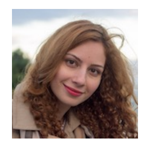
AAAI-25

Neighborhood-Aware Negative Sampling for Student Knowledge and Behavior Modeling

Siqian Zhao

szhao2@albany.edu

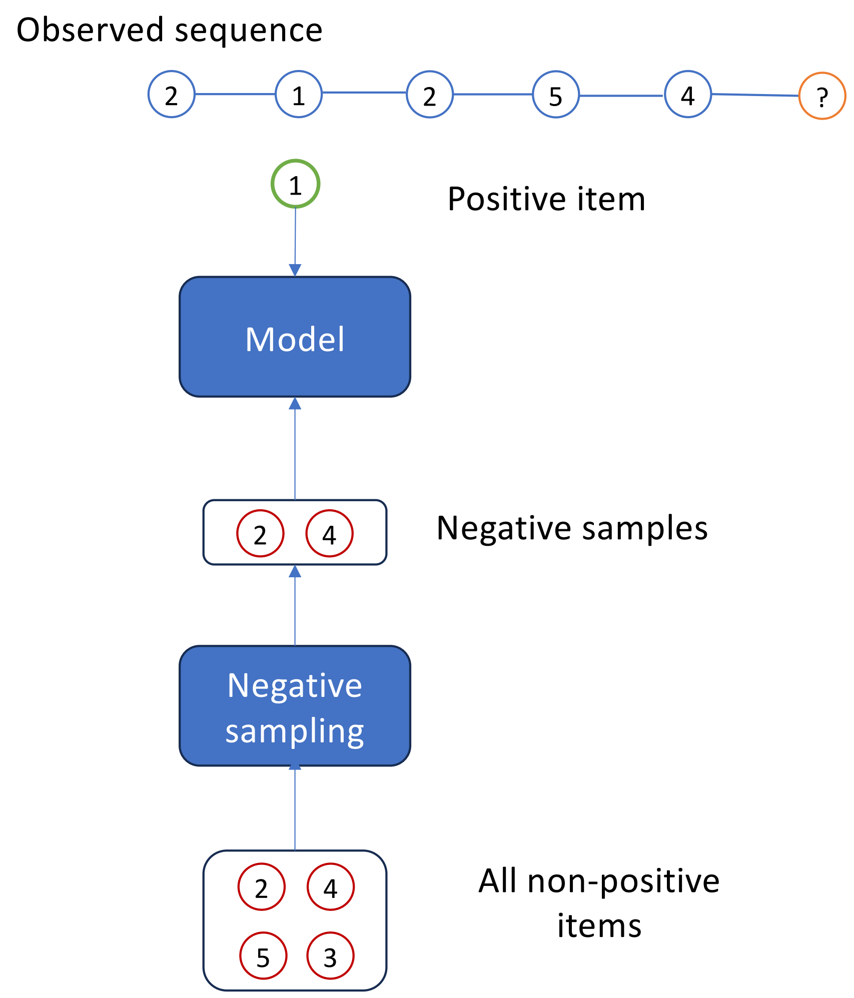


Sherry Sahebi

ssahebi@albany.edu

General problem

- Negative sampling
  - Distinguish informative vs. non-informative samples
  - To predict the next sequence item with complex patterns and interrelations
  - Using coarse-grained sequence information



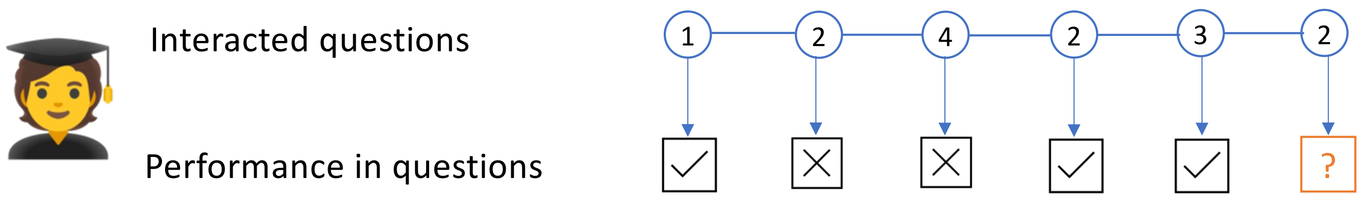
Application Context: Student Modeling

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

- Essential EDM problems
  - Student Knowledge Tracing (KT)
  - Student Behavior Modeling (BM)

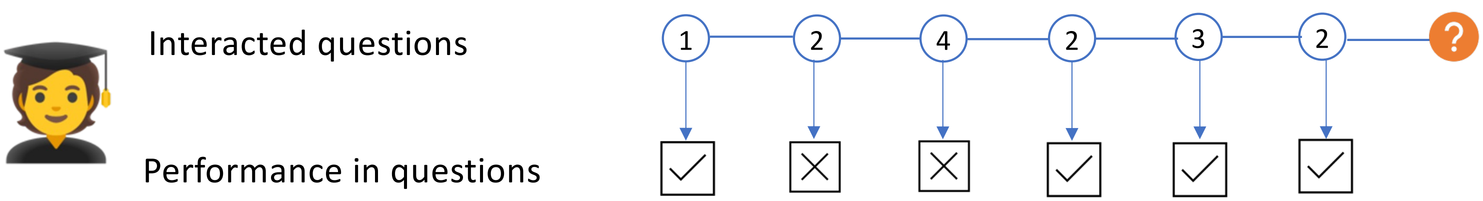
Student Knowledge Tracing (KT)

- Objectives
  - Given observed student history and performance
  - Quantify student knowledge level while learning
  - Predict students' future performance in a given question



Student Behavior Modeling (BM)

