ABSTRACT In online intelligent education systems, to offer proactive studying services to students (e.g., learning path recommendation), a crucial demand is to track students’ knowledge mastery levels over time. However, existing works ignore the impact of learning transfer on knowledge tracing and only track knowledge proficiency. Knowledge proficiency alone cannot fully reflect students’ knowledge mastery levels. A student’s knowledge structure (the similarities and differences within knowledge concepts) and abstract principle mastery level (common attributes among knowledge concepts, such as learning methods) also need to be tracked. To this end, we propose a novel multilevel Knowledge Tracing model with Learning Transfer (KTLT) to track students’ knowledge mastery levels. First, we clarify the relationships among abstract principles, knowledge structure, and knowledge proficiency by utilizing the learning transfer theory in educational psychology. Then, we associate each problem with a knowledge vector in which each element represents an explicit knowledge concept by leveraging educational priors (i.e., a Q-matrix). Correspondingly, each student is represented as a knowledge vector (knowledge proficiency) and an abstract principle vector (abstract principle mastery). Given a student’s knowledge and abstract principle vector over time, we use the learning and forgetting curve as priors to capture the student’s knowledge proficiency and abstract principle mastery level over time. Furthermore, we embed the knowledge concept by using a student’s abstract principle mastery level and obtain a personalized knowledge relevance matrix (a student’s knowledge structure) by calculating the cosine similarity among the knowledge embedding results. Finally, we design a probabilistic matrix factorization framework by combining student and problem priors for tracking a student’s knowledge mastery. Extensive experiments on two real-world datasets demonstrate both the effectiveness and explanatory ability of the KTLT.

INDEX TERMS Knowledge tracing, learning transfer, data mining, education priors, dynamic modeling, probability graph model.
the students that have learned C++ have better performance when learning python. What affects learning transfer? Cognitive structure migration theory [14] proposes that the occurrence of learning transfer is mainly affected by three aspects: the availability, stability and discernability of the existing knowledge structure. The availability of an existing knowledge structure refers to whether students can extract the abstract principles (common attributes between knowledge concepts, such as learning methods, learning styles, and learning habits) contained in the knowledge concepts. The stability of the existing knowledge structure indicates whether students have high knowledge proficiency. The discernability of an existing knowledge structure indicates whether students have a good knowledge structure and are aware of correlations within knowledge concepts (the similarities and differences within the knowledge concepts). When applying learning transfer theory to knowledge tracing, there are several challenges: How can cognitive structure migration theory be leveraged under the background of knowledge tracing? How can a student’s abstract principle mastery level (the common attributes within knowledge concepts), knowledge proficiency and knowledge structure (the correlations within knowledge concepts understood by the student) be modeled? How can a framework be established that can track a student’s abstract principle mastery level, knowledge proficiency and knowledge structure at the same time?

To solve the above challenges, we first associate each student with a knowledge proficiency vector and an abstract concept mastery level vector. We use the learning and forgetting curve as priors to capture the change in a student’s knowledge proficiency and abstract concept mastery level over time. Specifically, we establish a D-matrix to represent the correlation between knowledge concepts and abstract principles. To track a student’s knowledge structure, we divide the student’s knowledge structure into two parts: one is a deep knowledge structure obtained through understanding and reasoning about abstract concepts, and the other is a shallow knowledge structure that represents the shallow correlation within knowledge concepts. We embed a knowledge concept by using the D-matrix and a student’s abstract principles mastery level and calculated the student’s deep knowledge structure by calculating the cosine similarity among the knowledge embedding results. Furthermore, the S-matrix, in which each element represents the correlation level within the knowledge concepts, is used to represent a shallow knowledge structure. To improve the explanatory ability of the KTLT, we exploit the Q-matrix, which is marked by educational experts, to depict the correlation between knowledge concepts and problems as priors to generate problem representations. After that, we design a probabilistic matrix factorization framework by combining item response theory, slipping and guessing factors, students priors and problems priors for tracking a student’s knowledge proficiency, abstract principle mastery level and knowledge structure over time.

The main contributions of this work are outlined as follows:

- To the best of our knowledge, the KTLT is the first knowledge tracing method that takes into account learning transfer and tracks abstract principle mastery level, knowledge proficiency and knowledge structure at the same time.
- We propose a multilevel probability graph model by applying cognitive structure migration theory in educational psychology to track a student’s knowledge proficiency, abstract principle mastery level, and knowledge structure over time.
- Experiments conducted on real-world datasets show that the KTLT is effective and clearly outperforms the state-of-the-art methods.

II. RELATED WORK

In this section, we demonstrate some recent works about knowledge tracing and then briefly introduce learning transfer theory. The related work can be classified into the following two categories: knowledge tracing and learning transfer.

