



UNIVERSITY  
AT ALBANY

State University of New York

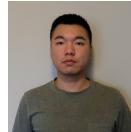
This paper is based upon work supported by  
the National Science Foundation under Grant  
No. 2047500.



# Transition-Aware Multi-Activity Knowledge Tracing



Siqian Zhao



Chunpai Wang



Shaghayegh (Sherry) Sahebi

Presented By Siqian Zhao (szhao2@Albany.edu)

2022 IEEE International Conference on Big Data (Big Data)



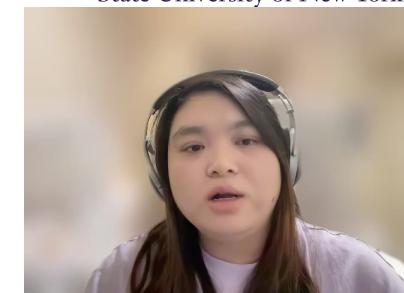
# Content

- Introduction
- Model
- Experiments
- Conclusions



# Introduction

- Online education systems
  - Enabling distance learning and abundant courses
  - Attracting more and more students
  - Promoting the development of Educational Data Mining (EDM)



# Introduction

- Online education systems
  - Enabling distance learning and abundant courses
  - Attracting more and more students
  - Promoting the development of Educational Data Mining (EDM)
- Student knowledge tracing
  - Quantifying student knowledge state
    - Understanding student learning
    - Creating a study plan
    - Recommending learning materials
    - Analyzing knowledge gaps





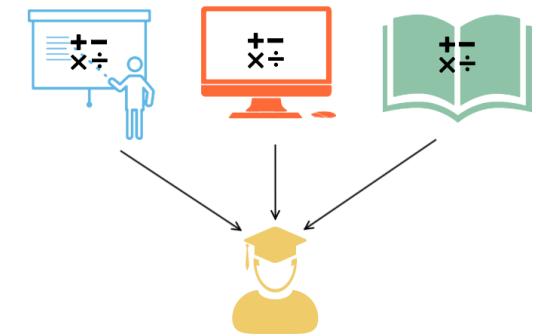
UNIVERSITY  
AT ALBANY

State University of New York



# Introduction

- Motivation and Limitation
  - Student learn by doing multiple types of activities
    - Solve questions (assessed), watch video lectures (non-assessed)



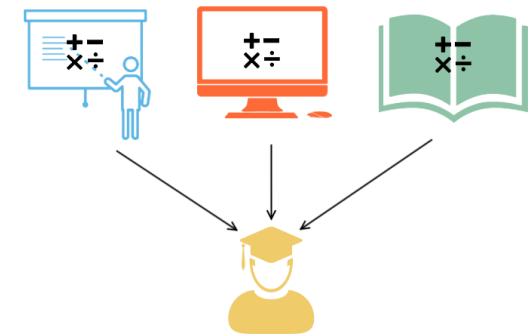
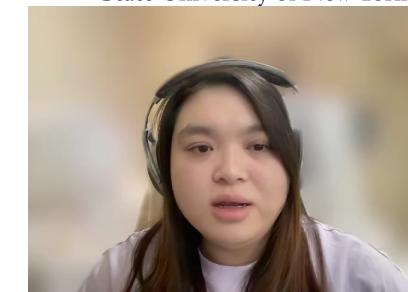


UNIVERSITY  
AT ALBANY

State University of New York

# Introduction

- Motivation and Limitation
  - Student learn by doing multiple types of activities
    - Solve questions (assessed), watch video lectures (non-assessed)
  - Most KT approaches rely on assessed activities (supervised sequence learning)
    - Linear and/or logistic regression, e.g. IRT [Frederic et al.], PFA [Philip et al.]
    - Hidden Markov model (HMM), e.g. BKT [Corbett et al.]
    - Recurrent Neural Networks, e.g. DKT [Piech et al.]





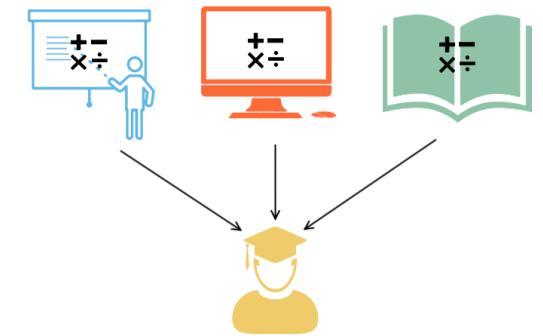
UNIVERSITY  
AT ALBANY

State University of New York



# Introduction

- Motivation and Limitation
  - Few multi-activity KT approaches
    - Factorization machine, e.g. MA-FA [Abdi et al.]
    - Elo-based learner model, e.g. MAM-Elo [Abdi et al.]
    - Tensor factorization, e.g. MVKM [Zhao et al.]





# Introduction

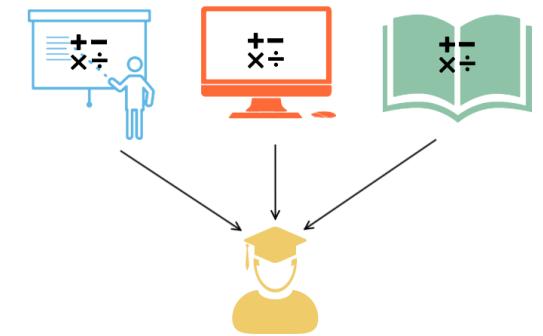
- Motivation and Limitation

- Few multi-activity KT approaches
  - Factorization machine, e.g. MA-FA [Abdi et al.]
  - Elo-based learner model, e.g. MAM-Elo [Abdi et al.]
  - Tensor factorization, e.g. MVKM [Zhao et al.]

- Knowledge transfer

- Different activities more/less helpful for other types of activities
  - Example:

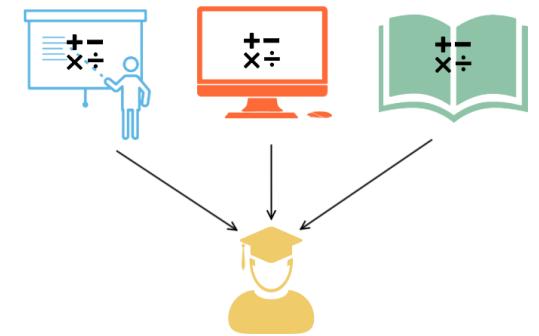
practicing summation problems → multiplication lecture  
multiplication lecture ↗ practicing summation problems
- Dynamics and realization of transfer depend on the transition order





# Introduction

- Motivation and Limitation
  - Few multi-activity KT approaches
    - Factorization machine, e.g. MA-FA [Abdi et al.]
    - Elo-based learner model, e.g. MAM-Elo [Abdi et al.]
    - Tensor factorization, e.g. MVKM [Zhao et al.]
  - Knowledge transfer
    - Different activities more/less helpful for other types of activities
      - Example:  
practicing summation problems → multiplication lecture  
multiplication lecture ↗ practicing summation problems
    - Dynamics and realization of transfer depend on the transition order
- Propose a transition-aware multi-activity component on top of LSTM (TAMKOT)

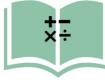
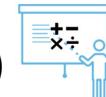


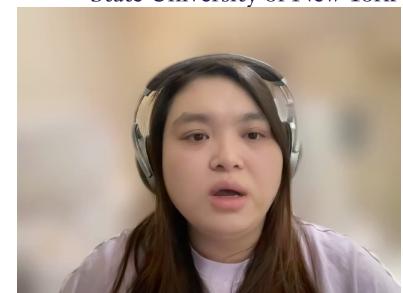
# Model

- Problem formulation
- TAMKOT model



# Problem formulation

- Suppose two materials types: Questions (Assessed),  Video lectures (Non-Assessed) 





# UNIVERSITY AT ALBANY

State University of New York

# Problem formulation

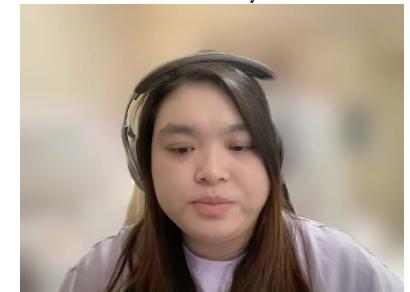
- Suppose two materials types: Questions (Assessed),  
Video lectures (Non-Assessed)
  - Each activity at learning point  $t$ :  $\langle (q_t, r_t), l_t, d_t \rangle$ 
    - $q_t$ : question 
    - $r_t$ : student performance: grades/correctness/ score  
    - $l_t$ : video lecture 
    - $d_t$ : material type: 1: assessed, 0: non-assessed





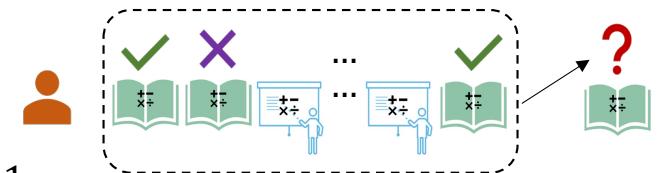
# UNIVERSITY AT ALBANY

State University of New York



# Problem formulation

- Suppose two materials types: Questions (Assessed),  
Video lectures (Non-Assessed)
  - Each activity at learning point  $t$ :  $\langle (q_t, r_t), l_t, d_t \rangle$ 
    - $q_t$ : question
    - $r_t$ : student performance: grades/correctness/ score
    - $l_t$ : video lecture
    - $d_t$ : material type: 1: assessed, 0: non-assessed
  - Goal: predict every  $r_{t+1}$ 
    - Given  $\{(q_1, r_1), l_1, d_1\}, \dots, \{(q_t, r_t), l_t, d_t\}\}$  and  $q_{t+1}$





