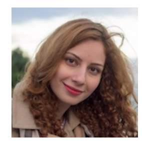
Graph-Enhanced Multi-Activity Knowledge Tracing

Siqian Zhao

szhao2@albany.edu



Shaghayegh Sherry Sahebi

ssahebi@albany.edu

UNIVERSITY AT ALBANY
State University of New York

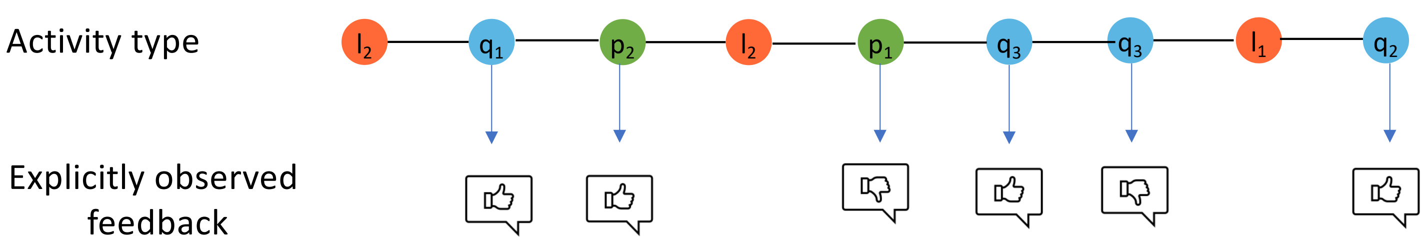
PersAI onalized

NSF

ECML PKDD 2023

General problem

- Modeling
  - fine-grained and coarse-grained transitions
  - in multi-type sequential data
  - between activities with implicit and explicit feedback



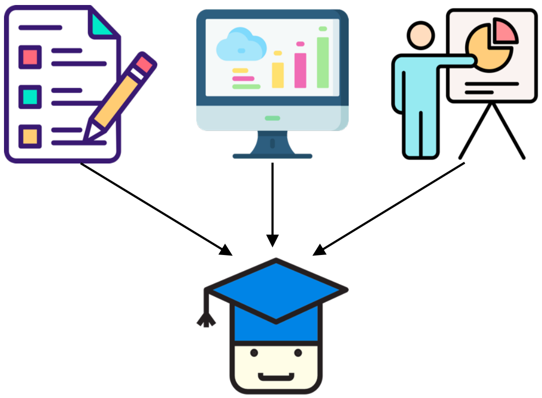
Application

- Online education systems
  - Facilitate distance education
  - Provide diverse courses and learning materials

- Student knowledge tracing (KT)
  - Quantifying student knowledge state, to
    - Predict student performance
    - Create a study plan
    - Recommend learning materials
    - Analyze knowledge gaps

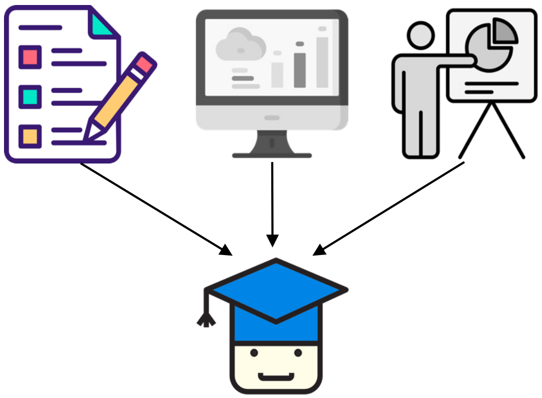
Introduction - Motivation and Limitation

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



Introduction - Motivation and Limitation

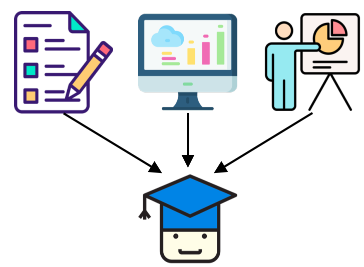
- Student learn by doing multiple types of activities
  - Solve questions (assessed), watch video lectures (non-assessed)



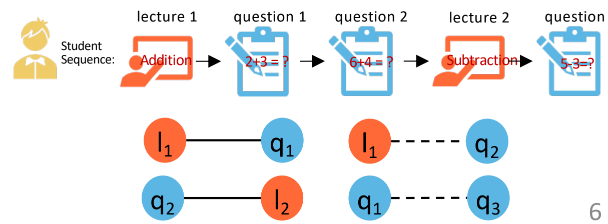
- Traditional KT approaches
  - Examples
    - Regression, e.g. IRT [Frederic et al.], PFA [Philip et al.]
    - Hidden Markov model (HMM), e.g. BKT [Corbett et al.]
    - Recurrent Neural Networks, e.g. DKT [Piech et al.]
  - Model assessed activities only and predict students' performance

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.]
    - Neural Networks, e.g. DMKT [Wang et al.], TAMKOT [Zhao et al.]

