J. COMBIN. MATH. COMBIN. COMPUT. 127b (2025) 3413--3428

Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

Predictive model for comprehensive development of vocational education students and optimal design of personalized learning path based on graph neural network

Qingbin Wei[1,✉]

[1] *Guangxi Vocational College of Water Resources and Electric Power, Nanning, Guangxi, 530023, China*

ABSTRACT

Graph neural networks are widely used in educational research, and have strong application potential in the prediction of students' comprehensive development and recommendation of personalized educational resources. In this paper, the information and characteristics of students are mined from massive learning data, and the prediction method of multi-topology graph neural network is used to realize the effective prediction of students' comprehensive development. Through the graph neural network, knowledge graph and cluster search algorithm and other technologies, the personalized learning path planning and optimization are completed, and the personalized learning path is designed. The research shows that the data accuracy of the student development trend prediction model in this paper reaches the qualifying value of 0.1, and the absolute maximum value of the error does not exceed 0.17, so the model constructed in this thesis is effective and robust. It can fulfill the task of student development direction prediction. The usage frequency of generating learning paths are more than 60%, so the learning path generation method proposed in this paper is practical. And the average grade of the users who use this method is 6.17 points higher than the average grade of the users who do not use this method.

Keywords: graph neural networks, personalized learning, attention mechanisms, recommender systems, prediction

1. Introduction

Since entering the 21st century, with the popularization of higher education and the increasing social demand for higher education, the national demand for quality education and academic qualifications has also been increasing. The scale of higher vocational education in China has been dramatically expanded, and now the scale of higher vocational education has reached nearly half of the proportion of higher education. As an organic part of higher education and an important part of higher education,

✉ Corresponding author.
E-mail address: weiqingbin191987@gxsdxy.edu.cn (Q. Wei).
Received 30 May 2024; Revised 15 August 2024; Accepted 10 January 2025; Published Online 16 April 2025.
DOI: 10.61091/jcmcc127b-190
© 2025 The Author(s). Published by Combinatorial Press. This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

higher vocational education is particularly important. However, the employment situation of vocational education in China still exists a more prominent structural contradiction, that is, the contradiction between the increasing number of vocational education graduates who have difficulties in choosing their future careers, unclear employment direction and unclear self-positioning, and the shortage of employment in enterprise units, where there is a shortage of workers and fewer workers [1-3]. From an educational perspective, this is caused by the mismatch between the job-seeking intentions of most graduates and the qualities and abilities they possess, or the incompatibility between employment opportunities and self-career planning [4].

With the rapid development of the information society, the types of work are becoming more and more diversified. Although the traditional talent training method ensures the employment rate of graduates of higher vocational colleges and universities to a certain extent, its shortcomings and students' irrational cognition of themselves make the students' employment quality low and their career stability bad [5-7]. And at the stage of training in higher vocational and technical colleges and universities, it is of great significance to make a prediction of the development trend of students and design personalized learning paths according to their specific conditions [8-9]. Because this will not only help students to clarify the direction of study and employment, but also allow them to clearly understand their shortcomings [10]. At the same time, this can also improve students' learning efficiency and make their learning methods more targeted [11]. In addition, making corresponding classification predictions and planning guidelines for the development trend of students will be of great help to the effective cultivation of talents in higher vocational education, improving the utilization of resources and enhancing the quality of employment [12-14]. In view of this, predicting the development trend of higher vocational students has become an urgent problem.

In the face of the increasingly severe employment situation of college students in China, a large number of scholars have studied the career development direction planning of graduates. Cho, K. Y. et al. used a hierarchical generalized linear model to investigate the systematic differences in student characteristics and discussed the factors influencing the employment decisions of professional vocational high school students [15]. Huang, C. explored the influencing factors of college students' employment in a local agricultural enterprise through theoretical research and empirical investigation, and introduced the Dematile method for ranking, and the results obtained effectively promoted the development of local agricultural enterprises [16]. Nie, L. used data mining methods to mine and analyze college students' career development behaviors, and then identified the relevant factors influencing students' career development choices, so as to provide reference and decision support for the management of college students' career development behaviors [17]. Sun, X. used the Apriori association rule algorithm to mine the association rules between student performance data and employment flow, and provided decision support for the personalized education path design of education managers through the management and analysis of employment data [18]. Jing, Y. et al. established a fuzzy clustering-based employment goal analysis system for college graduates, which can obtain scientific clustering results from student information and employment information data, so as to identify the main factors affecting college students' employment [19]. However, these studies only stop at the study of the influencing factors of college students' employment, and do not further establish employment prediction models or further refine the influencing factors, resulting in the low practicability of the final research results.

Some scholars have also predicted the employment of graduates by establishing various employment data models, evaluation index systems and employment prediction models, with a view to providing personalized guidance for the school's teaching management and students' career planning.Yang, Y. et al. showed that college students' career planning helps students to carry out accurate self-assessment and career orientation, and put forward a college students' career development prediction method based on the grey prediction model that revealed diverse career development needs and provided data support for educators to carry out personalized guidance [20].

Huang, F. et al. proposed a predictive model for college students' career direction based on sequence algorithm, which forms a data-driven career development plan and provides more accurate and personalized career guidance based on processing students' relevant data [21]. Shen, J. based on the big data analysis technology to establish a two-way employment prediction model for graduates, which accurately predicts the employment prospects of college students through the analysis of students' comprehensive quality and the current closing situation, so as to generate accurate employment guidance programs [22]. Fok, W. W. et al. examined the application of artificial intelligence and deep learning models in the construction of student development prediction models, which can not only analyze the students' disciplinary performance, but also can assess students' non-subject knowledge ability, so as to accurately and timely provide them with personalized curriculum and future development suggestions [23]. Li, H. Y. et al. investigated the application of decision tree algorithm in college students' employment prediction, and its analysis results can provide a fine-grained guidance for the employment direction of college students in many aspects [24]. Kumar, D. et al. evaluated the role of deep statistical computing and machine learning algorithms play a performance in predicting the employment status of students, which can help the subject objects such as students or teachers to prepare for the future development accordingly [25].

