Personalized Exercise Recommendation via Implicit Skills

Wenying Kang
School of Information Science and Technology, Northwest University
Xi'an, China
kwy@stumail.nwu.edu.cn

Lei Zhang
School of Information Science and Technology, Northwest University
Xi'an, China
zhlei@nwu.edu.cn

Bo Li
School of Information Science and Technology, Northwest University
Xi'an, China
libo_ceasar@stumail.nwu.edu.cn

Jing Chen
School of Information Science and Technology, Northwest University
Xi'an, China
jamiechen@stumail.nwu.edu.cn

Xia Sun
School of Information Science and Technology, Northwest University
Xi'an, China
raindy@nwu.edu.cn

Jun Feng
School of Information Science and Technology, Northwest University
Xi'an, China
fengjun@nwu.edu.cn

ABSTRACT

Cognitive diagnosis methods need to assess the students' skills to provide personalized exercise recommendation. To perform this assessment, an initially hand built Q-matrix are presented to students, which would affect the recommendation results in intelligence education. However, very few previous studies have examined this exercise recommendation task on utilizing the implicit skills among exercises, in which the opinions of implicit skill might carry extra specific knowledge. We propose a data-driven frame to reconstruct Q-matrix automatically from implicit skills perspective and explore the utility of Dynamic Key-Value Memory Networks to solve this task. Experimental results demonstrate that our method has a guiding significance in pedagogical theory.

KEYWORDS

Personalized Exercise Recommendation; Q-matrix; Dynamic Key-Value Memory Networks

ACM Reference format:

Wenying Kang, Lei Zhang, Bo Li , Jing Chen, Xia Sun and Jun Feng. 2019. Personalized Exercise Recommendation via Implicit Skills. In *Proceedings of ACM TURC 2019 (SIGCSE China) conference. ACM, Chengdu, China, 6 pages.* https://doi.org/10.1145/3321408.3322849

1 INTRODUCTION

Online education platforms have gained great popularity in recent years like massive open online course (MOOC) [1], ASSISTment system [2] and cognitively diagnostic assessment [3], which provide a good opportunity for both data-driven educational assessment and intelligent tutoring.

As an important type of learning resource, exercises play an important role in these learning platforms. However, due to the massive amount of exercises, it is impossible for students to practice all exercises in a limited time. Rather than wasting their time on too easy or too hard exercises, students need to practice more with appropriate difficulty range. Therefore, how to help students to find the suitable exercises has become an important issue. In order to solve this 'Information Overload' problem, some works have focused on personalized exercise recommendation [4, 5]. Most studies on recommendation use historical data such as answer records to conduct student's analysis and predict their preferences. Usually, the effectiveness of recommendation is validated by predicting the possible scores of exercises from students' performance. So, a crucial step of the recommendation is how to get the predicted scores.

In educational, scholars believe that exercises are associated with a group of skills. Cognitive diagnosis methods (CDMs) is one of the most popular recommendation methods whose recommendation results on exercises is affected by the exercise-skill mapping. In the field of education data mining, this kind of mapping is referred to as Q-matrix [6]. Whether students get exercises right or wrong associated with the structure of the Q-matrix [7].

However, a known limitation of Q-matrix is the assumption of skill independence in exercises that involve multiple skills [8]. In other words, the Q-matrix just takes into account the relationship between the exercise and the skill defined in advance by humans, which ignores the implicit skills among exercises, that is to say, the individual skills also are connected to each other by meaningful relations. For example, in mathematics teaching in junior high school, lines and rays exist a relation called referential relation. Broek [9] raised that referential and causal/logical relations is very important and significant to human cognition, but existing cognitive diagnosis researches [10, 11] who only apply the Q-matrix ignore implicit skills information, leading to large deviation in exercise recommendation.

