Learning Behavior-oriented Knowledge Tracing

Bihan Xu[1,2], Zhenya Huang[1,2], Jiayu Liu[1,2], Shuanghong Shen[1,2],

Qi Liu[1,2], Enhong Chen[1,2], Jinze Wu[3], Shijin Wang[2,3]

1.Anhui Province Key Laboratory of Big Data Analysis and Application,

University of Science and Technology of China

2.State Key Laboratory of Cognitive Intelligence

3. iFLYTEK Research, iFLYTEK Co., Ltd, Hefei, China

CONTENTS

01 Background

02 Architecture

03 Experimental Results

04 Conclusion

CONTENTS

01 Background

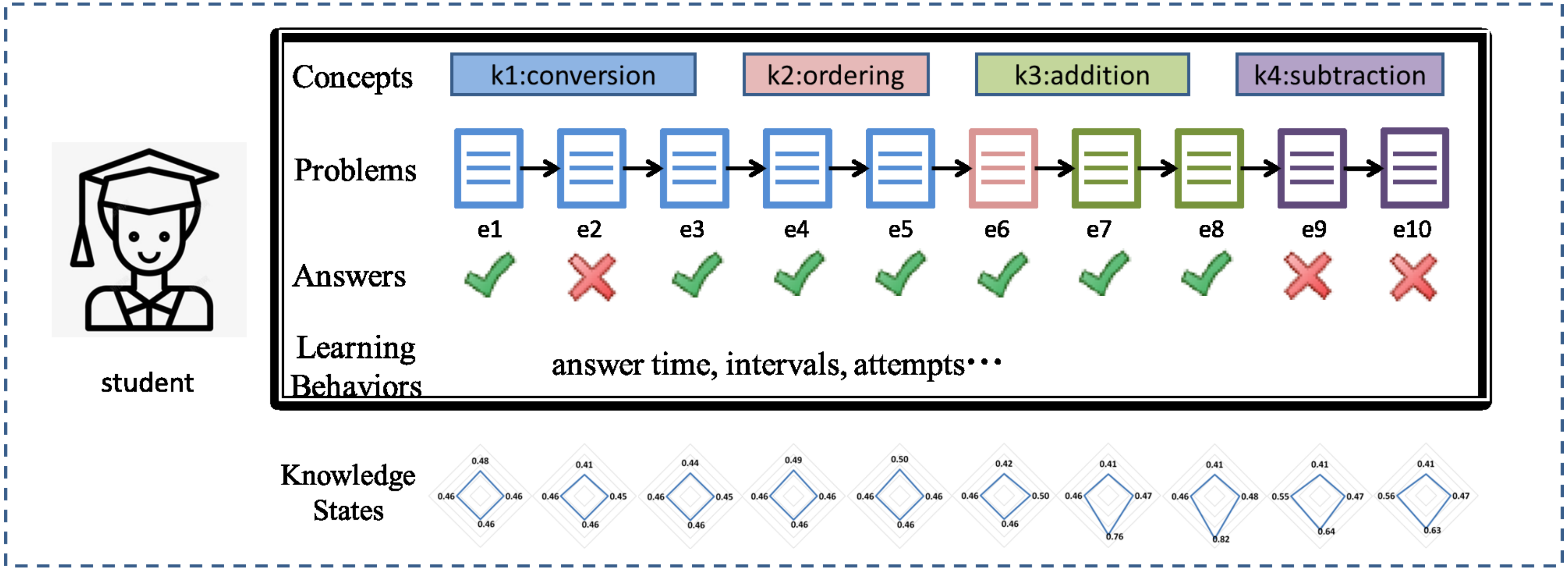
02 Architecture

03 Experimental Results

04 Conclusion

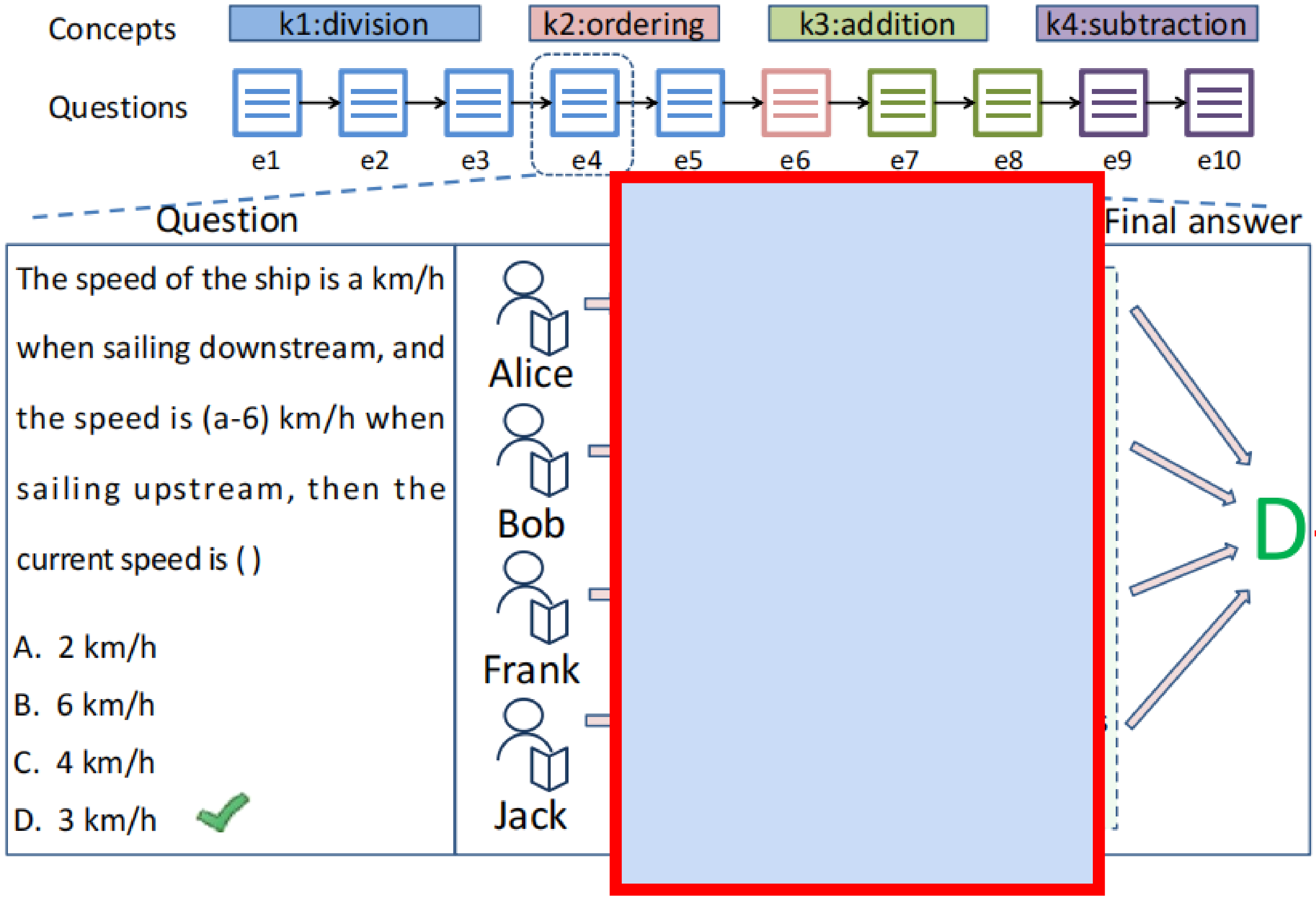
Knowledge Tracing

- Estimate students' knowledge states based on their historical learning interactions.
- Help students realize their weakness and improve learning efficiency.