Nowadays, because of the unique information processing ability of artificial neural network, it is able to learn and analyze from a large number of data samples, and then find some difficult to parse the change rule, which can well deal with some nonlinear problems, and it has high application value in predicting the development direction of students. Therefore, starting from graph neural network, designing and realizing the comprehensive development prediction model of college vocational education students and carrying out the prediction on the direction of college students' career development can provide scientific and objective decision-making basis for the personalized employment guidance work and teaching work of college educators.

In this study, students' academic information is mined, screened, and processed through big data-related technologies to obtain students' basic information, learning behavior data, and so on. These data provide an effective basis for the subsequent comprehensive development prediction of students based on multi-topological graph neural network. Two vocational colleges are selected as research objects to verify the effectiveness and data accuracy of the student comprehensive development prediction model in this paper. In order to realize personalized learning in vocational education, this paper proposes a planning method for personalized learning paths based on graph neural network, combined with knowledge graph and cluster search optimization algorithm. Higher vocational students are randomly selected as research objects to conduct empirical research to explore the effect of their personalized learning path generation.

2. An analysis of the graph neural network approach to predicting students' overall development

2.1. Artificial Intelligence and BP Neural Networks

In the era of big data, the amount of data is huge and diverse, and traditional data processing methods can no longer meet the demand. Artificial intelligence technology enables more accurate and intelligent data processing and analysis by utilizing the information and patterns provided by big data for learning and decision making. A neural network is a common artificial neural network with a back propagation algorithm that is commonly used for tasks such as pattern recognition, classification, and prediction. Neural networks are trained by propagating errors layer by layer, constantly adjusting weights and biases to achieve more accurate outputs. Driven by big data, neural networks can be used for training and prediction. With inputs from big data, neural networks can continuously learn and optimize weights and biases to improve accuracy and generalization. Big data contains a large amount of information, and neural networks can help extract key features from it to better analyze and process the data. When dealing with large-scale data, neural networks can be used as one of the modeling and optimization methods to discover patterns and regularities in the data through the training process of neural networks. The relationship between big data-driven AI and neural networks is complementary. By combining data mining or K-means clustering techniques and neural network techniques, data-driven intelligent analysis and decision-making can be better realized, thus promoting the development and progress in the field of education.

2.2. Techniques related to graph neural networks

1) Graph convolutional neural network. Graph Convolutional Neural Networks use the idea of clustering methods to aggregate the features of neighboring nodes of a node layer by layer on the graph data and update the feature representation of their own node based on the adjacency between the nodes.

For the input graph data $G=(V,E)$, where $v\in V$ represents the nodes, $e\in E$ represents the edges connecting the nodes to each other, $X\in R^{d_0}$ represents the features of the nodes, and d_0 represents the node feature dimensions, the objective function of the feature matrix of the graph data after l convolutional layers is expressed as:

$$H^{(l+1)}=\sigma(AH^{(l)}W^{(l)}) \tag{1}$$

The structure of graph convolutional neural is schematically shown in Fig. 1. Where, $H^{(l+1)}\in\mathbb{R}^{N\times d_{l+1}}$ denotes the output layer feature matrix, N denotes the number of nodes, d_{l+1} denotes the feature dimension of the node output, $A\in R^{N\times N}$ denotes the adjacency matrix formed by the input graph, $H^{(l)}\in R^{N\times d_1}$ denotes the lth layer feature matrix, $W^{(l)}\in R^{d_1\times d_{l+1}}$ is the weight matrix of the lth to the $l+1$th implicit layer, and σ denotes the activation function.

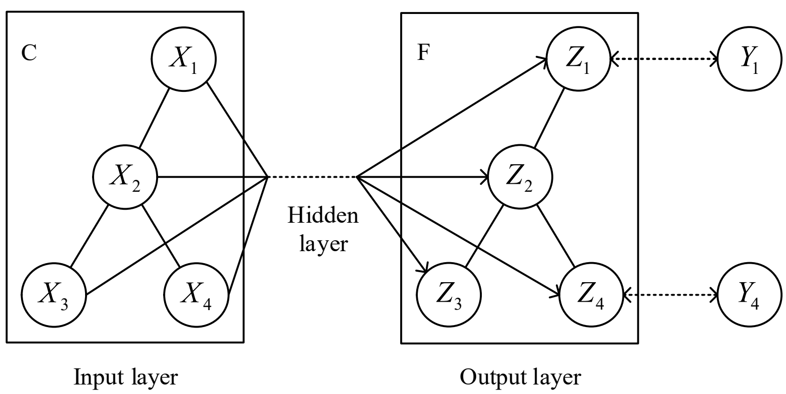

Fig. 1. Convolutional neural network structure

2) Graph Attention Networks. Graph Convolutional Neural Networks are limited to processing undirected graph data, which cannot detect the direction of edges between nodes in a directed graph. However, graph attention networks introduce an attention mechanism based on graph neural networks, which can extract features between nodes based on the null domain and utilize the self-attention mechanism of the hidden layer thus overcoming the shortcomings of the frequency-domain based graph convolutional neural network approach. The graph attention network greatly improves the flexibility and convenience of the algorithm by assigning appropriate weights to the neighboring nodes of each node without relying on complex matrix calculations.

