Received October 11, 2018, accepted November 28, 2018, date of publication December 7, 2018,
date of current version January 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2885339

A Concept Map-Based Learning Paths Automatic Generation Algorithm for Adaptive Learning Systems

YANCONG LI[1], ZENGZHEN SHAO[1,2], XIAO WANG[1], XUECHEN ZHAO[2], AND YANHUI GUO[2]

[1]School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China
[2]School of Data and Computer Science, Shandong Women's University, Jinan 250002, China

Corresponding author: Zengzhen Shao (shaozengzhen@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 71672154, in part by the Humanities and Social Sciences Foundation of the Ministry of Education of China under Grant 16YJA630038, in part by the Joint Research Project between Southwest Jiaotong University and Cornell University under Grant 268SWJTU15WTD01, in part by the China Postdoctoral Science Foundation under Grant 2016M592697, in part by the Key Science and Technology Project of Shandong Province of China under Grant 2014GGH201022, and in part by the Chang Jiang Scholars Program of China.

ABSTRACT A concept map is an important knowledge visualization tool in adaptive learning systems. The weak concepts of students can be identified by analyzing the concept maps to provide learning guidance and teaching suggestions to students and teachers. However, in the current researches, it is difficult to analyze effective information from the concept maps when the number of concepts is large and the associations between concepts are complicated. It rarely reflects the learning performance of different student groups, which do not reflect the characteristics of the adaptive learning systems. A learning path automatic generation algorithm, Learning paths generation (LPG) algorithm, for adaptive learning systems is proposed. The LPG algorithm fully considers the learning performance of different student groups. The concept maps with different learning features are generated by the clustering algorithm and association rules mining, and several simplified learning paths are generated by using the topological sorting algorithm. Experiments show that the LPG algorithm has a good discrimination for the student groups and can generate different learning paths according to the students' learning performance.

INDEX TERMS Concept map, adaptive learning system, learning path, automatic generation.

I. INTRODUCTION

Adaptive learning system is widely studied in the field of E-learning and intelligent education [1]. It is a learning system that can be adaptively adjusted according to the results of student modeling [2]. As an effective knowledge visualization tool in adaptive learning systems, the concept map has become a hot topic in current research [3]. The concept map can express concepts in a structured form, which can not only promote meaningful learning for students [4], but also can make adaptive adjustment of the conceptual structure based on the learning performance of students [5].

Researchers have conducted a lot of research on concept maps and many methods derived from concept maps have been proposed. Lee *et al.* [6] used association rules mining algorithm to analyze the association rules between concepts, and automatically generated concept maps. A Remedial-Instruction Path (RIP) was established for students based on these automatically generated concept maps. However, the instruction path they established was for all those students who were analyzed by them. Atapattu *et al.* [7] proposed a Natural Language Processing (NLP) algorithm that uses natural language processing techniques to generate concept maps from slides as an active alternative method for experts to manually generate concept maps. However, the generated concept map is static and does not reflect students' learning performance. Huang *et al.* [8] used the classification method to classify students into groups and generate concept maps based on the association rules method. Generated concept maps reflect the overall performance of students and do not have the ability to distinguish students. Acharya *et al.* [9] proposed a diagnostic learning method based on concept maps, using direct hash and pruning algorithms to construct concept maps. Redundancies in the concept map are deleted to generate a learning sequence. Based on this learning sequence, a prototype learning system was developed using the Android simulator. This learning sequence does not reflect the mastery of different concepts in different groups of students. In general, researchers have done a lot of researches based on concept maps and have achieved many achievements. However, these studies also have some limitations. These studies failed to distinguish and analyze students based on their data of learning behaviors. All students are guided by a unified analysis and cannot reflect the characteristics of the adaptive learning system. Further analysis of concept maps relies mainly on the labor force. When the number of association rules between concepts increase, it is difficult to extract valid information from the concept maps.

In this paper, a concept map-based learning path automatic generation algorithm LPG (Learning Paths Generation) algorithm for adaptive learning systems is proposed. The LPG algorithm aims at the above limitations. It is based on the level of mastery of students in the concepts of computer science, using clustering and association rules mining algorithms to generate concept maps with different students' learning features, and using topological sorting algorithm to automatically generate learning paths, which is accordance with the characteristics of adaptive learning systems.

The main contributions of this study are stated as follows:

(i) Several concept maps are generated based on the data of students' learning performance. This study can overcome the limitations of existing researches that do not distinguish students based on their learning behaviors.

(ii) Students are clustered into groups with different learning features, and each group of students is analyzed separately. Students after grouping can be guided in different ways. The lengths of time that students need to learn concepts after being grouped are different, reflecting the adaptability of LPG algorithm.

(iii) Topological sorting algorithm is applied to the analysis of concept maps to automatically generate several different learning paths. The generation of learning paths does not require labor force, which saves labor costs. Teachers can arrange teaching sequence and teaching duration according to learning paths, and students in each group are guided by different teaching methods.

The remaining sections of this paper are organized as follows. Related literature is reviewed in Section 2. The basic ideas and steps of the LPG are discussed in Section3. The computational experiments and analysis are conducted in Section 4. And we conclude our results and point out the future research directions in Section 5.

II. RELATED WORK

Novak and Gowin [10] from Cornell University proposed the concept map for the first time, using nodes to represent concepts and using directed edges to represent association rules between concepts. Concept maps are widely used in adaptive learning systems [11]. Adaptive learning system is a system that takes into account students' learning performance and emphasizes personalized learning [1]. As an intuitive knowledge visualization tool, concept maps can be adaptively adjusted based on students' learning performance in adaptive learning systems [12]. Valuable information can be mined by analyzing concept maps [13], [14].

