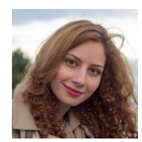
Multi-Task Modeling of Student Knowledge and Behavior

Siqian Zhao

szhao2@albany.edu

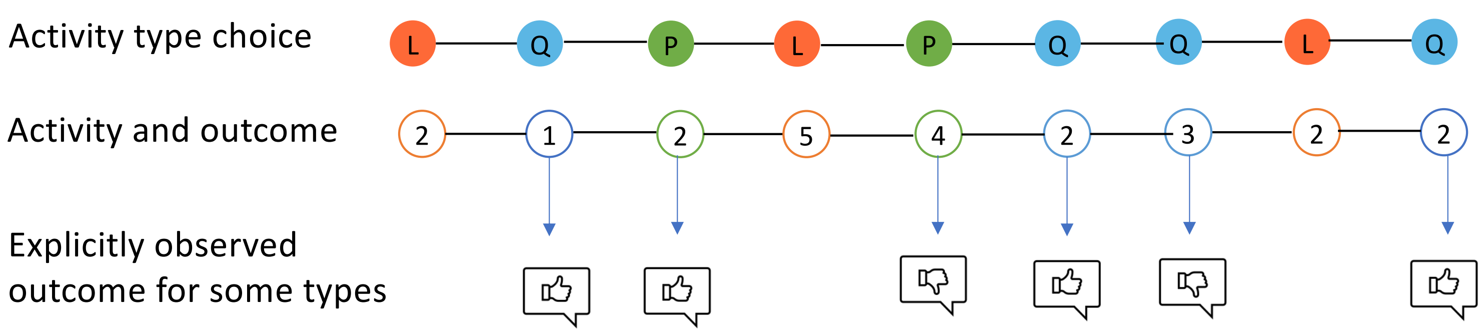


Sherry Sahebi

ssahebi@albany.edu

General problem

- Modeling
  - interrelations between parallel input sequences
  - in multi-type sequential data
  - between activities with implicit and explicit observations
- Predicting
  - Next sequence item (task 1) and outcome (task 2)



Application Context: Student Modeling

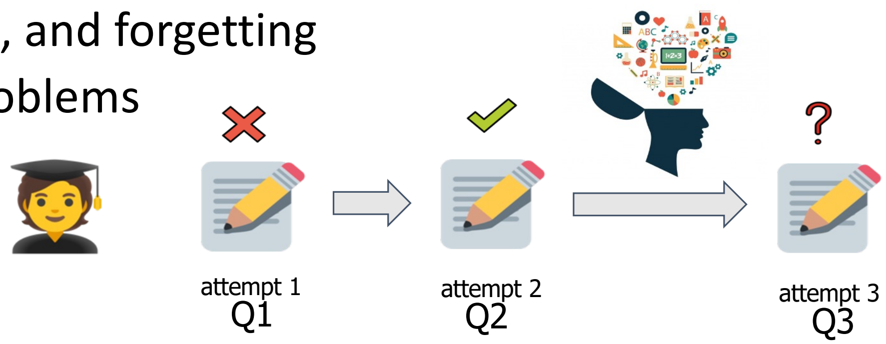
- Online education systems
  - Enabling distance learning and abundant courses
  - Attracting more and more students
  - Promoting the development of Educational Data Mining (EDM)

- 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

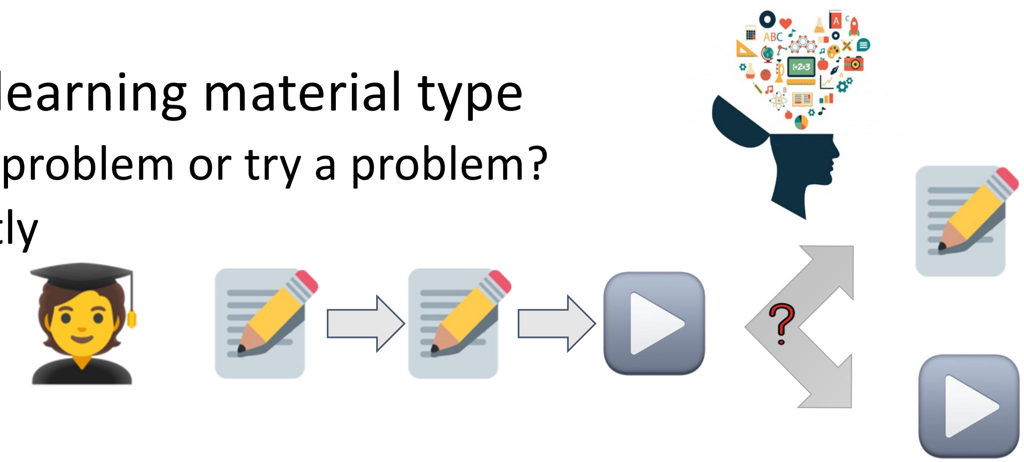
- Challenges
  - Multi-type: learning from non-assessed and assessed
  - Sequence: knowledge increase, learning, and forgetting
  - Sparsity: too few attempts, too many problems
  - Noise: slipping and guessing answers



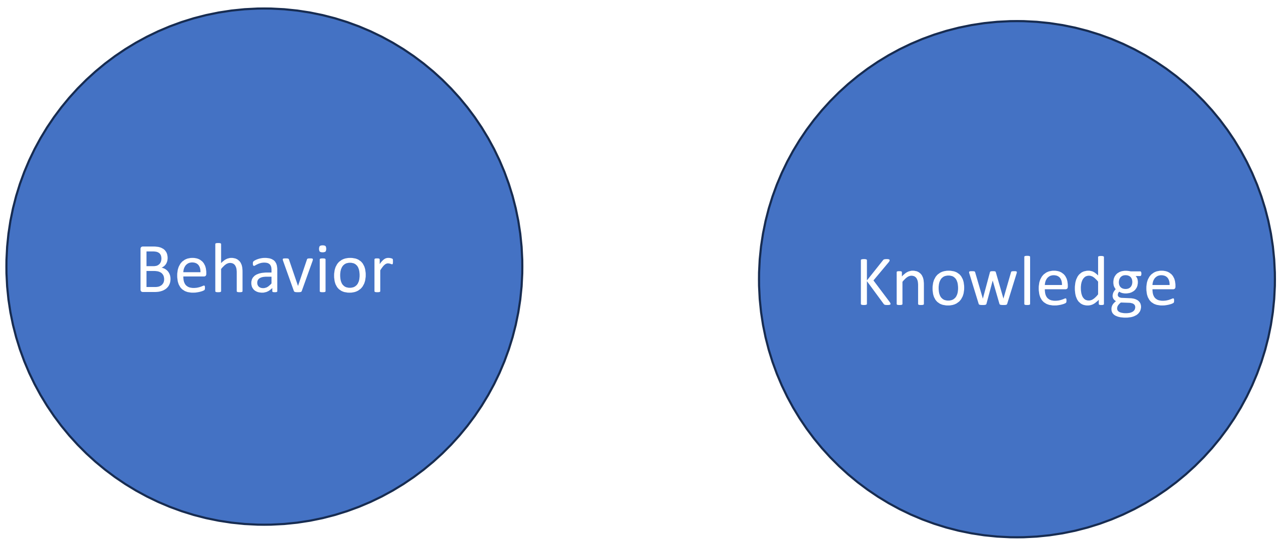
Student Behavior Modeling

- Understanding students' behavior patterns during the learning process
  - Preference for learning materials
  - Engagement
  - Procrastination

