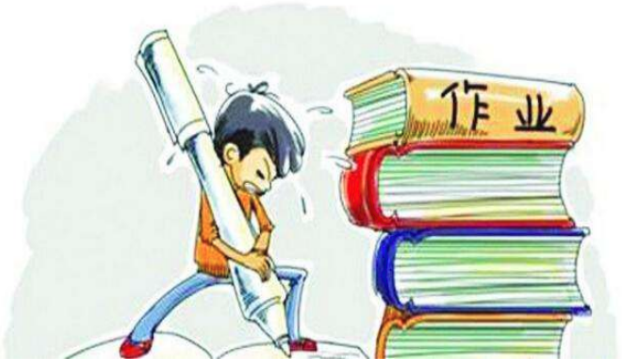
EKT: Exercise-aware Knowledge Tracing for Student Performance Prediction

Qi Liu, Zhenya Huang, Yu Yin, Enhong Chen, Hui Xiong, Yu Su and Guoping Hu

Outline

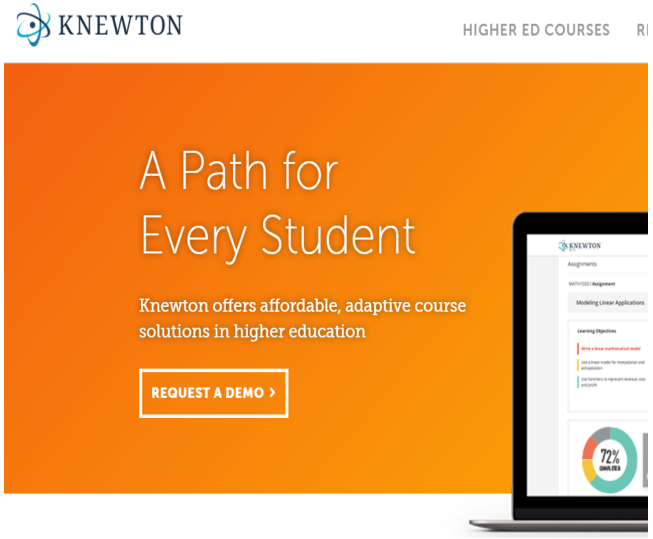
Background

- Traditional Learning
  - Classroom & Homework & Examination
- Limitations
  - Resources
  - Share
  - Personalized



Background

- Online Education Systems
  - MOOC, ITS, OJ

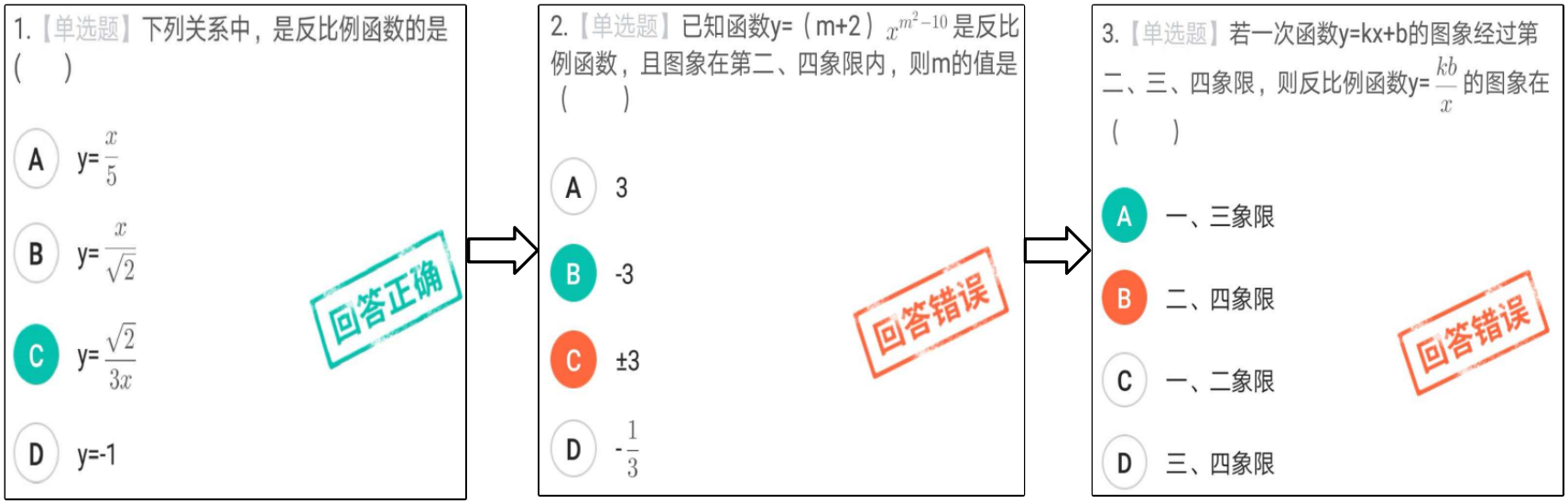


松鼠AI 1对1

TAL 好未来

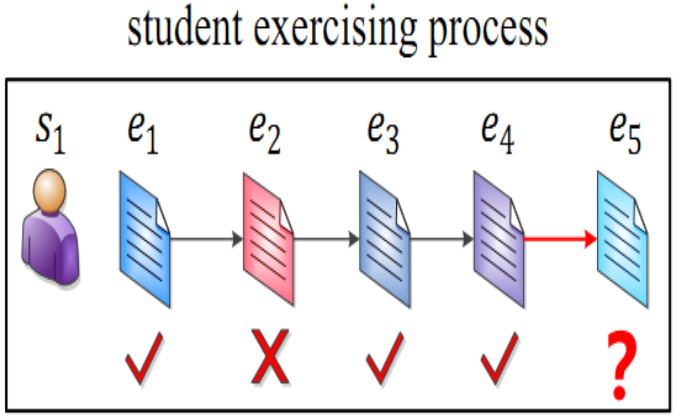
Background

- Student can choose exercises individually according to their needs and acquire necessary knowledge during exercising



Background

- A fundamental problem
  - Predict student performance in the future



Exercise	Exercise Content	Knowledge Concept
e_1	If function $f(x) = x^2 - 2x + 2$ and $x \in [0,3]$, What is the range of $f(x)$?	Function
e_2	If four numbers are randomly selected without replacement from set $\{1,2,3,4\}$, what is the probability that the four numbers are selected in ascending order?	Probability
e_3	What is the y-intercept of the graph of equation $y = 2 \times \|4 \times x - 4\| - 10$?	Function
e_4	What is the value of x If the inequality $\frac{2x-1}{x+2} \leq 3$?	Inequality
e_5	If function $f(x) = 2x - 2$ and $\frac{2x-1}{3x+2} \leq 4$, what is the range of $f(x)$?	Function, Inequality

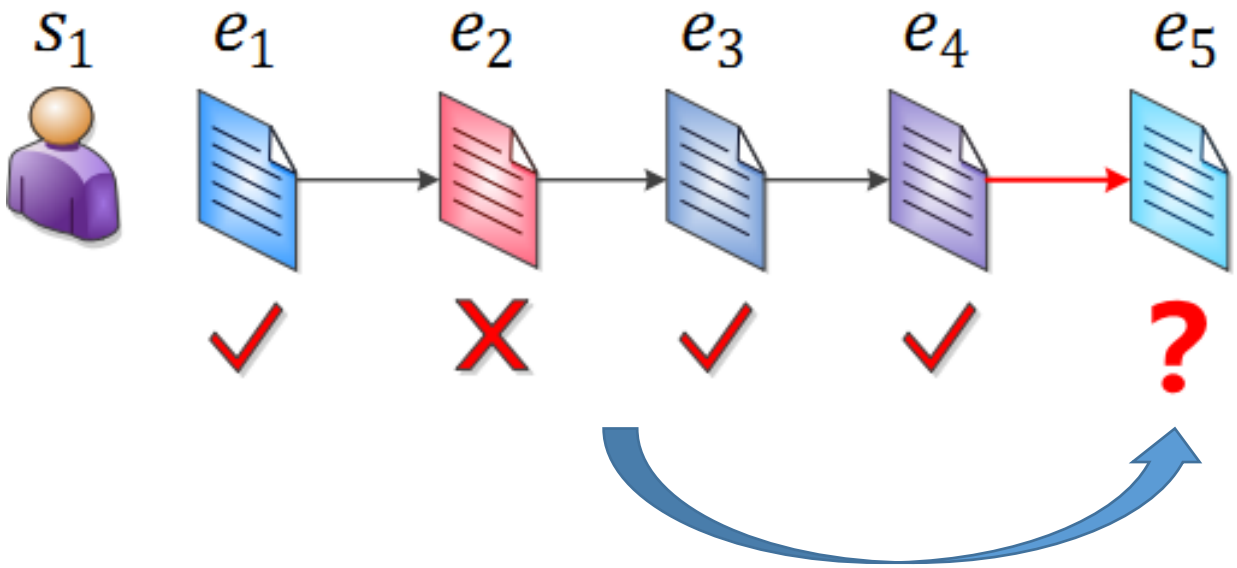
Challenge 1

- Requires a unified way to automatically understand and represent exercises from a semantic perspective
  - Diverse text expressions of exercises

	Exercise Texts	Knowledge Concept
1	**Can you guess the texts of the 3 exercises?**	Function
2		Function
3		Function

