Knowledge Tracing Model with Learning and Forgetting Behavior

Conference Paper · October 2022

DOI: 10.1145/3511808.3557622

6 authors, including:

Quanlong Guan
Jinan University (Guangzhou, China)

Liangda Fang
University of Jinan (Jinan, China)

Knowledge Tracing Model with Learning and Forgetting Behavior

Mingzhi Chen
College of Information Science and Technology, Jinan University
Guangzhou, China
hawks19970722@stu2020.jnu.edu.cn

Quanlong Guan*
Guangdong Institute of Smart Education, Jinan University
Guangzhou, China
gql@jnu.edu.cn

Yizhou He
College of Management, Jinan University
Guangzhou, China
hyz@jnu.edu.cn

Zhenyu He
College of Information Science and Technology, Jinan University
Guangzhou, China
tzhenyuhe@jnu.edu.cn

Liangda Fang
College of Information Science and Technology, Jinan University
Guangzhou, China
fangld@jnu.edu.cn

Weiqi Luo
Guangdong Institute of Smart Education, Jinan University
Guangzhou, China
lwq@jnu.edu.cn

ABSTRACT

The Knowledge Tracing (KT) task aims to trace the changes of students' knowledge state in real time according to students' historical learning behavior, and predict students' future learning performance. The modern KT models have two problems. One is that these KT models can't reflect students' actual knowledge level. Most KT models only judge students' knowledge state based on their performance in exercises, and poor performance will lead to a decline in knowledge state. However, the essence of students' learning process is the process of acquiring knowledge, which is also a manifestation of learning behavior. Even if they answer the exercises incorrectly, they will still gain knowledge. The other problem is that many KT models don't pay enough attention to the impact of students' forgetting behavior on the knowledge state in the learning process. In fact, learning and forgetting behavior run through students' learning process, and their effects on students' knowledge state shouldn't be ignored. In this paper, based on educational psychology theory, we propose a knowledge tracing model with learning and forgetting behavior (LFBKT). LFBKT comprehensively considers the factors that affect learning and forgetting behavior to build the knowledge acquisition layer, knowledge absorption layer and knowledge forgetting layer. In addition, LFBKT introduces difficulty information to enrich the information of the exercise itself, while taking into account other answering performances besides the answer. Experimental results on two public datasets show that LFBKT can better trace students' knowledge state and outperforms existing models in terms of ACC and AUC.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Applied computing** → **Computer-assisted instruction**.

KEYWORDS

Educational Data Mining; Knowledge Tracing; Learning and Forgetting

ACM Reference Format:
Mingzhi Chen, Quanlong Guan, Yizhou He, Zhenyu He, Liangda Fang, and Weiqi Luo. 2022. Knowledge Tracing Model with Learning and Forgetting Behavior. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management (CIKM '22), October 17–21, 2022, Atlanta, GA, USA.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3511808.3557622

1 INTRODUCTION

The online education system advocates teaching students in accordance with their aptitude, real-time tracing of students' knowledge state is essential for personalized online education [1]. KT can model users' answers to assess their knowledge state and predict their answers at the next moment [2].

In recent years, in the field of KT research, a large number of KT models based on deep learning have been proposed. Deep Knowledge Tracing (DKT) based on recurrent neural network (DKT) [3] uses the hidden vector of recurrent neural network to represent the knowledge state of students, and predicts students' performance accordingly. Dynamic Key-Value Memory Networks for Knowledge Tracing (DKVMN) [4] borrows ideas from memory networks [5] to predict student performance. In the above KT models, there are two problems: one is that the knowledge state of students in each learning interaction is directly determined by their answering performance in exercise. When a student answers incorrectly, the model will think his/her knowledge state on the corresponding knowledge concept (KC) will decline, which isn't consistent with neurological theory. Brain scientists believe that it is normal to make mistakes, it is a kind of information, from a neurological point of view, it is necessary, myelin growth the necessary condition is that mistakes must be corrected [6]. The essence of students' learning process is the process of constantly making mistakes and correcting them, growing and acquiring knowledge. Second, the above KT model ignores the influence of students' forgetting behavior in the learning process. In the field of KT, only a few studies have considered the forgetting behavior of students. The extension of DKT [7] considers interval time to simulate forgetting behavior. Attentive Knowledge Tracing (AKT) [8] and Aware Self-Attention