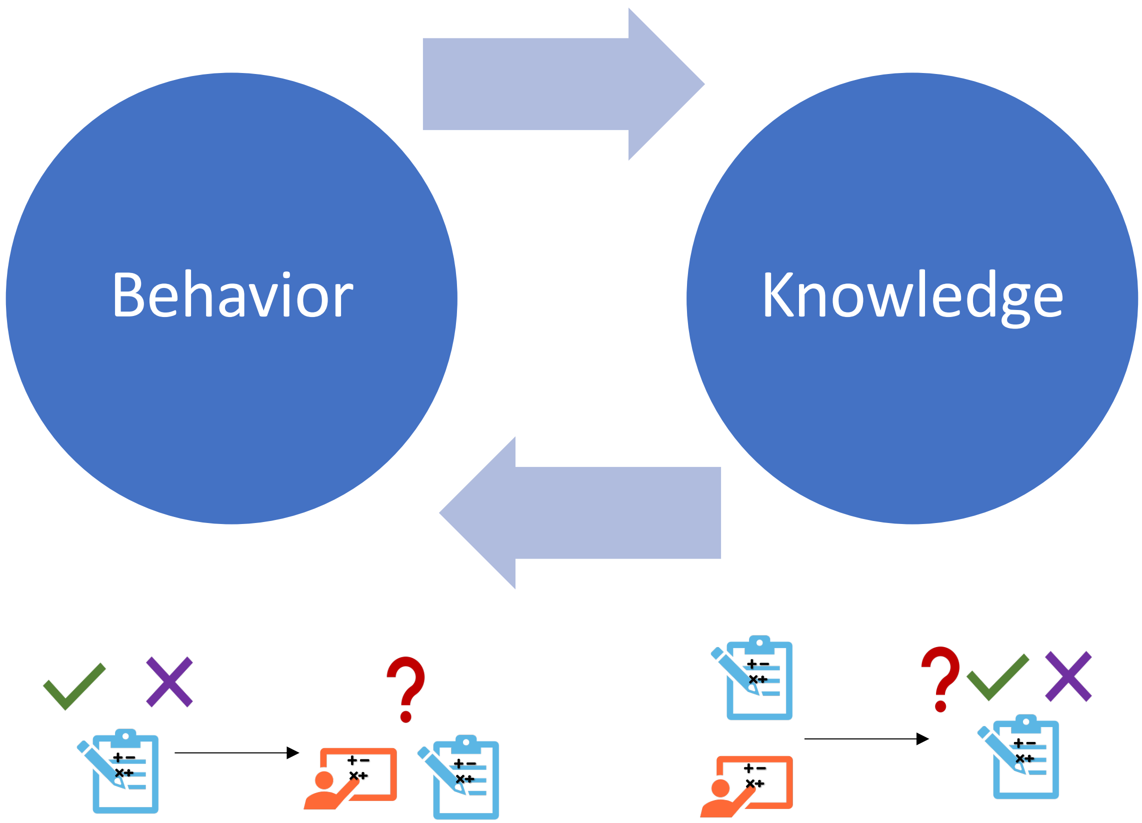
- Modeling choice / preference of learning material type
  - Read a book chapter after failing a problem or try a problem?
  - Different students choose differently



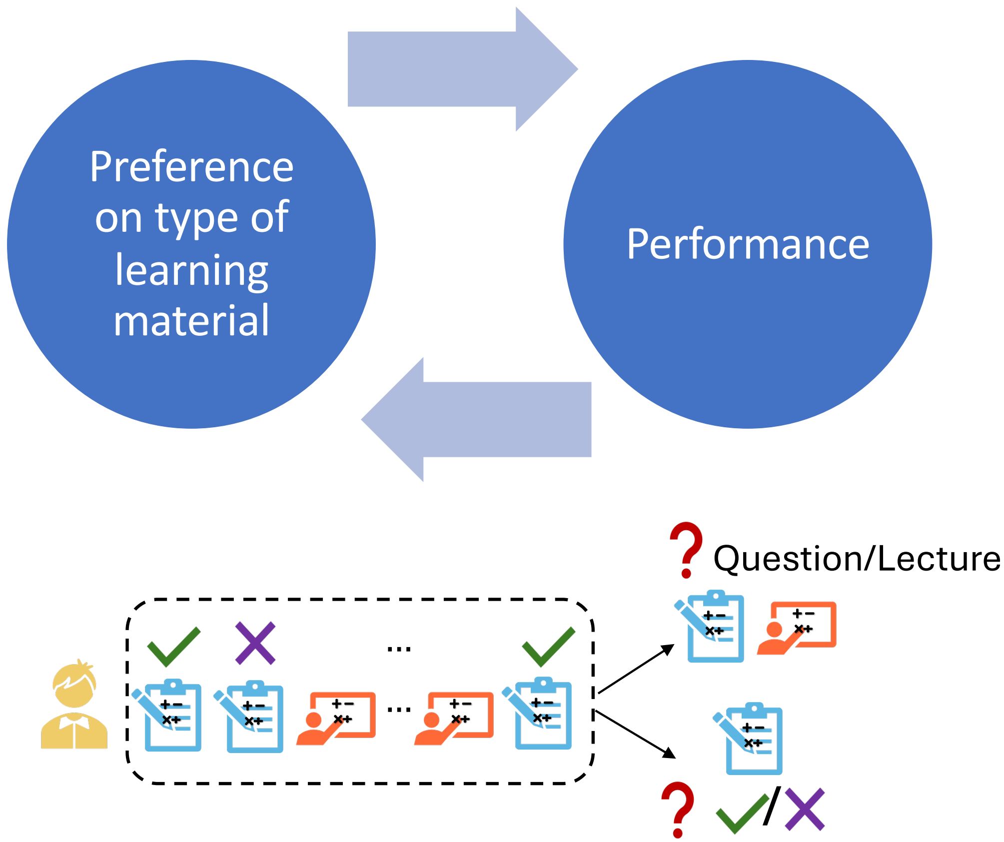
KT and BM have been addressed independently



What about their interrelations?

