The 38st Association for the Advancement of Artificial Intelligence (AAAI'17)
2018/02/02 - 02/07, New Orleans, Louisiana

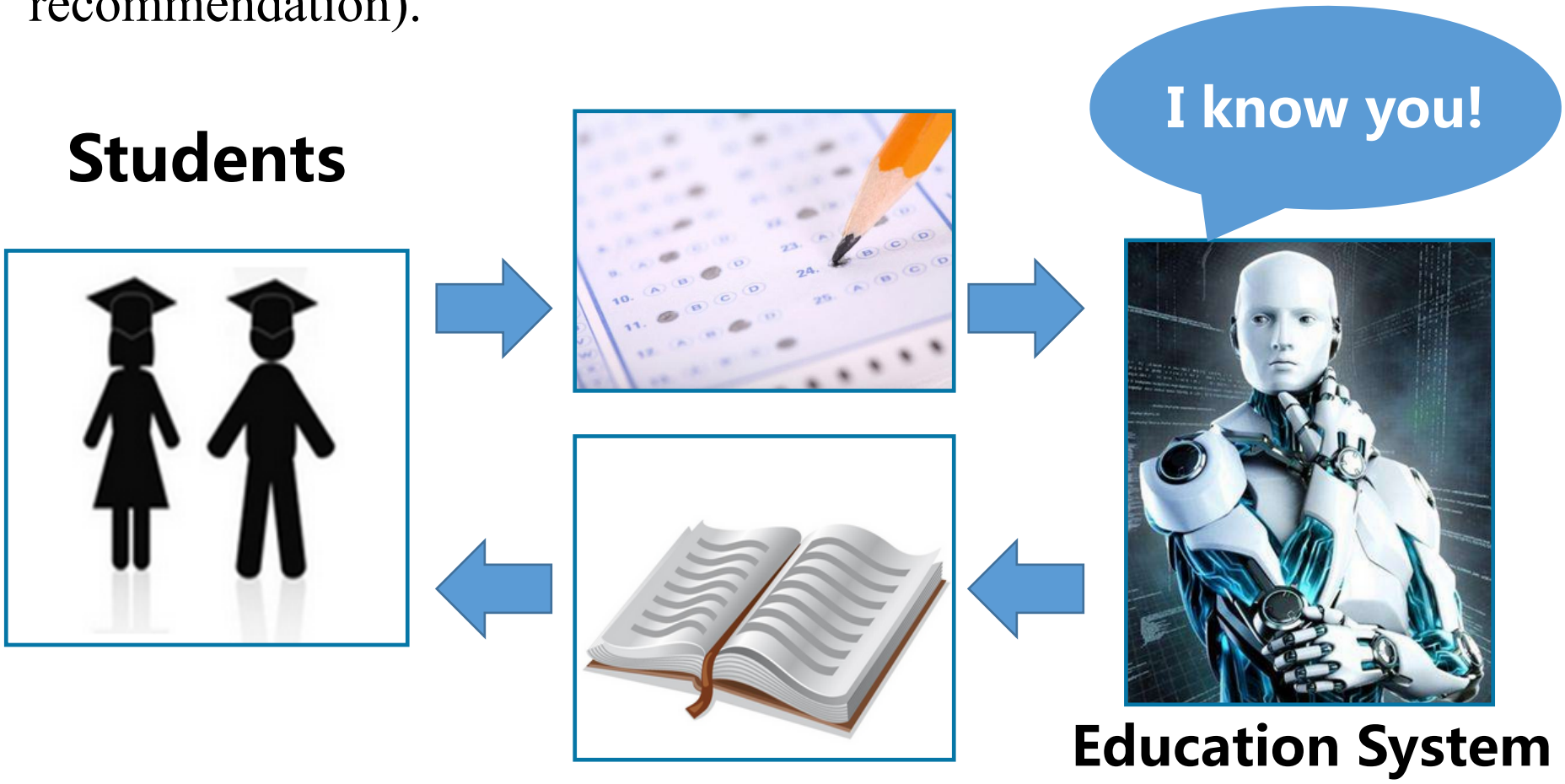
Exercise-Enhanced Sequential Modeling for Student Performance Prediction

Yu Su[†§], **Qingwen Liu**[§‡], **Qi Liu**[‡*], **Zhenya Huang**[‡], **Yu Yin**[‡],
Enhong Chen[‡], **Chris Ding**[♭], **Si Wei**[§], **Guoping Hu**[§]
[†]School of Computer Science and Technology, Anhui University
[‡] Anhui Province Key Lab. of Big Data Analysis and Application, University of Science and Technology of China
[§]iFLYTEK Research, [♭]CSE Department, University of Texas at Arlington
{yusu, qwliu}@iflytek.com, qiliuql@ustc.edu.cn, {huangzhy, yxonic}@mail.ustc.edu.cn
cheneh@ustc.edu.cn, chqding@uta.edu, {siwei, gphu}@iflytek.com

Outline

Background- Online Education System

Online education systems provide students with open access for self-learning (e.g., learning remedy suggestion and personalized exercise recommendation).



Background- Predict Student Performance

How can we say an education system understand a student?

I can predict student's performance(score)

Education System

Recommend Exercises

too hard

appropriate

too easy

Research Problem

- Urgent issue: Predict Student Performance(PSP)
  - How to automatically predict student performance without manual intervention?
  - This work regards score as performance.
- Opportunity
  - Exercises records of students
  - Text materials of exercise

Exercises records of one student

Exercise	Exercise Texts
e_1	If function $f(x) = x^2 - 2x + 2$ and $x \in [0,3]$, What is the range of $f(x)$?
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?
e_3	What is the y-intercept of the graph of equation $y = 2 \times \|4 \times x - 4\| - 10$?
e_4	What is the value of x If the inequality $\frac{2x-1}{x+2} \leq 3$?
e_5	If function $f(x) = 2x - 2$ and $x \in [-1,1]$, what is the range of $f(x)$?