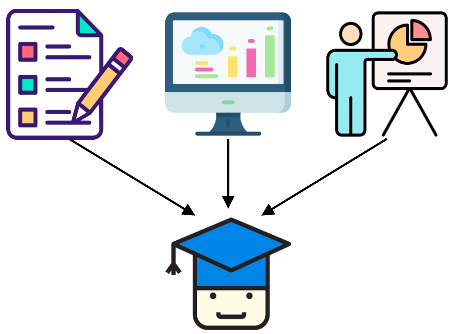


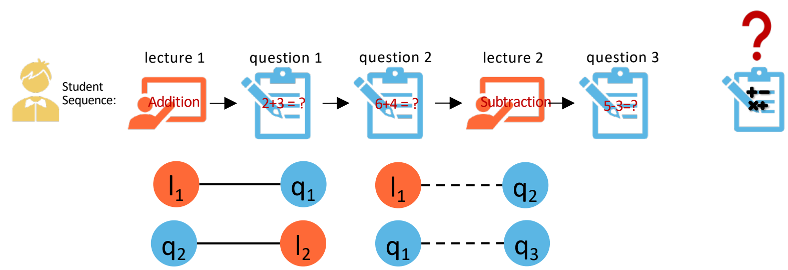
- Supervised sequence learning
    - Non-assessed activities not explicit in the objective function
- Markovian assumptions of materials (fine-grained latent-concept spaces)
    - Lost long-range and coarse-grained associations between materials



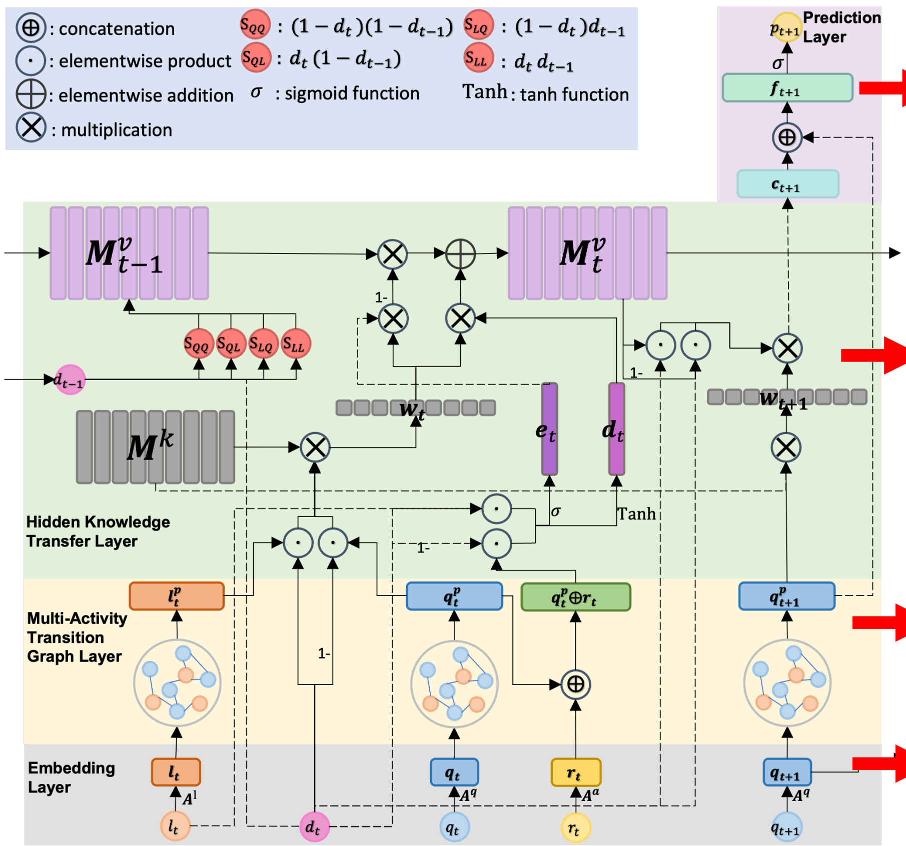
Our solution

- Graph-enhanced Multi-activity Knowledge Tracing (GMKT)
  - Introduce transition-aware graph neural network
  - Formulate as semi-supervised learning problem