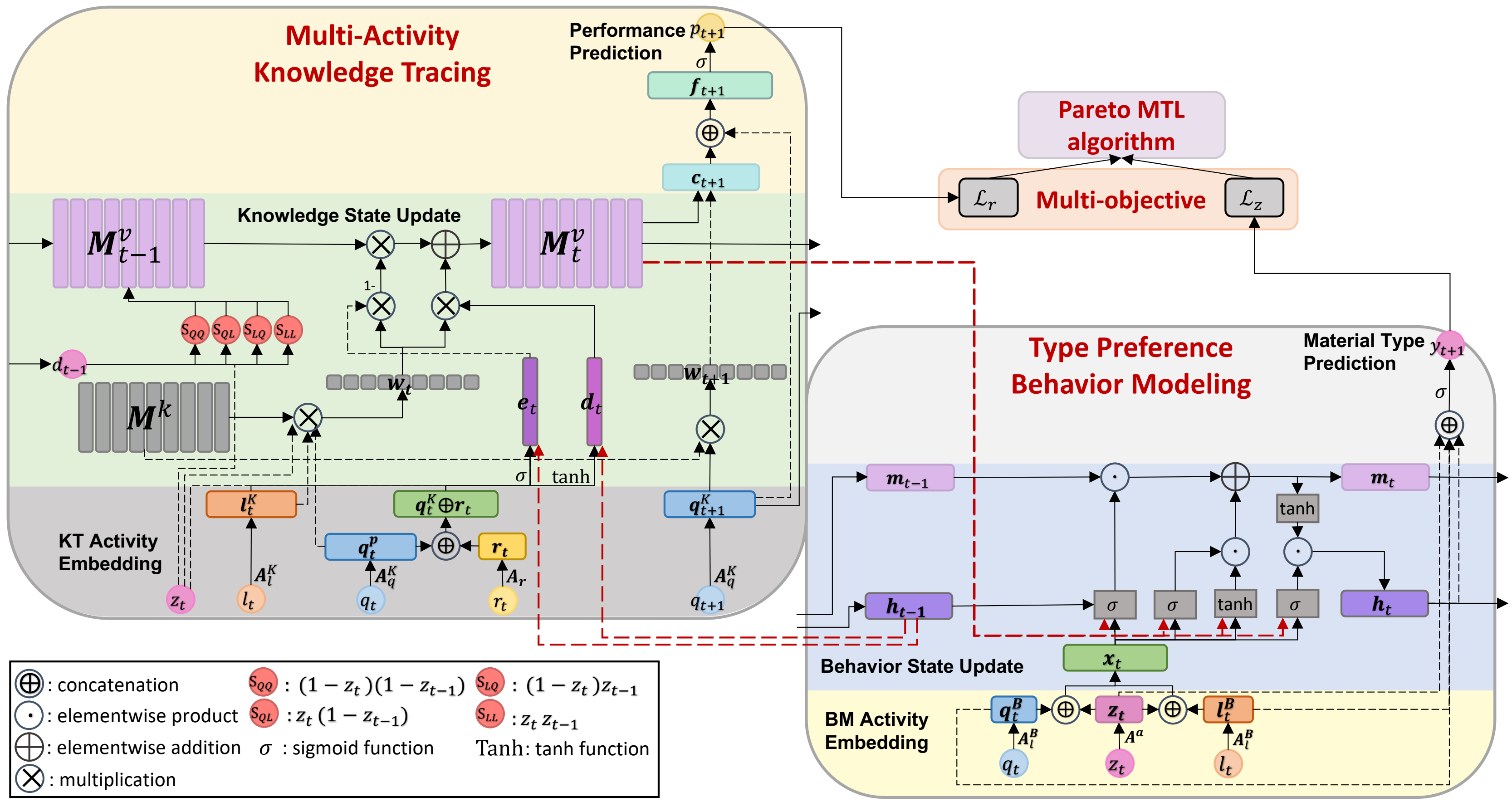


Let’s model knowledge and behavior simultaneously

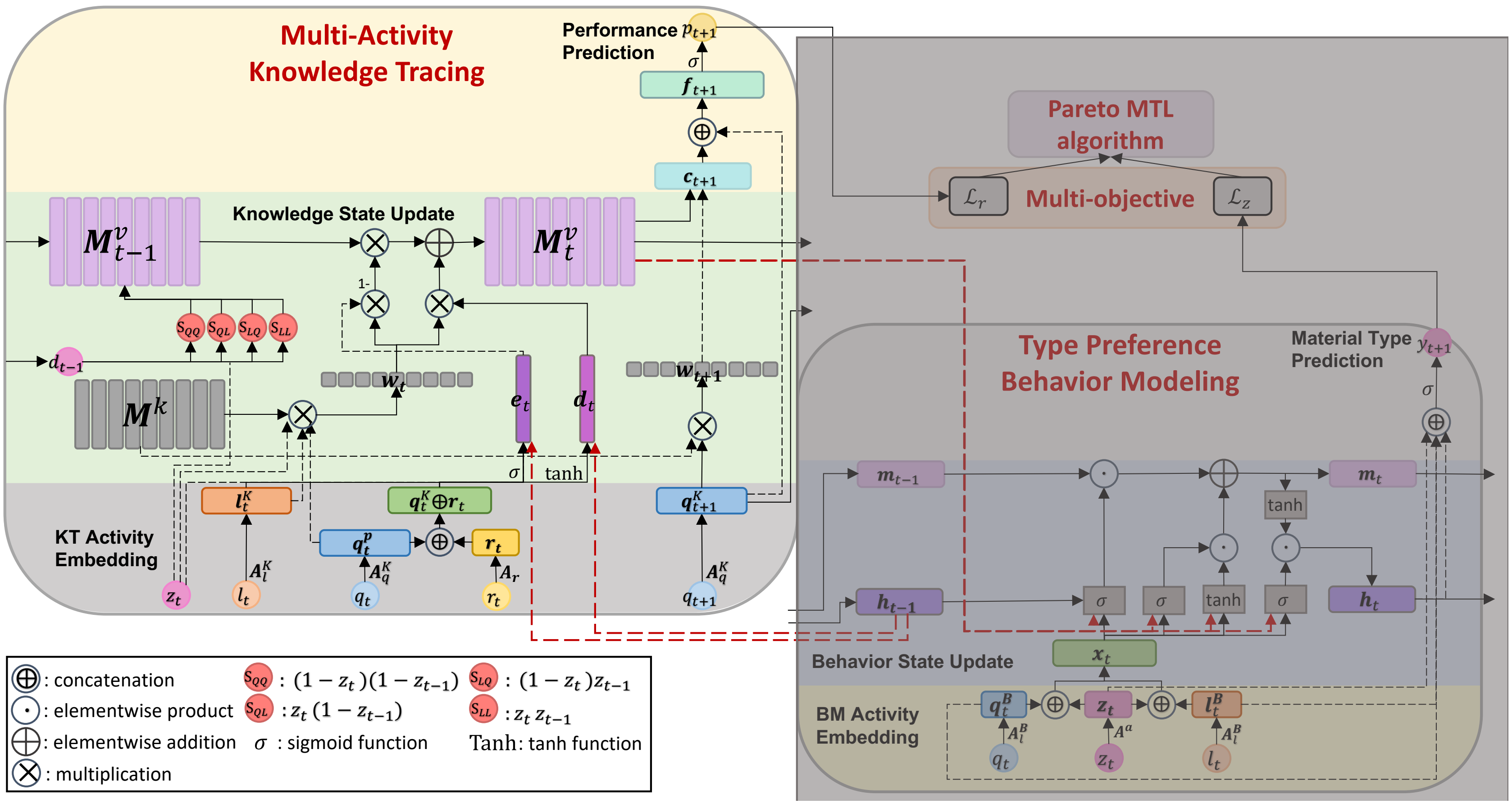


- Challenges:
  - representing knowledge and preference behavior states
  - modeling information transfer between the two tasks
  - balancing task objectives for mutual benefit

