Virtual Conference KDD2021 August 14th – 18th

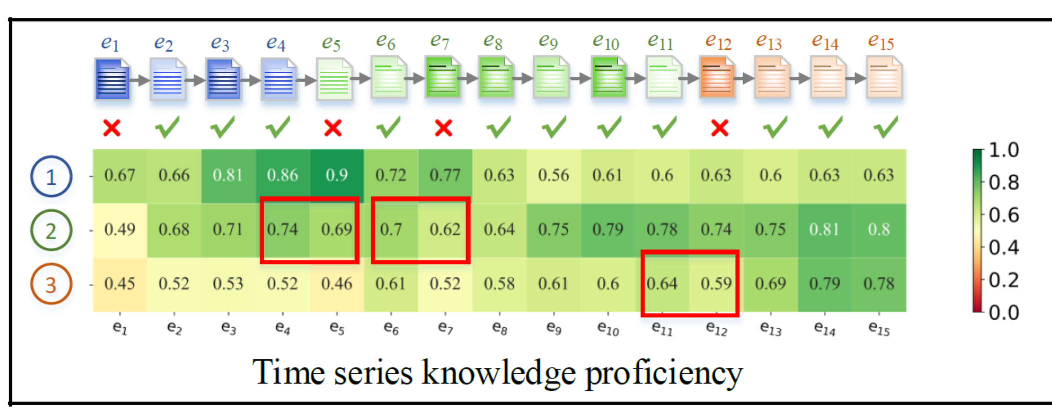
Learning Process-consistent Knowledge Tracing

Shuanghong Shen[1], Qi Liu[1], Enhong Chen[1,*], Zhenya Huang[1], Wei Huang[1], Yu Yin[1], Su Yu[2] , Shijin Wang[2]

[1]Anhui Province Key Laboratory of Big Data Analysis and Application, , School of Data Science & School of Computer Science and Technology, University of Science and Technology of China, [2]iFLYTEK Research, iFLYTEK, Co., Ltd

Introduction

Motivation: Knowledge Tracing (KT) is a fundamental and critical task of online education, which aims to trace students' changing knowledge state during their learning process. Most of the existing KT methods pursue high accuracy of student performance prediction but neglect the consistency of students' changing knowledge state with their learning process. They argue that once the student has answered wrongly, his/her knowledge state on corresponding knowledge concepts will decline.



Exercises	Knowledge Concepts
e_1, e_2, e_3, e_4	① Absolute Value
$e_5, e_6, e_7, e_8, e_9, e_{10}, e_{11}$	② Addition and Subtraction Integers
$e_{12}, e_{13}, e_{14}, e_{15}$	③ Ordering Integers

Figure 1: A toy example of the evolution process of a student's knowledge states that are traced by DKT, where the student has answered 15 exercises on 3 knowledge concepts. In the left figure, the color of the heatmap or the number in the small box refers to the knowledge state of the student after answering the exercise. The red boxes indicate that DKT thinks the knowledge state will decline after wrong answers. The right table gives the relations between exercises and knowledge concepts.

Key Problem: Mistakes are seen as natural elements of learning processes , and students can learn from errors and foster learning progress through a favorable error climate. How to keep the consistency of students' learning process in knowledge tracing and give equal attention to both right and wrong learning interactions?

Challenges:

- How to define the learning process and convert it into a proper form for modeling.
- The learning gain, which represents the knowledge that students acquire in learning, is implicit and changeable in learning process.
- students' knowledge will also decrease over time, which commonly manifests as forgetting, is also necessary to be considered in the KT task.

Problem Statement

Given:

- Students' learning sequence $\boldsymbol{x} = \{ (e_1, at_1, a_1), it_1, (e_2, at_2, a_2), it_2, \dots, (e_t, at_t, a_t), it_t \}$

Goal:

- To monitor students' changing knowledge state during the learning process and predict their future performance at the next learning step t+1

Datasets

ASSIST2012:https://sites.google.com/site/assistmentsdata/home/2012-13-school-data-withaffect

ASSISTchall:https://sites.google.com/view/assistmentsdatamining/dataset

EdNet-KT1: http://ednet-leaderboard.s3-website-ap-northeast-1.amazonaws.com/

Statistics	Datasets		
	ASSIST2012	ASSISTchall	EdNet-KT1
Students	29,018	1,709	784,309
Exercises	53,091	3,162	12,372
Concepts	265	102	141
Answer Time	26,747	1,326	9,292
Interval Time	29,748	2,839	41,830
Avg.length	93.45	551.68	121.48

Table 1: Statistics of all datasets.

