Dynamic Key-Value Memory Networks for Knowledge Tracing

Jiani Zhang[1,2] Xingjian Shi[3] Irwin King[1,2] Dit-Yan Yeung[3]

[1]The Chinese University of Hong Kong, Hong Kong, China

[2]Shenzhen Research Institute, The Chinese University of Hong Kong, China

[3]Hong Kong University of Science and Technology, Hong Kong, China

April 21st, VALSE 2017

Motivation

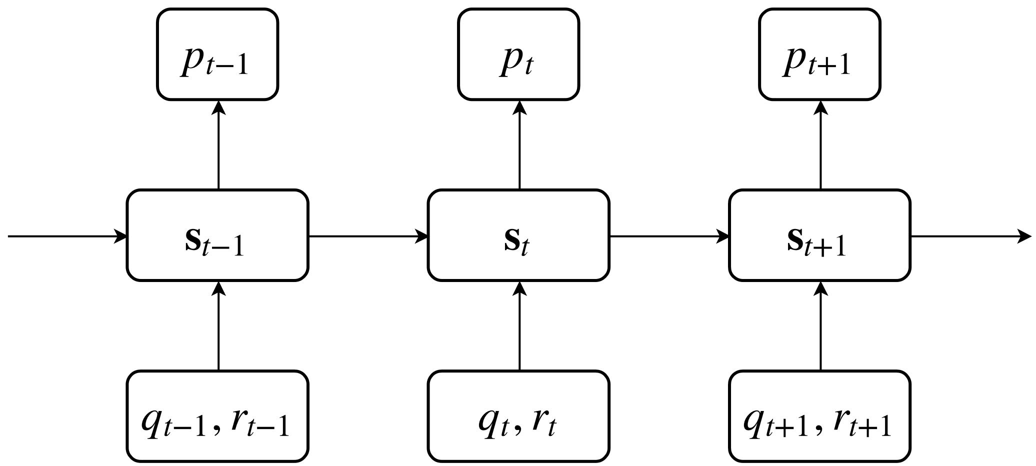
Massive Open Online Course

- Goal: Personalize the learning of students (个性化学习)
- Elementary Task: Knowledge Tracing (知识追踪)

Importance and Challenge of Knowledge Tracing

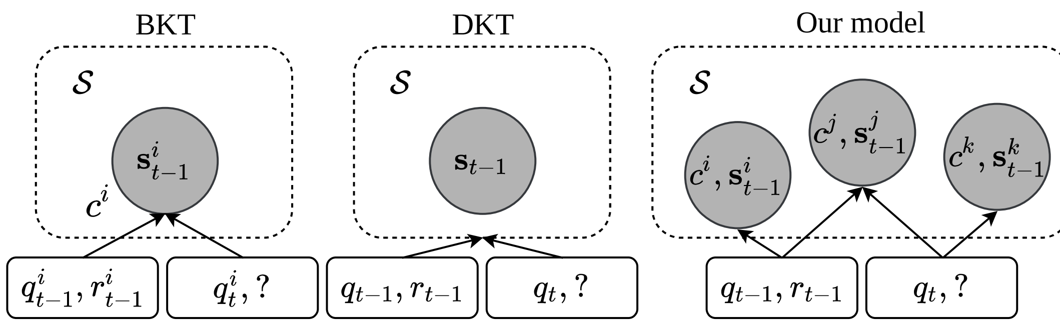
- Trace the student's knowledge level as the learning process proceeds.
- An essential task in online learning platforms.
  - (Tutors) Give proper hints and tailor the sequence of practice exercises
  - (Students) Devote more energy to less familiar concepts
- Challenges:
  - How to model the knowledge state?
    - Latent variable
    - Changes as the student learns
  - Hard to evaluate