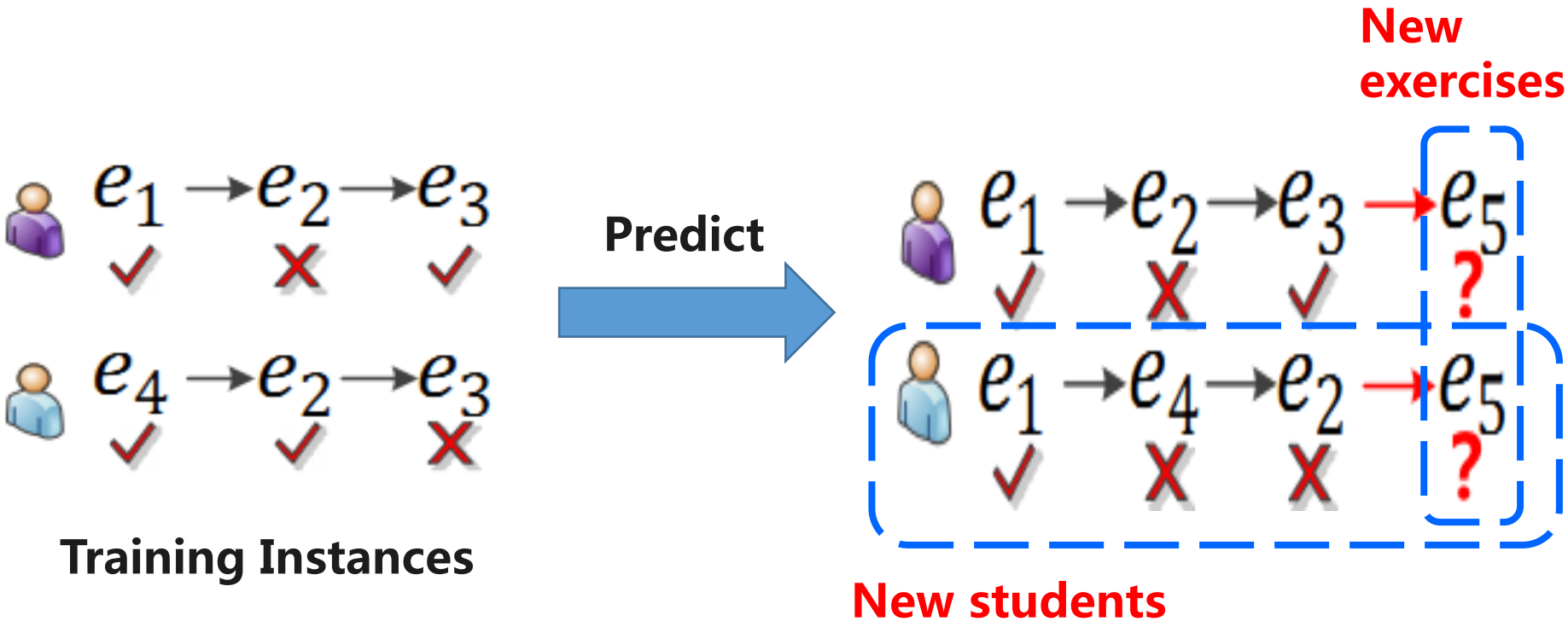
Challenge 2

- How to track the historically focused information for the exercising records of students
  - Long-term historical exercising



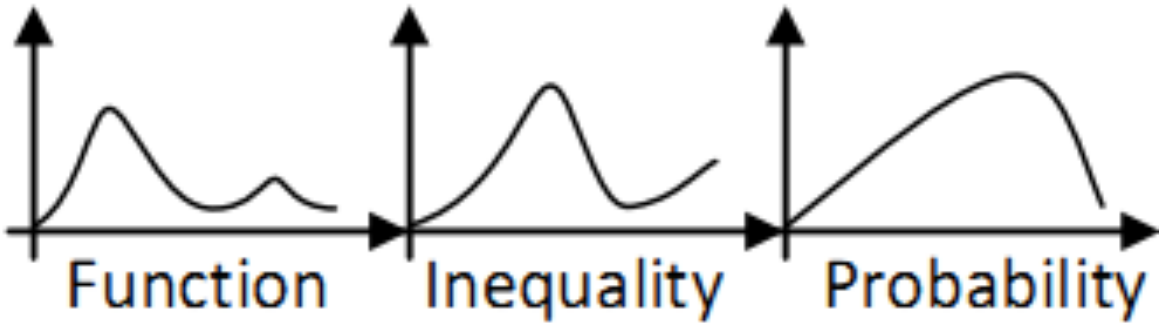
Challenge 3

- Cold start problem
  - We have to make predictions for new students and new exercises



Challenge 4

- Tracking knowledge acquisition
  - Students usually care about not only what they need to learn but also wonder why they need it.
  - Remind them how much they have already learned about each knowledge concepts.



Related Work

- Cognitive Diagnosis
  - IRT: Item Response Theory

$$P(X_{ij}=1|\theta_j)=c_i+\frac{1-c_i}{1+\exp[-1.7a_i(\theta_j-b_i)]}$$

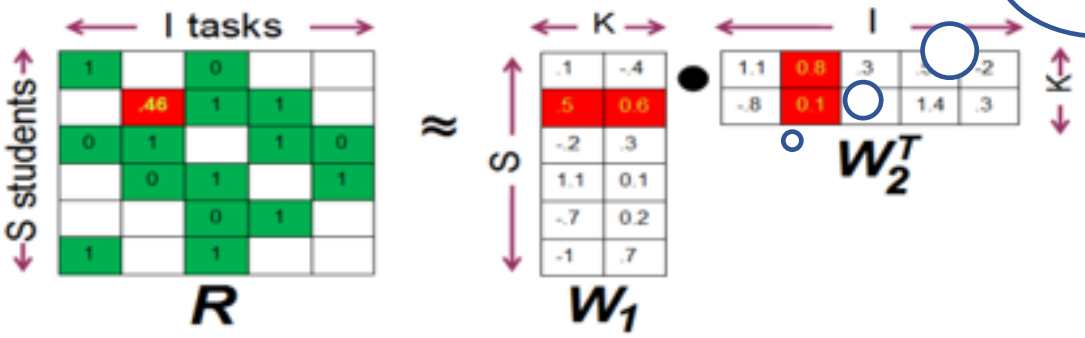
  - DINA: Deterministic Inputs, Noisy-And gate model

$$P(Y_{ij}=1\mid\boldsymbol{\alpha}_i)=(1-s_j)^{\eta_{ij}}(g_j)^{1-\eta_{ij}}$$

- Matrix Factorization
  - projects students and exercises into latent f

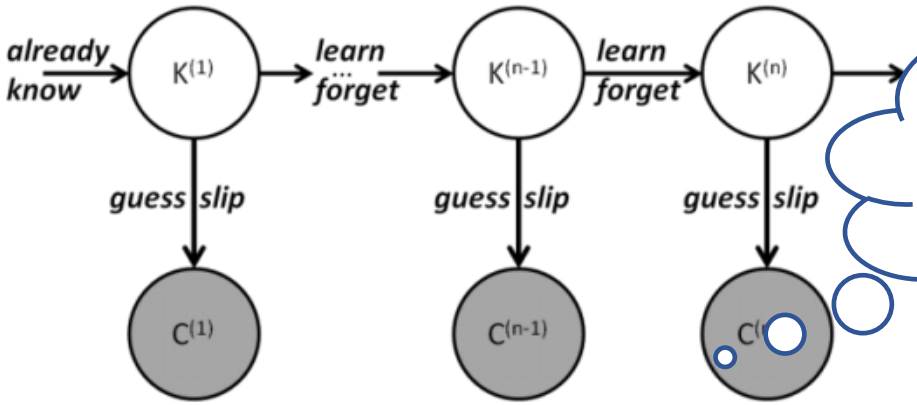
Limited Performance

Lack of Interpretability



Related Work

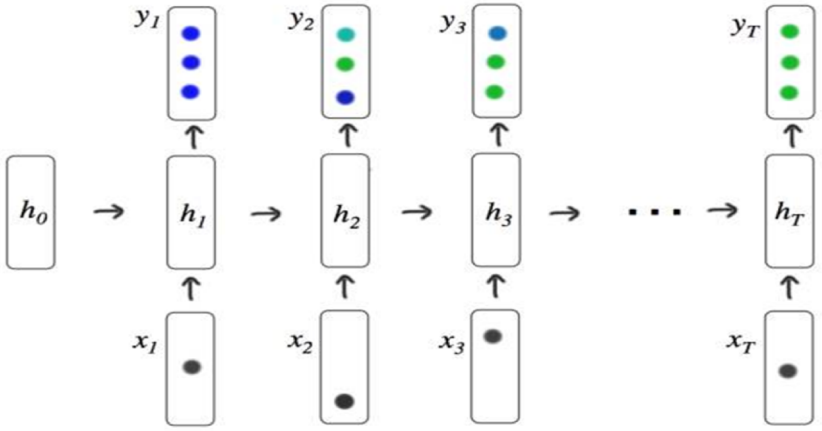
- Bayesian Knowledge Tracing



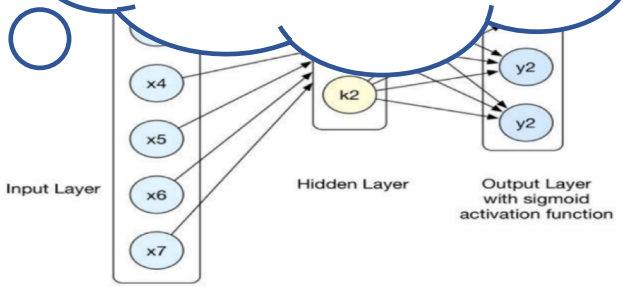
1. Single skill tracing
2. Mastered or non-mastered results

