

Discovering Knowledge-Point Importance From the Learning-Evaluation Data

Hongfei Guo, Shandong Normal University, China*

Xiaomei Yu, Shandong Normal University, China

Xinhua Wang, Shandong Normal University, China

Lei Guo, Shandong Normal University, China

Liancheng Xu, Shandong Normal University, China

Ran Lu, Shandong Normal University, China

ABSTRACT

As students in online courses usually show differences in their cognitive levels and lack communication with teachers, it is difficult for teachers to grasp student perceptions of the importance of knowledge-points and to develop personalized teaching. Though recent studies have paid attention to this topic, existing methods fail to calculate the importance of every knowledge-point for each student. Moreover, some studies are based on expert analysis, are not data-driven, and hence, are inapplicable to large-scale online scenarios. To address these issues, this article proposes a personal topic rank (PTR) as a solution, which links students and concepts to generate a personalized knowledge concept map. Then, the authors present a novel PTR method to calculate the importance of knowledge-points, wherein student mastery of knowledge-points, student understanding, and the knowledge-point itself are considered simultaneously. This article conducts extensive experiments on a real-world dataset to demonstrate that the method can achieve better results than baselines.

KEYWORDS

Concept Map, Distance Education, Learning Analysis, Learning-Evaluation Data, Online Courses, Personalized Difference, Personalized Teaching, Random Walk

INTRODUCTION

Nowadays, the popularization of online course learning systems, such as Xuetangx (www.xuetangx.com) and Canvas (www.canvas.net), makes it possible for students to receive high-level continuing education regardless of time and location limitations. However, given the massive data generated in these online platforms and the lack of face-to-face communication between teachers and students, it is difficult for a teacher to know a student's learning state. Students have different learning backgrounds and knowledge, which also aggravates this lack of understanding. For example, some teachers believe that students who spend more time watching the teaching videos by the teacher in online courses

DOI: 10.4018/IJDET.302012

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

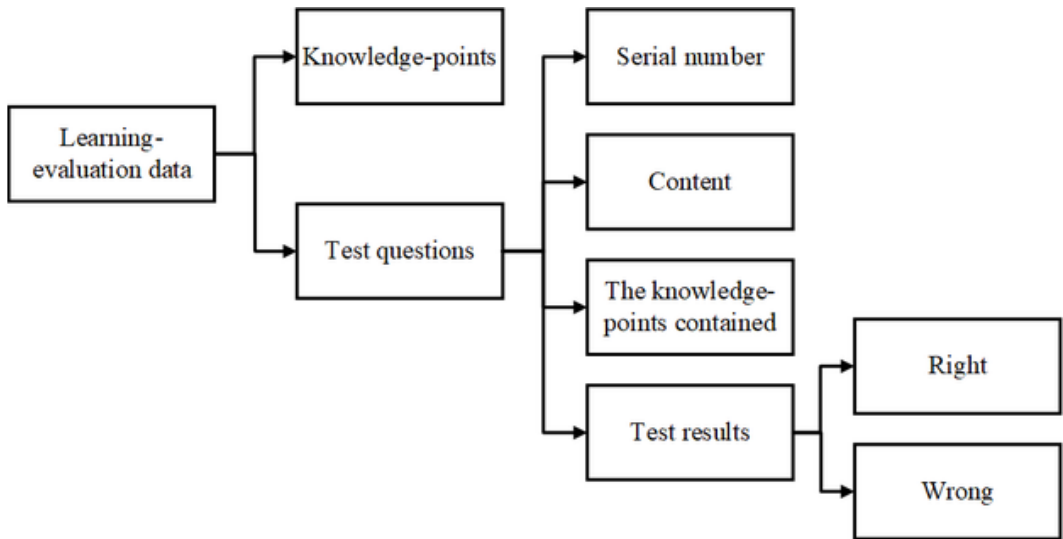
can master the course well, but that is not always the case, probably due to learners' lack of required background knowledge in the course subject matter. Moreover, as courses are composed of sets of concepts or knowledge-points, directly modeling the course information will ignore how students learn different concepts and hence cannot consider knowledge-points from the perspective of different students in order to assess the state of their learning. Therefore, the objective of our work is to discover the importance of different knowledge-points to different students and to achieve it by designing a personalized method. Intuitively, if teachers can discover the importance of each knowledge-point to different students, it will help teachers carry out personalized teaching. Based on the learned knowledge-point importance, teachers are able to further suggest other related knowledge to the student to improve her learning efficiency. For example, if teachers know *list* (a knowledge-point in Python) is important to a student, they may recommend the related *cycle* knowledge-point. Moreover, the teacher may also need to adjust the teaching strategy for different students, since students have different learning priorities.

In recent decades, applying machine learning and artificial intelligence technology to study student learning processes and improve personalized teaching has always been a research hotspot. However, only a few works are focused on identifying the importance of personalized knowledge-points. For example, Leake et al. (2004) modeled the importance of concepts in concept maps by assessing how a series of potential structural factors combine to affect human judgments of the importance of concepts. Wu et al. (2007) took the value of Hub as the concept importance according to the graph structure characteristics of ontology. They calculated the weight of concept importance by using the iterative method of mutual enhancement of concepts and relations. Wang, Zhang, et al. (2020) constructed a knowledge map of university physics based on a mind map of university physics, in which an expert questionnaire and natural language analysis method were adopted to obtain the importance of knowledge-points. A mind map is an approach to the organization of the human mind that prepares the ground for thinking (Baghestani et al., 2021). Though previous studies represent outstanding achievements, there are still some shortcomings to be addressed. First, in the online learning scenario, such as the Xuetangx learning platform, there is no upper limit on the number of students enrolled in each course, resulting in a large number of students participating in courses. Previous studies are mainly based on expert analysis, are not data-driven, hence inapplicable to the large-scale online scenarios. Secondly, existing methods fail to calculate the importance of every knowledge-point for each student. In addition, as a record of student progress in practice or test, learning-evaluation data (the content of which is shown in Figure 1) provides us a new research opportunity to study the importance of knowledge-points to each student.

To better discover the personalized importance of each knowledge-point for every student, the authors propose the personal ranking method, or PTR, as a solution. Precisely, to capture the structure of knowledge-points, and link students and concepts in a unified graph, through the frequency rule mining algorithm, this article first adaptively generates a *personalized knowledge concept map* (PCM) from student learning-evaluation data with multiple types of nodes and correlations. Then, this article presents a novel random-walk ranking algorithm to calculate the arrival probability between student nodes and any concept node, which is further deemed as the personalized knowledge-point importance for every student or participant. Compared with traditional ranking methods on graphs, solutions can simultaneously consider the degrees of student mastery and understanding and the knowledge-point itself. Finally, the authors conduct extensive experiments on a real-world dataset to demonstrate that the proposed method can achieve better results than other related baselines.

The main contributions of this work are listed as follows. Firstly, the authors have adaptively generated a PCM from student learning-evaluation data with multiple types of nodes and correlations. PCM not only can link students and knowledge-points in a unified graph but also can capture the structure of knowledge-points, laying the foundation for calculating the personalized importance of knowledge-points. Secondly, the authors further devise a novel random-walk ranking method based on the PCM, namely the PTR method. This method is used to calculate the personalized importance

Figure 1. The content of the learning-evaluation data



of knowledge-points for each student or participant, wherein the degrees of student mastery and understanding and the knowledge-point itself are simultaneously considered. Finally, the authors conducted extensive experiments on a real-world dataset, and the experiment results demonstrate the advantages of the proposed method compared with several state-of-the-art baselines.