Knowledge Tracing as a Sequence Learning Problem



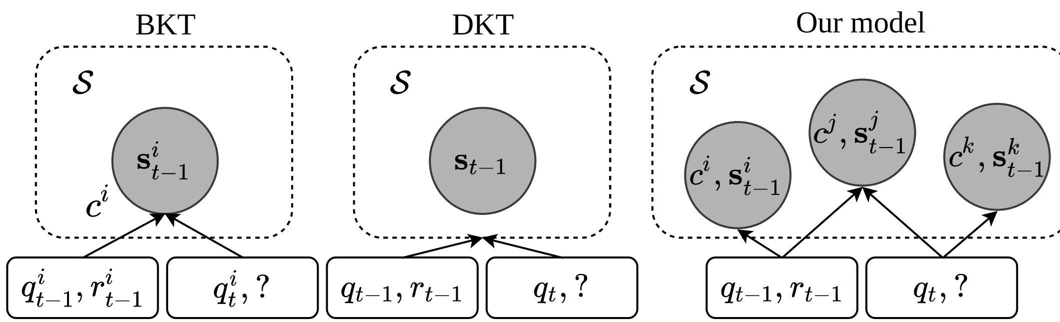
- A supervised sequence learning problem.
- **Input**: the student's past exercise performance $\mathcal{X} = \{(q_1, r_1), (q_2, r_2), \dots, (q_{t-1}, r_{t-1})\}$
- **Output**: the probability that the student will answer a new exercise correctly $p(r_t = 1 | q_t, \mathcal{X})$
- **Latent Variables**: Knowledge state $\mathcal{S} = \{\boldsymbol{s}_1, \boldsymbol{s}_2, \dots, \boldsymbol{s}_{t-1}\}$

Bayesian Knowledge Tracing (BKT) and Deep Knowledge Tracing (DKT)



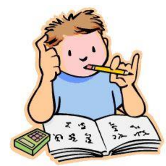
- **BKT（贝叶斯知识追踪）** [Albert et. al., UMUAI, 1995] models the knowledge state of each concept separately
  - Hidden Markov Models (HMMs)
  - Cannot capture the relationship among different concepts
- **DKT（深度知识追踪）** [Piech et. al., NIPS, 2015] summarizes a student's knowledge state of all concepts in one hidden state.
  - Long Short-Term Memory (LSTM)
  - Cannot pinpoint which concepts a student is good at or unfamiliar with

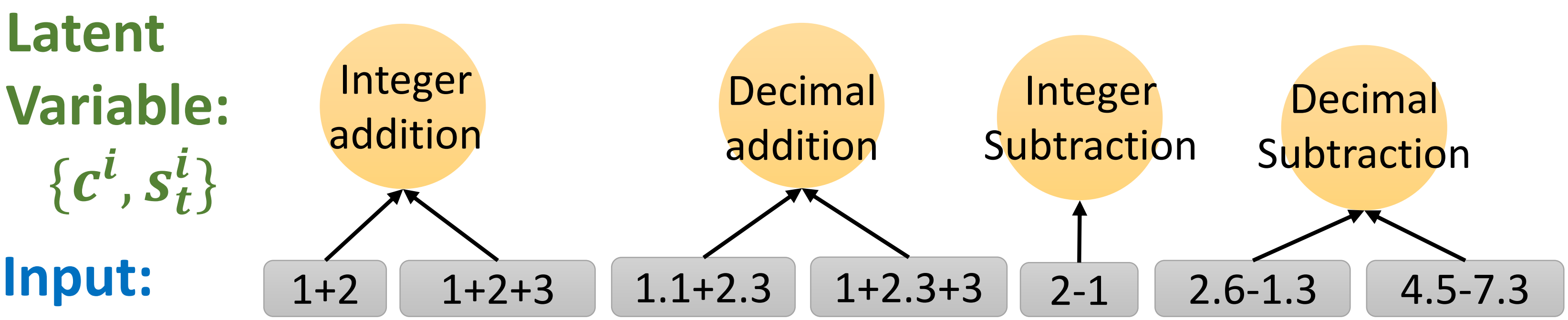
Our Model



- **Our Model**:
  - Interpretability（解释性）:
    - Directly output a student's mastery level of each concept
  - Higher performance（更好的效果）:
    - Exploit the relationships between underlying concepts
    - Keep the high representation ability of neural networks

