**What Will You Code Next? - Deep knowledge tracing with recurrent neural nets on open-ended student responses**

Modeling student knowledge while students are acquiring new concepts is a crucial stepping stone towards providing personalized automated feedback at scale. This task is also referred to as “knowledge tracing” and has been explored extensively on exercises where student responses can only be correct or incorrect. However, knowledge tracing on open-ended problems where answers extend beyond binary solutions is still mostly unexplored. We believe a rich set of information about a student’s learning hides within their responses to open-ended problems. This is a challenging task, but with recent advances in machine learning, there are more promising techniques to take this on. In our work, we train [Recurrent Neural Nets](http://karpathy.github.io/2015/05/21/rnn-effectiveness/) (LSTMs) to predict a student’s performance and their next move while solving a coding exercise on code.org, an online learning platform for computer programming. Our work shows promising results and brings us a step closer to building automated feedback systems.

Automated feedback is a major challenge for open-ended questions since correct and incorrect solutions can take a variety of forms. This makes personalized feedback that is specific to each student’s answer even more critical, in order to allow the students to understand their performance and steps for improvement. With the inception of massive open online courses (MOOCs), educators from around the world can reach millions of students by disseminating course videos and content through online classrooms. However, in the current model of online courses, scaling feedback for these open-ended questions remains cost-inhibitive and difficult.

Robust knowledge tracing is a crucial step towards personalized feedback. [Piech et al.](https://web.stanford.edu/%7Ecpiech/bio/papers/deepKnowledgeTracing.pdf) applied deep learning to predict student performance on multiple-choice math exercises on [Khan Academy](https://www.khanacademy.org/), and found that RNNs are particularly suitable for this task. However, exercises with open-ended answers like coding problems are much harder to model. In our research, we extend Piech et al.’s work on “deep knowledge tracing” by modeling students’ learning trajectories as they solve [Code.org](https://code.org/)'s open-ended coding problems. More concretely, our deep learning model trains on a student’s history of code submissions and predicts firstly the student’s performance on the next problem and secondly the next line of code that the student will write. We train our model on code submissions because they contain rich information about a student’s knowledge of both code logic and style. Since code submissions are challenging to represent directly in a feature space, we also trained an RNN to generate embeddings for code submissions. These embeddings are then fed into a second RNN, which performs the final predictions over a series of exercises represented by the embeddings.

Source code: [Github](https://github.com/ange3/deepcode)