A. KNOWLEDGE TRACING

Over the past years, many algorithms have been proposed for knowledge tracing. Most of them fall into three broad categories: traditional knowledge tracing, knowledge tracing based on data mining, and knowledge tracing based on deep learning.

The representative algorithms for traditional knowledge tracing are Bayesian Knowledge Tracing (BKT) [6] and Temporal Item Response Theory (TIRT) [7]. BKT assumes that a student’s knowledge proficiency is represented as a set of binary variables, in which every knowledge concept is either mastered by a student or not, and models a student’s knowledge proficiency as a latent variable in a hidden Markov model. TIRT incorporates forgetting factors into Item Response Theory (IRT) [15] to track a student’s knowledge proficiency. Knowledge tracing based on data mining includes FuzzyCDF [3] and KPT [8]. FuzzyCDF combines fuzzy set theory and educational hypotheses to model a student’s knowledge proficiency and predicts a student’s performance by considering both the skipping and guessing factors. KPT is an explanatory probabilistic method that tracks the
knowledge proficiency of students over time by leveraging educational priors (i.e., a Q-matrix). Knowledge tracing based on deep learning research works mainly includes DKT [10], DKVMN [16], and EERNN [11]. DKT is the first model that applies deep learning algorithms to knowledge tracing. DKT uses flexible recurrent neural networks that are ‘deep’ in time to track students’ knowledge proficiency. Since then, EERNN based on DKT takes full advantage of both students’ learning records and the text of each problem. DKVMN applies key-value memory networks to exploit the relationship among the underlying knowledge and directly outputs a student’s knowledge proficiency. However, all the above methods fail to take into account the impact of learning transfer on knowledge tracing and only track a student’s knowledge proficiency and ignore the abstract principles (the common elements and learning methods behind knowledge concepts) and knowledge structure (the correlations within knowledge concepts understood by students).

### B. LEARNING TRANSFER

In the field of educational psychology, learning transfer (also called transfer of learning) refers to the impact of one learning on another. The original theory of learning transfer is the identical elements theory [17], which claims that one kind of learning can affect another kind of learning because they have the same elements. However, the identical elements theory was quickly rebuked by the general principle [18]. The general principle holds that the main reason for learning transfer does not lie in the common elements among tasks but in summarizing the common principles within knowledge concepts. Cognitive structure migration theory [14] believes that the occurrence of learning transfer is mainly affected by three aspects: the availability of an existing knowledge structure, the stability of an existing knowledge structure and the discernability of an existing knowledge structure. Metacognitive learning transfer theory states that a student’s metacognitive level is an important factor affecting learning transfer. Metacognition [19] refers to a student’s understanding of the cognitive process. Compared to metacognitive learning transfer theory, cognitive structure migration theory is more suitable for optimizing knowledge tracing. Schema learning transfer theory [20] holds that both schema acquisition and rule automation are indispensable in learning transfer.

### III. KNOWLEDGE TRACING MODEL WITH LEARNING TRANSFER

In this section, we first formally introduce the knowledge tracing task and our study overview. Then we provide the technical details of the KTLT. Finally, we specify the parameter learning in the KTLT.

#### A. PROBLEM DEFINITION

Suppose there are $N$ students, $M$ problems, and $K$ knowledge concepts in a learning system. In this system, the students’ exercise logs, as shown in Table 1, indicate that students exercise at different times. The students’ exercise responses are represented as a tensor $R \in \mathbb{R}^{N \times M \times T}$, where $R_{ij}^t$ denotes student $i$’s response for problem $j$ in time window $t$. The students’ abstract principle mastery levels, knowledge structures, and knowledge proficiencies are represented as three tensors, $L \in \mathbb{R}^{N \times T \times K}$, $S \in \mathbb{R}^{N \times K \times K \times T}$, and $U \in \mathbb{R}^{N \times T \times K}$, where $L_{ij}^t$ denotes student $i$’s mastery of abstract principle $j$ in time windows $t$, $S_{k_1,k_2,i}^{t_1,t_2}$ denotes the similarity between knowledge concept $k_1$ and $k_2$ understood by student $i$ at time windows $t_1$ and $t_2$, and $U_{ik}^t$ denotes student $i$’s proficiency in knowledge concept $k$ in time window $t$. Additionally, we have a Q-matrix, which is represented as binary matrix $Q \in \mathbb{R}^{M \times K}$. $Q_{ik} = 1$ means that problem $j$ relates to knowledge concept $k$ and vice versa. Without loss of generality, the problem can be formulated as:

**Problem Formulation:** Given the students exercise response tensor $R$ and the Q-matrix provided by educational experts, our goal is three fold: (1) modeling the change in the students’ abstract principle mastery levels $L$, knowledge proficiencies $U$ and knowledge structures $S$ from time 1 to $T$; (2) predicting the students’ abstract principle mastery levels $L^{T+1}$, knowledge proficiencies $U^{T+1}$ and knowledge structures $S^{T+1}$ in time window $T + 1$; (3) predicting the students’ responses $R^{T+1}$ in time window $T + 1$.

