Knowledge Tracing Using Deep Learning Methods

Ghodai Abdelrahman Zaki

School of Computing
College of Engineering and Computer Science
Australian National University, Canberra, Australia

ghodai.abdelrahman@anu.edu.au

April 5, 2023

Agenda

- Human instructors can intelligently track the knowledge state of each student with respect to learning concepts and customize the learning experience accordingly.

- Human instructors can intelligently track the knowledge state of each student with respect to learning concepts and customize the learning experience accordingly.
- However online learning systems lack the ability to track the student's knowledge state.

- Human instructors can intelligently track the knowledge state of each student with respect to learning concepts and customize the learning experience accordingly.
- However online learning systems lack the ability to track the student's knowledge state.
- This asks the question: **can machines trace the student's knowledge state like a human instructor?**

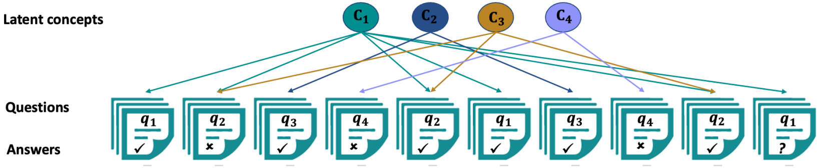
Knowledge tracing (KT): the process of estimating a student's learning state over underlying learning concepts given the previous practice history.

Knowledge Tracing Problem

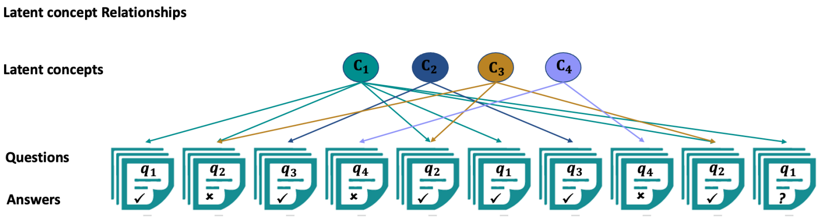

Questions
Answers
q_1 ✓
q_2 ×
q_3 ✓
q_4 ×
q_2 ✓
q_1 ✓
q_3 ✓
q_4 ×
q_2 ✓
q_1 ?

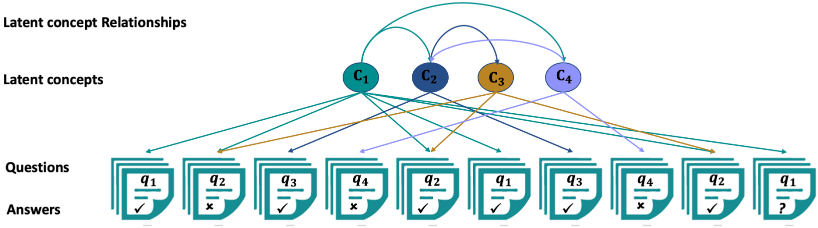
Knowledge Tracing Problem

Knowledge Tracing Problem

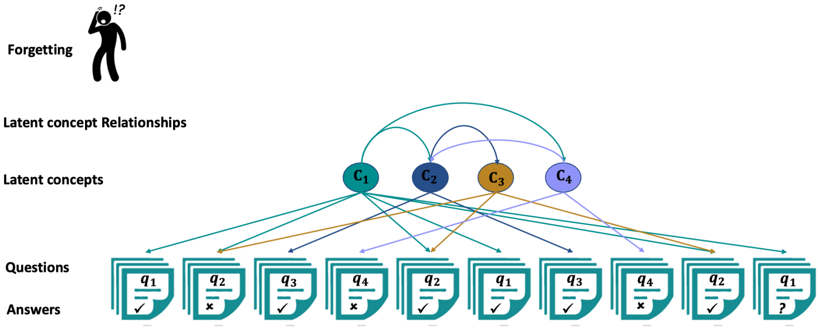


Knowledge Tracing Problem

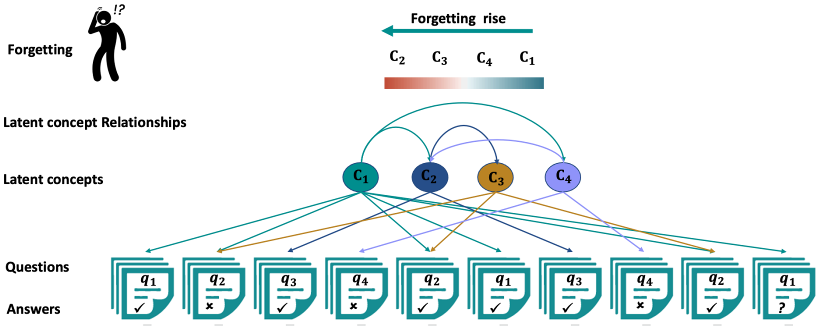



Latent concept Relationships
Latent concepts
Questions
Answers
C_1
C_2
C_3
C_4
q_1 ✓
q_2 ×
q_3 ✓
q_4 ×
q_2 ✓
q_1 ✓
q_3 ✓
q_4 ×
q_2 ✓
q_1 ?