The LPKT Model

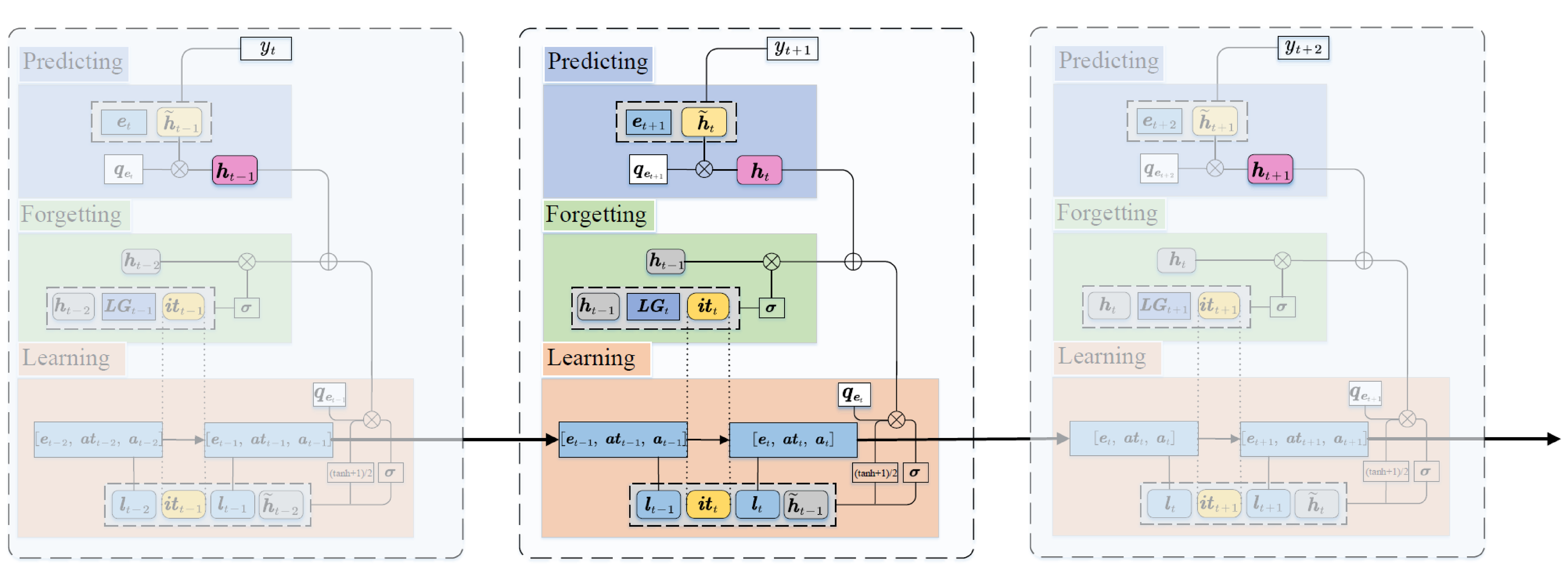


Figure 2: The architecture of the LPKT model.

Modules

- **Learning Module:** model learning gains compared with the previous learning interaction.
- **Forgetting Module:** measure how much knowledge will be forgotten as time goes on. Then, the learning gains and forgotten knowledge will be taken advantage of to update the student's previous knowledge state for achieving their latest knowledge state.
- **Predicting Module:** predict the student's performance on the next exercise according to his/her latest knowledge state.