Figure 1: Single-skill knowledge tracing architecture

- Deep Knowledge Tracing



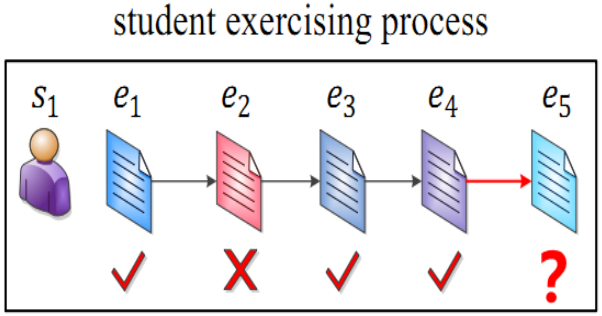
Cannot distinguish exercises without content



Outline

Problem Definition

- Given:
  - Student exercising sequence: $s = \{(k_1, e_1, r_1), (k_2, e_2, r_2), \ldots, (k_T, e_T, r_T)\}$
  - Exercise Content: word sequence: $e = \{w_1, w_2, \ldots, w_M\}$
  - Knowledge Concept: $k \in K$

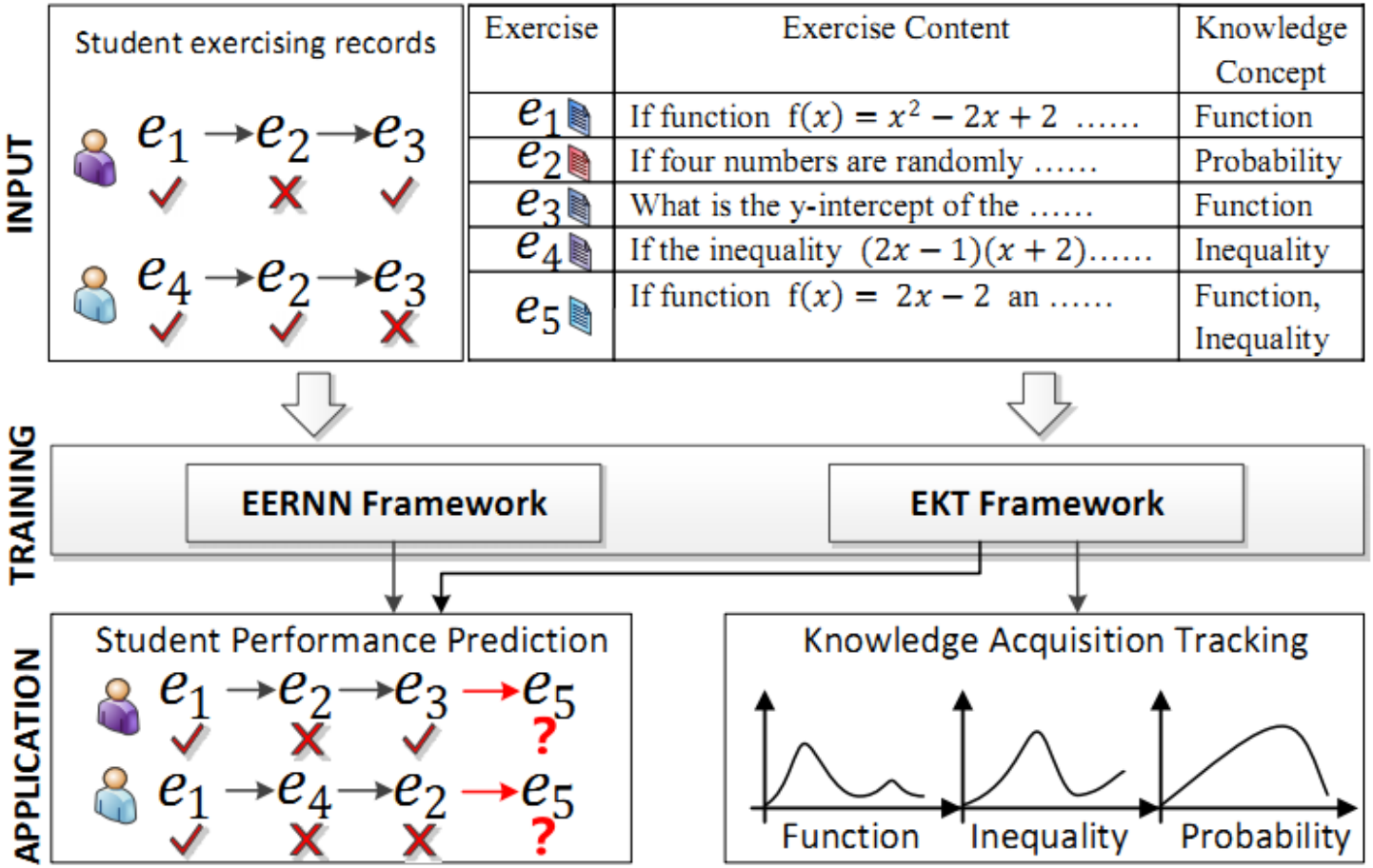


Exercise	Exercise Content	Knowledge Concept
e_1	If function $f(x) = x^2 - 2x + 2$ and $x \in [0,3]$, What is the range of $f(x)$?	Function
e_2	If four numbers are randomly selected without replacement from set $\{1, 2, 3, 4\}$, what is the probability that the four numbers are selected in ascending order?	Probability
e_3	What is the y-intercept of the graph of equation $y = 2 \times \lvert 4 \times x - 4 \rvert - 10$?	Function
e_4	What is the value of x If the inequality $\frac{2x-1}{x+2} \leq 3$?	Inequality
e_5	If function $f(x) = 2x - 2$ and $\frac{2x-1}{3x+2} \leq 4$, what is the range of $f(x)$?	Function, Inequality

- Goal:
  - Track the mastery level of student's knowledge states on K concepts
  - Predict student performance on future exercises e_{T+1}

Study Overview

- Overview solution



Outline

EKT framework

- Exercise-aware Knowledge Tracing
  - Learning a unified exercise representations from its text/formula content

    — Exercise Embedding Module
  - Exploring the impacts of each exercise on improving student states from exercise's knowledge concepts

    — Knowledge Embedding Module
  - Modeling student exercising states with LSTM architecture

    — Student Embedding Module
  - Two prediction strategies
    - EKTM with Markov property
    - EKTA with Attention mechanism