GMKT Model – overview



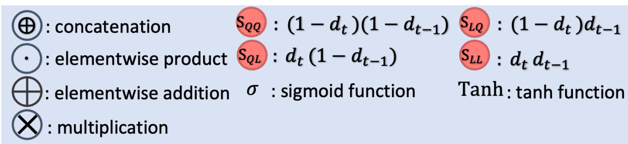
Prediction Layer

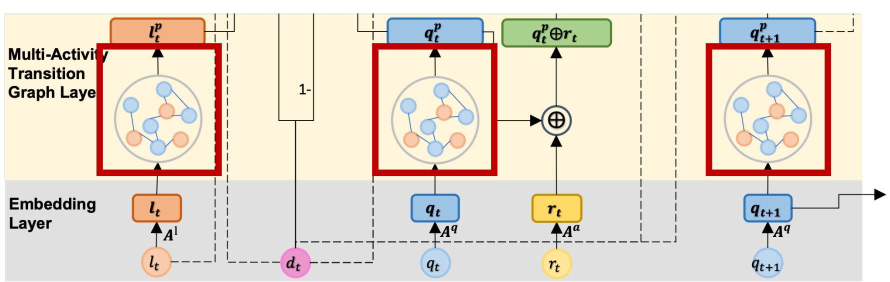
Hidden Knowledge Transfer Layer

Multi-Activity Transition Graph Layer

Embedding layer

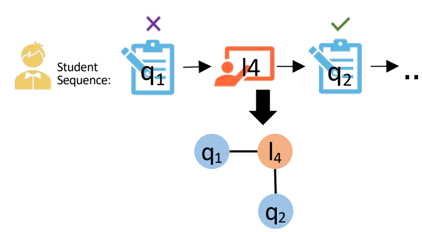
GMKT Model - Multi-Activity Transition Graph



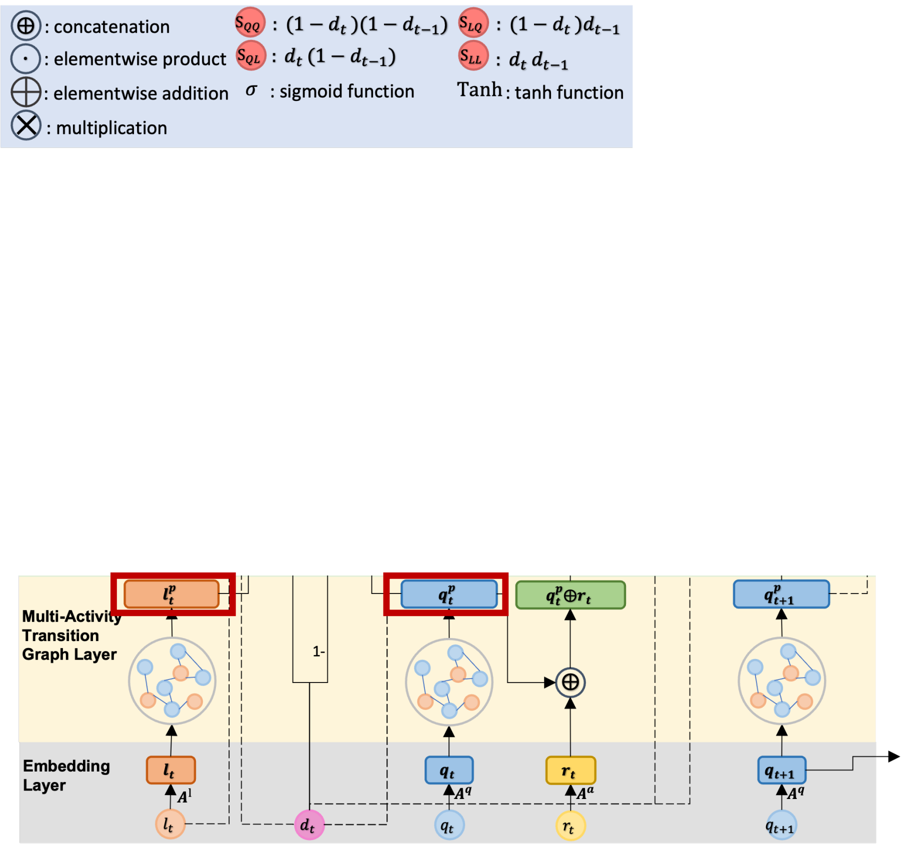


- Incorporates the coarse-grained long-range patterns among materials

- Construct a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
  - $\mathcal{V}$: all assessed and non-assessed materials
  - $\mathcal{E}$: undirected edges between materials
  - An edge exists if a student has consecutively interacted with materials