Learning Behaviors

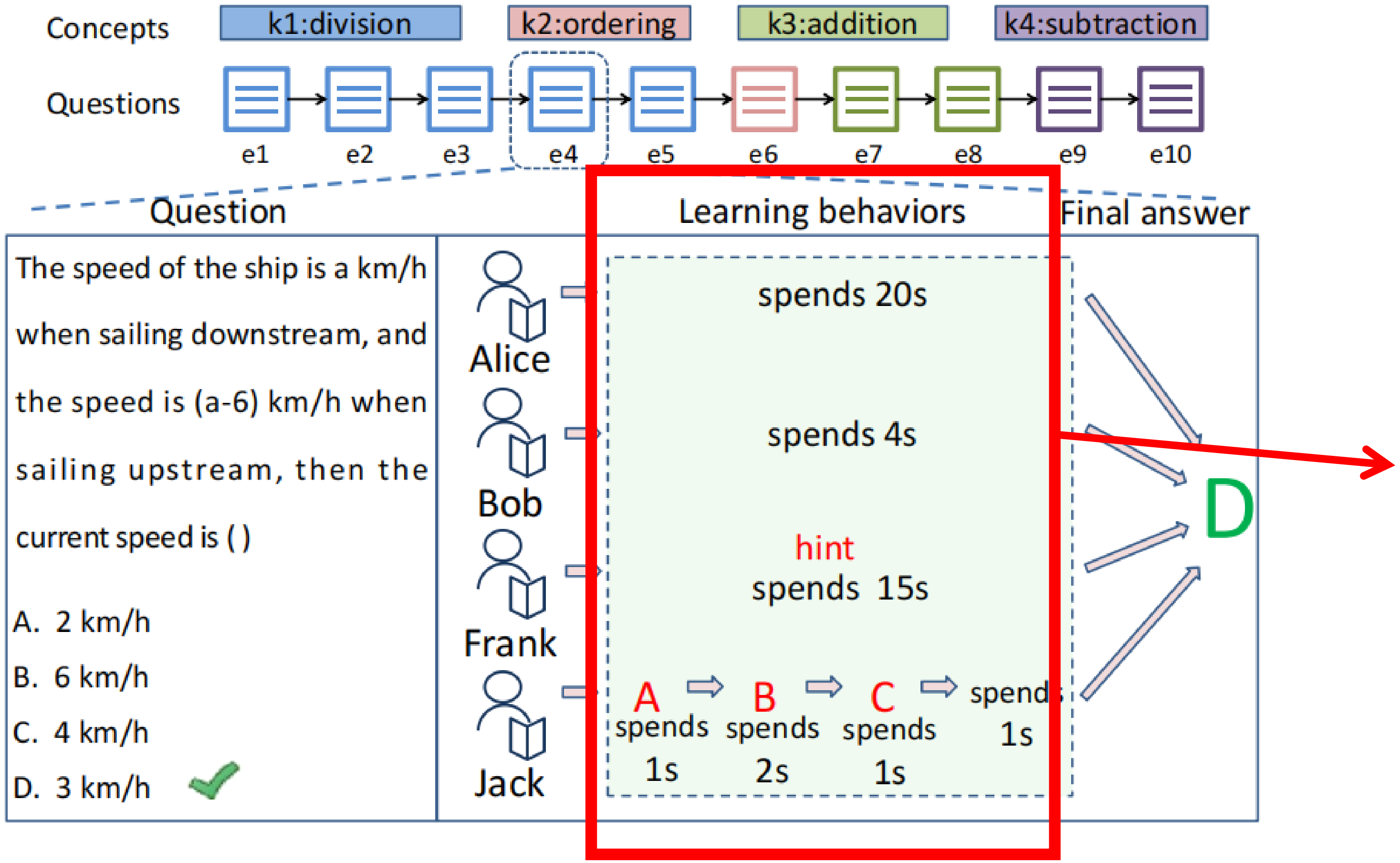
- Previous works focus on exploring learners' question-response pairs to track their knowledge mastery while neglecting the critical learning behaviors.



The four learners have equally benefited from this question-solving process.

Learning Behaviors

- Only consider the question-response pairs would lead to misleading estimation results.

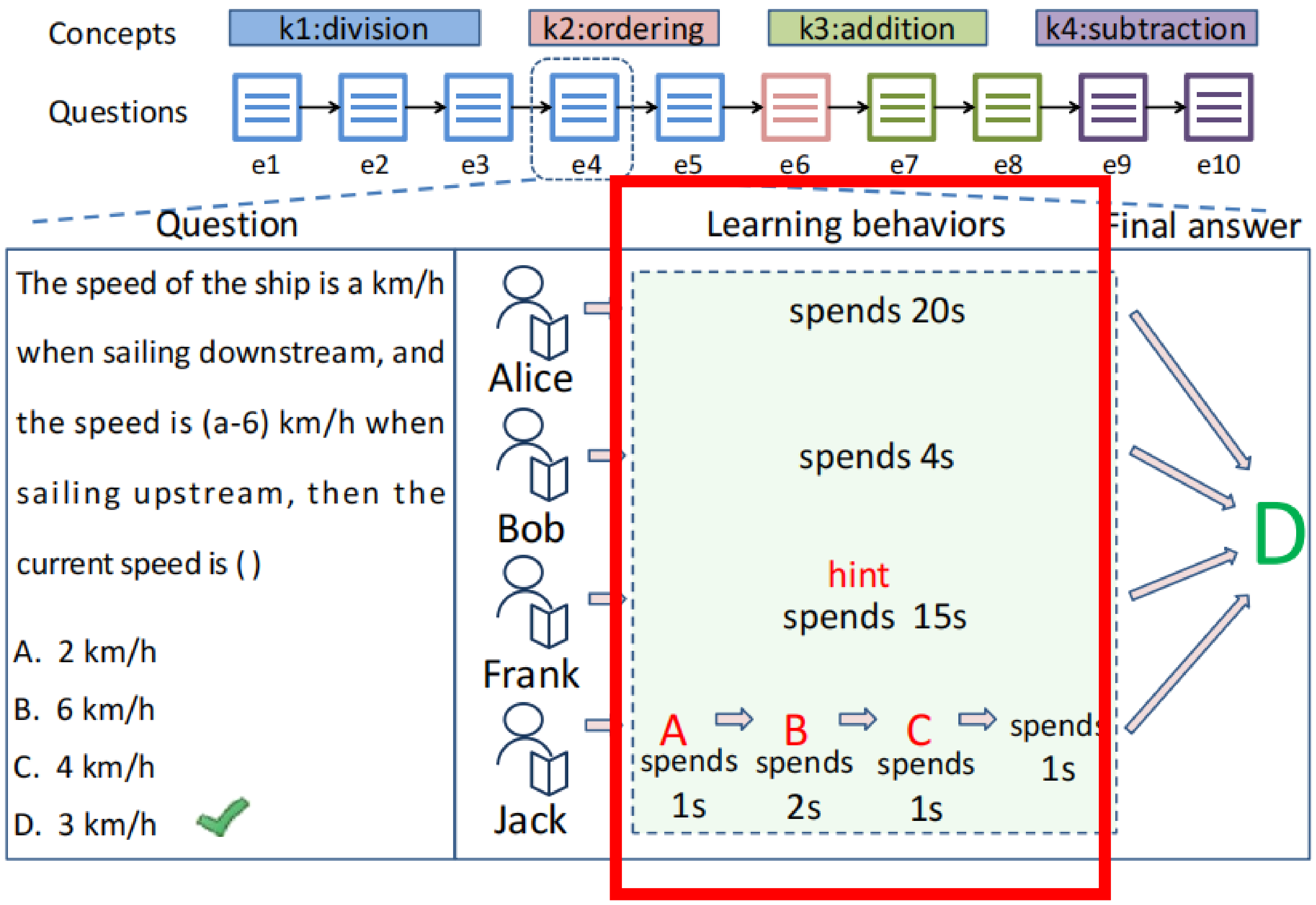


The four learners who show extremely different behaviors should have quite different knowledge acquisition.

Learning Behaviors

We summarize three typical behaviors and investigate their complex effects on assessing learners' knowledge states:

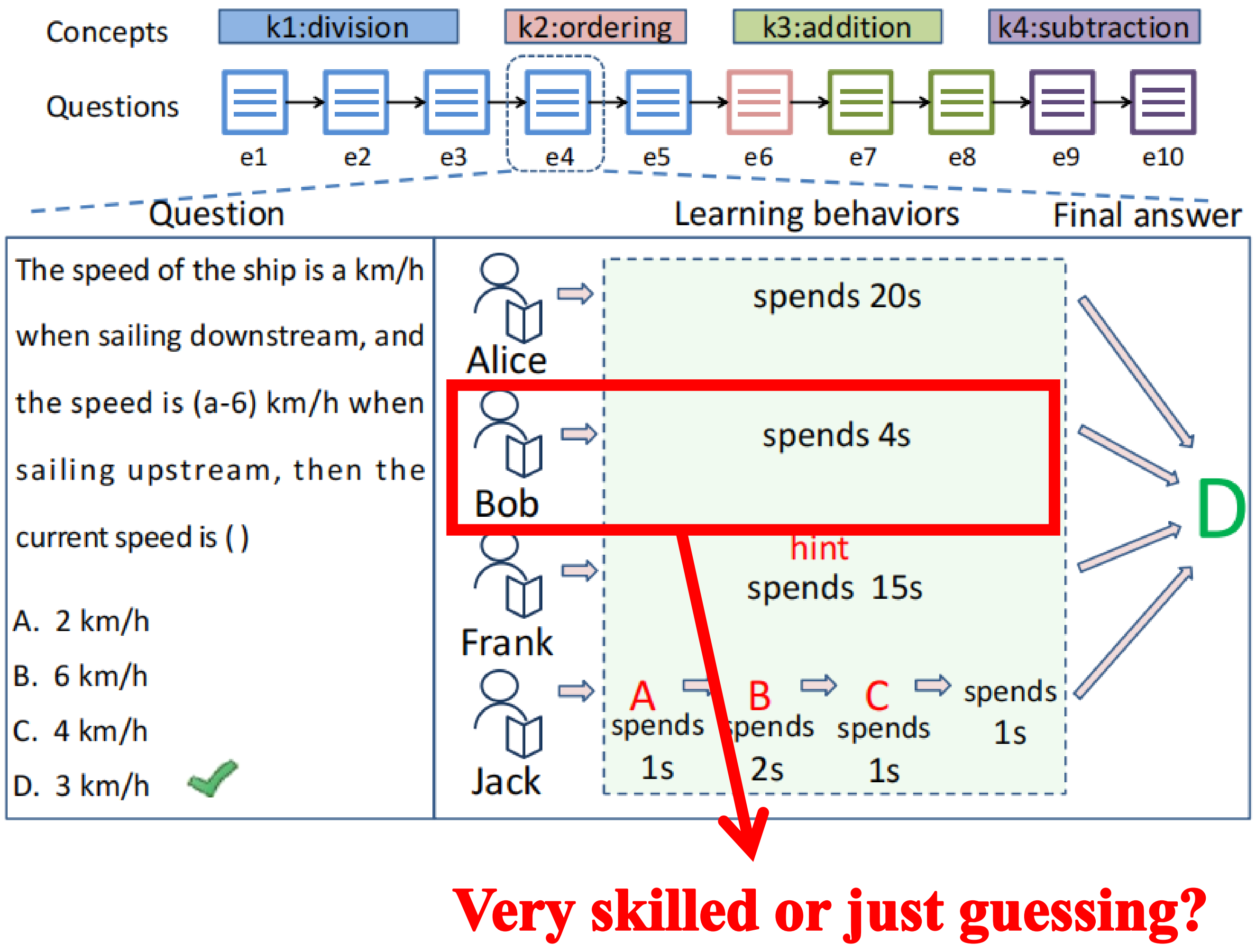
- **Speed**. The response time for a learner to answer a question.
- **Attempts**. The number of attempts to answer a question.
- **Hints**. The number of requested hints to answer a question.



