

비선형 회귀를 이용한 학습도우미 애플리케이션

장은영 김강우 김민식 류다은 박승묵 고병철

계명대학교

okjj0217@naver.com kim49197@naver.com ckmy9355@naver.com
moderato2036@gmail.com 3mist@naver.com niceko@kmu.ac.kr

Learning Assistant Application Using Non-Linear Regression

Jang, Eun-yeong Kim, Kang-Woo Kim, Min-Sik Ryu, Da-Eun Park, Seung-Mook Ko,
 Byung-Chul
 Keimyung University

요약

코로나 19로 대학교 강인들이 비대면 방식으로 전환되고 있는데, 기존의 교수학습 지원센터는 웹 환경만을 제공한다. 따라서 본 논문에서는 모바일 애플리케이션을 통해 수강생들이 교수학습 지원센터에 쉽게 접근할 수 있도록 도와주는 시스템을 개발하였다. 애플리케이션에서 학생들의 강의 시간 및 시험, 과제 등의 일정을 관리해주고, 푸시 알림을 제공해주는 학습 도우미의 역할을 수행한다. 뿐만 아니라 직관적인 인터페이스, 다크 모드, scroll-to-top 버튼 등을 고려한 디자인으로 사용자의 편리함을 도모한다. 학습 도우미 애플리케이션의 가장 핵심기능 중 하나는 머신러닝 기법 중 비선형 회귀(Non-Linear Regression)을 이용해 성적 데이터를 분석해주는 차별화된 기능이다. 이를 위해 최종적인 성적을 종속변수, 일정 기간까지의 성적을 독립변수로 설정하여 기존의 성적 데이터를 바탕으로 종속변수인 최종성적을 랜덤 포레스트 비선형 회귀분석으로 예측하는 알고리즘을 제시하고자 한다.

1. 서론

코로나 19로 인하여 대면으로 진행되던 대학교의 기존 강인들이 대다수 비대면 강의로 전환되었다. 비대면 강의는 실시간 강의-동영상 강의 등의 형태로 이루어지고 있으며 강의의 과제, 시험 등 전반적인 부분이 교수학습지원시스템에 의지하고 있다. 대면 강의를 접해 온 학생들은 교수학습지원시스템에 의존적인 새로운 수업방식이 익숙하지 않아 교과목의 공지를 놓치거나 과제 제출기한을 넘겨 제출하지 못하는 등의 어려움을 겪고 있다. 이러한 문제점에 착안해 모바일 웹 환경으로만 제공되던 기존의 교수학습지원시스템을 별도의 애플리케이션으로 제작의 필요성을 생각하게 되었다. 애플리케이션을 제작함으로써 교수학습지원시스템의 접근성을 향상하는 것을 목적으로 한다.

애플리케이션에서는 교수학습 지원센터의 기존 문제점들을 개선하여 학생들이 공지사항과 같은 중요사항들을 보다 빨리 접할 수 있게 된다. 학생들의 강의 시간 및 시험, 과제 등의 일정을 애플리케이션에서 관리해 줌으로써 학생들의 학업 편의를 도모한다. 공지사항이 업로드되는 즉시 푸시 알림을 통해 알 수 있으므로, 교수진 측에서는 SMS를 통한 별도의 안내를 하지 않아도 되고 학생 측에서는 빠르게 공지사항을 확인할 수 있게 되며 그에 따른 시간적, 경제적 절약을 기대할 수 있다. 애플리케이션이 사용자에게 예상되는 성적을 보여주는 등의 다양한 서비스를 제공하여 학습 도우미 역할을 수행해 학업능률을 증진한다.