RELATED RESEARCH

The learning-evaluation data is a record of student progress that allows us to understand student cognition and the importance of each knowledge-point. In this section, the authors present related work, that is, designing discriminant models for knowledge-importance discovery and the learning diagnosis models based on this kind of data.

Knowledge-Point Importance Discovery Methods

Existing research on knowledge-point importance discovery is either based on small, fabricated samples or non-personalized methods.

In the first type of method, the research is mainly based on a questionnaire or well-designed instructional design. For example, Li et al. (2019) found important knowledge-points in the course by focusing on the logical relationship, correlation mechanisms, and the guiding relationship between knowledge-points. Yao (2017) analyzed how to construct essential knowledge in teaching from four perspectives: creating inquiry space, selecting resources, positioning logic, and designing activities. Wang et al. (2021) designed the preference index of gamification elements in the course, and they found that students preferred the importance of activity elements and mechanism elements through questionnaire surveys. However, the questionnaires and well-designed instructional design hinder the above studies from being applied to large-scale online learning platforms because teachers had a high degree of participation in the questionnaires or instructional design, limiting the scalability of these methods.

Another type of study is concept-map-based methods. For example, Kardan and Razavi (2014) proposed an evaluation method for knowledge level based on a concept map. They used the score and total score of each concept to calculate the importance of each concept through a neural network.

Koponen and Nousiainen (2014) used centrality and similarity measures to find key knowledge-points in concept maps, and they proved that their method can reliably identify a group of important knowledge-points in the maps. Fernández-Álvarez et al. (2021) designed an importance ranking method on a complex knowledge graph to rank the class importance. In natural language, Ren et al. (2021) evaluated the importance of knowledge from a semantic perspective. Zhong et al. (2021) analyzed the correlation between candidate concepts through semantic similarity to calculate the semantic weight of concepts and extract important concepts. An et al. (2018) built the domain knowledge network based on knowledge association relationships. They follow and analyze the time series with relevant indexes of centrality and clustering to discover essential knowledge and knowledge clusters. Some of these works are aimed mainly at discovering the importance of knowledge-points, and some of them study this task as an intermediate step for other purposes. Moreover, existing studies used non-personalized methods, and they could only estimate the importance of knowledge-points at the course level, not at a personal level.

In summary, the above studies at least have two shortages: (a) the concept maps are manually made by students and teachers, which limits the application of these methods, and (b) due to the lack of personalized information on the data used or the method designed is non-personalized, so the discovered knowledge-point importance is non-personalized, which hinders us from knowing a student's learning state. Therefore, it is essential for us to develop a personalized ranking method to discover the importance of knowledge-points from student perspectives.

Personalized Learning Diagnosis on Learning-Evaluation Data

As far as the authors know, only a few works focus on mining student learning status from the learning-evaluation data. In this section, the authors take the learning diagnosis models based on learning-evaluation data as the related works. Learning-evaluation data is a subset of education big data, which is the relevant evaluation data generated and collected in the whole learning process, including test questions, test results, and other information (Wang et al., 2019). Through the measuring tools of the learning system, the learning-evaluation data can be recorded, measured, and evaluated student learning processes, the state of knowledge, and learning ability in the learning situations to help the student learning state (Mou & Li, 2019). As one of the diagnostic tools, the learning-evaluation data of students is objective (Lee et al., 2009). In addition, because each student's learning-evaluation data is different, these data contain rich, personalized information.

Learning diagnosis is defined as the diagnostic evaluation of an individual's learning status. It includes knowledge mastery (Cheng et al., 2019), learning motivation intensity (Chu & Hung, 2015), and learning strategy (Makhambetova et al., 2021). When focusing on knowledge and skills, learning diagnosis can be regarded as applying cognitive diagnosis to learning evaluation (Zhan et al., 2020). The learning diagnosis model uses personalized student information to carry out learning diagnoses. In essence, it is a personalized model (Chu et al., 2010).

Learning-evaluation data has been used successfully in learning diagnosis models and modeled students from a personal perspective. There are many cases of learning diagnosis models based on learning-evaluation data. Wang et al. (2020) projected student and learning-evaluation data onto the factor vector, modeled their interaction using multiple neural layers, and proposed a neural network cognitive diagnosis model. Using the correlation between knowledge-points in the learning-evaluation data, Jin et al. (2020) updated student learning statuses, to rank the questions and recommend them to the students. According to the learning-evaluation data of students, Chen and Sue (2013) used an association rule algorithm to automatically generate concept maps without expert intervention, that is, to calculate the frequency of correctly answered questions between two knowledge points. The concept maps show positive influences on student academic achievements and quality of education (Hafeez, 2021). Based on the work of Chen and Sue (2013), Kim et al. (2020) proposed the optimization method of the associative knowledge graph using TF-IDF-based ranking scores. As stated in Pinandito et al. (2021), Fatawi et al. (2020), and Ma and Shi (2016), concept maps are applied in classroom

teaching and learning systems. Li et al. (2018) used helpful information in the learning-evaluation data to analyze a concept map. They diagnosed the weak knowledge-points of students and provided remedial paths for them, with the promotion of online courses and the continuous increase of users.

With progress in a course, the differences in student cognition of the importance of each knowledge-point in the course become increasingly apparent. Moreover, the learning-evaluation data has the characteristics of objectivity and diversity, which can model students from a personal perspective. Therefore, we believe that it is feasible to use learning-evaluation data to design and develop a personalized discrimination model of knowledge-point importance, and the feasibility is verified at the methodology and experimental level.

METHODOLOGY

In this section, the authors summarize the overall structure of the proposed method and then present how we implement it.

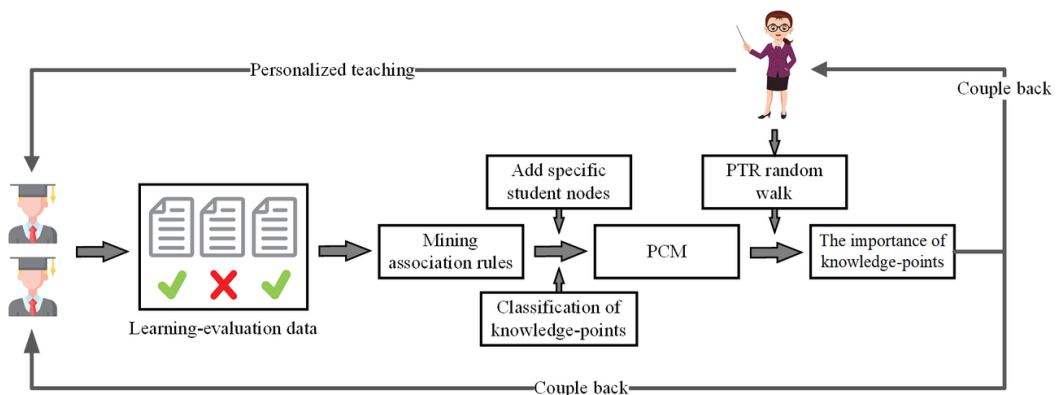
Overview of Our Solution

Figure 2 shows the overview of the proposed method, which consists of two components, i.e., constructing PCM, and the PTR ranking method based on that.

