Learning Ability Community
for Personalized Knowledge Tracing

Juntao Zhang, Biao Li, Wei Song, Nanzhou Lin, Xiandi Yang(✉),
and Zhiyong Peng

School of Computer Science, Wuhan University, Wuhan, Hubei, China
{juntaozhang,biao.li,songwei,linnzh,xiandiy,peng}@whu.edu.cn

Abstract. Knowledge tracing is an essential task that estimates students' knowledge state as they engage in the online learning platform. Several models have been proposed to predict the state of students' learning process to improve their learning efficiencies, such as Bayesian Knowledge Tracing, Deep Knowledge Tracing, and Dynamic Key-Value Memory Networks. However, these models fail to fully consider the influence of students' current knowledge state on knowledge growth, and ignore the current knowledge state of students is affected by forgetting mechanisms. Moreover, these models are a unified model that does not consider the use of group learning behavior to guide individual learning. To tackle these problems, in this paper, we first propose a model named Knowledge Tracking based on Learning and Memory Process (LMKT) to solve the effect of students' current knowledge state on knowledge growth and forgetting mechanisms. Then we propose the definition of learning capacity community and personalized knowledge tracking. Finally, we present a novel method called Learning Ability Community for Personalized Knowledge Tracing (LACPKT), which models students' learning process according to group dynamics theory. Experimental results on public data sets show that the LMKT model and LACPKT model are effective. Besides, the LACPKT model can trace students' knowledge state in a personalized way.

Keywords: Personalized Knowledge Tracing · Learning ability community · Knowledge growth · Knowledge state

1 Introduction

With the development of Intelligent Tutoring Systems (ITS) [1] and Massive Open Online Courses (MOOCs) [15], a large number of students are willing to learn new courses or acquire knowledge that they are interested in through the online learning platform. When students have completed some exercises or courses, the online learning platform can obtain the knowledge state they mastered according to their learning records. For example, when a student tries to solve the exercise "$x^2 - 2x + 1 = 0$", the online learning system can estimate

the probability that the student will answer this exercise correctly according to the student whether had mastered arithmetic operations and quadratic equation with one unknown. In online learning platform, the task of knowledge tracing is to model the learning process of students to trace their future knowledge state based on their historical performance on exercises and the underlying knowledge concepts they had mastered. Knowledge tracing can be formalized as: given observations of interactions $X = \{x_1, ..., x_t\}$ taken by a student on a past exercise task, predict the probability that the student answers the next interaction x_{t+1} correctly. In the knowledge tracing, interactions take the form of a tuple of $x_t = (q_t, r_t)$ that combines a tag for the exercise being answered q_t with whether or not the exercise was answered correctly r_t [18]. Therefore, knowledge tracing technology can help teachers to teach according to their aptitude and give students personalized guidance and can help students to strengthen training on unfamiliar or less familiar knowledge concepts, which is very meaningful in the teaching process.

Knowledge tracing is inherently difficult as the complexity of the human brain learning process and the diversity of knowledge. Several models have been proposed to model the knowledge state of students in a concept specific manner, such as Bayesian Knowledge Tracing (BKT) [2,5], Deep Knowledge Tracing (DKT) [18], Dynamic Key-Value Memory Networks (DKVMN) [24]. BKT divides students' knowledge states into different concept states and assumes the concept state as a binary latent variable, *known* or *unknown*, and uses the Hidden Markov Model to update the posterior distribution of the binary concept state. Although BKT can express each knowledge concept state for each student, it requires students to define knowledge concepts in advance. Moreover, BKT assumes that once students have mastered the knowledge, they will never forget it, which limits its ability to capture the complicated relationship between different knowledge concepts and model long-term dependencies in an exercise sequence.

In recent years, inspired by the successful application of deep learning [12], the deep learning model is beginning to be used to solve the knowledge tracing problem. DKT is the first deep knowledge tracing model, which exploits Recurrent Neural Networks (RNNs) to the problem of predicting students to exercise based on students' previous learning sequences. Because of the RNNs has hidden layers with a large number of neurons, it can comprehensively express the relationships between different concepts to improve the accuracy of prediction. However, DKT summarizes a student's knowledge state of all concepts in one hidden state, which makes it difficult to trace how much a student has mastered a certain concept and pinpoint which concepts a student is good at or unfamiliar with [11]. To address this deficiency, Zhang et al. proposed the DKVMN model based on Memory Augmented Neural Networks (MANNs) [19,21], which can discover the correlation between input exercises and underlying concepts and reveal the evolving knowledge state of students through using a key-value memory. However, there are still some knowledge tracking problems that need to be addressed, as follows:

– The knowledge growth of DKVMN is calculated by multiplying students'
  exercises $x_t = (q_t, r_t)$ with a trained embedded matrix, which means that
  knowledge growth is only related to the absolute growth of this exercise.
  According to human cognitive processes, students' knowledge growth is
  related to their current knowledge states [3], so the calculation method of
  knowledge growth has limitations.
– The reading process of the DKVMN model ignores the effect of forgetting
  mechanism on students' knowledge state. According to the research that for-
  getting occurs along an exponential curve [7], forgetting means that a stu-
  dent's knowledge decreases over time, we believe that the knowledge state of
  students is affected by the forgetting mechanism.
– Existing knowledge tracing models assume that all students have the same
  learning ability without considering their inherent differences, and then con-
  struct a unified model to predict the probability that students answer the
  exercise correctly at the next moment, which lacking personalized ability.

