International Journal for Housing Science and Its Applications
Publish August 3, 2025. Volume 46, Issue 3 Pages 61-77

https://doi.org/10.70517/ijhsa46305

Research on Optimized Construction and Path Recommendation Algorithm of Same-Class Heterogeneous Curriculum Resources Supported by Knowledge Graph

Yuqi Li[1] and Chunsu Zhang[1,*]
[1] College of Computer Science and Technology, Beihua University, Jilin, Jilin, 132000, China
Corresponding authors: (e-mail: zhangchunsu23@163.com).

Abstract With the in-depth development of education informatization, the optimal construction of co-curricular heterogeneous course resources and the recommendation of personalized learning paths have become an important research field in the education sector. In this paper, we propose an algorithm for the optimal construction of heterogeneous course resources and learning path recommendation based on subject knowledge mapping. Firstly, based on the linear chain conditional random field model and the relational multi-classification model, the recognition of named entities and the extraction of entity relations are accomplished respectively. Then the relationships between the obtained subject knowledge points are inputted into Neo4j graphical database to complete the visualization of subject knowledge graph. The smart learning E-GPPE-C model and the constructed subject knowledge graph are combined to construct the learner portrait and complete the design of the learning path recommendation algorithm from four cognitive levels. The entity recognition and relationship extraction models in this paper both obtain more excellent experimental results, and the learning path conforms to the law of gradual progression in education, obtaining much higher than the comparative algorithm P@20 value, which reaches 75.45. The method in this paper can effectively explore more scientific learning paths, provide a method for the optimization of the resources of co-curricular heterogeneous courses, and provide support for personalized education.

Index Terms subject knowledge graph, linear chain conditional random field model, named entity recognition, knowledge graph visualization, learning path recommendation

I. Introduction

With the rapid development of information technology, digital transformation has gradually become a reality that all industries need to face. The field of education is also advancing with the trend of the times, and gradually began the transformation of digital education. With the help of modern advanced science and technology, educators optimize the design of the curriculum, and through the rich and diverse forms of curriculum teaching and teaching resources, stimulate students' senses, focus students' attention, expand students' knowledge system, which not only can improve the teachers' own teaching level, but also can improve the students' learning effect and ability to think [1]-[3]. Among other things, digital transformation has enabled educational resources to be more widely shared and open. The traditional education model is limited by time and space, while the digital transformation breaks this limitation, so that learners can access the required educational resources anytime and anywhere, such as online databases, online courses, e-books and other open educational resources, and such a change improves the popularity and equality of education, so that more people can enjoy high-quality educational resources [4]-[7]. The construction and application of digital curriculum resources is an important part of the “classroom revolution” [8].

Educational information resources show a trend of massive growth, how to quickly and accurately obtain the required content from these resources has become a major challenge in the field of education. And recommendation algorithm is the best way to solve the above challenges. Through recommendation algorithms, on the one hand, educational information resources can be more reasonably distributed and utilized, so that quality resources can better serve learners and improve the quality of education. On the other hand, educational information resources can also be adjusted and optimized in real time according to learners' feedback and behavioral data to provide personalized services for learners to enhance their experience and learning satisfaction [9], [10]. Digital course resources have become the most basic means of supporting education at present. At this stage, in order to better enhance the quality of education and improve the independent learning ability of students, institutions compete to introduce digital curriculum resources to create an advanced and diverse learning atmosphere for student education, so as to achieve the teaching goal of independent learning [11], [12]. At the same time, educational curriculum resources are also facing the overload of similar teaching resources, and the high level of content differences between resources of the same class, and the low degree of matching between the knowledge structure and the knowledge system associated with it, which makes it difficult for students to accurately access the required resources, and increases the cost of time for resource finding [13]-[15]. In addition, the recommendation program of course resources does not match the current demand in real time, which is easy to cause the recommended resources to deviate from the demand, and the usability of the program is reduced [16].