Text contents of corresponding exercise

Challenge 1 for PSP

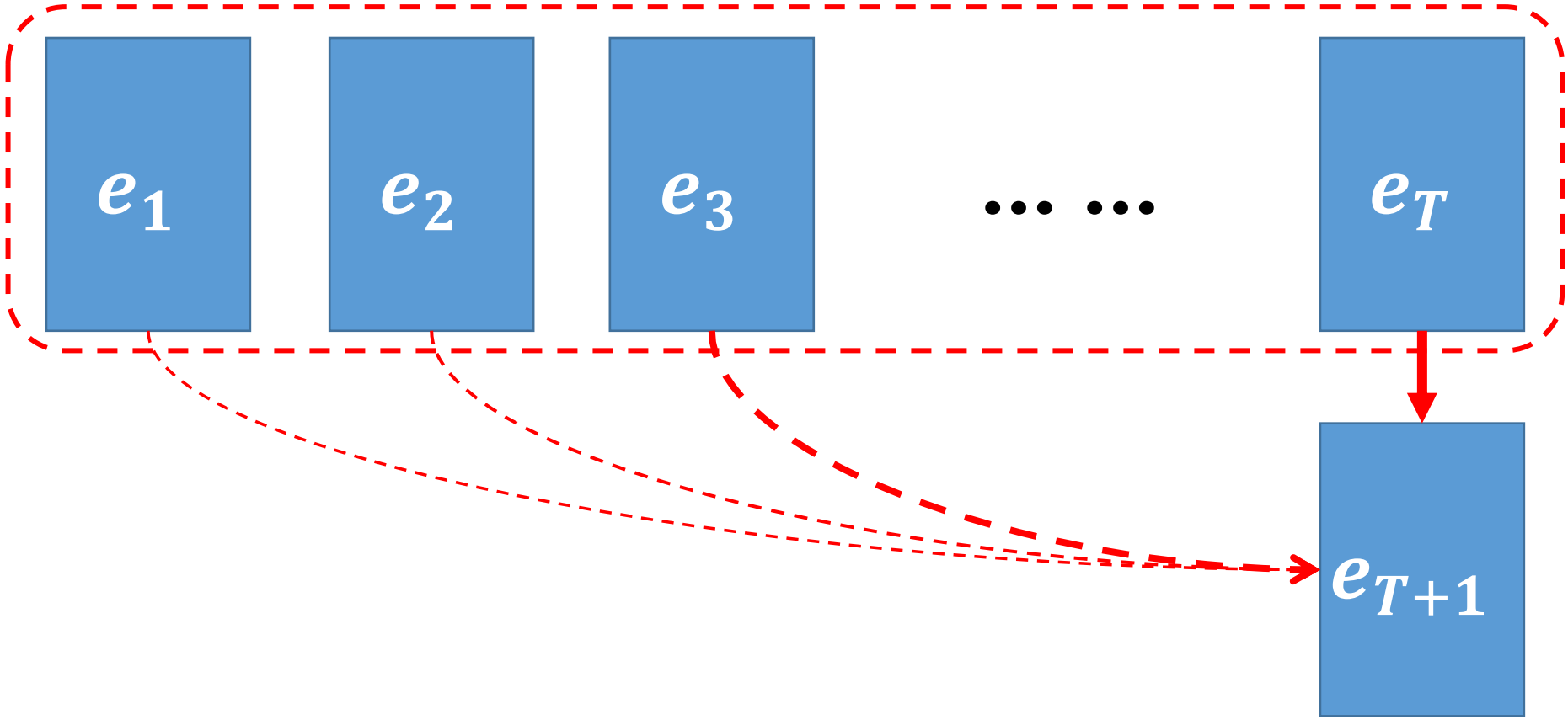
- Diverse expressions of exercises

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

- Need a unified way to understand and represent them automatically.

Challenge 2 for PSP

- Long-term historical exercising



Challenge 3 for PSP

- Cold start problem

Training Data

New Exercise

learn

predict

Education System

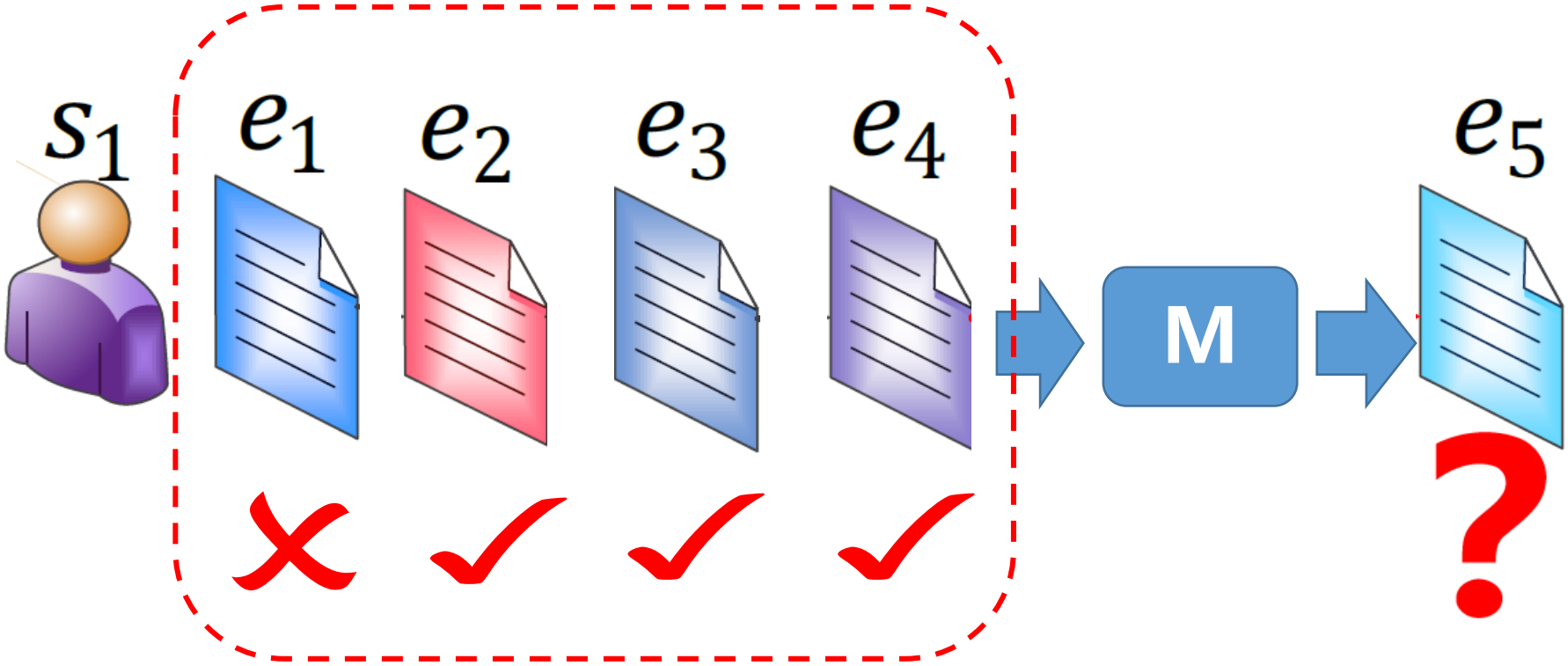
Related Work for PSP

- Education Psychology
  - **IRT** (Item Response Theory)
    - models student exercising records by a logistic-like function
  - **BKT** (Bayesian Knowledge Tracing)
    - traces them with a kind of hidden Markov model
- Machine Learning and Data Mining
  - **PMF**
    - projects students and exercises into latent factors
- Deep Learning
  - **DKT**
    - deep learning method uses RNN to model student exercising process for prediction .