LPKT details

- **Embeddings:**
  - **Time embedding:** representing the discretized interval time and answer time by two embedding matrices.
  - **Learning embedding:** the embedding of the basic learning cell (e_t, at_t, a_t),
  - **Knowledge embedding:** storing and updating the knowledge state of students during the learning process
- **Learning:**
  - two continuous learning embeddings, their interval time and students' previous knowledge state

$$\boldsymbol{lg}_t = tanh(\boldsymbol{W}_2^T[\boldsymbol{l}_{t-1} \oplus \boldsymbol{it}_t \oplus \boldsymbol{l}_t \oplus \widetilde{\boldsymbol{h}}_{t-1}] + \boldsymbol{b}_2),$$

$$\boldsymbol{\Gamma}_t^l = \sigma(\boldsymbol{W}_3^T[\boldsymbol{l}_{t-1} \oplus \boldsymbol{it}_t \oplus \boldsymbol{l}_t \oplus \widetilde{\boldsymbol{h}}_{t-1}] + \boldsymbol{b}_3),$$

$$LG_t = \boldsymbol{\Gamma}_t^l \cdot ((\boldsymbol{lg}_t + 1)/2),$$

$$\widetilde{LG}_t = \boldsymbol{q}_{e_t} \cdot LG_t,$$

- **Forgetting:**
  - students' previous knowledge state, present learning gains, and the interval time

$$\boldsymbol{\Gamma}_t^f = \sigma(\boldsymbol{W}_4^T[\boldsymbol{h}_{t-1} \oplus LG_t \oplus \boldsymbol{it}_t]) + \boldsymbol{b}_4),$$

$$\boldsymbol{h}_t = \widetilde{LG}_t + \boldsymbol{\Gamma}_t^f \cdot \boldsymbol{h}_{t-1}.$$

- **Predicting:**
  - using the related knowledge state to infer the student's performance

$$y_{t+1} = \sigma(\boldsymbol{W}_5^T[\boldsymbol{e}_{t+1} \oplus \widetilde{\boldsymbol{h}}_t] + \boldsymbol{b}_5),$$

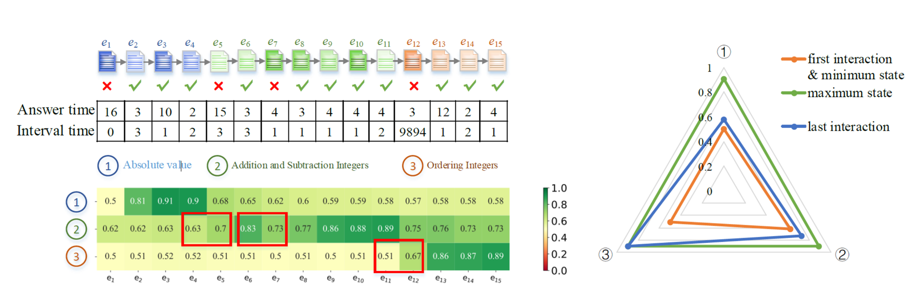
Loss Function

$$\mathbb{L}(\theta) = -\sum_{t=1}^{T}(a_t \log y_t + (1 - a_t)\log(1 - y_t)) + \lambda_\theta ||\theta||^2,$$

Experiments

Knowledge State Visualization

- LPKT can capture reasonable knowledge state of students, which is in consistent with their learning process.



(a) Time series knowledge proficiency

(b) Radar diagram

Figure 3: The evolution process of a student's (the same student in Figure 1) knowledge states traced by LPKT. In sub-figure (a), the top part indicates his/her performance at each time step, the answer time and interval time. Sub-figure (b) is the radar diagram of the student's knowledge state at the first interaction and the last interaction, his/her maximum and minimum knowledge state in learning process are also depicted on it.

Student Performance Prediction

- LPKT outperforms all other KT methods on all datasets and metrics.

Methods	ASSIST2012				ASSISTchall				EdNet-KT1			
	RMSE	AUC	ACC	r^2	RMSE	AUC	ACC	r^2	RMSE	AUC	ACC	r^2
DKT	0.4241	0.7289	0.7360	0.1468	0.4471	0.7213	0.6907	0.1425	0.4508	0.6836	0.6889	0.1008
DKT+	0.4239	0.7295	0.7254	0.1497	0.4502	0.7101	0.6842	0.1308	0.4601	0.6429	0.6733	0.0635
DKVMN	0.4261	0.7228	0.7329	0.1398	0.4503	0.7108	0.6842	0.1302	0.4538	0.6741	0.6843	0. 0913
SAKT	0.4258	0.7233	0.7339	0.1403	0.4626	0.6605	0.6694	0.0822	0.4524	0.6794	0.6862	0.0964
CKT	0.4234	0.7310	0.7365	0.1497	0.4455	0.7263	0.6924	0.1488	0.4519	0.6811	0.6871	0.0984
AKT	0.4100	0.7740	0.7554	0.2035	0.4317	0.7655	0.7141	0.2015	0. 4241	0. 7701	0.7287	0.2059
LPKT	**0.4069**	**0.7772**	**0.7583**	**0.2145**	**0.4153**	**0.8008**	**0.7424**	**0.2609**	**0.4234**	**0.7721**	**0.7300**	**0.2085**

Table 2: Results of comparison methods on student performance prediction. LPKT outperforms all baselines on all datasets.