*Corresponding authors

CIKM '22, October 17–21, 2022, Atlanta, GA, USA

ACM ISBN 978-1-4503-9236-5/22/10…$15.00
https://doi.org/10.1145/3511808.3557622

for Knowledge Tracing (RKT) [9] design an exponentially decaying kernel function to simulate forgetting behavior. However, the modeling of forgetting behavior in these studies only considers the time factor, which is insufficient to simulate complex forgetting behavior. In the field of educational psychology, many scholars have recognized the forgetting behavior of human beings and explored the factors that affect human forgetting. The Ebbinghaus forgetting curve theory [11] and the theory of memory trace decay [12] in educational psychology theory propose that students will forget what they have learned. The interval time and previous knowledge state will affect students' forgetting degree.

In order to solve the above two problems, on the basis of DKT [7], Learning Process-consistent Knowledge Tracing (LPKT) [10] and other models, focusing on students' learning and forgetting behavior and combining pedagogical theory, we propose a KT model, LFBKT, which models students' learning and forgetting behavior to trace the changes of students' knowledge state and predict students' performance. Based on educational psychology, LFBKT incorporates three factors that affect knowledge learning and forgetting into modeling: the number of repeated learning of KC, interval time, and students' current knowledge state. In order to pay attention to the impact of students' performance in the process of answering exercises, the answer time and the total number of hints requested are taken into consideration. Besides, inspired by Item Response Theory (IRT) [13], the difficulty information of exercise is considered to enrich the information of exercise itself. The experimental verification results on two real online education datasets show that: LFBKT can effectively model students' learning behavior and forgetting behavior, trace students' knowledge state in real time, and the prediction performance of LFBKT model is better than existing models.

2 THE LFBKT MODEL

We define $E = \{e_1, e_2, \dots, e_t, \dots\}$ as the set of exercises, and the knowledge points involved in each exercise are defined in the Q matrix. $\{(e_1, p_1), it_1, (e_2, p_2), \dots, (e_t, p_t), it_t, \dots\}$ is student's learning history, p_t refers to the student's answering performance, and it_t refers to the interval time. Student acquires knowledge during the learning process and also forgets some knowledge in the interval between two studies, which leads to the growth and decline of the knowledge state. Also, the performance of e_{t+1} can be predicted by the student's knowledge state s_t.

The overall framework of LFBKT model is shown in Figure 1.

2.1 Learning Performance Unit

We denote a basic unit of student's learning performance by lu_t, which includes answering performance p_t and exercise information. p_t reflects the proficiency of students in applying knowledge, and includes answers $a_t \in \mathbb{R}^{d_a}$, answer time $at_t \in \mathbb{R}^{d_k}$, and the total number of hints requested $ht_t \in \mathbb{R}^{d_k}$, d_k and d_a are dimensions of the vectors. The mainstream approach to characterizing exercise information is to directly represent the exercise as the skill examined by the exercise, which ignores the information of exercise itself. Inspired by the IRT, we introduce the difficulty information

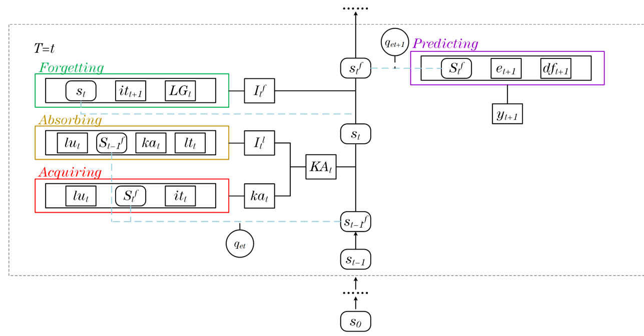

Figure 1: LFBKT framework

of exercise to enrich the information of exercise itself.

