PROPOSAL FOR CS 490 - INDEPENDENT STUDY - FALL 2024

Arsh Parmar, Hongmin Li
Department of Computer Science
California State University East Bay, Hayward
aparmar11@horizon.csueastbay.edu

June 19, 2024

ABSTRACT

Knowledge Tracing using Deep Learning techniques has great potential to improve present online learning systems in almost all subjects, especially, University-based and Massively Open Online Courses, which have always been considered inferior in comparison with physical classroom teaching due to their lack of personalized learning. This project would implement a Deep Knowledge Tracing (DKT) model and a question recommendation system that tailors quizzes to students' individual knowledge levels. This model would then be evaluated using the ASSISTments dataset, a benchmark dataset in the area with more than 25000 questions evaluated on 4000+ students. It would serve to promote student tailored learning in and outside the CSUEB. This project will be conducted by the undergraduate student working the supervision of Professor Hongmin Li.

1 Introduction and Motivation

In recent years, especially post-COVID, there has been massive growth in the number of online courses, from K12 up to the professional sphere, and these continue to grow in popularity due to ease of access, low cost, and remote modality.

However, in terms of credibility, these are largely still seen as ineffective, as on average, students mostly lose interest within a short span, leaving most of the course unattended. This is largely due to a lack of adaptiveness to the student's learning levels and also the limited engagement to keep the motivation aloft [1]. With the rise of Deep Learning and AI, the educational industry also awaits revolutionary change, using said technologies to tailor online course materials to each student's needs.

To this end, multiple approaches can be used to improve the personalized nature of course materials and questionnaires. Older models include those based on Bayesian Inference, a stochastic model. Machine learning based models include

Logistic Regression and the newer Deep Learning models, which we will work on. These use Neural Networks to compute probabilities for each individual based on multiple factors [2].

In summary, the motivation for exploring Deep Learning in online education is clear. It offers the potential to overcome current limitations, provide highly personalized learning experiences, and significantly improve student outcomes. By embracing these advancements, the educational industry can move towards a more inclusive, efficient, and effective future.

2 Methodology

2.1 Model for Knowledge Tracing

Deep Knowledge Tracing aims to understand the progress level of a student in a particular area, and based on initial parameters and sequentially processed responses to questions, it calculates probabilities pertaining to their understanding, which can then be used by the student to work on weak areas and improve them. Our implementation will utilize Recurrent Neural Networks, particularly LSTM (Long Short Term Memory) to model such data and interactions.

The data will first be preprocessed, which includes encoding question information (outcome, other metadata). This data will be fed to the LSTM to capture dependencies in the interactions with the students. The output provides a probability of answering the question correctly next time.

2.2 Dataset

The dataset (tentative) to be used for this project will be the ASSISTments dataset. ASSISTments is an online learning and assessment tool designed for teachers to provide personalized feedback to students in Math and Science.

The dataset is derived from the work completed in 2012, and is open-sourced by the authors for use in evaluation of models, competitions etc [3].

It consists of:

- Records of students-system interactions, capturing responses to various problems
- Timestamps for each interaction, enabling time-based analysis of learning behaviors.
- Information about each problem, including IDs, skill tags, and other relevant content.
- Correctness data, hints used, etc.

The dataset is available in CSV format. There are around 180,000 questions evaluated on around 46,000 students [2]. Overall, it is a valuable and reliable resource to evaluate the model that we would implement.

2.3 Implementation and Deployment

We will use TensorFlow to develop a basic RNN based model, and the decided dataset to train it. After this, we will create APIs using Node, where we will utilize the TensorFlow.js library to integrate our recommendation APIs with the implemented model. This will be deployed on a local server, and should be able to integrate with UI platforms.

3 Timeline

Period	**Tasks**
May-July 2024	Preliminary Research
August-September 2024	Development of ML Model
October 2024	Testing of model, Development of API
November 2024	Testing model-API integration
December 2024	Final Report

Table 1: Project Timeline

4 Outcomes and Deliverables

The minimum target would be successful implementation and dataset training of the DL model, and the implementation of question recommendation APIs.

A more desirable target would add a UI implementation to the mix, and possibly an analytics database.

References

[1] S. Loeb, "How effective is online learning? what the research does and doesn't tell us (opinion)," Apr 2024. [Online]. Available: https://www.edweek.org/technology/opinion-how-effective-is-online-learning-what-the-research-does-and-doesnt-tell-us/2020/03#:~:text=Some%20students%20do%20as%20well,to%20struggle%20even%20more%20online.

[2] G. Abdelrahman, Q. Wang, and B. Nunes, "Knowledge tracing: A survey," *ACM Comput. Surv.*, vol. 55, no. 11, feb 2023. [Online]. Available: https://doi.org/10.1145/3569576

[3] N. Wattiez, "Assistments data set 2012-2013," Feb 2021. [Online]. Available: https://www.kaggle.com/datasets/nicolaswattiez/skillbuilder-data-2009-2010