Outline

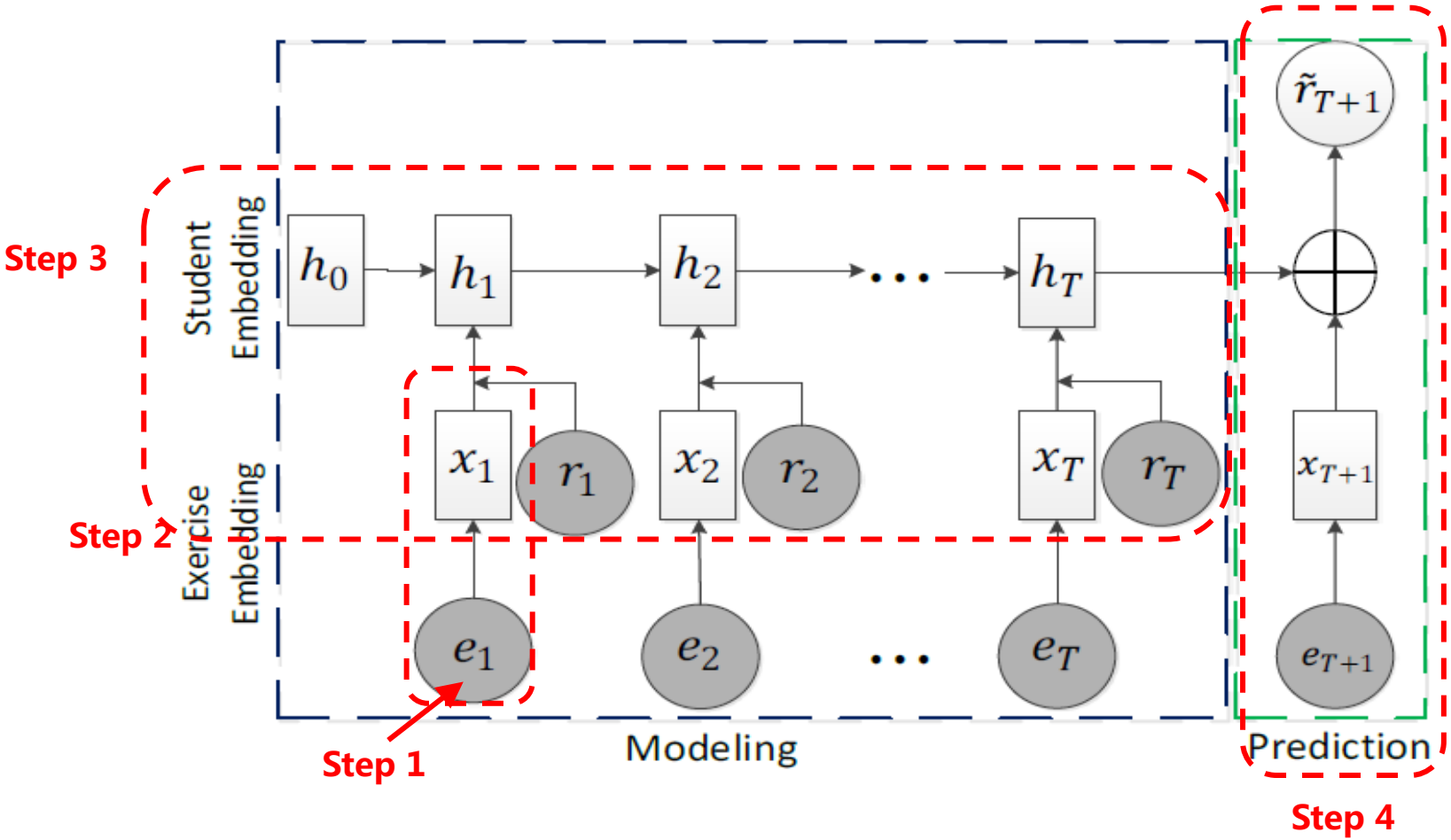
Problem Definition

- Given: the exercising records of each student and the text descriptions of each exercise from 1 to T。
- Goal: train a unified model M, predict the scores on the next exercise e_{T+1} of each specific student.



