

학습자 정밀 평가를 위한 개인 맞춤형 테스트셋 생성 시스템

김연찬, 운봉영, 아가르왈 판카즈
(주)태그하이프

ychan@tag-hive.com, wbyoung@tag-hive.com, pankaj@tag-hive.com

Personalized Test Set Generating System for Advanced Student Assessment

Kim Yeoun Chan, Woon Bong Young, Agarwal Pankaj
TagHive, Inc.

요 약

본 논문은 코로나 시대에 주목을 받고 있는 에듀테크 시장에서 인공지능 기술의 주요 역할인 학습자 평가를 개별 문제 풀이가 아닌 다양한 난이도와 문제 배열 순서를 지닌 시험 상황에서 정밀 평가하기 위한 테스트셋 생성 시스템을 제안한다. 본 시스템은 Random Forest 분류 알고리즘을 사용한 정답률 예측 모델과 문제 배열 순서를 최적화하는 모델로 구성하였다. 본 연구에서 제안한 시스템을 적용할 경우 적용하지 않은 비교군에 비하여 예상 시험 점수에서 높은 성능을 확인하였다.

I. 서론

최근 COVID-19의 영향으로 비대면 교육의 수요가 증가하면서 에듀테크(Edu-Tech) 시장에 대한 관심이 급격하게 높아지고 있다. 4차 산업혁명의 혁신적인 기술을 교육 분야에 접목하는 연구가 다방면으로 진행되고 있으며 특히 인공지능은 온라인 환경에서 학습 도중 발생하는 정답 유무, 문제 풀이 시간 등 다양한 데이터를 분석하여 개인 맞춤형 학습 및 고도화된 개인 평가를 제공하는 서비스에 주로 사용되고 있다[1]. 학습자의 평가에 있어 개별 문제 풀이 과정을 이용하는 연구[2,3]는 다양하게 진행되고 있지만 다양한 난이도와 문제 배열 순서의 변수를 가진 시험 상황에서의 연구는 부족한 편이다. 본 논문에서는 시험 상황에서 학습자를 정밀하게 평가하기 위한 개인 맞춤형 테스트셋 생성 방법을 제안한다.

II. 개인 맞춤형 테스트셋 생성 시스템

본 논문에서는 그림 1과 같이 정답률 예측 모델과 문제 배열 순서 최적화 모델 기반의 학습자 정밀 평가가 목적인 개인 맞춤형 테스트셋을 생성하는 시스템을 제안한다.

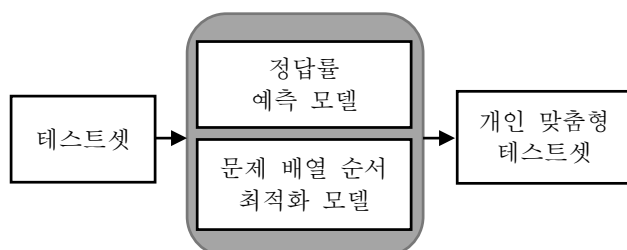


그림 1. 개인 맞춤형 테스트셋 생성 시스템 블록도

A. 정답률 예측 모델

본 모델은 분류 알고리즘을 유저들의 과거 학습 데이터를 이용하여 최적화한 후 정답 확률을 예측하도록 설계했다. 입력 데이터는 학습자의 개념 이해도, 문제의 난이도, 과거 테스트셋 내에서의 문제의 순번, 평균 문제 풀이 시간으로 이루어져 있으며 정답률을 예측하도록 학습한다. 개념 이해도는 학습자별 해당 문제에 요구되는 개념의 이해도를 0과 1 사이의 실숫값으로 수치화 하였으며 개념의 이해도가 높아질수록 1에 가까워진다. 문제의 난이도 역시 0과 1 사이의 실숫값으로 1에 가까울수록 높은 난이도를 의미한다. 평균 문제 풀이 시간은 각 문제의 전체 유저 학습 데이터를 기반으로 계산되었으며 온라인 환경에서 발생할 수 있는 접속 지연 문제, 디바이스 에러 등의 변수를 고려하여 상 하위 10%를 제외한 평균값을 사용한다. 과거 테스트셋에서 문제의 순번은 n 개의 문제로 이루어진 과거의 테스트셋에서 해당 문제가 배열된 순번을 의미한다. 입력 데이터는 인도의 6-10학년 수학, 과학 과목이 대상인 본 연구팀의 학습 서비스 "Class Saathi"의 모의 시험 데이터를 사용하였으며 모의 시험은 학습자들이 시험 범위를 설정하면 해당 범위 내의 테스트셋을 제공한다. 제시된 입력 데이터를 이용하여 Random Forest, Support Vector Machine, K-Nearest Neighbors, Naïve Bayes 분류 알고리즘을 적용하여 성능을 비교하였다. 분류 알고리즘은 확률을 예측하여 50% 이상의 확률을 가지면 1, 반대의 경우 0을 출력하도록 초기 설계되어 있으나 분류 알고리즘의 신뢰도를 높이기 위하여 60% 이상의 확률을 가질 때 1, 반대의 경우 0을 출력하도록 설계한다. 성능 측정 결과, 표 1과 같이 Random Forest 분류 알고리즘을 적용하였을 때 가장 높은 성능을 보이며 정확도는 83.33%로 측정된다. 따라서 본 모델은 Random Forest 분류 알고리즘을 사용하여 정답률을 예측한다.