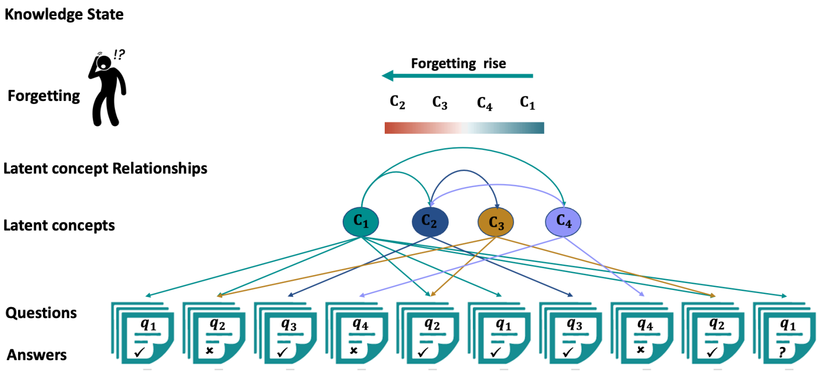
Knowledge Tracing Problem



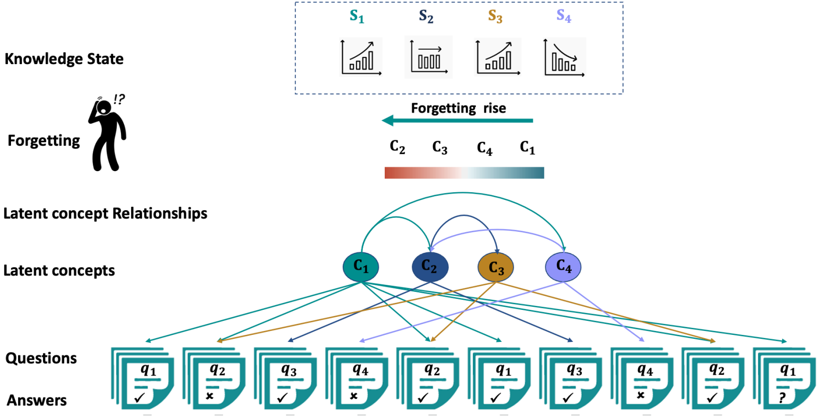
Knowledge Tracing Problem



Knowledge Tracing Problem



Knowledge Tracing Problem



Knowledge Tracing Problem

Given

- A set of question tags $Q = \{q_1, q_2, \dots, q_n\}$ relating to a hidden set of learning concepts.
- A student's practice history $\mathrm{X} = \langle (\mathrm{q}_1, \mathrm{y}_1), (\mathrm{q}_2, \mathrm{y}_2), \dots, (\mathrm{q}_{\mathrm{t}-1}, \mathrm{y}_{\mathrm{t}-1}) \rangle$, where $y_i \in \{0, 1\}$.

The KT problem is to predict the probability of correctly answering the next question q_t by the student:

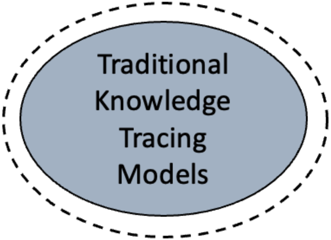
$$p_t = (y_t = 1 | q_t, X)$$

Related Work

The related work in knowledge tracing (KT)[1] problem can be categorized into two main categories:

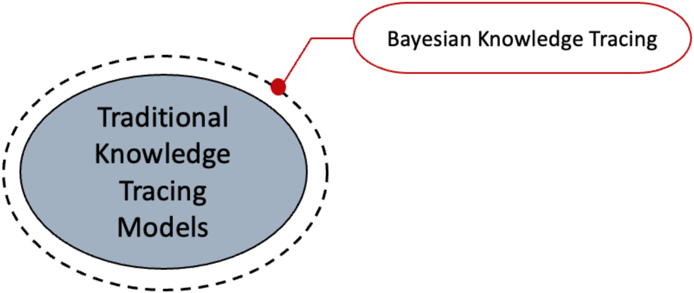
[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work



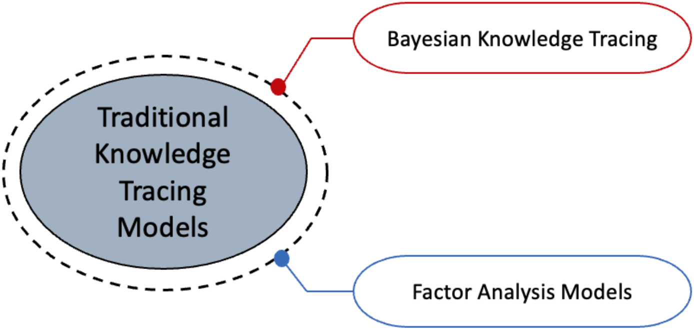
[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work



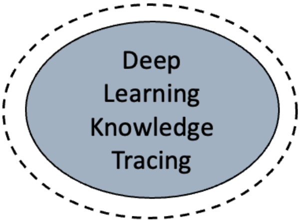
[1]Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work

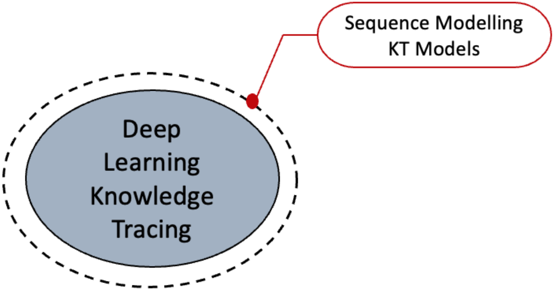


[1]Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work

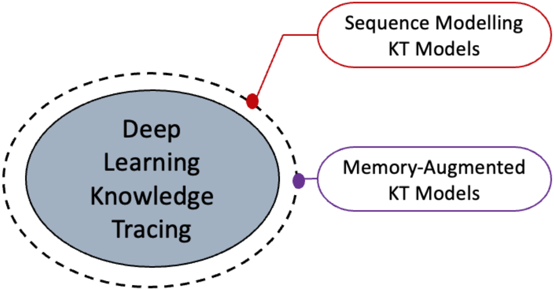


[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

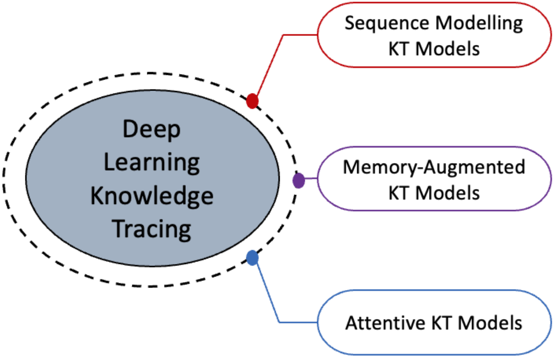


[1]Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work

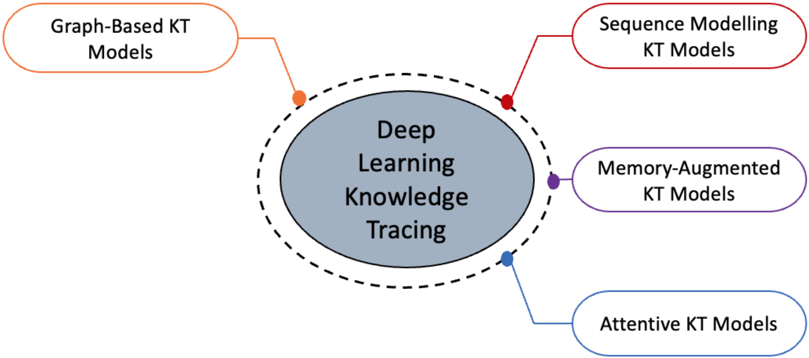


[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)