Learning Behaviors

We summarize three typical behaviors and investigate their complex effects on assessing learners' knowledge states:

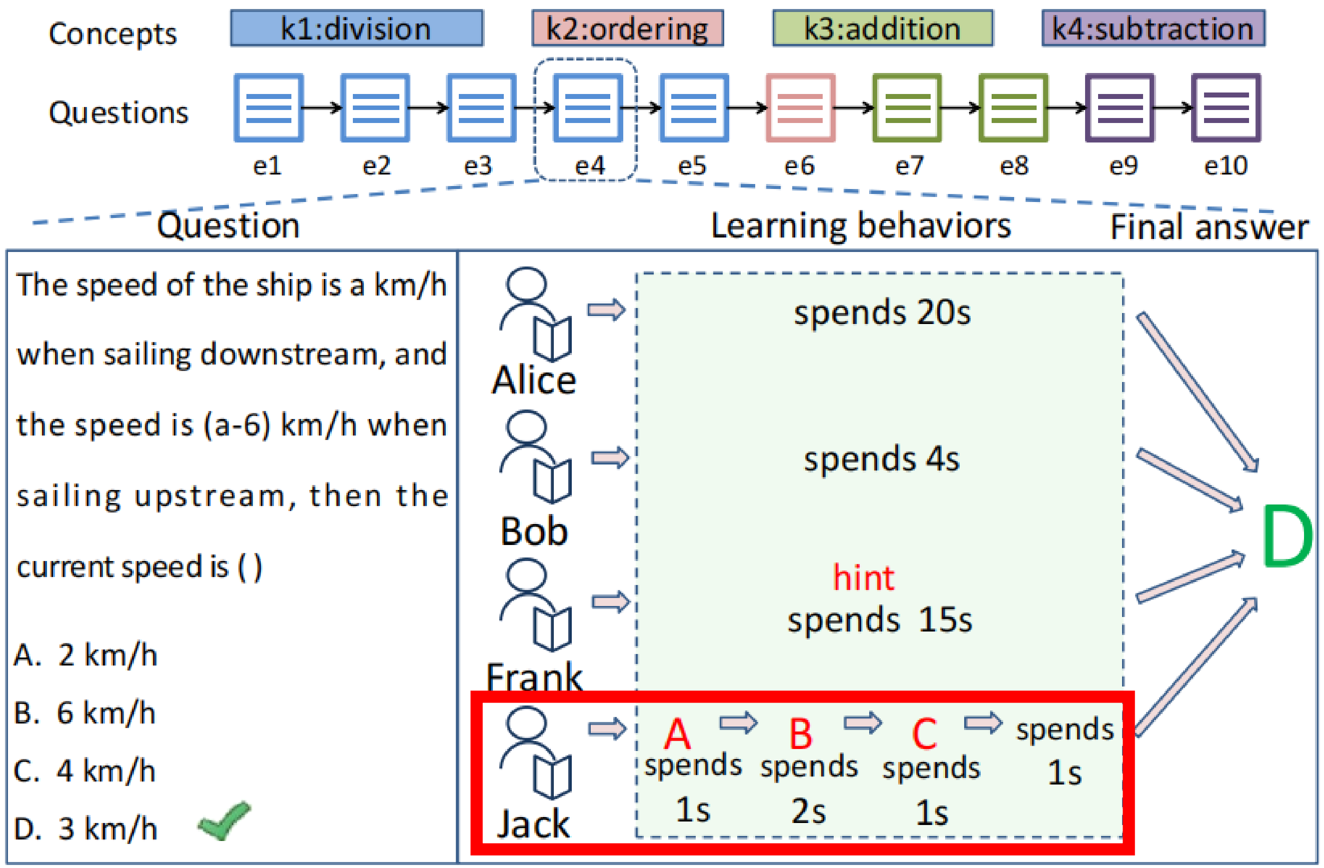
- **Speed**. The response time for a learner to answer a question.
- **Attempts**. The number of attempts to answer a question.
- **Hints**. The number of requested hints to answer a question.



Learning Behaviors

We summarize three typical behaviors and investigate their complex effects on assessing learners' knowledge states:

- **Speed**. The response time for a learner to answer a question.
- **Attempts**. The number of attempts to answer a question.
- **Hints**. The number of requested hints to answer a question.

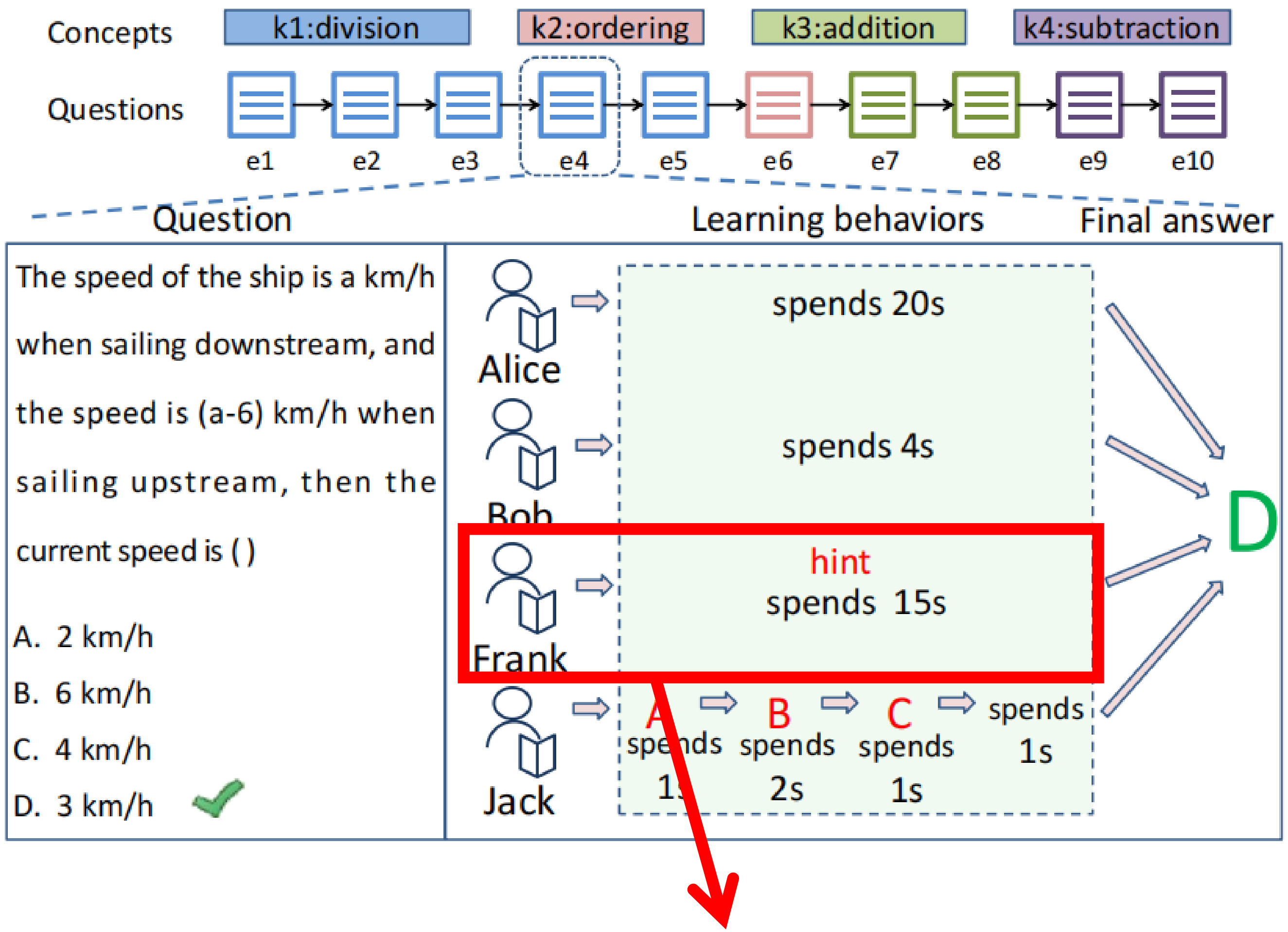


Learning or trying out the right answer?