Knowledge graph (KG) is a knowledge network that describes the real-world entities and their interrelationships, where entity characteristics are represented in the form of “attribute-value” pairs, and relationships between entities are represented by edges in the network [17]. It can form a common understanding of domain knowledge and realize the representation, sharing and reuse of knowledge. In the field of educational resources, Huang and Yang [18] used KG technology to construct geography curriculum resources, which efficiently demonstrated the correlation between geography knowledge points and facilitated the use of resources by students with diverse learning needs. Li et al [19] combined AI and KG technologies to mine classification and knowledge association of teaching resources respectively, so as to realize the integration of educational resources and facilitate efficient access and utilization of teaching resources. Niu et al [20] proposed a learning resource recommendation method for online courses using KG and collaborative filtering by simultaneously considering students' learning needs and their preferences for different modal learning resources. Wei and Yao [21] co-constructed a course learning resource recommendation model based on KG technology for knowledge association of resources and incorporating personalized interest similarity calculation. Ma et al [22] designed a new model for learning resource recommendation, mainly based on KG technology, applying graph embedding algorithm, graph matching technology, and learning style theory to achieve more accurate resource recommendation from resource similarity identification and student interest. These researches provide theoretical support for KG to optimize the construction of resources for co-curricular courses.

In this paper, a named entity recognition algorithm based on CRF model is designed, and the algorithmic framework for named entity recognition is designed in detail from three levels: corpus preprocessing, feature extraction and model training. On the basis of the traditional entity relationship extraction algorithm, positional features are added to distinguish different entity relationships, and feature enhancement method is used to optimize the effect of entity relationship extraction. Subsequently, Neo4j graph database is introduced to complete the construction and visualization of subject knowledge graph. A learner portrait portrayal method combining the E-GPPE-C model framework is then proposed from four aspects, and the recommendation process of learning paths is refined from four different directions: current cognitive state annotation, target learning state annotation, group learning path planning, and personalized learning path recommendation. Comparative experiments are used to verify the performance of the named entity recognition and entity relationship extraction models of this paper, and the effectiveness of this paper's method on the optimization of co-curricular and heterogeneous course resources and the recommendation of learning paths is verified on three different disciplinary datasets.

II. Machine learning-based knowledge extraction of co-curricular curriculum resources

II. A. Named Entity Recognition Based on Linear Chain Conditional Random Fields

II. A. 1) Principles of Linear Chain Conditional Random Field Modeling

Let X and Y be random variables, $P(Y|X)$ be the conditional probability distribution of Y given X, and a conditional random field is an undirected graphical model trained by maximizing the conditional probability distribution $P(Y|X)$, i.e., it satisfies Markov property. Linear chain conditional random fields are often used for sequence labeling, where X denotes the observation sequence or input sequence, and Y denotes the corresponding state sequence or output sequence. For example, in a secondary school chemistry textbook, let the observation sequence X be “Aluminum reacts with sodium hydroxide solution”, and its corresponding state sequence Y be “Aluminum/Knowledge point reacts with/other sodium hydroxide/Knowledge point solution/other/other”.

Given an observation sequence X, the conditional probability of a state sequence Y, where $Z(x)$ is the normalization factor, t_k and s_i are the eigenfunctions, and λ_k and μ are the weights, is said to indicate that the conditional random field is determined by the eigenfunctions and their weights.

$$P(y|x)=\frac{1}{Z(x)}\exp(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{i,j}\mu_i s_i(y_i,x,i)) \tag{1}$$

$$Z(x)=\sum_y \exp(\sum_{i,k}\lambda_k t_k(y_{i-1},y_i,x,i)+\sum_{ij}\mu_i s_i(y_i,x,i)) \tag{2}$$

Given an observation sequence x, the conditional probability of predicting the state sequence y. The CRF model is trained by maximizing the conditional probability $P(y|x)$, i.e., the maximization of the probability of correctness of the prediction of the state sequence. The eigenfunctions t_k are called transfer features and depend on the state at the previous and current positions. The characteristic function s_i, called the state feature, depends only on the state at the current position. Both depend on the position, are localized feature functions, and usually take the value of 0 or 1, 1 when the condition is met, and 0 when it is not met.