The process of building PCM is shown as follows. The authors first collected student learning-evaluation data from an online course platform of a university in Shandong Province, China. These data were collected in September–November 2020 and September–November 2021. As shown in Figure 1, from the collected data, authors can obtain student answers to specific questions and the test results, as well as the knowledge-points contained in questions. According to student answers to questions and the knowledge-points contained in questions, we can further infer their answers to the related knowledge-points, which are divided into three difficulty levels: low, ordinary, and high level. The authors linked the students who have answered them and the knowledge-point in a graph with multiple types of nodes. The student nodes in the graph are utilized to conduct personalized random walks. To this end, the constructed PCM not only reflects the coupling relationship between knowledge-points but also student mastery of different knowledge-points.

To learn the importance of knowledge-points to different students, the authors further developed a personalized ranking method based on PCM. That is, the authors treat the arrival probability from student node to knowledge-point node as the measure of how valuable the knowledge-points are to different students. To achieve this goal, rather than applying a traditional node ranking method, a

Figure 2. Overview of the proposed method



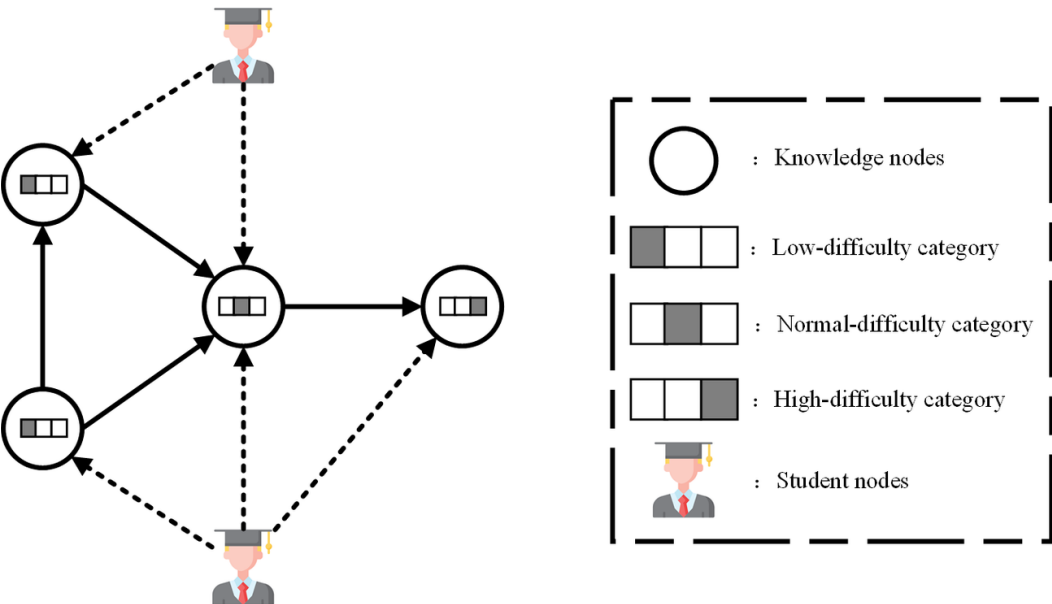
novel knowledge-point importance ranking method is developed, where student mastery of knowledge-points, student understanding, and the difficulty of knowledge-points are considered simultaneously.

Constructing the Personalized Knowledge Concept Map (PCM)

In this section, the authors present how to construct PCM from the learning-evaluation data and show the process in Figure 3, which shows that a PCM consists of two kinds of nodes: a student node and knowledge-point node. Each knowledge node has three difficulty levels: high, standard, and low difficulty levels. The connections between the student node and knowledge-point node indicate the student can master this knowledge-point well. The relationships between two knowledge-points denote whether these two nodes are correlated or not. In the following, the authors detail how to construct a PCM from the learning-evaluation data.

First, the authors collected student learning-evaluation data from the learning platform, mainly including the knowledge-points contained in the test question and the corresponding test result. Here, the test result refers to each student answer to all the questions in the test question, which can be divided into correct answers and wrong answers. In addition, before learning-evaluation data generation, the knowledge-points contained in questions has been extracted from the test questions by domain experts. In this work, in order to be more realistic, we assumed that each question has at least one knowledge-point. To build connections among knowledge-points, the authors used the frequency rule mining algorithm. That is, the authors treated question pairs as the 2-itemset between questions and their answer results (both questions are answered correctly, or both are answered incorrectly) as the instances of these 2-itemsets. Here, question pairs, such as (Q_x, Q_y) and (Q_y, Q_x) , are determined a priori according to the sequence of test questions in a test. This article calculates the confidence that both questions are correctly answered according to the question pairs. The formula for calculating the confidence is shown in Eq. (1):

Figure 3. An example of the PCM

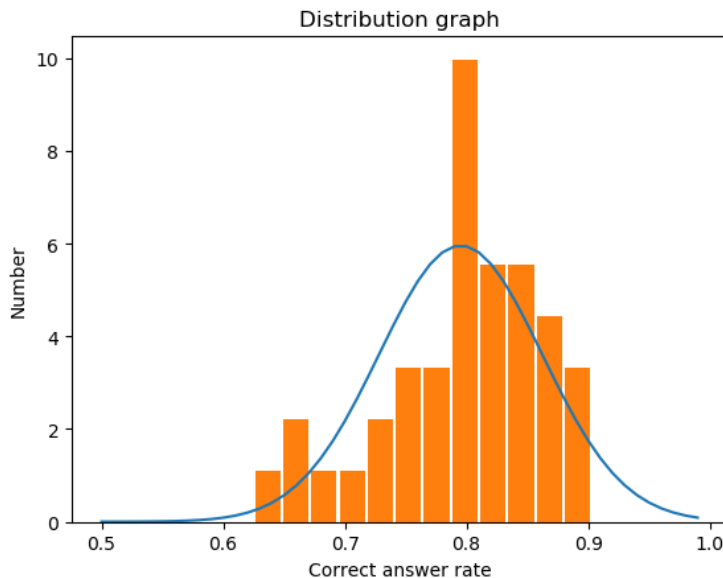


$$Confidence(Q_x \rightarrow Q_y) = \frac{Support(Q_x, Q_y)}{Support(Q_x)} \quad (1)$$

where Q_x represents the question x , Q_y represents the question y , $Support(Q_x, Q_y)$ denotes the support of the question pairs 2-itemset (Q_x, Q_y) , $Support(Q_x)$ denotes the support of the 1-itemset Q_x , and $Confidence(c)$ is the confidence of the association rule $Q_x \rightarrow Q_y$. Based on this, the authors further deduced the confidence level of knowledge-point pairs by mapping knowledge-points to the related questions and deeming the knowledge-point pairs as association rules. Moreover, suppose the confidence level of an association rule is smaller than the minimum confidence threshold c . In that case, the authors would delete it and only keep the rules that satisfy the minimum confidence level as the edges in the concept graph. In addition, if there are conflicting association rules, such as knowledge-point $a \rightarrow b$ and $b \rightarrow a$, the authors only retain one of them with the greatest confidence level.

