Deeper Knowledge Tracing

Overview

Knowledge tracing in general is the task of evaluating a student's level of proficiency in the subject matter. This evaluation comes from the student's interactions with the course. Data such as quiz scores, number of attempts at a question, or time since the last attempt of reviewing a certain topic are used as inputs to the knowledge tracing model, which predicts how likely the student is to answer questions from a certain topic correctly. This prediction is then further used to measure other metrics such as student overall performance.

In the past, people tried many approaches. In general, knowledge tracing approaches can be categorized into two main categories: traditional models and deep learning models. Traditional models mainly include Bayesian Knowledge Tracing (BKT) and factor analysis, while deep learning models can be further divided into five approaches. Memory structures, attention mechanisms, and graph representation learning are approaches that are based on advancements in deep learning. There are also textual features and forgetting features, which are approaches that emphasize the role of two different aspects of the model [1]. The work of Piech et al. [2] pioneered the wave of implementing deep learning in the field of knowledge tracing. They used a combination of a Recurrent Neural Network (RNN) and a Long Short-Term Memory (LSTM) model. Their work was followed by others like Xiong et al. [3]. Today, state-of-the-art models are all deep learning models.

In a previous semester, namely Fall 2021, another group worked on a similar project. They used data from different sources to predict the probabilities of students giving correct answers to questions. We plan to build on that work.

Project description

Knowledge tracing is a machine learning problem that focuses on modeling and predicting a learner's knowledge state over time as they interact with a sequence of educational items or exercises. The goal is to understand what a student knows and doesn't know at different points in their learning journey.

While state-of-the-art knowledge tracing models have shown notable advancements, there is a limited amount of research conducted on multimodal knowledge tracing. Very few datasets and models are capable of handling diverse question formats, which include images and mathematical equations. This hinders the development of more informative embedding representations for knowledge-tracing problems.

Problem to be solved and/or interesting ideas to be researched

1. Understanding the knowledge tracing problem domain.

2. Literature survey on different kinds of knowledge tracing models (i.e., Bayesian models and factor models)
3. Exploring multimodal and informative representation learning and datasets
4. Explore the use of the attention mechanism in knowledge tracing

Project Goals

By the end of this semester, we hope to have implemented the knowledge tracing model in many different modes. We will perform extensive research into the implementation of this model by both Recurrent Neural Networks and a Long Short Term Memory model. After performing this extensive research, we will begin implementing the model and train it using the ASSISTment data, data gathered from students using skill builder problem sets, and potentially ITS data. If we are able to have a functioning model working before the end of the semester, we hope to turn the model into a helpful product by possibly using the textbook data. We will work towards developing a model that achieves a reasonable accuracy. We will strive to meet all the deadlines we set in the below timeline for the project, especially the numbered milestones.

Projected Timeline

Weekly Tasks	Due date
Team Formation, Project Planning	02/05
Form Repo, initial project files, Research, Learn Basics, Conceptualize a demo	02/12
Continued research, data collection, data preprocessing, start implementation of minimal model	02/19
Deal with issues in initial model implementation	02/26
Finalize implementation of initial model (RNN and LSTM) - milestone 1	03/04
Experiment with a more advanced model	03/11
(spring break)	03/18
Experiment with a more advanced model	03/25
Implement a more advanced model - milestone 2	04/01

Start working on demo	04/08
Prepare and finalize a demo - Milestone 3	04/15
Finalize presentation	04/22
Final Presentation - Milestone 4	04/24

Proposed solution / foreseeable problems and potential pitfalls:

- Getting up to speed with new software tools for those not used to it
- Limited potential applications of the project
- Over-ambitious deadlines
- Getting a robust multimodal dataset if we decide to go in that direction

Solutions:

- Weekly meetings
- Staying true to deadlines
- Group chat to maintain clear communication between group members
- Sharing of useful resources

Implementation

Datasets:

ASSISTments2009:

https://sites.google.com/site/assistmentsdata/home/2009-2010-assistment-data

ASSISTments2012:

https://sites.google.com/site/assistmentsdata/datasets/2012-13-school-data-with-affect

ITS Data

Software:

Python, PyTorch, Pandas, Matplotlib, Jupyter Notebook, Docker

Resources

- https://youtube.com/playlist?list=PLblh5JKOoLUIxGDQs4LFFD--41Vzf-ME1&si=q_YuRgTbYK74Kpcd - ML basics playlist
- https://www.youtube.com/watch?v=8ITtYnhslvE - KT overview lecture
- http://neuralnetworksanddeeplearning.com/index.html
- https://github.com/bigdata-ustc/EduKTM

- https://youtube.com/playlist?list=PLqnslRFeH2UrcDBWF5mfPGpqQDSta6VK4&si=cQ3Wspa9DGeLtbnM - PyTorch tutorial

Group Membership

Member	Participation	Skills	Responsibilities
Abdulaziz Memesh	3 credits	Deep Learning, Python, PyTorch, Scikit-learn, Pandas, Matplotlib	Gleaning insights from research papers, building the model, data preprocessing
Wenying Wu	2 credits	Machine learning (Tensorflow), Web application development (Front-end: React, Back-end: Java, Python)	LIterature survey, playing with datasets and models, exploring some multimodal learning
Sabina Ajjan	2 credits	Machine learning, PyTorch (just started to learn in computer vision), Python	Researching,implementing model and turning model into product
Will Lash	1 Credit	Python, Java, C, C++, SQL, open to learning new tools	Generally fill in where needed, implement model
Zac Danziger	1 Credit	Machine Learning (In Progress), Python, Java, C++	Assist in implementing the model, fill in as needed

Communication, Collaboration, and File Management

- Weekly team meeting: 8 pm on Mondays
- Communication: Teams
- File management: GitHub, Teams, Google Drive

References

1. Ghodai Abdelrahman, Qing Wang, and Bernardo Nunes. 2023. Knowledge Tracing: A Survey. ACM Comput. Surv. 55, 11, Article 224 (November 2023), 37 pages. https://doi.org/10.1145/3569576
2. https://stanford.edu/~cpiech/bio/papers/deepKnowledgeTracing.pdf
3. https://www.educationaldatamining.org/EDM2016/proceedings/paper_133.pdf
4. https://link.springer.com/article/10.1007/s10489-022-04095-x