GMKT Model - Multi-Activity Transition Graph

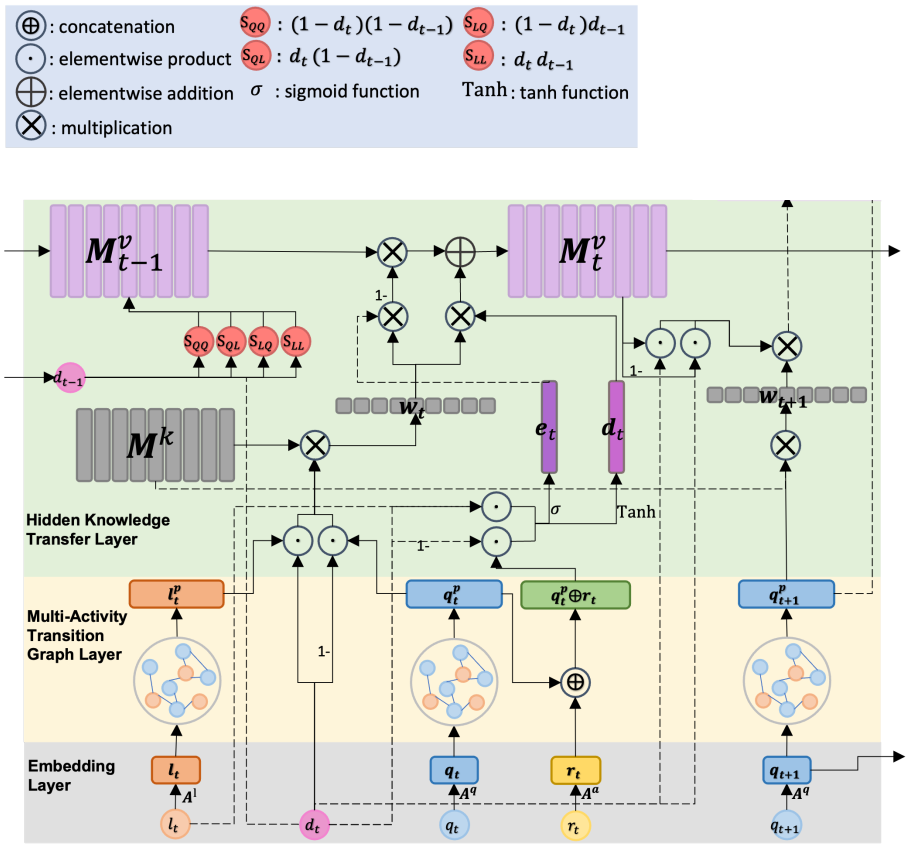


- Multi-Activity GNN to propagate learning material embeddings among neighbors
  - Models how different activity types relate
  - Learns separate transition matrices between different material types

$$\boldsymbol{q}_t^p = V_Q^{\mathrm{T}}\left[\boldsymbol{q}_t + \frac{1}{|\mathcal{N}_{q_t}^Q|}\sum_{i\in\mathcal{N}_{q_t}^Q} \boldsymbol{G}_{QQ}^T \boldsymbol{q}_i + \frac{1}{|\mathcal{N}_{q_t}^L|}\sum_{j\in\mathcal{N}_{q_t}^L} \boldsymbol{G}_{QL}^T \boldsymbol{l}_j\right] + \boldsymbol{b}_Q$$

$$\boldsymbol{l}_t^p = V_L^{\mathrm{T}}\left[\boldsymbol{l}_t + \frac{1}{|\mathcal{N}_{l_t}^L|}\sum_{i\in\mathcal{N}_{l_t}^L} \boldsymbol{G}_{LL}^T \boldsymbol{l}_i + \frac{1}{|\mathcal{N}_{l_t}^Q|}\sum_{j\in\mathcal{N}_{l_t}^Q} \boldsymbol{G}_{LQ}^T \boldsymbol{q}_j\right] + \boldsymbol{b}_L$$