$$diff = 1 - \frac{\sum_{i=1}^{N} G_i}{N} \tag{1}$$

We made a statistic on the answers of each exercise on the entire data set, and calculated the probability that it was answered correctly by the students. As shown in Equation 1, $\sum_{i=1}^{N} G_i$ is the total score of the students on this exercise, and a correct answer is 1 point, a wrong answer is 0 point, and N is the total number of students. We use the normalization method to map the probability of the correct answer to an interval in the range $(0, 10)$ as exercise difficulty. Then, the difficulty embedding of each exercise is expressed as $df_t \in \mathbb{R}^{d_k}$. Therefore, lu_t is expressed as:

$$lu_t = W_1^T \left[e_t \oplus df_t \oplus ht_t \oplus at_t \oplus a_t \right] + b_1 \tag{2}$$

where $e_t \in \mathbb{R}^{d_e}$ is the exercise embedding and d_e is dimension of the vector, $W_1 \in \mathbb{R}^{(d_e + 3 \times d_k + d_a) \times d_k}$ is the weight matrix, $\oplus$ is the connection operation, $b_1 \in \mathbb{R}^{d_k}$ is the bias term. The following W_i is the weight matrix of the corresponding network layer, and b_i is the corresponding bias term.

2.2 Knowledge Acquisition Layer

Students will update their knowledge mastery state after learning, and will forget a certain proportion of the knowledge they have learned between two learning intervals, resulting in a decline in the degree of knowledge mastery. Therefore, the knowledge state s_{t-1} at time $t-1$ is reduced to s_{t-1}^f. In order to pay attention to the relevant knowledge state of the current exercise, we first multiply s_{t-1}^f and the KC vector q_{e_t} of the current exercise to obtain the relevant knowledge state S_{t-1}^f:

$$S_{t-1}^f = q_{e_t} s_{t-1}^f \tag{3}$$

where $q_{e_t} \in \mathbb{R}^{d_s}$ is obtained from the Q matrix, d_s refers to the total number of KCs, so q_{e_t} is the KC involved in the question. $s_{t-1}^f \in \mathbb{R}^{d_s \times d_k}$ represents the student's knowledge state, $S_{t-1}^f \in \mathbb{R}^{d_k}$, which indicates the student's knowledge state of the KC examined for this exercise.

The growth of students in the learning process can be assessed by the amount of knowledge acquired. Not all students have the same knowledge acquisition. Knowledge acquisition is directly related to the exercises that students hava done. Exercises that examine different KCs and exercises with different levels of difficulty will bring students different degrees of gain. Exercise information and answering performance are included in lu_t. Then, prior knowledge state S^f_{t-1} also affects students' outcomes, for example, students with lower mastery have more room for improvement. Based on the above theories, we introduce lu_t and S^f_{t-1} to design the knowledge acquisition layer. Besides, students can gain knowledge even if they get the wrong answer. Errors are regarded as a natural factor in the learning process [14]. Students can learn from mistakes and promote learning progress through a good atmosphere of errors [15]. Therefore, we set the knowledge acquisition to be always positive. The range of output value of tanh activation function is $(-1,\ 1)$, so knowledge acquisition ka_t on the corresponding KC can be expressed as:

$$ka_t = \left(\tanh\left(W_2^T\left[lu_t \oplus S^f_{t-1}\right] + b_2\right) + 1\right)/2 \tag{4}$$

2.3 Knowledge Absorption Layer

Not all learning knowledge can be completely absorbed and transformed into the growth of students' knowledge state. The Ebbinghaus curve theory shows that the absorption rate of knowledge will be affected by the following two aspects: the number of repeated learning of KC and the interval time. The interval time can be divided into the interval time of repeated KC learning and the interval time of sequential learning. Considering that the interval time of repeated KC learning is already included in the interval time of sequential learning, we only consider the interval time of sequential learning. The theory of extinction interference inhibition [16] shows that the mutual interference between learning materials will affect students' learning, which indicates that the interval time is a factor affecting learning, and it is often more efficient to learn a knowledge point continuously. Based on the above theories, we design learning gates to model the absorption rate of knowledge I^l_t:

$$I^l_t = \sigma\left(W_3^T\left[lu_t \oplus S^f_{t-1} \oplus ka_t \oplus lt_t \oplus it_t\right] + b_3\right) \tag{5}$$

where $lt_t \in \mathbb{R}^{d_k}$ is the number of repeated learning, $it_t \in \mathbb{R}^{d_k}$ is interval time, σ is the nonlinear *sigmoid* activation function.