ACM TURC 2019, May 17–19, 2019, Chengdu, China

ACM ISBN 978-1-4503-7158-2/19/05…$15.00
https://doi.org/10.1145/3321408.3322849

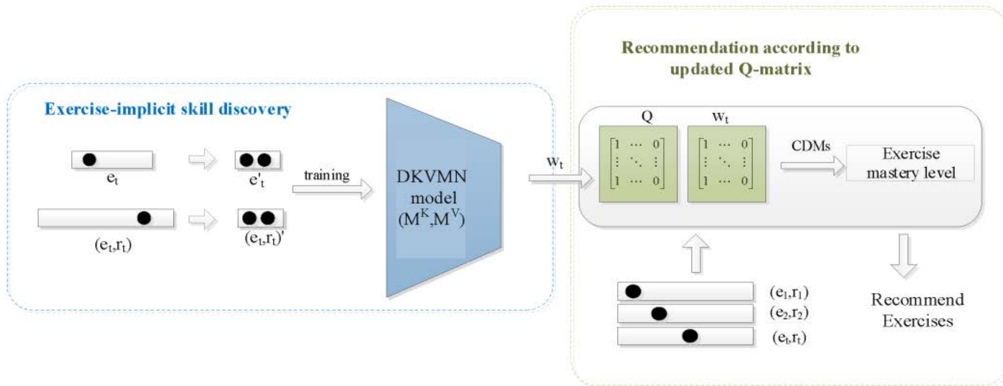

Figure 1: Methodology for the proposed exercise recommendation system

So how are the existing Q-matrix updated to incorporate the implicit skills? We think that this relation may be "signaled" by the students performance data.

In this work we study techniques for updated Q-matrix using student exercises data in cognitive diagnosis. Since it is difficult to generate implicit skills directly from students' exercises data, we can use the existing Q-matrix as a bridge. We propose personalized exercise recommendation via implicit skills (PERviaIS) framework to use exercise-implicit skill pairs with the aid of Dynamic Key-Value Memory Networks [12] (DKVMN), which is a deep learning memory networks to solve the knowledge tracing [13, 14] problem. Then we predict students' performance by combining the updated Q-matrix with cognitive diagnosis model. Finally, we recommend exercises to students according to their predicted performance.

2 RELATED WORK

In recent years, with the growing number of online choices, recommendation systems are becoming more and more indispensable. Thus, various algorithms for recommendation systems have been developed [5, 14]. Exercises performance data of students is one of the most important sources, so scholars have tried to apply these data and technology of the recommendation to the exercise recommendation.

In educational psychology, CDMs [7, 17] is a technique to predict student performance and recommend exercises by discovering student states from their answer records. CDMs firstly model students' abilities through various models. Let r_{ij} be the response of student i to exercise j, i = 1, ..., M, j = 1, ..., N, and let $\partial_i = \{\partial_{ik}\}$ be the student's binary skills vector, k = 1, ..., K, where 1 on the k_{th} element denotes mastery of skill k and 0, non-mastery of the skill. When a general interpretation is intended, the generic term attribute can be used to subsume a skill. Implementation of CDMs requires the construction of a Q-matrix [18, 6], which is a J × K matrix of zeros and ones, and the element on the j_{th} row and k_{th} column of the matrix. q_{jk}, indicates whether skill k is required to correctly answer exercise j. Table 1 shows an example of Q-matrix. Each row represents an exercise, and each column represents a skill. From table 1 we can see that exercise e_1 associates with $Skill_1$ and $Skill_3$. A Q-matrix can be viewed as a cognitive design matrix that explicitly identifies the cognitive specification for each exercise. And then, CDMs introduce the noise: slip and guessing parameters to obtain accurate performance. That is, students who possess all the required skills for an exercise can slip and miss the exercise, and students who lack at least one of the required skills can guess and still answer the exercise correctly with typically nonzero probabilities. Zhu [4] has proposed a three-step exercise recommendation method called PMF-CD, which apply the above cognitive diagnosis results combined with probabilistic matrix factorization.

In spite of importance of previous studies, there are still some limitations in existing methods. Determining which skill are involved in an exercise can prove both tedious and difficult [18]. Machine learning methods often outperform humans over a range of complex tasks for it can offer an effective method to get the correct exercise-skill mapping. So, we want to automatically identify significant implicit skill from student's performance data and construct the Q-matrix based on a data-driven frame.

Table 1 An example of Q-matrix

Exercise	$Skill_1$	$Skill_2$	$Skill_3$
e_1	1	0	1
e_2	0	0	1
e_3	0	1	0

3 Personalized Exercise Recommendation Framework With DKVMN

In this research, a novel data-driven exercise recommendation framework named PERviaIS is proposed, which is divided into three steps. In the first step, we use a novel method to extract the exercise-implicit skill pairs from student's performance data automatically. In the second step, we integrate the exercise-implicit skill pairs into cognitive diagnosis model's Q-matrix as the updated Q-matrix. In the third step, we recommend exercises to students according to their predicted performance by using CDMs. The overall model architecture is shown in Figure 1.

3.1 The utilization of Dynamic Key-Value Memory Networks

In this section, we first introduce the basic DKVMN model. Then we describe how our method discover implicit skills with DKVMN.