Since the concept map was proposed, researchers have carried out a lot of researches on the generation and application of concept maps. Chen *et al.* [15] generated concept maps for e-learning from academic articles. They used e-learning journal articles and conference papers as data sources, and used text mining techniques to automatically generate concept maps. Qasim *et al.* [16] proposed a cluster-based method for generating concept maps from text documents. They have used a variety of techniques to process text documents, such as natural language processing, clustering, information retrieval, and structural similarity measuring. The generated concept maps can assist in teaching. Lee *et al.* [17] proposed a method for automatically generating a concept map from a single document using the bursts of words, and the method was verified by them. Wang *et al.* [18] proposed a joint optimization model for generating concept maps from textbooks. The model generates a concept map using the relationships between conceptual keywords implicit in the textbook. Lai *et al.* [19] proposed a system based on information retrieval technologies, which can extract keywords from each section of the book to generate keywords concept maps about this book. Santos [20] has classified the concepts based on natural language processing and machine learning techniques from the abstract, and analyzes the relationships between concepts to automatically generate concept maps. These studies use text analysis, natural language processing, machine learning, and other techniques to analyze unstructured texts and automatically generate concept maps. The generated concept maps can provide teachers with a useful reference, design teaching materials for teachers, and help students understand the structure of concepts. These studies extract key items from the text and generate concept maps. However, the generated concept maps do not analyze students' learning behaviors, and do not reflect students' mastery of concepts, which make it difficult to guide students in accordance with their aptitude.

There are also many papers that consider students' learning behaviors, generate and analyze concept maps. Lee *et al.* [6] proposed a system for generating concept maps using the Apriori algorithm and using concept maps to diagnose teaching obstacles. The system has taken into account the students' learning performance and has given students a unified guidance. Bai *et al.* [21] proposed an automatic construction method of concept map based on fuzzy rules [22] and student answer records. Using fuzzy rules and fuzzy inference techniques, the concept map is automatically constructed and the degrees of associations between concepts are evaluated. The generated concept map indicates that a concept should be learned before another concept. Chen *et al.* [23] proposed an automatic generation of concept maps for adaptive learning systems based on data mining techniques. Their method can dynamically generate the concept map based on students' answer records, which have certain reference significance.

Huang *et al.* [8] used the association rules method to analyze all the test records of students to automatically generate a concept map. The generated concept map reflects the learning performance of all students. Schroeder *et al.* [24] studied the effects of learning using concept maps. They applied a concept map to all students and verified that the concept map has a continuing advantage in learning. Romero *et al.* [25] proposed a learning strategy based on concept maps. The strategy creates a concept map by students and compares it with the concept map created by teachers to discover the weaknesses in the students' grasp of concepts. The strategy focuses on students' learning experience and is subjective. These studies have achieved great success and are consistent with the characteristics of adaptive learning systems. And they have certain guiding significance for teaching and learning. However, these studies also have limitations. Generated concept maps reflect all students' learning performance. All students are guided by the same learning path and learning duration. Further analysis of concept maps usually requires the assistance of the labor force. It is time consuming and is not conducive to quickly extracting information from concept maps.

In general, researches on the concept map were summarized as two types. The first type of researches utilize the text analysis [26] technology to automatically generate the concept map. However, generated concept maps lack further analysis and do not reflect students' learning performance, nor the characteristics of adaptive learning systems. The second type of researches have considered students' learning performance. However, generated concept maps reflect the features of all students, and cannot distinguish students based on their mastery of concepts. The diagnoses are also applicable to all students and do not have the ability to distinguish students. And the analysis of concept maps is time consuming due to the assistance of the labor force. In this study, the clustering algorithm is used to divide students into groups according to the mastery of concepts for students, and combined with the association rules mining method to generate several concept maps. Each concept map corresponds to a group of students' learning features, which can effectively distinguish students. Using the topological sorting algorithm, concept maps are analyzed separately, and learning paths with different student learning features are automatically generated. The learning sequence and learning duration of each learning path are different, which can provide teachers with teaching suggestions and provide students with learning guidance.

III. LPG LEARNING PATH AUTOMATIC GENERATION ALGORITHM

The LPG (Learning Paths Generation) algorithm uses students' test records, combined with the data of relationships between questions and concepts, to transform the students' mastery of questions into the students' mastery of concepts, and students' learning features can be extracted. The clustering method is used to divide students into different groups based on the learning features of students, and then generate several different concept maps through the association rules mining method. Based on these several concept maps, the topological sorting algorithm is used to analyze the associations between concepts, and different concept maps can generate different learning paths. These learning paths are simplified based on the level of mastery of each concept for each group of students. Different simplified learning paths that match students' learning situations can be generated. Students can learn concepts in a targeted manner based on these simplified learning paths. The flow chart of the LPG algorithm is shown in Figure 1. In addition, we use the notations in TABLE 1 throughout this paper.

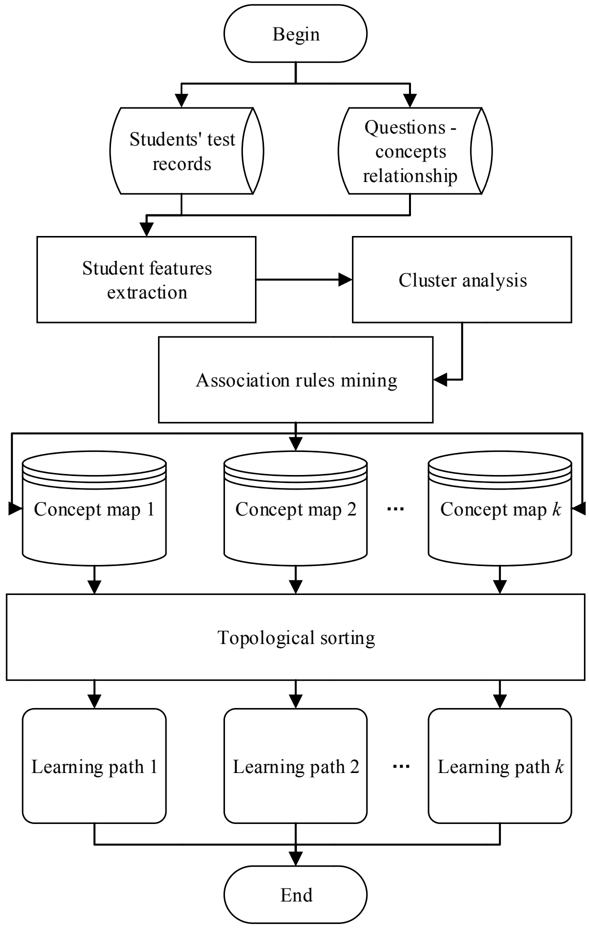


FIGURE 1. The flow chart of the LPG algorithm.

A. CONCEPT MAPS WITH STUDENT FEATURES GENERATION

Based on the error rate of each student's answer to each concept in an exam, the student features are extracted and will be transformed into discrete forms. Student features reflect each student's mastery of each concept. Using the clustering technique in data mining, students can be clustered into several groups, combined with association rules mining method, several concept maps are generated. Each concept

TABLE 1. Notations.