Outline

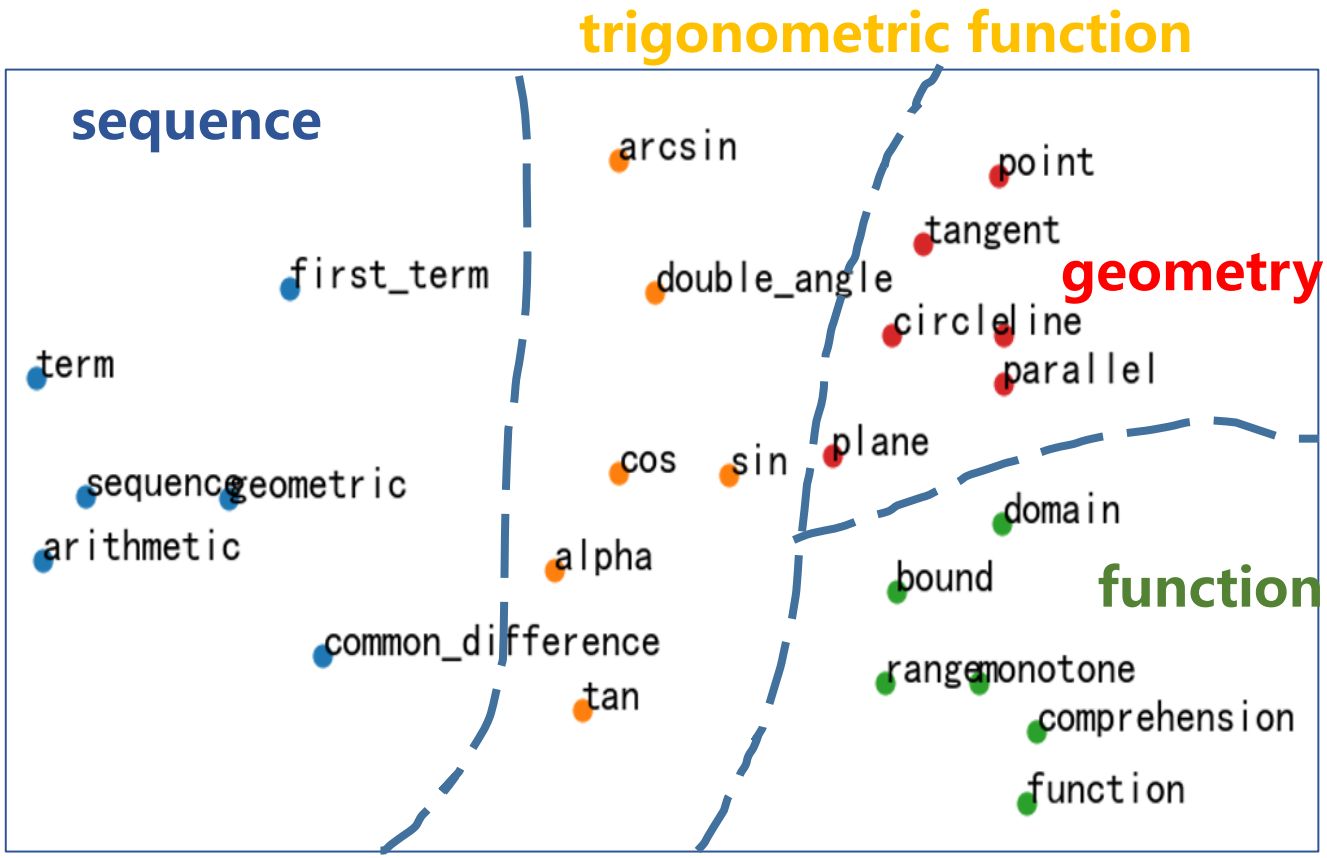
Exercise-Enhanced Recurrent Neural Network(EERNN)



Step 1: Word Embedding

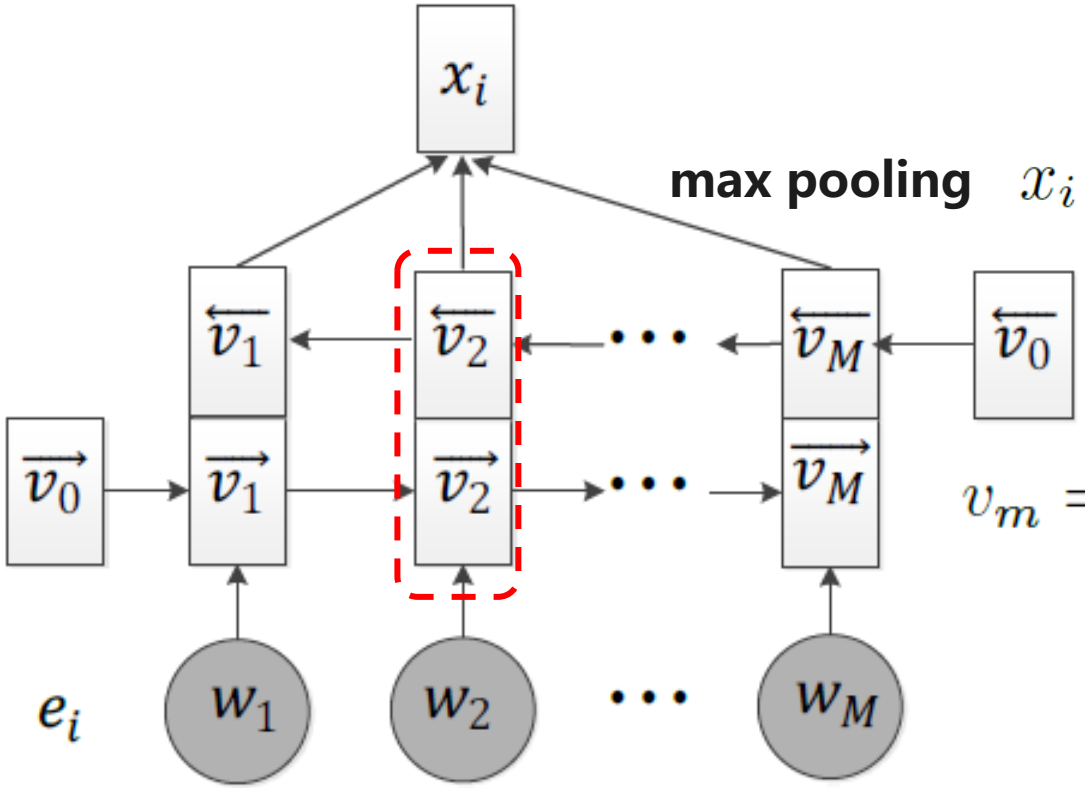
- Goal: learn word representations from semantic perspective in math exercise.

- Main procedures:
  - Word split
  - Latex to feature
  - Word to vector



Step 2: Exercise Embedding

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



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

BiLSTM

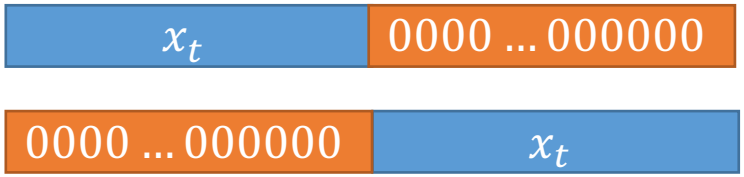
$v_m = concatenate(\overrightarrow{v}_m, \overleftarrow{v}_m)$