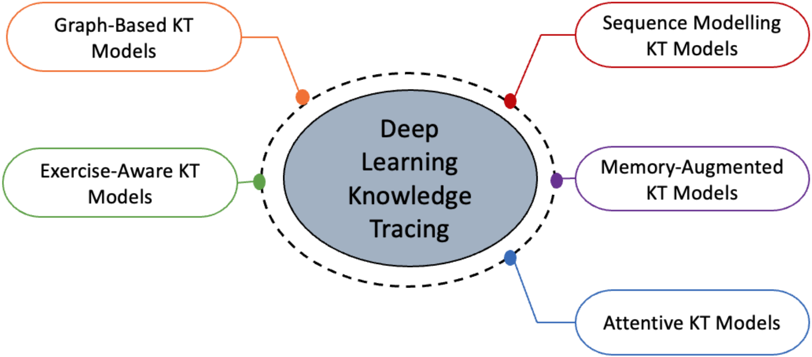
[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work

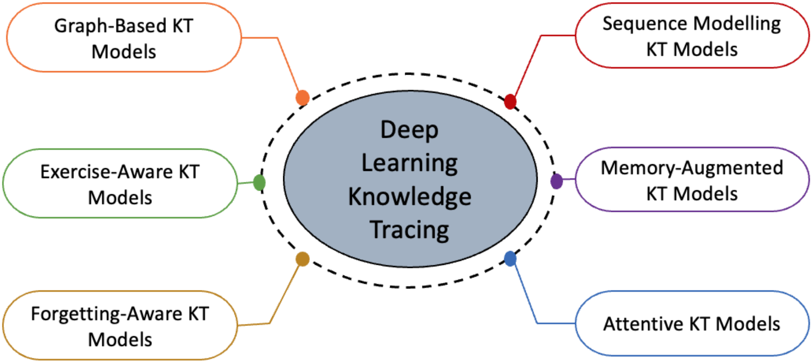


[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Related Work



[1]Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)



[1] Knowledge Tracing: A Survey, G. Abdelrahman, Q. Wang, and B. Pereira Nunes, Journal publication accepted (ACM Comp. Surveys)

Research Goals and Challenges

Research Goals

1. Conduct a detailed literature review for KT literature to guide future directions in the field.

Research Goals

1. Conduct a detailed literature review for KT literature to guide future directions in the field.
2. Propose novel and effective methods for tackling the KT problem.

Research Goals

1. Conduct a detailed literature review for KT literature to guide future directions in the field.
2. Propose novel and effective methods for tackling the KT problem.
3. Collect and formulate a new KT dataset to overcome limitations in existent datasets.

- **RQ1**: How to effectively represent and track a student's knowledge state given exercise observations?

Research Questions

- **RQ1**: How to effectively represent and track a student's knowledge state given exercise observations?
- **RQ2**: How to represent and tackle the student's forgetting behaviour during knowledge tracing?

- **RQ1**: How to effectively represent and track a student's knowledge state given exercise observations?
- **RQ2**: How to represent and tackle the student's forgetting behaviour during knowledge tracing?
- **RQ3**: Could teaching strategies be learned via knowledge tracing?

- **RQ1**: How to effectively represent and track a student's knowledge state given exercise observations?
- **RQ2**: How to represent and tackle the student's forgetting behaviour during knowledge tracing?
- **RQ3**: Could teaching strategies be learned via knowledge tracing?
- **RQ4**: What are the current limitations in KT datasets? How to address them?

Contributions and Publications

Contributions and Publications

- **Deep Knowledge Tracing**
  - **SKVMN** - G. Abdelrahman and Q. Wang. Knowledge Tracing with Sequential Key-Value Memory Networks. Conference publication accepted (SIGIR 2019).
  - **DGMN** - G. Abdelrahman and Q. Wang. Deep Graph Memory Networks for Forgetting-Robust Knowledge Tracing. Journal publication accepted (TKDE 2022).
- **Knowledge-aware Teaching Strategy Learning**
  - **KADT** - G. Abdelrahman and Q. Wang. Learning Data Teaching Strategies Via Knowledge Tracing. Under review journal publication (Knowledge-Based Systems).
- **New KT Dataset**
  - **DBE-KT22** - G. Abdelrahman, S. Mohamed, Q. Wang and Y. Lin. DBE-KT22: A Knowledge Tracing Dataset Based On Online Student Evaluation. Under review journal publication (UMUAI).
- **Comprehensive KT Survey**
  - G. Abdelrahman, Q. Wang, and B. Pereira Nunes. Knowledge Tracing: A Survey. Journal publication accepted (ACM Comp. Surveys)

Methodology

- Sequence Key-Value Memory Network (SKVMN)
- Deep Graph Memory Network (DGMN)
- Learning Knowledge Augmented Data teaching Strategies (KADT)
- Database Exercises Dataset for Knowledge Tracing (DBE-KT22)

[2] G. Abdelrahman and Q. Wang. **Knowledge Tracing with Sequential Key-Value Memory Networks**. (SIGIR 2019).

- **Contribution**

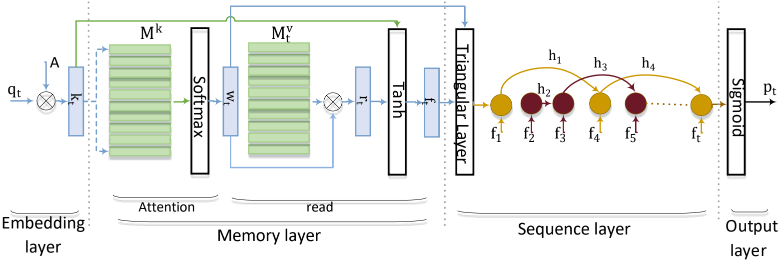
1. Knowledge tracing over multiple KCs.
   - A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M^v_t}(\textit{dynamic})\rangle$
2. Recurrent hopping mechanism.
   - Hop-LSTM
3. Knowledge-informed memory update.
   - Attention-based memory read and write

- **Contribution**

1. Knowledge tracing over multiple KCs.
   - A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M^v_t}(\textit{dynamic})\rangle$
2. Recurrent hopping mechanism.
   - Hop-LSTM
3. Knowledge-informed memory update.
   - Attention-based memory read and write

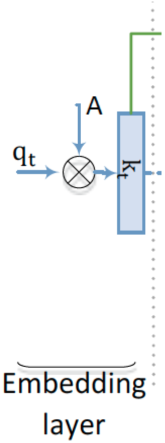
- **Contribution**

1. Knowledge tracing over multiple KCs.
   - A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M_t^v}(\textit{dynamic})\rangle$
2. Recurrent hopping mechanism.
   - Hop-LSTM
3. Knowledge-informed memory update.
   - Attention-based memory read and write

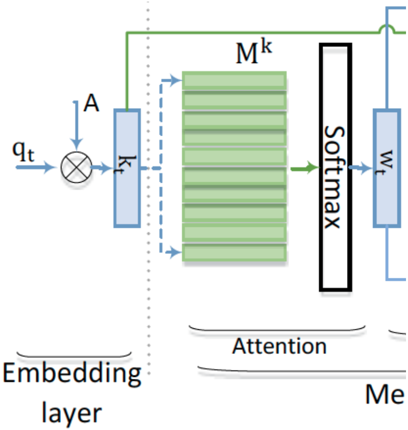


- Four layers in this model:
  1. **Embedding** layer maps a question into a continuous vector.
  2. **Memory** layer involves two processes: attention and read.
  3. **Sequence** layer consists of recurrently connected LSTM cells.
  4. **Output** layer yields probability of correctly answering a question.