#### B. MODELING THE STUDENT’S EXERCISE RESPONSES

To better model the students’ exercise responses, we put forward four assumptions based on learning transfer theory [18], item response theory [15], and the slipping and guessing factors [21], [22] as follows.

**Assumption 1:** Students understand and master knowledge concepts $A$ and $B$. Then, the students have a high probability of mastering knowledge concept $C$, which is highly correlated with knowledge concepts $A$ and $B$.

**Assumption 2:** The higher the students’ mastery level of knowledge concepts contained in the problem is, the better the students’ mastery level for the problem.

**Assumption 3:** The higher the students’ mastery level for the problem is and the lower the difficulty of the problem is, the stronger the students’ ability to do the problem.

**Assumption 4:** Students’ responses in the real world are affected by their slipping and guessing.

We omit the proofs for the above assumptions since they are straightforward. Based on the above four assumptions, we model the students’ exercise responses, which are shown in Figure 2. The KTLT is a generation process that starts with the students’ abstract principles $L$, followed by determining the students’ deep knowledge structures $S$ by using the D-matrix, which represents the correlation between the knowledge concepts and abstract principles. Furthermore, the KTLT computes the students’ knowledge structures $S^t$.
by considering the S-matrix, which represents the correlation among the knowledge concepts, and then calculates the students’ knowledge mastery level $\hat{U}$ by combining the students’ knowledge proficiencies $U$. Finally, the KTTLT obtains the students’ problem mastery levels $\theta$ and generates the students’ problem responses $R$ by considering the problem difficulty $b$, slipping $s$ and guessing $g$ factors.

### C. PROBABILISTIC MODELING WITH PRIORS

In this section, we introduce the technical details of the KTTLT as shown in Figure 2 (from bottom to top). For a better illustration, some notations are summarized in Table 2.

#### 1) MODELING STUDENTS’ EXERCISE RESPONSES WITH SLIPPING AND GUESSING

Inspired by many existing works [8], [23], for each student and each problem, we model a student’s exercise response tensor $R$, as follows:

$$p(R|\mu, s, g) = \prod_{t=1}^{T} \prod_{i=1}^{N} \prod_{j=1}^{M} \mathcal{N}(R_{ij}^{t}((1-s_{j})\mu_{ij}^{t} + g_{j}(1-\mu_{ij}^{t}), \sigma_{R}^{2}))$$

(1)

where $\mathcal{N}(\mu, \sigma^{2})$ is a Gaussian distribution with mean $\mu$ and variance $\sigma^{2}$, and $s_{j}$ and $g_{j}$ denote the slipping and guessing factors of problem $j$. $I$ is an indicator tensor and $I_{ij}^{t}$ equals 1 if student $i$ does the exercise of problem $j$ in time window $t$, and vice versa.

#### 2) MODELING STUDENTS’ PROBLEM SOLVING ABILITIES WITH ITEM RESPONSE THEORY

Formally, following an IRT-like high-order logistic model [3], a student’s problem solving ability $\mu_{ij}$ is defined as:

$$\mu_{ij}^{t} = \frac{1}{1 + \exp[-1.7a_{j}(\theta_{ij}^{t} - b_{j})]}$$

(2)

This definition implies that the ability of student $i$ to solve problem $j$ ($\mu_{ij}$) depends on the difference between the student’s mastery level for problem $j$ ($\theta_{ij}$) and the properties of the problem $j$, including difficulty ($b_{j}$) and discrimination ($a_{j}$).  

#### 3) MODELING STUDENTS’ PROBLEM MASTERY LEVELS WITH LEARNING TRANSFER

According to cognitive structure migration theory [14], we model a student’s knowledge mastery level $\hat{U}$ and a student’s problem mastery level $\theta$ as follows:

$$\theta_{ij}^{t} = \hat{U}_{ij}^{t} * V^{T}$$

$$\hat{U}_{ij}^{t} = U_{ij}^{t} + \sum_{k \neq j} \sigma(U_{ik}^{t}) * S_{kji}^{t}$$

(3)

where $U_{ij}^{t}$ is the knowledge proficiency of student $i$ on problem $j$ in time window $t$, $\sigma(\ast)$ denotes the sigmoid function, $S_{kji}^{t}$ is student $i$’s knowledge structure in which each element...
represents the correlation level between knowledge concept \( k \) and knowledge concept \( j \) understood by student \( i \) in time windows \( t \), and \( V_{ij} \) represents the correlation level between problem \( i \) and knowledge concept \( j \).