Denote the feature vector corresponding to node v_i in layer l of the graph as $h=\{\vec{h}_1,\vec{h}_2,\ldots,\vec{h}_N\}$, where $\vec{h}_1\in R^F$, N denote the number of nodes and F denotes the feature dimension of each node. Considering the attention weight of its neighbor node j on i, the formula is calculated as follows:

$$e_{ij} = Leaky\,\mathrm{Re}\,LU\,(a^T[W\vec{h}_i \| W\vec{h}_j]) \tag{2}$$

where vector $W \in R^{F \times F}$ represents the weight matrix which can be used to linearly transform the features of each node, vector $\vec{a} \in R^{2F}$ represents the attention weights, e_{ij} is the computed attention coefficients, and $\|$ represents the stitching of the vectors.

After obtaining the attention values for all neighboring nodes of node $\vec{h}_i$, e_{ij} is normalized using softmax, and the computational formula is expressed as:

$$\alpha_{ij} = softmax(e_{ij}) = \frac{\exp(e_{ij})}{\sum_{\vec{h}_i \in N_i} \exp(e_{ik})} \tag{3}$$

where N_i denotes the set of all neighboring nodes of node i. After calculating the normalized weight coefficients α_{ij} of all the nodes, the information extraction of the nodes is carried out by using the graph attention layer and the output feature vector is represented as:

$$\overrightarrow{h_{i'}} = \sigma(\sum_{\overrightarrow{h_j} \in N_i} \alpha_{ij} W \vec{h}_j) \tag{4}$$

where σ denotes the activation function that traverses all neighboring nodes adjacent to node i, and the output node features are defined as weighted linear combinations of neighboring nodes under the attention mechanism.

2.3. Personalized learning path planning and optimization based on graph neural networks

2.3.1. Time-Series Neural Networks. A temporal neural network is a neural network model capable of processing time-series data, capturing long-term dependencies in sequential data, and modeling complex nonlinear patterns. Commonly used temporal neural network algorithms include long and short-term memory networks and gated recurrent units.

1) Long Short-Term Memory Networks. Long Short-Term Memory (LSTM) networks are a special type of recurrent neural network suitable for processing important events with relatively long intervals and delays in a time series, and have been widely used since then due to their unique ability to deal with long term dependencies in sequential data.

The memory cell update of LSTM is controlled by forgetting gate and input gate. Assuming that the memory cell of the previous moment is C_{t-1}, the output of the forgetting gate of the current moment is f_t, the output of the input gate is i_t, and the new candidate memory cell of the current moment is $\tilde{C}_t$, the update of the memory cell is computed in equation (5):

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{5}$$

Among them, the computational update rules for the forgetting gate, input gate and output gate are as follows:

Oblivion gate calculation is shown in equation (6):

$$f_t = \sigma\left(W_f \cdot \left[h_{t-1}, x_t\right] + b_f\right) \tag{6}$$

The input gate is calculated as shown in equation (7):

$$i_t = \sigma\left(W_i \cdot \left[h_{t-1}, x_t\right] + b_i\right) \tag{7}$$

The outputs are calculated in equation (8):

$$o_t = \sigma\left(W_o \cdot \left[h_{t-1}, x_t\right] + b_o\right) \tag{8}$$

and has equation (9):

$$h_t = o_t \cdot \tanh\left(C_t\right) \tag{9}$$

σ is the sigmoid activation function, $\tanh$ is the hyperbolic tangent activation function, W_f, W_i, and W_o are the learnable weight matrices, and $\left[h_{t-1}, x_t\right]$ denotes the splicing together of the hidden state of the previous moment and the input of the current moment.

In the learning path planning study presented in this paper, LSTM model is used to extract temporal features.

2) Gated Recurrent Unit. Gated Recurrent Unit (GRU) and Long Short-Term Memory Network LSTM are both improved versions of Recurrent Neural Network RNN used to process sequential data.GRU introduces update gate z_t and reset gate r_t to control the updating of information. The hidden state at the current moment is h_t, the hidden state at the previous moment is h_{t-1}, and the input at the current moment is x_t. The computation of the update gate and the reset gate are shown in equations (10) and (11), respectively:

$$z_t = \sigma\left(W_z \cdot \left[h_{t-1}, x_t\right]\right) \tag{10}$$

$$r_t = \sigma\left(W_r \cdot \left[h_{t-1}, x_t\right]\right) \tag{11}$$

where σ is the sigmoid activation function and the GRU uses the candidate hidden state $\tilde{h}_t$ (see Eq. (12)) and the hidden state h_t at the current moment (see Eq. (13)) to update the hidden state by resetting the gate:

$$\tilde{h}_t = \tanh\left(W_h \cdot \left[r_t \odot h_{t-1}, x_t\right]\right) \tag{12}$$

$$h_t = \left(1 - z_t\right) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{13}$$

where $\odot$ denotes element-wise multiplication.

In the learning path planning study presented in this paper, the GRU model is used to extract temporal features.

2.3.2. Cluster search algorithms. Beam Search is a search algorithm for finding the most probable sequence in a search space, and was originally commonly applied to natural language processing tasks such as machine translation and language model generation. Unlike greedy search, Beam Search considers multiple alternative sequences, called “beams” (beams), and retains some alternative solutions during the search process. In the application scenario of this paper, the subgraphs recommended for learners at each time node contain a huge number of knowledge concepts, and it is very difficult to form a specific knowledge planning path by linking the recommendation results of all time nodes. And cluster search can provide an effective method for forming specific paths of several knowledge concepts.