Challenge 3: Cold-start problem

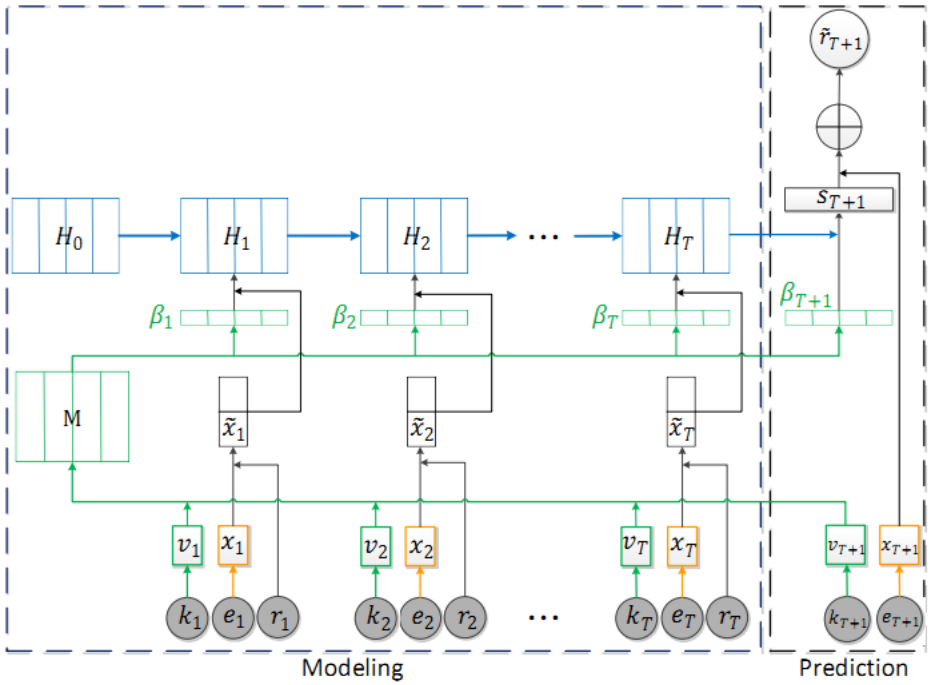
Challenge 1: content presentation

Challenge 4: knowledge tracking

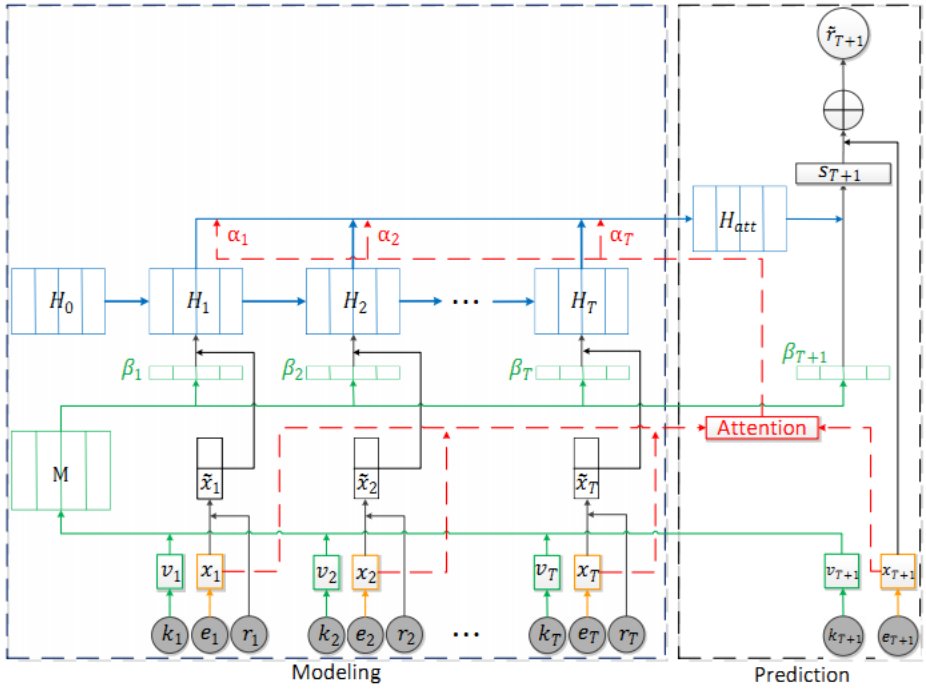
Challenge 2: Long-term focused states

EKT framework

- Framework architecture
  - EKTM with Markov property
  - EKTA with Attention mechanism
  - Both have same modeling process and different prediction strategies



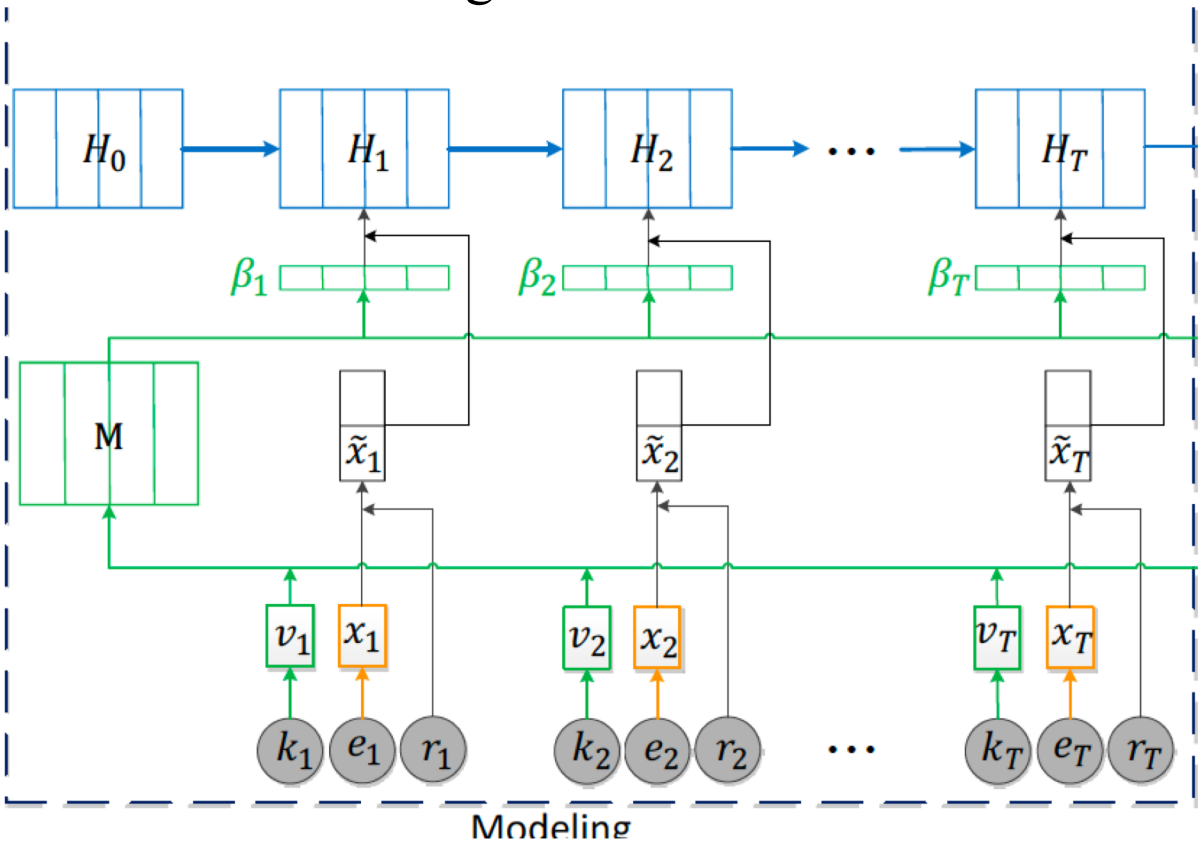
(a) EKTM with Markov property



(b) EKTA with Attention mechanism

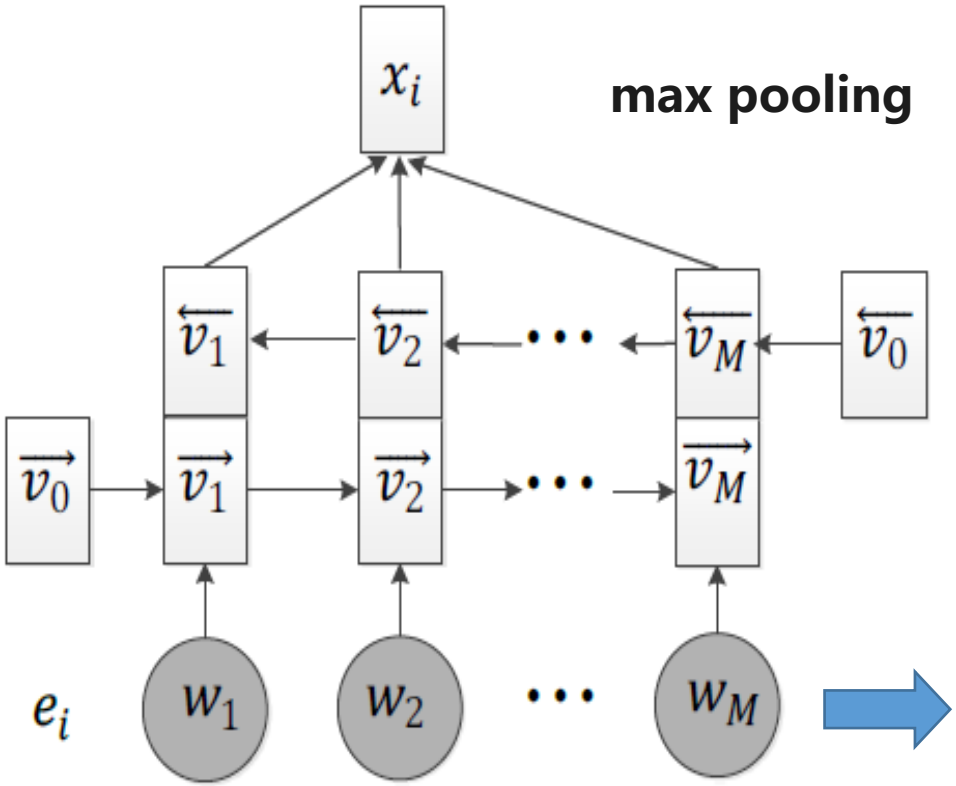
EKT framework

- Modeling process
  - Orange: Exercise content Embedding
  - Green: Knowledge Embedding
  - Blue: Student Embedding



EKTM: Step 1

- Exercise Embedding Module
  - Goal: learns the semantic representation of each exercise x_i from its text content e_i.



