Knowledge Tracing with Sequential Key-Value Memory Networks

Ghodai Abdelrahman and Qing Wang

THE AUSTRALIAN NATIONAL UNIVERSITY

Australian National University
Research School of Computer Science
Canberra, Australia

{ghodai.abdelrahman, qing.wang}@anu.edu.au

June 17, 2021

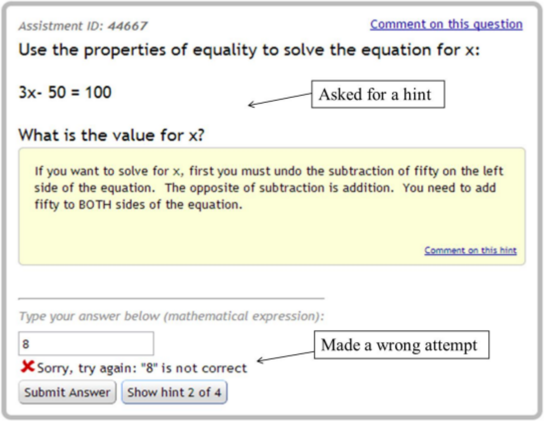
Knowledge Tracing – Introduction

- **Question:** Can machines trace human knowledge like humans?

(Image source: The New York Academy of Sciences)

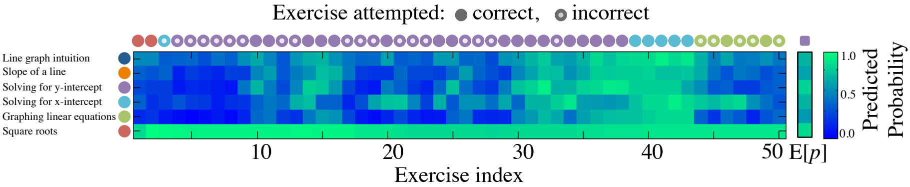
Knowledge Tracing – Introduction

- Human educators can intelligently track the knowledge of a student.



Knowledge Tracing – Introduction

- Knowledge tracing aims to build a model that can trace the knowledge of students as they interact with coursework items.



(Image source: Deep Knowledge Tracing, NIPS 2015)

- This can help enable personalised learning experiences for students.

Related Work - Overview

Deep Learning KT

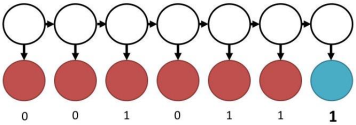
- 2017 Dynamic Key-Value Memory Networks (DKVMN) [Zhang et al.]
- 2015 Deep Knowledge Tracing (DKT) [Piech et al.]

Probabilistic KT

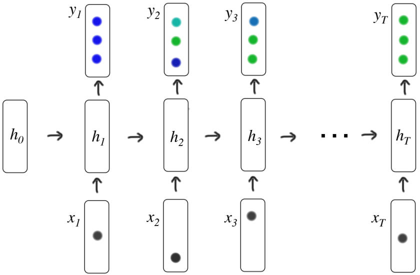
- 2009 Performance Factors Analysis (PFA) [Pavlik Jr, Cen & Koedinger]
- 2006 Learning Factors Analysis (LFA) [Cen, Koedinger & Junker]
- 1995 Bayesian Knowledge Tracing (BKT) [Corbett & Anderson]

Related Work - Bayesian Knowledge Tracing

- Built upon Hidden Markov Model (HMM)
- Represent a knowledge state as a set of binary latent variables, each for student understanding of a concept (i.e., *known* and *unknown*):
  - How can we capture the relationship between different concepts?

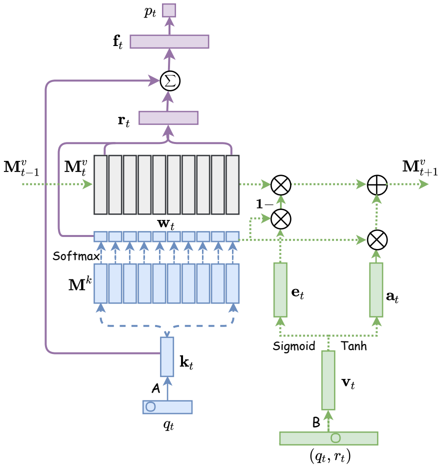


- Built upon Recurrent Neural Network (RNN)
- Summarise a student's knowledge state of all concepts in one hidden state, thus difficult to trace:
  - How much a student has mastered certain concepts?