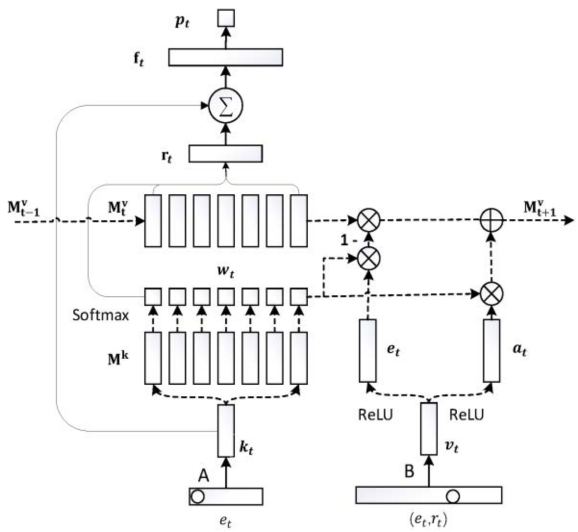

Figure 2: Architecture for Dynamic Key-Value Memory Networks

3.1.1 Dynamic Key-Value Memory Networks Model. The DKVMN model was primarily designed to solve the knowledge tracing task, which is formulated as a supervised sequence learning problem: given a student's past exercise performance data $X = \{x_1, x_2 ... x_t\}$, predict the probability that the student will answer a new exercise correctly, i.e., $p\ (r_t = 1|r_t;\ X)$. Input $x_t = \{e_t;\ r_t\}$ contains the exercise e_t, which student attempts at the time stamp t, and the correctness of the student's answer rt. Figure 2 shows the architecture for Dynamic Key-Value Memory Networks.

DKVMN has one static matrix called key, which stores the skill representations and the other dynamic matrix called value, which stores and updates the students skill state of each skill. We assume there are N implicit skills $\{c^1, c^2, ..., c^N\}$. So these skills are stored in the key matrix M^k (of size $N \times d_k$) and the students mastery levels of each skill, i.e., skill states $\{s^1, s^2 ..., s_t^N\}$ are stored in the value matrix M^v (of size $N \times d_v$), which changes over time. The output p_t is the probability that the student will answer each exercise correctly in the next time stamp. During training, both exercise embedding matrices, as well as other parameters and the initial value of M^k and M^v are jointly learned by minimizing a standard cross entropy loss betweenp_t and the true value r_t.

$$L= -\sum_t (r_t \log p_t + (1 - r_t) \log (1 - p_t)) \quad (1)$$

3.1.2 Automatic exercise-implicit skill pairs discovery. After training the DKVMN well, we can use an intermediate parameter called correlation weight (w_t), which implying the strength of the inner relationship between the exercise and the skill. We can get the exercise-implicit skill pairs by:

$$w_t = \text{Softmax}(\hat{e} M^{k(T)}), \quad (2)$$

where $\hat{e}$ is the exercise embedding results that encoded, stored, and carefully updated through the DKVMN. M^k is the N implicit skill embedding. By assigning the exercise to the implicit skill with the largest correlation weight value, we can get exercise-implicit skill pair for each exercise.

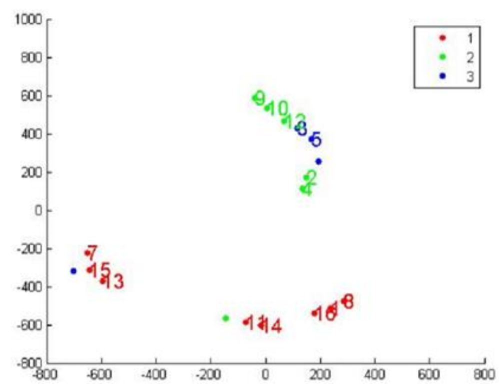

Figure 3: Exercise-implicit skill pairs when N is set to be 3

3.2 The Method of Personalized Exercise Recommendation Framework

Exercising with exercises is a necessary process for students to acquire new knowledge. Currently online exercise banks are commonly chosen by teachers for task assignments or by students for autonomous study. But when there are too many exercises in an exercise bank, teachers may not be able to tailor the task according to students' ability and students may waste their time in doing the exercises that are too easy for them. Therefore, how to help students to find the suitable exercises has become an important issue. We propose the following method to recommend exercises with the same implicit skills of the exercise.