Symbol	Meaning	Illustration
Q	Test questions	
QC	Relationships between questions and concepts	
R	Students' test records	
F	Student features	$F = \{f_1, f_2, \cdots, f_x, \cdots, f_n\}$
f_x	The x-th student feature in F	
n	The dimension of the space vector of student features	
F′	New student features	$F' = \{f'_1, f'_2, \cdots, f'_{x'} \cdots, f'_n\}$
f'_x	The x-th new student feature in F′	
G	Student groups	$G = \{G_1, G_2, \cdots, G_i, \cdots, G_k\}$
G_i	The i-th group in G	
k	Number of clustered groups	
M	k concept maps	$M = \{M_1, M_2, \cdots, M_i, \cdots, M_k\}$
M_i	The i-th concept map in M	
c	A concept	
Diff_{ij}	The differentiation degree between the i-th group and the j-th group	
E_{ci}	The error rate of the c-th concept for students in the i-th group	$E_{ci} \in [0,1]$
T_{ci}	The duration of the concept c which students in group G_i need to learn	
T_i	The duration for students in group G to learn all concepts	

map reflects the learning performance of the corresponding student group.

1) STUDENT FEATURES EXTRACTION AND FORM TRANSFORMATION

Before extracting student features, two data sets need to be introduced. The first data set is the relationships between questions and concepts. The relationships between questions and concepts indicates the concepts that questions belong to in a test paper. When a question corresponds to multiple concepts, if a student answers incorrectly in this question, it is difficult to determine which concept the student has insufficiently mastered. Therefore, in this algorithm, one question can only belong to one concept, but one concept may contain many questions. For convenience, relationships between questions and concepts are expressed as QC. QC is usually given by experts in related fields [8], [9], [23]. There are also papers that use text analysis algorithms to automatically classify test questions into concepts [27]. To ensure the correctness of this data set, in this paper, QC is given by experts in related fields. Another data set is students' test records. Students' test records indicate whether students answered correctly on each question. It is expressed as R. In the LPG algorithm, the two data sets will be used to generate learning paths.

Based on QC and R, students answering questions correctly or incorrectly can be translated into error rates on each concept for students. Student features can be extracted,

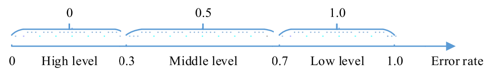


FIGURE 2. Student features transformation description.

expressed as $F = \{f_1, f_2, \cdots, f_x, \cdots, f_n\}$, where n is the number of concepts. f_x represents the x-th feature, which is the error rate of a student's answer on concept x, $f_x \in [0, 1]$. The larger the value of f_x, the worse the student's mastery of concept x.

Based on students' error rates in each concept, student features F can be transformed into the level of mastery of each concept, expressed as three forms: high level, medium level, and low level. The new student features are expressed as $F' = \{f'_1, f'_2, \cdots, f'_x, \cdots, f'_n\}$, where $f'_x = \{0, 0.5, 1.0\}$. As shown in Figure 2, if $f_x \in [0, 0.3]$, the student is considered to be at a high level on the concept x, let $f'_x = 0$; if $f_x \in [0.7, 1.0]$, the student is considered to be at a low level on the concept x, let $f'_x = 1.0$; if $f_x \in (0.3, 0.7)$, the student is considered to be at a medium level on the concept x, let $f'_x = 0.5$.

Each new student feature F′ constitutes the features of all students, and each F′ is an n-dimensional space vector. In the next step, the clustering algorithm will be used to analyze new student features.

2) STUDENT FEATURES CLUSTERING

Clustering [28] is an unsupervised machine learning method. Based on student features, students can be divided into several different groups, so that students in the same group are as similar as possible, that is, students in the same group are as similar as possible in mastering concepts. K-Means [29] is a classic algorithm in clustering methods. The number of clusters of the K-Means algorithm needs to be given before clustering. Therefore, students can be divided into several groups according to actual needs. Different groups of students can be analyzed separately. The K-Means algorithm is flexible and has been chosen by this paper.

Use the Euclidean distance [30] to calculate the distance between student features. If the distance calculation method can distinguish students, it indicates that the K-Means algorithm and the Euclidean distance calculation formula are applicable to the LPG algorithm.

Students are divided into groups after clustered, expressed as $G = \{G_1, G_2, \cdots, G_i, \cdots, G_k\}$. G_i represents the i-th group, and k represents the number of groups, that is, the number of clusters. Each student has a corresponding group label. In next steps, each group will be analyzed separately.

3) CONCEPT MAPS GENERATION

Based on the group to which each student belongs, the test record R is divided into $R = \{R_1, R_2, \cdots, R_i, \cdots, R_k\}$. Combine each test record with QC and use association rules mining to generate k concept maps.

This step refers to the concept maps generation algorithm proposed by Chen *et al.* [23]. The algorithm uses Apriori [31] algorithm to analyze each test record in R and calculate the frequent 2-itemsets between questions. With QC, frequent 2-itemsets between questions are mapped to association rules between concepts. The algorithm proposed by Chen *et al.* [23] is a classic concept maps generation algorithm, and its correctness is verified.

In order to avoid the relevant degree between concepts from being too small, the association rules [32] are not credible. Therefore, when calculating the association rules between concepts, the association rules with the relevant degree less than 0.5 are filtered, and k concept maps are generated ultimately, which expressed as M $=\{M_1, M_2, \cdots, M_i, \cdots, M_k\}$. M_i represents the i-th concept map, and k represents the number of concept maps. Each concept map reflects the learning performance of the corresponding group of students. The concept maps M will be analyzed separately in next steps.

B. LEARNING PATHS GENERATION FROM CONCEPT MAPS

Concept maps represent the associations between concepts, indicating that a concept should be learned before another concept. In a concept map, if the association rule between concepts C_1 and C_2 is $C_1 \rightarrow C_2$, it means that concept C_1 should be learned before concept C_2. Concept maps can be analyzed by topological sorting [33] algorithm algorithms to generate learning paths and different concept maps generate different learning paths. Calculate the level of mastery of each concept for each group of students. If a group of students has a low error rate in a concept, it means that the group of students has a high level of this concept and does not need to learn it again, so the concept can be removed from the overall learning path, and the simplified learning path is generated ultimately. The error rate of a group of students at a concept reflects the length of time that should be learned. Students in each group can learn concepts based on the corresponding simplified learning path, and teachers can also guide different groups of students according to the simplified learning paths.