Related Work - Dynamic Key-Value Memory Networks

- Built upon Memory-Augmented Neural Network (MANN)
- Acquire knowledge growth through the most recent exercise and fail to capture long-term dependencies
- Ignore the latest knowledge state during the memory write process



Knowledge Tracing - The Present Work

- **Research question**

  Can we build a knowledge tracing model that can alleviate the limitations of the previous work?

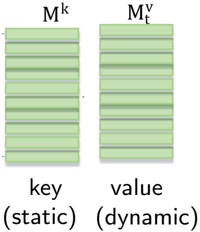
- **Problem definition**

  Given a student's exercise history $\mathrm{X} = \langle (\mathrm{q_1}, \mathrm{y_1}), \ldots, (\mathrm{q_{t-1}}, \mathrm{y_{t-1}}) \rangle$, we want to predict the probability of correctly answering the next question q_t by the student:

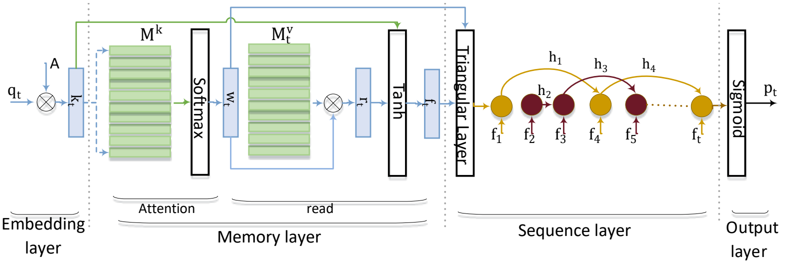
$$p_t = (y_t = 1 | q_t, X)$$

- **Assumptions**
  - Each question associates with one or more latent concepts.
  - The concept state of each latent concept is a random variable describing the mastery level of the student on this latent concept.
  - At each time step t, the knowledge state of a student is modelled as a set of all concept states of the student.

- **Key ingredients**
  1. A key-value memory $\langle \mathbf{M^k}, \mathbf{M^v_t} \rangle$
  2. Hop-LSTM
  3. Effective mechanisms for attention, read and write



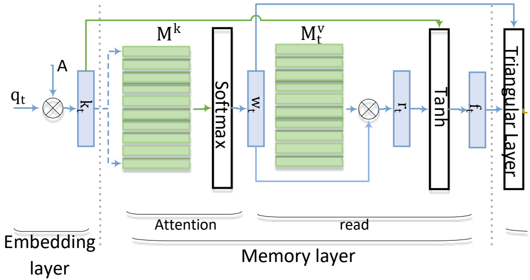
Sequential Key-Value Memory Networks (SKVMN)



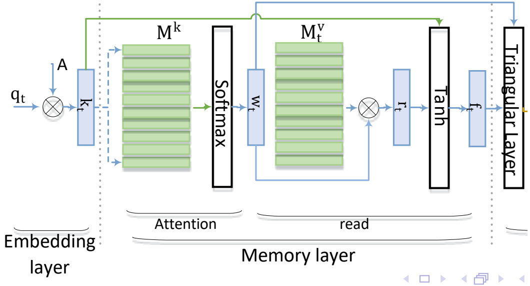
- Four layers in this model:
  1. **Embedding** layer maps a question into a continuous vector.
  2. **Memory** layer involves two processes: attention and read.
  3. **Sequence** layer consists of recurrently connected LSTM cells.
  4. **Output** layer yields probability of correctly answering a question.

