Deep Knowledge Tracing and Engagement in MOOCs

Klint Kanopka, David Lang, and Kritphong Mongkhonvanit

Stanford University

Abstract

MOOCs and online courses have notoriously high attrition. One challenge is that it can be difficult to tell if students fail to complete because of disinterest or because of course difficulty. Starting from the Deep Knowledge Tracing framework, we account for student engagement by including course interaction covariates. With these, we find that we can predict a student's next item response with over 88% accuracy. Using these predictions, targeted interventions can be offered to students and targeted improvements can be made to courses.

Introduction

Deep Knowledge Tracing is a way to estimate student knowledge through a course and has been explored with MOOCS in several ways:

- Predict and identify latent skills pertaining to items and tasks [1]
- Clickstream interactions e.g.(pause,plays, and video interactions) are a feature that improves grade prediction [2]
- Student Sentiment in MOOC forums can predict student dropout [3]

We are combining past item response information with video engagement interactions to predict the probability a student will answer the next item correctly.

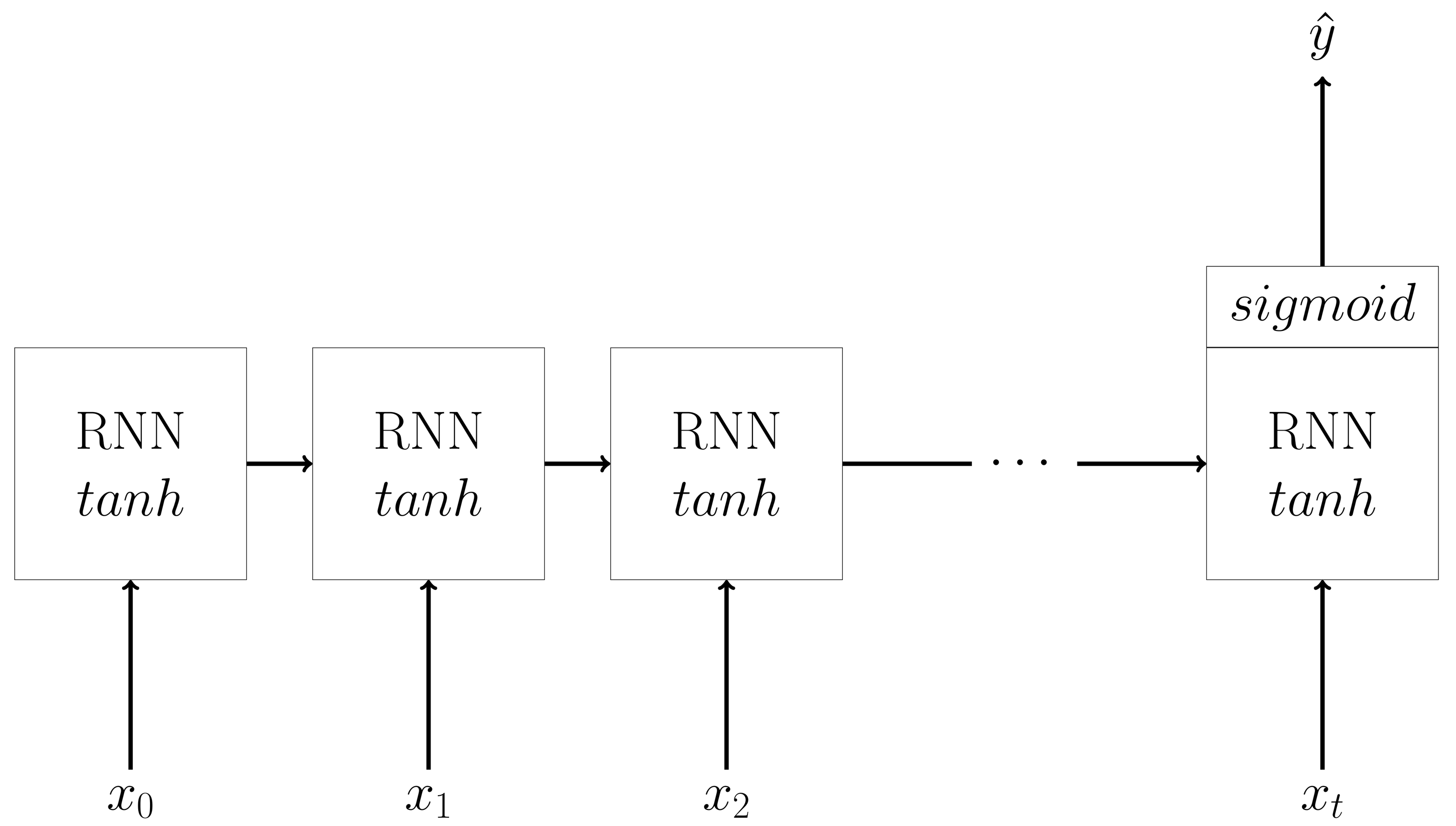
Data

anon_name	feature	index1	index2	⋯	index103
38fqh9dy	items attempted	103	⋯		
38fqh9dy	correct	0	1	⋯	1
38fqh9dy	playback_speed	1	1.25	⋯	0
38fqh9dy	pauses	1	0	⋯	0
38fqh9dy	seek_back	0	0	⋯	0
38fqh9dy	seek_forward	0	0	⋯	0
38fqh9dy	video_completed	0	0	⋯	0
38fqh9dy	attempt	1	1	⋯	0
38fqh9dy	quiz	0	0	⋯	0

Figure 1: Example entry in the data set for one student

- Data comes from 12,007 students in a MOOC course on statistics.
- Item responses and clickstream interactions for each student.
- Clickstream interaction data contains seeks, pauses, average playback speed, and video completion. Videos are tied to items.

