Learning Representations of Student Knowledge and Educational Content

Siddharth Reddy, Igor Labutov, Thorsten Joachims

Department of Computer Science, Cornell University, Ithaca, NY, USA

Summary of Contributions

- Demonstrate the ability of an embedding model to successfully predict assessment results
- Introduce an offline methodology as a proxy for assessing the ability of a model to recommend personalized lesson sequences

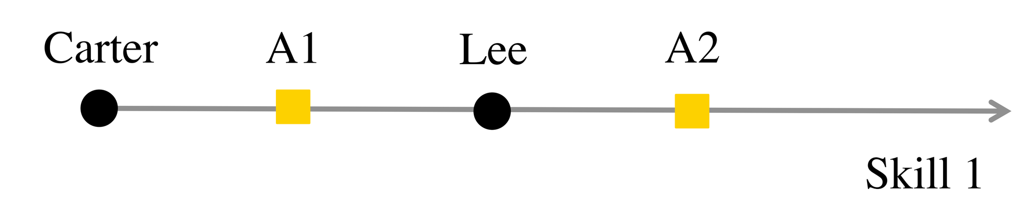
Model Representation

- **Student** = a set of d latent skill levels $\vec{s} \in \mathbb{R}_+^d$ that vary over time
- **Lesson module** = a vector of skill gains $\vec{\ell} \in \mathbb{R}_+^d$ and a set of prerequisite skill requirements $\vec{q} \in \mathbb{R}_+^d$
- **Assessment module** = a set of skill requirements $\vec{a} \in \mathbb{R}_+^d$
- A student can be tested on an assessment module, which has a **pass-fail result** $R \in \{0, 1\}$. The likelihood of passing should be high when a student has skill levels that exceed the assessment requirements, and vice-versa.
- A student can complete lesson modules to learn over time, though the **skill gains** $\vec{\ell}$ from a lesson module are modulated by **prerequisite knowledge** $\vec{q}$

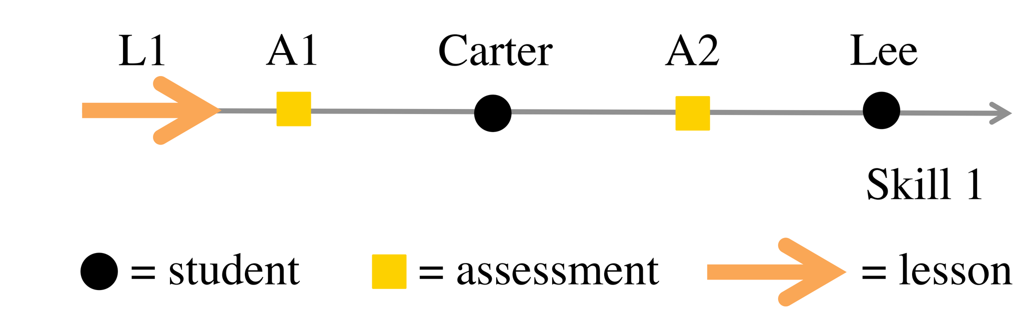
Examples

A simple example in one dimension to illustrate the model in action:

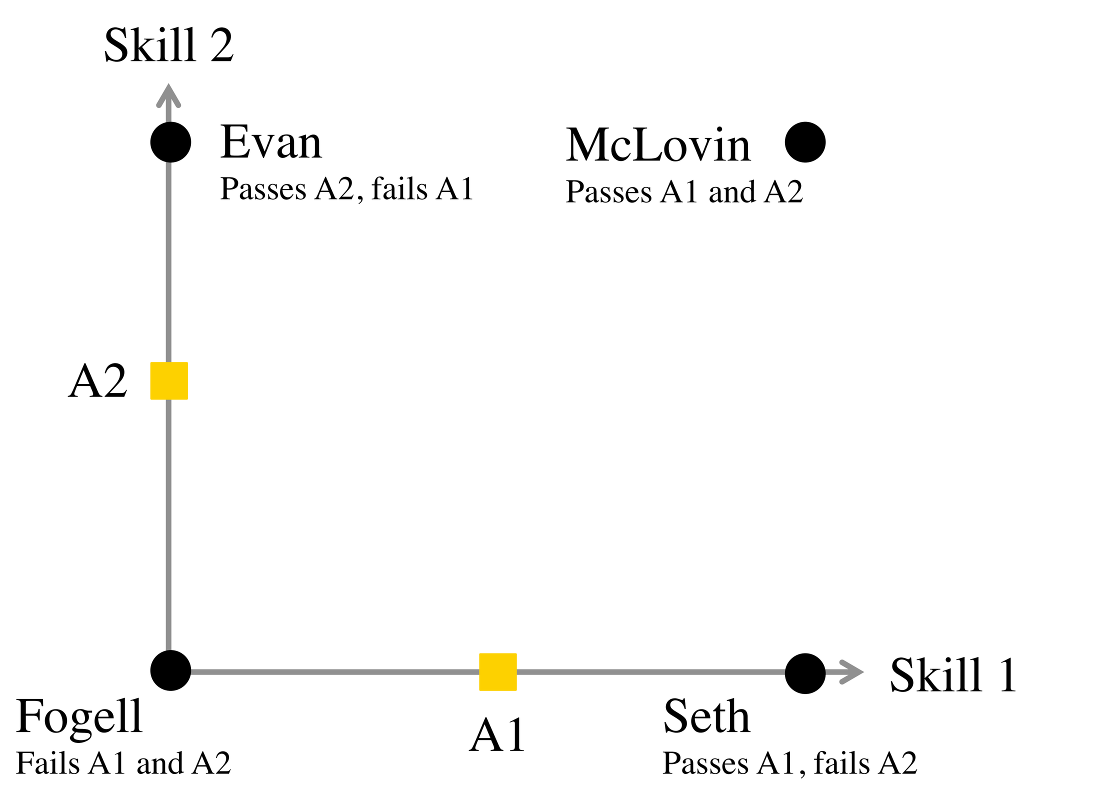
At t = 1,
- Lee passes A1, fails A2
- Carter fails A1 and A2



At t = 2,
- Lee completes lesson L1, then passes A1 and A2
- Carter completes lesson L1, then passes A1, fails A2