제안하는 학습도우미 애플리케이션은 다양한 차별성을 가진다. 먼저 전과목 공지를 한 번에 확인할 수 있는 공간이 있어 공지 확인 시간이 단축된다. GUI를 적극적으로 활용해 일정을 한눈에 보여줌으로써 과제 진행의 순서, 시험 공부계획 등을 세우는 데에 있어 도움을 준다. 또한 사용자별로 개인화 서비스를 제공한다. 이 기능은 개인별로 과목의 수강 진행도를 보여주거나 같이 수강하는 학생들의 데이터를 토대로 시험 예상 성적을 보여주어 학습에 도움을 준다. 2장과 3장에서는 성적 데이터 분석에 활용되는 머신러닝 기법과 성적예측 알고리즘을 소개하고 4장에서는 학습도우미 서버 프로그램에 대해서 설명한다.

2. 머신러닝을 통한 성적 예측

성적은 여러 변수의 영향을 받는다. 사용자의 학습 시간, 이해도, 논리력, 전달성 등 매우 다양하며, 이러한 변수들에 영향을 끼치는 요소들을 포함한다면 고려하여야 하는 요소들이 매우 방대해진다. 이러한 요소들을 모두 고려하여 성적을 예측하는 것은 매우 어려운 일이 될 것이다. 성적 예측에 있어 존재하는 이러한 어려움을 해결하기 위하여, 본 연구에서는 머신러닝(Machine learning) 기법 중 랜덤 포레스트(Random Forest)를 사용한 회귀 분석(Regression analysis)을 통하여 최종적인 성적을 종속변수, 일정 기간까지의 성적을 독립변수로 삼아 이전까지의 성적 데이터와 비교하여 종속변수인 최종성적을 예측하는 방법을 제시하고자 한다.

3. 성적 예측 알고리즘

성적 데이터를 분석할 때 최종성과 각 항목 간의 관계, 부가적으로 최종성적을 제외한 각 항목 간의 인과관계 총 두 방법을 분석하고 예측하였다. 이때 사용한 머신러닝모델은 랜덤 포레스트 회귀기(Random

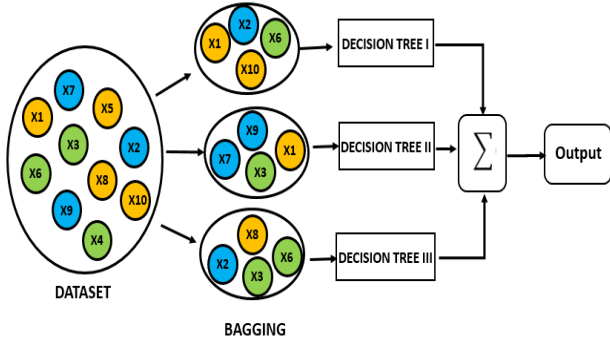


그림 1 . 랜덤 포레스트 회귀기(Random Forest Regressor)

Forest Regressor)이다[1].

첫 번째 방법으로 기존에 존재하는 성적 데이터 셋을 훈련 셋과 테스트 셋으로 반의 비율로 나누어 랜덤 포레스트 모델에 종속변수로 최종성적을, 독립변수로 그 외 항목들을 적용해 봄으로 얼마나 잘 학습하는지 알아보았다. 그 결과 결정 계수의 값이 0.95 이상이 나오므로 학습이 잘 되었다고 볼 수 있다. 독립변수로 중간고사까지의 과제포함 성적을 넣어 학습한 결과, 결정 계수의 값이 0.87이 나와 최종성적을 예측 가능한 수준까지 학습할 수 있었다. 또 랜덤 포레스트로 각 특징(변수)의 상대적 중요도(importance)를 측정할 수 있다. 여기서 실제 데이터에서 가장 중요한 중요도와 머신러닝을 통해 밝혀낸 중요도가 다르다는 점을 확인했다. 예로 최종성적을 산출할 때 기말 40%, 중간 30%로 비중을 두었다면, 랜덤 포레스트로 산출한 중요도는 기말 56%, 중간 18%로 값이 나왔다는 점이다. 이러한 결과를 통해 항목별로 적용된 비중이 실제 적용된 비중과 차이를 수치화하여 나타낼 수 있었다.