Learning Behaviors

We summarize three typical behaviors and investigate their complex effects on assessing learners' knowledge states:

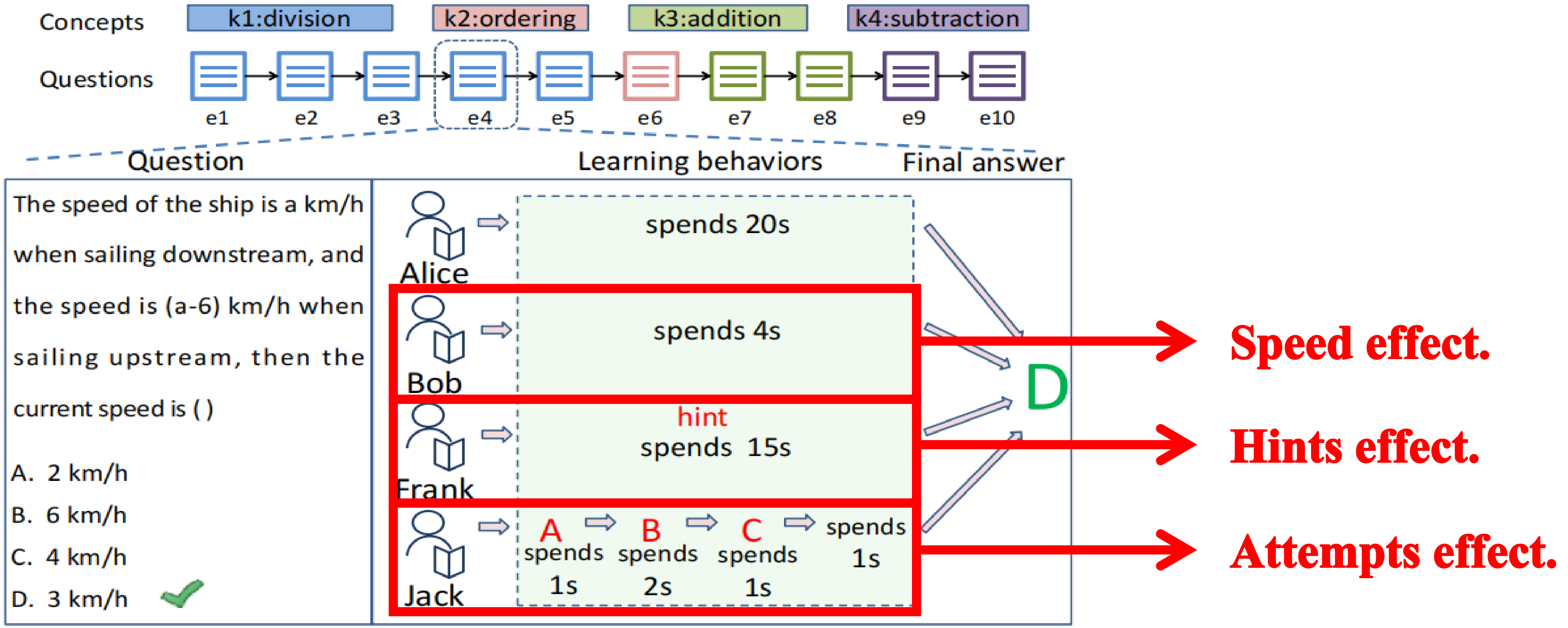
- **Speed**. The response time for a learner to answer a question.
- **Attempts**. The number of attempts to answer a question.
- **Hints**. The number of requested hints to answer a question.



Inspired or abusing the hints?

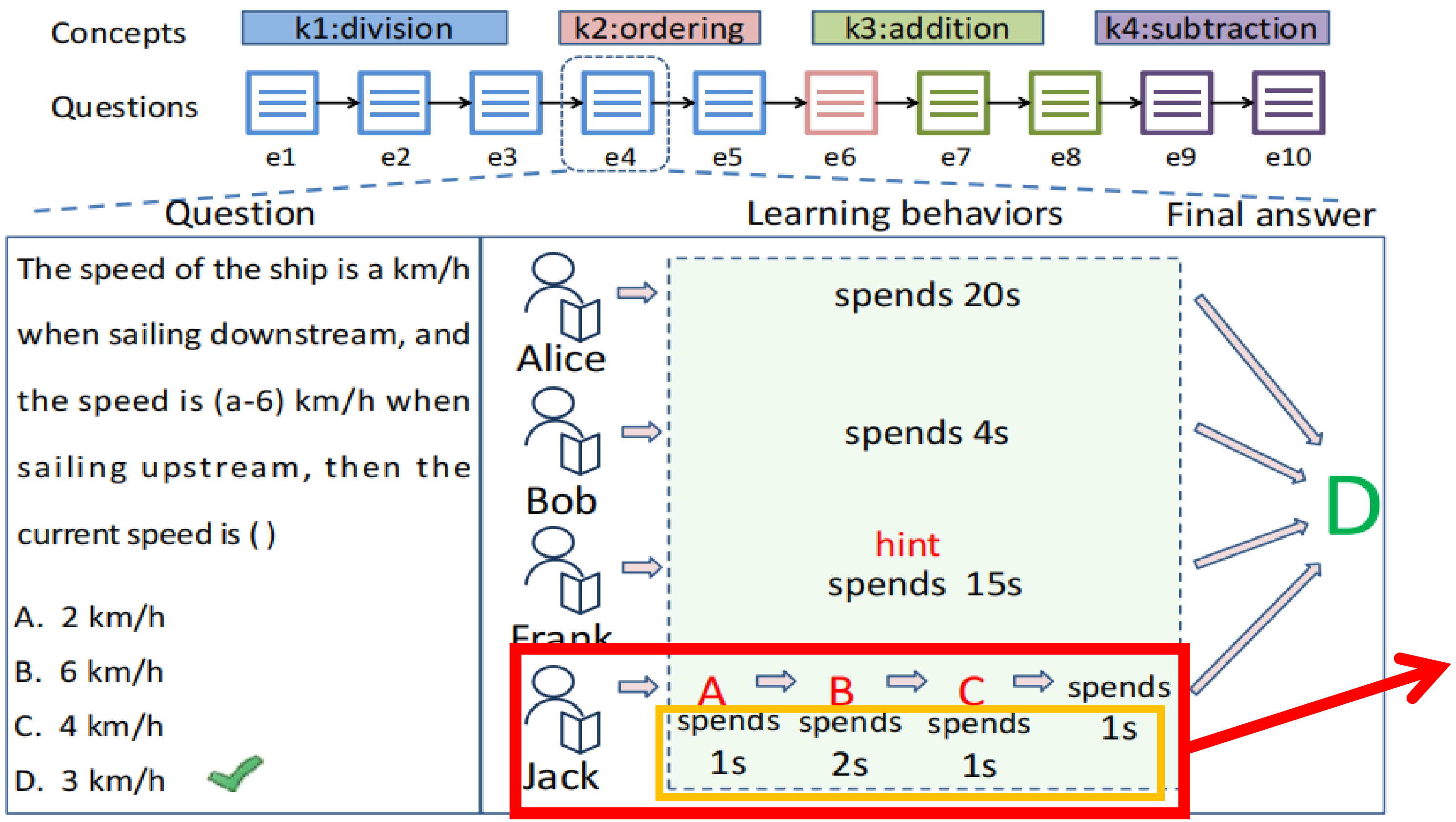
Challenges

- **How to quantify each behavior's effect.**
  - It is difficult to quantify the distinctive effect mechanisms of each behavior on assessing learners' knowledge acquisition.



Challenges

- **How to measure the fused effect of multiple behaviors.**
  - It is a great challenge to capture the complex dependent patterns of multiple learning behaviors.



Combining the Speed and Attempts!

Learning or trying out the right answer?

The latter is the case!