Second, based on student answers to the questions and the knowledge-points contained in questions, the authors can infer the correct answer rate of the knowledge-points themselves. Based on the distributions of correct answer rates on the whole data, the authors find that the correct answer rate of knowledge-points and the number of corresponding knowledge-points are approximately normal distribution, as shown in Figure 4. There are fewer knowledge-points with high and low correct answer rates, while there are more knowledge points with medium correct answer rates. The authors interviewed 16 college teachers with rich teaching experience in a face-to-face manner, and they agreed that the following proportion was the most reasonable. That is, according to expert experience, authors set the top 20% knowledge-points with the highest correct answer rate, the 20% knowledge points with the lowest correct answer rate, and the remaining 60% knowledge points as low-difficulty, high-difficulty, and normal-difficulty, respectively.

Figure 4. An approximately normal distribution graph



Finally, to learn personalized student learning states, the authors linked them with the knowledge-points that they have answered with high accuracy in the knowledge graph (the threshold is set as hyper-parameter in experiments). The authors introduce students as the nodes in the graph because the goal is to learn the importance of every knowledge-point from a student's view. Based on this, the authors can develop personalized knowledge-point importance ranking methods further.

The Knowledge-Point Importance Discovery Method PTR

Although the importance of knowledge-points can be calculated on PCM by the traditional random walk methods (Brin & Page, 1998; Haveliwala, 2003; Tong et al., 2008), i.e., calculating the arrival probability for every node pair. But the above methods cannot be directly applied to the personalized knowledge-point importance ranking task, since it ignores the difficulty of knowledge-points themselves and the differences of students.

In this work, the authors take finding the personalized importance of knowledge-points to students as the goal and innovatively propose a personalized knowledge-point significance calculation method based on the random walk process, namely PTR. This method makes full use of the personalized information reflected by student nodes on PCM. And more importantly, in PTR, the calculated importance of knowledge-points not only reflects knowledge-point difficulty levels but also the differences between degrees of student mastery. The symbols used in the proposed method are shown in Table 1.

Overview of PTR

The walking strategy of the PTR algorithm starts from the student node, which prepares to generate the importance of knowledge-points and carries out random walks with the probability of according

Table 1. The symbols utilized in the proposed method

Symbol	Meaning	Description
N	The number of all nodes in PCM	Including student nodes and knowledge nodes
α	Hyper-parameter	$\alpha \in [0, 1], \alpha + q \in [0, 1]$
q	Hyper-parameter	$q \in [0, 1], \alpha + q \in [0, 1]$
e	Difficulty level	e selected from the set {low-difficulty, normal-difficulty, high-difficulty}
s_e	Difficulty level matrix	s_e is a 0-1 matrix of $N \times 1$
o	Student code	Student code from 1 to o
r_o	Student o ' student matrix	r_o is a student matrix of $N \times 1$
M	The adjacency matrix of PCM	M is a $N \times N$ matrix, in which the sum of each column is equal to 1
R	Knowledge-points' importance matrix of students	$R = \{R_1, R_2, \dots, R_o\}$, R_o is the personal importance matrix of student o to each knowledge-point

to the network structure of PCM (i.e., the outgoing degree and incoming degree of every node). It randomly jumps to the knowledge-point node according to the difficulty level with a probability of q , while it jumps back to the initial student node with the probability α until the iteration ends after it converges. The difficulty level can be selected by teachers who want to know the degrees of student mastery in different learning levels.

Compared with the traditional random walk-based algorithms, the PTR method can simultaneously consider a knowledge-point's difference difficulty levels and a student's personal mastery. In the PTR method, the personalized characteristic is implemented by jumping back to the student node during the random walk process. The difficulty level is leveraged via jumping to the knowledge-point nodes that have the same difficulty level as the preliminary settings. The proposed method reaches the stable distribution state when the walking probability of every node does not change with time. The matrix form of the PTR method is shown as follows:

$$R_o^{(t)} = (1 - \alpha - q)MR_o^{(t-1)} + q \frac{s_e}{|s_e|} + \alpha r_o \quad (2)$$

where α and q are probabilities of jumping to the next node during the random walk, q determines the influence of difficulty levels, α indicates the influence of differences in students' mastery of knowledge-points. $R_o^{(t)} \in \mathbb{R}^{N \times 1}$ and $R_o^{(t-1)} \in \mathbb{R}^{N \times 1}$ are the arrival probability matrices of student o to all the other nodes in iteration t and iteration $t-1$, respectively. As defined in Eq. (3), each row ($I_i^{(t-1)}$) in $R_o^{(t-1)}$ represents the arrival probability from o to all the other nodes in the t -th walking step, that is, the importance of every node to student o . Note that, At the initial stage, the importance of all nodes to o is the same, and their sum is 1. Moreover, though authors can deduce the arrival probability from other student nodes to o , this work only focuses on the knowledge-point nodes, since this work's purpose is to discover their importance to each student. The symbols in Eq. (2) are defined as follows.

The arrival probability from node o to other nodes in $t-1$ -th step is defined as:

$$R_o^{(t-1)} = \left[I_1^{(t-1)}, I_2^{(t-1)}, \dots, I_i^{(t-1)}, \dots, I_N^{(t-1)} \right]^T \quad (3)$$

where $I_1^{(t-1)}$ denotes the arrival probability from o to node 1 in the $t-1$ -th step. The adjacency matrix M that reflects the outgoing and incoming degrees of every node in PCM is shown as:

$$M = \begin{bmatrix} \ell(p_1, p_1) & \dots & \ell(p_1, p_N) \\ \vdots & \ell(p_i, p_j) & \vdots \\ \ell(p_N, p_1) & \dots & \ell(p_N, p_N) \end{bmatrix} \quad (4)$$

where $\sum_{i=1}^N \ell(p_i, p_j) = 1$ if there is a directed edge from node j to i , otherwise, $\ell(p_i, p_j) = 0$.

The difficulty level of every node ($s_e \in \{0-1\}^{N \times 1}$) is defined as:

$$s_e = [X_1, X_2, \dots, X_i, \dots, X_N]^T \quad (5)$$

where e represents the difficulty level as {low-difficulty, normal-difficulty, high-difficulty}. In real-world applications, e is usually appointed by the teachers who want to know students' learning state in different learning levels. As shown in Eq. (5), if node i is a node with difficulty level e , then $X_i = 1$ in s_e . Otherwise, if node i is not a node with difficulty level e or node i is a student node, then $X_i = 0$. Compared with the Random Walk with Restart method (Tong et al., 2008), our method reflects the influence of the difficulty levels of knowledge-points on the walking results by jumping to the knowledge-points of corresponding difficulty levels with a certain probability.

The matrix of whether a node is a student node (r_o) is defined as:

$$r_o = [L_1, L_2, \dots, L_i, \dots, L_N]^T \quad (6)$$

In r_o , if node i is a student, then $L_i = 1$; otherwise, $L_i = 0$. Compared with the Topic-Sensitive PageRank method (Haveliwala, 2003), the authors add the student node as the starting node and have the chance to jump back to the student node during the walk, rather than selecting the start node randomly. By doing this, the authors can deduce the arrival probability of student o to other nodes from a personalized way, i.e., the personalized importance of knowledge-points.