표 1. 분류 알고리즘의 성능 비교

Metric	Random Forest	SVM	KNN	Naïve Bayes
F1-score	0.8895	0.8543	0.8313	0.7103
AUC	0.7707	0.5008	0.5673	0.5619
ACC	0.8333	0.746	0.7288	0.7103

B. 문제 배열 순서 최적화 모델

본 모델은 학습자의 개념 이해도와 테스트 범위를 기준으로 무작위의 순서로 선별된 n 개의 문제를 A에서 제시한 정답률 예측 모델을 사용하여 가장 고득점이 예상되는 문제 배열 순서의 개인 맞춤형 테스트셋을 생성한다. 해당 모델에서는 정답률 예측 모델이 1을 출력할 경우 정답 예상 문제, 0을 출력할 경우 오답 예상 문제로 정의한다. 본 모델은 무작위로 배치된 n 개의 문제에 대하여 아래의 순서로 진행된다.

- 1) 초기 문제 배열 순서는 테스트셋에 임의의 배열 순서를 $n/2$ 번 부여하며 그 중 예측 정답률의 평균이 가장 높은 문제 배열 순서를 테스트셋에 부여한다
- 2) (1)의 과정 후 정답 예상 문제의 순번을 고정하고 나머지 순번을 오답 예상 문제에 무작위로 부여한 후 정답률을 예측한다
- 3) (2)의 과정을 오답 예상 문제의 수가 최소화될 때까지 반복한다
- 4) (3)의 과정 후 오답 예상 문제의 개수가 최소 두 개 이상이면 나머지 순번을 무작위로 부여하며 오답 예상 문제 집합의 예측 정답률 평균이 가장 높아지도록 순번을 부여한다

III. 실험 및 결과

본 시스템의 성능을 검증하기 위하여 해당 시스템 외 난이도의 오름차순 및 내림차순을 활용한 문제 배열 순서에 관한 연구[4]를 참고하여 아래와 같이 비교군을 정의한다.

- 1) 무작위로 순번을 배열한 경우
- 2) 난이도의 오름차순으로 배열한 경우
- 3) 난이도의 내림차순으로 배열한 경우

총 100개의 테스트셋에 제안 방법을 포함한 네 가지 문제 배열 순서를 적용한 후 정답률 예측 모델을 사용하여 각 테스트셋 별 시험 점수를 예측한다.

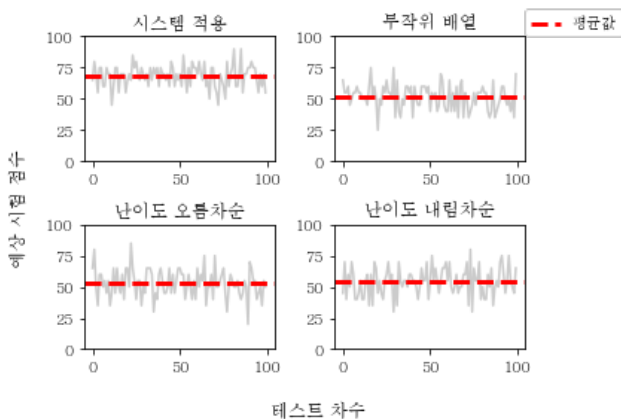


그림 2. 문제 배열 순서 별 예상 시험 점수

실험 결과, 그림 2와 같이 본 시스템을 적용하였을 때 예상 시험 점수의 평균값이 가장 높다는 점을 확인하였다. 문제 배열 순서 별 각 시험 점수의 평균값은 표 2와 같으며 본 시스템의 경우 시스템을 적용하지 않은 경우 대비 평균 29.42% 높은 성능을 확인하였다.

표 2. 문제 배열 순서에 따른 예상 시험 점수 평균

Metric	시스템 적용	무작위 배열	난이도 오름차순	난이도 내림차순
예상 시험 점수 평균	68.3	51.4	52.8	54.2

IV. 결론

본 논문에서는 학습자의 정밀 평가를 개별 문제 풀이 과정이 아닌 난이도가 다양하고 문제 배열 순서의 변수를 가진 시험 상황에서 적용할 수 있는 방법으로 과거 학습 데이터 기반의 정답률 예측 모델과 문제 배열 순서 최적화 모델을 사용한 개인 맞춤형 테스트셋 생성 시스템을 설계하였다. 본 시스템을 통한 개인 맞춤형 테스트셋의 경우 비교군 대비 29.42% 우수한 성능을 확인하였다. 추후 개인 맞춤형 테스트셋의 신뢰도 증가를 위해 정답률 예측 모델의 분류 알고리즘을 개선하여 성능을 향상하고 추가로 개인 맞춤형 테스트셋을 통한 정밀 평가 방식을 연구할 계획이다.

참 고 문 헌

- [1] Luan, Hui, et al. "Challenges and future directions of big data and artificial intelligence in education." *Frontiers in psychology* 11 (2020).
- [2] 김성훈, et al. "설명 가능한 AI 학습 지원 시스템 개발." *컴퓨터교육학회 논문지* 24.1 (2021): 107-115.
- [3] Beck, Joseph E., and Beverly Park Woolf. "High-level student modeling with machine learning." *International Conference on Intelligent Tutoring Systems*. Springer, Berlin, Heidelberg, 2000.
- [4] 김재철, 김성훈, and 김선희. "문항배열순서가 변별도와 난이도에 미치는 영향." *교육평가연구* 22 (2009): 409-427.