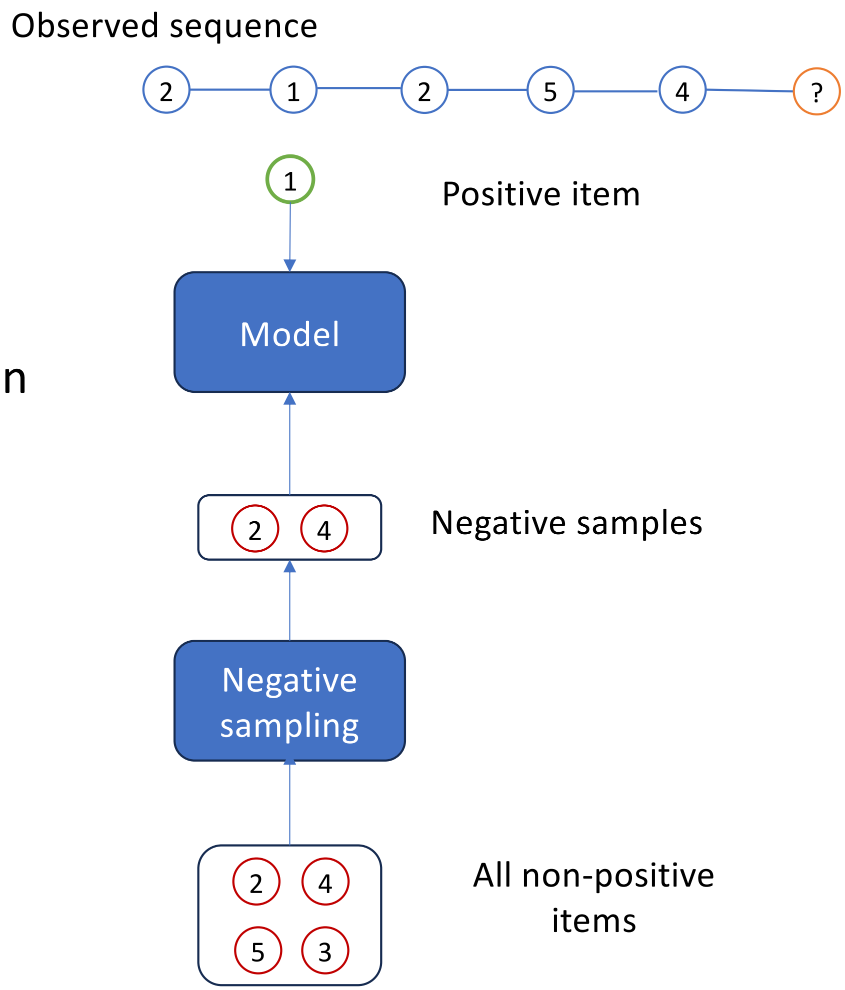
- Understanding students' behavior patterns during the learning process
  - Engagement
  - Procrastination
  - Preference for learning materials
    - Predict the next question student chooses to interact with, given history
    - Many questions to choose from → needs negative sampling



Challenge

- How to use sequence information in selecting informative negative samples
  - Using coarse-grained sequence information

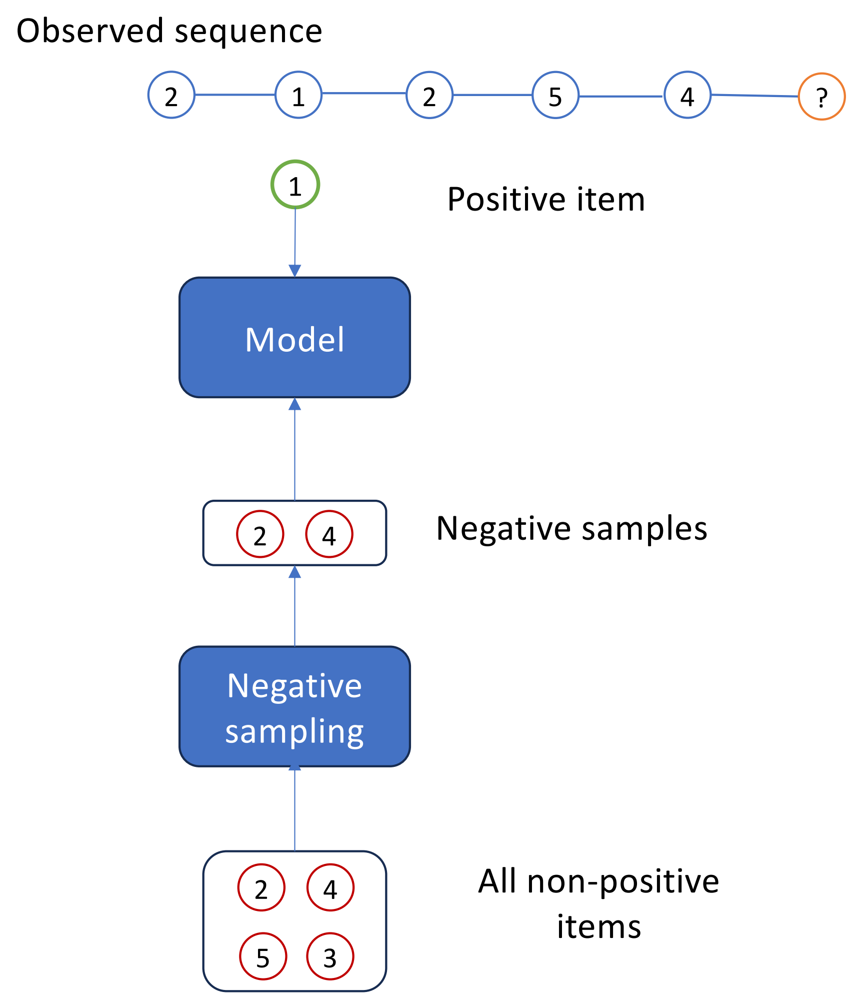
- While learning both student knowledge (KM) and Behavior (BM)
  - Using fine-grained sequence information



