

Technology Review

on

NLP/AI techniques in Education

Due: 2022-11-06

CS 410: Text Information Systems

Fall 2022

by

ChengXiang Zhai

Professor

Department of Computer Science

University of Illinois at Urbana-Champaign

Submitted By

Ayush Khanna

akhanna6@illinois.edu



Contents

1	Project Topic	2
2	Introduction	2
3	Overview of AI/NLP techniques and tools used to build AI Education (AIEd)	2
4	Conclusion	5
5	References	5

1 Project Topic

Overview of AI Education (AIEd) and NLP techniques used by Riid, which are helping the Education System

2 Introduction

Natural language processing/NLP is one of the main areas in Artificial intelligence whose main focus is to make computers understand human language in a better way and then also recommend suggestions. Nowadays NLP techniques are used in every domain such as Medical, Science, Engineering and Education. Here, we focus on the Education domain.

NLP is the main focus in improving the education system with the help of innovating new ways of learning and helping both students and teachers. There are several companies which are investing in improving the education systems across the globe and today this article focuses on one such leading Startup named **Riid**¹. This startup is focusing on building NLP/AI-driven solutions to improve education systems. Riid is currently focusing on 3 key areas of research: *AI Education*, *Natural Language Processing* and *Computer Vision*. This technology review focuses on giving an overview of NLP/AI-driven technologies developed by Riid to provide an AI Education system (AIEd), which are broadly categorized into following three research areas: **Passage Scoring**, **Question Generation**, and **Text Summarization**.

3 Overview of AI/NLP techniques and tools used to build AI Education (AIEd)

- **Use of Content-based Collaborative Filtering (CCF) techniques for Topic Recommendations:** The need for an AI-based education system is to make recommendations on subjects/topics to students, so it requires a good recommender technique. Collaborative filtering (CF) aka. User Similarity filtering is a popular technique to find the relationship between User and the Items User clicked or referenced. However, this disregards the item-user-specific information (insights on student interests), which sometimes leads to a *Cold Start* problem where the recommender system fails to make predictions for new students/items/subjects. To overcome this limitation, Riid NLP approach is to use Content-Based Collaborative Filtering (CCF), which improves the recommendation performance and also addresses the cold start problem. The two examples where this technique is useful are: *Topic Recommendation*, and *Knowledge Tracing* (predicting whether the student will answer questions correctly based on the previously answered responses).
- **GRAM Language Model:** As per the research paper submitted by Riid team on GRAM², if we directly use Pre-trained Language Models (LM) for CCF, the system requires large computational resources (high GPU) due to CCF's multi-model nature, where both user similarity (CF = Collaborative filter) and content similarity (CE = Content encoder) are considered. CF predicts the user's response to a given item based on the user's past interactions. CE gives the item representation. To overcome this problem, Riid proposed **Single-Step GRAM** (**GR**radient **A**ccumulation for **M**ulti-Modality in CCF) and **Multi-Step GRAM**. The main

¹<https://riid.com/>

²<https://arxiv.org/abs/2204.04179>

idea of GRAM(s) is to perform efficient training on content-based collaborative filtering models by splitting the CE and CF modules during training and also accumulating the items appearing repeatedly to improve the training performance.

- **Santa Project:** Santa ³ is Riiid's English proficiency Intelligent Tutoring System (ITS) in multi-language. It utilizes AI and NLP techniques to help students prepare for TOEIC (Test of English for International Communications). It uses Collaborative filtering (CF) and Assessment modelling approaches to predict the score of students (Academic Performance Prediction) based on their responses and also provides insights on Students Engagement.
 - **Predicting Academic Performance using NLP techniques:** The idea here is to predict Student's academic grades, recommend courses, provide insights on who are low-grade students so that teachers focus more on them, etc. All these results require a good prediction model. Riiid team recommended an **Assessment Modeling (AM)**, a pre-training method for an Intelligent tutoring system. AM model approach is based on *pre-training* and *fine-tuning* paradigms. In the pre-training phase, the model is trained to predict whether a student provides a correct answer to a given question (response correctness) and whether the students answer each question within the given time window (timeliness). In the fine-tuning phase, the model is fine-tuned with unseen scenarios to improve the performance of the AM model. Using AM modelling approach Santa has shown great Score prediction performance with MAE (Mean Absolute Error) as low as 49.84. ⁴. Another approach for score prediction used by Riiid is the Collaborative Filtering based approach which gives a MAE of 78.9. **AI tutor** is one such application built by Riiid to provide recommendation and predicts learner's scores.
 - **Student Engagement:** As per the paper submitted by Riiid (4) there is a correlation between score prediction and student engagement. According to the study, a more precise score leads to a higher level of student engagement when using Santa services. It is based on the fact that if we provide more accurate predictions to students in comparison to their actual real-world scores, there will be more engagement of students to use the Santa App. So, the research was focused on improving student engagement by building an engagement model with the help of Collaborative/Content-based filtering techniques.
- **EdNet Data:** For any company to build an AI and NLP driven Education system, the main challenge is to collect the real students' interaction data, which is needed to train AI models to provide better results. Riiid is helping the world to tackle this problem by providing the World's largest Hierarchical Dataset in Education, **EdNet** ⁵. EDNet contains more than 131 million student interactions and activities in 2 years while using Santa (a multi-platform) application. Making its data public, Riiid is helping to help companies to do more innovations in AI Education.