$$t_k(y_{i-1}, y_i, x, i) = \begin{cases} 1 & y_{i-1} \text{ and } y_i \text{ satisfy some combination condition} \\ 0 & \text{other} \end{cases} \tag{3}$$

From the above, the problems that need to be solved for training CRF models are feature selection and parameter estimation. Feature selection is to choose the appropriate set of features for the CRF model with respect to the corpus, and the features influence each other and depend on each other. Parameter estimation refers to selecting appropriate weights for different features from the training data, and the larger the weight, the greater the influence of the feature. CRF models generally use maximum likelihood estimation [23] to adjust the parameters.

II. A. 2) Named Entity Recognition Framework for Secondary School Subject Areas

The named entity recognition framework based on linear chain conditional random field in this chapter is shown in Fig. 1, which consists of three main parts, namely corpus preprocessing, feature extraction and model training.

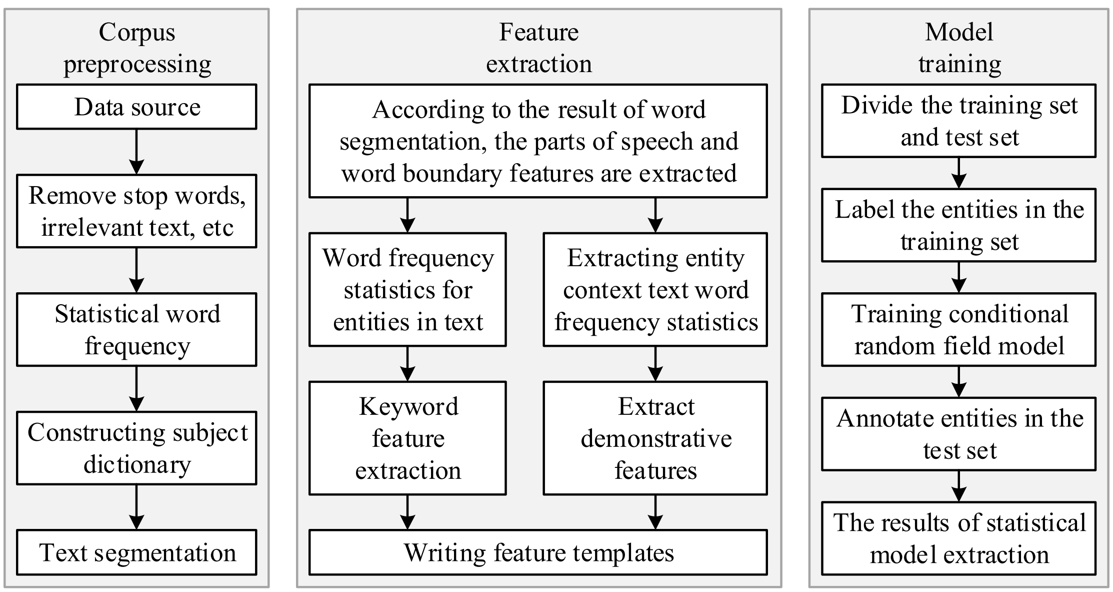

Figure 1: Entity recognition framework and process

II. A. 3) Feature extraction for linear chain conditional random field models

In this chapter, domain lexicon, lexical features, word boundary features, indicator features, keyword features, transfer features and context window features are extracted while training the CRF model. Lexical features i.e., the lexical properties of each word in the labeled text. Since in general the extracted entities are nouns, but some of the entities extracted in this chapter are combinations of English and numbers, lexicality can be used as a key piece of information to distinguish entities from non-entities.

There are three choices for the description of word boundaries, which are two-word position labeling set, four-word position labeling set and six-word position labeling set.

In this chapter, the four-word position description method is selected. This feature provides effective information about the position of entity words since it preserves the positional relationships between words in each word and between different words.