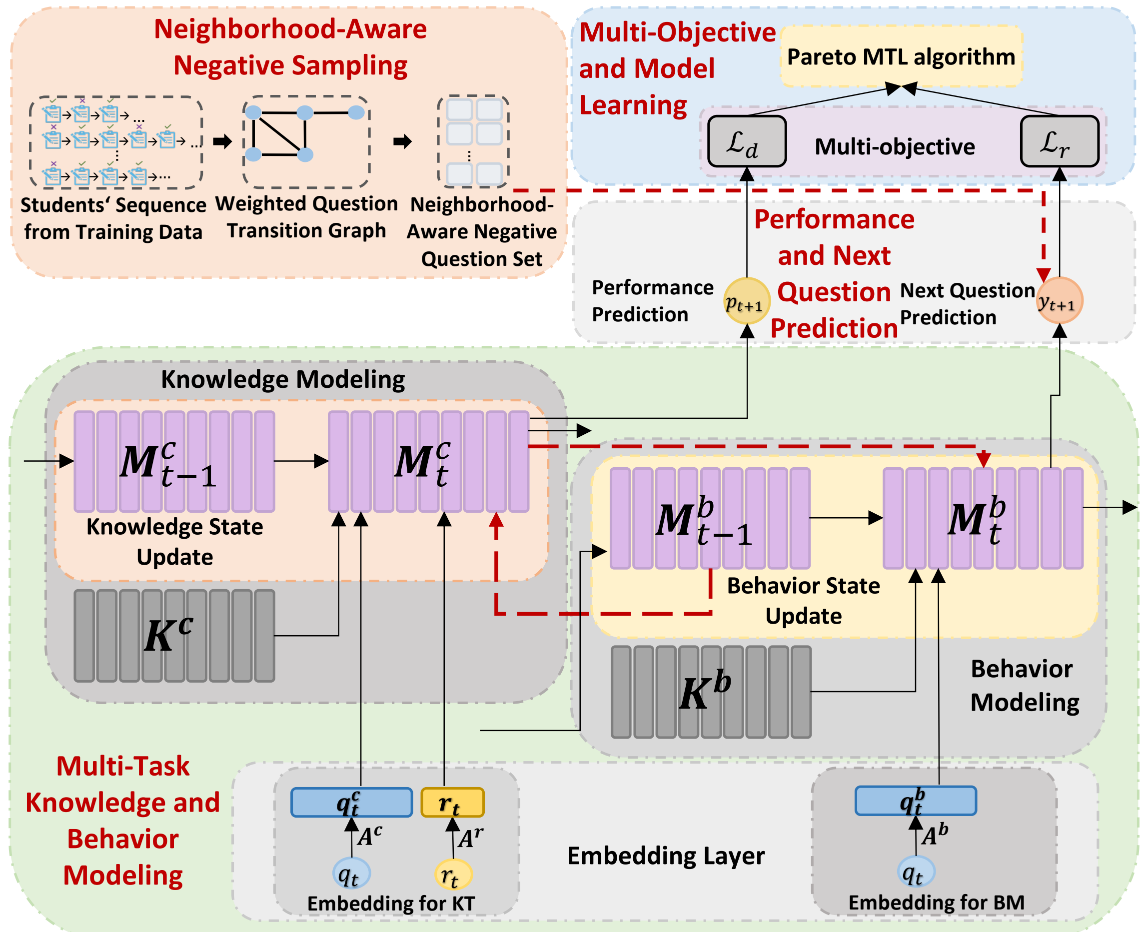
Our solution

Multi-objective multi-task key-value memory networks (KoBeM)

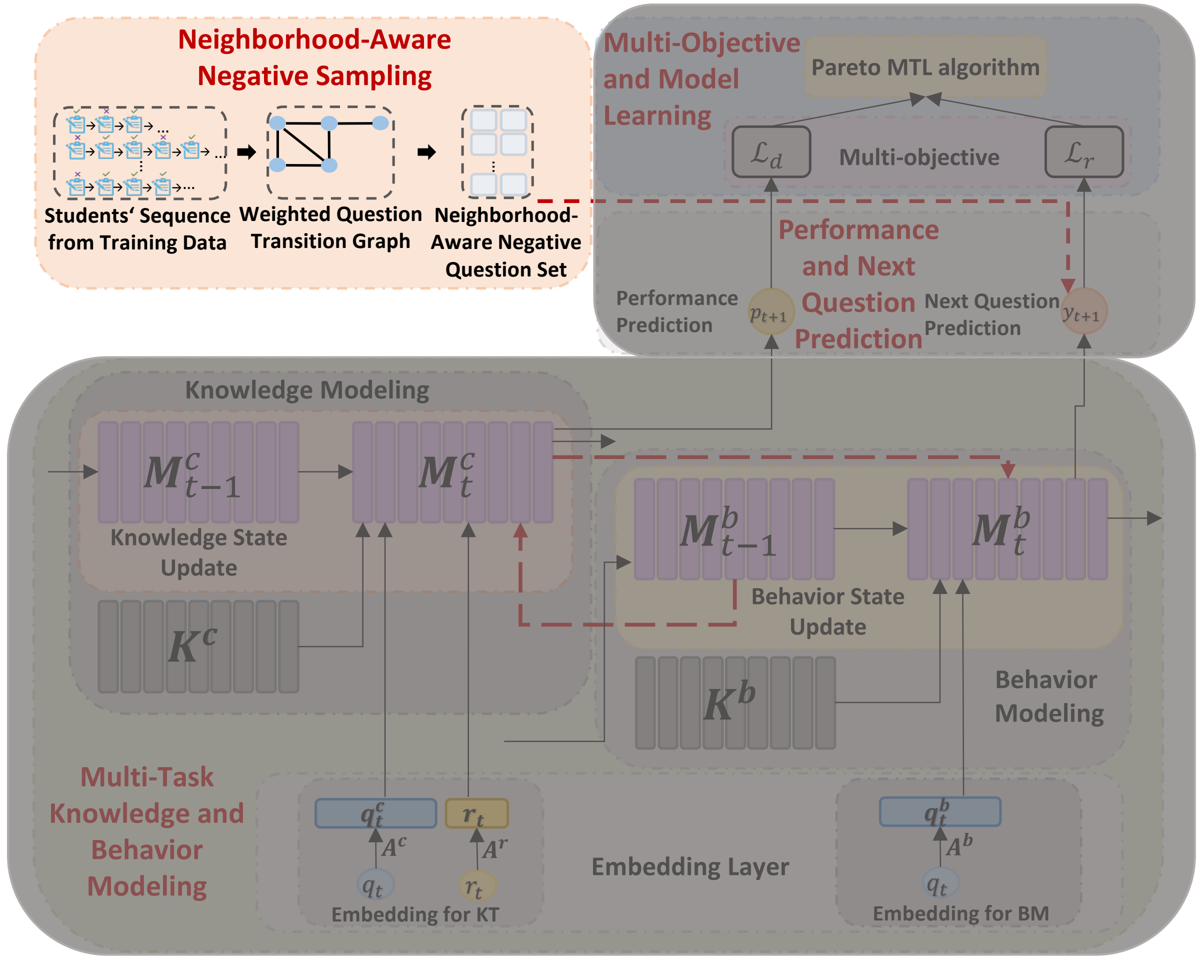
Neighborhood-aware negative sampling (NANS)



Neighborhood-Aware Negative Sampling (NANS) with Knowledge and Behavior Modeling (KoBeM)