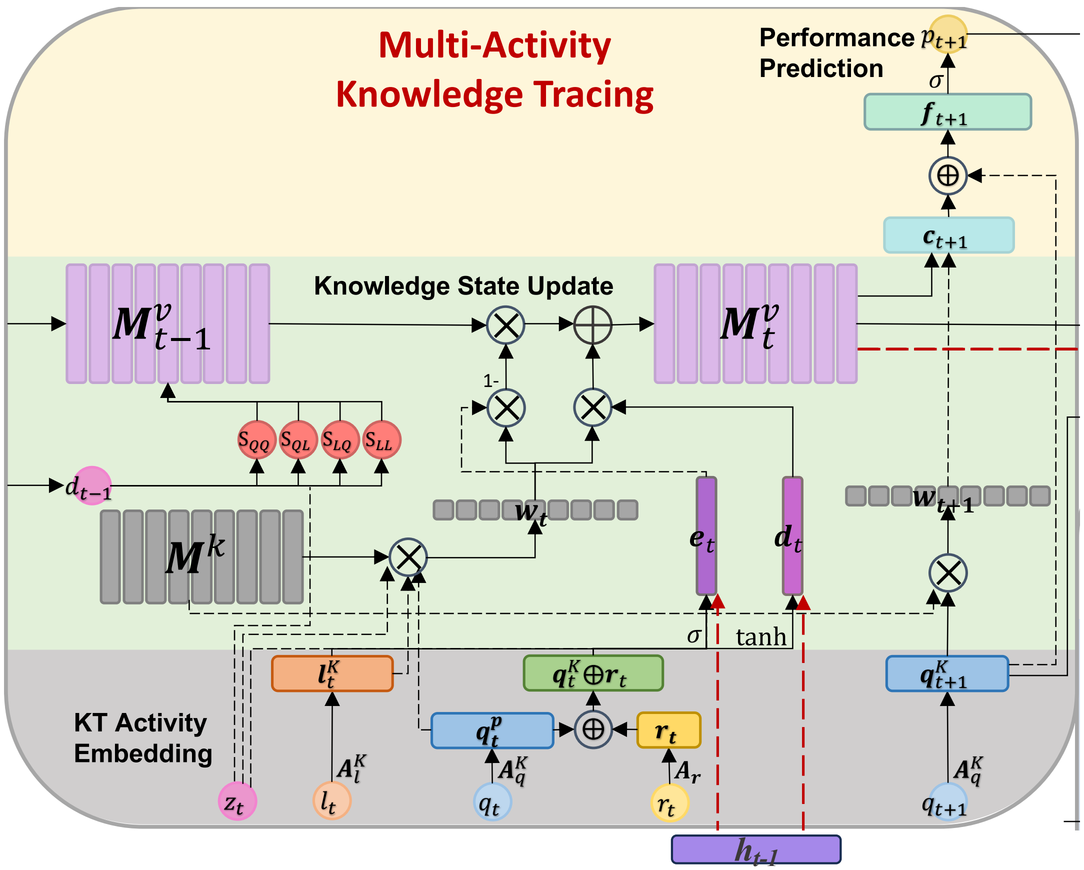
Multi-Task Student Knowledge and Behavior Model (KTBM)



Multi-Activity Knowledge Tracing



Multi-Activity Knowledge Tracing



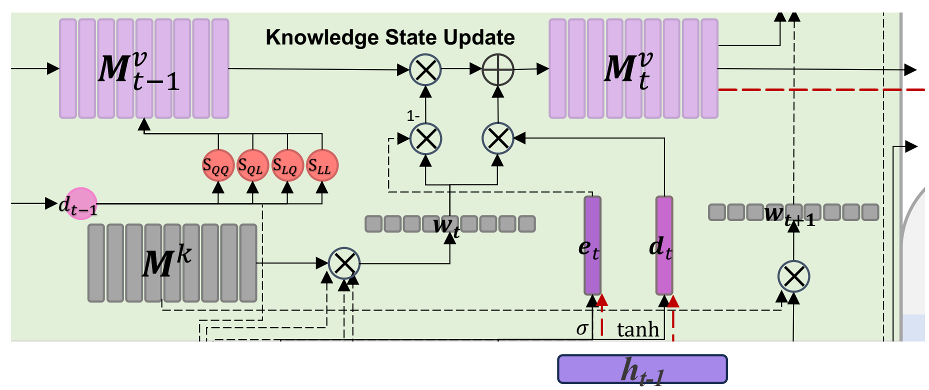
Performance Prediction Layer

Hidden Knowledge Transfer Layer

Embedding layer

Multi-Activity Knowledge Tracing – Knowledge Transfer

- Represents dynamic student knowledge
  - Based on dynamic key-value memory networks
    - M^k: latent concept features
    - M_t^v: student's mastery state
- Models how knowledge transfers between different activity types
  - Using separate knowledge transfer matrices enabled by transition indicators
  - E.g., $S_{QL} = z_t(1 - z_{t-1})$
- Considering behavior state
  - h_{t-1}: From the BM component



Multi-Activity Knowledge Tracing – Knowledge Transfer

Erase-followed-by-add to update Mastery $\boldsymbol{M}_t^v$

Erase:

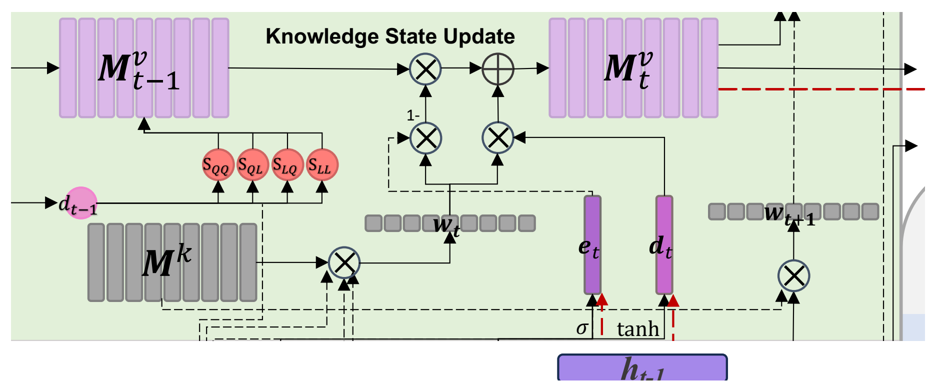
$$\boldsymbol{e}_t = \sigma\Big((1 - z_t)\cdot \boldsymbol{E}_q^T[\boldsymbol{q}_t^k \oplus \boldsymbol{r}_t] + z_t \cdot \boldsymbol{E}_l^T \boldsymbol{l}_t^k + \boldsymbol{E}_b^T \boldsymbol{h}_{t-1} + \boldsymbol{b}_e\Big)$$

$$\widetilde{\boldsymbol{M}}_t^v(i) = [S_{QQ}\cdot T_{QQ}\boldsymbol{M}_{t-1}^v + S_{LL}\cdot T_{LL}\boldsymbol{M}_{t-1}^v + S_{QL}\cdot T_{QL}\boldsymbol{M}_{t-1}^v + S_{LQ}\cdot T_{LQ}\boldsymbol{M}_{t-1}^v](i)\cdot[\boldsymbol{1} - w_t(i)\boldsymbol{e}_t]$$