두 번째로 각 항목 간의 인과관계를 알아보기 위해 훈련 셋과 테스트 셋을 같은 방법으로 정리한 후에 종속변수에 기말고사 점수를, 독립변수에 중간시험까지의 과제포함 성적을 넣어 학습한 결과를 산출하였다. 그 결과 결정계수가 0.43으로 예측한다고 하지 못할 값을 보여주었다. 실제 기말점수가 52점이라면, 예측 기말점수는 99점까지도 값이 나왔다. 문제점을 보완하기 위해 ICE (Individual Conditional Expectation : 개별조건부 기대치) plot을 통하여 종속변수에 독립변수 하나씩 개별의 그래프를 나타내어 보여주므로 변수 간의 인과관계를 자세히 탐색할 수 있었다. 이 결과 랜덤 포레스트로 정확히 예측하지 못한 결과도 그래프 분석을 통해 변수 간의 대략적인 관계성을 알 수 있었다. 사용자는 ICE plot으로 만들어진 그래프를 활용하여 예측된 성적을 시각화하여 볼 수 있다.

그림 2와 같이 x축인 기말고사(last)에 따라 y축인 예측성적(Predicted Score)의 결과가 증가하는 것을 알 수 있다. 가는 실선은 결정트리에서 생성된 개별 회귀자들이며 굵은 붉은색 실선은 각 결정트리 회귀자의 평균을 취한 최종 회귀함수이다.

4. 학습도우미 애플리케이션 Server

4.1 Back end

웹 또는 애플리케이션에서 받은 데이터를 저장 및 처리 할 수 있는

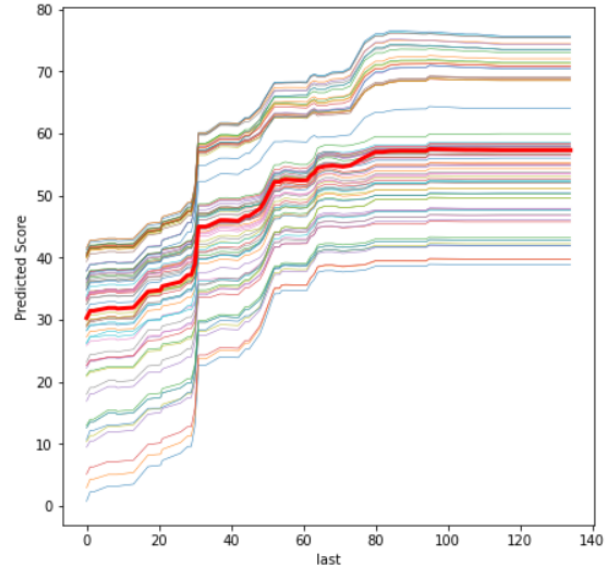


그림 2 . 랜덤 포레스트 회귀 결과

데이터베이스를 구축하기 위해 AWS EC2에 MySQL을 직접 설치하였다. MySQL서버에 생성한 data table의 모습은 그림 3 과 같다. 교수학습지원시스템에 필요한 데이터로는 기본적으로 로그인(로그인ID, ID, PW, 권한), 학생(학번, 학년, 이름, 전공), 교수(교수코드, 이름, 전공), 과목(과목코드, 과목명, 강의실, 시간, 학생수, 교수코드, 실시간URL) 정보가 있고, 그 외에는 과제(과제ID, 성적ID, 과제명, 과제내용, 날짜, 점수, 제출여부), 시험(시험ID, 성적ID, 날짜, 강의실, 점수, 응시여부), 출석률(출석ID, 성적ID, 날짜, 출결, 성적(성적ID, 과목코드, 학번, 최종성적, 석차, 반영비율), 공지사항(공지ID, 과목코드, 제목, 내용, 날짜)이 있다.