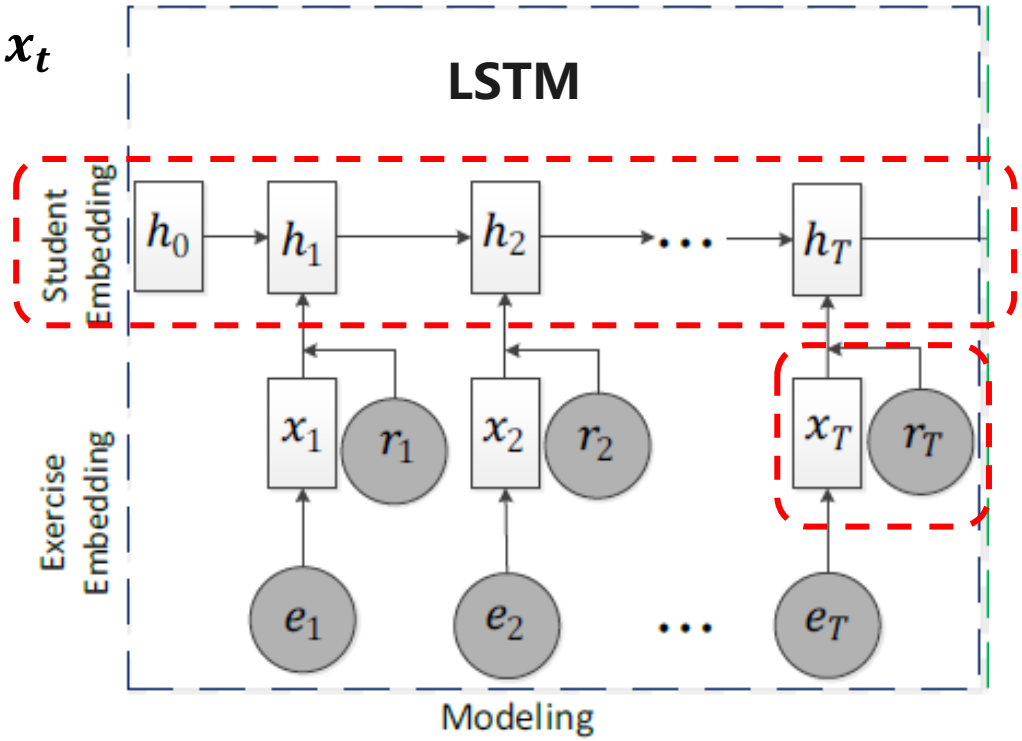
Step 3: Student Embedding

- Goal: Student Embedding aims at modeling the whole student exercising process and learning the hidden representations of students.

Combine the score and embedding x_t

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

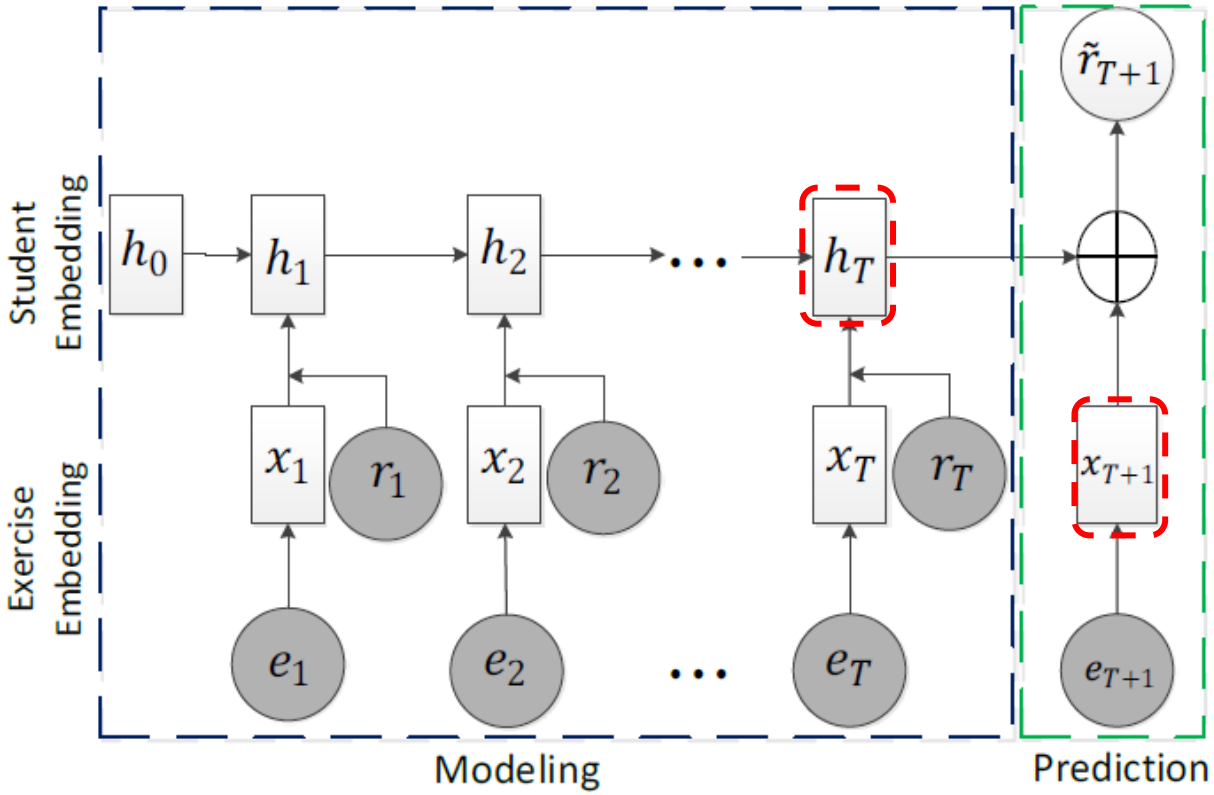




Step 4: Prediction (Two Strategies)