To solve the above problems, inspired by the literature [17,20], we propose a
novel method called Learning Ability Community for Personalized Knowledge
Tracing (LACPKT) based on the learning ability community and learning and
memory process. The main contributions of our paper are summarized as follows:

– We present a knowledge tracing model called Knowledge Tracking based on
  Learning and Memory Process (LMKT), which solves the impact of the cur-
  rent knowledge state of students on knowledge growth and consider the effect
  of forgetting mechanisms on the current knowledge state of students.
– We first define the learning ability degree and learning ability community and
  propose the definition of personalized knowledge tracking according to group
  dynamics theory.
– We propose a novel method called Learning Ability Community for Person-
  alized Knowledge Tracing (LACPKT), which models the learning process of
  students in a personalized way.
– We conduct experiments on four public datasets to verify the effectiveness of
  the LMKT model and LACPKT model.

2   Related Work

There are many kinds of research to estimate the knowledge state of students.
Bayesian Knowledge Tracing (BKT) [5] is the most popular knowledge tracing
model based on machine learning, which is also a highly constrained and struc-
tured model. BKT models every knowledge concept for every student, and each
knowledge concept only changes from the unmastered state to the mastery state.
Some variants of BKT have also raised. For example, Yudelson et al. [23] proposed
an individualized Bayesian knowledge tracing models. Baker et al. [2] presented a
more accurate student state model by estimating the *P(G)* and *P(S)* contexts in
BKT. Pardos et al. [16] added the item difficulty of item response theory (IRT) [10]

to the BKT to increase the diversity of questions. Other information or technologies [6,8,9] have also been introduced into the Bayesian network framework.

Deep Knowledge Tracing (DKT) [18] applied the vanilla Recurrent Neural Networks (RNNs) to trace knowledge concept state, reports substantial improvements in prediction performance. DKT uses the RNNs with a hidden layer map an input sequence of vectors $X = \{x_1, x_2, ..., x_t\}$ to an output sequence of vectors $Y = \{y_1, y_2, ..., y_t\}$ to model student learning and predict student's state of all knowledge concepts. Zhang et al. [25] improved the DKT Model by incorporating more problem-level features and then proposed an adaptive DKT model structure that converts the dimensional input into a low dimensional feature vector that can effectively improve accuracy. Cheung et al. [4] proposed an automatic and intelligent approach to integrating the heterogeneous features into the DKT model that can capture students behaviors in the exercises. Nagatani et al. [14] extended the DKT model behavior to consider forgetting by incorporating multiple types of information related to forgetting, which improves the predictive performance. Memory Augmented Neural Networks (MANNs) have made progress in many areas, such as question answering [21] and one-shot learning [19]. MANNs consists of two operations, reading and writing that are achieved through additional attention mechanisms. Because of the recurrence introduced in the read and write operations, MANNs is a special variant structure of RNNs, it uses an external memory matrix that stores the knowledge state of a student. Zhang et al. put forward a Dynamic Key-Value Memory Networks (DKVMN) [24] that uses the concepts of Memory Augmented Neural Networks (MANNs) to reveal the evolving knowledge state of students and learn the relationships between concepts. DKVMN with one static *key* matrix that stores the concept representations and one dynamic *value* matrix that stores and updates the students understanding concept state of each concept, thus it has more capacity to handle knowledge tracking problems.

3   The Proposed Model

In this section, we first give the formalization of definitions and then introduce Knowledge Tracking based on Learning and Memory Process. Finally, we propose a novel method named Learning Ability Community for Personalized Knowledge Tracing to model the learning process of students. In description below, we assume a student's exercise sequence $X = \{x_1, x_2, ..., x_t\}$ contains N latent knowledge concepts $C = \{c_1, c_2, ..., c_N\}$, where $x_t = (q_t, r_t)$ is a tuple containing the question q_t and the correctness of the students answer r_t, and all exercise sequences of M students $U = \{u_1, u_2, ..., u_M\}$ are $\boldsymbol{X} = \{X^1, X^2, ..., X^M\}$. The details are elaborated in the following three subsections.

3.1   Definition Formulation

Learning ability represents the internal quality of an individual that can cause lasting changes in behavior or thinking, which can be formed and developed

through certain learning practices. In the knowledge tracing problem, because of the cognitive level of each student is different, the result of each student's exercise sequence reflect their learning ability. We introduce the definition of learning ability degree δ according to the exercise sequence of students.

Definition 1. *Learning Ability Degree.* *We assume that the exercise of student u_i is $x_t = (q_t, r_t)$ contains knowledge concept c_j, the learning ability degree of student u_i is $\delta_{u_i}^{c_j} = s_{max}^{c_j}/s_{length}^{c_j}$ that represents the learning ability of student u_i to learn the knowledge concept c_j in question q_t.*

Where a big $\delta_{u_i}^{c_j} \in [1, \mathbf{s}_{max}^{c_j}]$ indicates that the student u_i has a strong ability to learn this question q_t, $s_{length}^{c_j}$ represents the number of times that the student u_i repeatedly learns the knowledge concept c_j, $s_{max}^{c_j}$ represents the maximum number of times that a student repeatedly learns the knowledge concept c_j. Because of the exercise sequence of student u_i is $X_i = \{x_1^i, ..., x_t^i\}$ contains knowledge concepts $\{c_1, ..., c_j\}$, the learning ability degree sequence of student u_i is $\delta_{u_i} = \{\delta_{u_i}^{c_1}, \delta_{u_i}^{c_2}, ..., \delta_{u_i}^{c_j}\}$. Therefore, according to the learning ability degree sequence of all students, we definite the Learning Ability Community is as follows:

Definition 2. *Learning Ability Community.* *We Suppose that the learning ability sequence of student u_i and student u_j are $\delta_{u_i} = \{\delta_{u_i}^{c_1}, \delta_{u_i}^{c_2}, ..., \delta_{u_i}^{c_j}\}$ and $\delta_{u_j} = \{\delta_{u_j}^{c_1}, \delta_{u_j}^{c_2}, ..., \delta_{u_j}^{c_j}\}$, if $|\delta_{u_i} - \delta_{u_j}| \leq \varepsilon$, we believe that student u_i and student u_j have similar learning abilities. In other words, they belong to the same learning ability community.*