³<https://www.aitutorsanta.com/>

⁴<https://arxiv.org/pdf/2005.05021.pdf>

⁵<https://riiid.com/blog/8>

EdNet data has following main characteristics: ⁶:

- **Large Scale:** Ednet has 131,441,538 student interactions starting the year 2017 with an average student interaction of 441.20. It also provides more than 13K problems and 1K lectures tagged with 293 types of skills.
- **Diversity:** EdNet dataset also includes student interactions on learning activities such as reading and watching lectures. Such information is usually not provided by any other market leader in providing AI Education datasets.
- **Hierarchy:** Ednet data is structured into four different hierarchical levels:
 - * **EdNet-KT1** - Stores most basic information, which consists of student's question-solving logs. This is useful to build deep knowledge and self-attentive knowledge tracing algorithms.
The next 3 levels of the dataset also provide the details of student's *action sequences*.
 - * **EdNet-KT2** - The simplest action-based dataset. Also contains information such as student's study mode.
 - * **EdNet-KT3** - It also incorporates information about different learning activities in which students participate aside from solving questions, such as watching lectures, reading through comments, etc.
 - * **EdNet-KT4** - This contains all student actions information, which are: "erase choice", "undo erase choice", "play and pause audio", "play and pause video", etc.
- **Multi-Platform:** Ednet dataset contains information gathered from both mobile (IOS, Android) and desktop users, which helps to build future multi-platform models.
- **Other AI Features and capabilities offered:** Riiid's R.inside is providing AI Education capabilities ⁷ on:
 - "Recommender System": Recommend learners different topics based on the state of the art of machine learning and natural language processing techniques such as Collaborative filtering and content filtering.
 - "Engagement Prediction": Optimize the learning experience to understand if there are risks of learners leaving the session.
 - "Knowledge Tracing": Access individual student's knowledge level.
 - "Option Choice Prediction": Riiid NLP models also predict student's performance on multiple choice questions such as how many times the user clicked on other choices before submitting with the final choice.
 - "Score Prediction": Predicts learners grades during the learning process based on student's interactions on different topics.
 - "Question Generation": Automatically generate questions based on reading passages. This feature helps students in grasping every topic in detail.
 - "Riiid for Classrooms" ⁸: AI-Powered teaching tool to enhance learner's ability by reducing inefficiency in creating and evaluating assessment content.

⁶<https://github.com/riiid/ednet>

⁷<https://riiid.com/solution?target=s2>

⁸<https://www.riiidlabs.com/home>

4 Conclusion

With the use of more digital tools and techniques in the education system, Natural language processing is becoming a key focus area. The main use of NLP/AI techniques in the education system is to help students in reading, writing and improving their skills in different topics. **Riiid** is one such leading startup in AI Education system, whose main focus is to build AI-driven education system to address different learner needs using their **Exponential Learning Experience**. In today's world, we are now starting to realize that every learner has his or her own pace in understanding any topic and every student requires a different level of attention. In a class full of students, this is not easy for a teacher to use different teaching methods to teach different students the same topics. So, with the help of AI and NLP-driven models, Riiid is building such infrastructure. Some key takeaways from this discussion are:

- NLP can help identify problems in student's reading and writing capabilities and can quickly provide feedback.
- NLP techniques helps to understand learners sentiments, actions during lectures to identify learners engagement.
- NLP for teachers can help find out the students that are struggling to understand the topic.
- Pre-trained Language Models are good with content and collaborative filtering (CCF) techniques but require more processing time in model training.
- GRAM is a more robust and efficient learning model using CCF techniques.

AI Tutor and AI Teacher's Assistant are some of the AI Learning solutions ⁹ built by Riiid to help the education system.

5 References

- **Riiid's Publications:**

- Score prediction and Student Engagement: <https://arxiv.org/abs/2005.05021>
- GRAM: <https://arxiv.org/abs/2204.04179>
- Assessment Modeling: <https://arxiv.org/abs/2002.05505>

- **Other Pages:**

- <https://github.com/riiid/ednet>
- <https://www.riiidlabs.com/ai-solutions>
- <https://riiid.com/blog/8>
- <https://riiid.com/research>
- <https://riiid.com/solution?target=s2>

⁹<https://www.riiidlabs.com/ai-solutions>