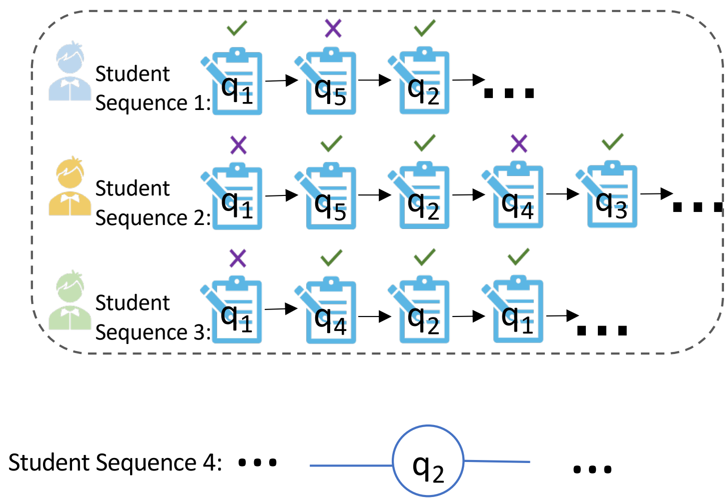


Neighborhood-Aware Negative Sampling (NANS)



Neighborhood-Aware Negative Sampling (NANS)

- Out of many similar questions, the student chooses one positive item
- Using “similar” questions as the difficult negative candidate set

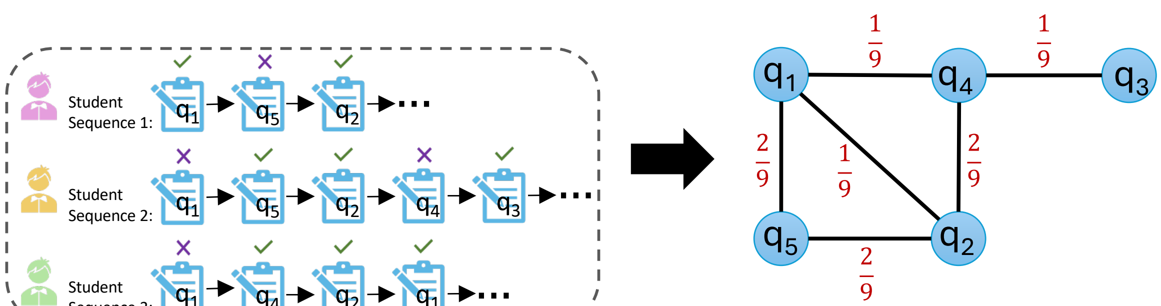


Most similar questions to q_2

$$\{ q_4 , q_5 \}$$

Neighborhood-Aware Negative Sampling (NANS) -

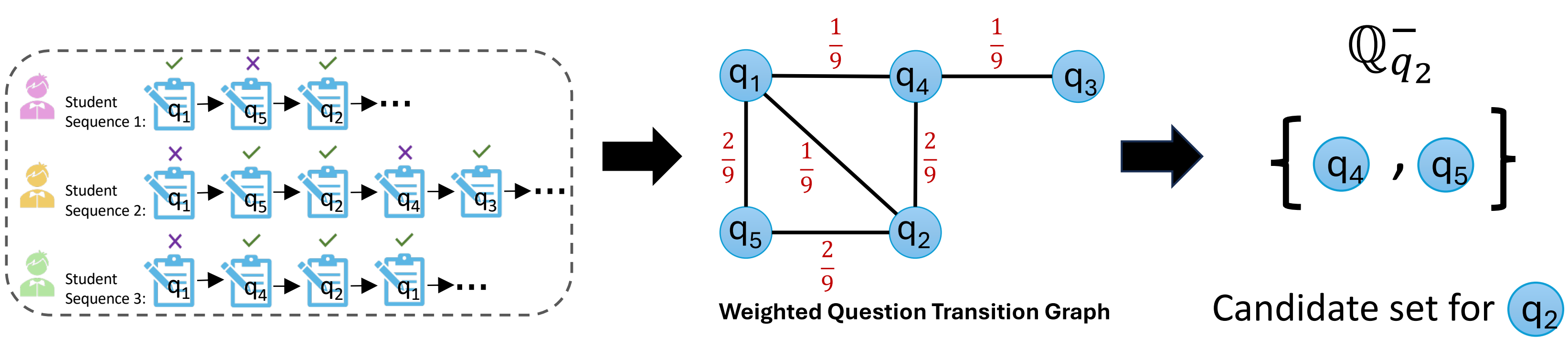
1. Weighted transition graph $w_{q_i,q_j} = \dfrac{\left|I_{q_i,q_j}\right|}{\sum_{i,j}\left|I_{q_i,q_j}\right|}$



Weighted Question Transition Graph

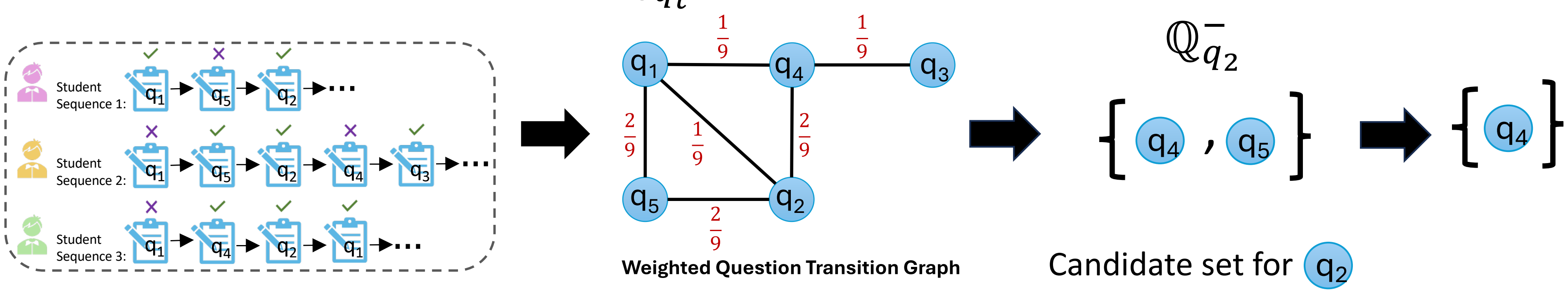
Neighborhood-Aware Negative Sampling (NANS) -

1. Weighted transition graph $w_{q_i,q_j} = \frac{|I_{q_i,q_j}|}{\sum_{i,j}|I_{q_i,q_j}|}$
2. Neighborhood-Aware Negative Candidate Set $(\mathbb{Q}^{-}_{q_t})$