An Example of Our Model





Output:

$$p(r_t = 1 | q_t, \mathcal{X})$$

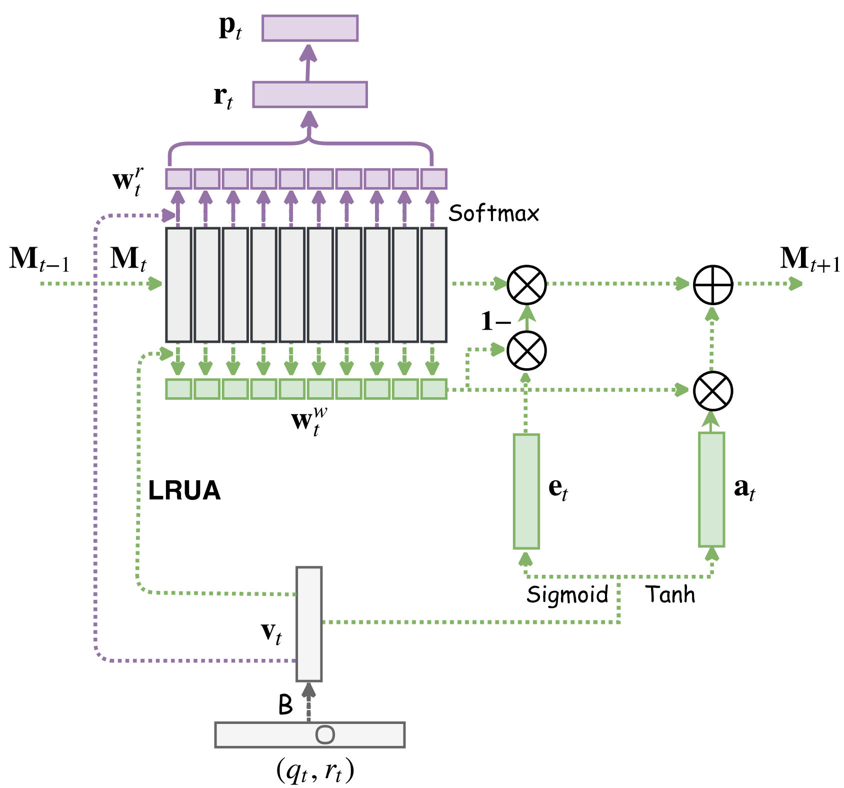
Our Contributions

- The utility of Memory-Augmented Neural Networks (MANNs 记忆增强神经网络) is exploited to better simulate the learning process of students
- A novel Dynamic Key-Value Memory Network (动态键值记忆网络) model with one static *key* matrix and one dynamic *value* matrix is proposed
- Our model can automatically discover concepts, a task that is typically performed by human experts, and depict the evolving knowledge state of students
- Our end-to-end trainable model consistently outperforms BKT and DKT(state-of-the-art) on one synthetic and three real-world datasets respectively

Memory-Augmented Neural Networks (MANN)

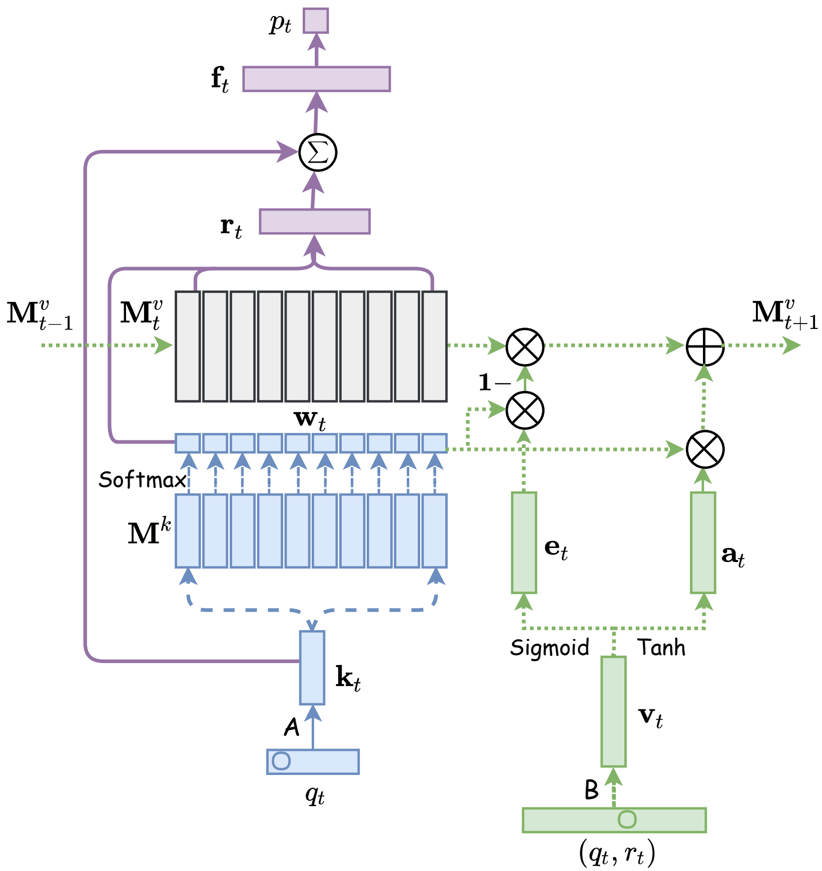
- Have a particular neural network module called external memory(外部存储器)
- Go beyond LSTM (长短期记忆神经网)
- Have a larger capacity than LSTM
- Standard MANNs: a single memory matrix or two static memory matrices
- Success in various areas:
  - **Question Answering** [Weston et. al., ICLR, 2015; Sukhbaatar et. al., NIPS, 2015; Grave et. al., Nature, 2016]
  - **Natural Language Transduction** [Grefenstette et. al.,NIPS, 2015]
  - **Algorithm Inference** [Graves et. al., Arxiv, 2014; Joulin et. al., NIPS, 2015]
  - **One-shot Learning** [Santoro et. al., ICML, 2016; Vinyals et. al., NIPS, 2016]