2.3.3. Educational Knowledge Mapping Construction. In the field of education, datasets have diverse data types, and the data cover numerous learners' learning behaviors and curriculum resources, etc. However, different data may adopt different data formats and storage methods. In order to organize the scattered information into a structured knowledge system, the construction of online education knowledge graph is carried out in this chapter. First, MOOCCubeX, an educational dataset, is analyzed as a preferred resource for constructing knowledge graphs, and its many features and advantages are described. Then, data preprocessing was carried out for the different formats of data in the MOOCCubeX dataset, and data consistency and usability were ensured through data cleaning, data transformation and data standardization. Finally, for the different storage methods in the MOOCCubeX dataset, this paper represents the knowledge graph in the form of ternary and uses the graph database Neo4j to store it.

2.4. *Planning and Optimization of Personalized Learning Paths*

In the face of the large and complex set of knowledge concepts in the learner mapping information, how to accurately indicate the learning direction for the learners and provide a truly meaningful learning path planning is the focus of this research paper. The learning path is shown in Figure 2. By the fusion of graph neural network and temporal neural network, simply stacking all relevant knowledge concepts together does not form an effective learning path.

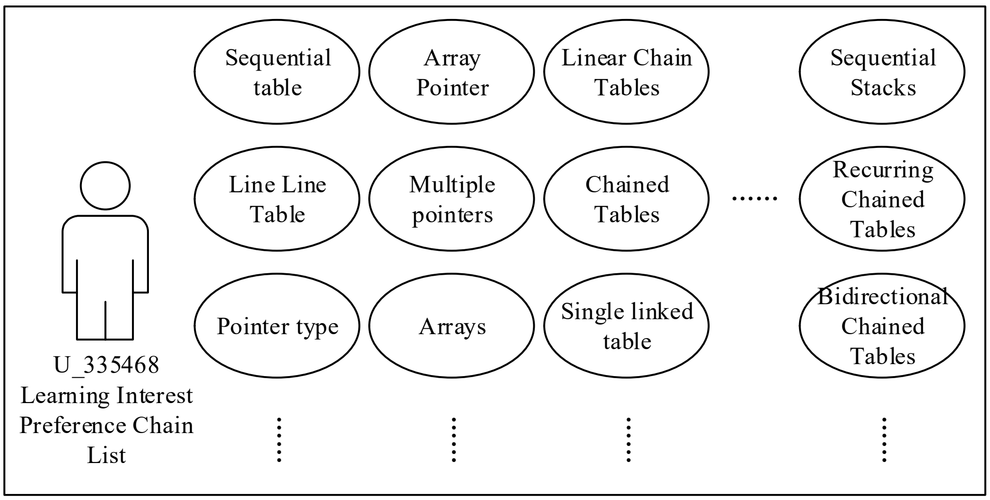

Fig. 2. Learning path

To solve this problem, this paper adopts a cluster search algorithm to optimize the learning path. The core idea of the cluster search algorithm is to keep a certain number (e.g., 3-5) of the highest scoring knowledge concepts at each time step, which are considered to be the most urgent need for learners to pay attention to and master at the moment. In this way, this study can ensure the diversity of learning paths while ensuring the quality of the selected paths. In addition to retaining the knowledge concepts with the highest scores, the knowledge concepts with the maximum probability for each time step were also obtained by means of softmax. The benefit of this approach is that it can more intuitively reflect the learner's learning focus at the current time step, thus providing a stronger basis for accurate planning of learning paths.

Ultimately, based on the sequential repair relationship and inclusion relationship between knowledge concepts, by concatenating the first TopK knowledge concepts and the maximum probability knowledge concepts retained at each time step, precise learning path planning and fuzzy learning path planning can be generated for learners. Precise planning provides learners with clear learning goals and steps, while fuzzy planning provides learners with a wider range of learning options and exploration space. This dual-planning scheme is designed to meet the individual needs of learners in different learning stages and scenarios, thus enhancing their learning efficiency and effectiveness.

3. Research on a method for predicting students' comprehensive development based on graph neural networks

Prediction of students' comprehensive development is one of the most important research problems in the field of educational data mining, which is crucial to help learners achieve learning outcomes. The purpose of student general development prediction is to utilize student-related information and data, obtain valuable information from them through existing data mining and analyzing techniques, and construct appropriate models to predict students' possible student general development in the next learning stage.

In this chapter, an academic achievement prediction method based on Multi-Topological Graph Neural Network (MTGNN) is proposed. In this study, the achievement prediction problem is modeled as a node classification task of graph neural networks. On the one hand, the potential relationships between students are defined on the graph, and for the process of graph structure characterization of learner relationships, the learners are regarded as nodes, the learner node linking algorithm based on similarity method is designed, and the potential relationship mining mechanism based on dynamic threshold discrimination is constructed to achieve the exploratory discovery of the edge relationships between nodes. On the other hand, for the problems such as the student relationship graph from a single perspective is not enough to mine the potential relationship of students, this study constructs the student relationship graph structure from different perspectives by improving and optimizing the graph neural network model, the model will consider the importance of the graph structure of each perspective, and merge the student relationship information of each graph structure.

The paper proposes a student achievement prediction model based on Multi-Topological Graph Neural Network (MTGNN). The model extensively considers the achievement relationships between students with similar characteristics and further improves the prediction of student achievement through these potential relationships. The framework of the model, which consists of three main phases: graph construction phase, information transfer phase and prediction phase. Multiple similarity learning methods are utilized in the graph construction phase to construct the student academic data into a graph structure that reflects the potential relationships among students. In the information transfer phase, a multi-topological graph neural network is designed based on the improvement of GNN, and the embedding information of the graph is extracted and fused through the attention strategy, and the fused embedding is used in the prediction phase.

4. Research on the prediction of student development based on graph neural networks

4.1. Status of student development in higher education

The survey was conducted in Guangdong Southern Vocational College and Guangdong ATV Performing Arts Vocational College. The survey was completed in October 2023 and October 2024, and the data collection is shown in Table 1. There are 320 graduates in the 2021 class of the two institutions, 310 people were sampled and 280 questionnaires were recovered, with a recovery rate of 90.23%. By the class of 2024, there were as many as 1657 students. The collected questionnaires were analyzed and the data were mined, screened and preprocessed to filter out the factors that have a greater impact on students' development.