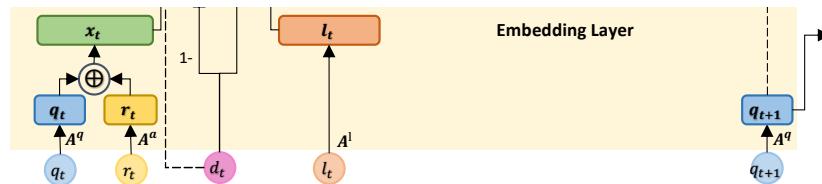
UNIVERSITY  
AT ALBANY

State University of New York



# TAMKOT Model

$S_{QQ}$	$(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$	$(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$S_{QL}$	$d_t(1 - d_{t-1})$	$S_{LL}$	$d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\bigoplus$	: elementwise addition

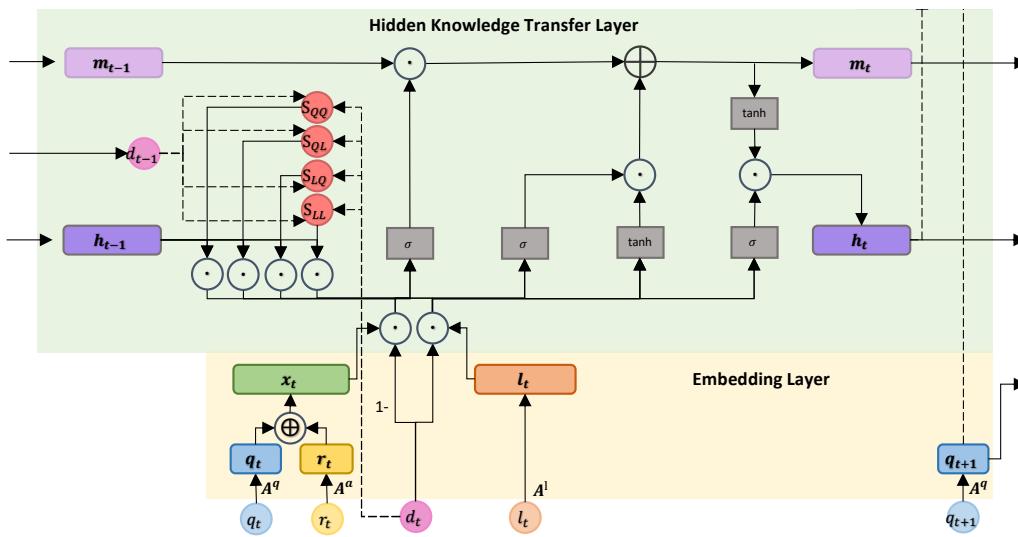


- Embedding layer
  - Map learning materials and student performance into latent concept space



# TAMKOT Model

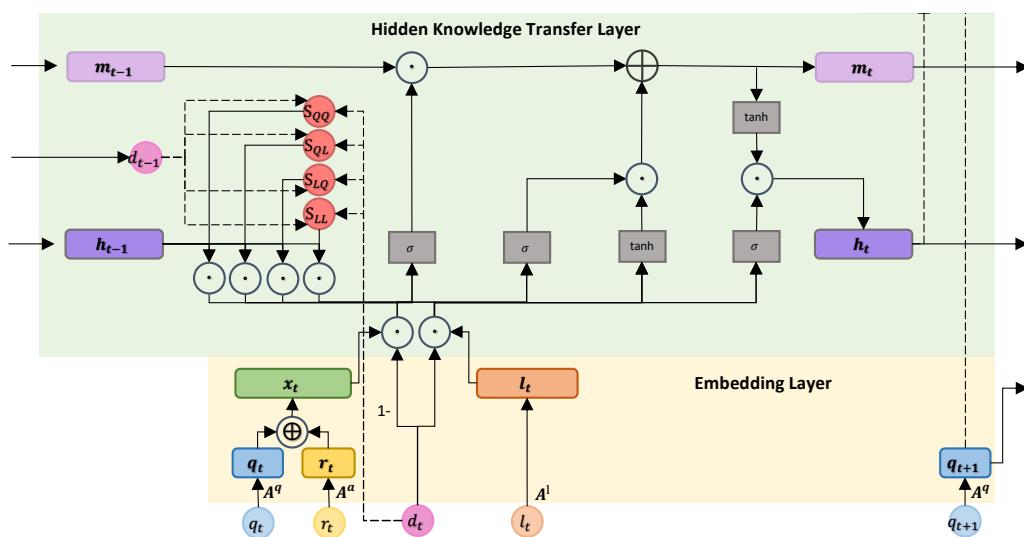
$s_{QQ}$	: $(1 - d_t)(1 - d_{t-1})$	$s_{LQ}$	: $(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$s_{QL}$	: $d_t(1 - d_{t-1})$	$s_{LL}$	: $d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\oplus$	: elementwise addition



- Hidden Knowledge Transfer Layer
  - Represent student knowledge
- Embedding layer
  - Map learning materials and student performance into latent concept space

# TAMKOT Model

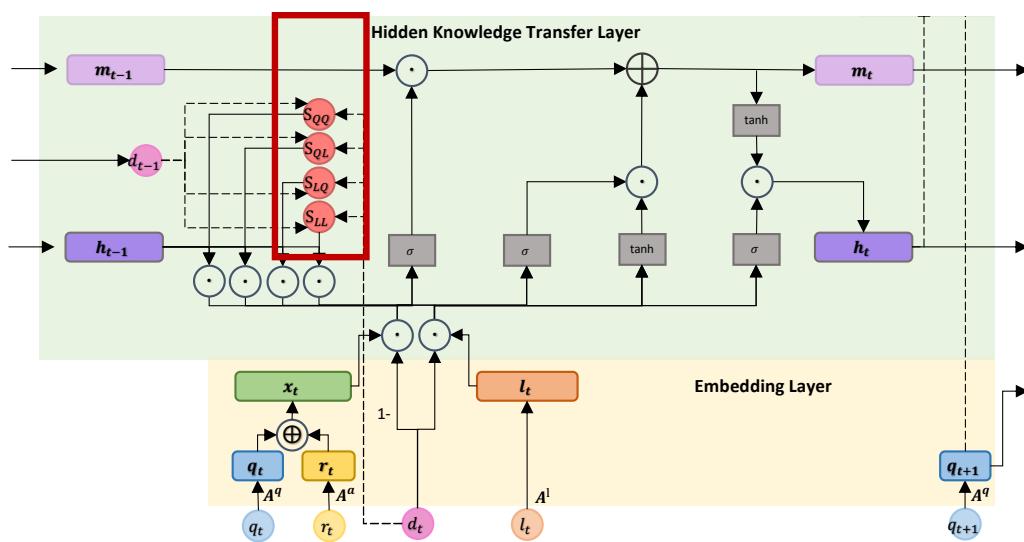
$S_{QQ}$	$(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$	$(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$S_{QL}$	$d_t(1 - d_{t-1})$	$S_{LL}$	$d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\oplus$	: elementwise addition



- Hidden Knowledge Transfer Layer
    - Memory cell
    - Input gate
    - Output gate
    - Forget gate
- Different weight matrices for different transition permutations

# TAMKOT Model

$S_{QQ}$	$(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$	$(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$S_{QL}$	$d_t(1 - d_{t-1})$	$S_{LL}$	$d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\oplus$	: elementwise addition



- Hidden Knowledge Transfer Layer

$d_t$ : material type: 1: assessed, 0: non-assessed

$$S_{QQ} = (1 - d_t)(1 - d_{t-1})$$

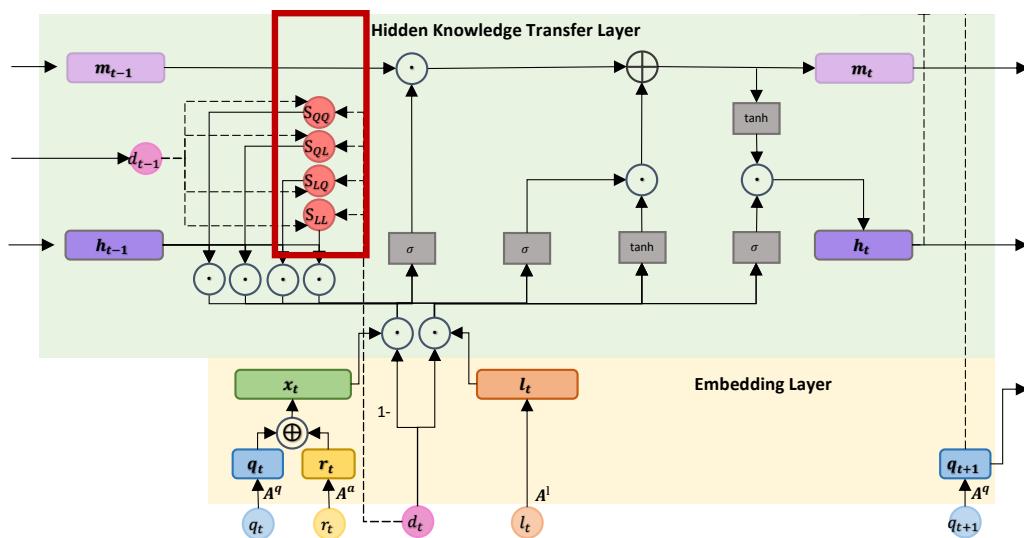
$$S_{QL} = d_t(1 - d_{t-1})$$

$$S_{LQ} = (1 - d_t)d_{t-1}$$

$$S_{LL} = d_t d_{t-1}$$

# TAMKOT Model

$S_{QQ}$	$(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$	$(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$S_{QL}$	$d_t(1 - d_{t-1})$	$S_{LL}$	$d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\oplus$	: elementwise addition



- Hidden Knowledge Transfer Layer

$d_t$ : material type: 1: assessed, 0: non-assessed

$$S_{QQ} = (1 - d_t)(1 - d_{t-1})$$

$$S_{QL} = d_t(1 - d_{t-1})$$

$$S_{LQ} = (1 - d_t)d_{t-1}$$

$$S_{LL} = d_t d_{t-1}$$

Only one of these four indicators equals to 1.