The absorption of student's knowledge ka^n_t can be obtained by multiplying the knowledge absorption rate I^l_t by the knowledge acquisition ka_t, which is expressed as:

$$ka^n_t = I^l_t \bullet ka_t \tag{6}$$

By multiplying $q^T_{e_t}$ by lg^n_t, the overall knowledge absorption amount KA_t is obtained, and the formula is expressed as:

$$KA_t = q^T_{e_t} ka^n_t \tag{7}$$

where $q_{e_t} \in \mathbb{R}^{d_s}$, $ka^n_t \in \mathbb{R}^{d_k}$, so $KA_t \in \mathbb{R}^{d_s \times d_k}$ represents the overall knowledge absorption.

After time t, the student's knowledge state is the original knowledge state plus the absorptive capacity of knowledge after learning at time t, so the knowledge state s^f_{t-1} is updated to s_t:

$$s_t = s^f_{t-1} + KA_t \tag{8}$$

2.4 Knowledge Forgetting Layer

Students will forget part of what they have learned, and the impact of forgetting is a decline in the degree of knowledge mastery. Simply setting a time-based exponential decay function is not sufficient to simulate complex forgetting behavior. The theory of memory trace decay suggests that forgetting is caused by the decay of memory traces, which occurs automatically over time [12]. The theory of memory trace decay also suggests that the previous knowledge state affects the degree of forgetting. Therefore, we consider the following factors to design the knowledge forgetting gate to measure the overall forgetting degree of knowledge state.

When $t = 0$, that is, the initial moment, regardless of forgetting behavior, when $t \neq 0$, the degree of forgetting can be expressed as:

$$I^f_t = \sigma\left(W_4^T\ [KA_{t-1} \oplus s_{t-1} \oplus it_t] + b_4\right) \tag{9}$$

We remove the student's forgotten knowledge by multiplying I^f_t by s_{t-1}. Knowledge state after forgetting is updated to:

$$s^f_{t-1} = I^f_t \bullet s_{t-1} \tag{10}$$

2.5 Predicting Layer

The purpose of the prediction layer is to predict the student's performance on the next exercise e_{t+1}. We concatenate the exercise embedding e_{t+1}, the exercise difficulty df_{t+1} and the student's relevant knowledge state S^f_t, and project them to a fully connected network with sigmoid activation. The predicted performance of student y_{t+1} can be expressed as:

$$y_{t+1} = \sigma\left(W_5^T\left[e_{t+1} \oplus df_{t+1} \oplus S^f_t\right] + b_5\right) \tag{11}$$

where y_{t+1} is the range of $(0,\ 1)$, representing the probability that the student provides the correct response to e_{t+1}. LFBKT judges the student's answer by the relationship between y_{t+1} and threshold.

3 EXPERIMENTS

3.1 Training Details

As shown in Equation 12, we optimize the parameters by minimizing the cross-entropy loss function between the predicted value of answer y_t and the real result of answer a_t, and train the model in all experiments using the Adam optimizer [17].

$$\mathbb{L}(\theta) = -\sum_{t=1}^{T}\left(a_t \log y_t + (1-a_t)\log(1-y_t)\right) + \lambda_\theta \|\theta\|^2 \tag{12}$$

We evaluate all models with 5-fold cross-validation. For both datasets, we split 70% of the datasets to train the models and test on the remaining datasets. We initialize all parameters in the distribution randomly and uniformly [18]. The dimension parameters d_k, d_e and d_a are all set to 128. To prevent overfitting, we add a dropout layer [19] with a dropout of 0.2. All model training is done on servers with RTX 3090 GPUs.

3.2 Datasets

To evaluate our model, we used two real-world datasets.

- ASSISTments 2012 (ASSIST2012) is collected by ASSISTments online tutoring platform and is widely used in the KT field [20].

Table 1: Results of comparison methods on performance prediction. LFBKT outperforms all baselines on both datasets.