The construction of domain lexicon is important for domain-specific named entity recognition, which relates to the accuracy of textual disambiguation and adds a priori knowledge for named entity recognition. Based on the curriculum standards, this chapter uses TF-IDF statistical word frequency method to obtain high-frequency occurrence of subject words for different types of entities extracted, and constructs domain dictionaries through subject words to assist Chinese word separation.

II. B.Feature Enhancement Based Multi-Categorical Entity Relationship Extraction

II. B. 1) Entity Relationship Extraction Algorithm Framework

In the process of constructing the knowledge graph [24], the links between entities can be obtained from the corpus through entity relationship extraction to form a net-like knowledge structure system. The extraction of relationships between knowledge points is accomplished from the corpus related to the experimental courses, so as to construct the knowledge graph of the experimental subjects in secondary schools.

The entity-relationship extraction framework designed in this chapter is mainly divided into three parts, which are corpus preprocessing, feature extraction and relationship multi-classification module, and the entity-relationship extraction framework is shown in Figure 2.

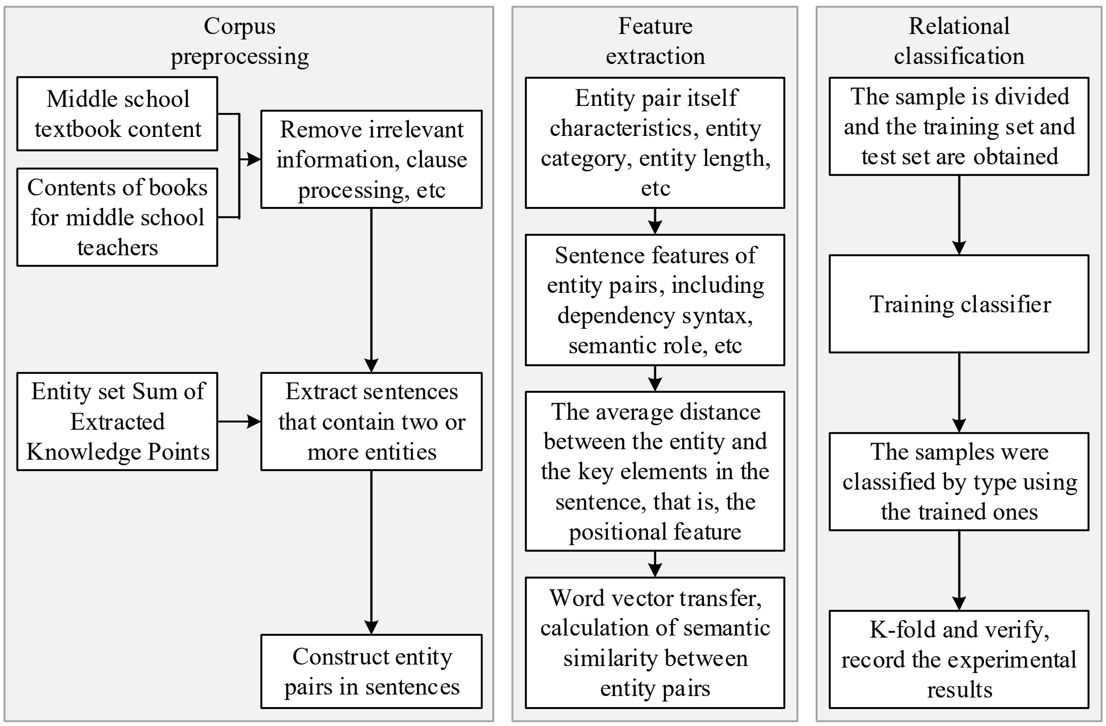

Figure 2: Physical relationship extraction framework

Most of the feature-based relation extraction classifiers use SVM, and in order to get better classification results, KNN classifier is also selected in this paper. Both types of classifiers are suitable for text classification, especially can get better classification effect on small sample training set. In addition, the KNN classifier only collects samples in the training model stage, and then carries out the classification process when it receives the test samples, so the time overhead of model training is zero, and the classification efficiency is high. Therefore, a two-layer machine learning classification approach was used for the extraction of relationships between knowledge points.