max pooling $x_i = \max(v_1, v_2, \ldots, v_M)$

BiLSTM

$v_m = concatenate(\overrightarrow{v}_m, \overleftarrow{v}_m)$

word: word2vec

formula: Tex code features

EKTM: Step 2

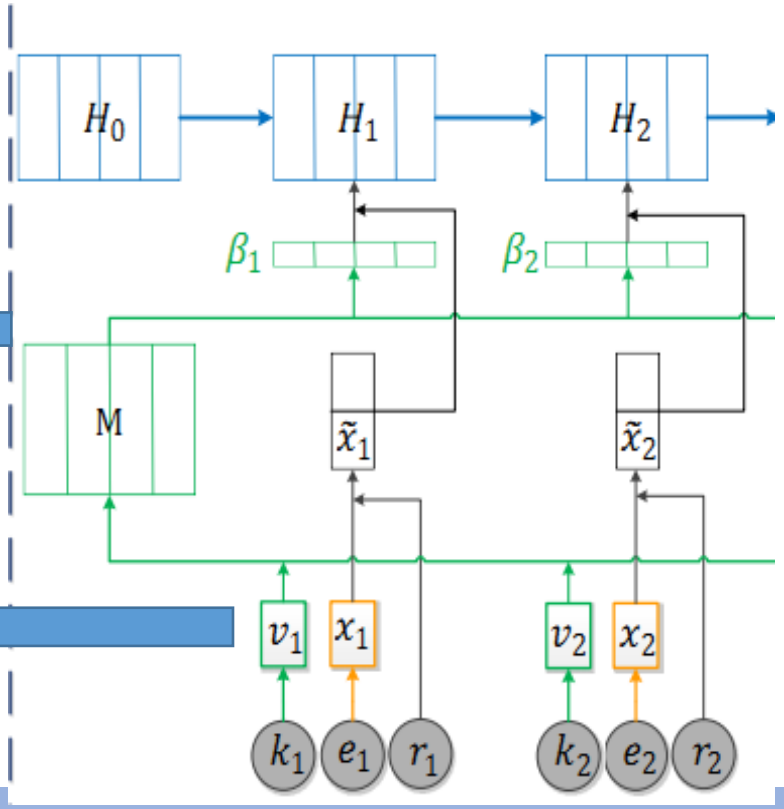
- Knowledge Embedding Module
  - Goal: Exploring the impacts β_t of each exercise on improving student states from exercise's knowledge concepts k_t
  - Intuition: Knowledge concepts are not isolated but contain correlations
  - Assumption: learning one concept could affect the acquisition of other ones

Knowledge impacts

$$\beta_t^i = \text{Softmax}(v_t^{\mathrm{T}}\mathbf{M}_i) = \frac{\exp(v_t^{\mathrm{T}}\mathbf{M}_i)}{\sum_{i=1}^{K}(\exp(v_t^{\mathrm{T}}\mathbf{M}_i))}.$$

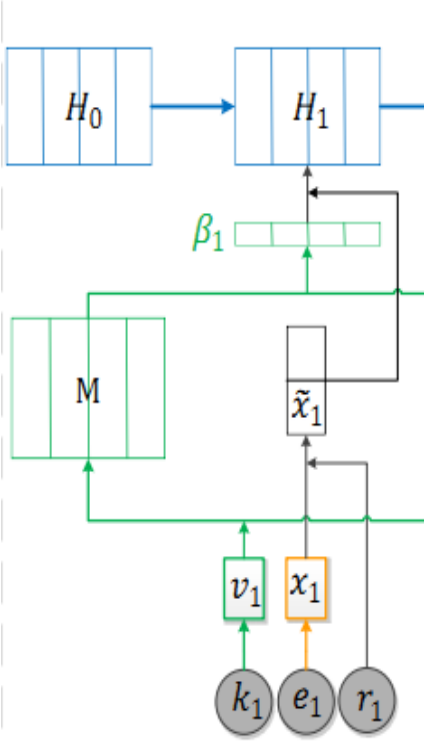
Knowledge initialization $v_t = \mathbf{W_k}^{\mathrm{T}} k_t$

k: one-hot encoding



EKTM: Step 3

- Student Embedding Module
  - Goal: modeling exercising process and learning the student states considering
    - Exercise content x_i
    - Knowledge impacts β_t
    - Score r_t

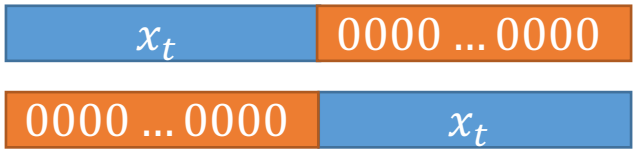


Modeling exercising process $H_t^i = LSTM(\tilde{x}_t^i, H_{t-1}^i; \theta_{H^i})$,

Combine with knowledge $\tilde{x}_t^i = \beta_t^i \tilde{x}_t$,

Combine content and score

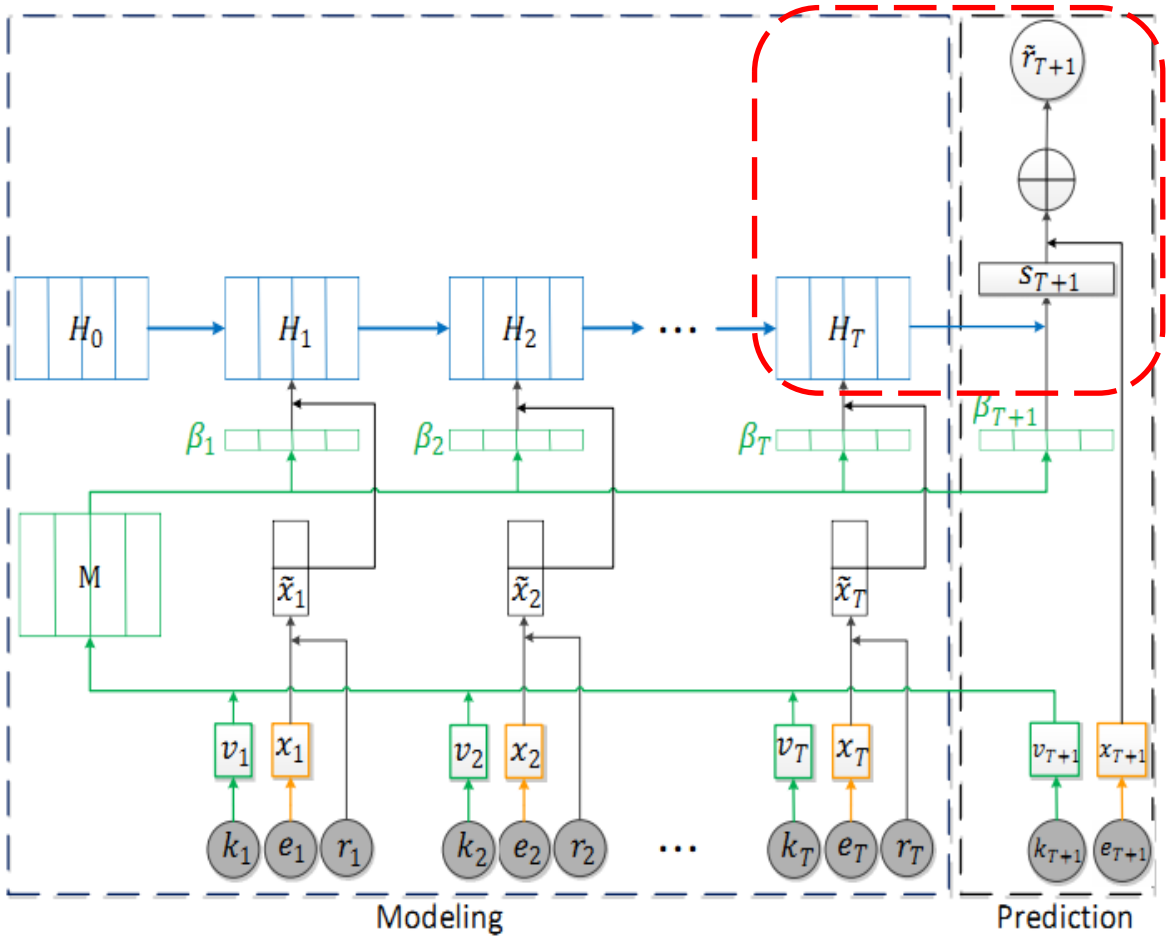
$$\tilde{x}_t = \begin{cases} [x_t \oplus 0] & \text{if} \quad r_t = 1, \\ [0 \oplus x_t] & \text{if} \quad r_t = 0, \end{cases}$$



students getting right response and wrong response to the same exercise actually reflect their different states

EKTM

- EKTM with Markov property
  - Assumption: student next state only depends on the current state