According to the exercise sequence of students and the definition of learning ability community, we use an unsupervised deep clustering algorithm to minimize ε to divide students into their range of learning ability through continuous iteration and acquire multiple different learning ability communities. In a learning ability community k, we input all exercise sequences into a basic knowledge tracing model for training and get a corresponding optimization model by adjusting the parameters of the basic model. Because all students have similar learning abilities in the learning ability community k, we can use group learning characteristics to guide individual learning. Therefore, we give the definition of the Personalized Knowledge Tracing is as follows:

Definition 3. *Personalized Knowledge Tracing.* *We Suppose that m students had already learned the exercise sequences $X = \{x_1, x_2, ..., x_T\}$ contain knowledge concepts $\{c_1, c_2, ..., c_j\}$ in the learning ability community k, if a student u_{m+i} wants to learn this exercise sequences, we are able to trace the personalized knowledge state of student u_{m+i} according to the knowledge state of m students. In other words, we can predict the probability that student u_{m+i} correctly answer this exercise sequence, which is called Personalized Knowledge Tracing.*

3.2   Knowledge Tracking Based on Learning and Memory Process

Despite being more powerful than DKT and BKT in storing and updating the exercise sequence of students, DKVMN still has deficiencies when solved the knowledge tracing problem. To solve the problem, we propose the model: Knowledge Tracking based on Learning and Memory Process (LMKT), its framework is shown in Fig. 1. We assume the *key* matrix $\mathbf{M}^k$ (of size $N \times d_k$) is a static matrix that stores the concept representations and the *value* matrix $\mathbf{M}_t^v$ (of size $N \times d_v$) is a dynamic value matrix that stores the student's mastery levels of each concept, meanwhile $\mathbf{M}_t^v$ updates over time. The task of knowledge tracing is completed by three mechanisms of LMKT: *attention*, *reading* and *writing*.

Attention. For the input exercise q_t of a student u_i, we utilize the attention mechanism to determine which concepts are relevant to it. Thus, we multiply q_t by embedding matrix $\mathbf{A}$ to get an embedding vector $\boldsymbol{k}_t$. Relevance probability of q_t belongs to every concept in $\mathbf{M}^k$ is computed by comparing the question to each *key* matrix slot $\mathbf{M}^k(i)$, which is defined as attention weight vector $\boldsymbol{w}_t$. $\boldsymbol{w}_t$ represents the weight of each student's attention between exercise and each concept and will be applied to read and write processes.

$$\boldsymbol{w}_t(i) = Softmax(\boldsymbol{k}_t^T \mathbf{M}^k(i)), \tag{1}$$
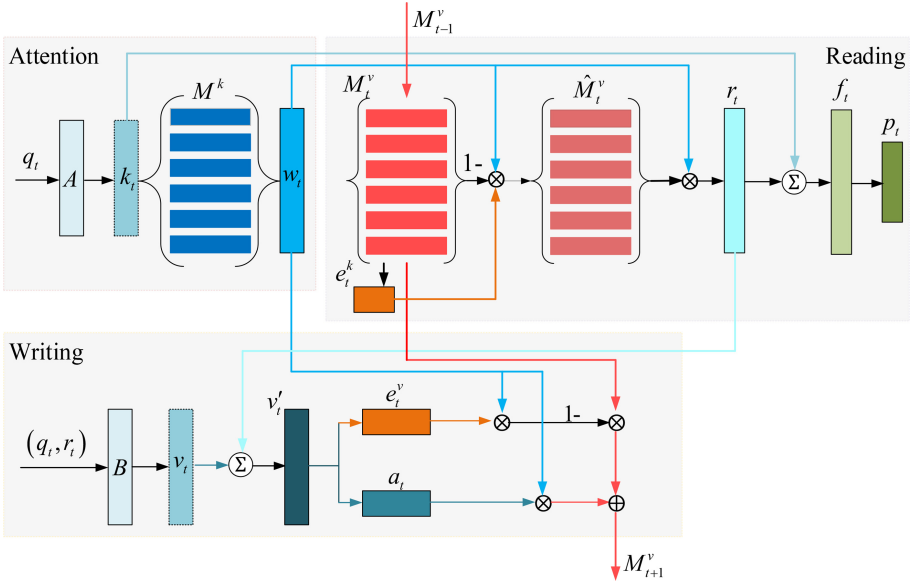
where $Softmax(x) = e^x / \sum_y (e^y)$.

Fig. 1. The framework of Knowledge Tracking based on learning and memory process

Reading. When an exercise q_t comes, the *value* matrix $\mathbf{M}_t^v$ of students' current knowledge state cannot remain unchanged, because of the influence of human forgetting mechanism. Therefore, we assume the *forgetting* vector $\mathbf{e}_t^k$ represents the forgetting of current knowledge state during the reading process, and students' current knowledge state $\mathbf{M}_t^v(i)$ will be updated as $\hat{\mathbf{M}}_t^v(i)$.

$$\mathbf{e}_t^k = Sigmoid(\mathbf{E}_k^T \mathbf{M}_t^v + \boldsymbol{b}_k), \tag{2}$$

$$\hat{\mathbf{M}}_t^v(i) = \mathbf{M}_t^v(i)[1 - \boldsymbol{w}_t(i)\mathbf{e}_t^k], \tag{3}$$

where $\mathbf{E}_k^T$ is a transformation matrix, the elements of $\mathbf{e}_t^k$ lie in the range $(0, 1)$. Then, according to the attention weight $\boldsymbol{w}_t$, we can calculate the read content $\mathbf{r}_t$ that stands for a summary of the student's mastery level of this exercise through the *value* matrix $\hat{\mathbf{M}}_t^v(i)$.