II. B. 2) Feature Extraction Based on Relational Multi-Classification Models

Based on the above analysis, in this paper, we will extract features from four aspects to train the multi-classifier: basic features of entity pairs syntactic semantic features, positional features and word vector based features.

Features of the entity pair itself. The features of the entity pairs themselves contain the basic information of the entities. The features of the entity pairs themselves selected in this chapter are the average length of the entity pairs, the distance between the entities, and the type of the entities, respectively.

Characteristics of the sentence in which the entity pair is located. Unlike the features of the entity pairs themselves, the sentence level features obtain long distance dependencies between words from the perspective of human understanding of natural language. In this paper, positional features are considered to enhance the categorization of entity relationships.

Word vector-based features show that the word vectors of similar words in the text are closer in the vector space, and richer semantic information of words can be obtained in the low-dimensional vector space. In this paper, the word vector model trained by textbook text is used to calculate the cosine similarity between word vectors of entity pairs in the vector space, and the similarity between knowledge points is used as a feature to train the classifier. The similarity is calculated as shown in the following equation, where w_i and w_j are two entities that are not the same in the text.

$$Sim(w_i, w_j) = \frac{w_i \cdot w_j}{|w_i| \cdot |w_j|} \tag{4}$$

There are two commonly used word2vec models, the CBOW model and the Skip-gram model.The CBOW model has the advantage of being efficient, however, it ignores the order of the context words, making this representation less than reasonable.The Skip-gram model, on the other hand, models the posterior probability of a target word to a context word by constantly adjusting the vector of context words through the target word, i.e., the solution formula:

$$p(e_{context} | w_m) = \prod_{w_i \in C} p(w_i | w_m) \tag{5}$$

Compared to the CBOW model, the Skip-gram model is less efficient but more accurate and more suitable for small sample training. Therefore, the Skip-gram model is used to train the word vector model in this section.

II. C.Entity Recognition Experiment Results and Analysis

II. C. 1) Data sets and evaluation indicators

In this chapter, subject domain datasets are used to validate the performance of the model, the datasets are obtained from the existing math textbook materials in the context of co-curricular heterogeneity as well as through crawlers, and the datasets are manually labeled, expert quality control, and the ratio of the division of the training set, the test set and the validation set is 8:1:1.

In this chapter, Precision (P), Recall (R) and FI values are used to consider the accuracy and completeness of model recognition in a comprehensive way, which can be referred to the following equation:

$$P = \frac{T_p}{T_p + F_p} \times 100\% \tag{6}$$

$$R = \frac{T_P}{T_P + F_N} \times 100\% \tag{7}$$

$$F1 = \frac{2PR}{P+R} \times 100\% \tag{8}$$

where F_p denotes the number of samples that are actually negative but that the model incorrectly predicts as positive, T_p denotes the number of samples that are actually positive and that the model correctly predicts as positive, and F_N denotes the number of samples that the model incorrectly predicts as negative when they are truly positive.

II. C. 2) Experimental results

The comparison models used in this chapter are specifically the following:

BiLSTM-CRF: Sentence context semantic features are extracted by BiLSTM, and CRF learns potential relationships between sequences and excludes labeled sequences that do not conform to the rules.

IDCNN-CRF: Sentence semantic features are extracted by IDCNN and CRF predicts labeled sequences.

BiLSTM-IDCNN-CRF: the sentence is modeled by BiLSTM, the obtained features are sent to the IDCNN layer, and finally the prediction results are corrected by CRF.

BERT-BiLSTM-CRF: a commonly used sequence annotation model that uses BERT model for pre-training and inputs word vectors into BiLSTM to extract features, and finally the prediction results are corrected by CRF.