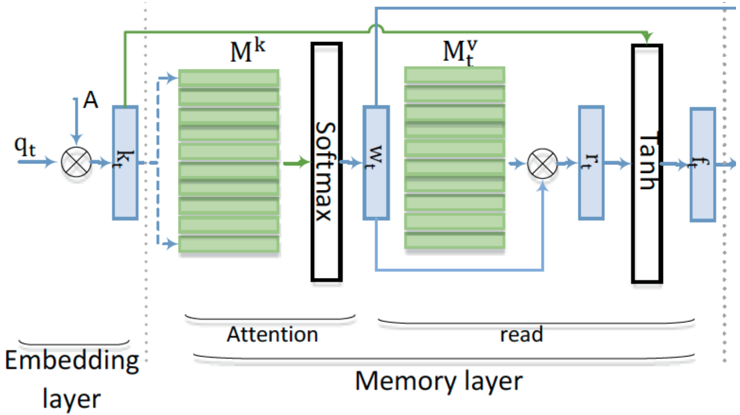
- **Embedding Layer:**



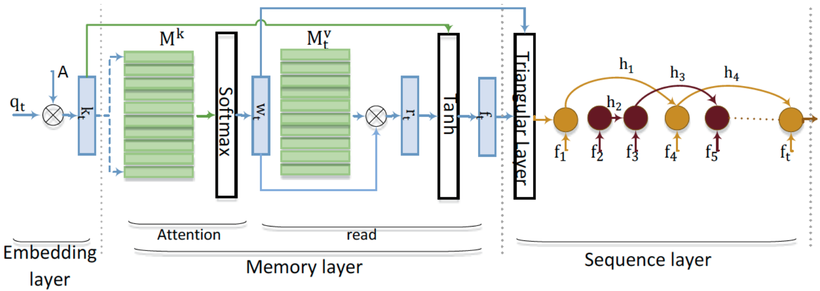
- **Memory Layer - Attention:**



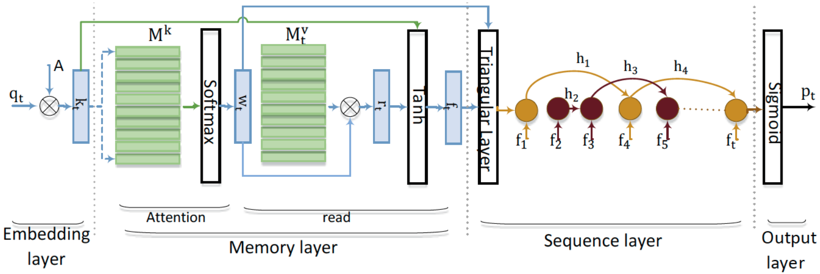
- **Memory Layer - Read:**



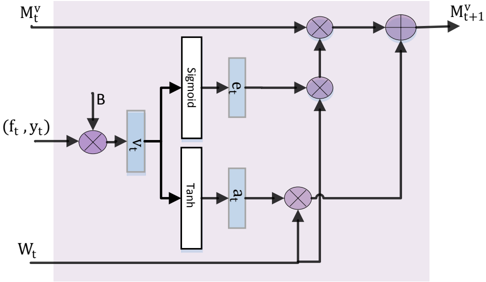
- **Sequence Layer:**



- **Output Layer:**



- **Memory Write:**



Experimental Setup

- 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	575	199,549	25,628,935	128

- Three baselines:
  - Bayesian knowledge tracing (BKT)
  - Deep knowledge tracing (DKT)
  - Dynamic key-value memory networks (DKVMN)

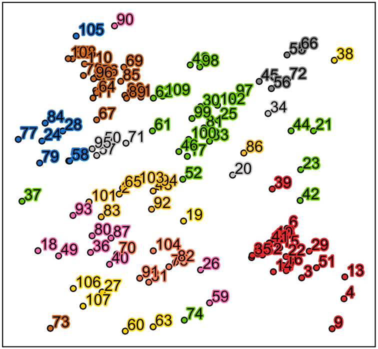
Experiments 1

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

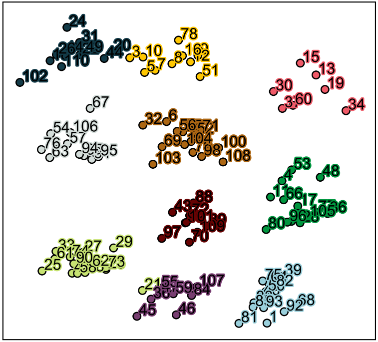
Dataset	BKT	DKT	DKVMN	SKVMN
Synthetic-5	62.0 ± 0.02	80.3 ± 0.1	82.7 ± 0.1	$\mathbf{84.0 \pm 0.04}$
ASSISTments2009	63.1 ± 0.01	80.5 ± 0.2	81.6 ± 0.1	$\mathbf{83.6 \pm 0.06}$
ASSISTments2015	64.2 ± 0.03	72.5 ± 0.1	72.7 ± 0.1	$\mathbf{74.8 \pm 0.07}$
Statics2011	73.0 ± 0.01	80.2 ± 0.2	82.8 ± 0.1	$\mathbf{84.9 \pm 0.06}$
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 traditional models.
  2. Memory-augmented models performed better than RNN-based models.
  3. Sequential dependencies among exercises in our model enhanced the prediction accuracy.