In this work, the initial importance of all nodes is set to $1/N$. The arrival probability R is iterated through the PTR method until its obtained value converges. Each value in R_o represents the importance of each knowledge-point to the student o . The higher the value of R_o , the more important the knowledge-point is. The process of conducting random walks on PCM is shown as Algorithm 1:

Algorithm 1. PTR algorithm

```

Input:
PCM,  $e$ ,  $\alpha$ ,  $q$ 
Output:
 $R_1, R_2, \dots, R_o$ 
1:  while TRUE do
2:    for each student  $o$  do
3:       $t = 1, \quad R_o^{(t-1)} = \left[ \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right]$ 
4:      for  $t$  to 10000 do
5:         $R_o^{(t)} = (1 - \alpha - q)MR_o^{(t-1)} + q \frac{s_e}{|s_e|} + \alpha r_o$ 
6:         $d = R_o^{(t)} - R_o^{(t-1)}$ 
7:        if  $d < \varepsilon$  do
8:           $R_o = R_o^{(t)}$ 
9:        end for
10:     end for
11:  end while
12: return  $R_1, R_2, \dots, R_o$ 

```

EXPERIMENTAL RESULTS AND ANALYSIS

Dataset and the Evaluation Method

To evaluate the proposed model, the authors collect the records of the Python language course on an online course platform for students in a university, September–November 2020 and September–November 2021. As a result, a total of 290 students participate in this course, and their 21,460 learning-evaluation records are selected from the online mid-term examination to generate the importance of knowledge-points. Based on 10-fold cross-validation, the authors further divided students into 10 pieces at random for training and testing in each cross-validation. The training set is used to train the hyper-parameters of the proposed method and related baseline. The testing set is used to calculate the personalized importance of knowledge-points and verify the accuracy rate of the proposed method. The statistics of this dataset are shown in Table 2.

To evaluate the ranking accuracy of the PTR method, the authors considered how many efforts a student intended to spend on a knowledge-point, with the assumption that students will pay more attention to important knowledge-points. More specifically, given the ranking list of the knowledge-points for every student, authors first selected k knowledge-points (in experiments, we set k to 5) with the highest-ranking scores and k knowledge-points with lowest-ranking scores. Then, the authors computed the effort gap that a student has given them, that is, the differences in the number of learning behavior records on the two groups of knowledge-points. To ensure student efforts are meaningful rather than useless, this evaluation method chooses learning behaviors related to learning performance. In this work, as shown in Table 2, the authors selected 33,930 learning behavior records of students on knowledge-points, including publishing discussions, replying to discussions, and doing homework, because these three types of learning behavior are related to learning performance. If the number of learning behavior records on all the highest-ranking knowledge-points is higher than that on all the lowest-ranking knowledge-points, then this work registers a hit (i.e., the ranking accuracy plus one). In this work, the final ranking accuracy is defined as the average ranking accuracy for all the students.

Implementation Details

The implementation details of the proposed method are as follows: The experiment was conducted among 290 students in the data set. The authors set the hyper-parameter μ , which determines whether a student will link to a concept node, as 0.882. And the minimum confidence threshold c of the association rule is set to 0.5 during PCM generation. The authors searched the hyper-parameter α , which indicates within [0.1, 0.4], and searched the hyper-parameter q , which indicates within [0.1, 0.4]. After 10-fold cross-validation training, the optimal values of α and q are 0.3 and 0.2 respectively. Here, the purpose of computation during training is to obtain the best hyper-parameters, while the purpose of computation during testing is to verify the Ranking Accuracy of the proposed method. The training experiment for specific hyper-parameters α and q is in the section titled “Impact of Hyper-Parameters.

Table 2. The statistics of the dataset

Statistics	Python Language Course
Number of students	290
Number of test questions	74
Number of learning-evaluation records	21,460
Number of knowledge-points	39
Number of difficulty-level	3
Number of learning behavior records	33,930

Finally, according to the obtained hyper-parameters, the PTR walk was carried out from three different difficulty levels (e), low-difficulty, normal-difficulty, and high-difficulty, to obtain students' personalized importance of knowledge-points when different e is selected in the testing set. In addition, to obtain the confidence interval of performances, the authors set the confidence level to 95%.

It is worth noting that since the PCM is generated in advance, the PCM used at training time is the same as that used at test time. Specifically, the training stage is to walk from the student nodes of the training set on PCM and calculate the best values of hyper-parameters α and q , because α and q are unknown at first. The test phase is to walk from the student nodes of the test set on PCM, and use the hyper-parameters α and q calculated by the training stage to calculate the personalized importance of knowledge-points and test the Ranking Accuracy of the PTR method. Therefore, the personalized importance of knowledge-points generated by a single student will not change in the training stage or the testing stage.

BASELINES

The authors compare the following methods to evaluate the performance of the proposed method:

- **PageRank (PR):** This method is successfully used by the Google search engine to calculate the importance of web pages and rank them. The walking strategy in it is to walk along with the directed edge and jump to the random node with a certain probability. The result can reflect the graph structure. (Brin & Page, 1998)
- **Topic-Sensitive PageRank (TSR):** This adds topic relevance to PageRank. Its walking strategy is to walk according to the directed edge and jump to the node consistent with the topic with a certain probability. The results can not only reflect the structure of the graph but also reflect the differences in related topics. (Haveliwala, 2003)
- **Random Walk With Restart (RWR):** Based on the PageRank method, this adds the probability of jumping back to the starting node. Its strategy is to walk from a starting node according to the directed edge and jump back to the starting node with a certain probability. The result can reflect not only the structure of the graph but also the differences of different starting nodes. (Tong et al., 2008)

As far as the authors know, authors are the first to study the concept ranking problem on learning-evaluation data. Hence, the authors mainly compare with the baselines that use different walking strategies. As comparative baselines, the hyper-parameters of the three traditional random walk methods are also trained according to the section titled "Impact of Hyper-Parameters." It is worth noting that PageRank and random walk with restart cannot reflect different difficulty levels in their algorithm structures, they are discussed regardless of difficulty levels and evaluated only according to learning behaviors. The comparison results can be seen in the Experimental Results section.

EXPERIMENTAL RESULTS

The experimental results with different evaluation behaviors are shown in Tables 3–5, from which the authors have made the following observations.

First, the ranking accuracy of PTR in different evaluation behaviors and different difficulty levels is better than all baselines, proving the advantage of the proposed method to solve the personalized importance of knowledge-points. Among them, PTR's ranking accuracy increases by about 6.0% to 19.9% compared with PR, 1.5% to 11.5% compared with RWR, and 2.5% to 17.1% compared with TSR.

Table 3. The comparison results with doing homework as the evaluation behavior

Method	Ranking Accuracy								
	Low-Difficulty			Normal-Difficulty			High-Difficulty		
	Mean	Var	SD	Mean	Var	SD	Mean	Var	SD
PR				0.775±0.036	0.002	0.048			
RWR				0.829±0.040	0.003	0.052			
TSR	0.814±0.044	0.004	0.059	0.786±0.032	0.002	0.042	0.832±0.032	0.002	0.042
PTR	0.861±0.043	0.003	0.056	0.918±0.047	0.004	0.062	0.871±0.036	0.002	0.048

Note. *Mean*=mean of Ranking Accuracy at 95% confidence level, *Var*=variance, *SD*=standard deviation.