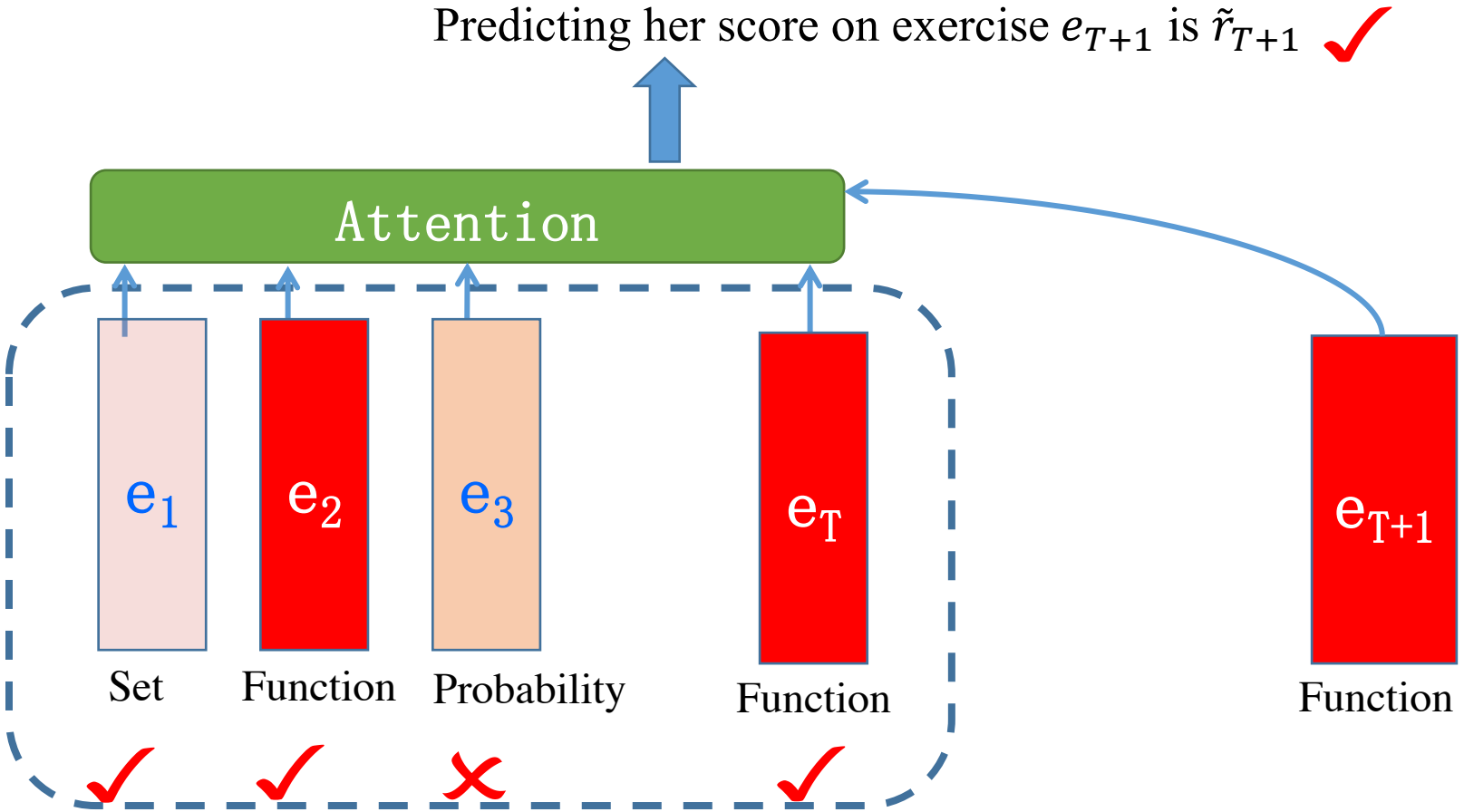


$$y_{T+1} = ReLU(\mathbf{W_3} \cdot [s_{T+1} \oplus x_{T+1}] + \mathbf{b_3}),$$
$$\tilde{r}_{T+1} = \sigma(\mathbf{W_4} \cdot y_{T+1} + \mathbf{b_4}),$$

$$s_{T+1} = \sum_{i=1}^{K} \beta_{T+1}^{i} H_T^i,$$

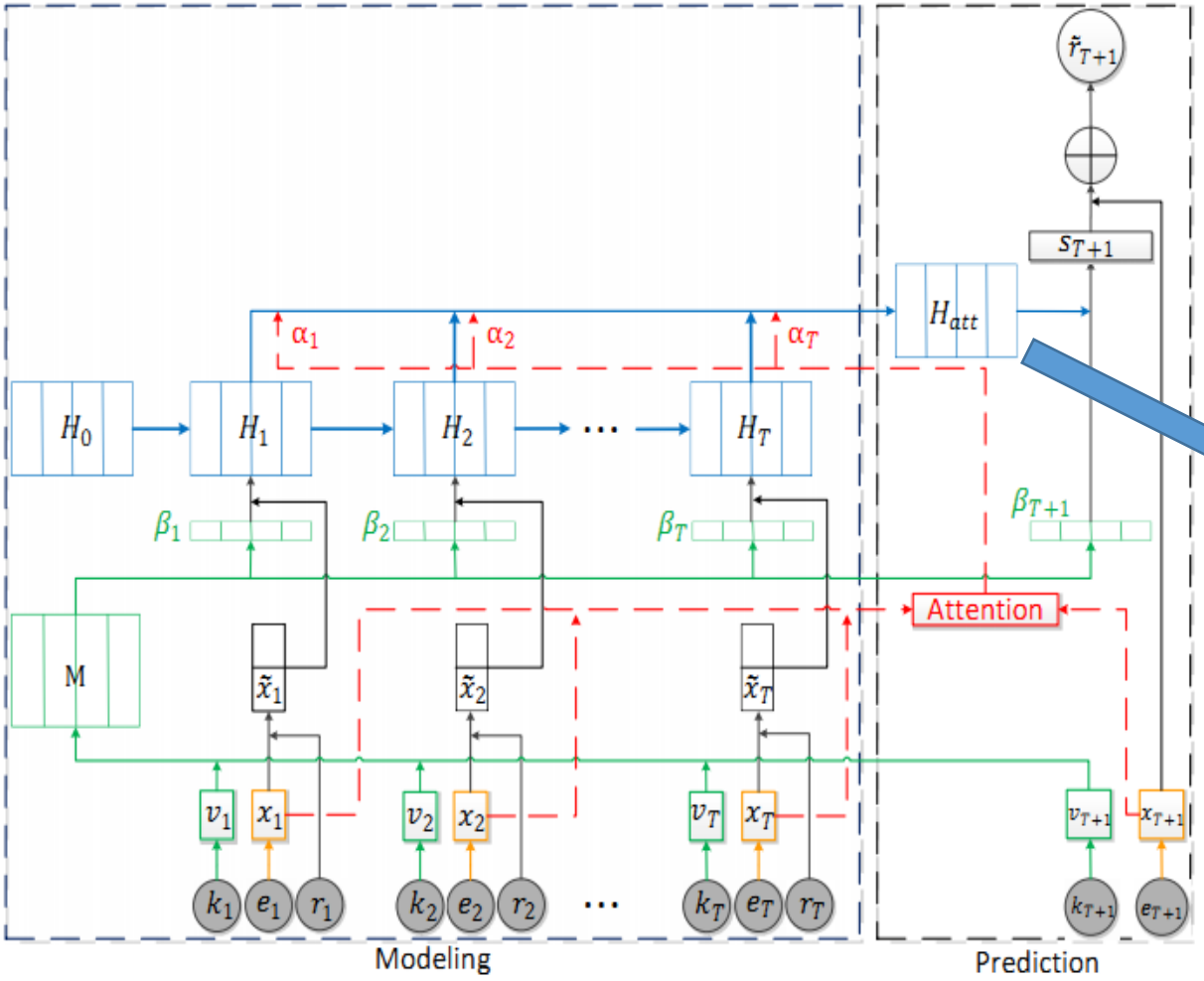
EKTM

- EKTM with Markov property
  - Problem: Vanish problem, ignoring the effects of historical states
  - Intuition: Students may get similar scores on similar exercises



EKTA

- EKTA with Attention mechanism
  - Assumption: student next state depends on the aggregated focused states



$$H_{att}^{i} = \sum_{j=1}^{T} \alpha_j H_j^i,\ \alpha_j = cos(x_{T+1}, x_j).$$

Applications

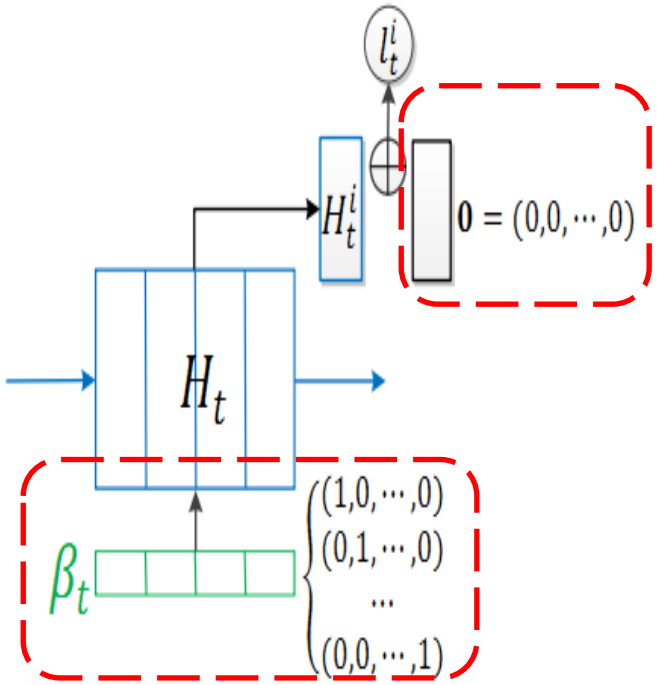
- Student performance prediction
  - Given: an individual exercising record

$$s^p = \{(k_1^p, e_1^p, r_1^p), (k_2^p, e_2^p, r_2^p), \ldots, (k_T^p, e_T^p, r_T^p)\}$$

  - Steps:
    - Apply model EKTM(A) to fit exercising process s^p to get the student state H_T^p (H_{att}^p) at step T
    - Extract exercise representation x_{T+1}^p and knowledge impact β_{T+1} with Exercise Embedding and Knowledge Embedding modules
    - Predict performance $\tilde{r}_{T+1}^p$
  - Cold start problems
    - Exercises can be new exercises
    - Students can be new students