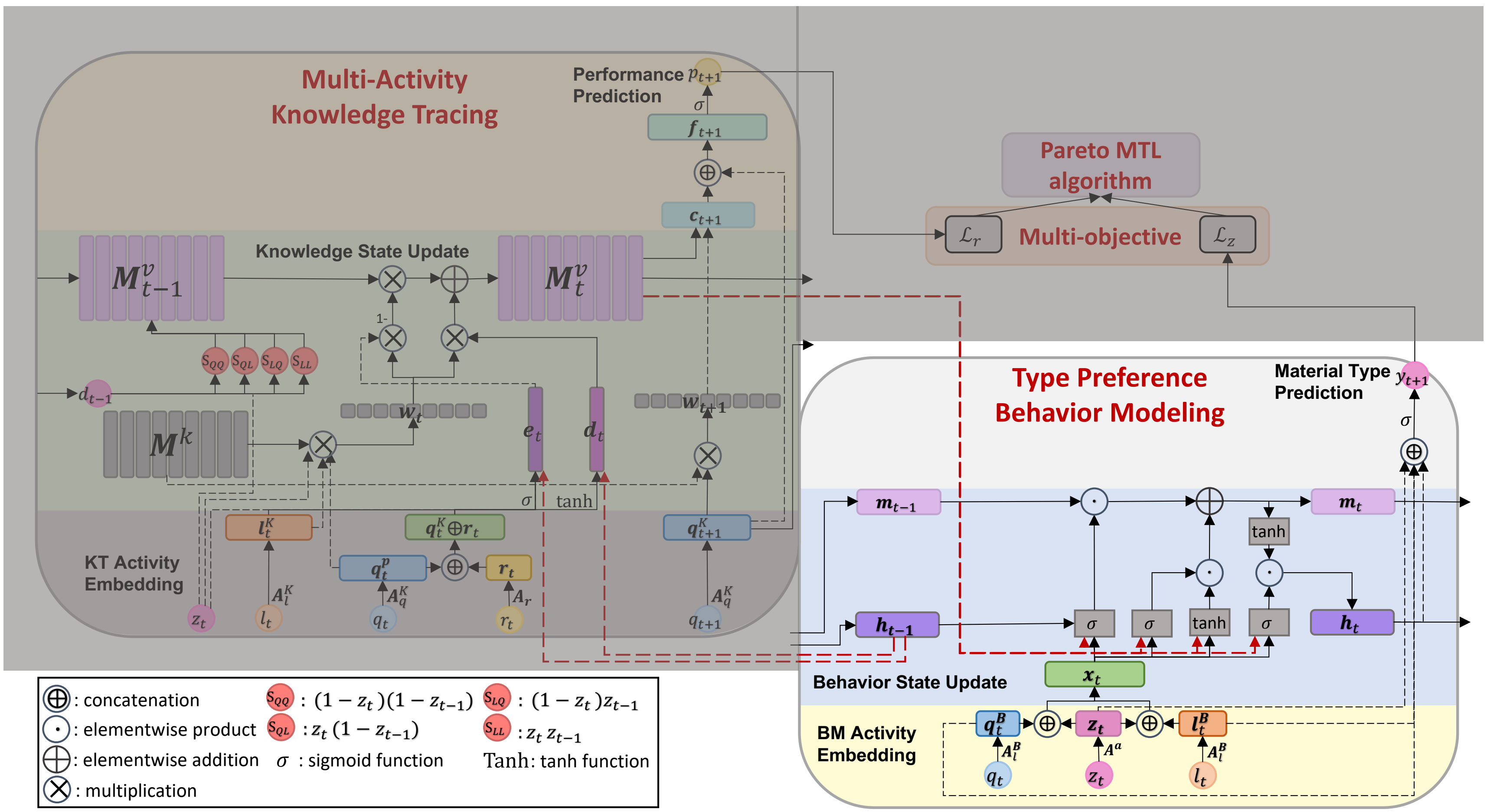
Add:

$$\boldsymbol{d}_t = Tanh\Big((1 - z_t)\cdot \boldsymbol{D}_q^T[\boldsymbol{q}_t^k \oplus \boldsymbol{r}_t] + z_t \cdot \boldsymbol{D}_l^T \boldsymbol{l}_t^k + \boldsymbol{D}_b^T \boldsymbol{h}_{t-1} + \boldsymbol{b}_d\Big)$$

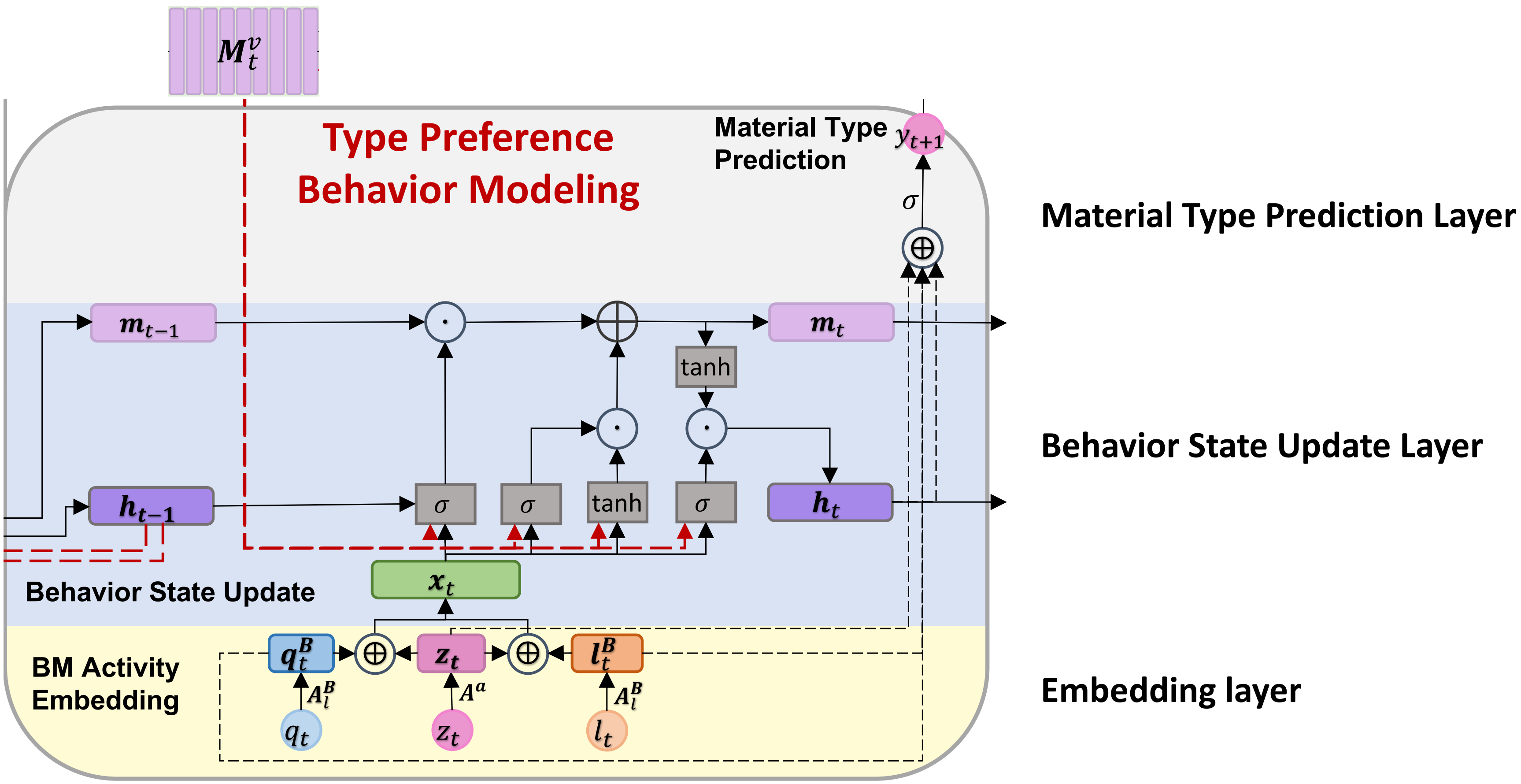
$$\boldsymbol{M}_t^v(i) = \widetilde{\boldsymbol{M}}_t^v\,(i) + w_t(i)\boldsymbol{d}_t$$