$$\boldsymbol{r}_t = \sum_i^N \boldsymbol{w}_t(i)\hat{\mathbf{M}}_t^v(i). \tag{4}$$

next, we concatenate the read content $\boldsymbol{r}_t$ and the input exercise embedding $\boldsymbol{k}_t$ and then pass through a multilayer perceptron with the Tanh activation function to get a summary of knowledge state vector $\boldsymbol{f}_t$, which contains the student's mastery level and the difficulty of exercise q_t.

$$\boldsymbol{f}_t = Tanh(\boldsymbol{W}_1^T[\boldsymbol{r}_t, \boldsymbol{k}_t] + \boldsymbol{b}_1). \tag{5}$$

finally, after $\boldsymbol{f}_t$ pass through the sigmoid activation function, we can get the predicted scalar $\boldsymbol{p}_t$ that represents the probability of answering the exercise q_t correctly.

$$\boldsymbol{p}_t = Sigmoid(\boldsymbol{W}_2^T \boldsymbol{f}_t + \boldsymbol{b}_2), \tag{6}$$

where $\boldsymbol{W}_1^T, \boldsymbol{W}_2^T$ stand for the weight and $\boldsymbol{b}_1, \boldsymbol{b}_2$ stand for the bias.

Writing. Writing process is the update process of students' knowledge state. In DKVMN model, The (q_t, r_t) embedded with an embedding matrix $\mathbf{B}$ to obtain the *knowledge growth* $\boldsymbol{v}_t$ of the students after working on this exercise [24], which is insufficient to express the actual gains in the learning process. However, concatenate the original *knowledge growth* $\boldsymbol{v}_t$ and the read content $\boldsymbol{r}_t$ and pass it through a fully connected layer with a Tanh activation to get the new *knowledge growth* $\boldsymbol{v}_t^{'}$.

$$\boldsymbol{v}_t^{'} = Tanh(\boldsymbol{W}_3^T[\boldsymbol{v}_t, \boldsymbol{r}_t] + \boldsymbol{b}_3). \tag{7}$$

before writing the student's *knowledge growth* into the *value* matrix $\mathbf{M}_t^v$, we should consider the forgetting according to human learning and cognitive processes. We assume that the *forgetting* vector $\mathbf{e}_t^v$ that is computed from $\boldsymbol{v}_t^{'}$.

$$\mathbf{e}_t^v = Sigmoid(\mathbf{E}_v^T \boldsymbol{v}_t^{'} + \boldsymbol{b}_v), \tag{8}$$

where $\mathbf{E}_v^T$ is a transformation matrix, the elements of $\mathbf{e}_t^v$ lie in the range (0,1). After the *forgetting* vector $\mathbf{e}_t^v$, the memory vectors component $\mathbf{M}_{t\text{-}1}^v(i)$ from the previous timestamp are modified as follows:

$$\widetilde{\mathbf{M}}_t^v(i) = \mathbf{M}_{t\text{-}1}^v(i)[\mathbf{1} - \boldsymbol{w}_t(i)\mathbf{e}_t^v], \tag{9}$$

where $\boldsymbol{w}_t(i)$ is the same as in the reading process. After forgetting, the *add* vector $\boldsymbol{a}_t$ is the actual gains of the new *knowledge growth* $\boldsymbol{v}_t'$, which is calculated as follows:

$$\boldsymbol{a}_t = Tanh(\boldsymbol{W}_a^T \boldsymbol{v}_t' + \boldsymbol{b}_a), \tag{10}$$

where $\boldsymbol{W}_a^T$ is a transformation matrix. Finally, the *value* matrix is updated at each time t based on $\widetilde{\mathbf{M}}_t^v(i)$ and $\boldsymbol{a}_t$.

$$\mathbf{M}_t^v(i) = \widetilde{\mathbf{M}}_t^v(i) + \boldsymbol{w}_t(i)\boldsymbol{a}_t. \tag{11}$$

Training. All parameters of the LMKT model, such as the embedding matrices $\mathbf{A}$ and $\mathbf{B}$ as well as other weight parameters, are trained by minimizing a standard cross entropy loss between the prediction label p_t and the ground-truth label r_t.

$$\mathcal{L} = -\sum_t \left(r_t \log p_t + (1 - r_t) \log (1 - p_t)\right) \tag{12}$$

3.3 Learning Ability Community for Personalized Knowledge Tracing

In this subsection, we introduce Learning Ability Community for Personalized Knowledge Tracing (LACPKT) based on the previous two subsections. The framework of LACPKT is shown in Fig. 2, the process of the LACPKT is as follows:

Firstly, we input the exercise sequences of all students into the LMKT model for training and get a basic $LMKT_0$ model suitable for all students.

Secondly, According to Definition 1, we process each student's exercise sequence to obtain their learning ability degree sequence $\{\delta_1, \delta_1, ..., \delta_L\}$.

Thirdly, we input the learning ability degree sequence of all students into the deep clustering network (DCN) [22], which joints dimensionality reduction and K-means clustering approach in which DR is accomplished via learning a deep neural network. According to the Definition 2, we assume that we obtain k learning ability communities (LAC) as follows:

$$\{LAC_1, LAC_2, ..., LAC_k\} \Leftarrow \{(\delta_1, \delta_1, ..., \delta_L), DCN\}. \tag{13}$$
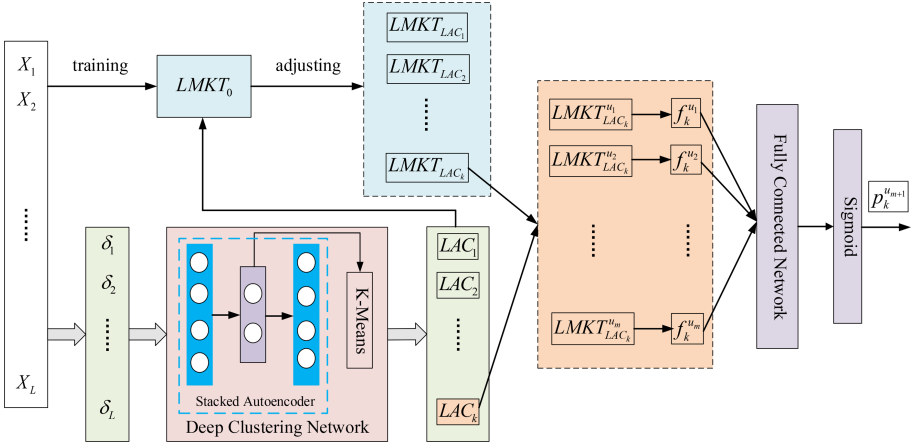
Fig. 2. The framework of learning ability community for Personalized Knowledge Tracing

Then, we input the exercise sequences of k learning ability communities into the basic $LMKT_0$ model for training, and acquire k optimized LMKT models by adjusting the parameters of the basic model, as shown in Eq. (14). In any learning ability community, we can trace the personalized knowledge state of each student.

$$\{LMKT_{LAC_1}, ..., LMKT_{LAC_k}\} \Leftarrow \{(LAC_1, ..., LAC_k), LMKT_0\}. \qquad (14)$$

Next, according to the Definition 3, if there are m students had already learned the exercise sequence $X = \{x_1, x_2, ..., x_t\}$ involves knowledge concepts $\{c_1, c_2, ..., c_j\}$ in the learning ability community k, we are able to construct m personalized knowledge models for these m students, and obtain these m students' current knowledge state $\{f_k^{u_1}, f_k^{u_2}, ..., f_k^{u_m}\}$, as shown in Eq. (15) and Eq. (16).

$$\{LMKT_{LAC_k^{u_1}}, ..., LMKT_{LAC_k^{u_m}}\} \Leftarrow \{(u_1, ..., u_m), LMKT_{LAC_k}\}, \qquad (15)$$

$$\{f_k^{u_1}, f_k^{u_2}, ..., f_k^{u_m}\} \Leftarrow \{LMKT_{LAC_k^{u_1}}, ..., LMKT_{LAC_k^{u_m}}\}. \qquad (16)$$

Finally, when student u_{m+1} wants to learn the exercise sequence $X = \{x_1, ..., x_t\}$, we concatenate these m students' current knowledge state $\{f_k^{u_1}, f_k^{u_2}, ..., f_k^{u_m}\}$ and then pass it through a fully connected network with the Sigmoid activation function to predict the probability $p_k^{u_{m+1}}$ that student u_{m+1} will correctly answer this exercise sequence.

$$p_k^{u_{m+1}} = Sigmoid(W_k^T[f_k^{u_1}, f_k^{u_2}, ..., f_k^{u_m}] + b_k), \qquad (17)$$

where $\boldsymbol{W}_k^T$ and $\boldsymbol{b}_k$ stand for the weight and bias of student u_{m+1}. The optimized objective function is the standard cross entropy loss between the prediction label $p_k^{u_{m+1}}$ and the ground-truth label r_t.

$$L_k^{u_{m+1}} = -\sum_t \left(r_t \log p_k^{u_{m+1}} + (1 - r_t) \log \left(1 - p_k^{u_{m+1}} \right) \right). \qquad (18)$$

4   Experiments

In this section, we first evaluate the performance of our LMKT model against the state-of-the-art knowledge tracing models on four experimental datasets. Then, we verify the validity of our LACPKT model and analyze the effectiveness of Personalized knowledge tracing.

4.1   Experiment Settings

Datasets. We use four public datasets from the literature [18,24]: Synthetic-5[1], ASSISTments2009[2], ASSISTments2015[3], and Statics2011[4], where Synthetic-5 is one synthetic dataset, the other three are the real-world datasets obtained from online learning platforms. The statistics of four datasets are shown in Table 1.

Implementation Details. First, we encoded the experimental datasets with one-hot encoding, and the length of the encoding depends on the number of different questions. In the synthetic-5 dataset, 50% of exercise sequences were used as a training dataset, but in the other three dataets, 70% of exercise sequences were used as a training dataset. A total 20% of the training dataset was split to form a validation dataset that was used to find the optimal model architecture and hyperparameters, hyperparameters were tuned using the five-fold cross-validation. Then, we constructed the deep clustering network consist of four hidden layers to cluster learning ability community, of which the number of neurons in the four hidden layers is 1000,800,500,100, in addition to every layer is pre-trained for 100 epochs and the entire deep network is further finetuned for 50 epochs. Next, in all experiments of our model, we trained with Adam optimizer and repeated the training five times to obtain the average test results. Finally, about evaluation indicators, we choose the $AUC \in [0,1]$, which is the area under the Receiver Operating Characteristic (ROC) curve, to measure the performance of all models.

---

[1] https://github.com/chrispiech/DeepKnowledgeTracing/tree/master/data/synthetic.

[2] https://sites.google.com/site/assistmentsdata/home/assistment-2009-2010-data/skill-builder-data-2009-2010.

[3] https://sites.google.com/site/assistmentsdata/home/2015-assistments-skill-builder-data.

[4] Statics2011:https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=507.

Table 1. The statistics of four datasets.

Datasets	Students	Questions	Exercises
Synthetic-5	4000	50	200,000
ASSISTments2009	4,151	110	325,637
ASSISTments2015	19,840	100	683,801
Statics2011	333	1223	189,297

4.2   Results and Analysis

LMKT Model Performance Evaluation. We evaluate our LMKT model against the other three knowledge tracing models: BKT, DKT and DKVMN. The test AUC results are shown in Table 2.

Table 2. The AUC results of the four models on four datasets.