1) LEARNING PATHS GENERATION WITH ALL CONCEPTS

In a concept map, when the number of concepts is large, the association rules between concepts are complex, and the labor force analysis concept maps is time consuming. In reality, the teacher's lecture sequence and the student's learning sequence are linear. Therefore, it is necessary to automatically convert complex inter-concept association rules into linear concept sequences. The topological sorting algorithm is a commonly used algorithm for analyzing directed acyclic graphs [34]. The algorithm can convert a directed acyclic graph into a sequence of nodes. Since there is acyclic between concepts, that is, between two concepts, they should not be learned before each other. It does not have any practical significance, so the topological sorting algorithm is feasible.

Analysis of each concept map in M, k learning paths are obtained. The k learning paths obtained in this step contain all the concepts, and the order of the concepts in each learning path is different. When the learning paths of different groups are different, it indicates that the distinguishing ability of the LPG algorithm is good. In order to verify that the LPG algorithm does have the ability to distinguish the student groups G, a definition is introduced, called the differentiation degree, and the formula is expressed as

$$\mathrm{Diff}_{ij} = \sum\nolimits_{1}^{n} |E_{ci} - E_{cj}| \tag{1}$$

Diff_{ij} indicates the differentiation degree between the i-th group and the j-th group in groups G on concept c, $c \in [1, n]$. n is the number of concepts. E_{ci} indicates the error rate of the c-th concept for students in the i-th group, and E_{cj} indicates the error rate of the c-th concept for students in the j-th group. The value of Diff is large, indicating that the previous steps have distinguished students from different features. This will be reflected in the comparison experiment.

2) SIMPLIFIED LEARNING PATHS GENERATION

The generated learning paths contains all the concepts, however, some concepts have been well mastered by students and do not need to be learned again. Therefore, the learning paths need to be simplified.

For concept c, if $E_{ci} \leq 0.3$, indicating that the students in group G_i have a high level on concept c, and it does not need to be learned again by the students in group G_i. Therefore, concept c can be removed from the learning path. If $E_{ci} \geq 0.7$, it means that the students in group G_i have a low level on concept c, and it needs to be focused on being learned again by the students in group G_i.

It can generally be considered that the larger the value of E_{ci}, the longer the students in group G_i should learn on concept c. Suppose that the duration T_{ci} of a group of students in G_i needs to learn concept c is only related to E_{ci} and is proportional. The greater the value of E_{ci}, the longer the students in group G_i need to learn concept c, and the larger the value of T_{ci}. Therefore, E_{ci} can be mapped to the learning duration. For convenience, simply map the value of E_{ci} to an integer interval as the value of T_{ci}, that is, let $T_{ci} = E_{ci} * 100$. Similarly, the learning duration for students in group G to learn all concepts is $\mathrm{T}_i = \sum_{1}^{n} T_{ci}$, where n is the number of concepts. The value of T_i and T_{ci} do not indicate the actual learning duration, but only provide a reference for teachers and students to arrange for teaching and learning. Simplified learning paths with learning durations are generated ultimately.

Based on the conceptual sequences of the learning paths, teachers can arrange teaching plans for each group of students. For a group of students, if the number of concepts in the learning path is large, it means that the number of concepts that the students cannot grasp is large, and teachers should give this group of students with a wider range of guidance. Similarly, if the value of the learning duration of a group of students is large, teachers should spend a lot of time to guide this group of students.

The pseudo-code of the LPG algorithm is as follows:

Algorithm LPG

Begin

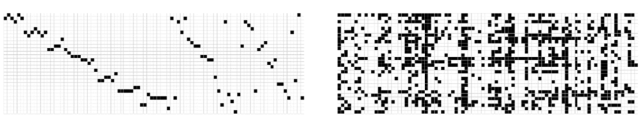
Input QC, R and k
For each student:
F ← error rates of students in each concept obtained by QC and R
F′ ← student's feature F
End for
Use K-Means to cluster students into G = $\{G_1, G_2, \cdots, G_i, \cdots, G_k$
R = $\{R_1, R_2, \cdots, R_i, \cdots, R_k \leftarrow G$
For each test records in R:
Calculate the frequent 2-itemsets between questions with R_i
Map the associations between questions to the associations between concepts with QC
Generate concept map M_i
End for
For each concept map in M:
Generation learning paths with all concepts
For each concept in concepts:
E_{ci} ← calculate the error rate of the concept
If $E_{ci} \leq 0.3$:
Remove concept c from learning path
End if
Calculate learning duration T_{ci} with E_{ci}
End for
End for
Output: Simplified learning path with learning durations

End

IV. EXPERIMENTS AND RESULTS ANALYSIS

A. DATA SOURCES AND EXPERIMENTAL ENVIRONMENT

The experiment selected 617,940 test records from 6,866 students in a large-scale quiz of the Computer Culture Foundation course as the experimental data set, including 90 questions covering 29 concepts. The data set was collected in December 2017 in the Shandong province in China. The relationships between test questions and concepts QC are given by authoritative experts in relevant fields, which guarantee correctness and authority. The data of QC are shown in Figure 3(a), and the partial data of R are shown in Figure 3(b).



(a) (b)

FIGURE 3. Visualization of the datasets used in this paper: (a) Data of QC; (b) Partial data of R.

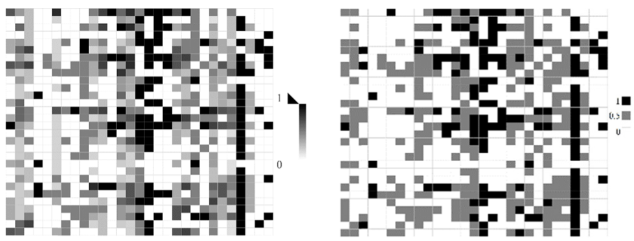
For convenience, in Figure 3(a), the black color indicates that the question belongs to the concept, and the white color indicates that the question does not belong to the concept. The abscissa represents concepts and the ordinate represents questions.

Similarly, in Figure 3(b), due to space limitations, we only show the test records of 30 students in R. The black color indicates that the student has answered the question incorrectly, and the white color indicates that the student answered the question correctly. The abscissa indicates the students, and the ordinate indicates the questions.