Model	ASSISTChall		ASSIST2012	
	AUC	ACC	AUC	ACC
DKT	0.7243	0.6927	0.7293	0.7371
DKVMN	0.7108	0.6842	0.7248	0.7346
SAKT	0.7013	0.6794	0.7253	0.7359
SAINT+	0.7345	0.7186	0.7673	0.7417
AKT	0.7563	0.7115	0.7735	0.7551
LPKT	0.7997	0.7405	0.7772	0.7583
LFBKT	0.8454	0.7643	0.7943	0.7651

- ASSISTments Challenge (ASSISTChall) is collected for a data mining competition held in 2017, and its average number of records per student is relatively rich.

For both datasets, we deleted the data with empty number of hints requested, KC, and question id.

3.3 Baseline Models

To demonstrate the effectiveness of LFBKT, we compare our model with the following state-of-the-art KT models.

- DKT [3] is the first model which predicts students' performance using a single layer LSTM model.
- DKVMN [4] uses key matrix and value matrix to represent the relationship between different KCs, and uses the value matrix to represent the students' mastery of each KC.
- SAKT [21] applies the transformer structure to the KT, and uses a self-attention mechanism to identify previously related exercises.
- SAINT+ [22] introduces exercise information and students' response to improve prediction performance of SAINT [23].
- AKT [8] is a transformer-like model with two self-attention encoders, which uses monotonic self-attention to model forget behavior and uses IRT [13] to generate question embeddings.
- LPKT [10] considers the effects of learning and forgetting, and models students' learning process to monitor knowledge state.

3.4 Experimental Results and Analysis

As can be seen from Table 1, LFBKT has different degrees of improvement compared with other models on both datasets, indicating that the learning and forgetting behavior that LFBKT focuses on are effective in KT modeling.

DKT uses the hidden vector of LSTM model to model students' overall knowledge state, and can't model the students' mastery of each KC. Therefore, the prediction performance of DKT on the two datasets is lower than that of LFBKT. Both DKVMN and LFBKT can model students' mastery of each KC, but DKVMN ignores students' forgetting behavior during learning, and defaults to students' mastery of unreviewed KCs. There are certain limitations, so the prediction performance of LFBKT is stronger than that of DKVMN. SAINT+ introduces exercise information and students' response information to optimize the model, which shows excellent prediction performance. This also demonstrates the validity of these considerations in this paper. The experimental results of LPKT and LFBKT outperformed SAKT, SAINT+ and AKT, which indicates that comprehensive modeling of changes in knowledge state during learning process is more effective and can more accurately predict their future performance. Compared with other models, LPKT and LFBKT comprehensively model learning and forgetting behavior during learning, and have clear advantages. LFBKT pays more attention to exercise information and students' answering performance than LPKT, showing better performance.

Table 2: Ablation experiment results on ASSISTChall.

Model	learning	forgetting	time	count	AUC	ACC
LFBKT(NL)	×	✓	✓	✓	0.7965	0.7375
LFBKT(NF)	✓	×	✓	✓	0.8264	0.7504
LFBKT(NT)	✓	✓	×	✓	0.8355	0.7548
LFBKT(NC)	✓	✓	✓	×	0.7978	0.7384

3.5 Ablation Experiments

In this section, we conduct some ablation experiments to show how each module and each parameter in LFBKT affects the final result.

- LFBKT (NL) doesn't consider knowledge absorption rate.
- LFBKT (NF) doesn't consider forgetting behavior.
- LFBKT (NT) doesn't use any time information.
- LFBKT (NC) doesn't use any number of count information.

The results in Table 2 show some interesting conclusions. First, the retention rate of knowledge plays a crucial role in the learning process, and if the retention rate of knowledge is not considered, it will lead to the largest drop in prediction results. Secondly, in KT, forgetting behavior also plays an indispensable role in modeling the knowledge state of students, because it can well simulate the phenomenon that students' knowledge naturally decreases with time, and can better reflect the dynamic changes of students' knowledge state, which improves the performance and inter-pretability of the model. Finally, time information and frequency information are essential information in the entire learning process, especially frequency information. The number of repetitions of learning reflects the process of students consolidating knowledge, and the total number of hints requested reflects the proficiency of students in applying knowledge. If the information is ignored, it will be detrimental to accurately model the learning process.

4 CONCLUSIONS

In this work, we propose a new KT model, LFBKT. Combining some educational psychology theories, the model introduces the difficulty factor, and models the students' learning and forgetting behavior according to various influencing factors, so as to better simulate the students' learning process and obtain higher prediction accuracy. Experimental results on two public datasets show that LFBKT outperforms state-of-the-art models in both ACC and AUC metrics, proving the effectiveness of LFBKT in KT.