Type Preference Behavior Modeling



Type Preference Behavior Modeling



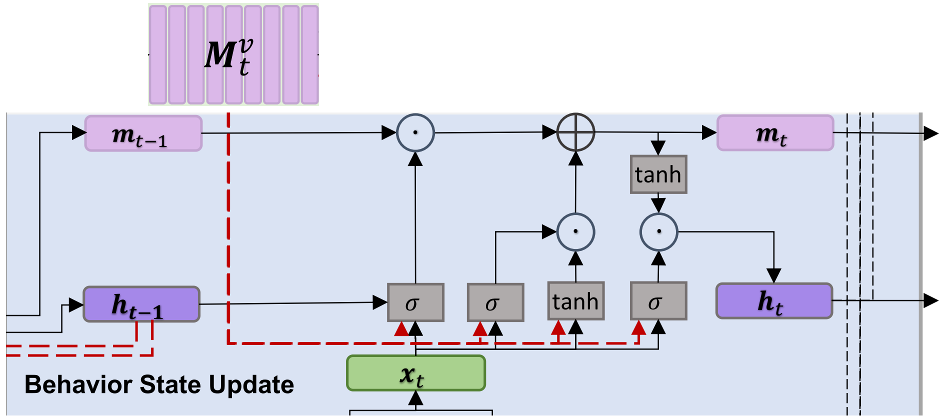
Type Preference Behavior Modeling - Behavior State Update

- Represents dynamic student behavior
  - Based on LSTM
  - $\boldsymbol{h_{t-1}}$: Behavior state
- Considering student knowledge
  - $\boldsymbol{M_t^v}$: student's mastery state from the KT component

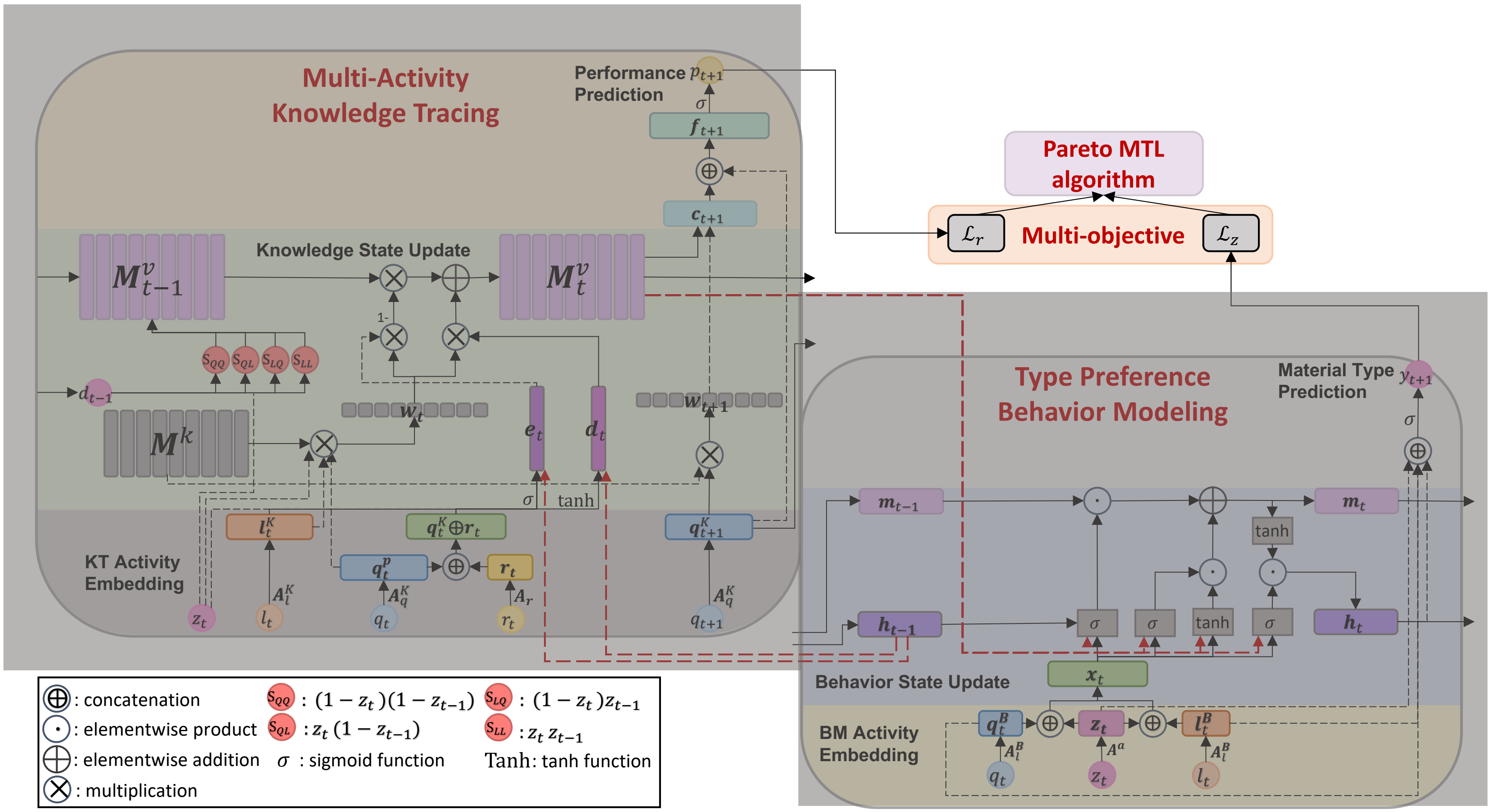
$$\boldsymbol{x}_t = (1 - z_t) \cdot \boldsymbol{X}_q^T[\boldsymbol{q}_t^B \oplus \boldsymbol{z}_t] + z_t \cdot \boldsymbol{X}_l^T[\boldsymbol{l}_t^B \oplus \boldsymbol{z}_t]$$

$$\boldsymbol{K}_t = \boldsymbol{W}_k^T \boldsymbol{M}_t^v + \boldsymbol{b}_k$$

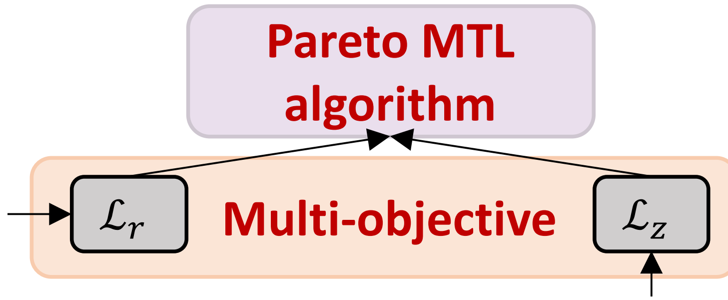
$$\boldsymbol{h}_t = LSTM(\boldsymbol{h}_{t-1}^b, \boldsymbol{K}_t, \boldsymbol{x}_t)$$



Multi-Objective and Pareto Optimization



Multi-Objective and Pareto Optimization

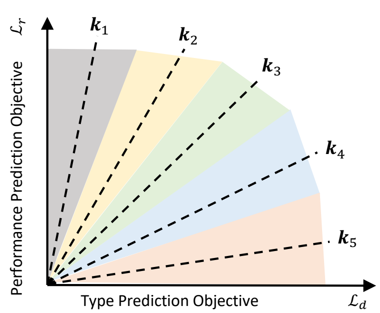


$$\mathcal{L}_r = \sum_t (r_t log p_t + (1 - r_t) log(1 - p_t))$$

Student performance prediction loss

$$\mathcal{L}_z = \sum_t (z_t log y_t + (1 - z_t) log(1 - y_t))$$

Student choice prediction loss



[Lin et al., 2019]