Table 4. The comparison results with replying in discussions as the evaluation behavior

Method	Ranking Accuracy								
	Low-Difficulty			Normal-Difficulty			High-Difficulty		
	Mean	Var	SD	Mean	Var	SD	Mean	Var	SD
PR				0.719±0.040	0.003	0.054			
RWR				0.803±0.034	0.002	0.046			
TSR	0.857±0.042	0.003	0.055	0.661±0.025	0.001	0.033	0.793±0.050	0.004	0.065
PTR	0.918±0.030	0.002	0.039	0.832±0.040	0.003	0.053	0.818±0.037	0.002	0.049

Note. *Mean*=mean of Ranking Accuracy at 95% confidence level, *Var*=variance, *SD*=standard deviation.

Table 5. The comparison results with publishing discussions as the evaluation behavior

Method	Ranking Accuracy								
	Low-Difficulty			Normal-Difficulty			High-Difficulty		
	Mean	Var	SD	Mean	Var	SD	Mean	Var	SD
PR				0.861±0.041	0.003	0.054			
RWR				0.918±0.042	0.003	0.055			
TSR	0.876±0.030	0.002	0.043	0.904±0.049	0.004	0.064	0.917±0.031	0.002	0.042
PTR	0.921±0.041	0.003	0.057	0.946±0.025	0.001	0.033	0.961±0.023	0.001	0.030

Note. *Mean*=mean of Ranking Accuracy at 95% confidence level, *Var*=variance, *SD*=standard deviation.

Second, when taking the publishing discussions as the evaluation behavior, the ranking accuracy of PTR is significantly higher than that of the other two learning behaviors. This is because the publishing discussion behavior can better reflect student attention to knowledge-points.

Third, when taking *doing homework* as the evaluation behavior, the authors choose the normal-difficulty, then the ranking accuracy of PTR is the highest. When replying to discussions as the evaluation behavior, the authors choose the low-difficulty, then the ranking accuracy of PTR is the highest. When publishing discussions as the evaluation behavior, the authors choose high-difficulty, then the Ranking Accuracy of PTR is the highest. This proves that the ranking accuracy of PTR is balanced at three difficulty levels.

Finally, in the baselines, except for a few cases, the ranking accuracy of RWR is higher than that of TSR in most cases. In comparison, the ranking accuracy of TSR is usually higher than that of PR. This is because in solving the personalized importance of knowledge-points, differences in student mastery play a more significant role than the difficulty differences of knowledge-points. The difficulty differences of knowledge-points play a more substantial role than only graph structure.

MODEL ANALYSIS

To explore the importance of introducing the student nodes and difficulty levels in the proposed model, the authors compare PTR with its three variants:

- **PTR-SD:** This is a simplified version of PTR that removes the student nodes and the difficulty levels of knowledge-points. Specifically, it only uses a basic knowledge concept map, without considering the differences in students' mastery and the difficulty level of knowledge-points. Then, the initial node of the random walk is the randomly selected knowledge node and can jump to any other node with probability q . This is to study the effectiveness of these two components.
- **PTR-S:** This is another variant of PTR that removes the student nodes, that is, excluding the differences in students' mastery, and utilizes PTR whose start-node is the randomly selected knowledge-point rather than students. This is to validate the importance of student mastery when PTR conducts random walking.
- **PTR-D:** This variant removes the difficulty level of knowledge-points from PTR, that is, the authors deem all the nodes have the same difficulty and jump to any other node with probability q when walking. This is to evaluate the effect of the difficulty level in discovering the importance of knowledge-points.

For convenience, the authors take the average ranking accuracy under the three difficulty levels as the ranking accuracy results, and the training and testing method is still 10-fold cross-validation. The experimental results are reported in Table 6, from which the authors have the following observations. First, PTR significantly outperforms PTR-S, PTR-D, and PTR-SD on the dataset used, indicating the importance of the student nodes and difficulty levels, and considering only one of them or none of them cannot get better results than combining them together. Second, PTR-D performs better than PTR-S, demonstrating the benefit of introducing the student node is better than considering the difficulty levels. Third, the authors also notice that PTR-D and PTR-S perform better than PTR-SD, demonstrating the effectiveness of the two components and the importance of exploiting student nodes to reflect student mastery as well as the benefit of exploiting difficulty levels to reflect the difficulty levels of knowledge-points.

Table 6. Importance of Components of PTR

Model	Ranking Accuracy								
	Doing Homework			Replying to Discussions			Publishing Discussions		
	Mean	Var	SD	Mean	Var	SD	Mean	Var	SD
PTR-SD	0.695±0.019	0.003	0.052	0.638±0.027	0.005	0.071	0.758±0.019	0.002	0.049
PTR-S	0.787±0.034	0.008	0.089	0.727±0.022	0.003	0.059	0.804±0.018	0.002	0.049
PTR-D	0.813±0.023	0.004	0.059	0.757±0.018	0.002	0.048	0.862±0.019	0.002	0.049
PTR	0.901±0.021	0.003	0.056	0.855±0.024	0.004	0.062	0.947±0.014	0.002	0.038

Note. *Mean*=mean of Ranking Accuracy at 95% confidence level, *Var*=variance, *SD*=standard deviation.

DISCUSSIONS

Besides the above experimental results analysis and model analysis, the authors also analyzed experimental results from the perspective of students. The authors select the low-difficulty level as an example to analyze the knowledge-points of 290 students in the data set. The analysis is as follows.

Firstly, the results are personalized for different students. Table 7 shows the partial results calculated by our method. The value of the results indicates student cognition of the importance of the knowledge-points. The larger the value, the more important the student thinks the knowledge-point is. Due to space constraints, the calculation results of the importance of four knowledge-points by three students are randomly listed in Table 7. As shown in Table 7, student 87 thinks that *insertion of the list* is more important than the other three knowledge-points. Both student 87 and student 144 think *insertion of the list* is more important than the other three knowledge-points, while student 88 thinks it is not important relative to the other three knowledge-points. Teachers can specify personalized learning strategies for each student according to their personalized cognition of the importance of knowledge-points.

Secondly, the findings of the results can reflect most of the student cognition to the knowledge-points. For example, in the results, the authors find that 233 students think that *for loop statement* is in the top five important knowledge-points, 206 students think that *insertion of the list* is in the top five, and 258 students think that *variable assignment* is in the bottom five. Teachers can diagnose the collective cognitive deviation of students and adjust teaching strategies according to most of the student cognition.

Finally, The authors can also find the differences in student cognition to the importance of knowledge-points. Among the 218 students who think *insertion of the list* belongs to the top five important knowledge-points, only nine students think *insert dictionary* also belongs to the top five, and only four student thinks *arithmetic operator* belongs to the last five. Teachers can analyze the differences in student cognition of the importance of knowledge-points and reflect on the teaching process, to give academic warning to some students.