ACKNOWLEDGMENTS

This research was supported by NSFC (62077028, 61877029), the Science and Technology Planning Project of Guangdong (2021B0101420003, 2020B1212030003, 2019A050510024, 2020ZDZX3013, 2019B1515120010), and the High Perfor-mance Public Computing Service Platform of Jinan University.

REFERENCES

[1] Qi Liu, Shuanghong Shen, Zhenya Huang, Enhong Chen, and Yonghe Zheng. A survey of knowledge tracing. *arXiv preprint arXiv:2105.15106*, 2021.

[2] Sergio Ivan Ramirez Luelmo, Nour El Mawas, and Jean Heutte. Machine learning techniques for knowledge tracing: A systematic literature review. 2021.

[3] Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J Guibas, and Jascha Sohl-Dickstein. Deep knowledge tracing. *Advances in neural information processing systems*, 28, 2015.

[4] Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web*, pages 765–774, 2017.

[5] Sainbayar Sukhbaatar, Jason Weston, Rob Fergus, et al. End-to-end memory networks. *Advances in neural information processing systems*, 28, 2015.

[6] Jo Boaler. Mistakes grow your brain. *Youcubed at Stanford University Graduate School of Education. Accessed April*, 14:2016, 2016.

[7] Koki Nagatani, Qian Zhang, Masahiro Sato, Yan-Ying Chen, Francine Chen, and Tomoko Ohkuma. Augmenting knowledge tracing by considering forgetting behavior. In *The world wide web conference*, pages 3101–3107, 2019.

[8] Aritra Ghosh, Neil Heffernan, and Andrew S Lan. Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2330–2339, 2020.

[9] Shalini Pandey and Jaideep Srivastava. Rkt: relation-aware self-attention for knowledge tracing. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 1205–1214, 2020.

[10] Shuanghong Shen, Qi Liu, Enhong Chen, Zhenya Huang, Wei Huang, Yu Yin, Yu Su, and Shijin Wang. Learning process-consistent knowledge tracing. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 1452–1460, 2021.

[11] Hermann Ebbinghaus. Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4):155, 2013.

[12] Charles D Bailey. Forgetting and the learning curve: A laboratory study. *Management science*, 35(3):340–352, 1989.

[13] Susan E Embretson and Steven P Reise. *Item response theory*. Psychology Press, 2013.

[14] Julia Käfer, Susanne Kuger, Eckhard Klieme, and Mareike Kunter. The significance of dealing with mistakes for student achievement and motivation: results of doubly latent multilevel analyses. *European journal of psychology of education*, 34(4):731–753, 2019.

[15] Gabriele Steuer and Markus Dresel. A constructive error climate as an element of effective learning environments. 2015.

[16] Oliver Kliegl and Karl-Heinz T Bäuml. The mechanisms underlying interference and inhibition: A review of current behavioral and neuroimaging research. *Brain Sciences*, 11(9):1246, 2021.

[17] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[18] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[19] Ghodai Abdelrahman and Qing Wang. Knowledge tracing with sequential key-value memory networks. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 175–184, 2019.

[20] Mingyu Feng, Neil Heffernan, and Kenneth Koedinger. Addressing the assessment challenge with an online system that tutors as it assesses. *User modeling and user-adapted interaction*, 19(3):243–266, 2009.

[21] Shalini Pandey and George Karypis. A self-attentive model for knowledge tracing. *arXiv preprint arXiv:1907.06837*, 2019.

[22] Dongmin Shin, Yugeun Shim, Hangyeol Yu, Seewoo Lee, Byungsoo Kim, and Youngduck Choi. Saint+: Integrating temporal features for ednet correctness prediction. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 490–496, 2021.

[23] Youngduck Choi, Youngnam Lee, Junghyun Cho, Jineon Baek, Byungsoo Kim, Yeongmin Cha, Dongmin Shin, Chan Bae, and Jaewe Heo. Towards an appropriate query, key, and value computation for knowledge tracing. In *Proceedings of the Seventh ACM Conference on Learning@ Scale*, pages 341–344, 2020.