- **Attention**:
  - Occur each time when a question q_t is taken.
  - Provide an addressing mechanism for the input question to allocate the relevant information from the key-value memory.
  - Obtain an attention vector $\mathbf{w_t}$ by applying Softmax to the inner product between $\mathbf{k_t}$ and each key slot $\mathbf{M^k}(\mathrm{i})$ in the key matrix $\mathbf{M^k}$:

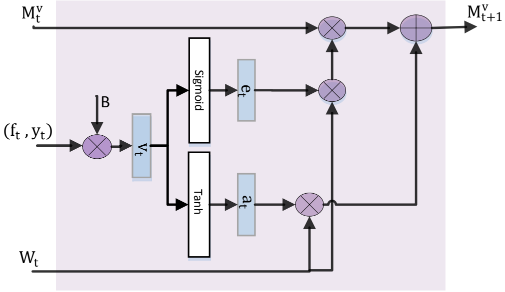
$$\mathbf{w_t}(\mathrm{i}) = \mathrm{Softmax}(\mathbf{k_t^T M^k}(\mathrm{i})).$$



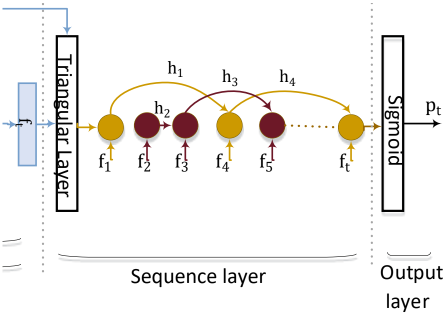
- **Read**:
  - Occur each time when a question q_t is taken.
  - Use the attention vector $\mathbf{w_t}$ to retrieve the concept states of the student w.r.t. the question $\mathrm{q_t}$ from the value matrix $\mathbf{M_t^v}$.
  - Yield the summary vector $\mathbf{f_t}$, i.e., how well the student has mastered latent concepts relevant to the question before attempting it.



- **Write**:
    - Occur each time after the student has attempted a question.
    - Update the concept states of the student using the knowledge growth $\mathbf{v_t}$ gained through attempting the question.
    - Lead to the transition of the value matrix from $\mathbf{M_t^v}$ to $\mathbf{M_{t+1}^v}$.



- **Hop-LSTM**:
  - An exercise history may contain a long sequence of exercises.
  - By hopping across irrelevant exercises, recurrent models can be applied on a shorter and more relevant sequence.
  - Sequential dependencies among exercises are identified via the attention vectors of their questions.



Experiments

- In our experiments, we aim to answer the following questions:

RQ1: What is the optimal size for a key-value memory of SKVMN?

RQ2: How does SKVMN perform on predicting a student's answers of questions, given an exercise history?

RQ3: How does SKVMN perform on discovering the correlation between latent concepts and questions?

RQ4: How does SKVMN perform on tracing the dynamics of a student's knowledge state?

Experiments

- Five datasets:

Dataset	#Questions	#Students	#Exercises	#Exercises per student
Synthetic-5	50	4,000	200,000	50
ASSISTments2009	110	4,151	325,637	78
ASSISTments2015	100	19,840	683,801	34
Statics2011	1,223	333	189,297	568
JunyiAcademy	722	199,549	25,628,935	128

- Three baselines:
  - Bayesian knowledge tracing (BKT) [UMUAI1994]
  - Deep knowledge tracing (DKT) [NIPS2015]
  - Dynamic key-value memory networks (DKVMN) [WWW2017]

[RQ1:]What is the optimal size for a key-value memory of SKVMN?

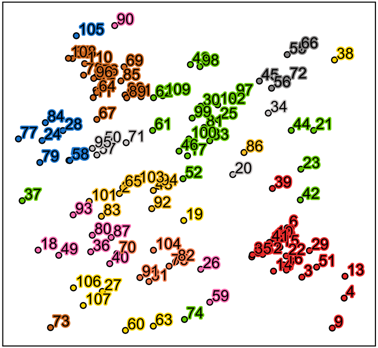
Dataset	d	N	SKVMN AUC (%)	SKVMN m	DKVMN AUC (%)	DKVMN m
Synthetic-5	10	50	83.11	15K	82.00	12k
	50	50	83.67	30k	82.66	25k
	100	50	**84.00**	57k	**82.73**	50k
	200	50	83.73	140k	82.71	130k
ASSISTments2009	10	10	**83.63**	7.8k	81.47	7k
	50	20	82.87	35k	**81.57**	31k
	100	10	82.72	71k	81.42	68k
	200	20	82.63	181k	81.37	177k
ASSISTments2015	10	20	**74.84**	16k	**72.68**	14k
	50	10	74.50	31k	72.66	29k
	100	50	74.24	66k	72.64	63k
	200	50	74.20	163k	72.53	153k
Statics2011	10	10	84.50	92.8k	82.72	92k
	50	10	**84.85**	199k	**82.84**	197k
	100	10	84.70	342k	82.71	338k
	200	10	84.76	653k	82.70	649k
JunyiAcademy	10	20	82.50	16k	79.63	14k
	50	10	82.41	31k	79.48	29k
	100	50	**82.67**	66k	79.54	63k
	200	50	82.32	163k	**80.27**	153k