Model



- Each RNN block contains 128 hidden units.
- L2 regularization with $\lambda = 0.01$.
- Binary cross-entropy loss.
- Adam optimizer.
- Each x_t contains the item response and clickstream data for the video associated with item t.
- $\hat{y}$ is the predicted item response (correct or incorrect) for item t.

Results

Inputs \ Type	GRU	LSTM	Simple RNN
Response only	0.8726	0.8826	0.8668
Response and clickstream	0.8818	0.8830	0.8834

- All models were trained for 150 epochs with an 80/10/10 split.
- All model types show at least marginal improvement with the inclusion of video interaction covariates.
- The simple RNN benefits the most from clickstream data and is also the fastest to train.
- LSTMs take significantly longer to train than other models.

Discussion

- Adding clickstream data as a covariate increases accuracy.
- Our results likely understate the accuracy gains of clickstream data as almost three-quarters of users did not interact with course videos.
- The method used is independent of the content of the course and requires no additional qualitative coding of items.
- At the student level, predicted item probability quickly identifies students who need support and informs appropriate interventions.
- At the course level, the average predicted probability of all students can be used to identify areas that can be targeted for improvement.

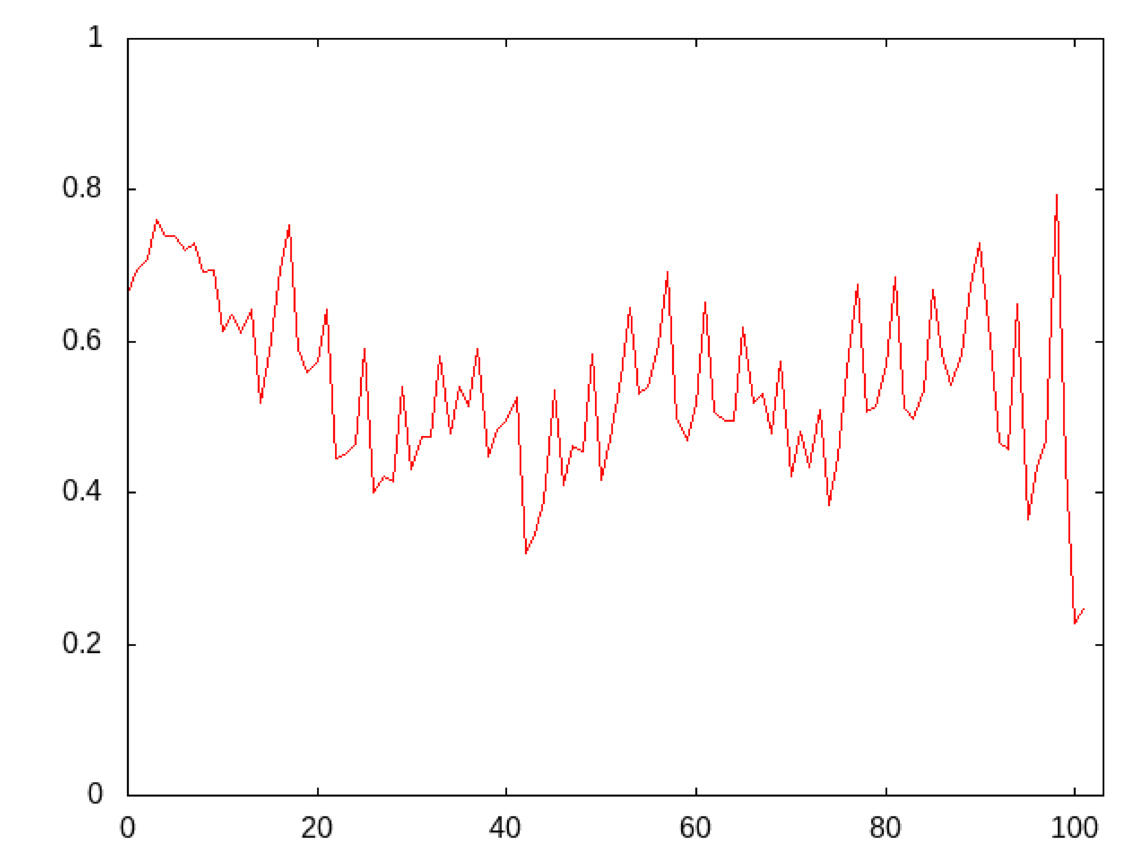


Figure 2: Example prediction graph for a student

Future Work

- Explore additional MOOCs with non-dichotomously coded items.
- Model similarity to "ideal engagement" to identify interventions.

References

[1] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein.
Deep knowledge tracing.
In *Advances in Neural Information Processing Systems*, pages 505–513, 2015.

[2] Tsung-Yen Yang, Christopher G Brinton, Carlee Joe-Wong, and Mung Chiang.
Behavior-based grade prediction for moocs via time series neural networks.
IEEE Journal of Selected Topics in Signal Processing, 11(5):716–728, 2017.

[3] Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim.
Predicting student attrition in moocs using sentiment analysis and neural networks.
In *AIED Workshops*, 2015.