The experimental running environment is Windows 10 operating system, the programming language is Python 3.6, and the software development environment is PyCharm Community Edition 2018 and SQL Server 2008. The hardware environment of the experiment is Core 7th generation 3.4 GHz CPU and 16G memory.

B. EXPERIMENT OF CONCEPT MAPS GENERATION

Based on QC and R, students' features F are extracted, and according to the students' mastery of concepts, F is converted into F' according to high level, medium level and low level. Due to space limitations, the features of all students are not displayed. 30 student features F is shown in Figure 4(a), and 30 student features after conversion F' is shown in Figure 4(b).



(a) (b)

FIGURE 4. Student features before and after conversion: (a) Student features before conversion; (b) student features after conversion.

In Figure 4, the value of student features are indicated by the shade of the color. The darker the color, the greater the value of student feature. It can be found that after the student features conversion, the color of different shades in Figure 4(a) is converted into three colors of black, gray and white in Figure 4(b). It means that the students' conceptual error rates are converted into three forms: high level, medium level and low level. New student features F' will be clustered in the next step.

In reality, students are generally not divided into more than 5 groups due to the number of teachers, time and geographical restrictions. Therefore, suppose the reality needs to divide students into 5 groups, i.e. $k = 5$. In addition, two control experiments were set up. The experimental group using the LPG algorithm was named A. Instead of doing feature extraction for students, students are randomly divided into 5 groups, and such a control experiment that generates 5 concept maps and generates 5 learning paths is named B. Students who are neither feature extracted nor grouped, such

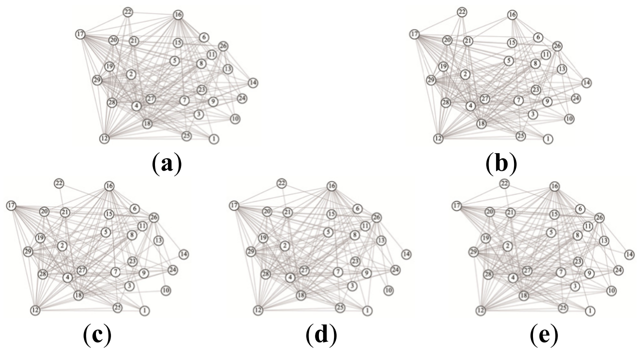


FIGURE 5. **Concept maps in experimental group A using the LPG algorithm.**

a control group that generates one concept map and generates one learning path is named C. In next steps, three experiments were analyzed separately.

In A, students are clustered into 5 groups by K-Means, combined with the association rules mining method, generated 5 concept maps as shown in Figure 5. In B, students are randomly divided into 5 groups, combined with the association rules mining method, generated 5 concept maps as shown in Figure 6. In C, students are not grouped, combined with the association rules mining method, then generated a concept map as shown in Figure 7.

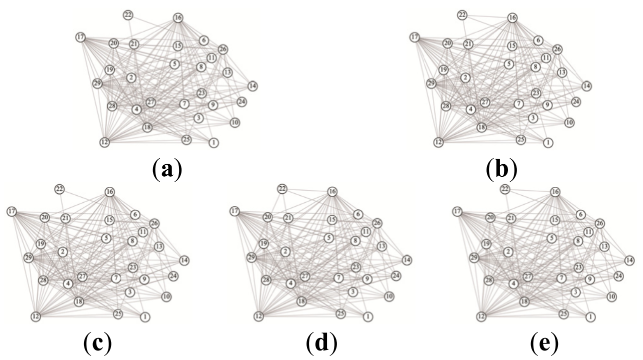


FIGURE 6. **Concept maps in control group B.**

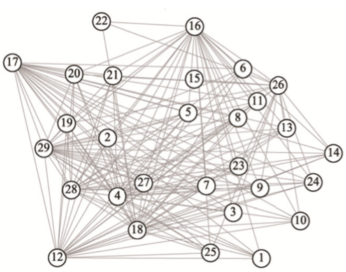


FIGURE 7. **The concept map in control group C.**

The correctness of the algorithm proposed by Chen *et al.* [23] has been verified by them, so the correctness of the above concept maps can be guaranteed. Generated concept maps indicate the associations between concepts, indicating which concept should be learned before which concept. However, it is difficult to visually discover the valuable information implied in these concept maps or the differences between them. Therefore, these concept maps will be analyzed in next steps.

C. EXPERIMENT OF LEARNING PATHS GENERATION

Each concept map in A, B, and C is analyzed using the topological sorting algorithm to generate several learning paths. The learning paths in A, B and C are expressed as shown in TABLE 2.

TABLE 2. **Learning paths in A, B and C.**

	Student group	Learning path
A	G1	27→25→24→23→22→21→20→19→14→13→11→10→9→8→7→6→5→3→2→1→16→18→28→29→17→12→4→15→26
	G2	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→26→16→1→29→18→28→12→17→4
	G3	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→4→12→29→16→17→26→18
	G4	27→25→24→23→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→18→17→28→4→29→12→16→22→26
	G5	27→25→24→23→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→18→28→29→12→4→16→22→26→17
B	G1	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17
	G2	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17
	G3	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17
	G4	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17
	G5	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17
C	G1	27→25→24→23→22→21→20→19→15→14→13→11→10→9→8→7→6→5→3→2→1→28→29→12→4→18→16→26→17

It can be found that the order of concepts in the five learning paths in A is different, and the order of concepts in the five learning paths in B is the same, and is the same as the learning path in C. It proves that the LPG algorithm can distinguish students from the level of mastery of concepts.

In order to further prove that the LPG algorithm can distinguish students, the differentiation degree is introduced. To calculate the degree of discrimination, it is necessary to

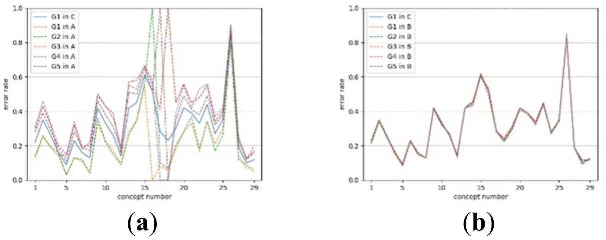


FIGURE 8. Error rates of each concept for each group of students in A, B and C: (a) The error rates of each concept for each group of students in A and the error rates of each concept for students in C; (b) The error rates of each concept for each group of students in B and the error rates of each concept for students in C.

first calculate the error rates of each concept for each group of students in A, B and C. The error rates of each concept for each group of students in A are shown in Figure 8(a), and the error rates of each concept for each group of students in B are shown in Figure 8(b). For a better comparison, the error rates of each concept for students in C are placed in both Figure 8(a) and Figure 8(b).