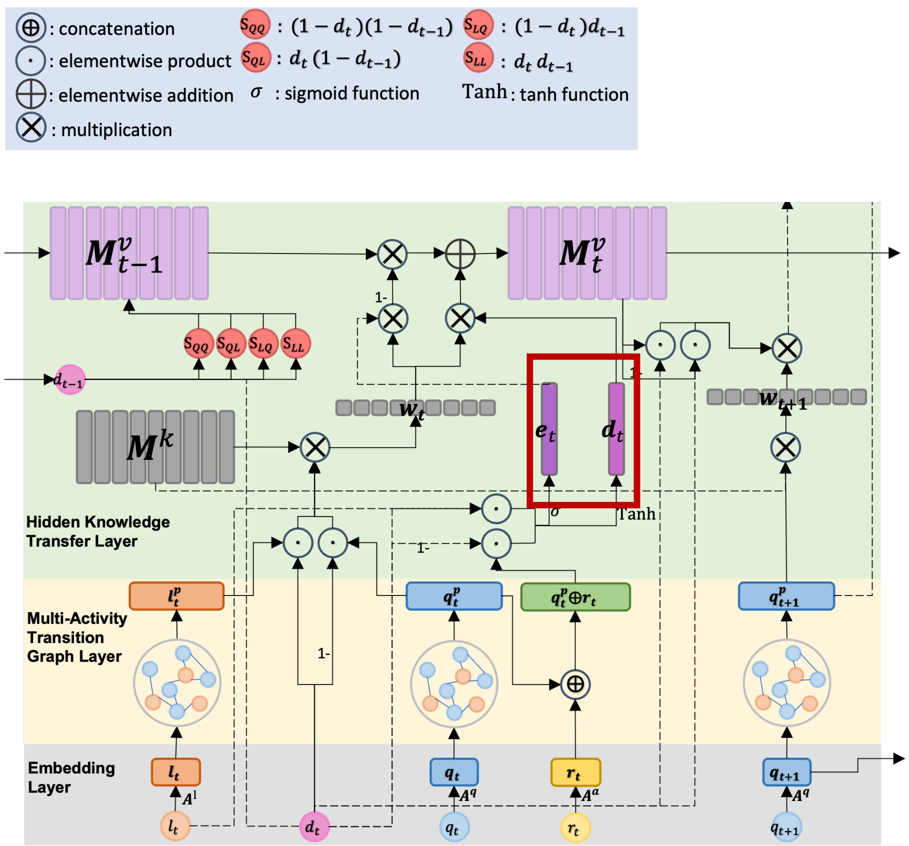
GMKT Model - Hidden Knowledge Transfer Layer



- Represents dynamic and fine grained student knowledge
  - Using 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., $S_{QL} = d_t(1-d_{t-1})$

GMKT Model - Hidden Knowledge Transfer Layer



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

Erase:

$$\boldsymbol{e}_t = \sigma\Big((1-d_t)\cdot \boldsymbol{E}_q^T[\boldsymbol{q}_t^p \oplus \boldsymbol{r}_t] + d_t \cdot \boldsymbol{E}_l^T \boldsymbol{l}_t^p + \boldsymbol{b}_e\Big)$$

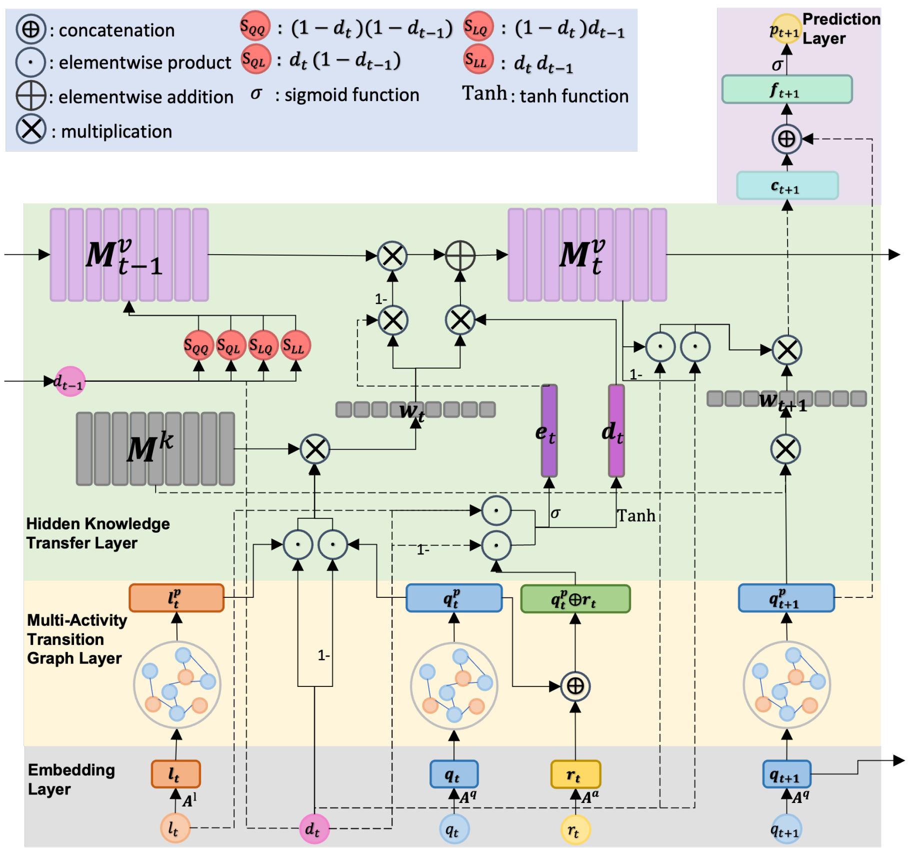
$$\widetilde{\boldsymbol{M}}_t^v(i) = [S_{QQ}\cdot T_{QQ}\boldsymbol{M}_{t-1}^v + S_{LL}\cdot T_{LL}\boldsymbol{M}_{t-1}^v + S_{QL}\cdot T_{QL}\boldsymbol{M}_{t-1}^v + S_{LQ}\cdot T_{LQ}\boldsymbol{M}_{t-1}^v](i)\cdot[\boldsymbol{1} - w_t(i)\boldsymbol{e}_t]$$