IMPACT OF HYPER-PARAMETERS

The hyper-parameters α and q represent the probabilities of jumping to the next node during the random walk. Among them, α indicates the influence of differences in student mastery of knowledge-points, and q determines the influence of differences in knowledge-point difficulty levels. A higher value of α indicates a greater probability of the proposed method jumping to the initial student node when walking, and the more it can reflect the difference in student mastery of knowledge-points. A

Table 7. Personalized importance of some knowledge-points

Students	Personalized importance of knowledge-points				
	Escape Character	Combining Strings	Insertion of the List	...	For Loop Statement
⋮	⋮	⋮	⋮	⋮	⋮
Student 87	0.0131	0.0126	0.0731	...	0.0079
Student 88	0.0051	0.0062	0.0034	...	0.0262
⋮	⋮	⋮	⋮	⋮	⋮
Student 144	0.0087	0.0006	0.0216	...	0.0069
⋮	⋮	⋮	⋮	⋮	⋮

higher value of q indicates a greater probability of jumping to the knowledge-points of the corresponding difficulty level during walking, and the more it can reflect the difference in knowledge-point difficulty levels.

Table 8 presents the experimental results. The reported results are the average ranking accuracy of the training set under 10-fold cross-validation for three evaluation behaviors and three difficulty levels. The analysis of the experimental results is as follows: The authors found that the best value of hyper-parameters: α is 0.3 and q is 0.2. Within the range [0.1,0.4], the performance of PTR first increases and then decreases with the increasing of α and q . When α and q exceed a certain threshold, the performance of PTR will even deteriorate due to considering more jump probability and ignoring the role of knowledge structure in the graph. For the best performance, the authors set $\alpha = 0.3$ and $q = 0.2$ for PTR. In addition, the Hyper-parameters of the related baseline are also trained according to the above method.

CONCLUSION

In this paper, the authors study the knowledge-point importance discovery problem due to the following issues. First, traditional knowledge-point importance discrimination model is challenging to apply to large-scale online courses due to the high degree of teacher participation. Second, standard random walk methods cannot reflect students' individual understanding of different knowledge-points. To overcome these challenges, the authors propose a random walk-based discriminant model for personalized importance discovery. Specifically, this approach uses student learning-evaluation data adaptively to generate a knowledge concept map PCM and then, based on it, calculates the importance of the knowledge-points by developing a novel random walk method. The experimental results on a real-world dataset show that the solution performs better than other related importance judging methods, demonstrating the proposed method's effectiveness in personalized knowledge-point importance discovery and the feasibility of leveraging the learning-evaluation data.

One limitation of this work is that this article evaluates the proposed method only via considering the relevance between the learning behaviors and the learning results, rather than the causality between

Table 8. Impact of hyper-parameter α and q

Hyper-parameters	$q = 0.1$		$q = 0.2$		$q = 0.3$		$q = 0.4$	
	Mean		Mean		Mean		Mean	
	Var	SD	Var	SD	Var	SD	Var	SD
$\alpha = 0.1$	0.813±0.013		0.816±0.015		0.801±0.014		0.705±0.013	
	0.004	0.061	0.005	0.072	0.004	0.063	0.004	0.062
$\alpha = 0.2$	0.847±0.014		0.856±0.016		0.823±0.012		0.714±0.011	
	0.004	0.065	0.005	0.072	0.003	0.055	0.003	0.053
$\alpha = 0.3$	0.875±0.008		0.892±0.010		0.842±0.013		0.765±0.017	
	0.001	0.038	0.003	0.051	0.004	0.060	0.006	0.078
$\alpha = 0.4$	0.839±0.014		0.843±0.012		0.802±0.013		0.708±0.010	
	0.005	0.070	0.003	0.056	0.004	0.061	0.002	0.048

Note. Mean=mean of Ranking Accuracy at 95% confidence level, Var=variance, SD=standard deviation.

them, resulting in inaccurate and incomplete evaluation results. Therefore, the authors' future work will focus on the utility of causal inference in model evaluation and its application in studying student learning behaviors to further improve model performance.

FUNDING AGENCY

The Open Access Processing fee for this article has been covered in full by the authors.

REFERENCES

- An, N., Teng, G. Q., Bai, S. C., & Han, S. X. (2018). Domain Knowledge Clustering from the Perspective of Network Centrality. *Information Studies: Theory & Application*, 41(8), 120–126.
- Baghestani, T. A., Sanatjoo, A., Behzadi, H., & Jamali, M. H. R. (2021). Use of mind mapping in search process to clarify information needs and improve search satisfaction. *Journal of Information Science*. Advance online publication. doi:10.1177/01655515211058041
- Brin, S., & Page, L. (1998). The Anatomy of a Large-Scale Hypertextual Web Search Engine. *Computer Networks and ISDN Systems*, 30(1-7), 107-117.
- Chen, S. M., & Sue, P. J. (2013). Constructing concept maps for adaptive learning systems based on data mining techniques. *Expert Systems with Applications*, 40(7), 2746–2755. doi:10.1016/j.eswa.2012.11.018
- Cheng, S., Liu, Q., Chen, E., Huang, Z., Huang, Z., Chen, Y., Ma, H., & Hu, G. (2019). DIRT: Deep Learning Enhanced Item Response Theory for Cognitive Diagnosis. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM'19)* (pp. 2397-2400). Association for Computing Machinery. doi:10.1145/3357384.3358070
- Chu, H. C., & Hung, C. M. (2015). Effects of the digital game-development approach on elementary school students' learning motivation, problem solving, and learning achievement. *International Journal of Distance Education Technologies*, 13(1), 87–102. doi:10.4018/ijdet.2015010105
- Chu, H. C., Hwang, G. J., & Huang, Y. M. (2010). An enhanced learning diagnosis model based on concept-effect relationships with multiple knowledge levels. *Innovations in Education and Teaching International*, 47(1), 53–67. doi:10.1080/14703290903525846
- Fatawi, I., Degeng, I. N. S., Setyosari, P., Ulfa, S., & Hirashima, T. (2020). Effect of online-based concept map on student engagement and learning outcome. *International Journal of Distance Education Technologies*, 18(3), 42–56. doi:10.4018/IJDET.2020070103
- Fernández-Álvarez, D., Frey, J., Labra Gayo, J. E., Gayo-Avello, D., & Hellmann, S. (2021). Approaches to measure class importance in Knowledge Graphs. *PLoS One*, 16(6), e0252862. Advance online publication. doi:10.1371/journal.pone.0252862 PMID:34111187
- Hafeez, M. (2021). Assessment of academic achievements and memorisation power of students by concept map strategy and traditional lecture approach. *International Journal of Learning and Teaching*, 13(3), 133–145. doi:10.18844/ijlt.v13i3.5922
- Haveliwala, T. H. (2003). Topic-sensitive PageRank: A context-sensitive ranking algorithm for Web search. *IEEE Transactions on Knowledge and Data Engineering*, 15(4), 784–796. doi:10.1109/TKDE.2003.1208999
- Jin, C., Zhou, Y., Ying, S., Zhang, C., & Wu, M. (2020). A Knowledge-Fusion Ranking System with an Attention Network for Making Assignment Recommendations. *Computational Intelligence and Neuroscience*, 3, 1–11. doi:10.1155/2020/6748430 PMID:33424959
- Kardan, A., & Razavi, N. (2014). Learner knowledge level calculation by concept map and concept weight estimation using neural networks. In *2014 4th International Conference on Computer and Knowledge Engineering (ICCCKE)* (pp. 62-67). IEEE. doi:10.1109/ICCCKE.2014.6993394
- Kim, H. J., Baek, J. W., & Chung, K. (2020). Optimization of Associative Knowledge Graph Using TF-IDF Based Ranking Score. *Applied Sciences (Basel, Switzerland)*, 10(13), 4590–4611. doi:10.3390/app10134590
- Koponen, I. T., & Nousiainen, M. (2014). Concept networks in learning: Finding key concepts in learners' representations of the interlinked structure of scientific knowledge. *Journal of Complex Networks*, 2(2), 187–202. doi:10.1093/comnet/cnu003
- Leake, D., Maguitman, A., & Reichherzer, T. (2004). Understanding knowledge models: Modeling assessment of concept importance in concept maps. In *Proceedings of the Annual Meeting of the Cognitive Science Society (COGSCI)* (pp. 795-800). California Digital Library.
- Lee, C. H., Lee, G. G., & Leu, Y. (2009). Application of automatically constructed concept map of learning to conceptual diagnosis of e-learning. *Expert Systems with Applications*, 36(2), 1675–1684. doi:10.1016/j.eswa.2007.11.049