CONTENTS

01 Background

02 Architecture

03 Experimental Results

04 Conclusion

Problem Definition

Input: Learner's learning records $X = \{(e_1, r_1, b_1), (e_2, r_2, b_2), \ldots, (e_T, r_T, b_T)\}$ at each time step.

e_t : question at time step t

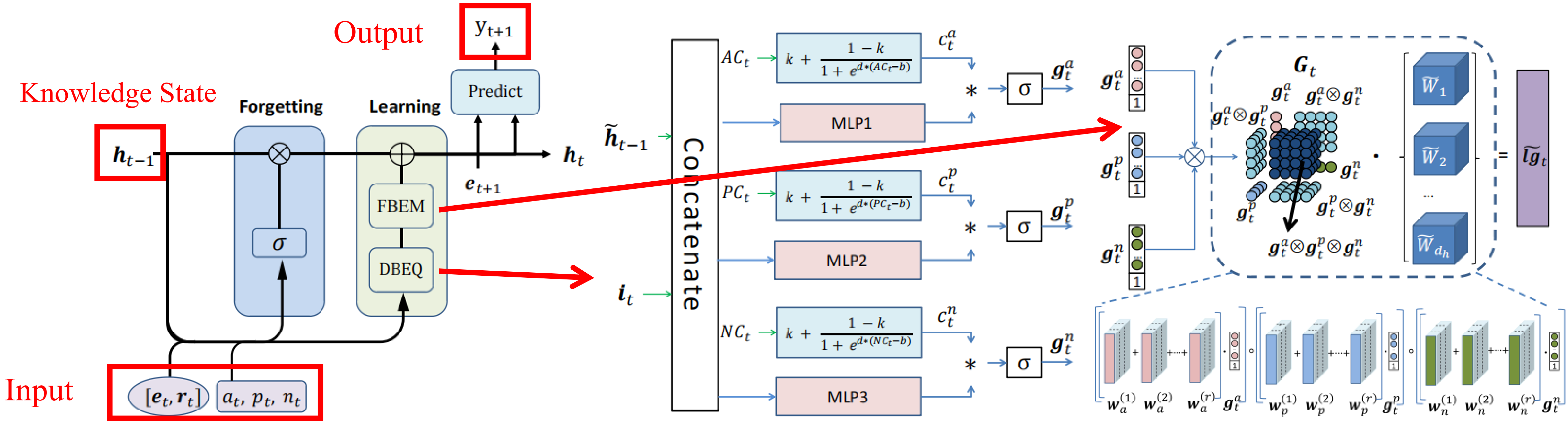
r_t : response at time step t

b_t : behaviors composed of (a_t, p_t, n_t) representing the Speed, Attempts, and Hints

Output: The predicted learner's performance on next question e_{t+1}.

Learning Behavior-oriented Knowledge Tracing Model

- Evaluate the distinctive and fused effect of multiple learning behaviors
- Combine the forgetting factor and knowledge acquisition.



(a) The framework of LBKT. (b) Differentiated Behavior Effect Quantifying (DBEQ). (c) Fused Behavior Effect Measuring (FBEM).

Learning Sequence Representation

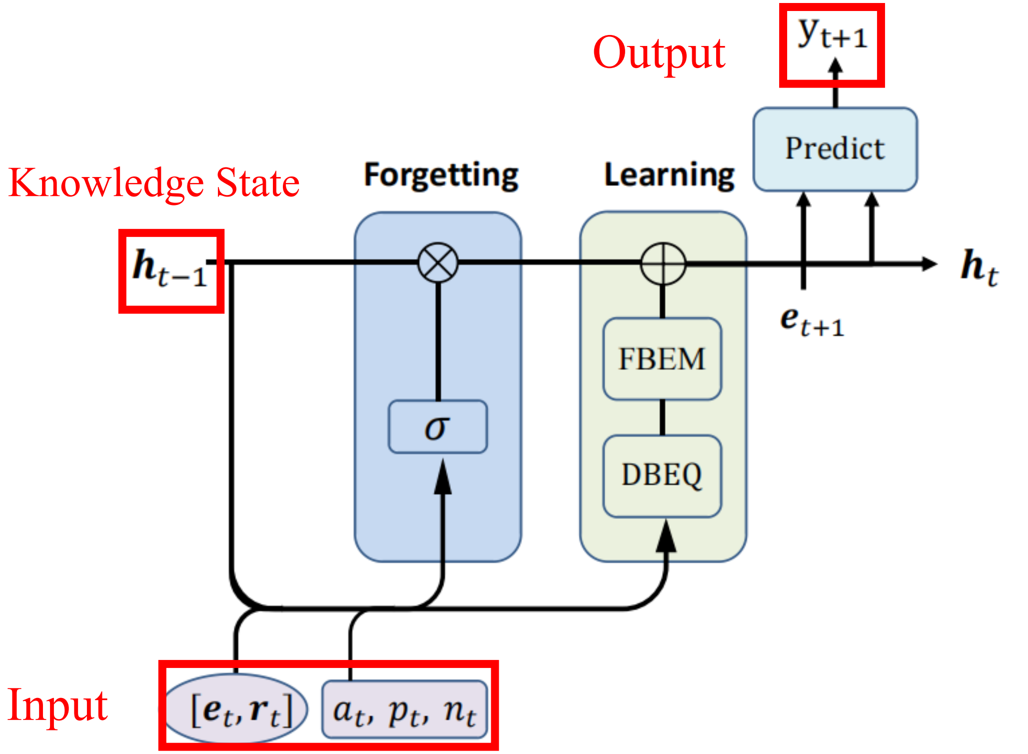
Knowledge State Embedding

$\boldsymbol{h}_t \in \mathbb{R}^{M \times d_h}$ M denotes the number of concepts.

Basic Interaction Representation

$\boldsymbol{e}_t$, $\boldsymbol{r}_t$ denote question and response embeddings respectively,

$$\boldsymbol{i}_t = ReLU(\boldsymbol{W}_1[\boldsymbol{e}_t \oplus \boldsymbol{r}_t] + \boldsymbol{b}_1)$$



(a) The framework of LBKT.

Differentiated Behavior Effect Quantifying

Speed Effect.

The response time of learner i on question j obeys:

$$\ln a_{ji} \sim \mathcal{N}(\mu_j, \sigma_j^2) \quad \textbf{Log-normal distribution}$$

we compute the Speed factor AC_{ji} as:

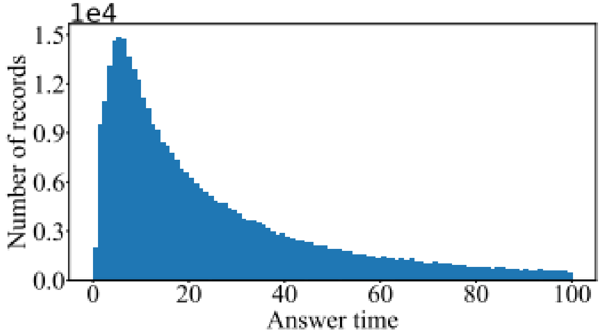
$$AC_{ji} = 1 - P(\mathcal{N}(\mu_j, \sigma_j^2) \le \ln a_{ji})$$

where higher speed correlates to higher AC_{ji}.

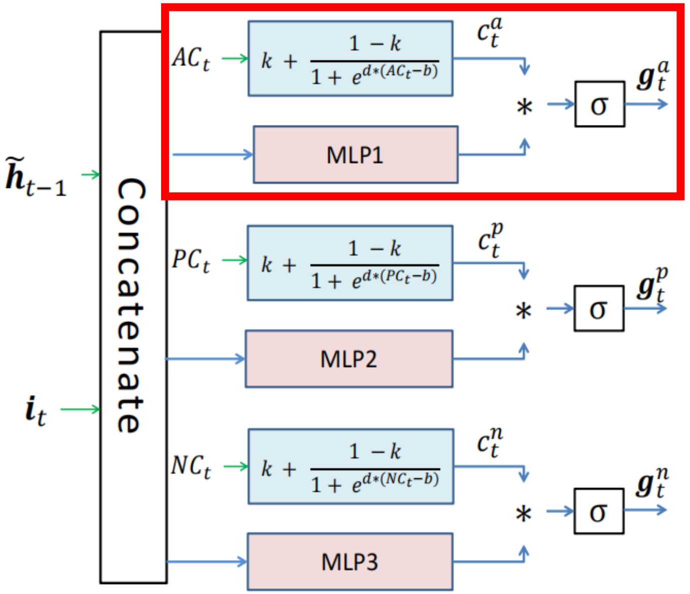
The knowledge acquisition vector $\boldsymbol{g}_t^a$ monitored by Speed factor AC_t at time step t is:

$$c_t^a = k + \frac{1-k}{1+e^{d\cdot(AC_t - b)}},$$

$$\boldsymbol{g}_t^a = \sigma(c_t^a \cdot (\boldsymbol{W}_2^a[\widetilde{\boldsymbol{h}}_{t-1} \oplus \boldsymbol{i}_t] + \boldsymbol{b}_2^a)),$$



(a) Distribution of time on ASSIST2009.



(b) Differentiated Behavior Effect Quantifying (DBEQ).

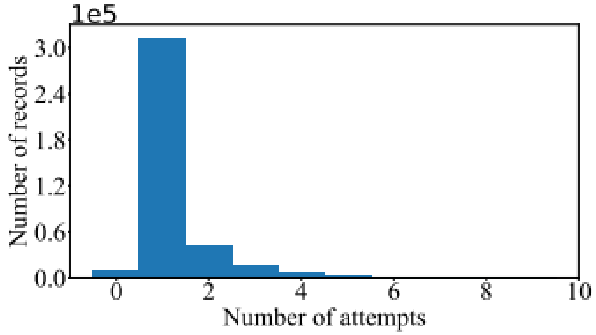
Differentiated Behavior Effect Quantifying

Attempts Effect.