Add:

$$\boldsymbol{d}_t = Tanh\Big((1-d_t)\cdot \boldsymbol{D}_q^T[\boldsymbol{q}_t^p \oplus \boldsymbol{r}_t] + d_t \cdot \boldsymbol{D}_l^T \boldsymbol{l}_t^p + \boldsymbol{b}_d\Big)$$

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

GMKT Model - Prediction Layer



- Predicts student performance at q_{t+1}

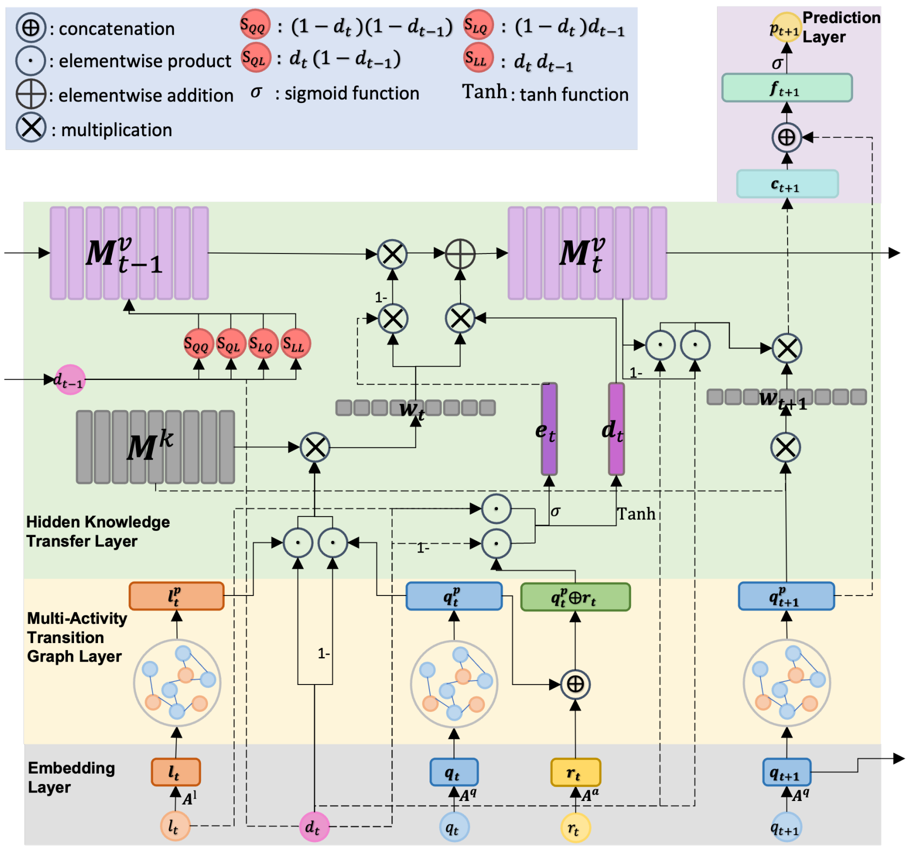
$$w_{t+1}(i) = softmax\left([\boldsymbol{R}_q^T \boldsymbol{q}_{t+1}^p]\boldsymbol{M}^k(i)\right)$$

$$c_{t+1} = \sum_{1}^{N} w_{t+1}(i)\,[(1-d_t)T_{QQ}\boldsymbol{M}_t^v + d_t \cdot T_{LQ}\boldsymbol{M}_t^v](i)$$

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

$$p_{t+1} = \sigma(\boldsymbol{W}_p^T \boldsymbol{f}_{t+1} + b_p)$$

GMKT Model - Objective Function



- Performance prediction objective:

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

- Activity type objective:
  - Learn from the unlabeled data (non-assessed activities)