Datasets	Test AUC (%)			
	BKT	DKT	DKVMN	LMKT
Synthetic-5	62	80.34	82.76	**82.95**
ASSISTments2009	63	80.53	81.52	**81.77**
ASSISTments2015	64	72.52	72.73	**72.94**
Statics2011	73	80.20	82.33	**82.76**

Since the performance of the BKT model and the DKT model is lower than the DKVMN model, the AUC results of these two models refer to the optimal results in [24]. For the DKVMN model, we set the parameters of it to be consistent with those in the original literature, then trained five times on four datasets to obtain the final average test results. In our LMKT model, we adjusted the model parameters to find the best model structure and obtain the best results according to the change of our model structure. It can be seen that our model outperformed the other models over all the four datasets. Although the performance of our model is improved by 0.25% to 0.5% compared to the DKVMN model, it proved the effectiveness of dealing with knowledge tracking problems based on human learning and memory processes. Besides, the state dimensions and memory size of the LMKT model are 10 or 20, it can get better results. However, the state dimensions and memory size of the DKVMN model are 50 or 100, which leads to problems such as more model parameters and longer training time. As shown in Fig. 3, the performance of the LMKT model is better than the DKVMN model, especially the LMKT model can reach the optimal structure of the model in fewer iterations. However, In the Statics2011 dataset, although the performance of the LMKT model and the DKVMN model are similar, the gap exists between the training AUC and the validation AUC of the DKVMN model is larger than the LMKT model, so the over-fitting problem of the LMKT model is smaller than the DKVMN model.
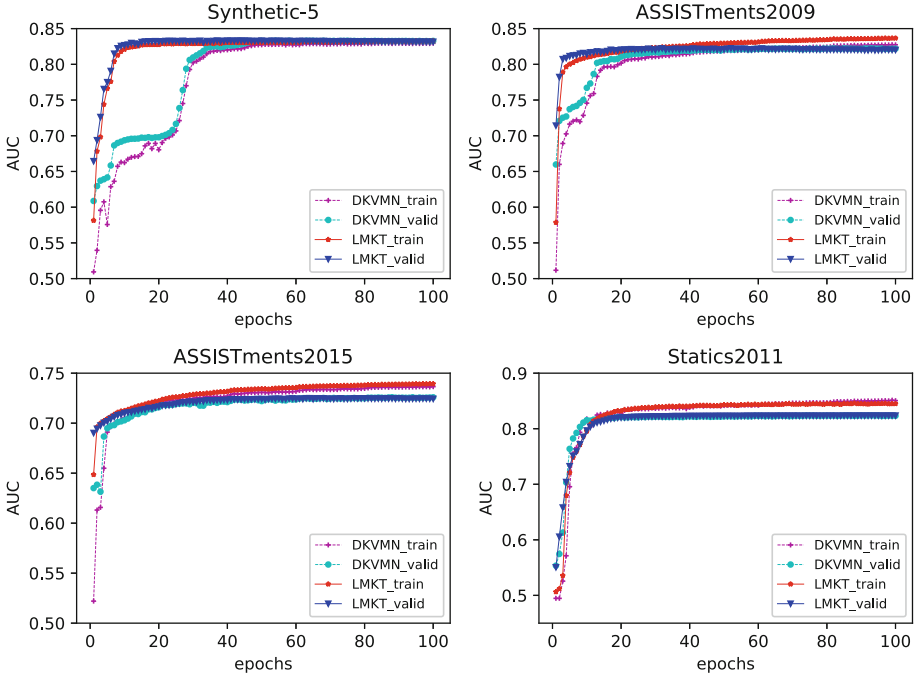
Fig. 3. Training AUC and validation AUC of DKVMN and LMKT on four datasets

Learning Ability Community Analysis. By analyzing the exercise sequences of the four experimental datasets, we divide each student into different learning ability communities thought the deep clustering network. According to Definitions 1 and 2, we knew that δ determines the difference in learning ability among students and then normalized the learning ability degree is between 0 and 1. However, we set up clusters of deep clustering networks to determine the number of learning ability communities and divide students into the learning ability communities to which they belong. We set the ASSISTments2015 dataset contains six different learning ability communities and the other three datasets contain five different learning ability communities, and visualize each dataset through t-SNE [13]. Figure 4 shows the results of the visualization of the learning ability community, where each color represents a learning ability community. In the Synthetic-5 dataset, because it is obtained by artificially simulating the student's answering process, the effect of dividing the learning ability community shows discrete characteristics. In ASSISTments2009 and ASSISTments2015 dataset, because of the exercise sequences are shorter in the datasets or the exercise questions are repeated multiple times in the same exercise sequence, some data is far from the center of the learning ability community. Moreover, the dataset ASSISTments2015 contains 19,840 students, so the number of learning ability communities set is not enough to meet the actual situation, so it has the