[RQ2:] How does SKVMN perform on predicting a student's answers of questions, given an exercise history?

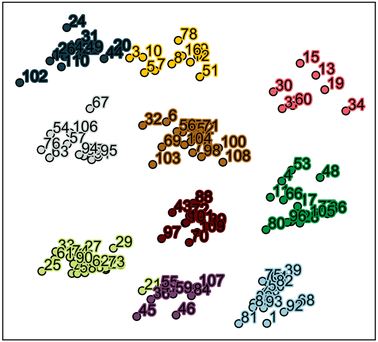
Dataset	BKT	DKT	DKVMN	SKVMN
Synthetic-5	62.0 ± 0.02	80.3 ± 0.1	82.7 ± 0.1	$\mathbf{84.0 \pm 0.04}$
ASSISTments2009	63.1 ± 0.01	80.5 ± 0.2	81.6 ± 0.1	$\mathbf{83.6 \pm 0.06}$
ASSISTments2015	64.2 ± 0.03	72.5 ± 0.1	72.7 ± 0.1	$\mathbf{74.8 \pm 0.07}$
Statics2011	73.0 ± 0.01	80.2 ± 0.2	82.8 ± 0.1	$\mathbf{84.9 \pm 0.06}$
JunyiAcademy	65.0 ± 0.02	79.2 ± 0.1	80.3 ± 0.4	$\mathbf{82.7 \pm 0.01}$

- Key observations:
  1. DL models generally performed better than probabilistic models.
  2. Memory-augmented models performed better than RNN-based models.
  3. The use of sequential dependencies among exercises in our model enhanced the prediction accuracy.

[RQ3:] How does SKVMN perform on discovering the correlation between latent concepts and questions?

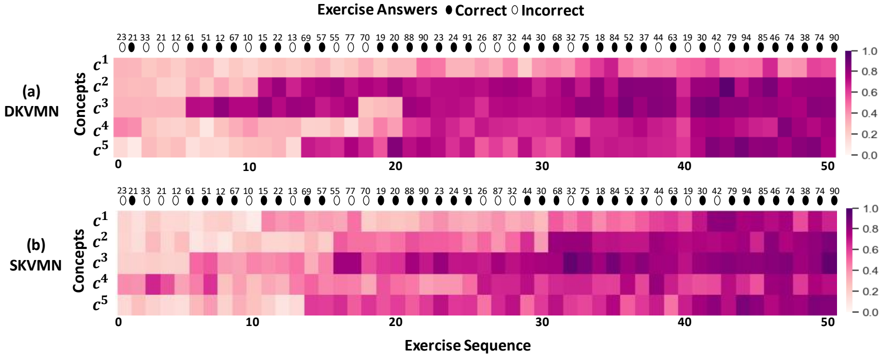


(DKVMN)



(SKVMN)

[RQ4:] How does SKVMN perform on tracing the dynamics of a student's knowledge state?



Conclusion

- We have proposed a novel knowledge tracing method called Sequential Key-Value Memory Network (SKVMN):
  - Trace knowledge states over multiple latent concepts;
  - Enhance the sequence modelling capacity;
  - Enhance memory read and writing capacity;
  - Discover the correlation between latent concepts and questions.
- Our experiments show that SKVMN outperformed the state-of-the-art methods on 5 datasets.