Experiments

- Prediction performance comparisons
  - Student performance
  - Student choice behavior
- Ablation studies
- Student group analysis
  - By average student score
  - By ratio of non-assessed to all activities
- Knowledge and behavior state visualization

Student Performance Prediction Comparison

	Methods	EdNet AUC	Junyi AUC	MORF RMSE
Assessed-only	DKT	0.6393**	0.8623**	0.1990**
	DKVMN	0.6296**	0.8558**	0.1995**
	SAKT	0.6334**	0.8053**	0.1975**
	SAINT	0.5205**	0.7951**	0.2190**
	AKT	0.6393**	0.8093**	0.2417**
	DeepIRT	0.6290**	0.8498**	0.1946**
Extended assessed-only	DKT+M	0.6372**	0.8652**	0.1942**
	DKVMN+M	0.6343**	0.8513**	0.2071**
	SAKT+M	0.6323**	0.7911**	0.1981**
	SAINT+M	0.5491**	0.7741**	0.2007**
	AKT+M	0.6404**	0.8099**	0.2226**
	MLP+M	0.6102**	0.7290**	0.2428**
Multi-type	MVKM	–	–	0.1936**
	DMKT	0.6394**	0.8561**	0.1856**
	TAMKOT	0.6786*	0.8745**	0.1857**
	GMKT	<u>0.6819</u>	<u>0.8960</u>	<u>0.1802*</u>
	KTBM	**0.6838**	**0.8989**	**0.1778**

- Information transfer between KT and BM helps
- Student knowledge is influenced by preference behavior

Student Preference Prediction Comparison

Methods	EdNet AUC	Junyi AUC	MORF AUC
LSTM	0.8768**	0.9069**	0.9221*
MANN	<u>0.8933*</u>	0.9299**	0.9223*
TAMKOT	0.8929**	0.9355*	0.9256*
GMKT	0.8932*	<u>0.9360*</u>	<u>0.9257*</u>
KTBM	**0.8992**	**0.9390**	**0.9272**

- Information transfer between KT and BM helps
- Preference behavior is influenced by student knowledge

Student Group Analysis – Average Performance

- KTBM shows improvement in all groups
- The better the student does, the easier to predict their performance
- Better-performing student scores make it easier to predict their material type selections

Range of Avg Performance	Student Performance AUC				Material Type AUC			
	DKT	TAMKOT	GMKT	KTBM	LSTM	TAMKOT	GMKT	KTBM
[0, 0.57]	0.6315*	0.6508*	0.6527	**0.6527**	0.8675*	0.8810	0.8819	**0.8825**
[0.57, 0.67]	0.6367*	0.6599*	0.6685	**0.6696**	0.8791*	0.8860	0.8869*	**0.8997**
[0.67, 1]	0.6304*	0.6604*	0.6718*	**0.6761**	0.8780*	0.8964*	0.8973*	**0.9094**

Most improvement group

Student Group Analysis – Activity Type Ratio

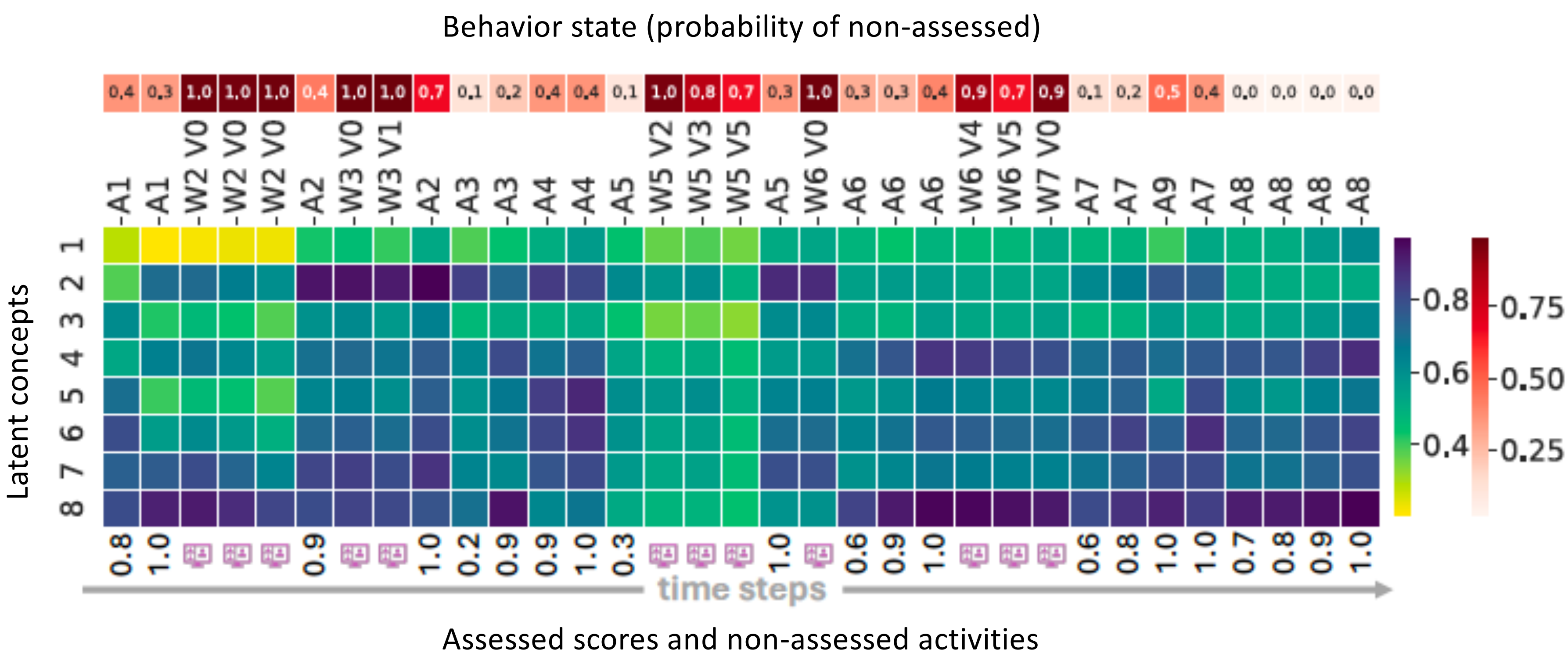
- KTBM shows improvement in all groups

Range of Non-Assessed Activity Ratio	Student Performance AUC				Material Type AUC			
	DKT	TAMKOT	GMKT	KTBM	LSTM	TAMKOT	GMKT	KTBM
[0, 0.4]	0.6761*	0.6823*	0.6844	**0.6845**	0.8177*	0.8269	0.8271	**0.8275**
[0.4, 0.48]	0.6359*	0.6837*	0.6849*	**0.6887**	0.8879*	0.8969*	0.8980*	**0.9073**
[0.48, 1]	0.6194*	0.6702*	0.6775*	**0.6821**	0.9038*	0.9120*	0.9131*	**0.9214**