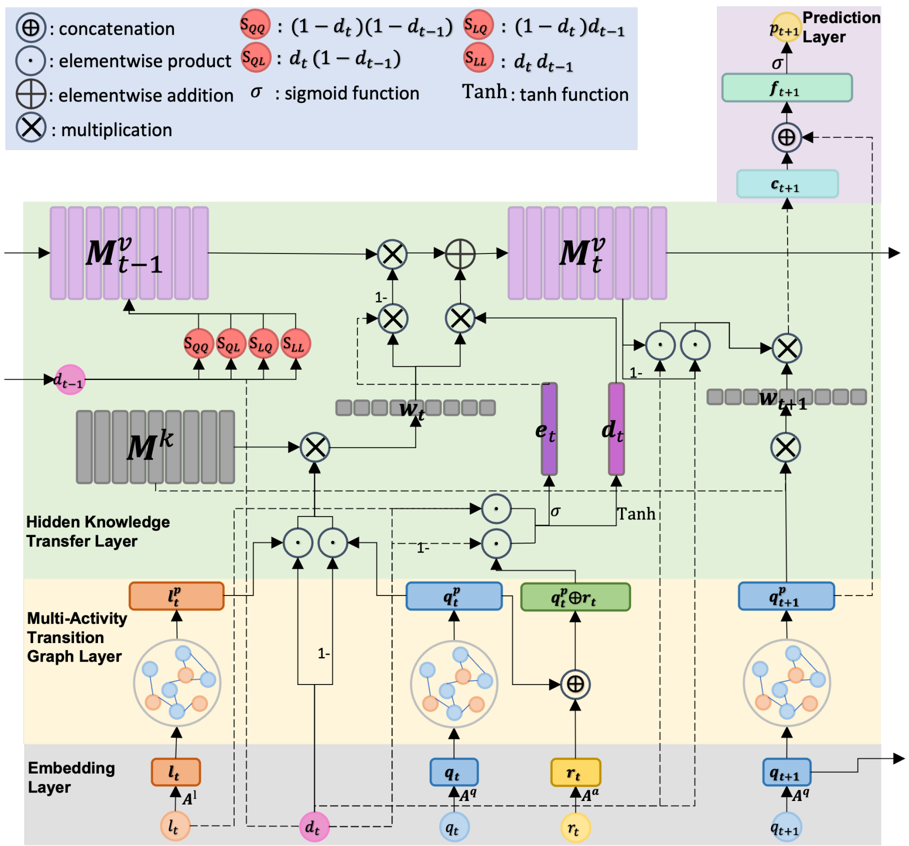
$$w^o_t(i) = softmax\big([(1-d_t)\cdot \boldsymbol{O}_q^T \boldsymbol{q}^p_t + d_t \cdot \boldsymbol{O}_l^T \boldsymbol{l}^p_t]\boldsymbol{M}^k(i)\big)$$

$$c^o_t = \sum_{1}^{N} w^o_t(i)\, \boldsymbol{M}^v_t(i)$$

$$p^o_{t+1} = \sigma(d_t \cdot \boldsymbol{W}^T_{oq} c^o_t + (1-d_t)\cdot \boldsymbol{W}^T_{ol} c^o_t + b_o)$$

$$\mathcal{L}^o = -\sum_t (d_t log p^o_t + (1-d_t)\log(1-p^o_t))$$

GMKT Model - Objective Function



Performance prediction objective:

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

Activity objective objective:

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

Final objective function:

$$\mathcal{L}_{total} = \mathcal{L} + \lambda_o \mathcal{L}^o + \underbrace{\lambda_\theta ||\theta||^2}_{\text{regularization}}$$

Experiments

- Two sets of experiments:
  - **Student Performance Prediction -** to validate if the model captures the variability of student performance
    - **Comparison with Baselines**
    - **Ablation Studies**
    - **Sensitivity Analysis**
  - **Knowledge Transfer Analysis -** to analyze the knowledge transfer between assessed and non-assessed learning material types

Comparison with Baselines

Methods	Ednet	Junyi	MORF
	AUC	AUC	RMSE
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**
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**
MVKM	–	–	0.1936*
DMKT	0.6394**	0.8561**	0.1856*
TAMKOT	0.6786	0.8745**	0.1857*
GMKT	**0.6819**	**0.8960**	**0.1802**

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

Comparison with Baselines

Methods	Ednet	Junyi	MORF
	AUC	AUC	RMSE
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**
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**
MVKM	–	–	0.1936*
DMKT	0.6394**	0.8561**	<u>0.1856*</u>
TAMKOT	<u>0.6786</u>	<u>0.8745**</u>	0.1857*
GMKT	**0.6819**	**0.8960**	**0.1802**

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

GMKT:

- outperforms all baseline models
  - more improvement in datasets w. less material associations and more transition variability
- outperforms MVKM and DMKT (multi-activity methods w/o knowledge transfer)
  - Highlights the importance of explicitly modeling activity-type transitions
- outperforms all the 'Multi-activity' setting models
  - Simply adding non-assessed activities does not help