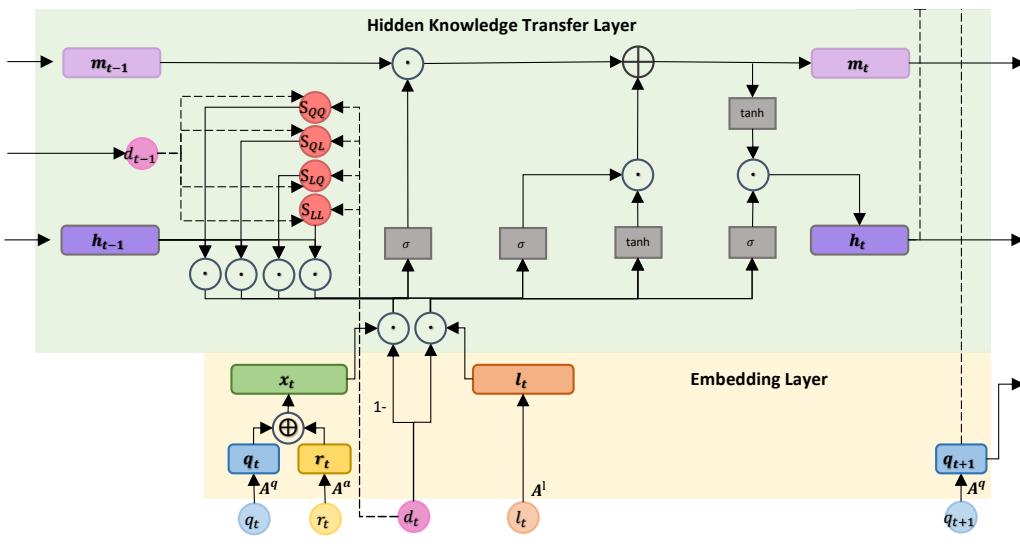
Example:  $d_t = 1, d_{t-1} = 0,$

$$S_{QL} = 1, S_{QQ} = S_{LQ} = S_{LL} = 0$$

# TAMKOT Model



$S_{QQ} : (1 - d_t)(1 - d_{t-1})$	$S_{LQ} : (1 - d_t)d_{t-1}$	$\oplus$ : concatenation
$S_{QL} : d_t(1 - d_{t-1})$	$S_{LL} : d_t d_{t-1}$	$\cdot$ : elementwise product
$\sigma$ : sigmoid function	$\tanh$ : tanh function	$\oplus$ : elementwise addition



- Hidden Knowledge Transfer Layer

$$i_t = \sigma \left( \begin{array}{l} (1 - d_t) \cdot x_t V_{iQ} + d_t \cdot l_t V_{iL} + \\ S_{QQ} \cdot h_{t-1} W_{iQQ} + S_{LL} \cdot h_{t-1} W_{iLL} + \\ S_{QL} \cdot h_{t-1} W_{iQL} + S_{LQ} \cdot h_{t-1} W_{iLQ} \end{array} \right)$$

$$g_t = \tanh \left( \begin{array}{l} (1 - d_t) \cdot x_t V_{gQ} + d_t \cdot l_t V_{gL} + \\ S_{QQ} \cdot h_{t-1} W_{gQQ} + S_{LL} \cdot h_{t-1} W_{gLL} + \\ S_{QL} \cdot h_{t-1} W_{gQL} + S_{LQ} \cdot h_{t-1} W_{gLQ} \end{array} \right)$$

$$f_t = \sigma \left( \begin{array}{l} (1 - d_t) \cdot x_t V_{fQ} + d_t \cdot l_t V_{fL} + \\ S_{QQ} \cdot h_{t-1} W_{fQQ} + S_{LL} \cdot h_{t-1} W_{fLL} + \\ S_{QL} \cdot h_{t-1} W_{fQL} + S_{LQ} \cdot h_{t-1} W_{fLQ} \end{array} \right)$$

$$o_t = \sigma \left( \begin{array}{l} (1 - d_t) \cdot x_t V_{oQ} + d_t \cdot l_t V_{oL} + \\ S_{QQ} \cdot h_{t-1} W_{oQQ} + S_{LL} \cdot h_{t-1} W_{oLL} + \\ S_{QL} \cdot h_{t-1} W_{oQL} + S_{LQ} \cdot h_{t-1} W_{oLQ} \end{array} \right)$$

$$m_t = f_t m_{t-1} + i_t g_t$$

$$h_t = o_t \tanh(m_t)$$

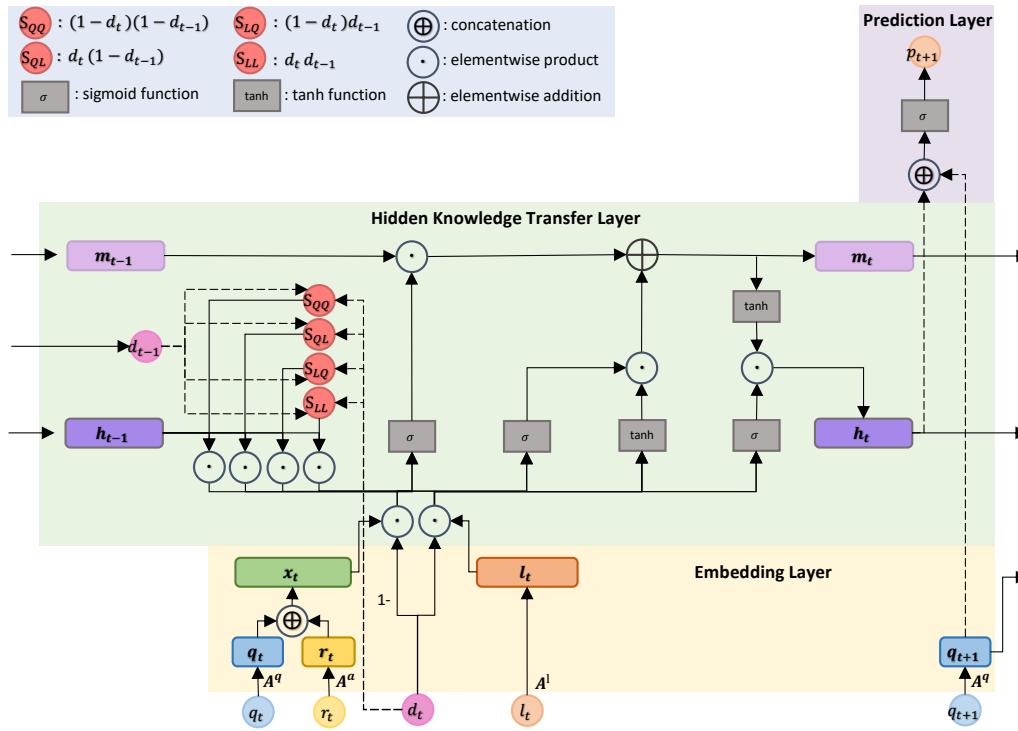
Only one is activated.

Example:  $S_{QL} = 1, S_{QQ} = S_{LQ} = S_{LL} = 0$

$$f_t = d_t \cdot l_t V_{fL} + S_{QL} \cdot h_{t-1} W_{fQL}$$

# TAMKOT Model

$S_{QQ}$ : $(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$ : $(1 - d_t)d_{t-1}$	$\oplus$ : concatenation
$S_{QL}$ : $d_t(1 - d_{t-1})$	$S_{LL}$ : $d_t d_{t-1}$	$\odot$ : elementwise product
$\sigma$ : sigmoid function	$\tanh$ : tanh function	$\oplus$ : elementwise addition



- **Prediction Layer**
  - Predict student performance of  $q_{t+1}$
- **Hidden Knowledge Transfer Layer**
  - Represent student knowledge
- **Embedding layer**
  - Map learning materials and student performance into latent concept space