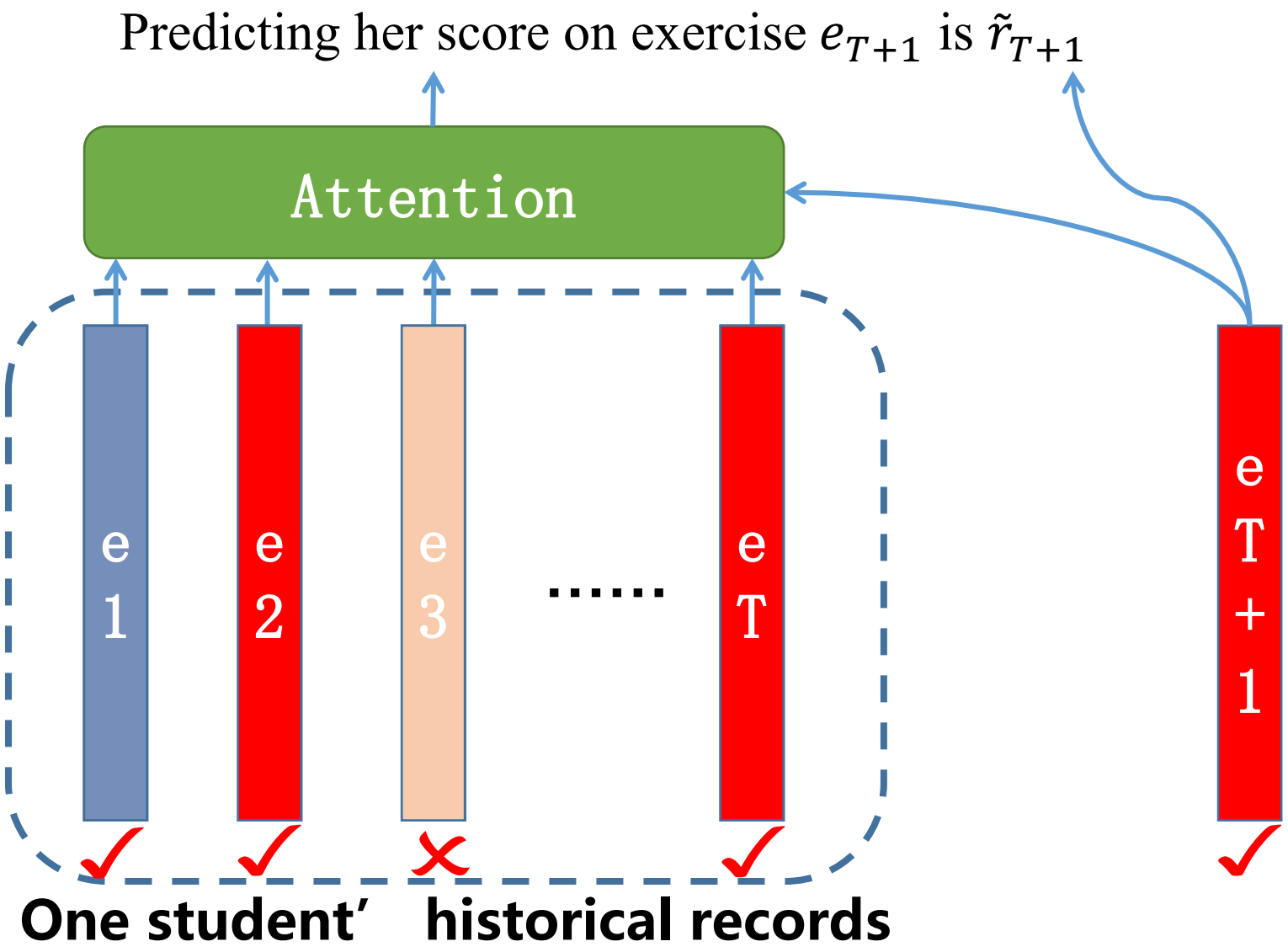
- Goal: Predicting her performance on exercise e_{T+1} at step T +1.

$$y_{T+1} = ReLU(\mathbf{W_1} \cdot [h_T \oplus x_{T+1}] + \mathbf{b_1})$$
$$\widetilde{r}_{T+1} = \sigma(\mathbf{W_2} \cdot y_{T+1} + \mathbf{b_2}),$$



EERNNM with Markov property

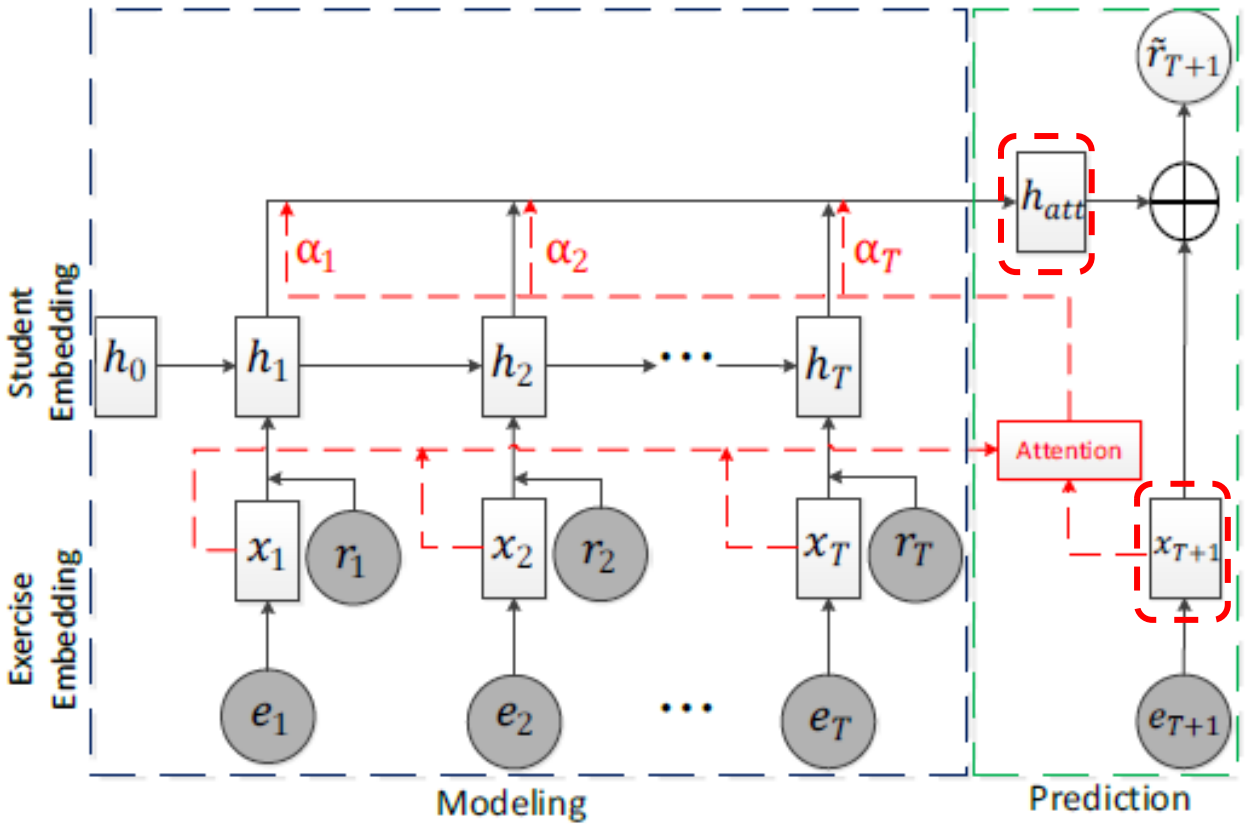
Step 4: Prediction (Two Strategies)



Step 4: Prediction (Two Strategies)

- Goal: predicting her performance on exercise e_{T+1} at step T +1.

$$h_{att} = \sum_{j=1}^{T} \alpha_j h_j,\ \alpha_j = cos(x_{T+1}, x_j)$$



EERNNA with Attention mechanism

Outline

Experiments

- Experiments dataset
  - Supplied by Zhixue, IFLYTEK
  - a widely-used online learning system, which provides senior high school students with a large exercise resources for exercising.