Firstly, get exercise-implicit skills pairs. All exercises can be grouped into different clusters according to the implicit skill size N from DKVMN. And we get this exercise-implicit skill pairs w_t from section 3.1. The exercise cluster graph in figure 3 is drawn using t-SNE [19] by projecting the multi-dimensional correlation weight of Math1 dataset to the 2-D points. The digital represents exercise tag and the exercises from the same implicit skill are labeled with the same color.

Secondly, update Q-matrix using exercise-implicit skill pairs discovered in the previous step. Since it is difficult to generate

Table 2 updated Q-matrix

Exercise	$Skill_1$	$Skill_2$	$Skill_3$	$\boldsymbol{Skill\_new_1}$	$\boldsymbol{Skill\_new_2}$
e_1	1	0	1	**0**	**1**
e_2	0	0	1	**1**	**0**
e_3	0	1	0	**0**	**1**

implicit skills directly, we can use the existing Q-matrix as a bridge. Specifically, we integrate the automatic mapping results of w_t into existing Q-matrix. Table 2 shows an example of updated Q-matrix from table1. The bold values indicate the cluster result of exercises. Clustering is based on the simple algorithm which assigns each exercise to one of the N clusters based on the maximum column value in matrix w_t. For example, *Skill_new₁* and *Skill_new₂* indicate that the N is two and e_1 and e_3 are belong to a cluster. e_2 is another cluster. Figure 4 is the process to get the updated Q-matrix.

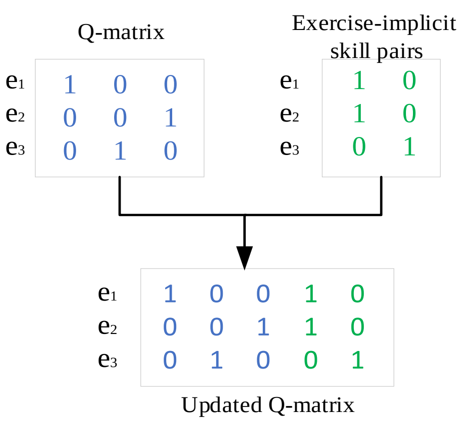

Figure 4: the process to get updated Q-matrix

Thirdly, using CDMs to predict student's performance and recommend exercise. More concretely, using our updated Q-matrix as CDMs Q-matrix to carry out experiments.

This method [4] improves the DINA model by changing the skills state to a continuous value from 0 to 1. Specifically, the posterior probability ∂_{ik} that will be obtained is as follows.

$$\partial_{ik} = P(\partial_{ik} = 1|r_i) = \frac{\sum_{\partial_{ik}=1} P(\partial_x|r_i)}{\sum_{x=1}^{2^K} P(\partial_x|r_i)} \tag{3}$$

After that, the mastery of each exercise S_{ij} is obtained by the probability of the student at each known skill state.

$$S_{ij} = \prod_{k=1}^{K} {In_{ijk}}^{\sum_{k=1}^{K} q_{jk}} \tag{4}$$

where the In_{ijk} is

$$In_{ijk} = \begin{cases} 1, & q_{jk} = 0 \\ \partial_{ik} \times q_{jk}, & q_{jk} = 1 \end{cases} \tag{5}$$

In the end, using the average mastery of the students' exercises S_{ij}, combined with guessing g_j and slip s_j of the exercises, the students' actual performance A_{ij} can be finally predicted as follows.

$$A_{ij} = \begin{cases} \frac{(1-s_j)S_{ij}}{(1-s_j)S_{ij}+g_j(1-S_{ij})}, & r_{ij} = 1 \\ \frac{s_j S_{ij}}{s_j S_{ij}+(1-g_j)(1-S_{ij})}, & r_{ij} = 0 \end{cases} \tag{6}$$

We set the student's exercise mastery level to be 0.3 or other values, that is, we will recommend the exercise that student's predicted performance is less than 0.3 or other values.

4 EXPERIMENTS

4.1 Datasets

The experiments are conducted on three real-world datasets [11]. The first dataset is composed of the scores of middle schools on fraction subtraction problems. The other two datasets are collected from two final mathematical exams from high school students. Each of the dataset is represented by a score matrix and a given Q-matrix by education experts. The brief summary of these datasets is shown in Table 3.

Table 3 Dataset Summary

Dataset	Student	Skill	Exercise
FrcSub	536	8	20
Math1	4209	11	20
Math2	3911	16	20

4.2 DKVMN Model Results

First, we train the DKVMN model on each dataset using AUC. We implement the models using MXNet [15] on a single CPU E5-2623. Figure 5 gives the AUC value on the testing data within 50 iterations.