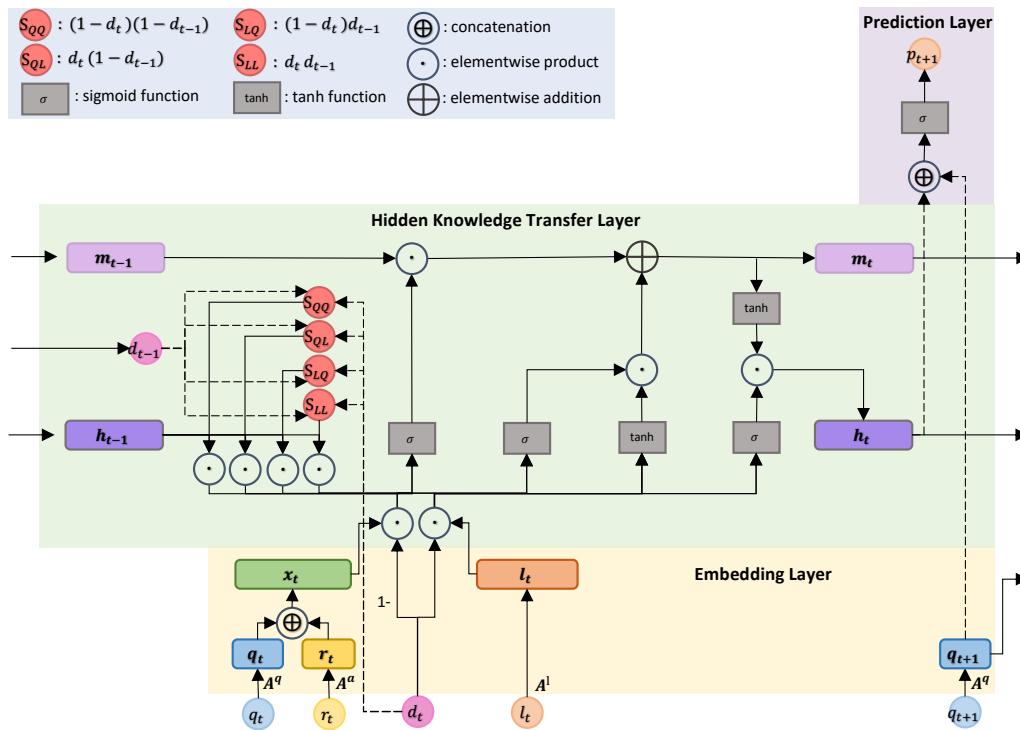
# TAMKOT Model



UNIVERSITY  
AT ALBANY

State University of New York

$S_{QQ}$	$(1 - d_t)(1 - d_{t-1})$	$S_{LQ}$	$(1 - d_t)d_{t-1}$	$\oplus$	: concatenation
$S_{QL}$	$d_t(1 - d_{t-1})$	$S_{LL}$	$d_t d_{t-1}$	$\cdot$	: elementwise product
$\sigma$	: sigmoid function	$\tanh$	: tanh function	$\oplus$	: elementwise addition



- **Prediction Layer**
  - Predict student performance of  $q_{t+1}$

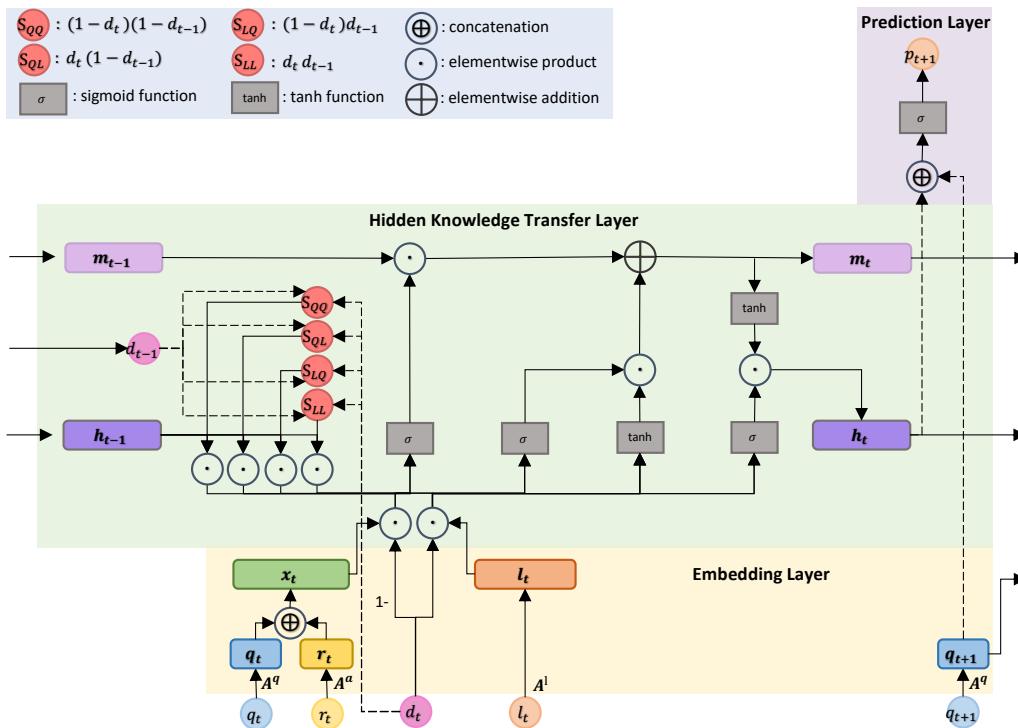
$$p_{t+1} = \sigma(\mathbf{W}_P^T[\mathbf{h}_t \oplus \mathbf{q}_{t+1}] + b_p)$$

# TAMKOT Model



UNIVERSITY  
AT ALBANY  
State University of New York

$S_{QQ} : (1 - d_t)(1 - d_{t-1})$	$S_{LQ} : (1 - d_t)d_{t-1}$	$\oplus$ : concatenation
$S_{QL} : d_t(1 - d_{t-1})$	$S_{LL} : d_t d_{t-1}$	$\cdot$ : elementwise product
$\sigma$ : sigmoid function	$\tanh$ : tanh function	$\oplus$ : elementwise addition



- **Prediction Layer**
  - Predict student performance of  $q_{t+1}$
- **Hidden Knowledge Transfer Layer**
  - Represent student knowledge
- **Embedding layer**
  - Map learning materials and student performance into latent concept space

Objective function:

$$\mathcal{L} = \sum_t (r_t \log p_t + (1 - r_t) \log(1 - p_t)) + \lambda_\theta \|\theta\|^2$$

regularization

# Experiments



- We evaluate our model with three sets of experiments:
  -  **Student Performance Prediction** - to validate if the model captures the variability of student performance
  -  **Knowledge Transfer Analysis** - to analyze the knowledge transfer between assessed and non-assessed learning material types
  -  **Student Knowledge State Visualization** - to check if our model represents student knowledge state



# Datasets & Experiment Setup

- Datasets

- MORF: Assignments and lecture
- EdNet: Questions and question explanations
- Junyi: Problems and hints

Dataset	#Users	#Questions	Question Records	Question Responses Mean	Question Responses STD	#Correct Question Responses	#Incorrect Question Responses	#Non-assessed materials	#Non-assessed Records
MORF	686	10	12031	0.7763	0.2507	N/A	N/A	52	41980
EdNet	1000	11249	200931	0.5910	0.2417	118747	82184	8324	150821
Junyi	2063	3760	290754	0.6660	0.2224	193664	97090	1432	69050

Statistics for each datasets





# Datasets & Experiment Setup

- Datasets

- MORF: Assignments and lecture
- EdNet: Questions and question explanations
- Junyi: Problems and hints

- Experiment Setup

- 5-fold student stratified cross-validation
  - Training set: 80% student sequences (20% sequences of training)
  - Testing set: 20% student sequences

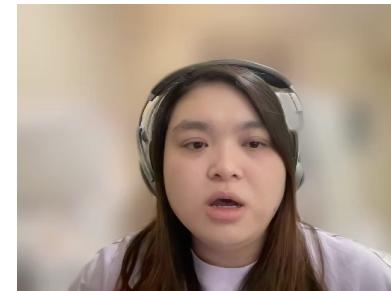
Dataset	#Users	#Questions	Question Records	Question Responses Mean	Question Responses STD	#Correct Question Responses	#Incorrect Question Responses	#Non-assessed materials	#Non-assessed Records
MORF	686	10	12031	0.7763	0.2507	N/A	N/A	52	41980
EdNet	1000	11249	200931	0.5910	0.2417	118747	82184	8324	150821
Junyi	2063	3760	290754	0.6660	0.2224	193664	97090	1432	69050

Statistics for each datasets





# Baselines & Metrics



- Baselines

- Single-activity: **DKT** [Piech et al.], **DKVMN** [Zhang et al.], **SAINT** [Choi et al.], **SAKT** [Pandey et al.], **AKT** [Ghosh et al.], **DeepIRT** [Yeung et al.]
- Multi-activity: **MVKM** [Zhao et al.], **DMKT** [Wang et al.],
- [method name]+MV: Baselines ‘Multi-activity’ setting
- **MLP+M**: simple multi-layer perceptron consider three recent assessed and three non-assessed activities as input



# Baselines & Metrics



- Baselines
  - Single-activity: **DKT** [Piech et al.], **DKVMN** [Zhang et al.], **SAINT** [Choi et al.], **SAKT** [Pandey et al.], **AKT** [Ghosh et al.], **DeepIRT** [Yeung et al.]
  - Multi-activity: **MVKM** [Zhao et al.], **DMKT** [Wang et al.],
  - [method name]+MV: Baselines ‘Multi-activity’ setting
  - **MLP+M**: simple multi-layer perceptron consider three recent assessed and three non-assessed activities as input
- Metrics
  - Area Under Curve (AUC)
  - Root Mean Squared Error (RMSE)