Table 1: The statistics of mathematics dataset.

Statistics	Original	Pruned
# of records	68,337,149	5,596,075
# of students	110,0726	84,909
# of exercises	1,825,767	15,045
# of knowledge concepts	550	447
Avg. exercises per student	\	65.9
Avg. words per exercise	\	27.3
Avg. knowledge concepts per exercise	\	1.8
Avg. exercises per knowledge concept	\	54.2

Experiments

- Baseline methods
  - Variants of EERNN: **LSTMM, LSTMA**
    - just utilize knowledge-specific representations
    - To validate the importance to incorporate exercise texts for the prediction in EERNN
  - Education Psychology**: IRT, BKT**
  - Machine Learning and Data Mining: **PMF**
  - deep learning method: **DKT**
    - The most similar method to ours

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

Experiments

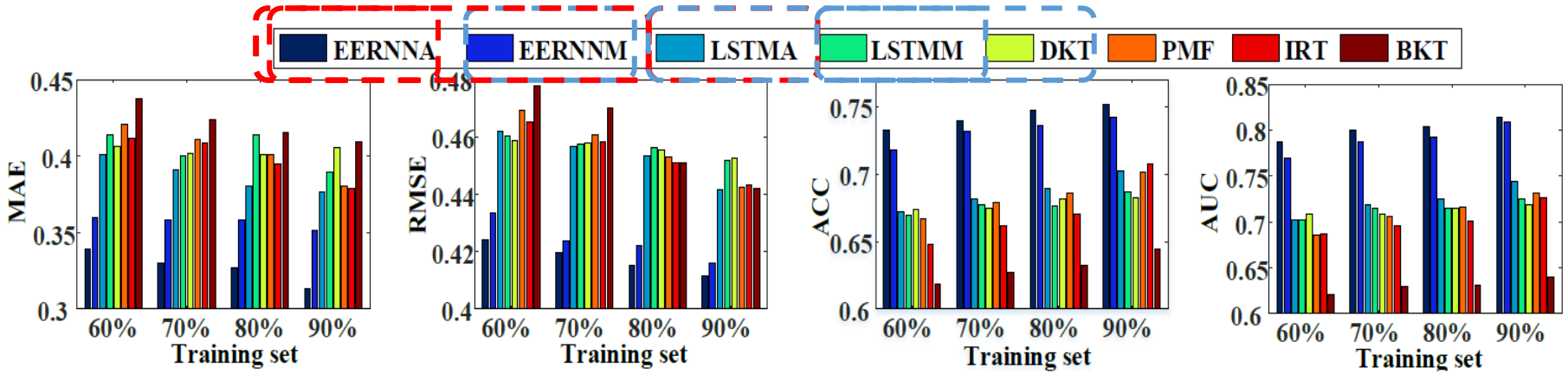


Figure 5: Overall results of student performance prediction on four metrics.

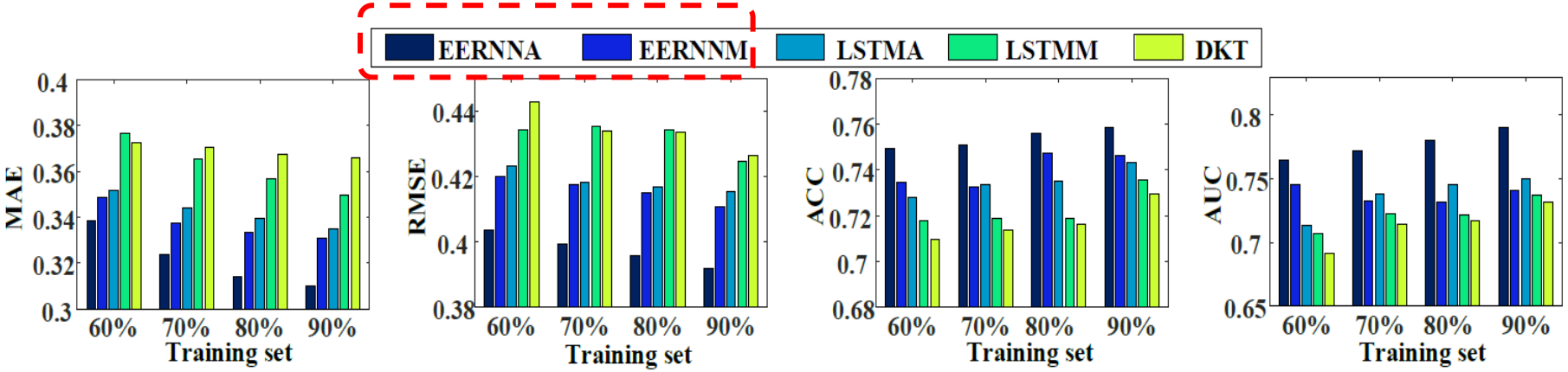
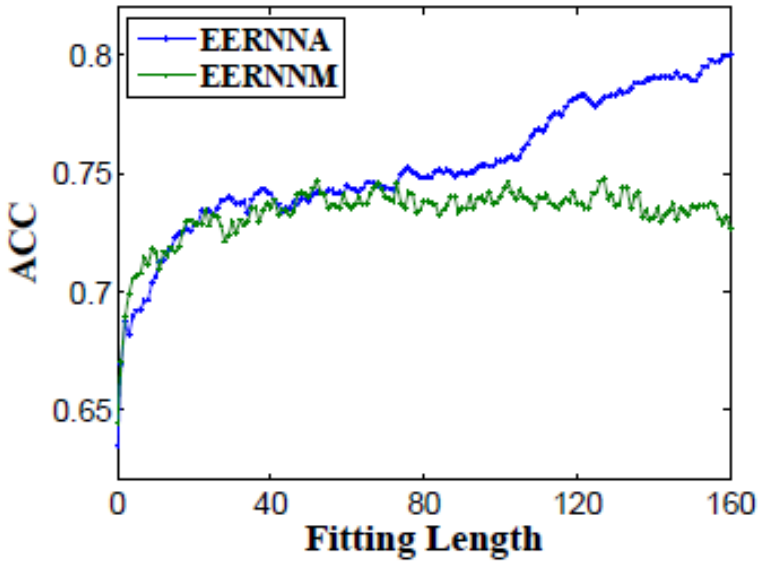
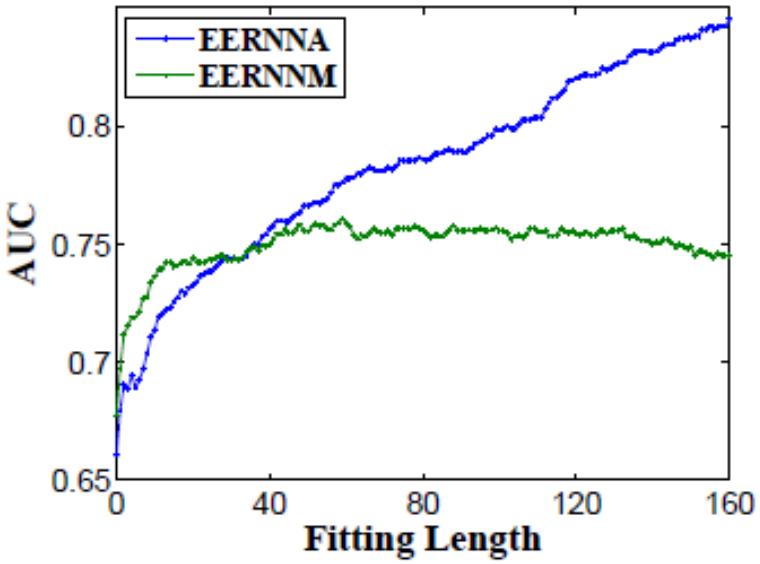