problem of overlapping learning ability communities. In the Statics2011 dataset, since it contains only 333 students, the overall clustering effect is sparse.
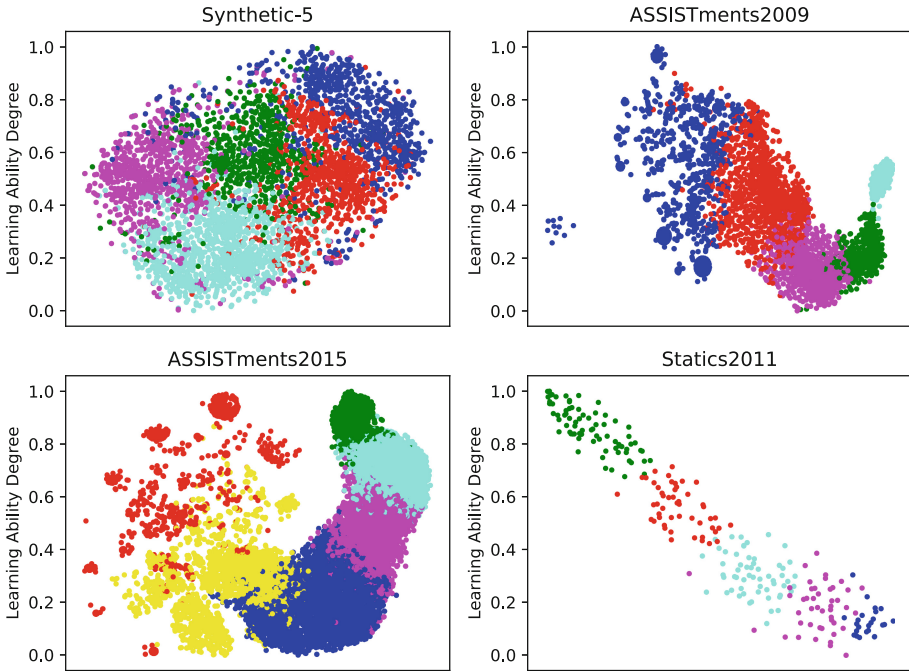

Fig. 4. Learning ability community

Personalized Knowledge Tracing. Because the dataset ASSISTments2015 is an updated version of the data set ASSISTments2009, and the data set Statics2011 has only 333 student exercise sequences, we chose the datasets Synthetic-5 and ASSISTments2015 for experimental verification of personalized knowledge tracing. Through the analysis of the dataset and the learning ability community, we set the number of learning ability communities of the datasets Synthetic-5 and ASSISTments2015 to 4 and 6. To ensure the sufficiency and reliability of personalized knowledge tracing experiment, we input different learning ability communities to DKVMN called the LAC_DKVMN model and compare it with LACPKT. According to Definition 3, we conducted experiments to validate the validity of the LACPKT model, the experimental results are shown in Fig. 5.

According to Fig. 5(a) and (b), we found that the personalized knowledge tracking capability of the LACPKT model is better than the LAC_DKVMN model in all learning capability communities of the two datasets. In dataset Synthetic-5, the AUC results of the LACPKT model in the four learning ability communities are 83.96%, 84.98%, 82.56%, 81.67%, respectively. In dataset

ASSISTments2015, the AUC results of the LACPKT model in the six learning ability communities are 76.25%, 70.34%, 73.23%, 71.87%, 70.68%, 72.72%, respectively. From the experimental results, when the LACPKT model tracks the knowledge status of students in different learning ability communities, its performance is shown to be different. The reason is that students learn different exercise sequence length and problems in different learning ability communities, so the performance of the LACPKT model is different in different learning ability communities. Figure 5(c) and (d) show the changing state of AUC of the LACPKT model in the validation dataset. In dataset Synthetic-5, because it is an artificial simulation dataset, the length and questions of each student's exercise sequence are the same, the test results and the verification results of the LACPKT model remain the same. However, in ASSISTments2015 dataset, because each student's exercise sequence length and questions are different, the performance of the LACPKT model in different learning communities is different. For different learning ability communities, the LACPKT model can model students' learning processes according to different learning abilities, and track students' personalized knowledge status based on the effect of group learning behavior on individuals. Therefore, the experimental results proved the effectiveness of dividing the learning ability community and the effectiveness of the LACPKT model in tracing students' personalized knowledge state.
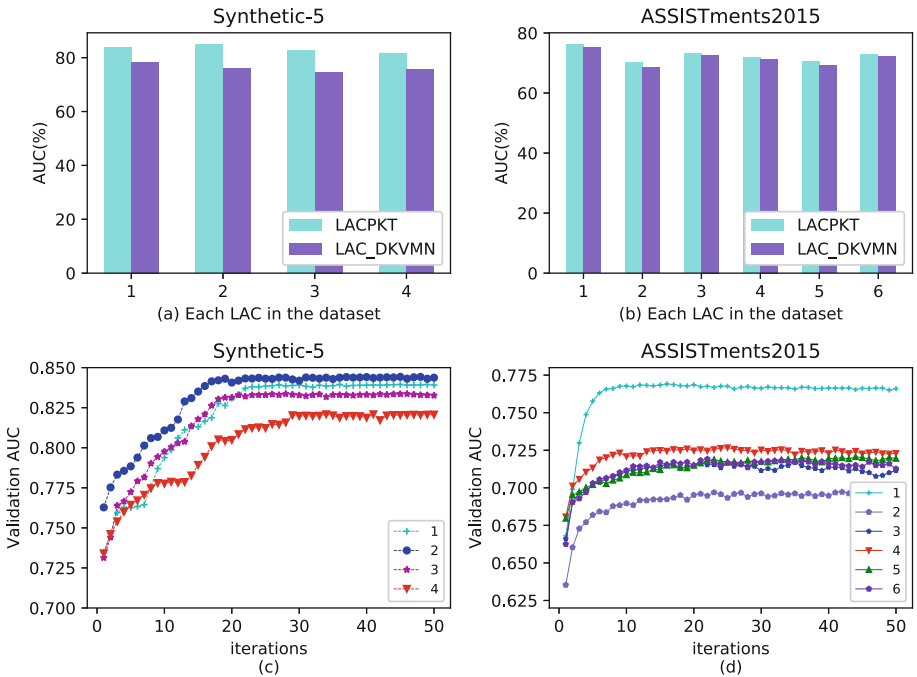

Fig. 5. The results of personalized knowledge tracing.

5   Conclusions and Future Work

In this paper, we propose a novel method that is called Learning Ability Community for Personalized Knowledge Tracing (LACPKT) to model the learning process of students and trace the knowledge state of students. The LACPKT model consists of two aspects, on the one hand, we propose the definition of learning ability community, which utilizes the effect of group learning on individuals to trace the personalized knowledge state of students; On the other hand, we propose a Knowledge Tracking based on Learning and Memory Process model that considers the relationship between the current knowledge state and knowledge growth and forgetting mechanisms. Finally, we demonstrate the effectiveness of the LACPKT model in tracing students' personalized knowledge on public datasets. For future work, we will integrate the content information of the problem to optimize the learning ability community and optimize the network structure to improve the prediction ability.