# Performance Prediction



Methods	MORF	EdNet	Junyi
	RMSE	AUC	AUC
DKT	0.1938*	0.6393**	0.8623**
DKVMN	0.2043**	0.6296**	0.8558**
SAKT	0.2113**	0.6334**	0.8053**
SAINT	0.2019**	0.5205**	0.7951**
AKT	0.2420**	0.6393**	0.8093**
DeepIRT	0.1946**	0.6290**	0.8498**
DKT+M	0.1928	0.6372**	<u>0.8652*</u>
DKVMN+M	0.2251**	0.6343**	0.8513**
SAKT+M	0.2085**	0.6323**	0.7911**
SAINT+M	0.1977**	0.5491**	0.7741**
AKT+M	0.2240**	<u>0.6404**</u>	0.8099**
MLP+M	0.2433**	0.6102**	0.7290**
MVKM	0.1936*	-	-
DMKT	<b>0.1754**</b>	0.6394**	0.8561**
TAMKOT	<u>0.1871</u>	<b>0.6786</b>	<b>0.8745</b>

Performance Prediction results, \*\* and \* indicate  
t-test p – value < 0.05 and p – value < 0.1



# Does Multi-Activity help improve prediction performance?

Methods	MORF	EdNet	Junyi
	RMSE	AUC	AUC
DKT	0.1938*	0.6393**	0.8623**
DKVMN	0.2043**	0.6296**	0.8558**
SAKT	0.2113**	0.6334**	0.8053**
SAINT	0.2019**	0.5205**	0.7951**
AKT	0.2420**	0.6393**	0.8093**
DeepIRT	0.1946**	0.6290**	0.8498**
DKT+M	0.1928	0.6372**	<u>0.8652*</u>
DKVMN+M	0.2251**	0.6343**	0.8513**
SAKT+M	0.2085**	0.6323**	0.7911**
SAINT+M	0.1977**	0.5491**	0.7741**
AKT+M	0.2240**	<u>0.6404**</u>	0.8099**
MLP+M	0.2433**	0.6102**	0.7290**
MVKM	0.1936*	-	-
DMKT	<b>0.1754**</b>	0.6394**	0.8561**
<b>TAMKOT</b>	<b>0.1871</b>	<b>0.6786</b>	<b>0.8745</b>

Performance Prediction results, \*\* and \* indicate  
t-test p – value < 0.05 and p – value < 0.1



# Does Multi-Activity help improve prediction performance? Yes!

Methods	MORF	EdNet	Junyi
	RMSE	AUC	AUC
DKT	0.1938*	0.6393**	0.8623**
DKVMN	0.2043**	0.6296**	0.8558**
SAKT	0.2113**	0.6334**	0.8053**
SAINT	0.2019**	0.5205**	0.7951**
AKT	0.2420**	0.6393**	0.8093**
DeepIRT	0.1946**	0.6290**	0.8498**
DKT+M	0.1928	0.6372**	<u>0.8652*</u>
DKVMN+M	0.2251**	0.6343**	0.8513**
SAKT+M	0.2085**	0.6323**	0.7911**
SAINT+M	0.1977**	0.5491**	0.7741**
AKT+M	0.2240**	<u>0.6404**</u>	0.8099**
MLP+M	0.2433**	0.6102**	0.7290**
MVKM	0.1936*	-	-
DMKT	<b>0.1754**</b>	0.6394**	0.8561**
<b>TAMKOT</b>	<b>0.1871</b>	<b>0.6786</b>	<b>0.8745</b>

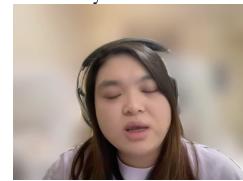
Performance Prediction results, \*\* and \* indicate  
t-test p – value < 0.05 and p – value < 0.1



# Does TAMKOT outperform Multi-Activity baselines?

Methods	MORF	EdNet	Junyi
	RMSE	AUC	AUC
DKT	0.1938*	0.6393**	0.8623**
DKVMN	0.2043**	0.6296**	0.8558**
SAKT	0.2113**	0.6334**	0.8053**
SAINT	0.2019**	0.5205**	0.7951**
AKT	0.2420**	0.6393**	0.8093**
DeepIRT	0.1946**	0.6290**	0.8498**
DKT+M	0.1928	0.6372**	<u>0.8652*</u>
DKVMN+M	0.2251**	0.6343**	0.8513**
SAKT+M	0.2085**	0.6323**	0.7911**
SAINT+M	0.1977**	0.5491**	0.7741**
AKT+M	0.2240**	<u>0.6404**</u>	0.8099**
MLP+M	0.2433**	0.6102**	0.7290**
MVKM	0.1936*	-	-
DMKT	<b>0.1754**</b>	0.6394**	0.8561**
<b>TAMKOT</b>	<b>0.1871</b>	<b>0.6786</b>	<b>0.8745</b>

Performance Prediction results, \*\* and \* indicate  
t-test p – value < 0.05 and p – value < 0.1



- Findings:

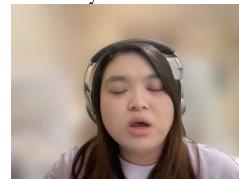
- TAMKOT significantly outperforms all the single activity models
- TAMKOT significantly outperforms all the ‘Multi-activity’ setting models
- TAMKOT significantly outperforms all the multi-activity baseline methods, except DMKT on MORF



# Does TAMKOT outperform Multi-Activity baselines?

Methods	MORF	EdNet	Junyi
	RMSE	AUC	AUC
DKT	0.1938*	0.6393**	0.8623**
DKVMN	0.2043**	0.6296**	0.8558**
SAKT	0.2113**	0.6334**	0.8053**
SAINT	0.2019**	0.5205**	0.7951**
AKT	0.2420**	0.6393**	0.8093**
DeepIRT	0.1946**	0.6290**	0.8498**
DKT+M	0.1928	0.6372**	<u>0.8652*</u>
DKVMN+M	0.2251**	0.6343**	0.8513**
SAKT+M	0.2085**	0.6323**	0.7911**
SAINT+M	0.1977**	0.5491**	0.7741**
AKT+M	0.2240**	<u>0.6404**</u>	0.8099**
MLP+M	0.2433**	0.6102**	0.7290**
MVKM	0.1936*	-	-
DMKT	<b>0.1754**</b>	0.6394**	0.8561**
TAMKOT	0.1871	<b>0.6786</b>	<b>0.8745</b>

Performance Prediction results, \*\* and \* indicate  
t-test p – value < 0.05 and p – value < 0.1



## Findings:

- TAMKOT significantly outperforms all the single activity models
- TAMKOT significantly outperforms all the ‘Multi-activity’ setting models
- TAMKOT significantly outperforms all the multi-activity baseline methods, except DMKT on MORF
  - DMKT’s complexity with MORF’s learning material complexity



Does the knowledge transfer from question to lecture is similar as from lecture to question?

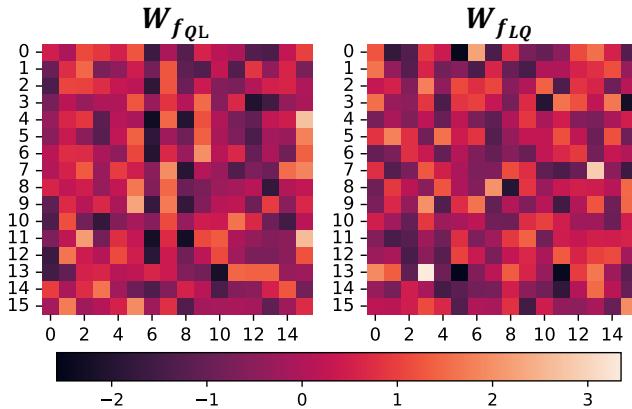
- Compared weight matrices  $W_{f_{QL}}$  and  $W_{f_{LQ}}$  of forget gate
  - Spearman correlation
  - Visualization

	MORF	EdNet	Junyi
Correlation	-0.03686	0.33680	0.38443
p-value	0.55714	3.30e-08	2.09e-37

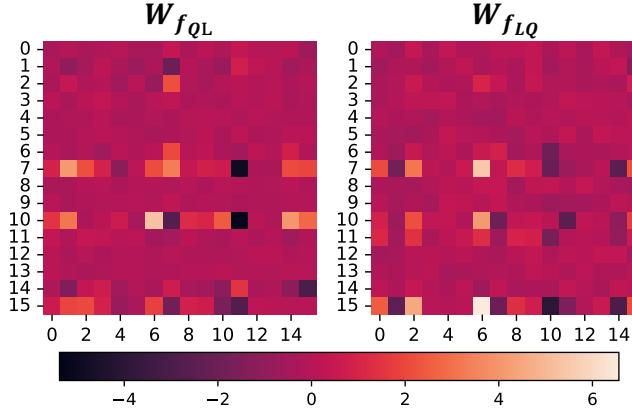
Spearman correlation coefficients with p-values



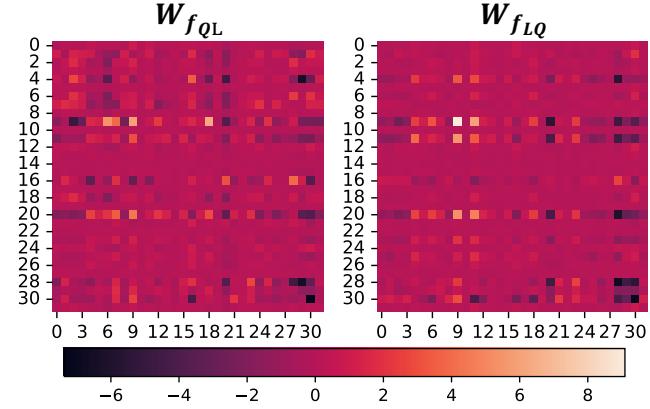
Does the knowledge transfer from question to lecture is similar as from lecture to question?