This paper focuses on seven factors affecting students' development: family educational background, gender, hobbies and career matching, learning ability, communication ability, and professional skills. Meanwhile, this paper plans the employment direction of graduates from higher vocational colleges and universities into the following six categories: software development, civil service, academic advancement, operation and maintenance customer service, network maintenance, and others. The main influencing factors are the independent variables, one is the family educational background, which is divided into above high school and below high school based on the parents' education. The second is gender, which is divided into male and female. The third is interest and major matching, which is divided into interest and major matching and not interested in the major studied. Fourth is learning ability, which is converted into a percentile system using college entrance examination results and school grades as criteria. Fifth, sociability, which is converted into a percentage system by taking students' participation in campus activities and the opinions of counselors and class teachers as references. Sixth, professional skills, which is transformed into a percentage system by the standard of being able to complete the work tasks corresponding to professional skills alone. The development direction of students in higher vocational colleges and universities is the dependent variable, which is divided into six categories. The main content of this paper is to use K-Means algorithm in clustering algorithm. Taking the data of graduates of higher vocational colleges and universities as the reference object, mining the intrinsic connection between multiple items, i.e., mining the influence of the corresponding value of the dependent variable when the independent variable takes a different value, and utilizing the K-Means algorithm so that it can be applied to the prediction of the development trend of the students in higher vocational colleges and universities.

Table 1. Questionnaire Sampling Form

	Total number	Sampling number	Recovery questionnaire	Recovery factor
2021	320	310	280	90.32%
2022	752	520	500	96.15%
2023	1154	520	500	96.15%
2024	1657	520	500	96.15%

Through the preliminary statistical analysis of the obtained data, this paper firstly analyzes the current situation of students' development in Guangdong ATV Performing Arts Vocational College and Guangdong Southern Vocational College. The match between the specialties engaged in by the two senior graduates of the case institutions and the various factors we talked about above can be obtained, and the situation is shown in Table 2: among the two graduates surveyed, the number of graduates engaged in the network, online store operation and maintenance is the largest 266 out of the six development trend classifications. Then this paper analyzes the impact of family factors on the future development of students. From the data, we can see that the proportion of students whose parents' education is above high school who take civil service, choose academic promotion and network maintenance is higher than that of students whose parents' education is below high school, and the proportion of students who choose software development, networking, online store operation and maintenance other is lower than that of students whose parents' education is below high school. There is no obvious linear correlation between them, and the main influence is the proportion of the number of people who have academic advancement, network maintenance and engaged in other related. However, on the whole, the vast majority of students whose parents' education in family background is above high school are engaged in computer application technology related majors after graduation, and it also shows the consistency and clarity between parents and students in choosing majors.

Table 2. Statistical Data of Graduates Engaged in Work and Various Factors

Influencing factor		Software	Civil servant	Promote	Operation and customer	Network maintenance	Other
Family background	High school	74	8	9	46	7	32
	Under high school	82	2	4	220	6	200
Gender	Male	150	8	8	230	12	210

	Female	6	2	5	36	1	22
Interest and professional matching	Match	156	5	9	180	13	50
	Mismatch	0	5	4	66	0	182
Learning ability	Pass	133	9	12	157	8	90
	No pass	23	1	1	109	5	142
Communicative ability	Pass	54	7	3	221	5	180
	No pass	102	3	10	45	8	52
Professional skill	Pass	144	3	6	33	12	2
	No pass	22	7	7	233	1	230

4.2. Analytical validation test of the model for predicting student development trends

According to the above-mentioned developmental status of higher vocational students, we classify the developmental direction of students into six categories, so in the K-Means clustering algorithm, we directly set the k-value to 6 without changing the size of the k-value, so as to carry out data mining in a targeted way. In this paper, we consider the factors of students' family background and gender, the matching factors of students' interests and majors and the influencing factors of students' learning ability, the factors of students' communicative ability, and the factors of students' professional ability as the inputs of students' development direction prediction model, and take the expected development direction as the outputs of the students' development direction prediction model finally constructed according to the graph neural network.

In this thesis, 500 sample data were selected as the training use data and 60 sample data were selected as the test use data, and the results showed that the prediction model of the development trend of higher vocational students has high accuracy, and the specific experimental results of the 3 groups are shown in Figure 3. After 112 steps of training, the data accuracy of the prediction model of the development trend of higher vocational students basically reaches the accuracy requirement of the absolute error limit value of 0.1. In this paper, a comparison experiment was designed to compare the experimental data of the prediction model of student development trend with the sample data of the actual study.