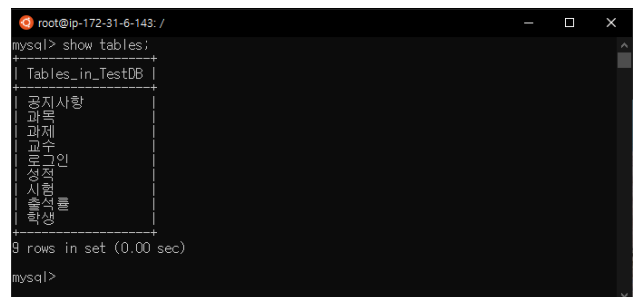


그림 3 . Database Table 목록

성적 테이블의 성적 ID는 과목 코드와 학번 정보로 개개인의 특정 과목 성적 데이터를 구분하는 기본 키이다. 따라서 성적과 관련된 과제 및 시험, 출석률 테이블들과 성적 ID를 외래키로 묶어 각각의 점수를 가져올 수 있게 하였다.

4.2 Front end

DB에 저장된 데이터를 이용하여 주요 일정 및 중요 알람과 출석, 과제, 시험에 해당하는 성적을 시각화하였다. 시각화는 그림 4 와 같이 크게 웹과 애플리케이션 2가지로 개발하였고, 그림 5에서 두 플랫폼 모두 다크 모드를 지원한다는 것을 알 수 있다.

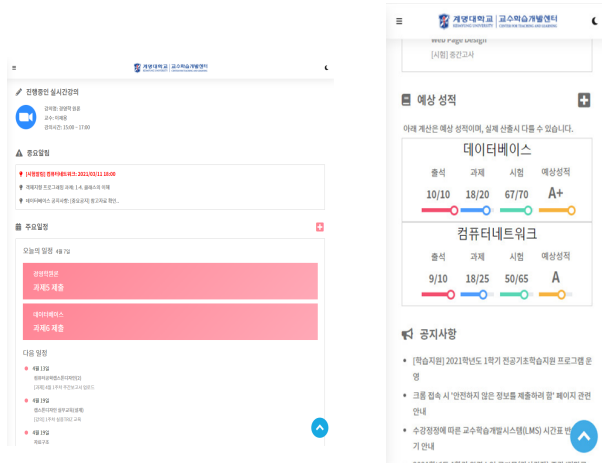


그림 4 . 웹 및 애플리케이션 시각화

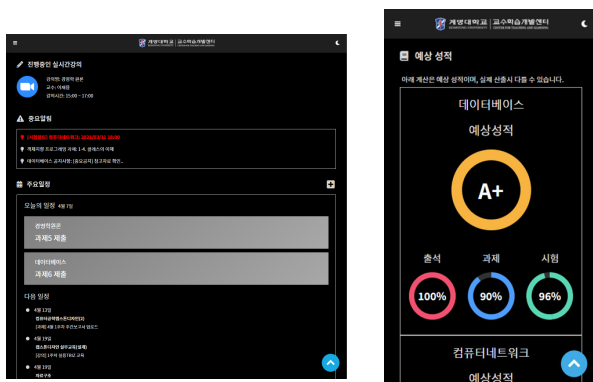


그림 5 . 다크모드

5. 결론

본 논문에서는 기존 교수학습지원시스템의 디자인을 개선하고 다크 모드, scroll-to-top 버튼 등 편의성을 강화하고 직관적인 인터페이스를 구성함으로써 사용자 경험을 향상한 새로운 교수학습지원시스템을 제시하였다. 또한 학업에 참고할 수 있는 성적 예측 기능을 지원하며, 신뢰도를 높이기 위해 머신러닝을 이용하여 기존 성적 데이터셋을 학습하고 이를 기반으로 현재 학생의 예상 성적을 제공하는 방법을 제시하였다. 다만, 개발단계에서는 학습과정에 사용한 실제 성적 데이터셋의 양이 많지 않아 실 서버 적용시 기대와 다른 학습결과가 나올 수 있으므로 반드시 더 많은 실제 성적 데이터셋을 통한 학습이 필요할 것이다.

감사의 글

본 논문은 교육부와 한국연구재단의 계명대학교 대학혁신지원사업비를 지원받아 수행된 것입니다.

참고문헌

- [1] Jocelyn D'Souza "A trip to Random Forest..."
<https://medium.com/greyatom/a-trip-to-random-forest-5c30d8250d6a>, 2018.03.20