Memory-Augmented Neural Networks (MANN) for KT



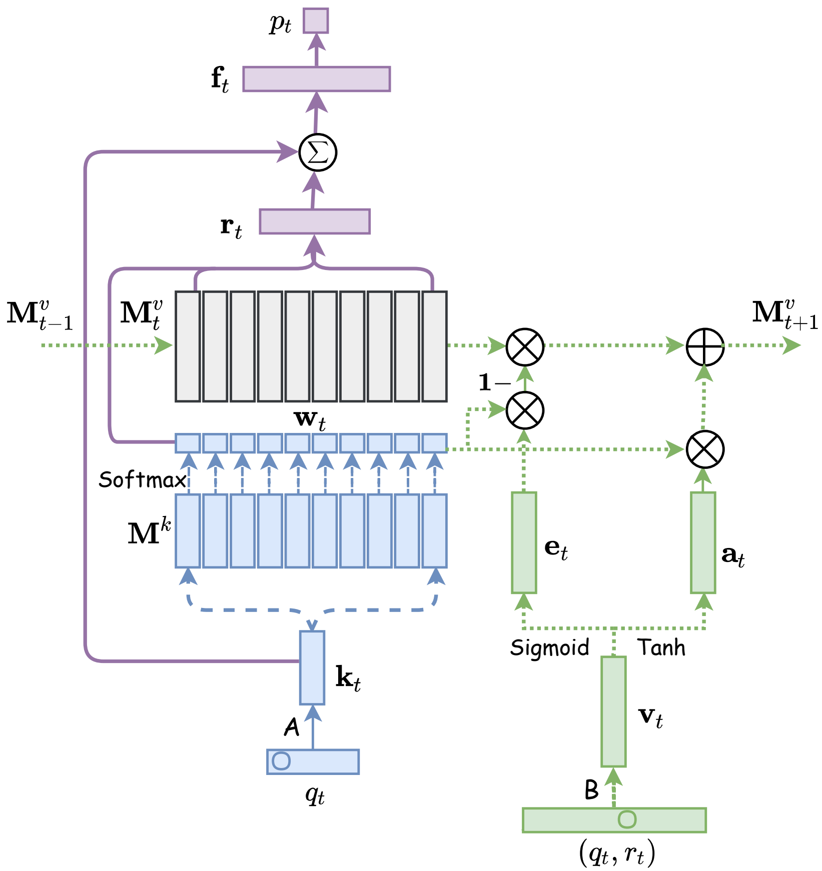
- **M** is treated as the knowledge state
- (q_t, r_t) will be used as the key to address memory locations
- Works as a list in python
- Drawbacks of MANN:
    1. The content we attend lies in the same space as the content we read and write (the addressing key?)
    2. Cannot model the underlying states of each concept (memory structure?)

Our Dynamic Key-Value Memory Networks (DKVMN) for KT



- Use *key-value* pairs for the memory structure.
  - The *key* $\mathbf{M}^k$ stores the concept representations.
  - The *value* $\mathbf{M}^v_t$ stores and updates the student's understanding (concept state) of each concept.
- Attend input q_t to the *key* $\mathbf{M}^k$, and read and write to the *value* $\mathbf{M}^v_t$
- Works as a dictionary in python.