Forget gate weight matrices for MORF



Forget gate weight matrices for EdNet



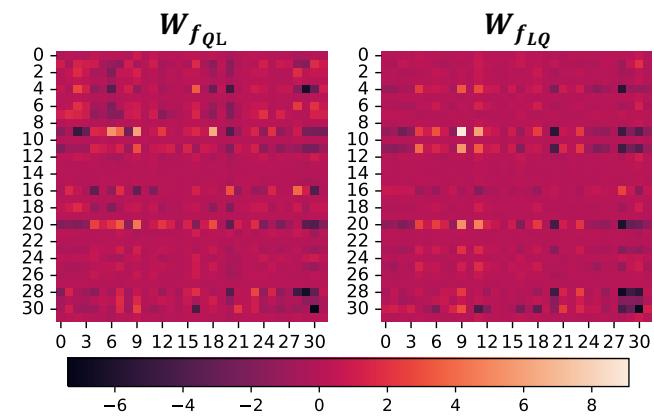
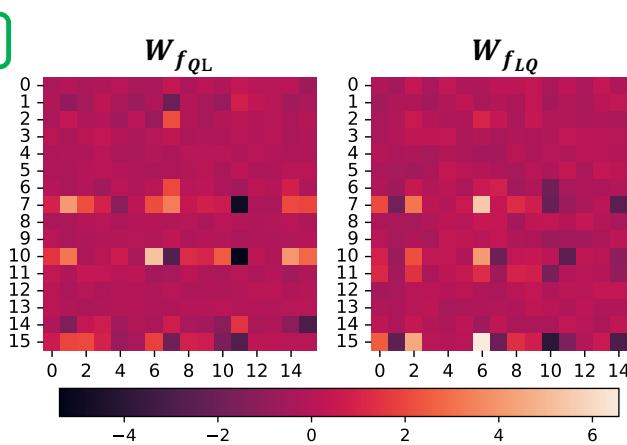
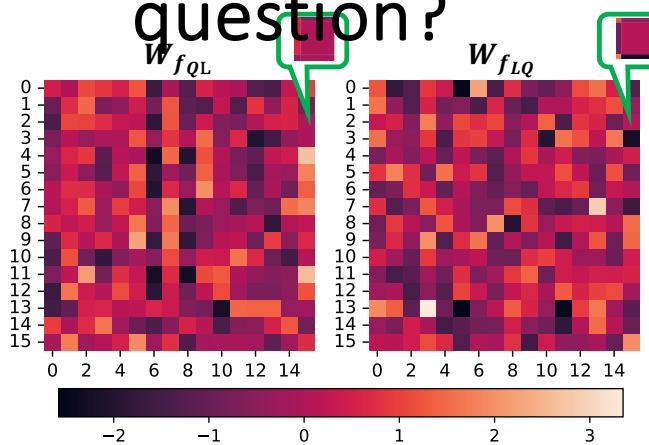
Forget gate weight matrices for Junyi

- Findings:
  - MORF: knowledge transfer is generally different with some are the same

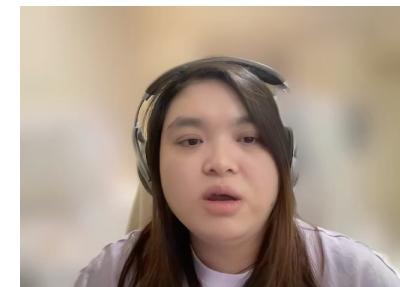




Does the knowledge transfer from question to lecture is similar as from lecture to question?

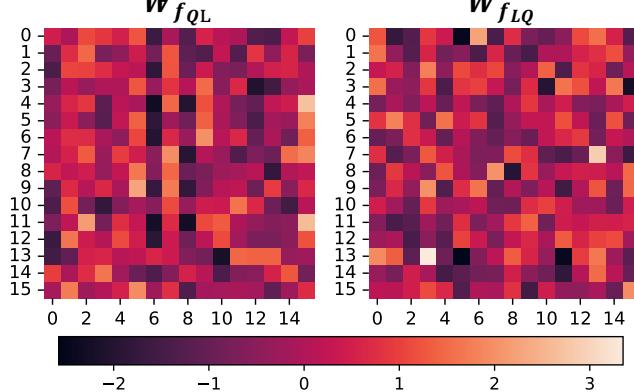


- Findings:
  - MORF: knowledge transfer is generally different with some are the same

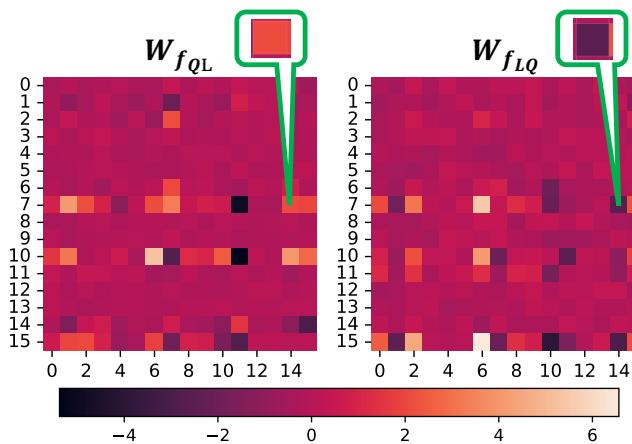




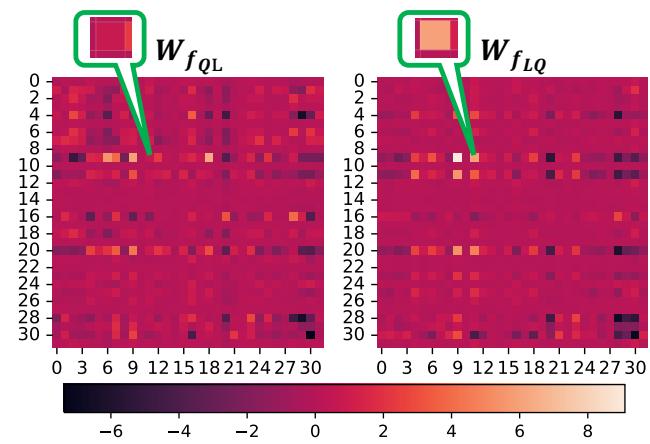
Does the knowledge transfer from question to lecture is similar as from lecture to question?



Forget gate weight matrices for MORF



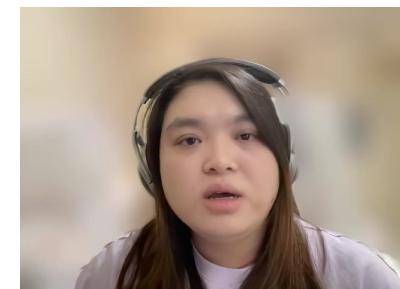
Forget gate weight matrices for EdNet



Forget gate weight matrices for Junyi

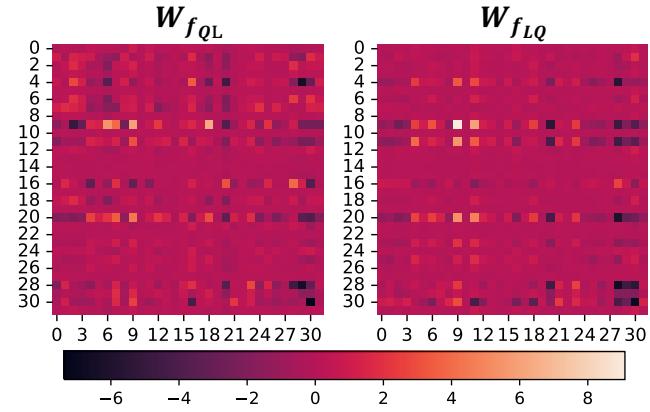
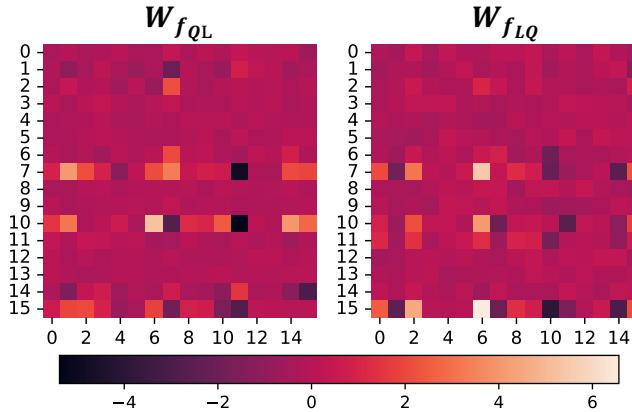
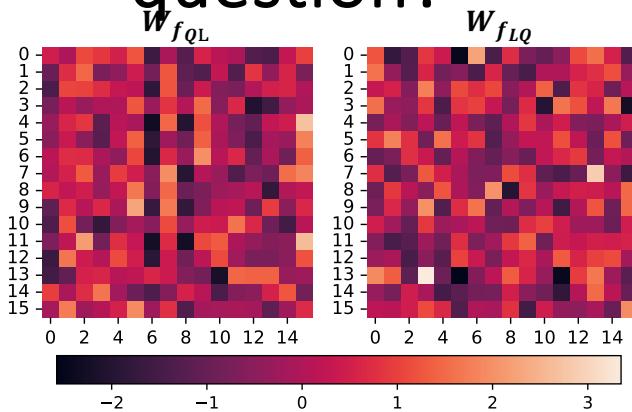
- Findings:

- MORF: knowledge transfer is generally different with some are the same
- Ednet and Junyi: Knowledge transfer is similar with exist difference





Does the knowledge transfer from question to lecture is similar as from lecture to question?