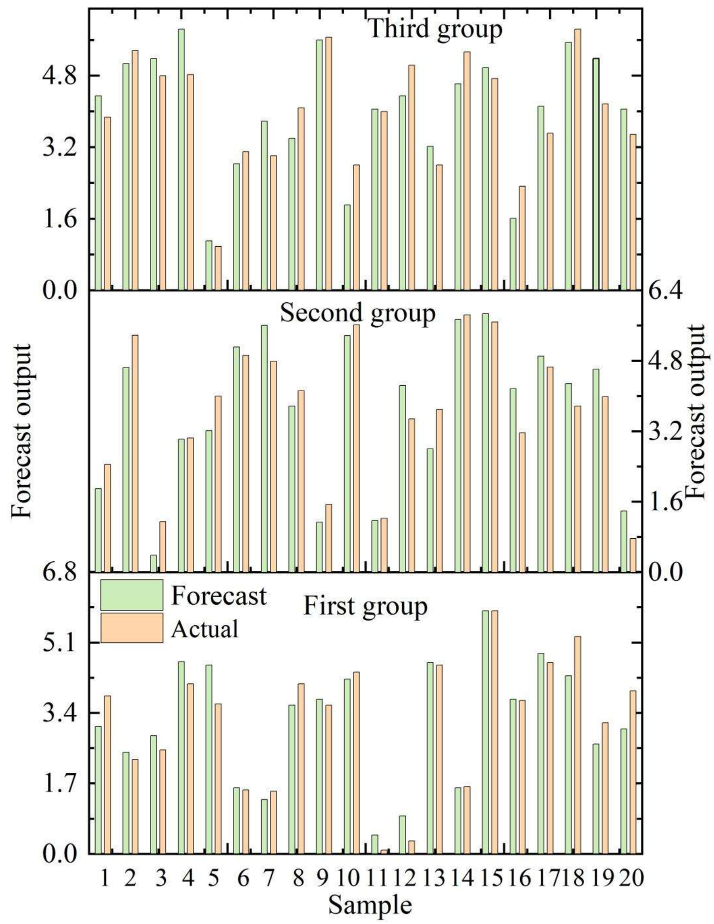

Fig. 3. Three specific experimental results

This paper further analyzes the experimental data, the unrefined, insufficiently comprehensive and specific problems in the selection of characteristic attributes and the selection of the direction of student development trend prediction. The error value is shown in Figure 4, from the overall point of view, the absolute maximum value of the error obtained from the student development trend prediction experiment in this paper does not exceed 0.17, which is within the acceptable error range, so the student development prediction model constructed in this thesis can accomplish the task of student development direction prediction, and it can preliminarily solve the problem of student development prediction, and it can satisfy the demand of development prediction of students in higher vocational colleges and universities, and it can be used to students in higher vocational colleges and universities to plan the direction of future development.

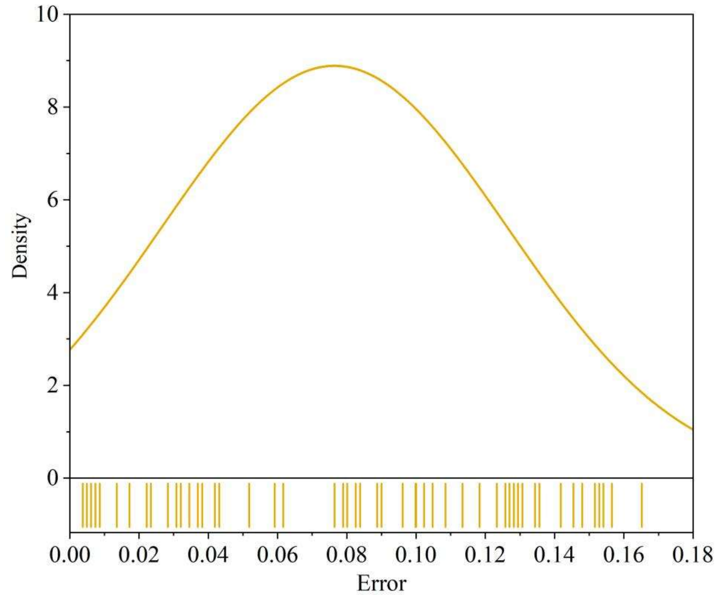

Fig. 4. Absolute Value Map of Experimental Results

5. Research on personalized learning paths in vocational education based on graph neural networks

5.1. Algorithm results and analysis

When generating the knowledge graph, whether there is a connection between two nodes is determined by the cosine similarity of the two exercise sentence vectors, and when the thresholds are different, the number of edges in the generated knowledge graph is different, and the results of the model are also different. In this paper, five different thresholds, 0.5, 0.6, 0.7, 0.8 and 0.9, are chosen to observe the experimental results. The experimental results corresponding to different thresholds are shown in Table 3. From the above experimental results, it can be found that with the continuous increase of the threshold value, the results of the model are getting better and better, when the threshold value is increased to 0.8, the model has the best effect, and when the threshold value continues to increase to 0.9, the effect of the model has begun to decline. This result is not difficult to understand, when the threshold is lower, the number of relationships in the knowledge graph is higher, there is more noise data in the knowledge graph, and many unnecessary features may be learned during the convolution operation, when the threshold is higher, the number of relationships in the knowledge graph is lower, and the features learned during the convolution operation are incomplete, and the model effect also decreases.

Table 3. The different thresholds correspond to the results of the experiment

Threshold value	The training set is accurate	Test set accuracy rate	Training set loss value	Test set loss value
0.5	0.92245	0.5822	0.5568	0.9322
0.6	0.9218	0.5822	0.8647	0.8042
0.7	0.8924	0.6107	0.8530	0.8043
0.8	0.8749	0.7143	0.5051	0.7046
0.9	0.8733	0.6377	0.4516	0.9887

From the experimental results, it can also be found that when the number of model iterations reaches about 300, the accuracy of the model on the training set reaches the highest, that is, the model has been fully learned, and continue to increase the number of iterations will cause the problem of overfitting, in the training of the model in this paper, the number of iterations is set to 300, in addition to the number of iterations in the training of the model, there are many other parameters that need to be set, such as the learning rate, the convolutional number of layers, number of neurons in the convolution layer, etc. The settings of these parameters are shown in Table 4. It is worth noting that the number of convolutional layers of the neural network should not be too many, because the convolution process of the graph convolutional neural network is to use the features of the surrounding nodes to represent this node, if the number of convolutional layers is too many, as the model is trained, the nodes undergo too many convolutional operations, and the representation of the individual nodes may become extremely similar, and the nodes do not have a lot of difference between each other, which in turn reduces the model effect.