Our Dynamic Key-Value Memory Networks for KT



- Addressing:
  - $w_t(i) = \mathrm{Softmax}(\mathbf{k}_t^T \mathbf{M}^k(i))$
- Read process:
  - $\mathbf{r}_t = \sum_{i=1}^{N} w_t(i)\, \mathbf{M}_t^v(i)$
  - $\mathbf{f}_t = \mathrm{Tanh}(\mathrm{FC}([\mathbf{r}_t, \mathbf{k}_t]))$
  - $p_t = \mathrm{Sigmoid}(\mathrm{FC}(\mathbf{f}_t))$
- Write process:
  - $\mathbf{M}_t^v(i) = \mathbf{M}_{t-1}^v(i)[\mathbf{1} - w_t(i)\mathbf{e}_t] + w_t(i)\boldsymbol{a}_t$

Experiment Results on Four Datasets

Table: Test AUC results for all datasets.

Datasets	Overview			Test AUC (%)				
	Students	Exercise Tags	Records	BKT	BKT+	DKT	MANN	DKVMN
Synthetic-5	4,000	50	200,000	62	80	80.3±0.1	81.0±0.1	**82.7±0.1**
ASSISTments2009	4,151	110	325,637	63	-	80.5±0.2	79.7±0.1	**81.6±0.1**
ASSISTments2015	19,840	100	683,801	64	-	72.5±0.1	72.3±0.2	**72.7±0.1**
Statics2011	333	1,223	189,297	73	75	80.2±0.2	77.6±0.1	**82.8±0.1**

Table: AUC with different parameters.

Model	Synthetic-5				ASSISTments2009				ASSISTments2015				Statics2011			
	s. dim	m. size	test auc	p. num	s. dim	m. size	test auc	p. num	s. dim	m. size	test auc	p. num	s. dim	m. size	test auc	p. num
DKT	10	-	80.06	2.4K	10	-	80.38	4.3K	10	-	72.40	4.0K	10	-	78.12	39K
	50	-	80.22	28K	50	-	**80.53**	37K	50	-	**72.52**	36K	50	-	79.86	205K
	100	-	**80.34**	96K	100	-	80.51	114K	100	-	72.49	111K	100	-	80.16	449K
	200	-	80.32	352K	200	-	80.43	388K	200	-	72.45	382K	200	-	**80.20**	1.0M
DKVMN	10	50	82.00	12K	10	10	81.47	7k	10	20	**72.68**	14K	10	10	82.72	92K
	50	50	82.66	25K	50	20	**81.57**	31k	50	10	72.66	29K	50	10	**82.84**	197K
	100	50	**82.73**	50K	100	10	81.42	68k	100	50	72.64	63K	100	10	82.71	338K
	200	50	82.71	130K	200	20	81.37	177k	200	50	72.53	153K	200	10	82.70	649K

Concept Discovery Results

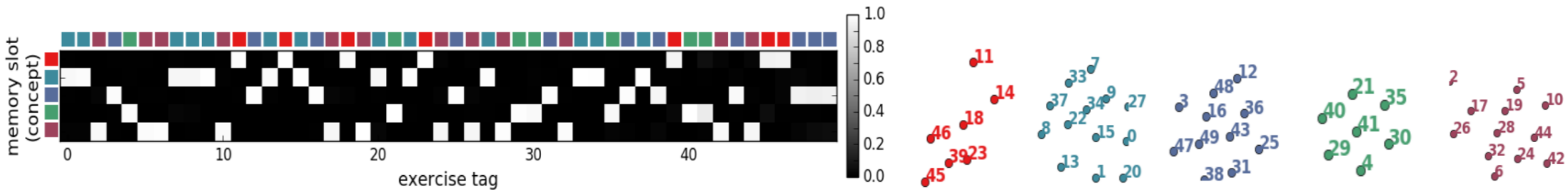


Figure: Concept discovery results on the synthetic-5 dataset when the memory size *N* is set to be 5.