Ablation Experiments Ⅰ

- forgetting plays a critical role in learning process, which can cause the biggest decline of the predictive results if we do not consider it.
- modeling the learning gain indeed performs better than modeling only the learning outcomes in knowledge tracing.
- the answer and interval time are essential and necessary in the learning process, which is harmful to accurately model the learning process if omitted.

Methods	learning	forgetting	time	RMSE	AUC	ACC	r^2
LPKT-L	✓		✓	0.4112	0.7659	0.7531	0.1980
LPKT-F		✓	✓	0.4087	0.7734	0.7554	0.2075
LPKT(no time)	✓	✓		0.4077	0.7759	0.7571	0.2115
LPKT	✓	✓	✓	**0.4069**	**0.7772**	**0.7583**	**0.2145**

Table 3: Results of ablation experiments on ASSIST2012.

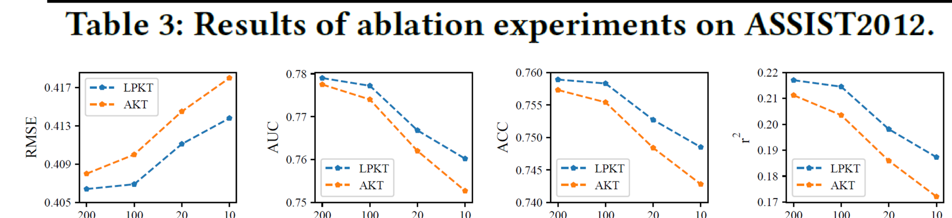


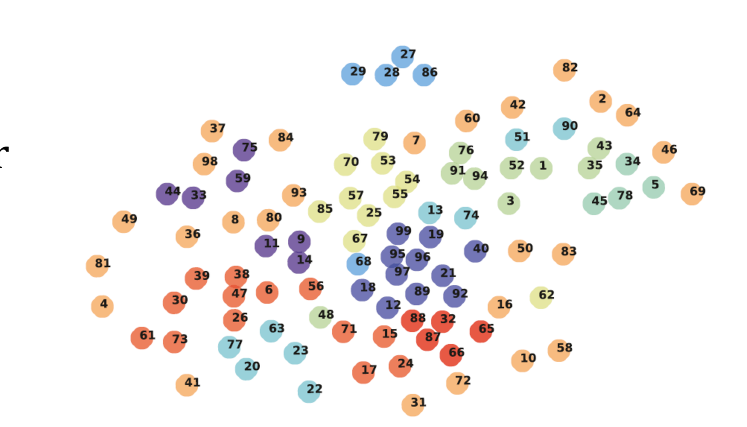
Figure 4: Comparison results of the influence of learning sequence length of LPKT and AKT on ASSISTchall.

Ablation Experiments Ⅱ

- LPKT is less affected by incomplete learning sequences, and indeed better models students' learning process.

Exercises Clustering

- LPKT can learn meaningful exercise embeddings automatically, which can serve as valuable supplements for the educational experts.



(a) Exercises clustering results. Randomly selected 100 exercises in the 3162 exercises of ASSISTchall are clustered into ten concepts. Exercises under the same concept are labeled in the same color and the number stands for the index of exercises.

Conclusions and Future Works

- ✓ A new paradigm for knowledge tracing through modeling learning process
- ✓ A novel model named Learning Process-consistent Knowledge tracing (LPKT)
- ✓ Formalized the learning process as the basic learning cell and interval time
- ✓ Modeled the learning gain and its diversity in learning process
- ✓ Designed a forgetting gate to determine the reduction of knowledge over time
- ✓ Extensive experiments on three public datasets demonstrated the interpretability and effectiveness of LPKT
- ➢ keep exploring better ways to model students' learning process
- ➢ study how to automatically learn the specific weights in the Q-matrix
- ➢ mathematical theory proof

Speaker: Shuanghong Shen

Email: closer@mail.ustc.edu.cn