Table 4. Parameter setting

Parameter	Learning rate	Iteration number	The number of neurons in the hidden layer	Dropout	Ealy_stopping	Layer number
Value	0.02	300	20	0.6	22	3

5.2. Effectiveness of Personalized Learning Path Generation

The learning paths in this study are learning resource-oriented paths with the ultimate goal of recommending a sequence of resources to be learned to the learner. In order to verify the effect of personalized learning path generation supported by educational knowledge graph and graph neural network, this study also randomly selected 30 higher vocational students from the developed system as the research object, and analyzed their 2-month learning process data, the specific steps of the analysis are as follows:(1) Group their learning records based on learner number, and sort the learning records in chronological order, so as to form a knowledge meta-learning sequence. (2) Remove the learning records in which the learner repeats the same knowledge element in consecutive time, and keep only the last learning record. (3) Extract the actual learning path of the learner from the de-weighted data records. (4) Take the last knowledge element in the learning path Lr as the learning goal, and use the cluster search optimization algorithm to automatically generate the learning path based on the learner's prior knowledge subgraph. (5) Comparing the order of knowledge elements on the actual learning path and the automatically generated learning path of each learner, the number of times that are the same is recorded as P1, and the number of times that are different is recorded as P2, then the frequency of the use of the generated paths can be expressed as $p = p_1/(p_1 + p_2)$. Based on the above analysis process, the frequency of the use of the system-generated learning paths by the 30 learners was counted, and the specific results are shown in Fig. 5. From the figure, it can be seen that the frequency of use of the generated learning paths is above 60%, and the average frequency of use is 75.5%, thus demonstrating that the learning path generation method supported by the educational knowledge graph and graph neural network proposed in this paper has certain practicality.

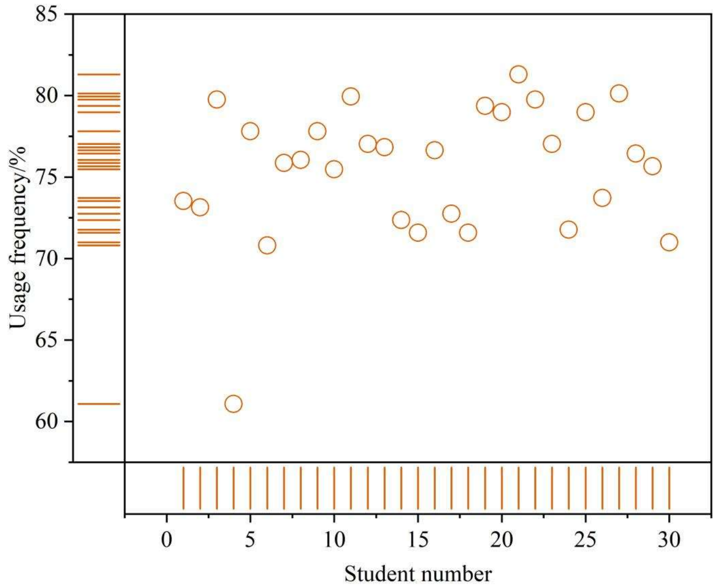

Fig. 5. The implementation of the path recommendation

5.3. Validity testing

In order to verify the validity of this model, in this paper, for the same school, 200 students were randomly selected and divided into two groups of 100 each. A detailed research was done on the students who used this model and those who did not use this model, and their test scores in the same exam were counted. The results are shown in Fig. 6, the failure rate of users who use this model is much lower than those who do not use this model. In addition, in some higher zones, such as 81-90 and above 90 points, the percentage of users using the model is higher than that of users not using the model. The average score of the users without the model is 71.75, and the average score of the users with the model is 77.92, which shows that the model can effectively improve the users' scores.

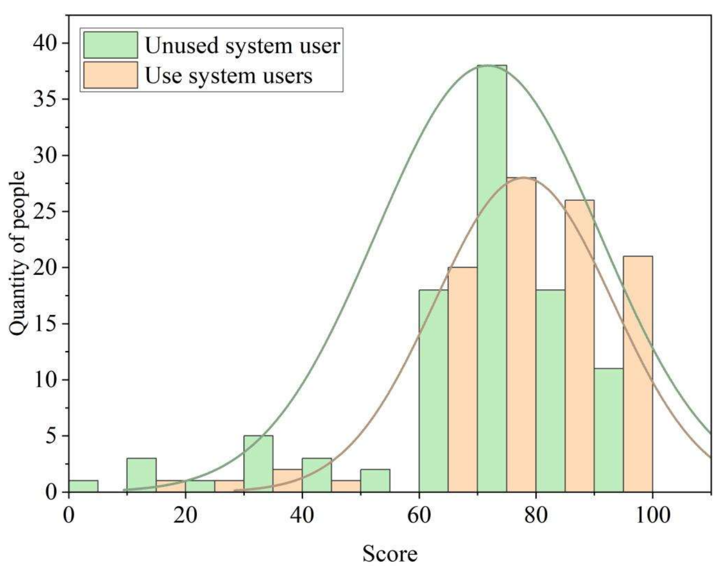

Fig. 6. System validity test

6. Conclusion

In this paper, we constructed a comprehensive student development prediction model and a personalized learning path planning model of multi-topology graph neural network based on graph neural network, so as to realize the effective prediction of student development and the optimal design of personalized learning path. The following conclusions can be drawn through the validation experiments:

1) By comparing the prediction data of students' comprehensive development with the sample data of the actual study, it is found that the absolute error limit value of the prediction model is 0.1, and the absolute maximum value of the error does not exceed 0.17. Therefore, the prediction model of students' comprehensive development in this paper can meet the requirements of experiment and application.