- Li, M. J., Chen, S. H., Wang, Y. H., & Qiu, X. H. (2019). A Study on the Wisdom Point-oriented Teaching Method under the Background of Emerging Engineering Education. *Research in Higher Education of Engineering*, 2, 183–187.
- Li, Y., Shao, Z., Wang, X., Zhao, X., & Guo, Y. (2018). A Concept Map Based Learning Paths Automatic Generation Algorithm for Adaptive Learning Systems. *IEEE Access : Practical Innovations, Open Solutions*, 7, 245–255. doi:10.1109/ACCESS.2018.2885339
- Ma, Z., & Shi, L. (2016). Application of Visual Interactive Concept Map in College English Writing Teaching. *International Journal of Emerging Technologies in Learning*, 11(11), 32–36. doi:10.3991/ijet.v11i11.6248
- Makhambetova, A., Zhiyenbayeva, N., & Ergesheva, E. (2021). Personalized Learning Strategy as a Tool to Improve Academic Performance and Motivation of Students. *International Journal of Web-Based Learning and Teaching Technologies*, 16(6), 1–17. doi:10.4018/IJWLTT.286743
- Mou, Z. J., & Li, Y. T. (2019). Research on Personalized Evaluation Modeling and Tool Design Based on Learning Evaluation Data. *e- Education Research*, 40(8), 96–104.
- Pinandito, A., Prasetya, D. D., Az-zahra, H. M., Wardhono, W. S., Hayashi, Y., & Hirashima, T. (2021). Design and Development of Online Collaborative Learning Platform of Kit-Build Concept Map. *Journal of Information Technology and Computer Science*, 6(1), 50–65. doi:10.25126/jitecs.202161294
- Ren, M., Huang, H., & Gao, Y. (2021). SKR-QA: Semantic Ranking and Knowledge Revise for Multi-Choice Question Answering. *Neurocomputing*, 459, 142–151. doi:10.1016/j.neucom.2021.06.076
- Tong, H., Faloutsos, C., & Pan, J. Y. (2008). Random walk with restart: Fast solutions and applications. *Knowledge and Information Systems*, 14(3), 327–346. doi:10.1007/s10115-007-0094-2
- Wang, C. C., Zhang, R., & Wu, T. G. (2020). The Construction of College Physics Knowledge Map and Its Application in Individualized Teaching. *Physics and Engineering*, 30(5), 77–82.
- Wang, C. L., Li, D. G., Zhang, Y. N., Xiao, Y., & Chen, S. Q. (2021). Research on Learners' Preferences for Gamification Elements in Online Learning. *e- Education Research*, 42(4), 68–75.
- Wang, F., Liu, Q., Chen, E., Huang, Z., Chen, Y., Yin, Y., & Wang, S. (2020). Neural Cognitive Diagnosis for Intelligent Education Systems. *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2020)*, 34(4), 6153-6161. doi:10.1609/aaai.v34i04.6080
- Wang, J. X., Yu, Z., Mou, Z. J., & Li, Y. T. (2019). Knowledge-based Learning Early Warning Modeling and Simulation Based on Learning and Evaluation Big Data. *Modern Distance Education*, 4, 28–37.
- Wu, G., Zhang, K., Li, J. Z., & Wang, K. H. (2007). Ranking by Mutually Reinforcing Concepts and Relations in Ontology. *Chinese Journal of Computers*, 30(9), 1490–1499.
- Yao, G. P. (2017). Experience scientific inquiry and construct important concepts. *Biology Teaching*, 42(11), 19–20.
- Zhan, P. D., Tian, Y. S., Yu, Z. H., Li, F. M., & Wang, L. J. (2020). A Comparative Study of Probabilistic Logic and Fuzzy Logic in Refined Learning Diagnosis. *Journal of Psychological Science*, 43(5), 1258–1266.
- Zhong, H., Ning, Z., Li, G., & Li, Z. (2021). A method of core concept extraction based on semantic-weight ranking. *Concurrency and Computation*, e6504.

Hongfei Guo is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University, China. His research interests include educational data mining and learning analysis.

Xiaomei Yu received her Master Degree and Doctoral Degree in 2004 and 2016 respectively from Shandong Normal University. She is currently an associate professor in School of Information Science and Engineering, Shandong Normal University, China. Her current research interests include computational intelligence, deep learning and educational data mining.

Xinhua Wang received the M.S. degree in applied mathematics from the Dalian University of Technology, China, in 2001, and the Ph.D. degree in management science and engineering from Shandong Normal University, China, in 2008. He was a Visiting Scholar with Shandong University, from 2002 to 2003, and a Senior Visiting Scholar with Peking University, from 2008 to 2009. He is currently a Professor and a Master Supervisor with the School of Information Science and Engineering, Shandong Normal University. His research interests include distributed networks and recommender systems. He is also a member of the Social Media Processing Committee.

Lei Guo received his Ph.D. degree in computer architecture from Shandong University, China, in 2015. He was a Research Intern with Microsoft Soft Research Asia (MSRA) in 2014 and a Visiting Scholar with The University of Queensland (UQ), Australia, from 2018 to 2019. He is currently an Associate Professor and a Master Supervisor with Shandong Normal University, China. His research grants include the National Natural Science Foundation of China and the China Postdoctoral Science Foundation. His research interests include information retrieval, social networks, and recommender systems. He is also a Director of Shandong Artificial Intelligence Society and a member of the Social Media Processing Committee of the Chinese Information Society and the Program Committees in IJCAI, CIKM, YSSNLP, DASSFA, etc.

Liancheng Xu received his PhD degree in Operational Research and Cybernetics from Dalian University of Technology in 2010. Since 1996, he has been an assistant at Shandong Normal University, and currently an associate professor. He is also a member of the China Computer Federation. His current research is mainly engaged in recommender system and user modeling, information retrieval, big data in education, and other fields.

Ran Lu received the master's degree from the Dalian University of Technology, in 2001. He is currently the Director of the Computer Teaching Department and an Associate Professor and a Master's Supervisor with the School of information science and engineering, Shandong Normal University, China. His research interests include network information processing, information security, and vocational technology education.