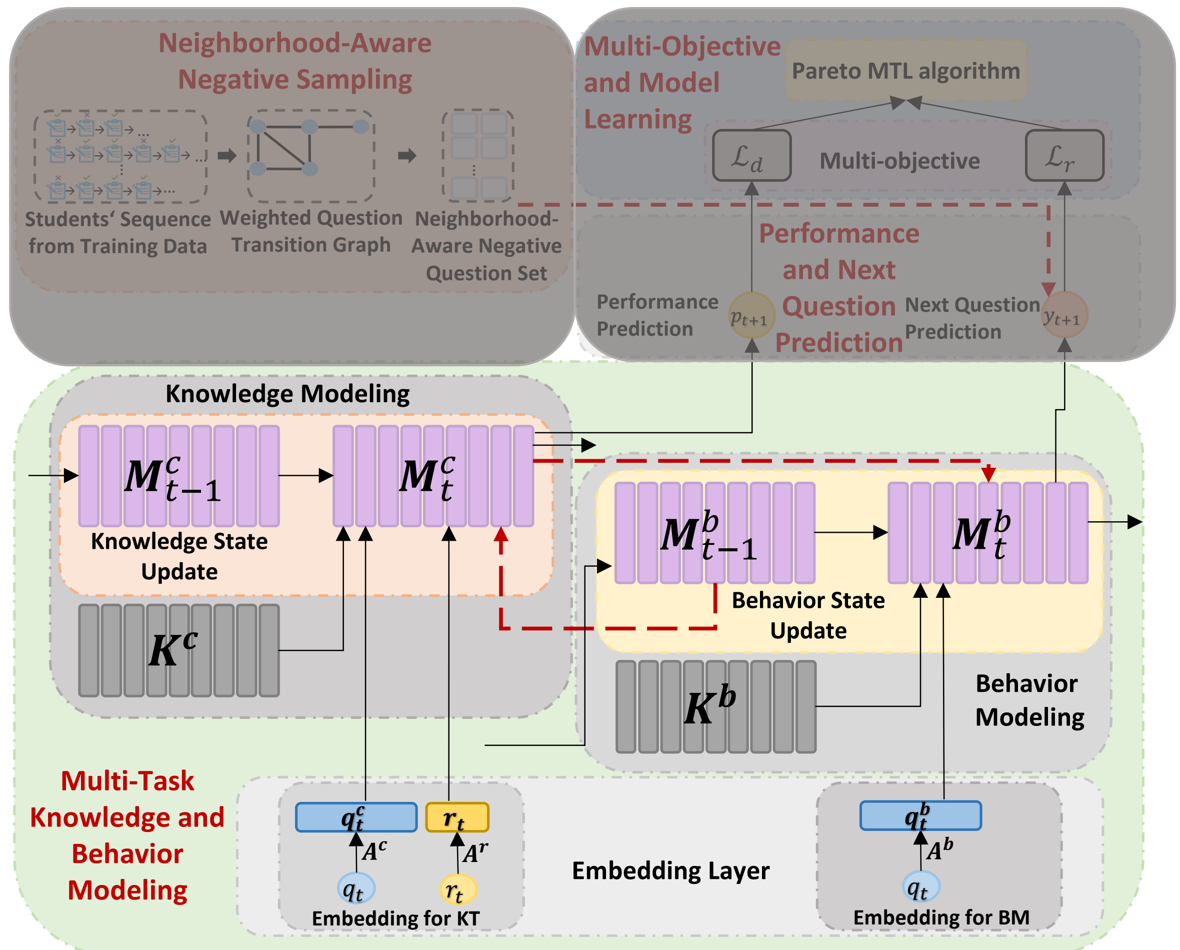


Neighborhood-Aware Negative Sampling (NANS) -

1. Weighted transition graph $w_{q_i,q_j} = \frac{|I_{q_i,q_j}|}{\sum_{i,j}|I_{q_i,q_j}|}$
2. Neighborhood-Aware Negative Candidate Set ($\mathbb{Q}_{q_t}^{-}$)
3. Randomly sample from $\mathbb{Q}_{q_t}^{-}$ to increase diversity of samples



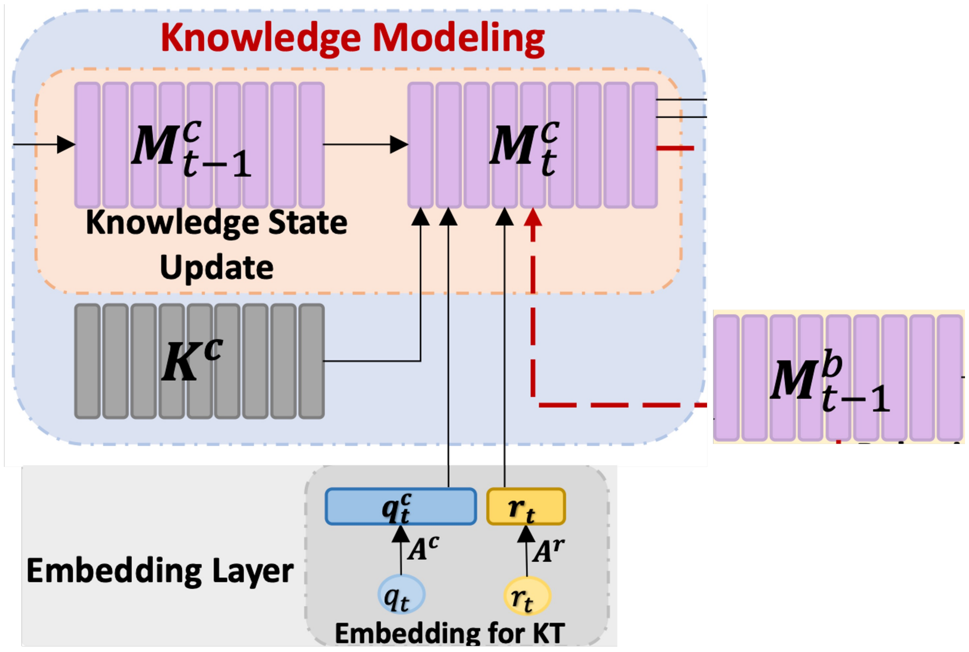
Multi-Task Knowledge and Behavior Modeling (KoBeM)



Behavior-Aware Knowledge Modeling

- Dynamic key-value memory networks for knowledge tracing
  - $\boldsymbol{K}^c$: latent concept features
  - $\boldsymbol{M}_t^c$: student's mastery state

- Update Mastery $\boldsymbol{M}_t^c$
  - erase-followed-by-add
  - considering student's behavior state from the BM component ($\boldsymbol{M}_t^b$)