2) The personalized learning path planning model, with the increasing threshold value, generates better results. Moreover, the learners' usage frequency of learning paths are over the passing line, and the average usage frequency reaches 75.5%. The learning path generation method based on the knowledge graph and graph neural network support has certain practicality, and can also effectively improve the user's performance.

Funding

This work was supported by Guangxi Education Science Planning Project-Digital Technology Empowers the Reform of Students' Comprehensive Quality Evaluation in Higher Vocational Colleges Practice and optimization strategy research.

References

[1] Cherednichenko, G. (2020). Employment and labor market outcomes of college and vocational school graduates. Voprosy obrazovaniya/Educational Studies Moscow, (1), 256-282.

[2] Cheng, W., & Nguyen, P. N. T. (2024). Academic motivations and the risk of not in employment, education or training: university and vocational college undergraduates comparison. Education+ Training, 66(10), 91-105.

[3] Perić, J., Oberman Peterka, S., & Getoš, Ž. (2020). The role of vocational education in developing entrepreneurial competences of students. Ekonomski pregled, 71(5), 463-492.

[4] Choi, S. J., Jeong, J. C., & Kim, S. N. (2019). Impact of vocational education and training on adult skills and employment: An applied multilevel analysis. International Journal of Educational Development, 66, 129-138.

[5] Kazi, A. S., & Akhlaq, A. (2017). Factors affecting students’ career choice. Journal of research and reflections in education, 2(2), 187-196.

[6] Ahmed, K. A., Sharif, N., & Ahmad, N. (2017). Factors influencing students’ career choices: empirical evidence from business students. Journal of Southeast Asian Research, 2017(2017), 1-15.

[7] Sibiya, A. T., & Nyembezi, N. (2018). Examining factors that shape Technical Vocational Education and Training engineering students’ understanding of their career choices. Transformation in Higher Education, 3(1), 1-6.

[8] Rosantono, I. G., Wijanarka, B. S., Daryono, R. W., & Nurtanto, M. (2021). Analysis of the influencing factor of vocational education students career decisions. Jurnal Pendidikan dan pengajaran, 54(3), 582-595.

[9] Yu, R., Li, Q., Fischer, C., Doroudi, S., & Xu, D. (2020). Towards Accurate and Fair Prediction of College Success: Evaluating Different Sources of Student Data. International educational data mining society.

[10] Yang, F. (2019). Decision tree algorithm based university graduate employment trend prediction. Informatica, 43(4).

[11] Zhang, Y., & Yang, X. (2022). Prediction of the Development Scale of Vocational Education Using Markov Algorithm and Countermeasures. Mobile Information Systems, 2022(1), 9932083.

[12] Wei, Y., Rao, X., Fu, Y., Song, L., Chen, H., & Li, J. (2023). Machine learning prediction model based on enhanced bat algorithm and support vector machine for slow employment prediction. Plos one, 18(11), e0294114.

[13] Saidani, O., Menzli, L. J., Ksibi, A., Alturki, N., & Alluhaidan, A. S. (2022). Predicting student employability through the internship context using gradient boosting models. Ieee Access, 10, 46472-46489.

[14] Krishnaiah, V., & Kadegowda, Y. H. (2022). Undergraduate engineering students employment prediction using hybrid approach in machine learning. International Journal of Electrical and Computer Engineering, 12(3), 2783.

[15] Cho, K. Y., & Jyung, C. Y. (2014). Factors affecting the determinants of employment decision making of vocational high school senior students. Journal of Agricultural Education and Human Resource Development, 46(1).

[16] Huang, C. (2023). Evaluation of the Factors Affecting the Employment of College Students in Agricultural Enterprises in Fujian ProvinceBased on Dematel Analysis. Adult and Higher Education, 5(5), 94-100.

[17] Nie, L. (2023, August). College students' career prediction model based on association rule mining algorithm. In EAI International Conference, BigIoT-EDU (pp. 378-384). Cham: Springer Nature Switzerland.

[18] Sun, X. (2024). A Big Data Analysis Model for College Employment Based on Decision Tree and Apriori. In Intelligent Computing Technology and Automation (pp. 1098-1106). IOS Press.

[19] Jing, Y., & Wang, J. (2024). Influencing Factors Analysis of College Employment Based on Fuzzy Clustering. In Intelligent Computing Technology and Automation (pp. 1206-1213). IOS Press.

[20] Yang, Y., Meng, F., Zhang, Y., Pan, H., & Su, S. (2024, October). Research on career planning prediction of college students based on grey model. In Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Intelligent Information Processing (pp. 134-139).

[21] Huang, F., & Feng, Z. (2024, June). Research on a College Student Career Development Direction Prediction Model Based on Sequential Algorithms. In Proceedings of the 2024 International Conference on Intelligent Education and Computer Technology (pp. 572-576).

[22] Shen, J. (2023). A bilateral employment situation prediction model for college students using GCN and LSTM. PeerJ Computer Science, 9, e1494.

[23] Fok, W. W., He, Y. S., Yeung, H. A., Law, K. Y., Cheung, K. H., Ai, Y. Y., & Ho, P. (2018, May). Prediction model for students' future development by deep learning and tensorflow artificial intelligence engine. In 2018 4th international conference on information management (ICIM) (pp. 103-106). IEEE.

[24] Li, H. Y., & Zhang, Y. (2020, July). Research on employment prediction and fine guidance based on decision tree algorithm under the background of big data. In Journal of Physics: Conference Series (Vol. 1601, No. 3, p. 032007). IOP Publishing.

[25] Kumar, D., Verma, C., Singh, P. K., Raboaca, M. S., Felseghi, R. A., & Ghafoor, K. Z. (2021). Computational statistics and machine learning techniques for effective decision making on student's employment for real-time. Mathematics, 9(11), 1166.