Figure 6: Results of student performance prediction on cold start new exercises on four metrics.

Experiments





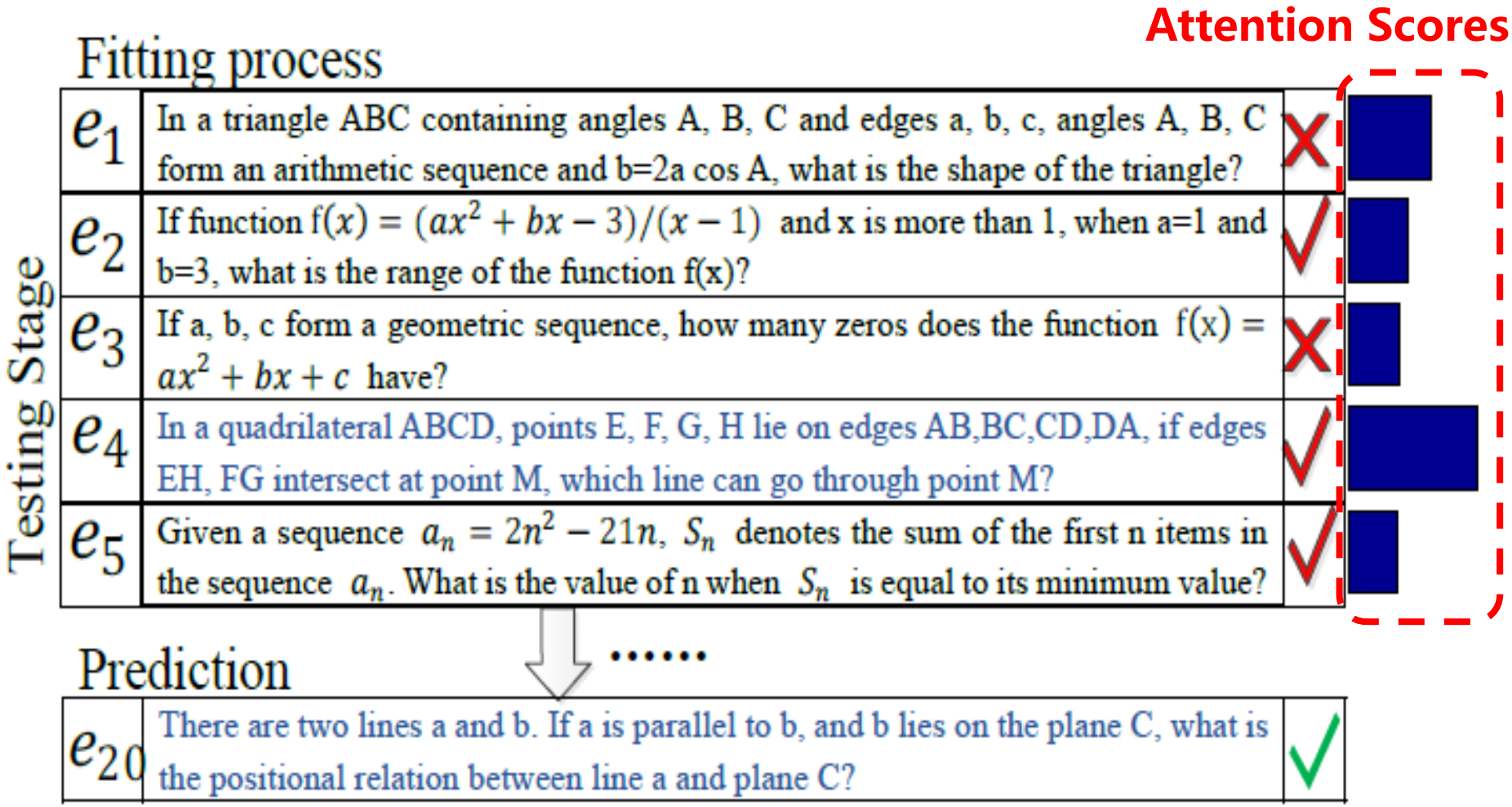
Information can be used by EERNNM

More Information can be used by EERNNM

Old Information forget by EERNNM

Focus useful information by EERNNA

Experiments



Outline

Conclusion

- Proposed a novel EERNN framework to predict student future performance.
- EERNN integrated two critical components, BiLSTM to extract exercise semantic representations from texts, LSTM architecture to trace student states.
- Proposed two strategies for prediction : EERNNM with Markov property and EERNNA with Attention mechanism.
- Experiments on real-world dataset demonstrated the effectiveness (specially cold start problem) of EERNN .

Future Work

- Different exercise types (e.g., the subjective exercises with continuous scores)

- Incorporate more information
  - Knowledge concepts
  - The time cost on exercises

- Integrate some educational theories
  - learning and forgetting curves
  - Guess and slip

Q & A

Thanks!