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

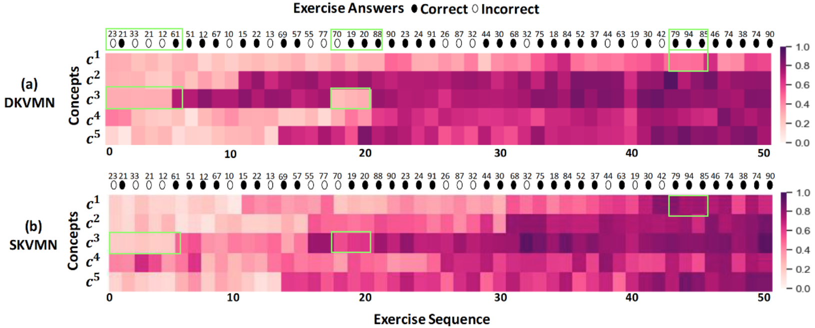


(DKVMN)



(SKVMN)

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



Methodology

- Sequence Key-Value Memory Network (SKVMN)
- Deep Graph Memory Network (DGMN)
- Learning Knowledge Augmented Data teaching Strategies (KADT)
- Database Exercises Dataset for Knowledge Tracing (DBE-KT22)

[3] G. Abdelrahman and Q. Wang. "Deep Graph Memory Networks for Forgetting-Robust Knowledge Tracing". (TKDE 2022).

- **Contribution**

1. Modelling forgetting over the knowledge space.
   – A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M^v_t}(\mathit{dynamic})\rangle$
2. KC graph representation learning.
   – Dynamic latent concept graph
3. Fusion of knowledge and forgetting features.
   – Forget Gating mechanism

- **Contribution**

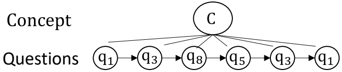
1. Modelling forgetting over the knowledge space.
   - A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M^v_t}(\textit{dynamic})\rangle$
2. KC graph representation learning.
   - Dynamic latent concept graph
3. Fusion of knowledge and forgetting features.
   - Forget Gating mechanism

- **Contribution**

1. Modelling forgetting over the knowledge space.
   - A key-value memory $\langle \mathbf{M^k}$ (static), $\mathbf{M^v_t}$(*dynamic*)$\rangle$
2. KC graph representation learning.
   - Dynamic latent concept graph
3. Fusion of knowledge and forgetting features.
   - Forget Gating mechanism

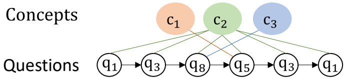
Deep Graph Memory Network (DGMN)

(a) Forgetting Over One Latent Concept

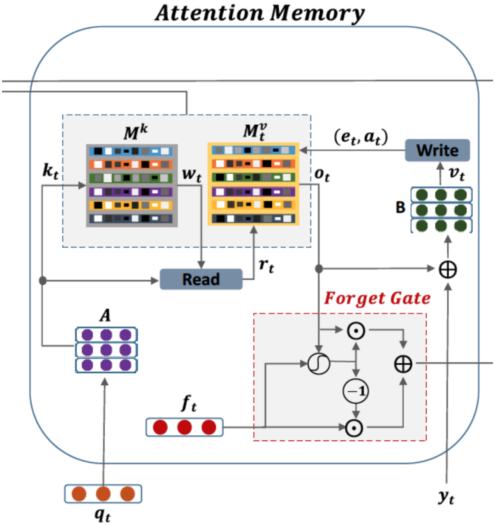


Time Step	t_1	t_2	t_3	t_4	t_5	t_6
Time Lapse	0	0	0	0	3	5
Trials	1	1	1	1	2	2

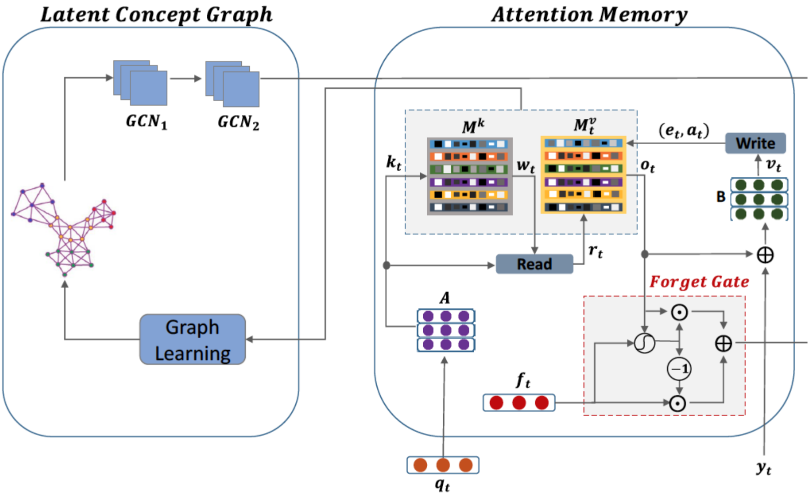
(b) Forgetting Over Multiple Latent Concepts



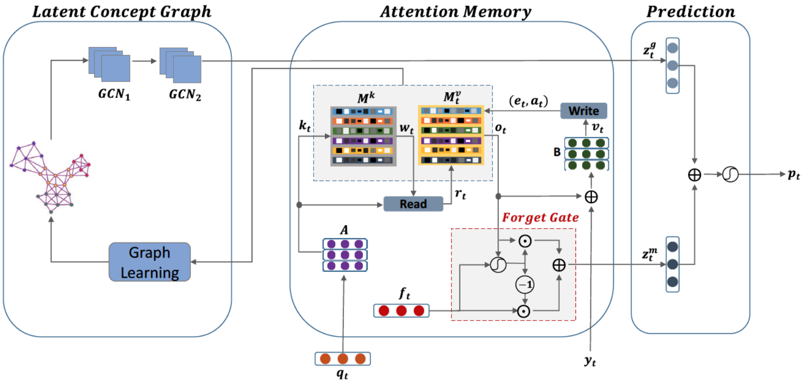
Time Step	t_1	t_2	t_3		t_4		t_5	t_6
Concepts	c_2	c_2	c_2	c_3	c_1	c_2	c_2	c_2
Time Lapse	0	1	1	0	0	1	1	1
Trials	1	2	3	1	1	4	5	6



Attention Memory
M^k
M_t^v
(e_t, a_t)
Write
k_t
w_t
o_t
v_t
B
r_t
Read
Forget Gate
A
-1
f_t
q_t
y_t


Latent Concept Graph
Attention Memory
GCN_1
GCN_2
Graph Learning
M^k
M_t^v
k_t
w_t
o_t
(e_t, a_t)
Write
v_t
B
Read
r_t
Forget Gate
A
f_t
q_t
y_t

Deep Graph Memory Network (DGMN)



Experimental Setup

- Four datasets:

Dataset	#Students	#Questions	#Exercises	#Concepts
ASSISTments2009	4,151	110	325,637	110
Statics2011	335	1,362	190,923	85
Synthetic-5	4,000	50	200,000	5
Kddcup2010	575	436	607,026	112

- Seven baselines:
  - Deep Knowledge Tracing (DKT) + forgetting(DKT+forget)
  - Dynamic key-value memory networks (DKVMN)
  - Graph-based Knowledge Tracing (GKT)
  - Self-Attentive Knowledge Tracing (SAKT)
  - Attentive Knowledge Tracing (AKT)
  - HawkesKT
  - Separated Self-Attentive Neural Knowledge Tracing (SAINT+).