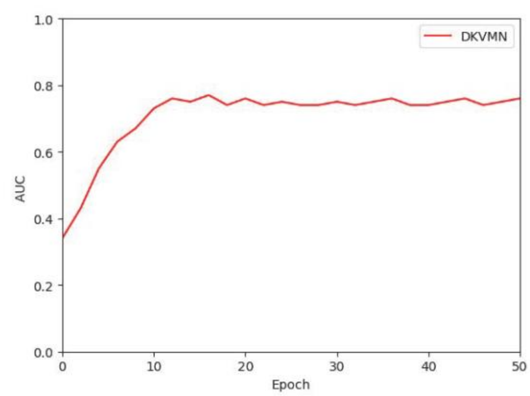

Figure 5: The AUC in testing time over 50 iterations for FrcSub dataset

4.3 State-of-art Methods Used for Performance Comparison

In the experiments, we compare our proposed approach with the following exercise methods.

DINA methods [10]. Diagnosis the knowledge state of the student according to DINA methods. And then recommend exercises to student associated with the skills according to the knowledge state of different master level.

FuzzyCDF (Fuzzy cognitive diagnosis framework) [11]. Extra the information from both objective and subjective exercises by fuzzifying the skill proficiency for students' cognitive modeling.

PMF (Probabilistic Matrix Factorization) [18]. A latent factor model projecting students and exercises into a low-dimensional space.

NMF (Non-negative matrix factorization) [16]. A latent non-negative factor model and can be viewed as a topic model.

PMF-CD (Probabilistic Matrix Factorization and Cognitive Diagnosis) [4]. Combine the complementary advantages of PMF and cognitive diagnosis, which takes both individual and common study status of students into account.

4.4 the Performance Results

4.4.1 Number of Implicit Skills. In section 3, we propose a novel method to extract the exercise-implicit skill pairs automatically, which has one parameter: N determines the number of latent skills. First, we vary N to study how does the N of exercise-implicit skill pairs impact the performance. For FrcSub dataset (Figure 6), the peak performance is achieved when N is 4; similarly, for Math1 dataset, the optimal N is 4. When N becomes smaller, the performance degrades significantly. This highlights the necessity of accounting for the Q-matrix when using CDMs for exercise recommendation. Moreover, when N is set too large, the performance also suffers. Based on this observation, we verified that what DeCarlo [7] has shown that the cognitive diagnosis results are largely associated with the structure of the Q-matrix. This indicates the effectiveness of our automatic exercise-implicit skill pairs discovery strategy. In the following experiments, we fix N according to the best performance evaluated by DKVMN.

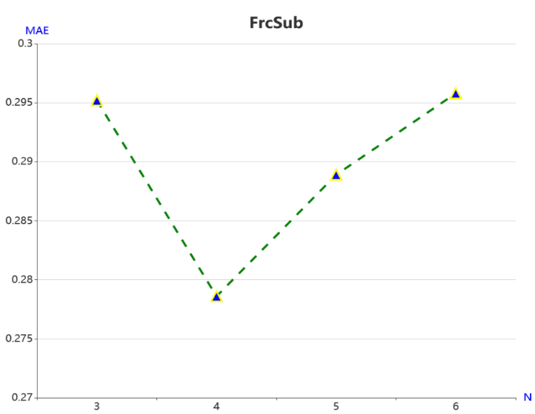

Figure 6: N vs. MAE

4.4.2 Performance Comparison. We train our model on the 80% of the interactions and holding out the last 20% for testing. We use mean absolute error (MAE) as the evaluation metrics. Figure 7 shows the performance results of the comparison. As shown in the performance results, our approach has achieved better performance than other methods because we apply the updated Q-matrix which offer a more objective and replicable means of getting the implicit skills from students' performance data, which evaluate students' strengths and weaknesses that can allow for effective measurement of student learning and progress, designing of better instruction.

5 CONCLUSION AND FUTURE WORK

The construction of a Q-matrix from students performance data is a highly desirable goal for tutoring systems. Our method allows a more effective means to build Q-matrix which from implicit skills perspective and apply it to cognitive diagnosis, as machine learning methods often outperform humans over a range of complex tasks. We except that our new method can provide more significant benefits when deployed in real-world tutoring systems. Not only serving as an attempt to bridge data mining with pedagogical and learning theory, our work also raises attention in the community for deeper, robust student learning. For future work, we will incorporate content information into the exercise embedding to further get the better exercise recommendation performance.