The error rates obtained in the previous step are then used to calculate the differentiation degrees between the groups in A and the differentiation degrees between the groups in B. At the same time, we compare each group in A with C and compare each group in B with C. The differentiation degrees between each group in A and C are shown in TABLE 3. The differentiation degrees between each group in B and C are shown in TABLE 4.

TABLE 3. The differentiation degrees between each group in A and C.

	G1 in A	G2 in A	G3 in A	G4 in A	G5 in A	G1 in C
G1 in A	0	1.41	5.92	4.94	4.70	2.97
G2 in A	1.41	0	5.79	4.81	4.65	2.90
G3 in A	5.92	5.79	0	2.04	2.70	2.97
G4 in A	4.94	4.81	2.04	0	2.08	2.71
G5 in A	4.70	4.65	2.70	2.08	0	2.07
G1 in C	2.97	2.90	2.97	2.71	2.07	0

It can be found that the differentiation degrees between concepts in A and C (TABLE 3) are much larger than the differentiation degrees between concepts in B and C (TABLE 4). It can further prove that LPG algorithm has good distinguishing ability. Analyze the five paths in A based on the student's error rate on each concept in the next step. If the error rate is less than or equal to 0.3, then the group of students can be considered to have a high level in this concept, and the concept can be removed from the learning path. The five learning paths in A are simplified and are sequentially shown in TABLE 5 and Figure 9. At the same time, in order to facilitate comparison of the concept sequences in A and C,

TABLE 4. The differentiation degrees between each group in B and C.

	G1 in B	G2 in B	G3 in B	G4 in B	G5 in B	G1 in C
G1 in B	0	0.21	0.18	0.20	0.22	0.13
G2 in B	0.21	0	0.19	0.25	0.19	0.12
G3 in B	0.18	0.19	0	0.24	0.18	0.15
G4 in B	0.20	0.25	0.24	0	0.22	0.17
G5 in B	0.22	0.19	0.18	0.22	0	0.11
G1 in C	0.13	0.12	0.15	0.17	0.11	0

TABLE 5. Simplified learning paths in A and C.

Student group	Learning path
G1 in A	23→21→14→9→15→26
G2 in A	23→21→15→14→9→26→16
G3 in A	25→24→23→22→21→20→19→15→14→13→11→10→9→6→3→2→16→17→26→18
G4 in A	25→24→23→21→20→19→15→14→13→11→10→9→6→3→2→1→16→22→26
G5 in A	25→24→23→21→20→19→15→14→13→11→10→9→2→16→22→26→17
G1 in C	25→23→22→21→20→15→14→13→10→9→2→16→26

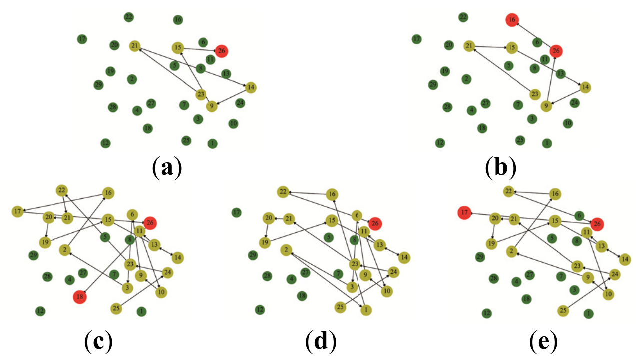


FIGURE 9. Concept maps with simplified learning paths.

the learning path in C is also shown in TABLE 5. In TABLE 4, it has been observed that the values of the differentiation degrees between B and C are small, indicating that the learning paths in B and C are basically the same, so B is no longer used for verification.

In TABLE 5, the sequence and number of concepts that each group needs to learn are different. It proves that the LPG algorithm not only can diagnose students' mastery of concepts, but also design several personalized learning paths. Calculate learning duration for each concept in simplified learning paths and total learning duration for each learning path, as shown in TABLE 6. At the same time, in order to facilitate comparison of the learning durations in A and C, the learning durations in C are also shown in TABLE 6. In Figure 8, it has been observed that the difference in error rates of concepts in B and C is small, indicating that the learning durations of B and C are basically the same, so B is no longer used for verification.

In Figure 9, the green color indicates the concepts that do not need to be learned by students. The yellow and red colors indicate the concepts that need to be learned by students. Based on the error rate, the error rate is expressed as the size of the concept node. The larger the size of the concept node, the longer the students need to spend on this concept. Simplified learning paths are generated ultimately.

TABLE 6. Conceptual learning durations in A and C.

Student group	Learning duration	Total learning duration
G1 in A	34→33→34→33→55→80	269
G2 in A	34→37→56→35→36→79→100	377
G3 in A	40→34→55→48→45→56→45→66→58→57→39→43→50→32→32→43→56→41→88→100	1028
G4 in A	42→37→56→42→55→39→67→53→55→36→43→49→34→33→46→31→58→53→90	919
G5 in A	38→32→49→41→49→35→65→50→51→31→38→46→39→52→38→90→100	844
G1 in C	35→44→33→39→42→61→45→42→33→42→35→52→85	588

In TABLE 6, it can be found that the learning paths have different learning durations. In comparison, it is found that the total learning duration of G1 in A is the smallest, which means that students in G1 in A have the best grasp of the concepts. It can also be found that the total learning duration of G3 in A is the largest, which means that the students in G3 in A have the worst grasp of the concepts, and the teacher needs to spend more time to guide the students in this group. If the students are not distinguished like G1 in C, the total learning duration is 588. For the students in G1 and G2 in A, the concepts are well mastered and do not need to be studied for so long. Reducing the learning duration and pruning the learning path can improve learning efficiency for students in G1 and G2 in A. For students in G3 and G4 and G5 in A, the concepts are poorly mastered and need to be learned longer. Increasing the learning duration and expanding the learning path can improve learning efficiency for students in G3 and G4 and G5 in A.

Combined with TABLE 5 and TABLE 6, concept 26 can be considered as a concept that should be focused on in all five groups in A. Teachers should focus on concept 26 when scheduling teaching progress. Teachers can provide personalized guidance to students based on these simplified learning paths with learning durations. Comprehensive analysis of the above results, some groups of students need to reduce learning durations, which can improve the learning efficiency. Some groups of students are exposed to more deficiencies and need to increase learning durations, which can also improve learning efficiency. Adaptability of the LPG algorithm has been validated. And it can be considered that the LPG algorithm can improve the learning efficiency.