How does our DGMN model perform against the state-of-the-art KT models?

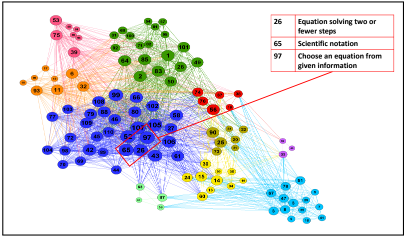
Model	Dataset			
	ASSISTments2009	Statics2011	Synthetic-5	Kddcup2010
GKT	72.3 ± 0.02	73.4 ± 0.03	74.2 ± 0.01	76.9 ± 0.01
DKT+forget	73.2 ± 0.02	74.5 ± 0.01	75.1 ± 0.03	79.0 ± 0.01
DKVMN	81.6 ± 0.03	82.8 ± 0.02	82.7 ± 0.01	79.8 ± 0.02
SAKT	83.7 ± 0.02	84.2 ± 0.01	81.9 ± 0.03	80.4 ± 0.02
HawkesKT	83.9 ± 0.04	83.8 ± 0.05	82.2 ± 0.05	79.5 ± 0.03
AKT	84.1 ± 0.03	85.0 ± 0.02	83.6 ± 0.03	81.5 ± 0.02
SAINT+	84.4 ± 0.04	85.1 ± 0.03	83.8 ± 0.03	81.6 ± 0.05
DGMN (ours)	$\mathbf{86.1 \pm 0.01}$	$\mathbf{86.4 \pm 0.02}$	$\mathbf{85.9 \pm 0.03}$	$\mathbf{83.4 \pm 0.01}$

How different components in the DGMN model affect on its performance?

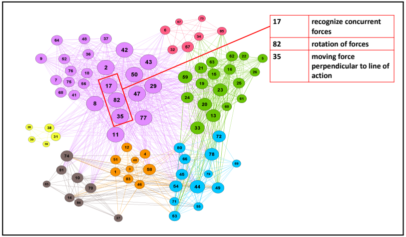
Component				Dataset			
Forget	Graph	Rank	Memory	ASSISTments2009	Statics2011	Synthetic-5	Kddcup2010
✓	Dynamic	✓	Key-Value	$\mathbf{86.1 \pm 0.01}$	$\mathbf{86.4 \pm 0.02}$	$\mathbf{85.9 \pm 0.03}$	$\mathbf{83.4 \pm 0.01}$
×	Dynamic	×	Key-Value	82.8 ± 0.01	83.9 ± 0.03	83.8 ± 0.04	80.7 ± 0.03
✓	×	×	Key-Value	85.2 ± 0.02	85.5 ± 0.01	84.7 ± 0.02	82.4 ± 0.01
✓	Dynamic	×	Key-Value	85.7 ± 0.02	85.9 ± 0.01	85.1 ± 0.02	82.7 ± 0.03
×	×	×	Key-Value	81.9 ± 0.01	82.9 ± 0.03	82.9 ± 0.01	80.0 ± 0.02
×	×	×	RNN	72.8 ± 0.08	73.9 ± 0.07	73.6 ± 0.04	76.2 ± 0.03
×	Static	×	Key-Value	75.4 ± 0.09	76.1 ± 0.08	75.7 ± 0.05	78.3 ± 0.05

Experiments 3

How effectively can a learned latent concept graph reflect latent concepts and their relationships?

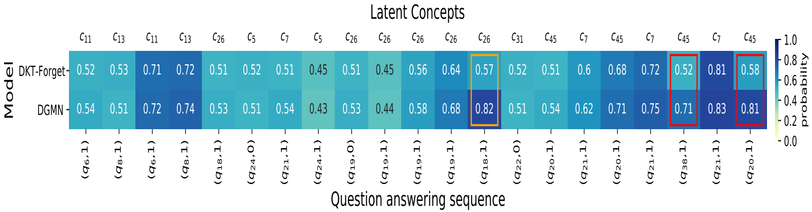


(ASSISTments2009)



(Statics2011)

How does modelling forgetting over multiple concepts in our model compare to other deep learning models assuming only a single concept?



- Sequence Key-Value Memory Network (SKVMN)
- Deep Graph Memory Network (DGMN)
- Learning Knowledge Augmented Data teaching Strategies (KADT)
- Database Exercises Dataset for Knowledge Tracing (DBE-KT22)

[4] G. Abdelrahman and Q. Wang. "Learning Data Teaching Strategies Via Knowledge Tracing". Under review journal publication(Knowledge-Based Systems).

- **Contribution**

1. Evolving data teaching strategies through RL.
2. Knowledge-informed state representation learning.
3. Adaptive training procedure.

- **Contribution**

1. Evolving data teaching strategies through RL.
2. Knowledge-informed state representation learning.
3. Adaptive training procedure.

- **Contribution**

1. Evolving data teaching strategies through RL.
2. Knowledge-informed state representation learning.
3. Adaptive training procedure.