Applications

- Knowledge Acquisition Tracking
  - Estimate her mastery of the i-th specific concept without any exercise input
    - Omit the input exercise embedding x_t
    - Construct the impact weight $\beta_t = (0, 0, \cdots, 1, 0, \cdots, 0)$



$$y_t^i = ReLU(\mathbf{W_3} \cdot [H_t^i \oplus \mathbf{0}] + \mathbf{b_3}),$$
$$l_t^i = \sigma(\mathbf{W_4} \cdot y_t^i + \mathbf{b_4}),$$

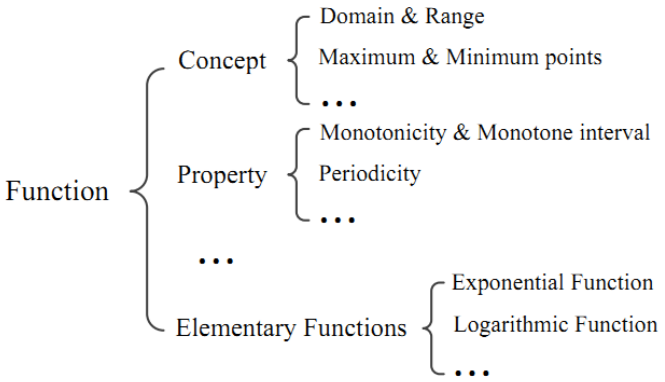
Outline

Experiments

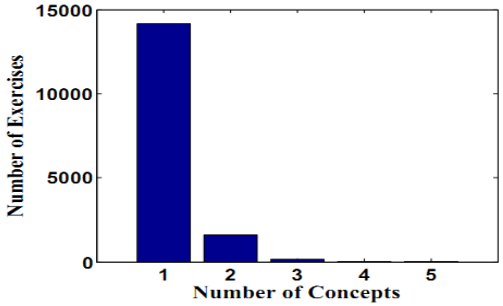
- Experiments dataset
  - Mathematical records of high school students

TABLE 1
The statistics of mathematics dataset.

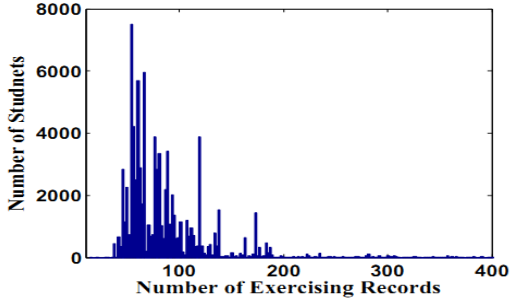
Statistics	Original	Pruned
# of records	68,337,149	5,596,075
# of students	1,100,726	84,909
# of exercises	1,825,767	15,045
# of knowledge concepts	37	37
# of knowledge features	550	447
Avg. exercising records per student	\	65.9
Avg. content length per exercise	\	27.3
Avg. knowledge concepts per exercise	\	1.12
Avg. knowledge features per exercise	\	1.8
Avg. exercises per knowledge concept	\	406.6



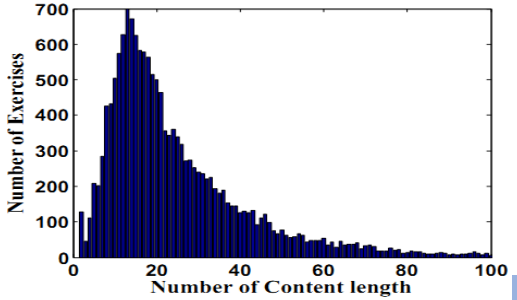
- Dataset analysis
  - Most exercises contain less than 2 knowledge concepts and features
  - One specific knowledge concept is related to 406 exercises on average
  - The average content length of each exercise is about 27



(a) Distribution of concepts



(b) Distribution of exercising records



(c) Distribution of content length

Experiments

- Baseline methods

TABLE 2
Characteristics of all models.

Model	Data Source			Prediction Scenario		Knowledge
	Score	Concept	Content	General	Cold-start	Tracking?
IRT [10]	✓	✗	✗	✓	✗	✗
BKT [7]	✓	✓	✗	✓	✗	✓
PMF [42]	✓	✗	✗	✓	✗	✗
DKT [34]	✓	✓	✗	✓	✓	✗
DKVMN [53]	✓	✓	✗	✓	✓	✓
LSTMM	✓	✓	✗	✓	✓	✗
LSTMA	✓	✓	✗	✓	✓	✗
EERNNM [38]	✓	✗	✓	✓	✓	✗
EERNNA [38]	✓	✗	✓	✓	✓	✗
EKTM	✓	✓	✓	✓	✓	✓
EKTA	✓	✓	✓	✓	✓	✓

- Evaluation metrics
  - Regression perspective: **RMSE**
  - classification perspective: **ACC**, **AUC**

Experiments

- Prediction Performance

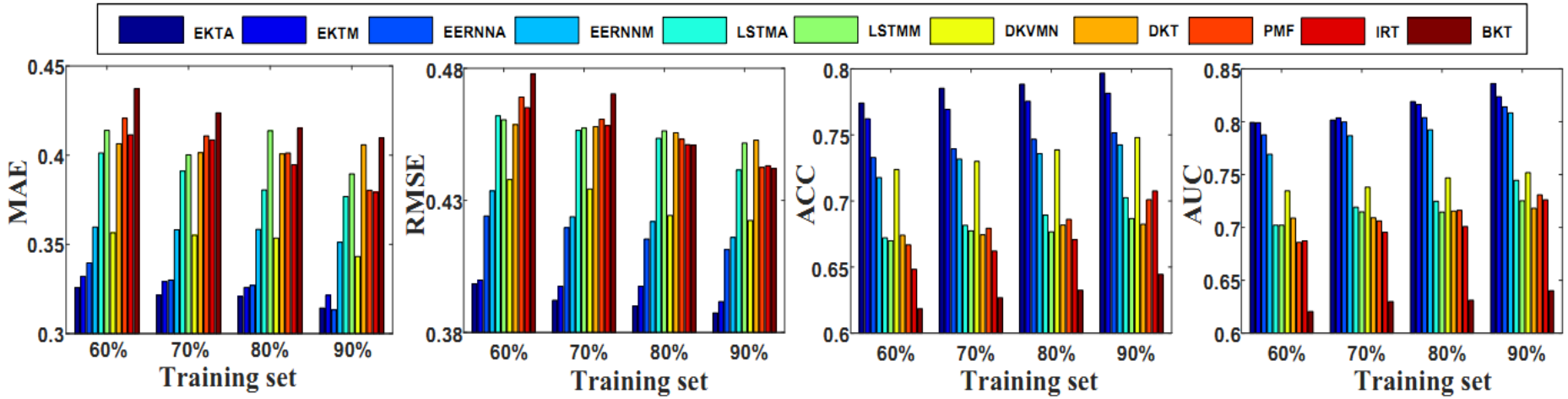


Fig. 9. Results of student performance prediction in general scenario under four metrics.

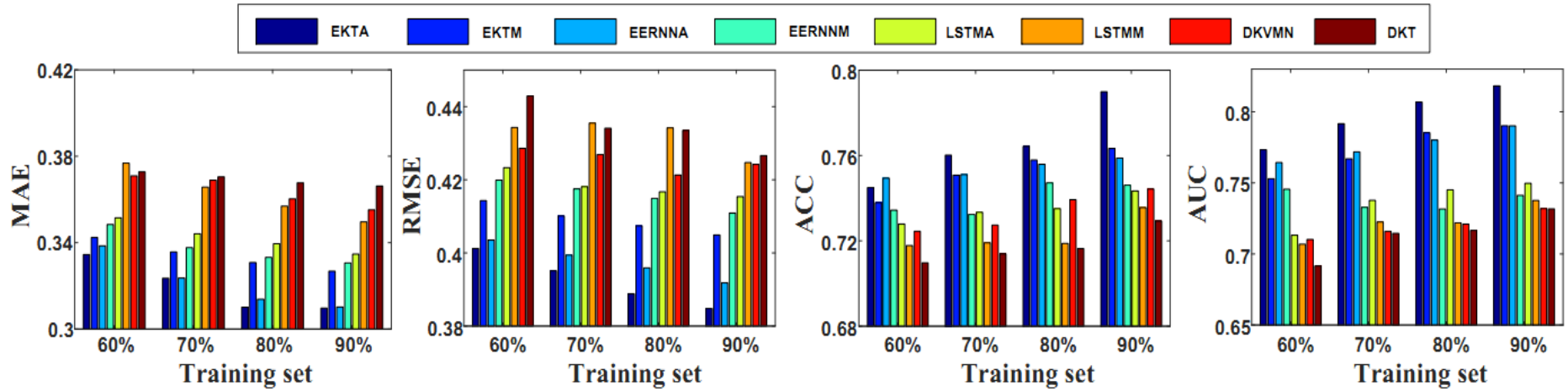
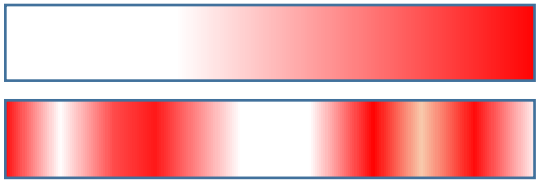


Fig. 10. Results of student performance prediction on cold-start (new) exercises under four metrics.