4) MODELING STUDENTS’ KNOWLEDGE STRUCTURES

For tracking the students’ knowledge structures, we model a student’s knowledge structure \( S_{ijk}^t \) and a student’s deep knowledge structure \( S_{ijk}^t \) as follows.

\[
S_{ijk}^t = S_{ijk}^t \ast \beta_k + S_{ij}^t \ast (1 - \beta_k)
\]

where \( S_{ijk}^t \) is student \( k \)’s deep knowledge structure in which each element represents the deep correlation level between knowledge concept \( i \) and \( j \) and understood by student \( k \) in time windows \( t \), \( \beta_k \) balances the two factors to capture the learning characteristics of the student’s knowledge structure, \( \sigma(*) \) denotes the sigmoid function, \( S_{ij}^t \) represents the shallow correlation level between knowledge concept \( i \) and \( j \), \( L_{ik}^a \) is the abstract principle mastery level of student \( k \) on abstract principle \( a \) in time window \( t \), and \( D_{ia}^t \) represents the correlation level between knowledge concept \( i \) and abstract principle \( a \).

5) MODELING U AND L WITH TWO DYNAMIC LEARNING THEORIES

The essence of knowledge tracing is to track the dynamic changes in students’ knowledge proficiencies and understanding. Therefore, inspired by many existing works [8], we combine the learning curve [24] and Ebbinghaus forgetting curve [25] as priors to model \( U \) and \( L \) as follows:

\[
p(U|\sigma_U^2, \sigma_{U|L}^2) = \prod_{i=1}^{N} \mathcal{N}(U_i^1|0, \sigma_U^2 I) \prod_{i=2}^{T} \mathcal{N}(U_i^t|\bar{U}_i^t, \sigma_{U|L}^2 I)
\]

\[
\bar{U}_i^t = \alpha_i f_i^t(*) + (1 - \alpha_i) f_s^t(*), \quad s.t. \ 0 \leq \alpha_i \leq 1
\]

\[
p(L|\sigma_L^2, \sigma_{L|U}^2) = \prod_{i=1}^{N} \mathcal{N}(L_i^1|0, \sigma_L^2 I) \prod_{i=2}^{T} \mathcal{N}(L_i^t|L_i^{t-1}, \sigma_{L|U}^2 I)
\]

where \( U_i^t \) and \( L_i^t \) are the knowledge proficiency and abstract principle mastery level of student \( i \) in time window \( t \), which follow Gaussian distributions with means \( U_i^t \) and \( L_i^{t-1} \) and variances \( \sigma_U^2 I \) and \( \sigma_L^2 I \). \( U_i^1 \) and \( L_i^1 \) follow a zero-mean Gaussian distribution. \( f_i^t(*) \) is the learning factor of a student’s knowledge proficiency, \( f_s^t(*) \) is the forgetting factor and \( \alpha_i \) balances the two factors to capture the learning characteristics of a student’s knowledge structure.

\[
l_i^t(*) = U_i^t \ast \frac{G \ast f_k^t}{\Delta t + r}
\]

\[
f_s^t(*) = U_i^t \ast e^{-\frac{\Delta t}{P}}
\]

where \( f_k^t \) denotes the frequency of knowledge concept \( k \) in time window \( t \), and \( \Delta t \) is the time interval between neighboring time windows. \( G, P \) and \( r \) are hyperparameters.

6) MODELING V WITH THE Q-MATRIX PRIOR

Many existing works [10], [11] suffer from the interpretation problem as the learned latent dimensions are unexplainable. To address this challenge, we model \( V \) with the Q-matrix prior [8], [26], [27] as follows:

\[
\ln p(V|Q) = \ln \prod_{(j,q,p)} p(q_{j+}^p | V) p(V)
\]

\[
= \prod_{j=1}^{M} \prod_{q=1}^{K} \prod_{p=1}^{N} I(q_{j+}^p) \ln \frac{1}{1 + e^{-1(V_{jq} - V_{jp})}} - \frac{1}{2\sigma_q^2} \|V\|_F^2
\]

where \( V \) follows a zero-mean Gaussian prior, \( I(q_{j+}^p) \) is an indicator matrix and \( I(q_{j+}^p) \) equals 1 if \( q_{j+}^p \) and vice versa. For problem \( j \), partial order \( j_{+} \) can be defined as:

\[
q_{j+}^p, \quad \text{if} \ Q_{jq} = 1 \quad \text{and} \ Q_{jp} = 0
\]

The above formula is based on the assumption that the knowledge concepts labeled 1 are more relevant to problems than the knowledge concepts labeled 0. Specifically, the knowledge concepts with the same mark are not comparable.