The number of attempts of learner i on question j obeys:

$$p_{ji} \sim \mathcal{P}(\lambda_j^p),$$

Poisson distribution



(b) Distribution of attempts on ASSIST2009.

we compute the Attempts factor PC_{ji} as:

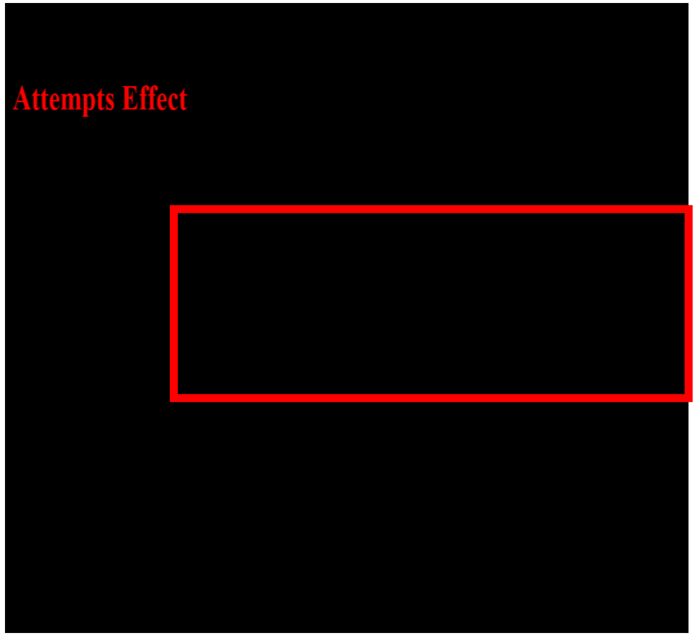
$$PC_{ji} = 1 - P(\mathcal{P}(\lambda_j^p) \geq p_{ji}),$$

where more attempts stands for higher PC_{ji}.

The knowledge acquisition vector $\boldsymbol{g}_t^p$ monitored by Attempts factor PC_t at time step t is:

$$c_t^p = k + \frac{1-k}{1+e^{d\cdot(PC_t - b)}},$$

$$\boldsymbol{g}_t^p = \sigma(c_t^p \cdot (\boldsymbol{W}_2^p[\widetilde{\boldsymbol{h}}_{t-1} \oplus \boldsymbol{i}_t] + \boldsymbol{b}_2^p)),$$

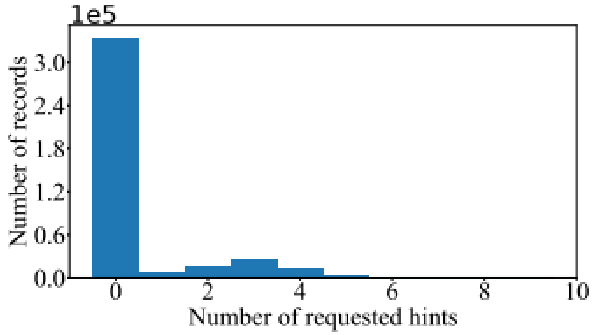


Differentiated Behavior Effect Quantifying

Hints Effect.

The number of requested hints of learner i on question j obeys:

$$n_{ji} \sim \mathcal{P}(\lambda_j^n) \qquad \textbf{Poisson distribution}$$



(c) Distribution of hints on ASSIST2009.

we compute the Hints factor NC_{ji} as:

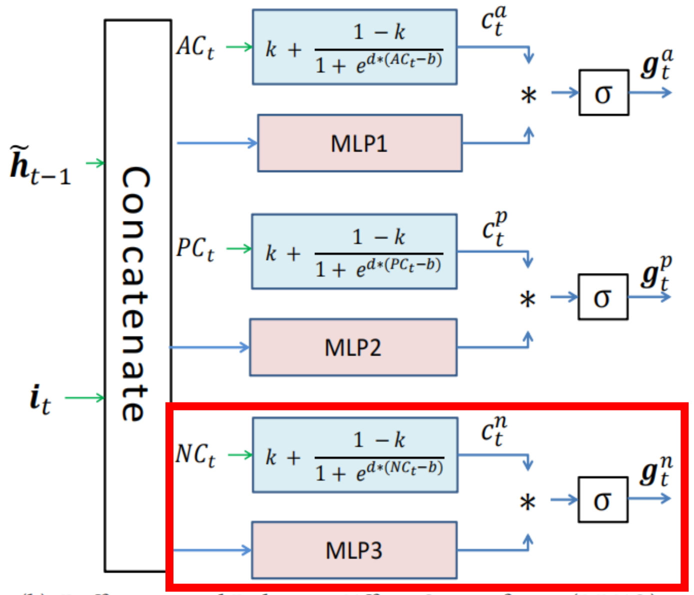
$$NC_{ji} = 1 - P(\mathcal{P}(\lambda_j^n) \geq n_{ji})$$

where more hints used equals to higher NC_{ji}.

The knowledge acquisition vector $\boldsymbol{g}_t^n$ monitored by Hints factor NC_t at time step t is:

$$c_t^n = k + \frac{1-k}{1+e^{d\cdot(NC_t - b)}},$$

$$\boldsymbol{g}_t^n = \sigma(c_t^n \cdot (\boldsymbol{W}_2^n[\widetilde{\boldsymbol{h}}_{t-1} \oplus \boldsymbol{i}_t] + \boldsymbol{b}_2^n)),$$



(b) Differentiated Behavior Effect Quantifying (DBEQ).

Fused Behavior Effect Measuring

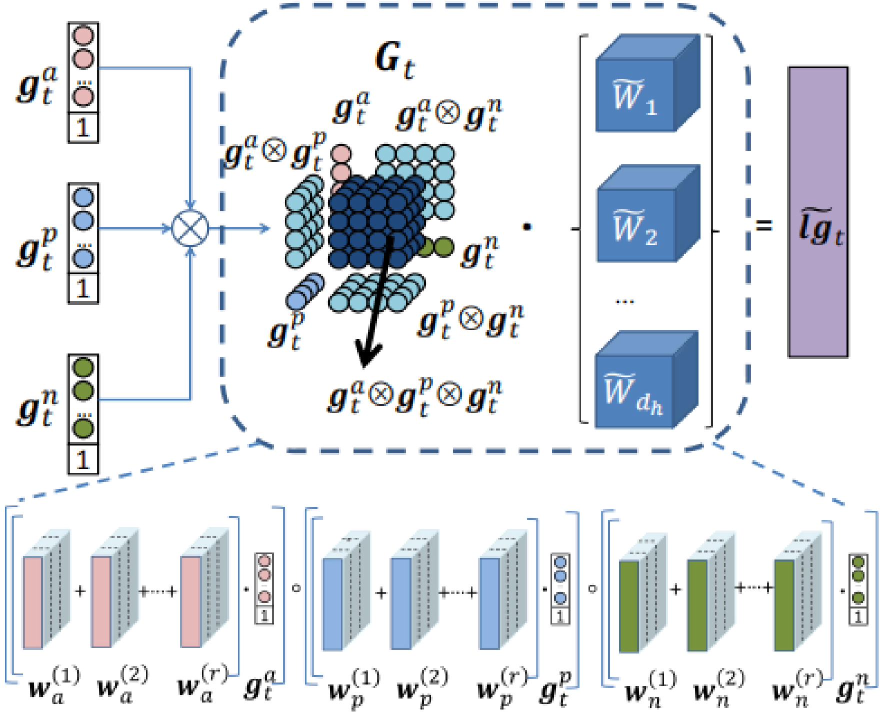
To capture the dependency among different behaviors.

$$G_t = \begin{bmatrix} g_t^a \\ 1 \end{bmatrix} \otimes \begin{bmatrix} g_t^p \\ 1 \end{bmatrix} \otimes \begin{bmatrix} g_t^n \\ 1 \end{bmatrix},$$

$$\widetilde{lg}_t = ReLU(W_3 \cdot G_t + b_3),$$

To save the space and computation overhead, we decompose $\boldsymbol{W}_3 \in \mathbb{R}^{d_h \times (d_h+1) \times (d_h+1) \times (d_h+1)}$ by:

$$\widetilde{\boldsymbol{W}}_k = \sum_{i=1}^{R} (\boldsymbol{w}_{a,k}^{(i)} \otimes \boldsymbol{w}_{p,k}^{(i)} \otimes \boldsymbol{w}_{n,k}^{(i)}),$$



(c) Fused Behavior Effect Measuring (FBEM).