Knowledge Augmented Data Teaching (KADT)

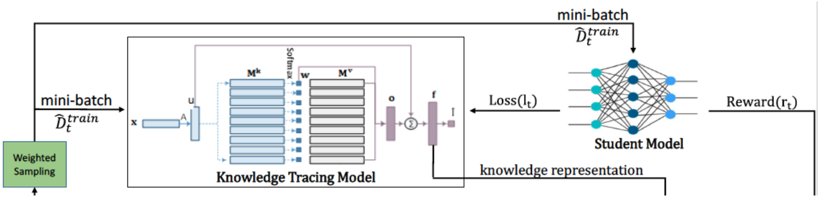
- **Problem definition**
  - **Student problem**: a conventional supervised learning problem to minimize a loss function given a labeled training dataset.

$$\theta^* = \underset{\theta \in \Theta}{\arg\min}\ \mathrm{L}(h_\theta, D^{\mathrm{train}})$$

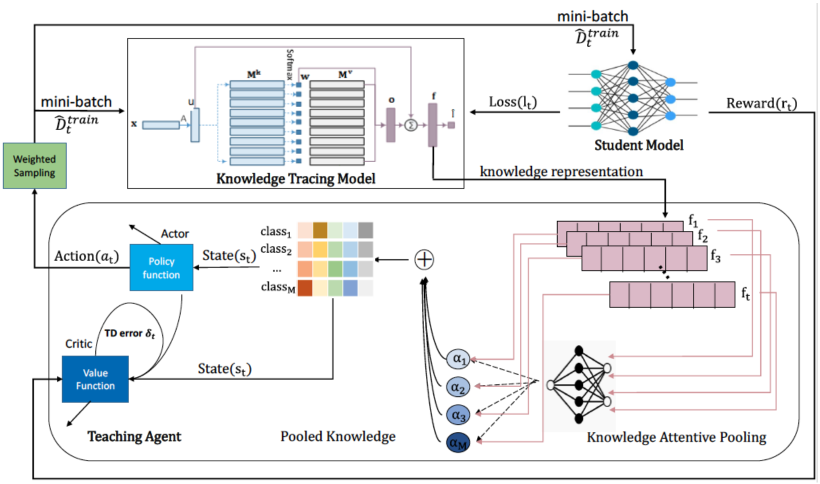
  - **Teacher problem**: a teacher agent g_ω aims at maximizing rewards from progress on a student's validation performance metric η

$$\omega^* = \underset{\omega \in \Omega}{\arg\max}\, \eta(h_\theta, g_\omega(D^{\mathrm{train}}), D^{\mathrm{valid}})$$

- **Student Knowledge Tracing:**



- **Learning Teaching Strategy:**



Experimental Setup

- Four datasets:

Task	Dataset	Student Model Setup [training→Testing] Similar Student Mode	Different Student Mode
Knowledge Tracing	ASSISTments2009	[SKVMN→SKVMN]	[SKVMN→DKT]
Sentiment Analysis	IMDB	[LSTM→LSTM]	[LSTM→SVM]
Movie Recommendations	MovieLens	[LSTM→LSTM]	[LSTM→MF]
Image Recognition	CIFAR-100	[MLP→MLP]	[MLP→CNN]

- Three baselines:
  - Self-Paced Learning (SPL)
  - Learning to Teach (L2T)
  - Random Sampling (RandomTeach)

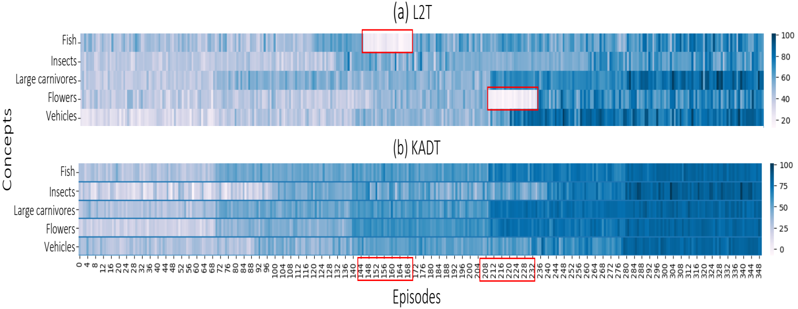
How effectively can our KADT method teach a student model across different learning tasks?

Dataset	Similar Student Mode			
	RandomTeach	**SPL**	**L2T**	**KADT**
ASSISTments2009	81.63 ± 0.4	83.45 ± 0.02	84.31 ± 0.03	$\mathbf{87.10 \pm 0.04}$
IMDB	88.54 ± 0.8	88.80 ± 0.05	89.46 ± 0.06	$\mathbf{92.24 \pm 0.05}$
MovieLens	75.12 ± 0.6	76.45 ± 0.06	77.23 ± 0.03	$\mathbf{80.31 \pm 0.02}$
CIFAR-100	62.18 ± 0.4	64.68 ± 0.04	65.33 ± 0.06	$\mathbf{68.75 \pm 0.03}$

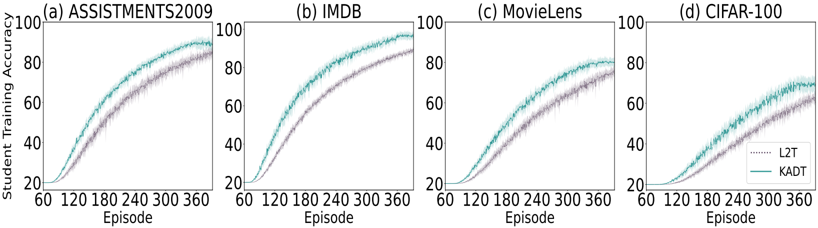
How well can our KADT method be generalized to teach different student models?

Dataset	**Different Student Mode**			
	RandomTeach	**SPL**	**L2T**	**KADT**
ASSISTments2009	79.82 ± 0.3	80.41 ± 0.06	81.29 ± 0.04	$\mathbf{84.26 \pm 0.04}$
IMDB	78.87 ± 0.5	81.04 ± 0.03	83.22 ± 0.05	$\mathbf{86.16 \pm 0.03}$
MovieLens	74.82 ± 0.8	76.87 ± 0.04	77.73 ± 0.06	$\mathbf{80.47 \pm 0.03}$
CIFAR-100	68.71 ± 0.5	70.27 ± 0.05	71.82 ± 0.03	$\mathbf{74.58 \pm 0.02}$

How well can our KADT method learn knowledge representation of a student model to improve the performance?



How well can our KADT method perform in comparison to the state-of-the-art RL teaching method?



- Sequence Key-Value Memory Network (SKVMN)
- Deep Graph Memory Network (DGMN)
- Learning Knowledge Augmented Data teaching Strategies (KADT)
- Database Exercises Dataset for Knowledge Tracing (DBE-KT22)

[5] G. Abdelrahman, S. Abdelfattah, Q. Wang and Y. Lin. "DBE-KT22: A Knowledge Tracing Dataset Based on Online Student Evaluation".Under review journal publication (UMUAI)

DBE-KT22 Dataset

- Data Collection
  - Based on exercise practicing in the Relational Databases course taught at the ANU.
  - Collected using the *CodeBench* online platform [1].
  - Spans over a three-years time period from 2018 to 2021.
  - Published through the *Australian Data Archive* (ADA) platform [2].

