

사전학습모델을 활용한 수학학습 도구 자동 생성 시스템

노명성^o

^o서울교육대학교 초등컴퓨터교육과

e-mail: gurwosh@gmail.com^o

Automatic Generation System of Mathematical Learning Tools Using Pretrained Models

Myong-Sung No^o

^oDept. of Computer education in elementary, Seoul National University Of Education

● 요약 ●

본 논문에서는 사전학습모델을 활용한 수학학습 도구 자동 생성 시스템을 제안한다. 본 시스템은 사전학습모델을 활용하여 수학학습 도구를 교과과정 및 단원, 유형별로 다각화하여 자동 생성하고 사전학습모델을 자체 구축한 Dataset을 이용해 Fine-tuning하여 학생들에게 적절한 학습 도구와 적절치 않은 학습 도구를 분류하여 학습 도구의 품질을 높였다. 본 시스템을 활용하여 학생들에게 양질의 수학학습 도구를 많은 양으로 제공해 줄 수 있는 초석을 다졌으며, 추후 AI 교과서와의 융합연구의 가능성도 열게 되었다.

키워드: 학습 도구(learning tools), 사전학습모델(pretrained models), 자동 생성(automatic generation)

I. Introduction

최근 교육현장에서는 코로나19의 여파를 거치면서 교육격차에 대한 논의가 활발하게 이루어지고 있다. 특히 수학 교과 같은 경우 양극화 양상이 나타나고 있다는 점에서 학생에게 양질의 학습 도구를 제공하는 것을 통해 이러한 문제를 해소하는 데 도움을 줄 수 있을 것으로 보인다. 하지만 양질의 학습 도구를 다량으로 생산하는 것은 전문인력이 필요하고 시간적, 물리적 비용이 크게 소모가 된다. 때문에 본 연구에서는 인공지능을 활용하여 수학학습 도구를 자동으로 생성하는 시스템을 구축하고 또한 부적절한 학습도구를 필터링하는 시스템을 추가적으로 구축하여 학생에게 적절한 수학학습 도구를 제공할 예정이다.

Table 1. Prompt Dataset Category

Id	Value(Example)
단원	세 자리수의 덧셈과 뺄셈
유형	문장형
영역	연산
난이도	상
주제	일상 속 이야기

II. System Development

1. 수학학습 도구 생성 자동화 과정

본 연구에서 수학학습 도구 생성을 자동화하기 위해 미리 준비된 Prompt Dataset을 이용하여 사전학습모델을 Fine-Tuning하여 활용하였다. 생성된 학습도구는 유형별, 교과과정별로 분류되어 데이터베이스에 저장된다.

2. Filtering Dataset 구축

1차적으로 사전학습된 모델을 통해 생성된 학습 도구는 학생에게 적절치 않은 내용이 포함될 수 있다. 예를 들면 전문적인 단어, 비속어 혹은 차별적인 단어, 교육과정에 적합하지 않은 문장등이 있을 수 있다. 때문에 이러한 데이터들을 모아 '적절한 것', '적절치 않은 것'으로 Labeling하여 Filtering Dataset을 구축하고 BERT기반의 언어모델에 Fine-Tuning하여 필터링 시스템을 구축할 수 있다.

{sentence: '철수의 키는 157cm이고 영희의 키는 148cm일때 키의 차이는?', label: '0'}
{sentence: '한 학회에서 1년간 발표된 논문이 15672편일때 하루평균 발표된 논문수는?', label: 'X'}
{sentence: '한변의 길이가 5cm인 정사각형의 넓이를 구하십시오.', label: '0'}
{sentence: '바보경진대회 우승자의 상금은 500달러이다 100달러를 가진이가 우승했을때 총 액수는?', label: 'X'}

Fig. 1. Filtering Dataset

3. Filtering System 구축

Filtering System은 사전학습모델을 통해 Fine-tuning된 언어모델을 활용하여 생성된 학습 도구를 검토하고 적절한 도구와 적절치 않은 도구로 분류하게 된다. 적절하게 판명된 학습 도구는 전문가의 최종 검토를 거쳐 최종 데이터베이스에 저장 후 학생에게 출력되고 그렇지 않은 도구는 Filtering Dataset으로 환원되어 언어모델의 성능향상에 이용된다.

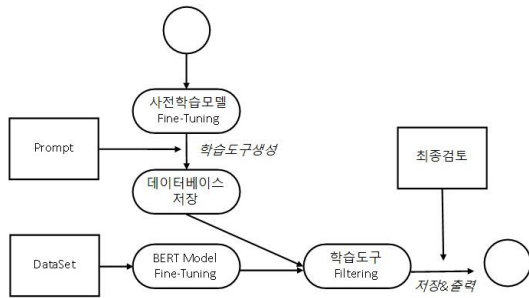


Fig. 2. System Architecture

III. Conclusions

해당 연구에서는 수학학습 도구를 사전학습모델을 활용하여 학습 도구 생성 과정을 자동화하고 이를 학습에 적절한 도구로 분류하여 학생들에게 유의미한 학습 도구를 제공할 수 있는 시스템의 초석을 다졌다. 추후에는 Prompt Dataset을 구체화하여 더욱 입체적인 수학 학습 도구를 생성하고 Filtering Dataset을 구체화하여 학생에게 보다 양질의 학습 도구를 제공해 줄 수 있도록 연구를 진행할 예정이다.

REFERENCES

- [1] Kim, Bum-Joo, "The Analysis of the Effect on Achievement of the COVID-19 Situation : Focusing on Middle Schools in the Metropolitan Area", The Korean Educational Administration Society, 39(4),107-129. October 2021.
- [2] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805. 2018..
- [3] Mosbach, Marius, Maksym Andriushchenko, and Dietrich Klakow. "On the stability of fine-tuning bert: Misconceptions, explanations, and strong baselines." arXiv preprint arXiv:2006.04884. 2020.
- [4] Brown, Tom, et al. "Language models are few-shot learners." Advances in neural information processing systems 33. 2020