Experiments

- Attention Effectiveness

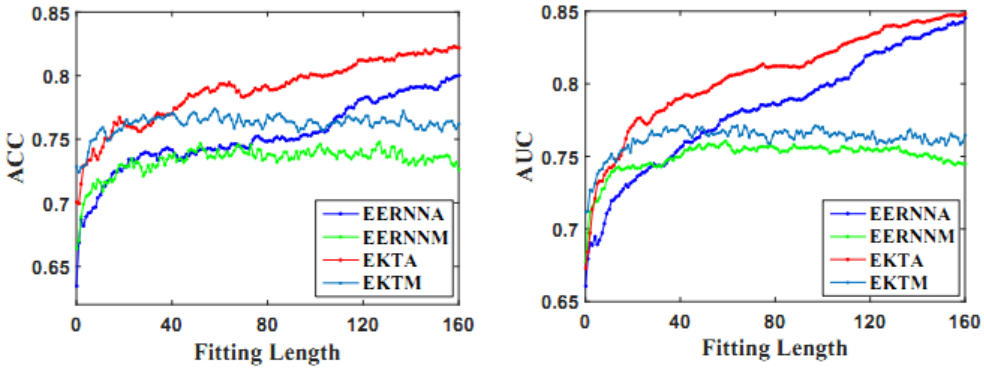
historical exercising states



EKTM

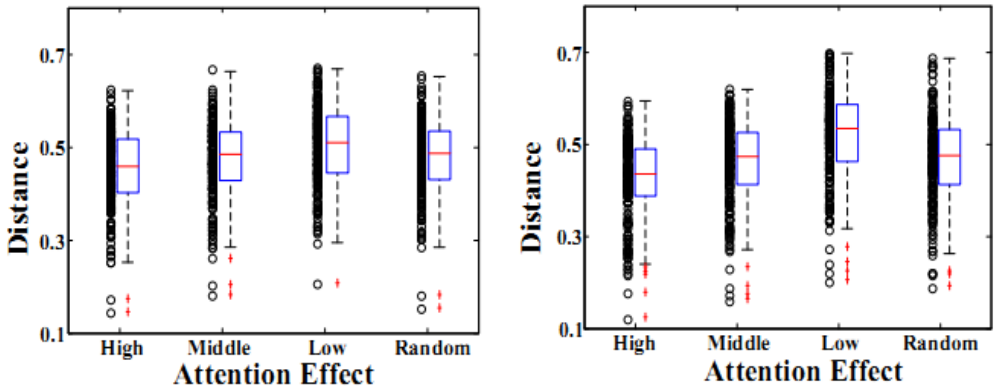
EKTA

- EKTA enhance the effect of some of student's historical states
- Cold-start students
- The higher the attention value is, the more contribution of this exercise will make when predictions



(a) ACC Comparison (b) AUC Comparison

Fig. 11. The effectiveness of attention in fitting process for testing.



(a) EERNNA (b) EKTA

Fig. 12. Performance over different attention values in proposed models

Experiments

- Knowledge Acquisition Tracking

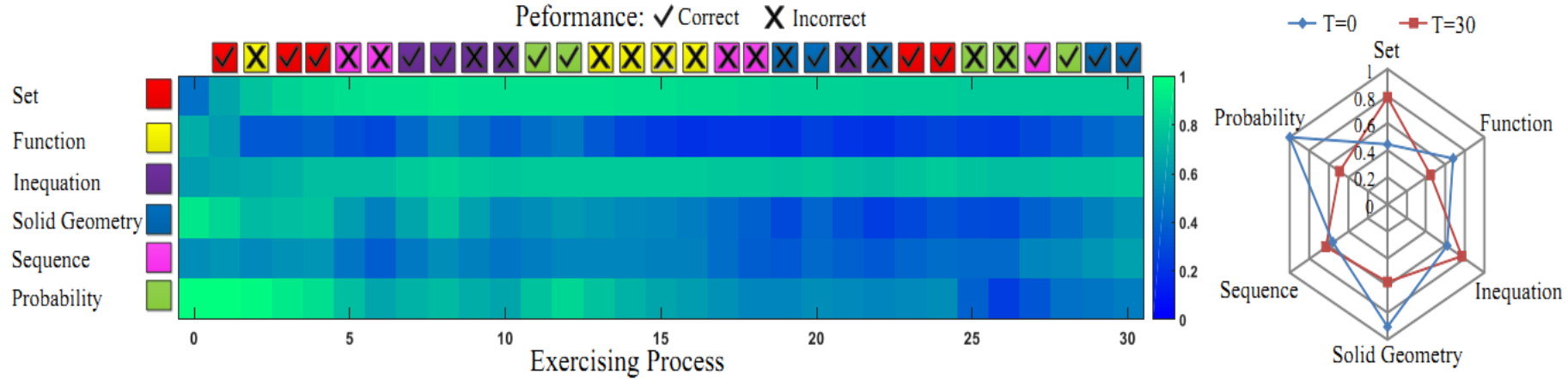


Fig. 13. An example of the knowledge mastery level tracking of a certain student on 6 concepts during her 30 exercising steps, which is painted in the middle matrix. Left side shows all concepts, which are marked in different colors. Top line records her performance on the 30 exercises. Right radar figure shows her knowledge mastery levels (in the range $(0, 1)$) on 6 concepts before (T=0) and after (T=30) exercising.

- The mastery levels on concepts change gradually during the process
- When she answers an exercise right (wrong), her knowledge state on the corresponding concept increases (decreases)

Experiments

- Prediction Case Study

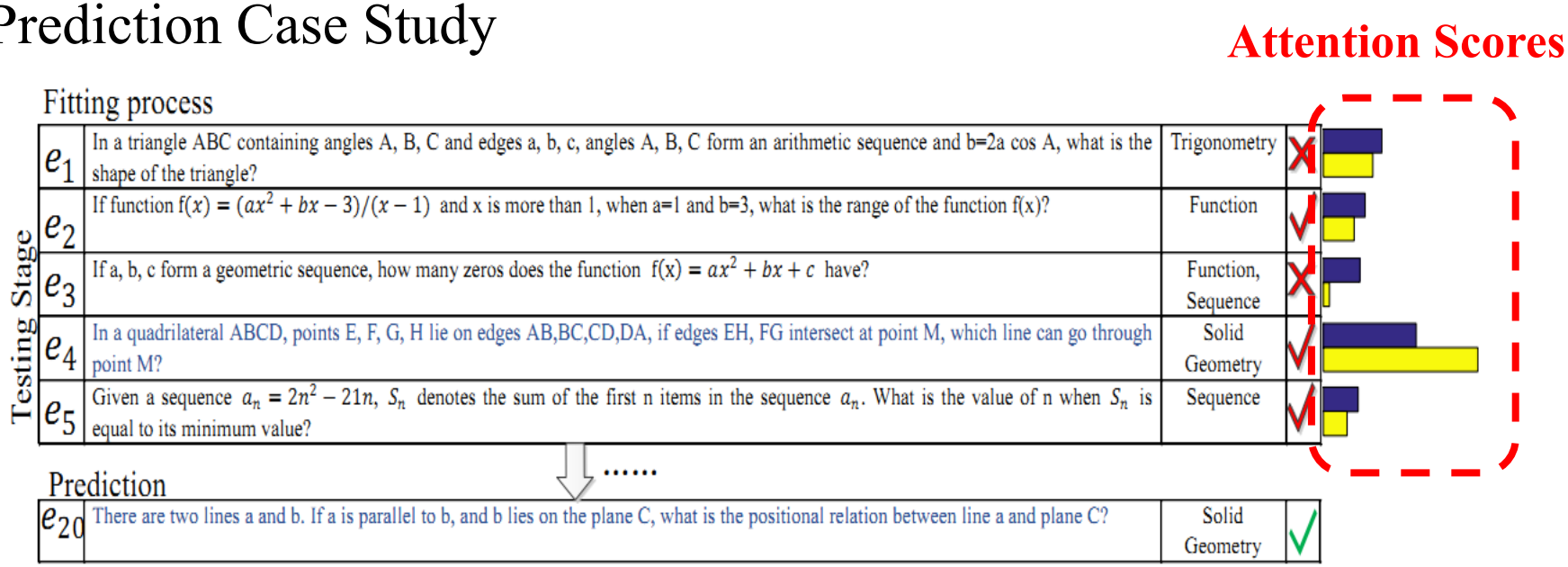


Fig. 14. Attention visualization in EERNNA and EKTA of an example student. We predict her performance on e_{20} based on her past 19 exercise records (we only show the first 5 exercises for better illustration). Right bars show the attention scores of two frameworks (i.e., EERNNA (blue) and EKTA (yellow)) for all exercises based on e_{20}.

- e_4 is actually much more difficult than e_{20}
- Both e_4 and e_{20} contain the same knowledge concept "Solid Geometry"
- EKTA endows a larger attention weight on e_4

Outline

Conclusion

- A novel EKT framework to track the mastery levels on multiple concepts and predict student future performance
- EERNN integrated three critical modules: Exercise Embedding, Knowledge Embedding, Student Embedding.
- Proposed two strategies for prediction : EKTM with Markov property and EKTA with Attention mechanism.
- Experiments on real-world dataset demonstrated the effectiveness and interpretability of EKT framework.

Q & A

Thanks!