[1] https://cs.anu.edu.au/dab/index.html
[2] https://ada.edu.au/

- Data Collection
  - Based on exercise practicing in the Relational Databases course taught at the ANU.
  - Collected using the *CodeBench* online platform [1].
  - Spans over a three-years time period from 2018 to 2021.
  - Published through the *Australian Data Archive* (ADA) platform [2].

[1] https://cs.anu.edu.au/dab/index.html

[2] https://ada.edu.au/

- Data Collection
  - Based on exercise practicing in the Relational Databases course taught at the ANU.
  - Collected using the *CodeBench* online platform [1].
  - Spans over a three-years time period from 2018 to 2021.
  - Published through the *Australian Data Archive* (ADA) platform [2].

[1] https://cs.anu.edu.au/dab/index.html

[2] https://ada.edu.au/

DBE-KT22 Dataset

- Data Collection
  - Based on exercise practicing in the Relational Databases course taught at the ANU.
  - Collected using the *CodeBench* online platform [1].
  - Spans over a three-years time period from 2018 to 2021.
  - Published through the *Australian Data Archive* (ADA) platform [2].

[1] https://cs.anu.edu.au/dab/index.html

[2] https://ada.edu.au/

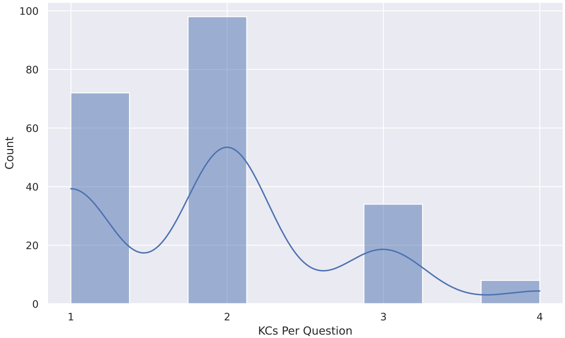
- Data Collection
  - Based on exercise practicing in the Relational Databases course taught at the ANU.
  - Collected using the *CodeBench* online platform [1].
  - Spans over a three-years time period from 2018 to 2021.
  - Published through the *Australian Data Archive* (ADA) platform [2].

[1] `https://cs.anu.edu.au/dab/index.html`
[2] `https://ada.edu.au/`

- Dataset facts

Fact	**Value**
Number of registered students	1810
Number of participating students	1361
Total number of questions	212
Total number of KCs	98
Total number of exercises	167222
Student participation ratio	75%
Number of course instructors	2
Hard questions percentage	44.8%
Medium questions percentage	39.6%
Easy questions percentage	15.6%
Percentage of questions with hint	33.5%

- Dataset distribution

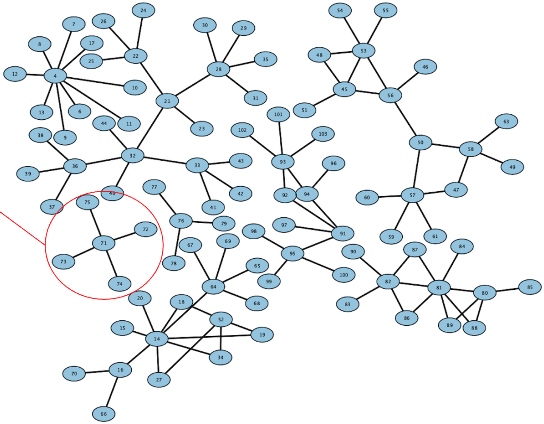


- Dataset statistics

Dataset	Question Text	Relationships		Question Difficulty	Answer Confidence
		Question - KC	KC - KC		
ASSISTments	×	✓	×	✓	×
STATICS	×	✓	×	×	×
Junyi Academy	×	✓	×	✓	×
Simulated-5	×	✓	×	×	×
KDDcup	×	✓	×	✓	×
EdNet	×	✓	×	✓	×
DBE-KT22	✓	✓	✓	✓	✓

- KC-KC relationship graph

75	Third normal form (3NF)
74	Boyce-Codd normal form (BCNF)
71	Normalization
73	Dependency preservation
72	Lossless Join



Conclusion & Future Work

In summary, this thesis proposed three novel models and a new dataset for knowledge tracing research including:

- SKVMN: Sequence Key-Value Memory Network.
- DGMN: Deep Graph Memory Networks.
- KADT: Knowledge Augmented Data Teaching.
- DBE-KT22: Database Exercises dataset.
- Comprehensive KT Survey.

Conclusion & Future Work

Future research directions to extend on our work includes:

- Investigating multi-modal representation learning methods for question, concept, and forgetting aspects.
- Exploring the potential of self-supervised and semi-supervised methods for dealing with limited training data and transfer learning across subjects.
- Exploring interactive (e.g., reinforcement learning) and collaborative (e.g., factorization models) methods to address the student's cold start challenge.
- Exploring methods that can fuse global (across the group level) and local (across individual level) knowledge features during modelling.

Conclusion & Future Work

Future research directions to extend on our work includes:

- Investigating multi-modal representation learning methods for question, concept, and forgetting aspects.
- Exploring the potential of self-supervised and semi-supervised methods for dealing with limited training data and transfer learning across subjects.
- Exploring interactive (e.g., reinforcement learning) and collaborative (e.g., factorization models) methods to address the student's cold start challenge.
- Exploring methods that can fuse global (across the group level) and local (across individual level) knowledge features during modelling.

Conclusion & Future Work

Future research directions to extend on our work includes:

- Investigating multi-modal representation learning methods for question, concept, and forgetting aspects.
- Exploring the potential of self-supervised and semi-supervised methods for dealing with limited training data and transfer learning across subjects.
- Exploring interactive (e.g., reinforcement learning) and collaborative (e.g., factorization models) methods to address the student's cold start challenge.
- Exploring methods that can fuse global (across the group level) and local (across individual level) knowledge features during modelling.

Conclusion & Future Work

Future research directions to extend on our work includes:

- Investigating multi-modal representation learning methods for question, concept, and forgetting aspects.
- Exploring the potential of self-supervised and semi-supervised methods for dealing with limited training data and transfer learning across subjects.
- Exploring interactive (e.g., reinforcement learning) and collaborative (e.g., factorization models) methods to address the student's cold start challenge.
- Exploring methods that can fuse global (across the group level) and local (across individual level) knowledge features during modelling.

I want to acknowledge my gratitude for the support and guidance that I got during my PhD from Dr. Yu Lin. I express my deepest sympathy to his family, friends, and the ANU community.

Thank you!