Ablation Studies

remove the **GNN** component

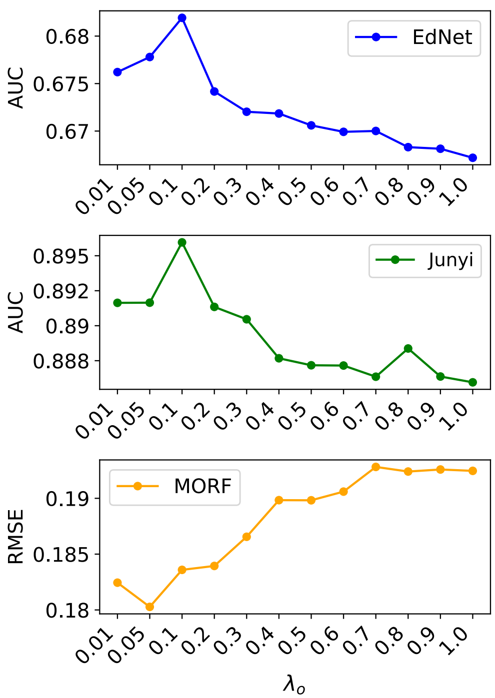
Methods	Ednet	Junyi	MORF
	AUC	AUC	RMSE
GMKT-G	0.6759	0.8909	0.1888
GMKT-O	0.6761	0.8911	0.1867
GMKT	**0.6819**	**0.896**	**0.1802**

Ablation study results

remove the **type objective**

- Removing either of components decreased the performance
- Modeling coarse-grained transitions (GNN) is more important for MORF
  - More complex materials and less type transition variability

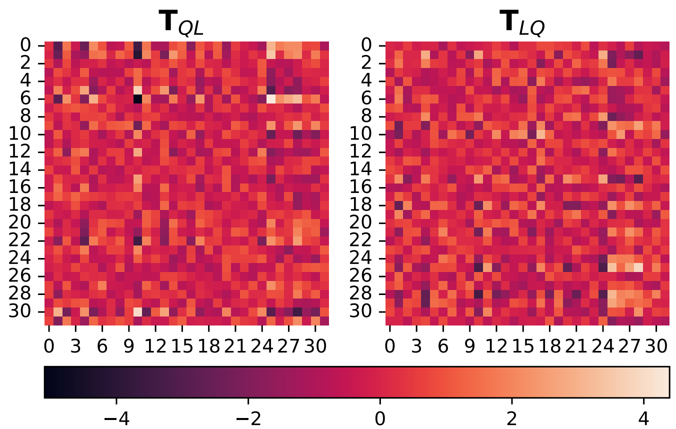
Sensitivity Analysis – The impact of type objective



Performance w.r.t. λ_o

- Performance initially improves, but gradually decreases after a certain λ_o value
  - The type objective helps, but, a balance is necessary
- Only a slight variation in the best λ_o across datasets

Knowledge Transfer Analysis –transfer from question to lecture (T_{QL}) vs. lecture to question (T_{LQ})

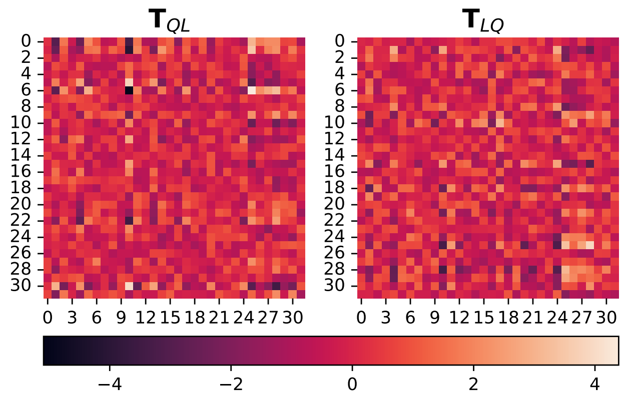


Transition weight matrices for MORF

Weight matrices are considerably different from each other

- Indicates that Knowledge transfer weights depend on the order of transition

Knowledge Transfer Analysis –transfer from question to lecture (T_{QL}) vs. lecture to question (T_{LQ})



- Spearman correlation
  - Small correlation and > 0.1 p-values

	EdNet	Junyi	MORF
Correlation	0.0357	-0.0128	-0.0504
p-value	0.2531	0.4120	0.1072

Spearman correlation coefficients with p-values

Conclusions

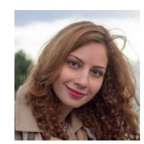
- Graph-enhanced Multi-activity Knowledge Tracing model (GMKT)
    - Can accurately represent student knowledge and predict their performance

- Models transition-aware knowledge transfers between activity types
- Captures materials’ coarse-grained associations by the transition-aware GNN
- Formulates multi-objective learning with implicit and explicit feedback

UNIVERSITY AT ALBANY
State University of New York

PersAI onalized

Thank you!
Q & A



szhao2@albany.edu ssahebi@albany.edu

Our code and supplementary material are available at GitHub: https://github.com/persai-lab/2023-ECML-PKDD-GMKT

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