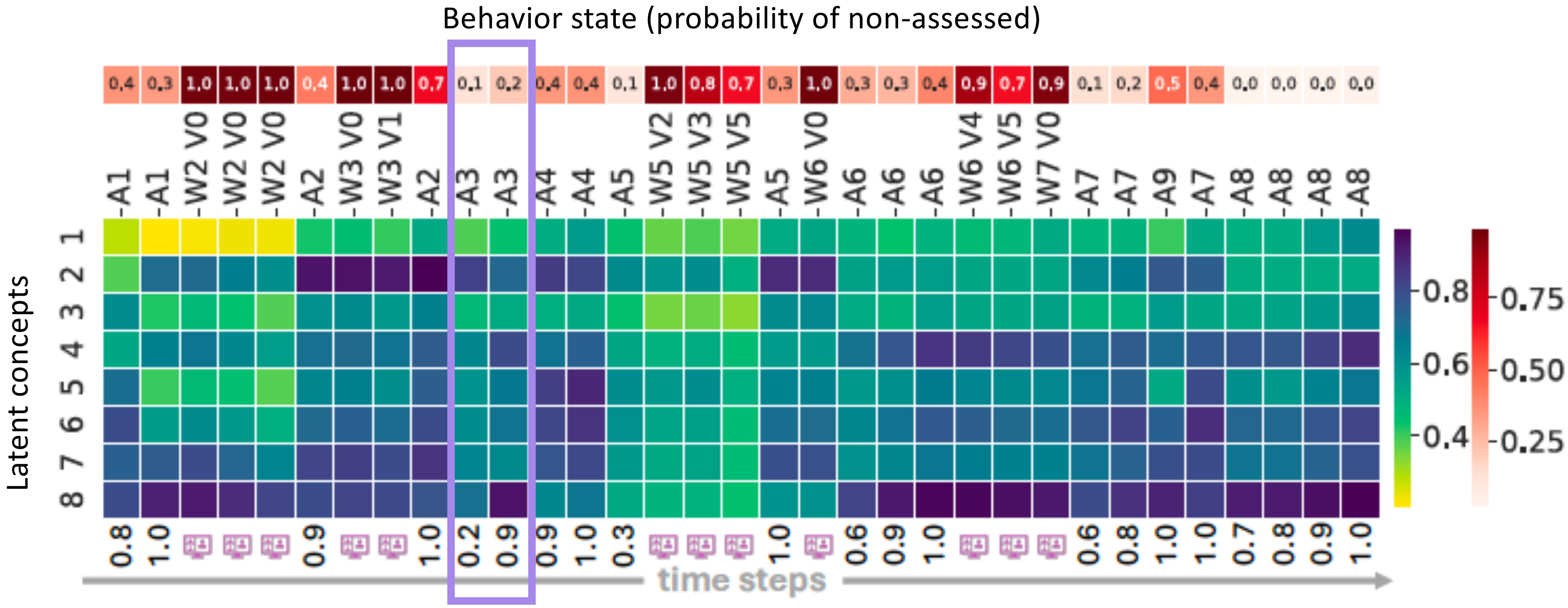
More non-assessed activities
Most difficult to predict
Most improvement!

Relatively balanced
Difficult to predict
Most improvement!

Knowledge and Behavior State Visualization

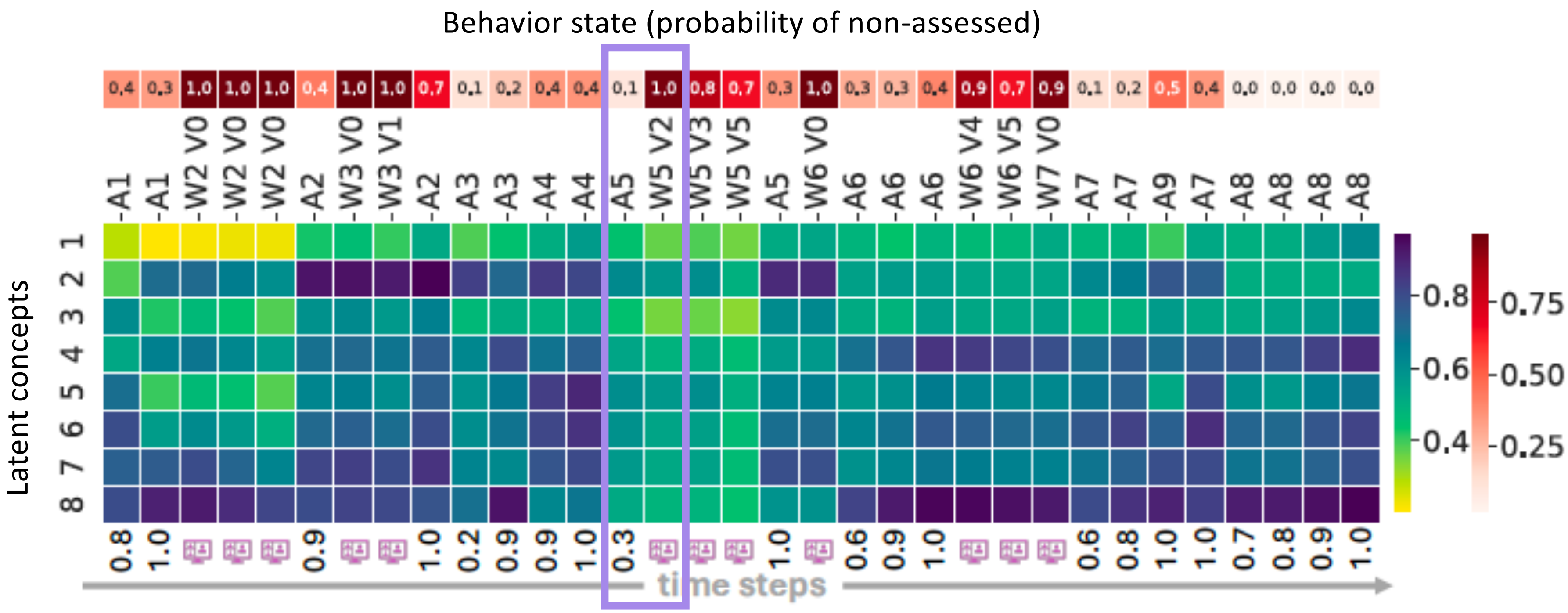


Knowledge and Behavior State Visualization



Decides to try the assignment again to get a higher score

Knowledge and Behavior State Visualization



Decides to watch video lectures after a low assignment score

Conclusions

- Proposed KTBM, a multi-task student knowledge and behavior model
- Effectively combining knowledge tracing and behavior modeling to enhance both tasks
- Modeling the interrelationships between student knowledge and material type preference behavior
- Interpretable by visualization
- Effective for predicting performance in the most challenging group: students engaged primarily in non-assessed activities

UNIVERSITY AT ALBANY
State University of New York

PersAI onalized

Thank you!
Q & A

szhao2@albany.edu ssahebi@albany.edu

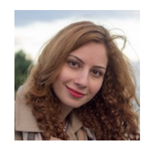
Our code and sample data are available at GitHub:
https://github.com/persai-lab/2024-CIKM-KTBM

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

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/2024-CIKM-KTBM

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

Model Complexity

4.3.1 ***Model Complexity.*** The time complexity of KTBM for a student activity sequence of length L_s is $O\Big(L_s \cdot \Big(N \cdot \max(d_q^K, d_l^K) \cdot d_c + N \cdot d_c + N \cdot d_v \cdot \max(d_q^K + d_r^K, d_l^K) + d_h \cdot (\max(d_q^B, d_l^B) + d_h) + d_h \cdot d_v\Big)\Big)$. Moreover, the time cost also depends on the number of dividing vectors set in the experiments.