Behavior-Aware Knowledge Modeling

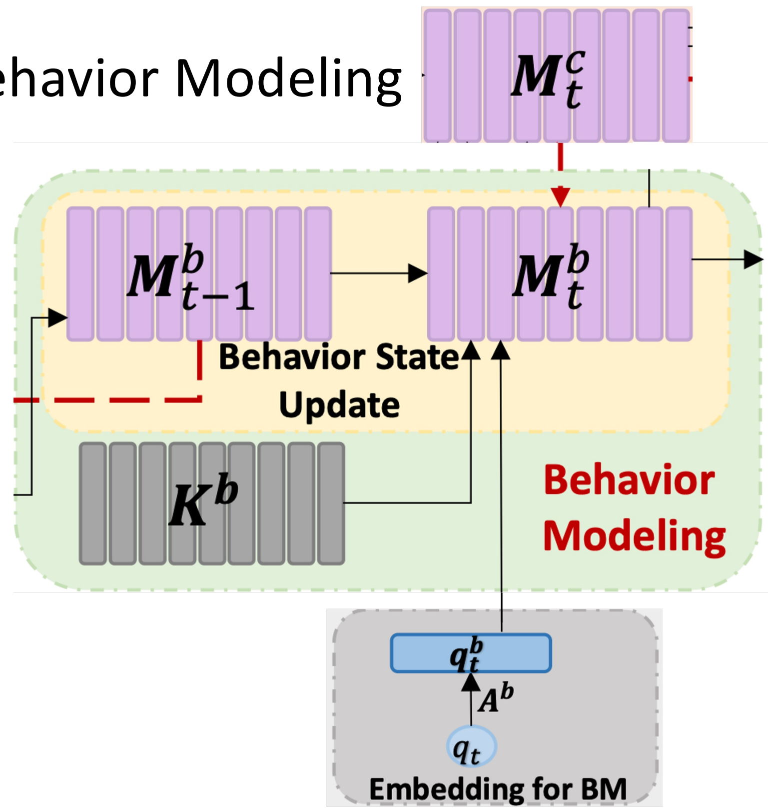
- Erase step:
  - $\boldsymbol{e}_t^c = \sigma\left(\boldsymbol{E}^{c^T}[\boldsymbol{q}_t^{c^T} \oplus \boldsymbol{r}_t] + \boldsymbol{E}_b^{c^T}\boldsymbol{M}_{t-1}^b + \boldsymbol{b}_e^c\right)$
  - $\widetilde{\boldsymbol{M}}_t^c = \boldsymbol{M}_t^c(i).[\mathbf{1} - w_t^c(i)\boldsymbol{e}_t^c]$

- Add step
  - $\boldsymbol{a}_t^c = Tanh(\boldsymbol{D}^{c^T}[\boldsymbol{q}_t^c \oplus \boldsymbol{r}_t] + \boldsymbol{D}_b^{c^T}\boldsymbol{M}_{t-1}^b + \boldsymbol{b}_d^c)$
  - $\boldsymbol{M}_t^c = \widetilde{\boldsymbol{M}}_t^c(i) + w_t^c(i)\boldsymbol{a}_t^c$

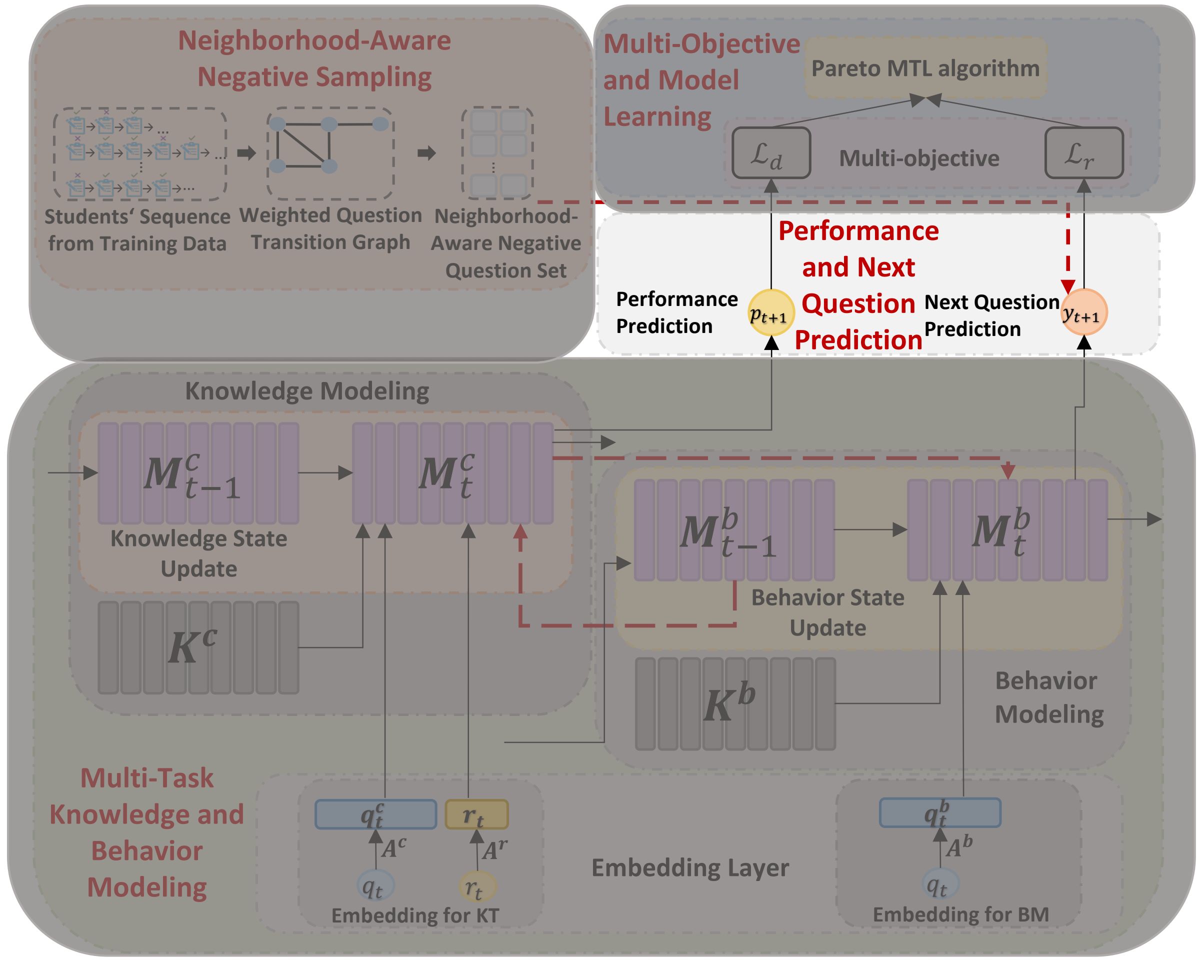
Knowledge-Aware Behavior Modeling

- Dynamic key-value memory networks for Behavior Modeling
  - $\boldsymbol{K}^b$: latent behavior features
  - $\boldsymbol{M}_t^b$: student's behavior state