V. CONCLUSIONS

The existing researches on concept maps have insufficient research on students' learning performance, and the results obtained are macroscopic and do not reflect the characteristics of adaptive learning systems. Aiming at these limitations, this paper proposes an automatic learning paths generation algorithm LPG algorithm based on concept maps for adaptive learning systems. The LPG algorithm uses clustering technology to automatically divide the students into several groups according to the mastery of concepts, and combines the association rules mining method to generate several concept maps, and uses the topology sorting algorithm to generate learning paths.

The simplified learning path with learning duration can provide educators with a students' learning diagnostic report and corresponding learning content recommendation plan. Students are adaptively assessed and given a corresponding learning path. As the number of quizzes increases, students can be continually diagnosed and the learning path can be automatically updated based on students' test records. Educators can continuously adjust the teaching plan based on the latest learning path, and students can further improve their learning efficiency. This study can provide teaching suggestions and planning for the adaptive learning system. As the data are continuously collected, the adaptive learning system can be updated in time to further enhance the adaptability.

The experiments show that the LPG algorithm has the following characteristics: 1) Good adaptability and high distinction between student groups. Several different learning paths can be generated based on students' mastery of concepts. 2) Low labor costs. The learning paths can be automatically generated and simplified. 3) High flexibility. The number of learning paths can be set according to actual needs.

Although the LPG algorithm performs well in automatically generating learning paths, it also has some limitations. Limited by factors such as time and space, this paper only verifies that LPG algorithm can distinguish students, does not observe students' long-term learning performance, and does not verify that LPG algorithm is helpful for improving students' grades. Moreover, in order to facilitate the extraction of students' learning features, relationships between questions and concepts can only be one-to-one, which reduces the flexibility of test papers. Collecting data from a wide time span and making LPG algorithm work on many-to-many relational data are our future work.

APPENDIX

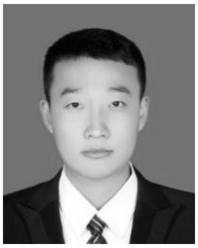
Supplementary data associated with this article can be found, in the online version, at https://github.com/diligentlee/LPG-algorithm-datasets.git.

REFERENCES

[1] Q. Li, S. Zhong, P. Wang, X. Guo, and X. Quan, "Learner model in adaptive learning system," *J. Inf. Comput. Sci.*, vol. 7, no. 5, pp. 1137–1145, 2010.

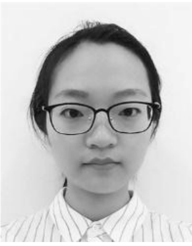
[2] B. Jia, S. Zhong, T. Zheng, and Z. Liu, "The study and design of adaptive learning system based on fuzzy set theory," in *Transactions on Edutainment IV*. 2010, pp. 1–11, doi: 10.1007/978-3-642-14484-4_1.