D. MODEL TRAINING

We summarize the graphical representation of the KTTLT in Figure 3, where the shaded and unshaded variables indicate the observed and latent variables, respectively. Given a student’s response tensor \( R \) and the Q-matrix provided by educational experts, our goal is to learn the parameters \( \Phi = [U, L, D, V, S^*, \alpha, \beta, a, b, s, g] \), and the posterior distribution over \( \Phi \) is:

\[
p(U, L, D, V, S^*, \alpha, \beta, a, b, s, g | R, Q) \propto p(R|R) \times p(V|Q) \times p(U|\sigma_U^2, \sigma_{U|L}^2) \times p(L|\sigma_L^2, \sigma_{L|U}^2) \times p(D|\sigma_D) \times p(S^*|\sigma_S)
\]
where $\hat{R}$ is defined as follows by combing eqs. 1, 2, 3 and 4.

$$\hat{R}_{ij} = (1 - s_j - g_j) \frac{1}{1 + \exp (-1.7a_i(U_{ij} - V_j - b_j))} + g_j \quad (10)$$

Maximizing posterior equation 9 is equivalent to minimizing the following loss function.

$$\min_{\Phi} E(\Phi)$$

$$= \frac{1}{2} \sum_{t=1}^{T} \sum_{i=1}^{N} \sum_{j=1}^{M} l_{ij}^2 [\hat{R}_{ij} - R_{ij}]^2$$

$$- \lambda_P \sum_{j=1}^{M} \sum_{p=1}^{P} \sum_{q=1}^{Q} I(q > j^+ p) \ln \frac{1}{1 + e^{-(V_p - V_q)}}$$

$$+ \frac{\lambda_D}{2} \sum_{t=2}^{T} \sum_{i=1}^{N} \left\| L_t^{i-1} - L_t^{i} \right\|_F^2 + \frac{\lambda_L1}{2} \sum_{i=1}^{N} \left\| L_t^{i} \right\|_F^2$$

$$+ \frac{\lambda_V}{2} \sum_{i=1}^{N} \left\| V_i \right\|_F^2 + \frac{\lambda_S}{2} \sum_{i=1}^{K} \left\| S_i^* \right\|_F^2 + \frac{\lambda_D}{2} \sum_{i=1}^{K} \left\| D_i \right\|_F^2 \quad (11)$$

where $\lambda_P = \frac{\sigma_p^2}{\sigma_P^2}$, $\lambda_U = \frac{\sigma_u^2}{\sigma_U^2}$, $\lambda_U1 = \frac{\sigma_u1^2}{\sigma_U1^2}$, $\lambda_L = \frac{\sigma_L^2}{\sigma_L^2}$, $\lambda_L1 = \frac{\sigma_L1^2}{\sigma_L1^2}$, $\lambda_V = \frac{\sigma_V^2}{\sigma_V^2}$, $\lambda_S = \frac{\sigma_S^2}{\sigma_S^2}$ and $\lambda_D = \frac{\sigma_D^2}{\sigma_D^2}$. $\lambda_P$ is a hyperparameter for balancing the response prediction loss and partial order loss. $\lambda_U$ and $\lambda_L$ are hyperparameters that measure how students’ knowledge proficiencies and abstract principle mastery levels change over time. $\lambda_U1$, $\lambda_L1$, $\lambda_V$, $\lambda_D$ and $\lambda_S$ are regularization parameters.

We implement the KTLT with TensorFlow\(^1\) for training our model and set the hyperparameters as $A = 10$, $P = 5$, $G = 2$, $r = 4$, $\lambda_P = 1.5$, $\lambda_U = \lambda_L = 3$ and $\lambda_S = \lambda_V = \lambda_D = \lambda_U1 = \lambda_L1 = 0.01$. The loss function is minimized using adam optimization [28].

### IV. EXPERIMENT

In this section, we validate the effectiveness of the KTLT with extensive experiments on real-world datasets.

\(^1\)www.tensorflow.org

### A. DATASET

We use two real-world datasets, POJ and HDU. The POJ and HDU datasets are crawled from the PKU OnlineJudge platform\(^2\) and HDU OnlineJudge platform\(^3\). The two datasets include submitted records from September 2018 to November 2018. In the OnlineJudge platform, students are allowed to resubmit their codes until they pass the problem. Thus, we can capture the students’ exercise logs, as shown in Table 1. Moreover, we retain 77 knowledge concepts in the HDU dataset and 10 knowledge concepts in the POJ dataset. Knowledge concepts, such as “dynamic programming”, “prim algorithm” and “depth first search”, are used to build the Q-matrix. We also filtered those students with fewer than 15 records, as well as the problems with fewer than 20 records. The statistics of the datasets after filtering are shown in Table 3.