Assuming that $\boldsymbol{W}_3$ is staked by $\widetilde{\boldsymbol{W}}_k \in \mathbb{R}^{(d_h+1) \times (d_h+1) \times (d_h+1)}, k = 1, \dots, d_h$

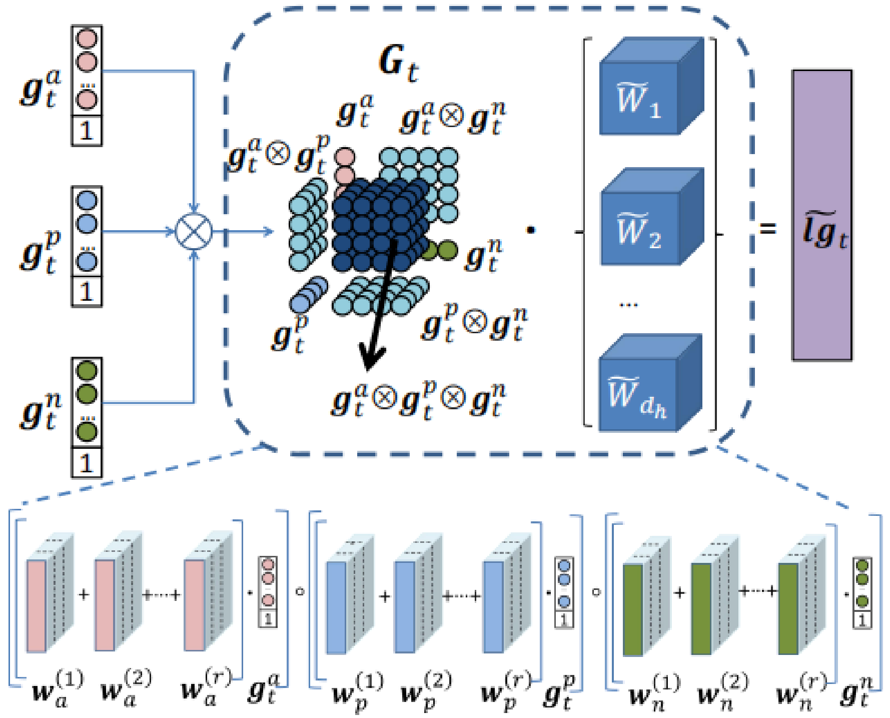
Fused Behavior Effect Measuring

To capture the dependency among different behaviors.

$$\widetilde{lg}_t = ReLU(W_3 \cdot G_t + b_3),$$

The high-order multiplication is transferred to:

$$\begin{aligned} W_3 \cdot G_t &= \left(\sum_{i=1}^{r} (w_a^{(i)} \otimes w_p^{(i)} \otimes w_n^{(i)})\right) \cdot G_t \\ &= \sum_{i=1}^{r} (w_a^{(i)} \otimes w_p^{(i)} \otimes w_n^{(i)}) \cdot G_t \\ &= \sum_{i=1}^{r} (w_a^{(i)} \otimes w_p^{(i)} \otimes w_n^{(i)}) \cdot \left(\begin{bmatrix} g_t^a \\ 1 \end{bmatrix} \otimes \begin{bmatrix} g_t^p \\ 1 \end{bmatrix} \otimes \begin{bmatrix} g_t^n \\ 1 \end{bmatrix} \right) \\ &= \left(\sum_{i=1}^{r} w_a^{(i)} \cdot \begin{bmatrix} g_t^a \\ 1 \end{bmatrix}\right) * \left(\sum_{i=1}^{r} w_p^{(i)} \cdot \begin{bmatrix} g_t^p \\ 1 \end{bmatrix}\right) * \left(\sum_{i=1}^{r} w_n^{(i)} \cdot \begin{bmatrix} g_t^n \\ 1 \end{bmatrix}\right), \end{aligned}$$



(c) Fused Behavior Effect Measuring (FBEM).

$w_a^{(i)}, w_p^{(i)}, w_n^{(i)} \in \mathbb{R}^{d_h \times (d_h+1)}$

$O\big(d_h \times (d_h+1) \times (d_h+1) \times (d_h+1)\big)$ **to** $O(r \times d_h \times (d_h+1))$.

Knowledge State Updating

Consider the influence of both the forgetting factor and knowledge acquisition:

$$f_t = \sigma(W_4[\boldsymbol{h}_{t-1} \oplus \boldsymbol{i}_t \oplus \boldsymbol{AC}_t \oplus \boldsymbol{PC}_t \oplus \boldsymbol{NC}_t] + \boldsymbol{b}_4),$$
$$\boldsymbol{h}_t = f_t * \boldsymbol{h}_{t-1} + \boldsymbol{lg}_t,$$

Performance Prediction

$$\widetilde{\boldsymbol{h}}_t = \boldsymbol{q}_{e_{t+1}} \cdot \boldsymbol{h}_t,$$
$$y_{t+1} = \sigma(W_5[\widetilde{\boldsymbol{h}}_t \oplus \boldsymbol{e}_{t+1}] + \boldsymbol{b}_5),$$

Training Objective

$$\mathbb{L} = -\sum_{t=1}^{T}(r_t \log y_t + (1 - r_t)\log(1 - y_t)).$$

CONTENTS

01 Background

02 Architecture

03 Experimental Results

04 Conclusion

Datasets

- **ASSIST2009** and **ASSIST2012** are both collected from the ASSISTments online tutoring system.
- **Junyi** is collected from Junyi Academy, a Chinese e-learning platform. We select 1000 most active learners.

Statistics	Datasets		
	ASSIST2009	ASSIST2012	Junyi
Records	297,343	2,622,857	4,316,340
Learners	3,006	22,397	1,000
Questions	9,798	37,413	701
Concepts	107	254	39
Avg.attempts	1.532	1.354	1.417
Avg.time (s)	51.220	54.322	208.398
Avg.hints	0.428	0.394	0.249

Records without concepts and learners whose answering sequence is less than 10 are removed. Questions answered less than 10 times are also removed.

Experimental Results

Baselines

- **RNN-based**: **DKT** (Piech et al., 2015), **DKT_concat** (an variant of DKT), **AT-DKT** (Liu et al, 2023)
- **Memory-based**: **DKVMN** (Zhang et al., 2017), **DMKT** (Wang et al., 2021)
- **Attention-based**: **SAKT** (Pandey et al., 2019), **AKT** (Ghosh et al., 2020)
- **Learning-forgetting paradigm**: **LPKT** (Shen et al., 2021)

Method	DKT	DKT_concat	AT-DKT	DKVMN	DMKT	SAKT	AKT	LPKT
With Behavior?	×	√	×	×	√	×	×	√

Learner performance prediction Results

Methods	ASSIST2009			ASSIST2012			Junyi		
	RMSE	AUC	ACC	RMSE	AUC	ACC	RMSE	AUC	ACC
DKT	0.4372	0.7430	0.7127	0.4226	0.7364	0.7333	0.3536	0.7586	0.8326
DKT_concat	0.4335	0.7508	0.7204	0.4193	0.7447	0.7407	0.3518	0.7655	0.8340
AT-DKT	0.4370	0.7574	0.7172	0.4162	0.7544	0.7440	0.3537	0.7581	0.8325
DKVMN	0.4416	0.7400	0.7038	0.4224	0.7351	0.7359	0.3544	0.7565	0.8324
DMKT	0.4370	0.7569	0.7196	0.4224	0.7377	0.7383	0.3538	0.7586	0.8336
SAKT	0.4545	0.7111	0.6885	0.4236	0.7335	0.7333	0.3544	0.7590	0.8323
AKT	0.4273	0.7766	0.7289	0.4084	<u>0.7760</u>	0.7559	0.3538	0.7593	0.8325
LPKT	<u>0.4236</u>	<u>0.7788</u>	<u>0.7325</u>	<u>0.4078</u>	0.7751	<u>0.7567</u>	<u>0.3509</u>	<u>0.7689</u>	<u>0.8344</u>
LBKT	**0.4203***	**0.7863***	**0.7380***	**0.4043***	**0.7823***	**0.7613***	**0.3494***	**0.7723***	**0.8362***

- Classic models which applies behaviors (DKT_concat, DMKT and LPKT) outperform other classic models.
- LBKT significantly outperforms all baseline methods on all datasets and evaluation metrics.

Experimental Results

Ablation Study

- **LBKT_None** : LBKT fed none behaviors.
- **LBKT_Attempt**: LBKT fed only Attempts.
- **LBKT_Speed** : LBKT fed only Speed.
- **LBKT_Hint**: LBKT fed only Hints.

Table 5: Results of learners' performance prediction for LBKT fed with different behaviors. The second best results are underlined and the best results are bold.

Methods	ASSIST2009			ASSIST2012			Junyi		
	RMSE	AUC	ACC	RMSE	AUC	ACC	RMSE	AUC	ACC
LBKT	**0.4203**	**0.7863**	**0.7380**	**0.4043**	**0.7823**	**0.7613**	**0.3494**	**0.7723**	**0.8362**
LBKT_None	0.4279	0.7733	0.7276	0.4082	0.7759	0.7549	0.3510	0.7681	0.8341
LBKT_Speed	0.4206	0.7854	0.7373	0.4062	0.7781	0.7583	0.3501	0.7699	0.8356
LBKT_Attempt	0.4224	0.7835	0.7336	0.4058	0.7792	0.7588	0.3506	0.7685	0.8342
LBKT_Hint	0.4223	0.7831	0.7335	0.4066	0.7786	0.7587	0.3503	0.7698	0.8354

- All the three behaviors are necessary compared with LBKT-None.
- The Speed behavior contributes most to LBKT

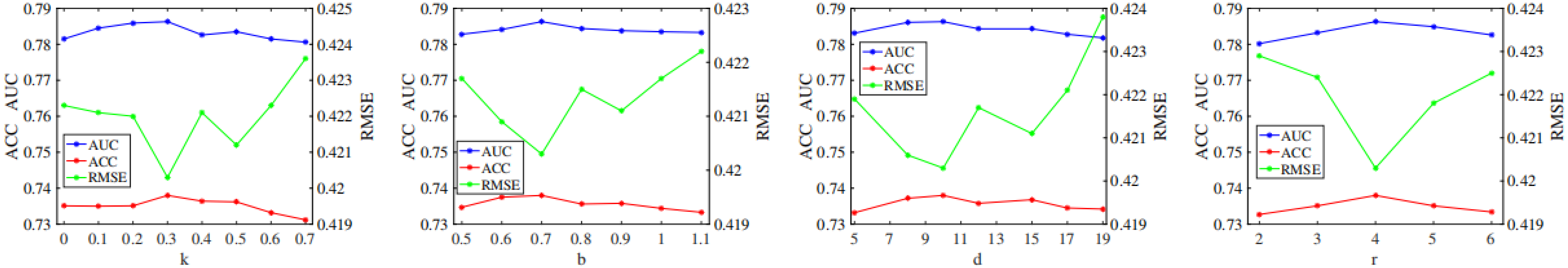
Parameter Sensitivity

$$c_t^a = \boxed{k} + \frac{1-k}{1+e^{\boxed{d}\cdot(AC_t - \boxed{b})}},$$

Differentiated Behavior Effect Quantifying Module

Fused Behavior Effect Measuring Module

$$W_3 \cdot G_t = \Big(\sum_{i=1}^{\boxed{r}} (w_a^{(i)} \otimes w_p^{(i)} \otimes w_n^{(i)})\Big) \cdot G_t$$



(a) Sensitivity of k. (b) Sensitivity of b. (c) Sensitivity of d. (d) Sensitivity of r.