[3] K. M. Markham, J. J. Mintzes, and M. G. Jones, "The concept map as a research and evaluation tool: Further evidence of validity," *J. Res. Sci. Teach.*, vol. 31, no. 1, pp. 91–101, 1994, doi: 10.1002/tea.3660310109.
[4] J. D. Novak, "Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning," *Instructional Sci.*, vol. 19, no. 1, pp. 29–52, 1990, doi: 10.1007/BF00377984.
[5] J. Grundspenkis, "Development of concept map based adaptive knowledge assessment system," in *Proc. IADIS Multi Conf. Comput. Sci. Inf. Syst., e-Learn. (MCCSIS)*, 2008, pp. 395–402.
[6] C.-H. Lee, G.-G. Lee, and Y. Leu, "Application of automatically constructed concept map of learning to conceptual diagnosis of e-Learning," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1675–1684, 2009, doi: 10.1016/j.eswa.2007.11.049.
[7] T. Atapattu, K. Falkner, and N. Falkner, "A comprehensive text analysis of lecture slides to generate concept maps," *Comput. Educ.*, vol. 115, pp. 96–113, Dec. 2017, doi: 10.1016/j.compedu.2017.08.001.
[8] X. Huang, K. Yang, and V. Lawrence, "Classification-based approach to concept map generation in adaptive learning," in *Proc. IEEE 15th Int. Conf. Adv. Learn. Technol.*, Jul. 2015, pp. 19–23, doi: 10.1109/ICALT.2015.149.
[9] A. Acharya and D. Sinha, "An intelligent Web-based system for diagnosing student learning problems using concept maps," *J. Educ. Comput. Res.*, vol. 55, no. 3, pp. 323–345, 2016, doi: 10.1177/0735633116667358.
[10] J. D. Novak and D. B. Gowin, *Learning How to Learn*vol. 26, no. 1. Cambridge, U.K.: Cambridge Univ. Press, 1984, pp. 416–428.
[11] S.-Y. Choi, "A concept map_based adaptive tutoring system supporting learning diagnosis for students with learning disability," in *Proc. 9th Int. Conf. ICCHP*, vol. 3118. Paris, France, Springer, Jul. 2004, pp. 194–201, doi: 10.1007/978-3-540-27817-7_29.
[12] S. Bhattacharya, P. Basu, A. Chakraborty, and S. Roy, "A framework for interactive pattern based adaptive recommender agent using concept map for personalized e-learning: IPBARA," in *Proc. IEEE Int. Conf. Technol. Enhanced Educ. (ICTEE)*, Jan. 2012, pp. 1–5, doi: 10.1109/ICTEE.2012.6208601.
[13] C. B. Williams, J. P. Moore, A. Johri, R. S. Pierce, and C. North, "Advancing personalized engineering learning via an adaptive concept map," in *Proc. ASEE Annu. Conf. Expo.*, 2012, pp. 25.137.1–25.137.12.
[14] T.-Y. Chen, H.-C. Chu, Y.-M. Chen, and K.-C. Su, "Ontology-based adaptive dynamic e-learning map planning method for conceptual knowledge learning," *Int. J. Web-Based Learn. Teach. Technol.*, vol. 11, no. 1, p. 20, 2016, doi: 10.4018/IJWLTT.2016010101.
[15] N.-S. Chen, Kinshuk, C.-W. Wei, and H.-J. Chen, "Mining e-Learning domain concept map from academic articles," *Comput. Educ.*, vol. 50, no. 3, pp. 1009–1021, 2008.
[16] I. Qasim, J.-W. Jeong, J.-U. Heu, and D.-H. Lee, "Concept map construction from text documents using affinity propagation," *J. Inf. Sci.*, vol. 39, no. 6, pp. 719–736, 2013, doi: 10.1177/0165551513494645.
[17] S. Lee, Y. Park, and W. C. Yoon, "Burst analysis for automatic concept map creation with a single document," *Expert Syst. Appl.*, vol. 42, no. 22, pp. 8817–8829, 2015, doi: 10.1016/j.eswa.2015.07.034.
[18] S. Wang *et al.*, "Using prerequisites to extract concept maps from-Textbooks," in *Proc. CIKM*, 2016, pp. 317–326, doi: 10.1145/2983323.2983725.
[19] C.-F. Lai, S.-Y. Chen, C.-W. Tsai, Y.-C. Chang, and Y.-S. Su, "Using information retrieval to construct an intelligent E-book with keyword concept map," *EURASIA J. Math., Sci. Technol. Educ.*, vol. 13, no. 10, pp. 6737–6747, 2017, doi: 10.12973/ejmste/78195.
[20] V. dos Santos, "Concept maps construction using natural language processing to support studies selection," in *Proc. 33rd Annu. ACM Symp. Appl. Comput.*, 2018, pp. 926–927, doi: 10.1145/2851613.2851735.
[21] S.-M. Bai and S.-M. Chen, "Automatically constructing concept maps based on fuzzy rules for adapting learning systems," *Expert Syst. Appl.*, vol. 35, nos. 1–2, pp. 41–49, 2008, doi: 10.1016/j.eswa.2007.06.013.
[22] D. Dubois and H. Prade, "What are fuzzy rules and how to use them," *Fuzzy Sets Syst.*, vol. 84, no. 2, pp. 169–185, 1996, doi: 10.1016/0165-0114(96)00066-8.
[23] S.-M. Chen and P.-J. Sue, "Constructing concept maps for adaptive learning systems based on data mining techniques," *Expert Syst. Appl.*, vol. 40, no. 7, pp. 2746–2755, 2013, doi: 10.1016/j.eswa.2012.11.018.
[24] N. L. Schroeder, J. C. Nesbit, C. J. Anguiano, and O. O. Adesope, "Studying and constructing concept maps: A meta-analysis," *Educ. Psychol. Rev.*, vol. 30, no. 2, pp. 431–455, 2017, doi: 10.1007/s10648-017-9403-9.
[25] C. R. García, M. Cazorla, and O. B. García, "Meaningful learning using concept maps as a learning strategy," *J. Technol. Sci. Educ.*, vol. 7, no. 3, pp. 313–332, 2017, doi: 10.3926/jotse.276.
[26] T. Matsumoto, W. Sunayama, Y. Hatanaka, and K. Ogohara, "Data analysis support by combining data mining and text mining," in *Proc. 6th IIAI Int. Congr. Adv. Appl. Inform.*, Jul. 2017, pp. 313–318, doi: 10.1109/IIAI-AAI.2017.165.
[27] Z. Shao, Y. Li, X. Wang, X. Zhao, and Y. Guo, "Research on a new automatic generation algorithm of concept map based on text analysis and association rules mining," *J. Ambient Intell. Humanized Comput.*, pp. 1–13, Jul. 2018, doi: 10.1007/s12652-018-0934-9.
[28] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: A review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999, doi: 10.1145/331499.331504.
[29] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010, doi: 10.1016/j.patrec.2009.09.011.
[30] R. Loohach and K. Garg, "Effect of distance functions on simple K-means clustering algorithm," *Int. J. Comput. Appl. (0975–8887)*, vol. 49, no. 6, Jul. 2012, doi: 10.5120/7629-0698.
[31] S. Singh, R. Garg, and P. K. Mishra, "Review of apriori based algorithms on MapReduce framework," in *Proc. Int. Conf. Commun. Comput. (ICC)*, Bangalore, India, 2014, pp. 593–604.
[32] M. Martínez-Ballesteros, A. Troncoso, F. Martínez-Álvarez, and J. C. Riquelme, "Obtaining optimal quality measures for quantitative association rules," *Neurocomputing*, vol. 176, pp. 36–47, Feb. 2016, doi: 10.1016/j.neucom.2014.10.100.
[33] D. Ajwani, A. Cosgaya-Lozano, and N. Zeh, "A topological sorting algorithm for large graphs," *J. Exp. Algorithmics*, vol. 17, 2012, Art. no. 3.2, doi: 10.1145/2133803.2330083.
[34] B. Shipley, "A new inferential test for path models based on directed acyclic graphs," *Struct. Equation Model., Multidisciplinary J.*, vol. 7, no. 2, pp. 206–218, 2000, doi: 10.1207/S15328007SEM0702_4.

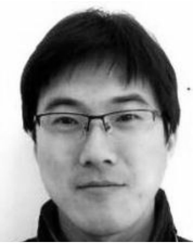


YANCONG LI was born in Zaozhuang, China in 1992. He is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University. His research interests are data mining and machine learning.

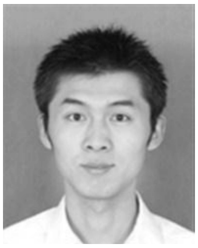
ZENGZHEN SHAO was born in Weifang, China in 1976. He received the Ph.D. degree from Shandong Normal University, China. He is currently an Associate Professor with the School of Information and Computer Science, Shandong Women's University. His main research interests are data analysis and prediction, machine learning, and emergency evacuation.



XIAO WANG was born in Taian, China in 1994. She received the B.S. degree from Shandong Jianzhu University, China, in 2017. She is currently pursuing the master's degree with the School of Information Science and Engineering, Shandong Normal University. Her research interest is big data processing.



XUECHEN ZHAO received the M.S. degree from the School of Information Science and Engineering, Shandong Normal University. He is currently an Assistant Teacher with the School of Data Science and Computer Science, Shandong Women's University. His main research direction is data mining.



YANHUI GUO is currently pursuing the Ph.D. degree with Shaanxi Normal University, Xi'an, China. He is currently a Lecturer with the School of Data and Computer Science, Shandong Women's University. His research interests include deep learning and pattern recognition.