### B. BASELINES FOR COMPARISON

We compare the KTLT with the following seven state-of-the-art methods with well-turned parameters.

- **BKT**: a kind of hidden Markov model that models students’ knowledge levels as a set of binary variables, each of which represents understanding or non-understanding of a single concept [6].
- **DINA**: a popular cognitive diagnostic model that leverages a binary vector with a Q-matrix to model each student’s knowledge levels [29].
- **IRT**: a popular cognitive diagnostic model to discover students’ knowledge levels through ranking with a logistic-like function [15].
- **PMF**: a probabilistic matrix factorization method that projects students and problems into low-rank latent factors [30].

\(^2\)poj.org

\(^3\)acm.hdu.edu.cn

### TABLE 3. The statistics of the datasets.

| Dataset | #Student | #Problem | #Record | Avg. records |
|---------|---------|----------|---------|-------------|
| HDU     | 5,714   | 2,434    | 366,836 | 64          |
| POJ     | 2,180   | 1,128    | 158,885 | 72          |

---

**FIGURE 4.** Diagnosis results of two students.
- **EERNN**: a recent deep learning method that predicts a future student’s performance by taking full advantage of the student’s exercise records and the texts of problems [11].
- **KPT**: a probabilistic matrix factorization method that tracks students’ knowledge levels by leveraging educational priors [8].

For the performance comparison with the baselines, we apply two widely used metrics, *mean absolute error* (MAE) and *root mean square error* (RMSE), as the evaluation metrics [8], [30].

### C. PERFORMANCE ANALYSIS

Figure 5 shows the overall results of all the models for predicting students’ exercise responses. There are several observations: First, our proposed KTLT model performs best on both datasets. Specifically, by combining learning transfer theory, the KTLT beats KPT and PMF and demonstrates the rationality of exploiting the cognitive structure migration theory as priors. Second, BKT, KPT and EERNN, as dynamic models, perform better than those with static assumptions (IRT, DINA, PMF), which demonstrates that it is more effective to diagnose students’ knowledge structures from an evolving perspective. Third, DINA does perform well when predicting students’ exercise responses. The performance of DINA is highly dependent on the sparseness of the Q-matrix. In our data, the Q-matrix is relatively sparse because only partial problems have been marked as knowledge concepts. This evidence demonstrates the effectiveness of the KTLT.

We present an example of the diagnosis results for two students in the HDU dataset, and the results are shown in Figure 4. Figure 4 (a) shows the two students’ mastery levels of the knowledge concepts. As a supplementary function of the KTLT, Figures 4 (b) and (c) show the knowledge structures of students A and B, respectively. The darker and wider the line connecting two knowledge concepts is, the higher the correlation between the two knowledge concepts. Compared with student A, student B is a beginner and has a lower mastery level of the knowledge concepts. Therefore, student B’s knowledge structure is not organized, and the student does not realize the differences and connections among the knowledge concepts. Note that the existing methods cannot obtain and visualize the knowledge structure for further analysis. The darker the color of the line connecting two knowledge concepts, the higher the correlation between the two knowledge concepts.

### D. IMPACTS OF DIFFERENT FACTORS

Through data analysis, we find that after repeated submissions, the probabilities for solving the problem are 89.9% (HDU dataset) and 76.6% (POJ dataset). Repeated submissions also lead to extreme imbalances between the positive and negative samples in the data. To address this problem, we sample some negative samples for each positive sample, following the negative sampling approach proposed in [31]. Figure 7 shows the performance of the KTLT at different sample rates. $\text{sampleRate} = n$ means that we sample $n$ negative samples for each positive sample. The experimental results show that as the sample rate increases, the model performance is improved. When the sample rate is 0, KTLT overfit occurs due to data imbalances. When the sample rate is greater than 5, the model performance tends
to be stable. Finally, we set sampleRate = 5 for the other experiments.

In addition, we also compare the effects of different factors on the experimental results. As shown in Figure 6, the slipping and guessing factors have little effect on the experimental results because programming is an objective problem that is less affected by the slipping and guessing factors. Using the Q-matrix as priors shows better performance on the HDU dataset because the HDU dataset has richer knowledge concept information than the POJ dataset. IRT can significantly improve performance on both the HDU dataset and POJ dataset. Specifically, the experimental results show that only considering the shallow knowledge structure hinders the performance, while considering both the shallow and deep knowledge structure can effectively improve the performance. It also verifies the correctness of the general principle [18] and cognitive structure migration theory [14].