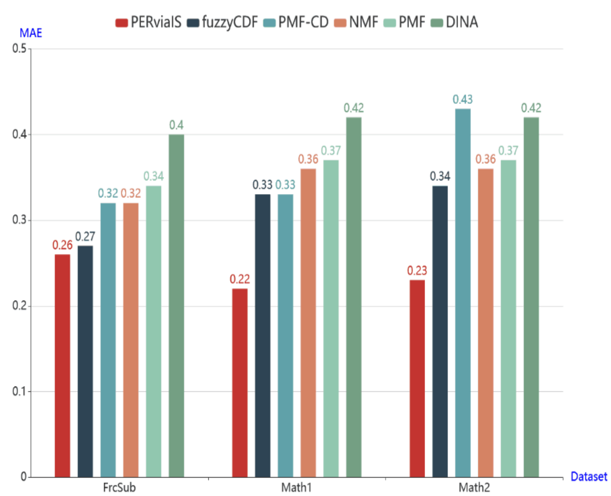

Figure 7: MAE Performance Comparison with Other methods

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (61877050), the Shaanxi Province Key Lab. of Satellite and Terrestrial Network Tech. R&D, and Support Returned Talented Scholars Research project in Shaanxi Province (202160002).

REFERENCES

[1] Anderson A, Huttenlocher D, Kleinberg J, et al. Engaging with massive online courses[C]//Proceedings of the 23rd international conference on World wide web. ACM, 2014: 687-698.
[2] Feng M, Heffernan N, Koedinger K. Addressing the assessment challenge with an online system that tutors as it assesses[J]. User Modeling and User-Adapted Interaction,2009, 19(3):243-266.
[3] Nichols P D, Chipman S F, Brennan R L. Cognitively diagnostic assessment[M]. Routledge, 2012.

[4] Zhu T Y, Huang Z Y, Chen E H, et al. Cognitive Diagnosis Based Personalized Exercise Recommendation. Chinese Journal of Computers,2017,41(1): 176-191.
[5] Fang L, Tuan L A, Hui S C, et al. Personalized question recommendation for English grammar learning[J]. Expert Systems, 2017: e12244.
[6] Tatsuoka K K E. Analysis of Errors in Fraction Addition and Subtraction Problems. Final Report[J]. 1984.
[7] DeCarlo L T. On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix[J]. Applied Psychological Measurement, 2011, 35(1): 8-26.
[8] Huang Y, Guerra-Hollstein J D, Brusilovsky P. Modeling Skill Combination Patterns for Deeper Knowledge Tracing[C]// Intl Workshop on Personalization Approaches in Learning Environments. 2016.
[9] Van den Broek P. Using texts in science education: Cognitive processes and knowledge representation[J]. Science, 2010, 328(5977): 453-456.
[10] Torre D L, J. DINA Model and Parameter Estimation: A Didactic[J]. Journal of Educational and Behavioral Statistics, 2008, 34(1):115-130.
[11] Liu Q, Wu R, Chen E, et al. Fuzzy cognitive diagnosis for modelling examinee performance[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2018, 9(4): 48.
[12] Zhang J, Shi X, King I, et al. Dynamic key-value memory networks for knowledge tracing[C]//Proceedings of the 26th international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2017: 765-774.
[13] Corbett A T, Anderson J R. Knowledge tracing: Modeling the acquisition of procedural knowledge[J]. User Modeling and User-Adapted Interaction, 1994, 4(4):253-278.
[14] Piech C, Spencer J, Huang J, et al. Deep Knowledge Tracing[J]. Computer Science, 2015, 3(3):págs. 19-23.
[15] Chen T, Li M, Li Y, et al. MXNet: A Flexible and Efficient Machine Learning Library for Heterogeneous Distributed Systems[J]. Statistics, 2015.
[16] Lee D D, Seung H S. Algorithms for non-negative matrix factorization[C]// Advances in neural information processing systems. 2001: 556-562.Conference Name: ACM Woodstock conference.
[17] Dibello L V, Roussos L A, Stout W. 31A Review of Cognitively Diagnostic Assessment and a Summary of Psychometric Models 1, 2[J]. Handbook of Statistics, 2006,26(06):979-1030.
[18] Sun Y, Ye S, Inoue S, et al. Alternating recursive method for Q-matrix learning[C]//Educational Data Mining 2014. 2014.
[19] Prezgonzlez A, Vergara M, Sanchobru J L, et al. Visualizing Data using t-SNE[J]. Journal of Machine Learning Research, 2008, 9(2605):2579-2605.