



Towards a Comprehensive AI Teaching Assistant Based on Course Forums

Bowen Gu¹, Hao Wang¹, Kaizhuo Chen¹, Prasun Dewan¹

¹University of North Carolina at Chapel Hill



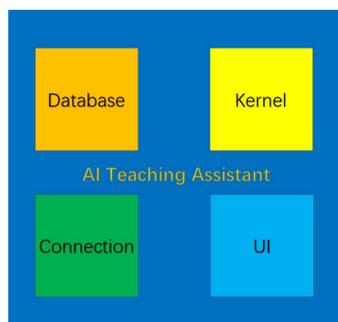
Why Building AI Teaching Assistant for course forums?

Managing and monitoring the forums is tedious and repetitive.

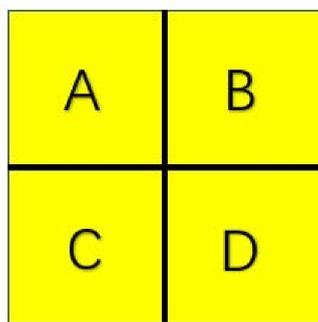
Jill Watson, an AITA for course forums developed by Dr. Ashok Goel in 2016 is a fully-automated program that reply to students' posts with high sophistication, but it was only capable for replying posts. A human TA's job also involves post processing, which can also be automated.

This shows the importance of the development of a multi-functional post processing AITA which serves as a complement of Jill Watson.

AITA Structure



AITA Kernel Structure



The fully developed AITA will contain the following major components:

- (a) A database that helps store and manipulate posts in different pre-designed categories.
- (b) A kernel that consists of different modules, in which different machine learning models are encapsulated and used to perform different tasks.
- (c) A connection module that fetches new posts and sends messages generated by the kernel to target students or instructors.
- (d) A user interface that makes the interaction between the instructors and the system easier.

Current work focuses on the development of different modules in the kernel. Now the following four modules have been implemented.

- A: Automatic labeling system
- B: Duplicate posts identifier
- C: Incomplete office hours requests identifier
- D: Automatic assignment assistant

Kernel Module Functionality and Implementation Illustration

A. Automatic Labeling System

The automatic labeling system is used to label the new posts into the pre-designed categories so that they can be processed by the following modules and be stored in the database.

B. Duplicate Posts Identifier

The duplicate posts identifier is used to identify duplicated forum posts and point students to the similar posts so that the student who asks the question does not need wait for the answers and the instructor does not need to answer the question again.

C. Incomplete Office Hours Requests Identifier

The incomplete office hours requests identifier is used to identify office hours requests that do not specify the meeting time or reason with the instructor so that the instructor do not need to spend the whole office hours session waiting for the student and can prepare for the student's questions in advance.

D. Automatic Assignment Assistant

The automatic assignment assistant is used to help students will code implementation assignments. It can provide a line by line English translation of the solution code to the student so that students can make progress on the assignment without violating the course policy.

Implementation Pipeline

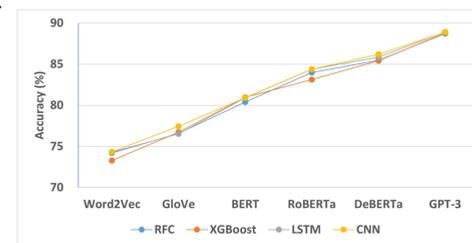
1. Collect data needed for each module
2. Clean data using different strategies
3. Extract features from the data, which is used as the learning material
4. Set up different prediction models
5. Train the prediction models using the features extracted and test using the test data



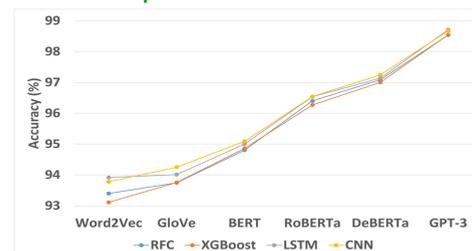
Module Performance

A. Automatic Labeling System

The figures on the right shows the accuracy of each language model - prediction model combination. The language models are the six shown in the x-axis, which are used to convert text in the forum posts into numerical vectors that can be learned by machines. The prediction models are the four shown in the legend, which are used to produce prediction results. The automatic assignment assistant module does not produce satisfying results and is not included in the figures.



B. Duplicate Posts Identifier



C. Incomplete Office Hours Requests Identifier



Summary and Future Work

1. This research presents different modules of a practical system that can be used to greatly alleviate instructors' burden on online education forum posts processing.
2. Most modules achieved a highest accuracy of over 90%
3. Future work should focus on further increasing the accuracy, completing the other components of the AITA, and adding more functionalities.

References (Partial)

1. Do, J., Dynamic Tagging of Educational Forum Posts Using NLP Post Embeddings, 2021.
2. Li, Y., Yang, T., Srinivasan, S., Word Embedding for Understanding Natural Language: A Survey, 2018.