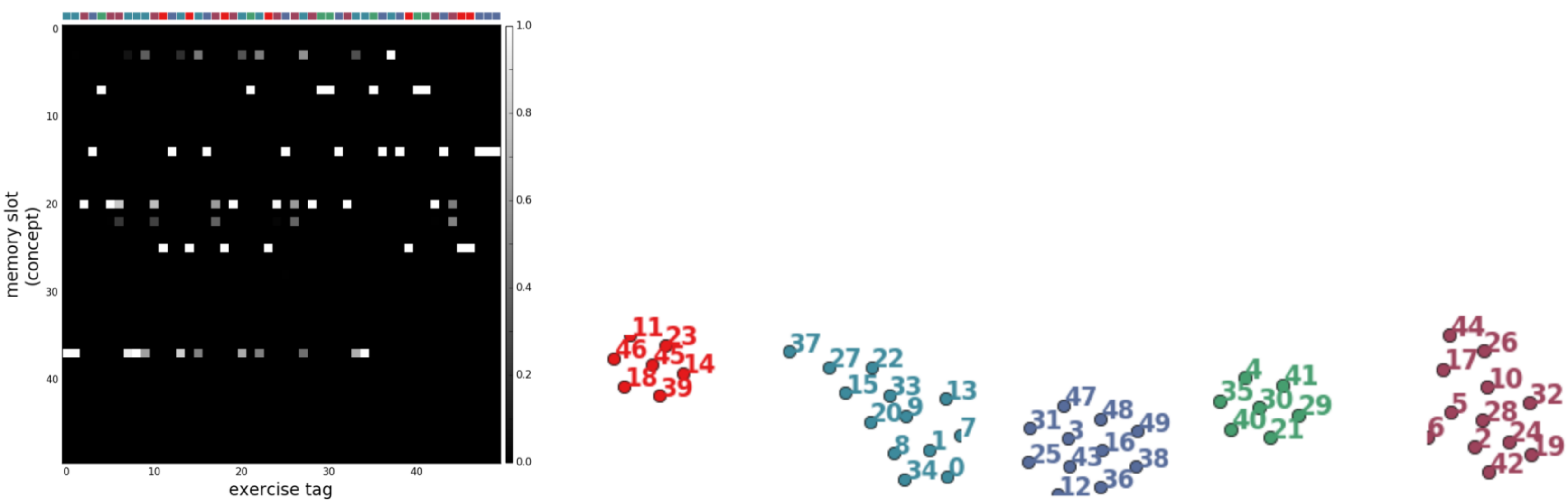


Figure: Concept discovery results on the synthetic-5 dataset when the memory size *N* is set to be 50.

Knowledge State Depiction

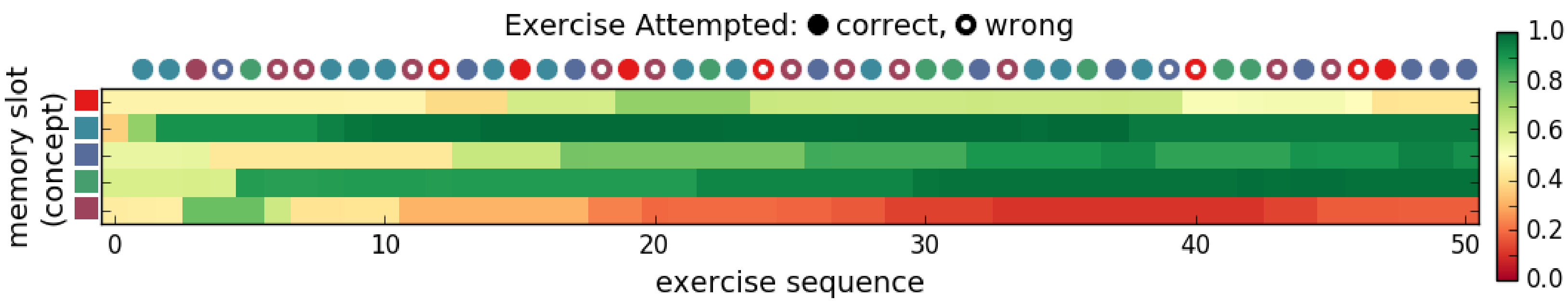


Figure: An example of a student's changing knowledge state on 5 concepts.

Conclusions and Future Work

1. The KT problem can be better simulated with the help of external memory.
2. Our DKVMN model with one static *key* matrix and one dynamic *value* matrix is now the state-of-the-art of KT.
3. We will incorporate content information into the exercise and concept embeddings to further improve their representations.
4. A new model with both dynamic *key* and *value* matricies can be designed to better simulate students' learning process.

Q & A

- Github
  - https://github.com/jennyzhang0215/DKVMN