V. CONCLUSION

In this paper, we explore the employment of learning transfer theory in educational psychology to optimize knowledge tracing for the first time. Specifically, we present a novel Knowledge Tracing model with Learning Transfer (KTLT) to track a student’s knowledge proficiency, abstract principle mastery level, and knowledge structure by applying the cognitive structure migration theory as priors. Extensive experiments on two real-world datasets demonstrate that the KTLT can effectively capture a student’s learning transfer and thus obtains superior effectiveness in predicting a student’s performance. However, the KTLT cannot track student knowledge mastery levels in real-time, and abstract principle mastery is not interpretative. In the future, we will design a probabilistic model based on the hidden Markov model, which can track a student’s knowledge proficiency and knowledge structure in real-time. In addition, we will refine the abstract principles by utilizing other theories in educational psychology and increase the interpretability of abstract principle mastery.

REFERENCES

[1] C. Geigle and C. Zhai, “Modeling MOOC student behavior with two-layer hidden Markov models,” in Proc. 4th ACM Conf. Learn. Scale (LS). New York, NY, USA: Association Computing Machinery, 2017, pp. 205–208.

[2] W. X. Zhao, W. Zhang, Y. He, X. Xie, and J.-R. Wen, “Automatically learning topics and difficulty levels of problems in online judge systems,” ACM Trans. Inf. Syst., vol. 36, no. 3, pp. 1–33, Apr. 2018.

[3] R. Wu, Q. Liu, Y. Liu, E. Chen, Y. Su, Z. Chen, and G. Hu, “Cognitive modelling for predicting examinee performance,” in Proc. 24th Int. Conf. Artif. Intell. Intell. Palo Alto, CA, USA: AAAI Press, 2015, pp. 1017–1024.

[4] Q. Liu, S. Tong, C. Liu, H. Zhao, E. Chen, H. Ma, and S. Wang, “Exploiting cognitive structure for adaptive learning,” in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA: Association Computing Machinery, Jul. 2019, pp. 627–635.

[5] F. Ai, Y. Chen, Y. Guo, Y. Zhao, Z. Wang, G. Fu, and G. Wang, “Concept-aware deep knowledge tracing and exercise recommendation in an online learning system,” in Proc. 12th Int. Conf. Educ. Data Mining, 2019, pp. 1–6.

[6] A. T. Corbett and J. R. Anderson, “Knowledge tracing: Modeling the acquisition of procedural knowledge,” User Model. User-Adapted Interact., vol. 4, no. 4, pp. 253–279, 1995.

[7] J. González-Brenes, Y. Huang, and P. Brusilovsky, “General features in knowledge tracing model multiple subskills, temporal item response theory, and expert knowledge,” in Proc. 7th Int. Conf. Educ. Data Mining, Pittsburgh, PA, USA: Univ. of Pittsburgh, 2014, pp. 84–91.

[8] Y. Chen, Q. Liu, Z. Huang, L. Wu, E. Chen, R. Wu, Y. Su, and G. Hu, “Tracking knowledge proficiency of students with educational priors,” in Proc. ACM Conf. Inf. Knowl. Manage. New York, NY, USA: Association Computing Machinery, Nov. 2017, pp. 989–998.

[9] L. Xiong, X. Chen, T.-K. Huang, J. Schneider, and J. G. Carbonell, “Temporal collaborative filtering with Bayesian probabilistic tensor factorization,” in Proc. SIAM Int. Conf. Data Mining, Apr. 2010, pp. 211–222.

[10] C. Piech, J. Bassen, J. Huang, S. Ganguli, M. Sahami, L. J. Guibas, and J. Sohl-Dickstein, “Deep knowledge tracing,” in Advances in Neural Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2015, pp. 505–513.

[11] Y. Su, Q. Liu, Q. Liu, Z. Huang, Y. Yin, E. Chen, C. H. Q. Ding, S. Wei, and G. Hu, “Exercise-enhanced sequential modeling for student performance prediction,” in Proc. 33th AAAI Conf. Artif. Intell. (AAAI), 2018, pp. 2435–2443.

[12] G. Chen, D. Davis, C. Hauff, and G.-J. Houben, “Learning transfer: Does it take place in MOOCs? An investigation into the uptake of functional programming in practice,” in Proc. 3rd ACM Conf. Learn. Scale (LS). New York, NY, USA: Association Computing Machinery, 2016, pp. 409–418.

[13] J. T. Folsom-Kovarik, B. Mostafavi, R. A. Sottilare, I. Davidson, R. Perez, and P. B. Walker, “Approaches to enhancing transfer of training using adaptive instructional systems,” in Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, New York, NY, USA: Association Computing Machinery, 2019, pp. 1–6.

[14] D. P. Ausuble, J. D. Novak, and H. Hanesian, Educational Psychology: A Cognitive View. New York, NY, USA: Traducción Española, 1978.