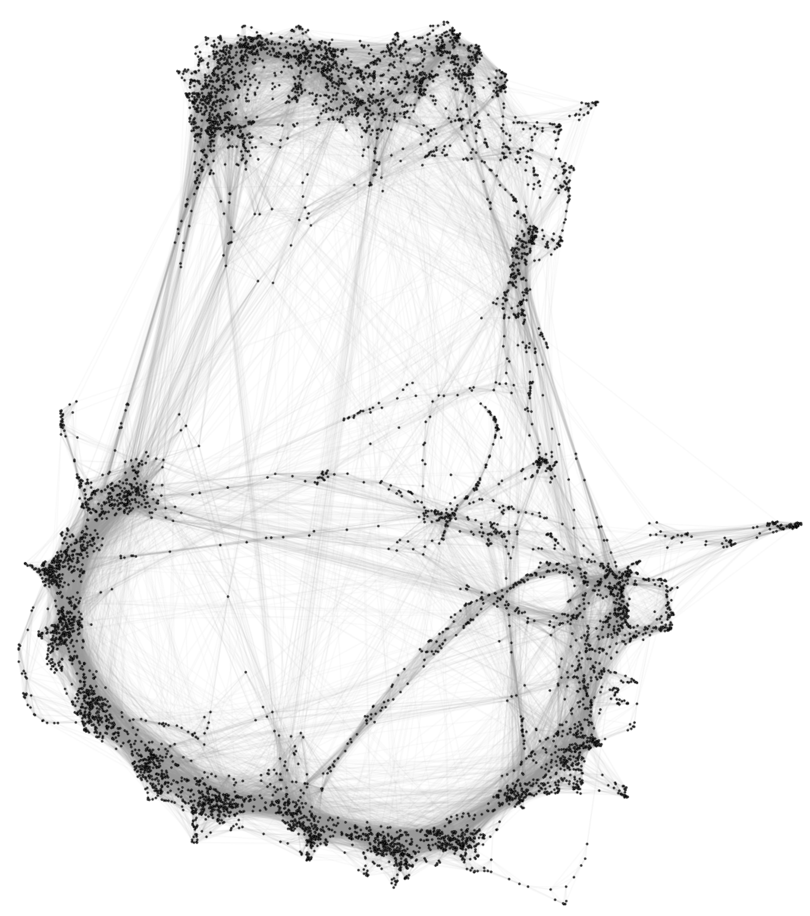


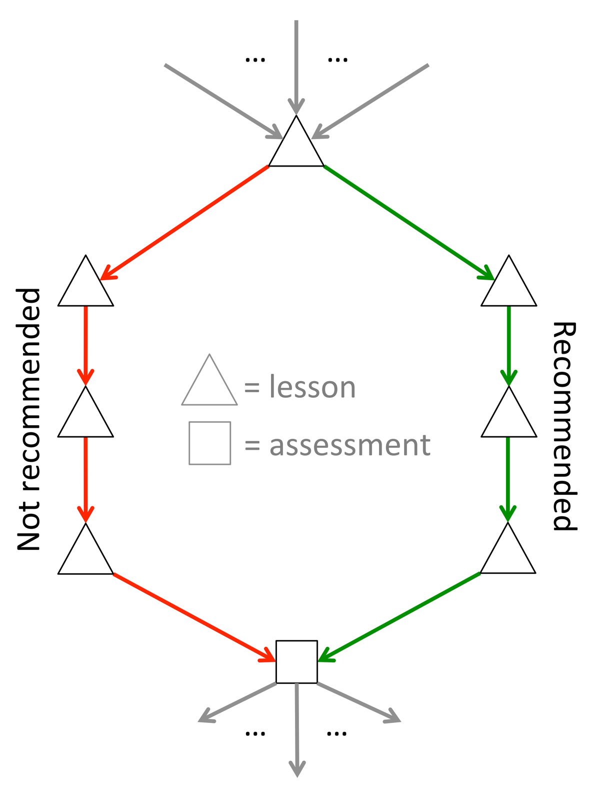
Generalizing the *latent skill space* to higher dimensions gives us the flexibility to describe more complicated scenarios, such as the one on the right.



Data

- The two data sets from Knewton contain 2,184,352 interaction logs from 1,939 classrooms over six months, for 7,034 students, 7,217 lessons, 7,287 assessments, and average assessment pass rates of 0.712 and 0.693
- Students take different paths through the set of lessons and assessments (the graph of all student paths is shown below). We observe many instances of student paths that share the same lesson module at the beginning and the same assessment module at the end, but contain different lessons along the way. We call these instances *bubbles* (a schematic is shown below), and use them to evaluate the a model's ability to recommend lesson sequences.





Model Dynamics

- **Assessment Results**
  For student $\vec{s}_t$, assessment $\vec{a}$, and result R,
  $$R \sim \text{Bernoulli}(\phi(\Delta(\vec{s}_t, \vec{a})))$$
  where ϕ is the logistic function and $\Delta(\vec{s}_t, \vec{a}) = \frac{\vec{s}_t \cdot \vec{a}}{||\vec{a}||} - ||\vec{a}|| + \gamma_s + \gamma_a$
- **Student Learning from Lessons**
  For student $\vec{s}$ who worked on a lesson with skill gains $\vec{\ell}$ and no prerequisites at time $t+1$, the updated student state is
  $$\vec{s}_{t+1} \sim \mathcal{N}\left(\vec{s}_t + \vec{\ell}, \boldsymbol{\Sigma}\right)$$
  where the covariance matrix $\boldsymbol{\Sigma} = I_d \sigma^2$ is diagonal. For a lesson with prerequisites $\vec{q}$,
  $$\vec{s}_{t+1} \sim \mathcal{N}\left(\vec{s}_t + \vec{\ell} \cdot \phi(\Delta(\vec{s}_t, \vec{q})), \boldsymbol{\Sigma}\right)$$
  where $\Delta(\vec{s}_t, \vec{q}) = \frac{\vec{s}_t \cdot \vec{q}}{||\vec{q}||} - ||\vec{q}||$

Parameter Estimation

- We compute MAP estimates of model parameters $\boldsymbol{\Theta}$ by maximizing the following objective function:

$$L(\Theta) = \sum_{\mathcal{A}} \log\left(Pr(R \mid \vec{s}_t, \vec{a}, \gamma_s, \gamma_a)\right) + \sum_{\mathcal{L}} \log\left(Pr(\vec{s}_{t+1} \mid \vec{s}_t, \vec{\ell}, \vec{q})\right) - \beta \cdot \lambda(\Theta) \qquad (1)$$

