Using Data Analytics for Personalization of Online Tutoring Systems

Introduction:

Applying data mining (DM) in education is an emerging interdisciplinary research field also known as educational data mining (EDM). Data Analytics is steadily being introduced into the online tutoring experience. However, personalization of the online tutoring experience has not been adopted across tutoring systems. The goal of this proposal is to integrate data analytics features such as clustering, classification and association mining in intelligent online tutoring systems. Our project aims at adding personalization features to an already existing **open source online tutoring system** such as Carnegie Mellon University's Cognitive Tutor (http://ctat.pact.cs.cmu.edu/). The three features that we wish to incorporate in the tutoring system are:

- Recommendation systems for personalized tutoring (using **Deep Learning**)
- Profiling of students (using **Expectation Maximization (EM) Clustering Algorithm**)
- Online chatbots as mentors (using **MOOCBuddy**)

Specific Aim 1: Personalized learning using Deep Learning: The goal of this aim is to personalize each student's learning experience based on performance data. We use a Deep Learning approach to map students and tutoring questions to a latent space where the similarity between students and their questions is maximized. We will then extend the model to jointly learn from behavioral features of students by introducing a ***multi-view Deep Learning model***. This rich student feature representation allows the model to learn relevant student behavior patterns and give useful recommended questions to each student in a personalized way. The combination of user features and questions in a single model for learning helps improve the recommendation quality for students. We will experiment our approach on two existing systems: Cognitive tutor by CMU [7] , and MetaTutor [6]

Methodology: We would begin by recording every tutoring session between the online tutor and the student, amounting to robust data collection. Such data would track the steps a student chose to solve a problem. The data collected would then be used to build a deep learning behavioral model. This would be particularly applicable to understanding various student learning styles and tailoring tutoring in an increasingly personalized way.

For example, consider a word problem involving 3 apples, 4 oranges and 5 carrots and a student is asked to calculate the total amount of fruits. The tutor would include "conceptually unrelated numbers (or distractors)." In this example the 5 carrots would be the distractor. The tutor would track the student's ability to solve similar problems which include distractors to achieve overall skill mastery in this area. If a student fails to eliminate the distractor, similar questions will be recommended in a personalized way.

Preliminary work

Our prior work in educational data mining has looked into mining student data by ensemble classification for prediction of student academic performance [1]. Such prediction is being used by educators for providing additional support to needy students based on the predictions.

Specific Aim 2: Profiling Students by Clustering:

The idea of clustering students according to their behavior in the context of learning systems has the potential it offers for the tutoring system or the human teacher (in a virtual classroom type of environment) to provide more adaptive scaffolding. Specifically, we can model self-regulated learning (SRL) skills in order to support their use during learning. We are interested in answering the following questions: (1) can we establish the existence of clusters of students according to their performance and interaction with a tutor? And if that's the case, (2) what are the characteristics that distinguish students belonging to those different clusters, and in particular, how do they relate to their use of self-regulated learning processes?

Methodology: In order to collect training data, students will be randomly assigned to either a prompt and feedback (PF) condition or a control (C) condition and asked to learn about a particular content. In the PF condition, participants will be prompted by the tutoring system to use specific planning, metacognitive monitoring, and learning strategies and will be given immediate feedback about the quality and accuracy of these processes. For example, after completing a quiz, participants in the PF condition will be given information about their performance on the quiz and, depending on their knowledge acquisition, will be prompted to either continue reviewing the content or progress to another sub-goal. The students assigned to the control section (C) will not receive any prompts or feedback, which would then be the test data.

We would like to use the Expectation-Maximization (EM) algorithm, as implemented in Weka 3.6.5, over the sub-sample of students in the PF condition, since they were the ones who interacted with a version of the system in which the tutoring agents provided them with the most adaptive and complex scaffolding of their SRL processes. To compensate for the sensitivity of EM to the choice of seed (i.e. the cluster initiator) for the algorithm, linked to its tendency to get stuck into local optima, we would run 200 different seeds to initialize it.

The next step would be to profile students based on the clusters produced by the EM algorithm. The cluster profiles provide us with an understanding of three different 'types' of learners, based on learner-driven variables. These clusters will provide us with insight on how these variables varied between groups. For example, Cluster 'n' learners also spent relatively less time than others reading and taking notes, while receiving the most prompts, etc. One possible future direction is to use the clusters that have been defined and characterized as input for a classifier to be used on-line (as opposed to the a posteriori only analysis done here), i.e. to be able to predict at any moment during the students' learning session with a tutor, the probability that they will be sorted into a specific cluster.

Preliminary work

Our prior work in data analytics in the field of education has looked into mining student data by ensemble clustering using k-means for profiling of students [2]. We wish to extend our knowledge by mining student-tutor data to better understand each student's creativity, logical thinking and confidence.

Specific Aim 3: Incorporating Chatbots as Mentors: Exploring the directories for chatbots and studying the current articles on this technology, one can note that education is seldom discussed as a domain for which chatbots could be built. Our aim is to incorporate a simple chatbot into an online tutoring system, which can provide quick answers to student questions that tend to recur each semester regarding a particular topic. In addition to answering questions, the chatbot can occasionally suggest links and further reading to students that pertain to their coursework. The chatbot can also suggest direct ways that a student can learn a topic by introducing computational thinking using hands-on projects. We would like to start by using MOOCBuddy or Api.ai, chatbots for personalized learning [5], as a starting point.

Preliminary work: Our prior work has shown that having mentors (both online and offline) helps in student retention of first year students [3]. Also, we have shown that incorporating computational thinking in students by hands-on projects has improved student performance [4].

References:

[1] **A. Satyanarayana** and M. Nuckowski (2016). Data mining using ensemble classifiers for improved prediction of student academic performance.

[2] **A. Satyanarayana** and G. Ravichandran. (2016). Mining Student data by Ensemble Classification and Clustering for Profiling and Prediction of Student Academic Performance.

[3] **A. Satyanarayana**, H. Li, and J. Braneky. (2014) "Improving retention by mentoring and tutoring freshmen students" *IEEE ASEE 2014 Zone 1 Conference*, University of Bridgeport, CT, April 3rd-5th, 2014.

[4] N.Benakli, B.Kostadinov, **A. Satyanarayana** and S.Singh, (2017) "Introducing computational thinking through hands-on projects using R with applications to calculus, probability and data analysis", *International Journal of Mathematical Education in Science and Technology (Taylor & Francis),*(2017), Vol 48(3), pp 393-427.

[5] Iftene, A., & Vanderdonckt, J. (2016). Moocbuddy: a chatbot for personalized learning with moocs. In *RoCHI–International Conference on Human-Computer Interaction* (p. 91).

[6] Azevedo, R., Witherspoon, A. M., Chauncey, A., Burkett, C., & Fike, A. (2009, October). MetaTutor: A MetaCognitive Tool for Enhancing Self-Regulated Learning. In *AAAI Fall Symposium: Cognitive and Metacognitive Educational Systems*.

[7] Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2006, June). The cognitive tutor authoring tools (CTAT): Preliminary evaluation of efficiency gains. In *International Conference on Intelligent Tutoring Systems* (pp. 61-70). Springer, Berlin, Heidelberg.