[15] F. Drasgow and C. L. Hulin, “Item response theory,” in Handbook of Industrial and Organizational Psychology, M. D. Dunnette and L. M. Hough, Eds. Consulting Psychologists Press, 1990, pp. 577–636.

[16] J. Zhang, X. Shi, I. King, and D.-Y. Yeung, “Dynamic key-value memory networks for knowledge tracing,” in Proc. 26th Int. Conf. World Wide Web, Apr. 2017, pp. 765–774.

[17] R. S. Woodworth and E. Thorndike, “The influence of improvement in one mental function upon the efficiency of other functions.,” Psychol. Rev., vol. 8, no. 3, p. 247, 1901.

[18] C. H. Judd, “The relation of special training and general intelligence,” Educ. Rev., vol. 36, pp. 28–42, Jan. 1908.
[19] J. H. Flavell and H. M. Wellman, “Metamemory,” in Perspectives on the Development of Memory and Cognition, R. V. Kail and J. W. Hagen, Eds. Hillsdale, NJ, USA: Erlbaum, 1977.

[20] G. Cooper and J. Sweller, “Effects of schema acquisition and rule automation on mathematical problem-solving transfer,” J. Educ. Psychol., vol. 79, no. 4, p. 347, 1987.

[21] R. Wu, G. Xu, E. Chen, Q. Liu, and W. Ng, “Knowledge or gaming?: Cognitive modelling based on multiple-attempt response,” in Proc. 26th Int. Conf. World Wide Web Companion (WWW Companion), 2017, pp. 321–329.

[22] R. S. J. de Baker, A. T. Corbett, and V. Aleven, “More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing,” in Intelligent Tutoring Systems (Lecture Notes in Computer Science), vol. 5091, Berlin, Germany: Springer, 2008, pp. 406–415.

[23] J. Tang, H. Gao, X. Hu, and H. Liu, “Exploiting homophily effect for trust prediction,” in Proc. 6th ACM Int. Conf. Web Search Data Mining (WSDM), New York, NY, USA: Association Computing Machinery, 2013, pp. 53–62.

[24] M. J. Anzanello and F. S. Fogliatto, “Learning curve models and applications: Literature review and research directions,” Int. J. Ind. Ergonom., vol. 41, no. 5, pp. 573–583, Sep. 2011.

[25] H. Von Foerster, Understanding Understanding: Essays on Cybernetics and Cognition. New York, NY, USA: Springer-Verlag, 2007.

[26] J. Liu, G. Xu, and Z. Ying, “Data-driven learning of Q-Matrix,” Appl. Psychol. Meas., vol. 36, no. 7, pp. 548–564, Oct. 2012.

[27] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), 2014, pp. 13–21.

[28] J. D. L. Torre, “DINA model and parameter estimation: A didactic,” J. Educ. Behav. Statist., vol. 34, no. 1, pp. 115–130, Mar. 2009.

[29] R. Salakhutdinov and A. Mnih, “Probabilistic matrix factorization,” in Proc. 26th Int. Conf. Neural Inf. Process. Syst. Red Hook, NY, USA: Curran Associates, 2007, pp. 1257–1264.

[30] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Proc. 26th Int. Conf. Neural Inf. Process. Syst. Red Hook, NY, USA: Curran Associates, vol. 2, 2013, pp. 3111–3119.

HENGyu LIU received the B.S. degree in computer science from Northeastern University, China, in 2017, where he is currently the Ph.D. degree in computer software and theory. His research interests include cognitive diagnosis, knowledge tracking, and educational data mining.

Tiancheng Zhang received the Ph.D. degree in computer science and theory from Northeastern University (NEU), China. He is currently an Associate Professor with the School of Computer Science and Engineering, NEU. His research interests include big data analysis, spatiotemporal data management, and deep learning.

Fan Li received the B.E. degree in computer science from Qinghai University in 2019. He is currently pursuing the master’s degree with the School of Computer Science and Engineering, Northeastern University, China. His research interest includes artificial intelligence in education.

Yu Gu (Member, IEEE) received the Ph.D. degree in computer science and theory from Northeastern University, China, in 2010. He is currently a Professor and a Ph.D. Supervisor with Northeastern University. His current research interests include big data analysis, spatial data management, and graph data management. He is a Senior Member of the China Computer Federation (CCF).

Ge Yu (Senior Member, IEEE) received the Ph.D. degree in computer science from Kyushu University, Japan, in 1996. He is currently a Professor and a Ph.D. Supervisor with Northeastern University, China. His research interests include distributed and parallel databases, OLAP and data warehousing, data integration, and graph data management. He is a member of the ACM and a Fellow of the China Computer Federation (CCF)