- Findings :
  - MORF: knowledge transfer is generally different with some are the same
  - Ednet and Junyi: Knowledge transfer is similar with exist difference
- Reason:
  - Close-knit associations between material types in Junyi and EdNet
  - Each assignment includes multiple problems in MORF

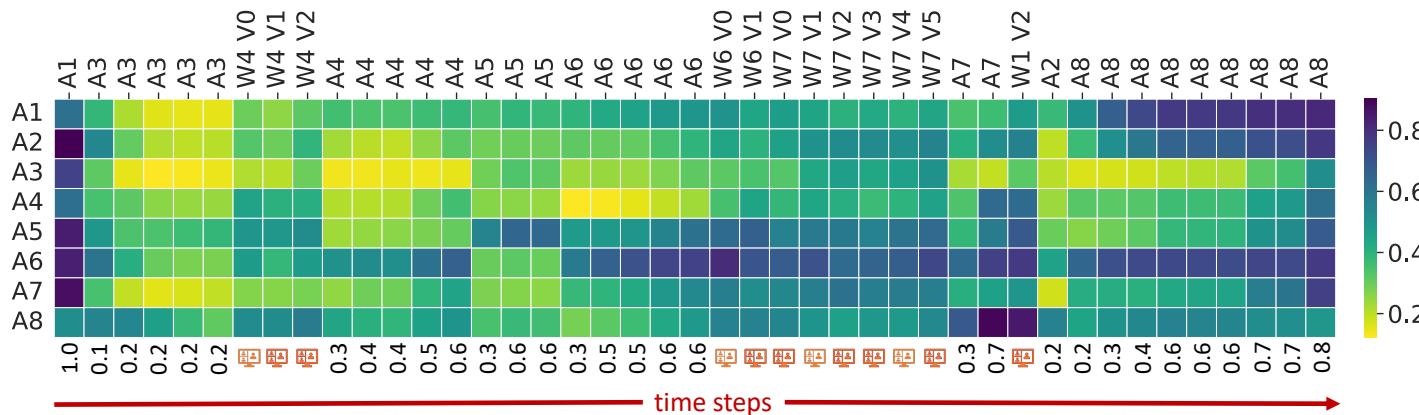




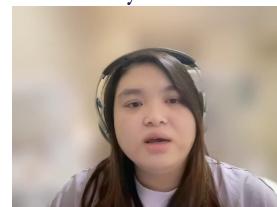
# Student Knowledge State Visualization



- Predicted performance in each assignment

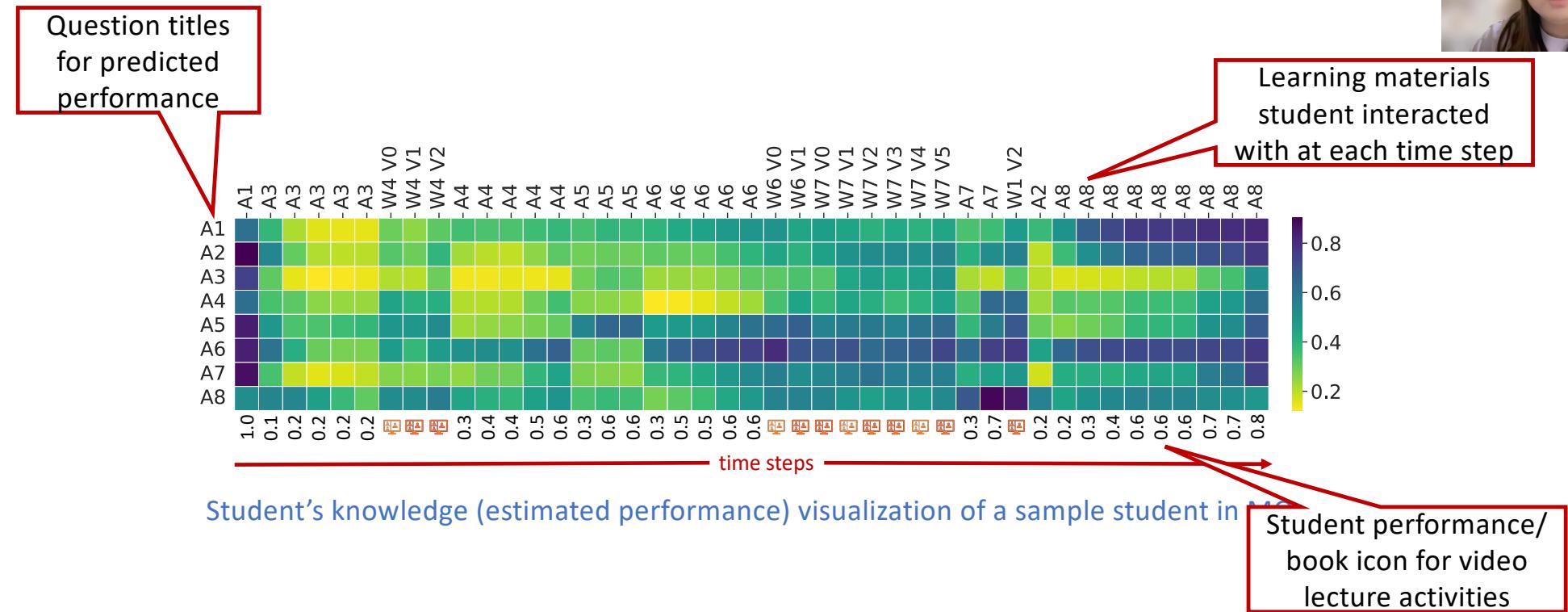


Student's knowledge (estimated performance) visualization of a sample student in MORF



# Student Knowledge State Visualization

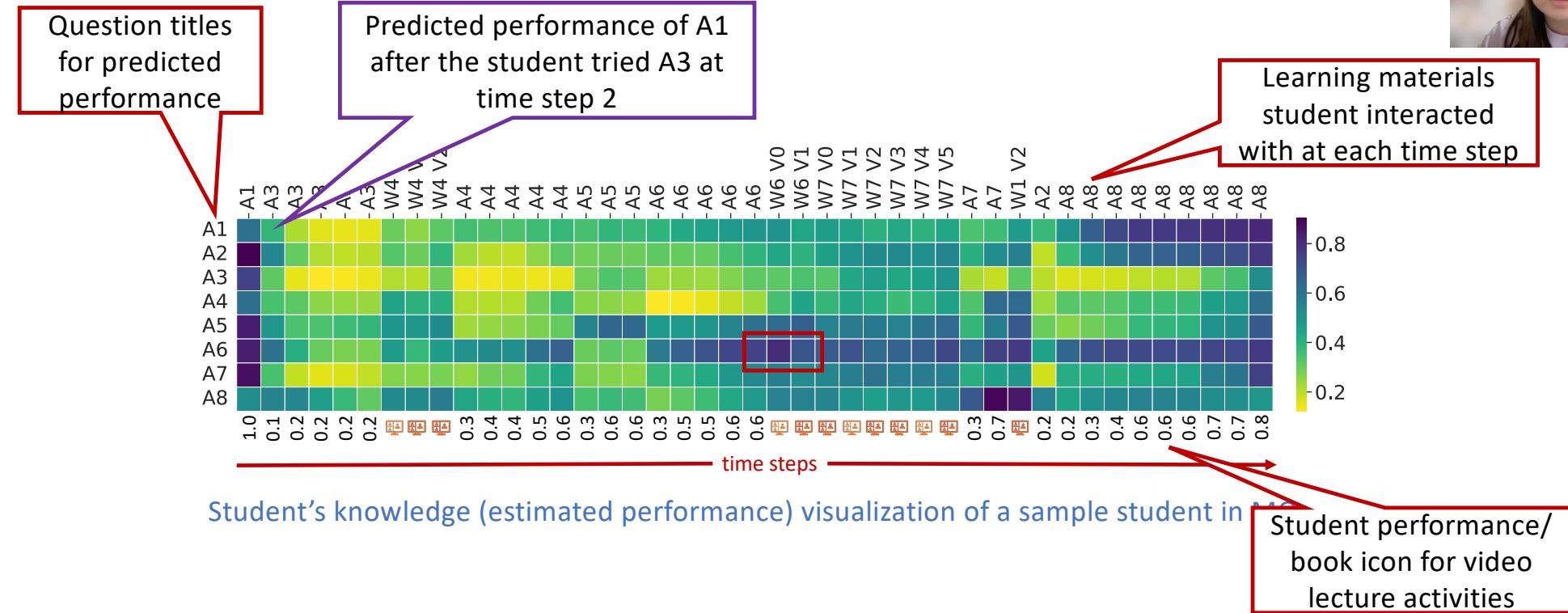
- Predicted performance in each assignment