References

1. Antunes, C.: Acquiring background knowledge for intelligent tutoring systems. In: Educational Data Mining 2008 (2008)
2. Baker, R.S.J., Corbett, A.T., Aleven, V.: More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 406–415. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-69132-7_44
3. Brod, G., Werkle-Bergner, M., Shing, Y.L.: The influence of prior knowledge on memory: a developmental cognitive neuroscience perspective. Front. Behav. Neurosci. **7**, 139 (2013)
4. Cheung, L.P., Yang, H.: Heterogeneous features integration in deep knowledge tracing. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, E.S. (eds.) ICONIP 2017. LNCS, pp. 653–662. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-70096-0_67
5. Corbett, A.T., Anderson, J.R.: Knowledge tracing: modeling the acquisition of procedural knowledge. User Model. User-Adapt. Interact. **4**(4), 253–278 (1994)
6. David, Y.B., Segal, A., Gal, Y.K.: Sequencing educational content in classrooms using Bayesian knowledge tracing. In: Proceedings of the Sixth International Conference on Learning Analytics & Knowledge, pp. 354–363. ACM (2016)

7. Ebbinghaus, H.: Memory: a contribution to experimental psychology. Ann. Neurosci. **20**(4), 155 (2013)
8. Feng, J., Zhang, B., Li, Y., Xu, Q.: bayesian diagnosis tracing: application of procedural misconceptions in knowledge tracing. In: Isotani, S., Millán, E., Ogan, A., Hastings, P., McLaren, B., Luckin, R. (eds.) AIED 2019. LNCS (LNAI), vol. 11626, pp. 84–88. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-23207-8_16
9. Hawkins, W.J., Heffernan, N.T., Baker, R.S.J.D.: Learning Bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (eds.) ITS 2014. LNCS, vol. 8474, pp. 150–155. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-07221-0_18
10. Johns, J., Mahadevan, S., Woolf, B.: Estimating student proficiency using an item response theory model. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 473–480. Springer, Heidelberg (2006). https://doi.org/10.1007/11774303_47
11. Khajah, M., Lindsey, R.V., Mozer, M.: How deep is knowledge tracing? In: Proceedings of the 9th International Conference on Educational Data Mining, EDM 2016, Raleigh, North Carolina, USA, 29 June–2 July 2016 (2016)
12. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
13. van der Maaten, L., Hinton, G.: Visualizing data using t-SNE. J. Mach. Learn. Res. **9**(Nov), 2579–2605 (2008)
14. Nagatani, K., Zhang, Q., Sato, M., Chen, Y.Y., Chen, F., Ohkuma, T.: Augmenting knowledge tracing by considering forgetting behavior. In: The World Wide Web Conference, pp. 3101–3107. ACM (2019)
15. Pappano, L.: The year of the MOOC. N. Y. Times **2**(12), 2012 (2012)
16. Pardos, Z.A., Heffernan, N.T.: KT-IDEM: introducing item difficulty to the knowledge tracing model. In: Konstan, J.A., Conejo, R., Marzo, J.L., Oliver, N. (eds.) UMAP 2011. LNCS, vol. 6787, pp. 243–254. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22362-4_21
17. Pardos, Z.A., Trivedi, S., Heffernan, N.T., Sárközy, G.N.: Clustered knowledge tracing. In: Cerri, S.A., Clancey, W.J., Papadourakis, G., Panourgia, K. (eds.) ITS 2012. LNCS, vol. 7315, pp. 405–410. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-30950-2_52
18. Piech, C., et al.: Deep knowledge tracing. In: Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, 7–12 December 2015, Montreal, Quebec, Canada, pp. 505–513 (2015)
19. Santoro, A., Bartunov, S., Botvinick, M., Wierstra, D., Lillicrap, T.: Meta-learning with memory-augmented neural networks. In: International Conference on Machine Learning, pp. 1842–1850 (2016)
20. Trivedi, S., Pardos, Z.A., Heffernan, N.T.: Clustering students to generate an ensemble to improve standard test score predictions. In: Biswas, G., Bull, S., Kay, J., Mitrovic, A. (eds.) AIED 2011. LNCS (LNAI), vol. 6738, pp. 377–384. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-21869-9_49
21. Weston, J., Chopra, S., Bordes, A.: Memory networks. In: 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, 7–9 May 2015, Conference Track Proceedings (2015)
22. Yang, B., Fu, X., Sidiropoulos, N.D., Hong, M.: Towards k-means-friendly spaces: simultaneous deep learning and clustering. In: Proceedings of the 34th International Conference on Machine Learning, vol. 70, pp. 3861–3870 (2017)

23. Yudelson, M.V., Koedinger, K.R., Gordon, G.J.: Individualized Bayesian knowledge tracing models. In: Lane, H.C., Yacef, K., Mostow, J., Pavlik, P. (eds.) AIED 2013. LNCS (LNAI), vol. 7926, pp. 171–180. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39112-5_18
24. Zhang, J., Shi, X., King, I., Yeung, D.Y.: Dynamic key-value memory networks for knowledge tracing. In: Proceedings of the 26th International Conference on World Wide Web, pp. 765–774. International World Wide Web Conferences Steering Committee (2017)
25. Zhang, L., Xiong, X., Zhao, S., Botelho, A., Heffernan, N.T.: Incorporating rich features into deep knowledge tracing. In: Proceedings of the Fourth (2017) ACM Conference on Learning@ Scale, pp. 169–172. ACM (2017)