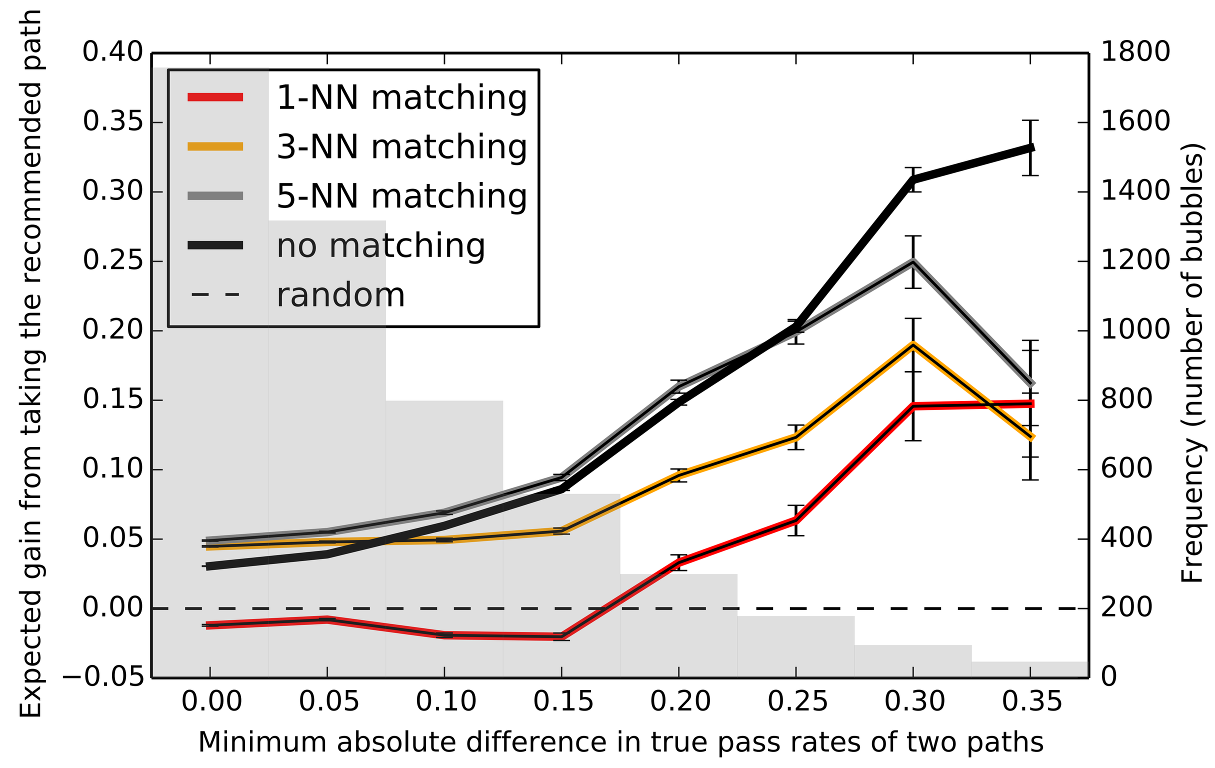
  where $\mathcal{A}$ is the set of assessment interactions, $\mathcal{L}$ is the set of lesson interactions, $\lambda(\Theta)$ is a regularization term that penalizes the L_2 norms of embedding parameters (not bias terms γ), and β is a regularization parameter. Non-negativity constraints on embedding parameters (not bias terms γ) are enforced.
- We solve the optimization problem with box constraints using L-BFGS-B and random parameter initializations

Model Evaluations

- For the **assessment result prediction** task, our performance measure is Area under ROC Curve (AUC)

							Book A				Book B			
							Random		Last		Random		Last	
	d	s	a	ℓ	q	γ	Train.	Val.	Train.	Val.	Train.	Val.	Train.	Val.
1	1	Y	N	N	N	N	0.659	0.605	0.659	0.626	0.723	0.696	0.723	0.686
2	1	N	Y	N	N	N	0.743	0.734	0.743	0.737	0.716	0.712	0.715	0.661
3	2	Y	Y	N	N	N	0.994	0.687	0.994	0.687	0.992	0.659	0.991	0.614
4	2	Y	Y	N	N	Y	0.995	0.673	0.994	0.705	0.993	0.745	0.992	0.726
5	2	Y	Y	Y	N	N	0.900	0.713	0.897	0.730	0.896	0.733	0.897	0.696
6	2	Y	Y	Y	N	Y	0.873	0.741	0.872	0.756	0.869	0.787	0.871	0.748
7	2	Y	Y	Y	Y	N	0.898	0.709	0.898	0.727	0.902	0.742	0.898	0.694
8	2	Y	Y	Y	Y	Y	0.890	0.732	0.889	0.753	0.872	0.787	0.882	0.749

- For the **lesson sequence discrimination** task, our performance measure is the expected gain from taking the recommended path, i.e., $\mathbb{E}\left[\frac{\mathbb{E}[R'] - \mathbb{E}[R]}{\mathbb{E}[R]}\right]$ where R' is the outcome at the end of the recommended path and R is the outcome at the end of the other path



- **Conclusions**: the embedding model performs well on both prediction tasks, largely due to the benefits of modeling skill gains from lessons and including bias terms to capture general pass rates for students and assessments; modeling lesson prerequisites has an insignificant effect on prediction accuracy.

http://siddharth.io

sgr45@cornell.edu