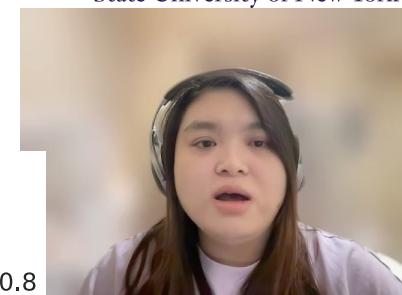
# Student Knowledge State Visualization

- Predicted performance in each assignment

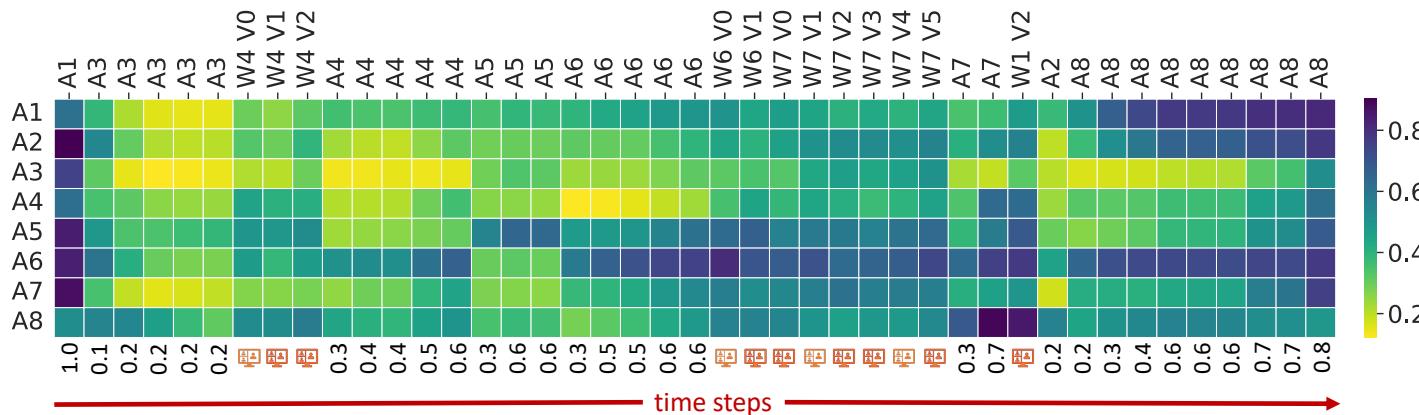




# Student Knowledge State Visualization



- Predicted performance in each assignment

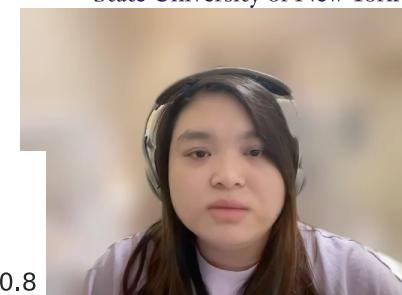


Student's knowledge (estimated performance) visualization of a sample student in MORF

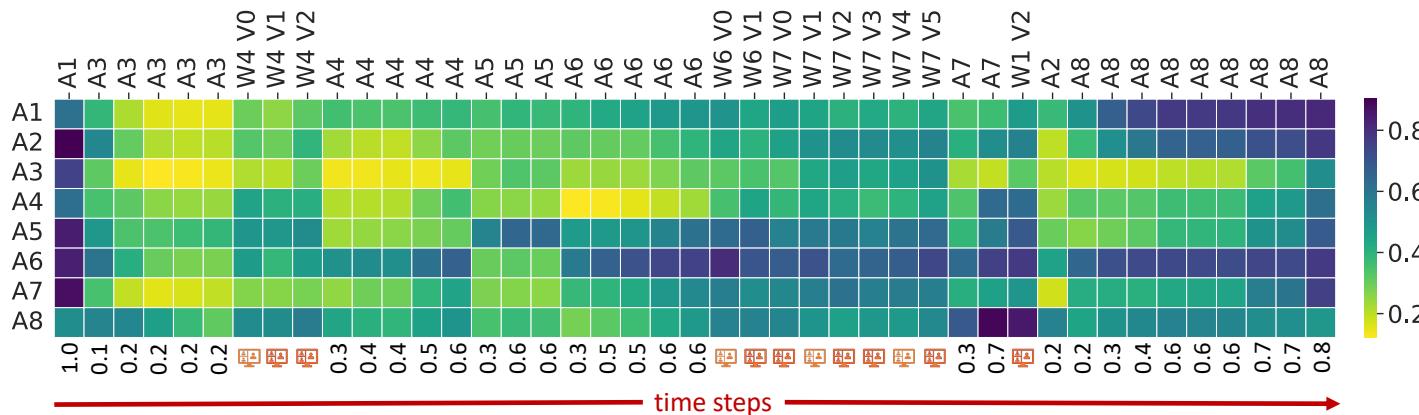
- Findings
    - Videos produce knowledge improvement for corresponding assignment



# Student Knowledge State Visualization



- Predicted performance in each assignment



Student's knowledge (estimated performance) visualization of a sample student in MORF

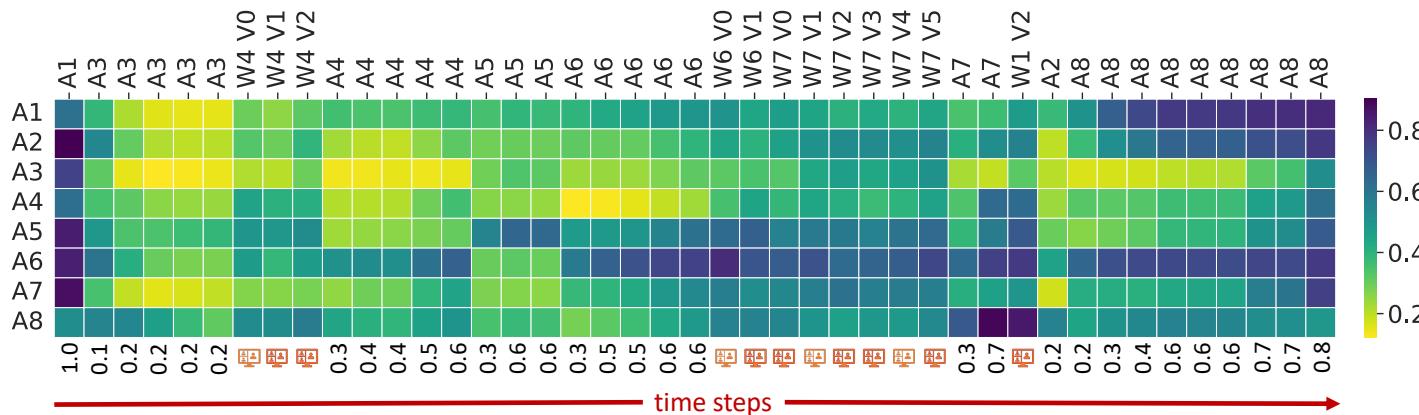
- Findings
    - Videos produce knowledge improvement for corresponding assignment



# Student Knowledge State Visualization



- Predicted performance in each assignment



Student's knowledge (estimated performance) visualization of a sample student in MORF

- Findings
    - Videos produce knowledge improvement for corresponding assignments
    - First attempt of video lecture has the largest improvement

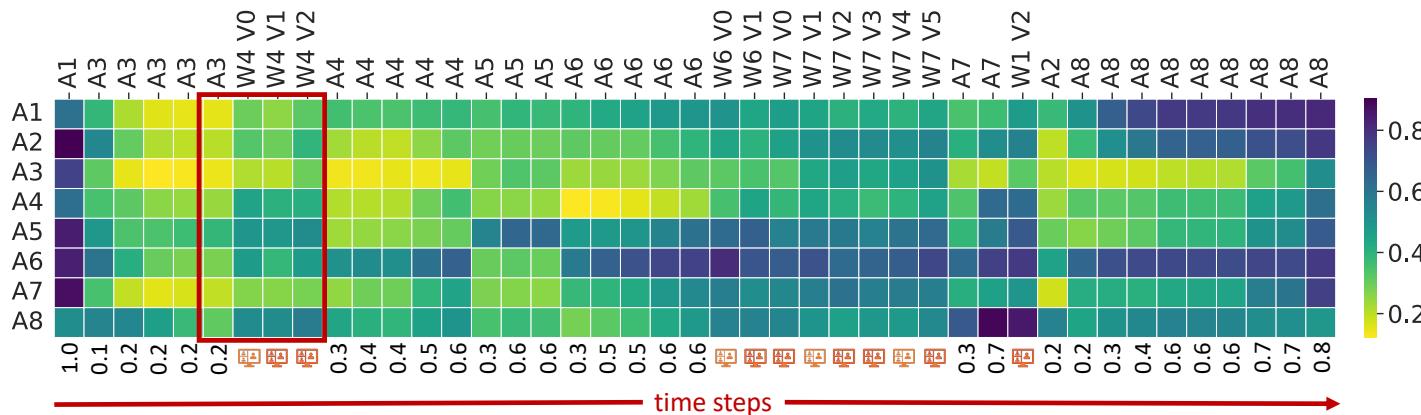


# Student Knowledge State Visualization



UNIVERSITY  
AT ALBANY  
State University of New York

- Predicted performance in each assignment



Student's knowledge (estimated performance) visualization of a sample student in MORF

- Findings
    - Videos produce knowledge improvement for corresponding assignments
    - First attempt of video lecture has the largest improvement

# Conclusions



- We proposed a Transition-Aware Multi-Activity Knowledge Tracing Model (TAMKOT)
  - Can accurately represent student knowledge and predict their performance
  - Can explicitly learn the knowledge transfer between different learning activity types.
- The knowledge transfer amount depend on the transition order
- For some students, the assessed activities were more helpful



This paper is based upon work supported  
by the National Science Foundation under  
Grant No. 2047500.

# Thank you! Q & A



Our code are available at GitHub:  
<https://github.com/persai-lab/BigData2022-TAMKOT>



[szhao2@albany.edu](mailto:szhao2@albany.edu)

[cwang25@albany.edu](mailto:cwang25@albany.edu)

[ssahebi@albany.edu](mailto:ssahebi@albany.edu)