-150, 2018