- Update Mastery $\boldsymbol{M}_t^b$
  - erase-followed-by-add
  - considering student's behavior state from the KT component ($\boldsymbol{M}_t^c$)



Prediction Layer



Prediction Layer

- Next question prediction

$$\boldsymbol{g}_t^b = \sum_{i=1}^{n_d} w_t(i)\boldsymbol{M}_t^b(i)$$

$$\boldsymbol{s}_t = Tanh(\boldsymbol{W}_s^{\mathsf{T}}[\boldsymbol{g}_t^b \oplus \boldsymbol{q}_t^b] + \boldsymbol{b}_s)$$

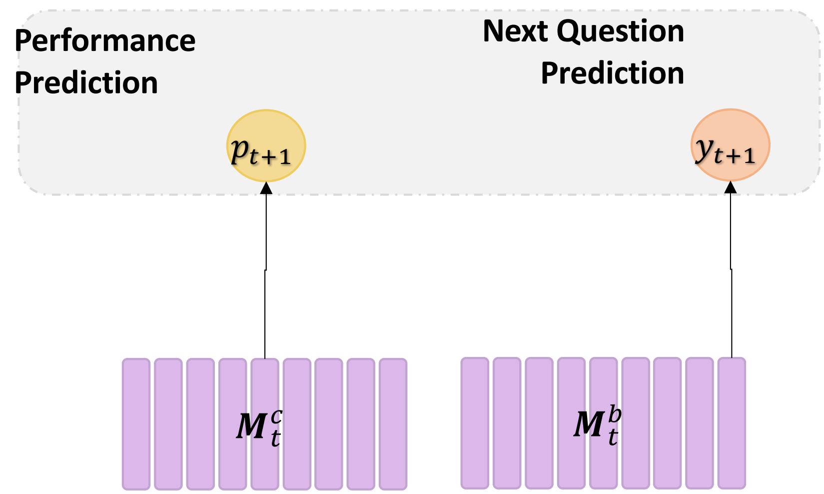
$$\boldsymbol{y}_{t+1} = sigmoid(\boldsymbol{W}_y^T \boldsymbol{s}_t + \boldsymbol{b}_y)$$

- Performance prediction

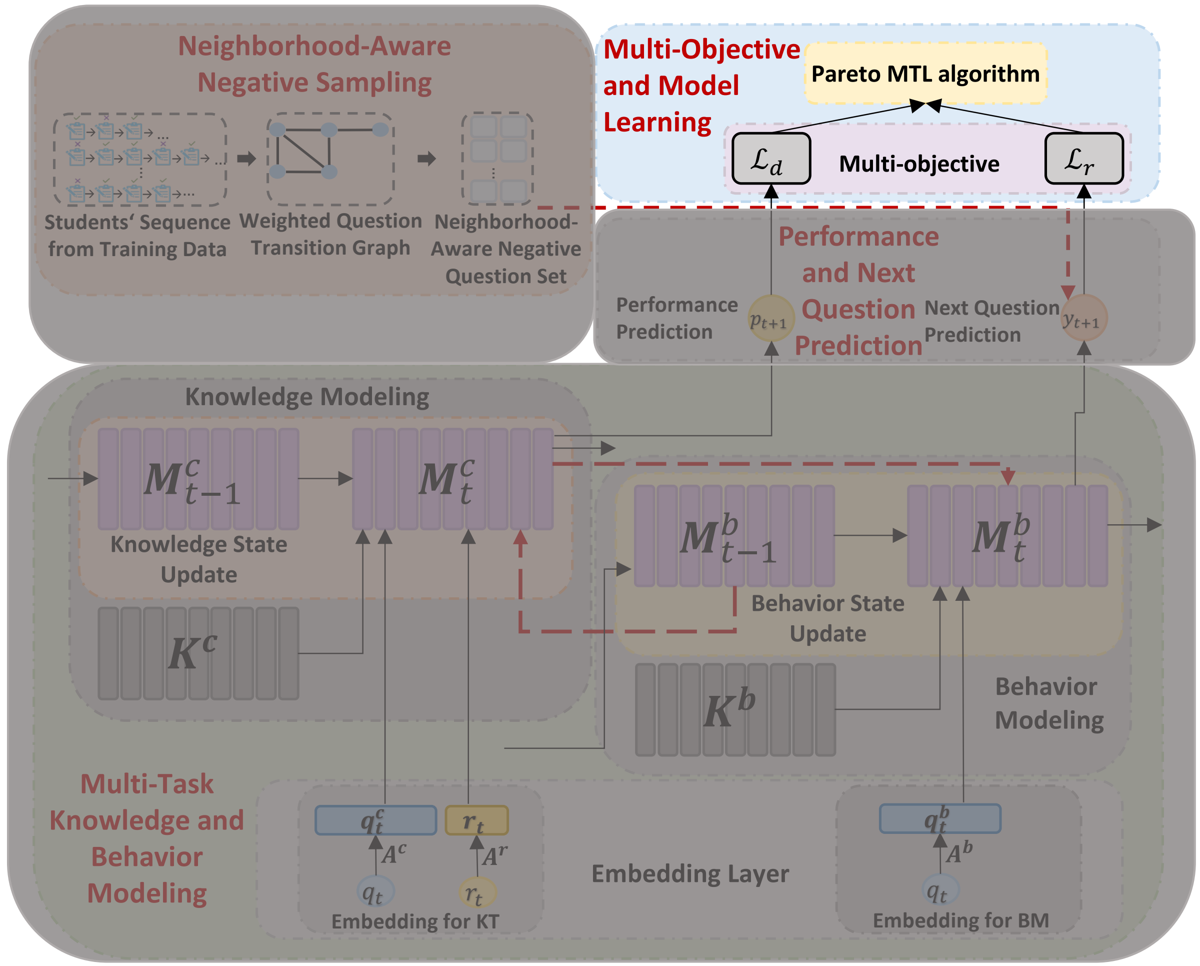
$$\boldsymbol{g}_{t+1}^c = \sum_{i=1}^{n_c} w_{t+1}(i)\boldsymbol{M}_t^c(i)$$

$$\boldsymbol{f}_{t+1} = Tanh(\boldsymbol{W}_f^T[\boldsymbol{g}_{t+1}^c \oplus \boldsymbol{q}_{t+1}^c] + \boldsymbol{b}_f)$$