- LBKT achieves best performance when k = 0.3, b = 0.7, d = 10, r = 4.
- LBKT shows stable ability to the different levels of parameters.

Association between behaviors and knowledge acquisition

Normalized Learning Gain (NL_Gain):

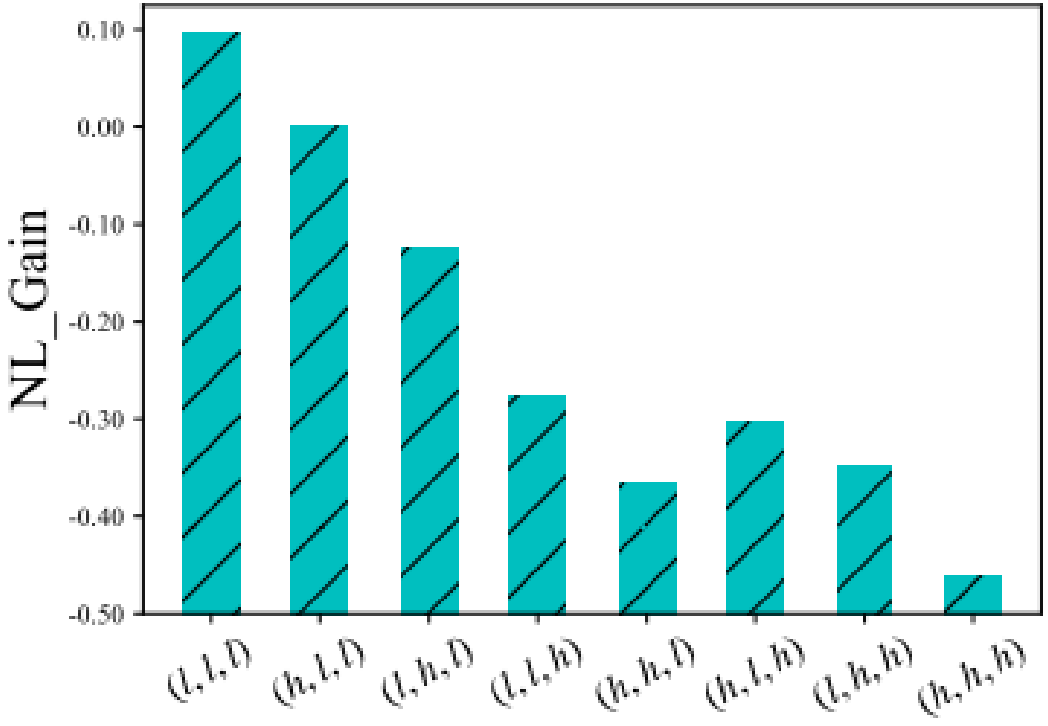
$$y_t = \sigma(\boldsymbol{W}_5 [\widetilde{\boldsymbol{h}}_{t-1} \oplus \boldsymbol{e}_t] + \boldsymbol{b}_5),$$

$$y'_t = \sigma(\boldsymbol{W}_5 [\widetilde{\boldsymbol{h}}_t \oplus \boldsymbol{e}_t] + \boldsymbol{b}_5),$$

$$NL\_Gain = \frac{y'_t - y_t}{1 - y_t},$$

Learning gain after answering a question.

For each behavior, we classify records into high and low groups.

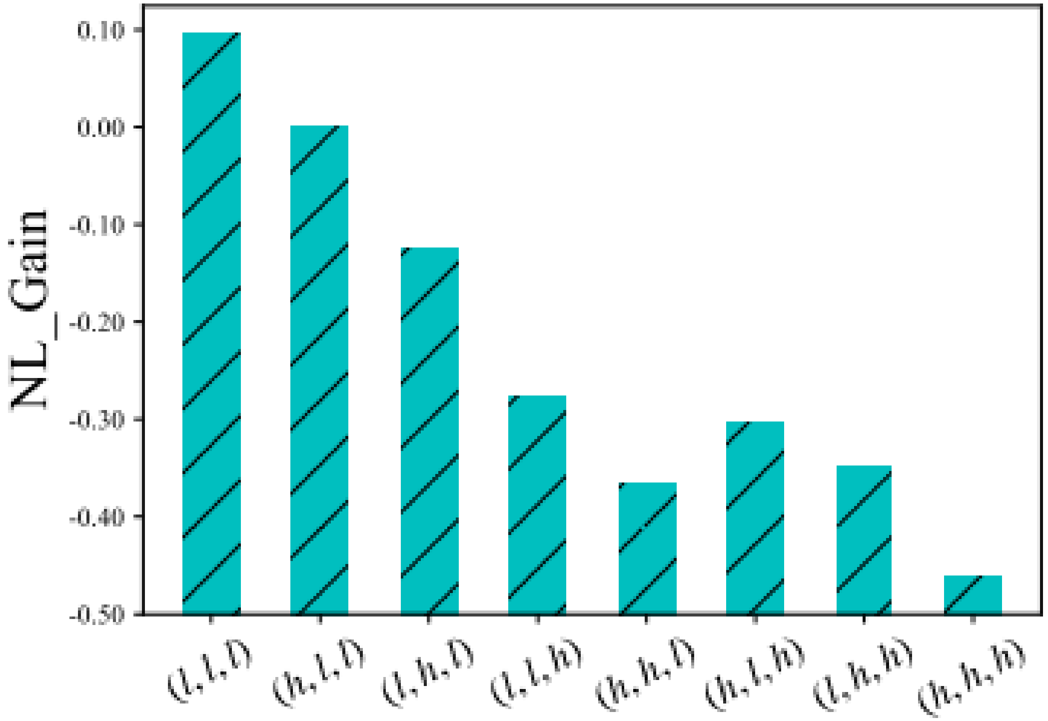


X axis: each group
Y axis: the NL_Gain value

Experimental Results

Association between behaviors and knowledge acquisition

- The NL_Gain of group (*l* ,*l* ,*l*) is the highest while group (*h*, *h*, *h*) is the lowest.
- The NL_Gain of groups that infer a high factor on just one behavior is higher than those groups that infer high factors on two or three behaviors.



Higher Speed, more attempts and more hints used correlate to poorer learning gain.

X axis: each group
Y axis: the NL_Gain value

Visualization of Proficiency Evolution

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
Concept	①	①	②	②	②	②	③	③	③
Resp	✓	✓	✓	✗	✓	✓	✓	✓	✗
①	0.91	0.87	0.84	0.84	0.83	0.83	0.83	0.83	0.83
②	0.44	0.74	0.65	0.64	0.78	0.71	0.68	0.63	0.62
③	0.59	0.59	0.60	0.59	0.60	0.60	0.72	0.69	0.59

Speed factor

	e_1	e_2	e_3	e_4	e_5	e_6	e_7	e_8	e_9
Time	30	25	14	66	61	19	121	6	8
Attempts	1	1	1	2	1	1	1	1	3
Hints	0	0	0	4	0	0	0	0	3
AC	0.76	0.79	0.89	0.36	0.41	0.80	0.04	0.82	0.97
PC	0.31	0.29	0.26	0.52	0.18	0.19	0.16	0.21	0.73
NC	0.00	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.99

LBKT tracks learners' proficiency not only based on their responses but also considers different behaviors' effect:

- Although e_3 is answered correctly, the proficiency on the corresponding concept decreases from 0.74 to 0.65 due to high speed.
- The decline of proficiency after answering e_9 is driven by the fused effect.

CONTENTS

01 Background

02 Architecture

03 Experimental Results

04 Conclusion

Conclusion

- We pointed out the significant effects of learning behaviors on knowledge tracing.
- We proposed LBKT model to quantify the distinctive and cooperative effects of behaviors on knowledge acquisition as well as the forgetting factor.
- Experimental results on three public datasets showed that LBKT outperformed previous classic KT methods.

- In the future, we will try to incorporate more behaviors and deeply mine how these behaviors affect learners' knowledge states.

Thank You for listening!

Author: Bihan Xu
Email: xbh0720@mail.ustc.edu.cn