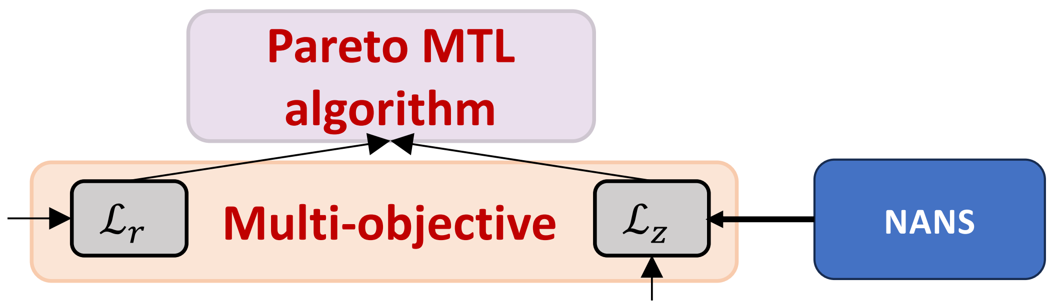
$$\boldsymbol{p}_{t+1} = Sigmoid(\boldsymbol{W}_p^T \boldsymbol{f}_t + \boldsymbol{b}_p)$$



Multi-Objective and Model Learning



Multi-Objective and Pareto Optimization

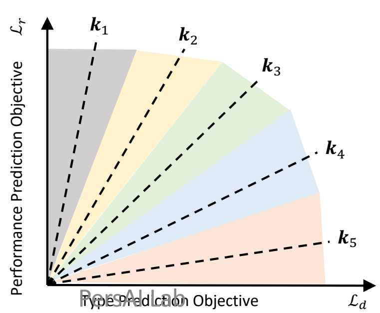


$$\mathcal{L}_r = \sum_t (r_t logp_t + (1 - r_t)log(1 - p_t))$$

Student performance prediction loss

$$\mathcal{L}_z = \sum_t \sum_{\tilde{y}_t \in y_t^p, y_t^n} (z_t log\tilde{y}_t + (1 - z_t)log(1 - \tilde{y}_t))$$

Student behavior prediction loss



[Lin et al., 2019]

Experiments

- Student behavior prediction
  - Compared to baselines
  - With alternative negative sampling methods

- Student performance prediction

Dataset	#Users	#Questions	Question Activities	Question Responses Mean	Question Responses STD	#Correct Question Responses	#Incorrect Question Responses
EdNet	1000	11249	200931	0.5910	0.2417	118747	82184
Junyi	2063	3760	290754	0.6660	0.2224	193664	97090

Student Behavior Prediction Comparison

- NANS-KoBeM structure and negative sampling performs well in predicting the next question, whether with or without knowledge modeling
- But, the information transfer between knowledge modeling and behavior modeling helps
  - Student behavior is influenced by preference knowledge

Methods	Ednet			Junyi		
	HR	NDCG	MRR	HR	NDCG	MRR
LSTM	0.0806**	0.0452**	0.0301**	0.4156**	0.2908**	0.2611**
MANN	0.0853**	0.0515**	0.0396**	0.4284**	0.3277**	0.2838**
NANS-KoBeM w/o KM	0.0877**	0.0555**	0.0449**	0.4391**	0.3387**	0.3051**
NANS-KoBeM	**0.0982**	**0.0622**	**0.0503**	**0.4919**	**0.3795**	**0.3422**

Student Behavior Prediction with Alternative Negative Sampling

- Neighborhood-aware negative sampling (NANS) is is more effective than
  - Random negative sampling (RD)
  - Weighted distribution negative sampling (WD)
  - Graph representation learning negative sampling (MCNS), and
  - Neighborhood overlap difference negative sampling (GNNO)

Methods	Ednet			Junyi		
	HR	NDCG	MRR	HR	NDCG	MRR
RD-KoBeM	0.0973*	0.0616*	0.0498*	0.4807*	0.3709*	0.3344**
WD-KoBeM	0.0969*	0.0614*	0.0496*	0.4355**	0.3490**	0.3200**
MCNS-KoBeM	0.0972*	0.0617*	0.0499*	0.4871*	0.3752*	0.3387**
GNNO-KoBeM	0.0974*	0.0618*	0.0499*	0.4875*	0.3756*	0.3389**
NANS-KoBeM w/o KM	0.0877**	0.0555**	0.0449**	0.4391**	0.3387**	0.3051**
NANS-KoBeM	**0.0982**	**0.0622**	**0.0503**	**0.4919**	**0.3795**	**0.3422**

Student Performance Prediction Comparison

Methods	EdNet AUC	Junyi AUC
DKT	0.6393**	0.8623**
SAKT	0.6334**	0.8053**
SAINT	0.5205**	0.7951**
AKT	0.6393**	0.8093**
DeepIRT	0.6290**	0.8498**
DKVMN (NANS-KoBeM w/o BM)	0.6296**	0.8558**
NANS-KoBeM	**0.6615**	**0.8779**

- NANS-KoBeM structure and negative sampling performs well in predicting student performance

- But, the information transfer between knowledge modeling and behavior modeling helps
  - Student knowledge is influenced by preference behavior

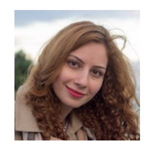
Conclusions

- Proposed NANS, a neighborhood-aware negative sampling method
  - Effectively capturing coarse-grained sequential data to produce difficult and diverse negative samples
- Proposed KoBeM, a multi-objective multi-task student knowledge and behavior model, and accordingly, proposed NANS-KoBeM
  - Effectively combining knowledge modeling and behavior modeling to enhance both tasks
  - Modeling the interrelationships between student knowledge and question preference behavior

UNIVERSITY AT ALBANY
State University of New York

PersAI
onalized

Thank you!
Q & A



szhao2@albany.edu ssahebi@albany.edu

Our code and sample data are available at GitHub:
https://github.com/persai-lab/2025-NANSKoBeM

AAAI-25

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