BERT-BiLSTM-IDCNN-CRF: This model uses the BERT model for pre-training and then inputs the trained word vectors into BiLSTM and IDCNN to extract features and combine them, and finally the prediction results are corrected by CRF.

LEBERT-BiLSTM-IDCNN-CRF: lexical information is fused into the sublayer of Transformer on the basis of BERT-BiLSTM-IDCNN-CRF.

SVR-BiGRU-CRF: The embedding layer uses word, lexical and positional information, and the coding layer uses a bi-directional gated recurrent unit with a combined CRF model.

The advantages and disadvantages of the proposed methods are analyzed by comparing the entity recognition results of the seven models. Table 1 shows the experimental comparison results of the seven models.

Although the number of parameters of the model proposed in this chapter increases, its precision rate, recall rate and F1 value reach the optimum, and compared with the better-performing SVR-BiGRU-CRF model, the precision rate, recall rate and F1 value of this paper's model are improved by 0.97, 0.71 and 0.99 percentage points, respectively, which effectively reduces the model's underreporting rate.

Table 1: Comparison of the target values of different models

Model	P	R	F1
BiLSTM-CRF	83.55	86.95	85.04
IDCNN-CRF	86.62	86.07	86.01
BiLSTM-IDCNN-CRF	86.67	88.12	86.99
BERT-BiLSTM-CRF	87.38	88.96	88.09
BERT-BiLSTM-IDCNN-CRF	88.53	88.98	88.96
LEBERT-BiLSTM-IDCNN-CRF	90.03	90.01	90.09
SVR-BiGRU-CRF	90.61	91.25	90.99
This model	91.58	91.96	91.98

II. D. Entity Relationship Extraction Experiment

II. D. 1) Data sets and evaluation indicators

Existing co-curricular heterogeneous mathematics subject data, after entity identification, slicing the entity-containing data into individual sentences and manually annotating them to determine the type of relationship between the entities in each sentence, results in a dataset that can be used for training relational extraction models. This chapter sets the data structure discipline dataset to 11 relationships.

The following metrics are used as evaluation metrics in this chapter:

Precision-Recall (P-R) curve: plotted according to different classification thresholds and used to evaluate the performance of the classification model, each point on the P-R curve represents the precision and recall under a classification threshold.

P@N: represents the accuracy of the first N predictions, which is used to measure the performance of the model in a given ranking range.

AUC value: the area under the P-R curve, if one P-R curve completely wraps around the other P-R curve, the learner represented by the former is considered to have better performance, when the two appear to cross, the size of the AUC value is generally used for comparison, the higher the AUC value, the better the classification effect.

II. D. 2) Experimental results

To validate the performance of the models in this chapter, the results are compared with those of the following four models and the advantages and disadvantages of the proposed methods are analyzed.

PCNN+MIL: Combining multi-instance learning in PCNN to cope with the data noise problem.

PCNN+ATT: Add sentence-level attention mechanism to assign weights to sentence vectors to cope with the data noise problem.

BGWA: Introduces attention at word level and sentence level respectively and uses BiGRU to extract sentence features.

EMSA+PCNN: uses multi-head self-attention to compute and assign weights for word-level attention allocation, and uses the semantic representations of the head and tail entities as the basis for weight allocation for sentence-level attention allocation.

Figure 3 shows the P-R curves of different models, Table 2 shows the comparison of AUC values and P@N and P@M values.

The P-R curves of this paper's method can maintain high precision and recall under different thresholds, which verifies the effectiveness of this paper's method in solving the problem of frequent mislabeling in remote entity relationship extraction. The P@N, P@M and AUC values of this paper's method are significantly improved compared with the better-performing EMSA+PCNN, and the P@M and AUC values are improved by 2.8 and 2.4 percentage points, respectively, compared with the EMSA+PCNN model, which indicates that this paper's entity-relationship extraction model is able to better capture the key semantic information in the input sequences, and alleviate the impact of packet-level noise on classification. .

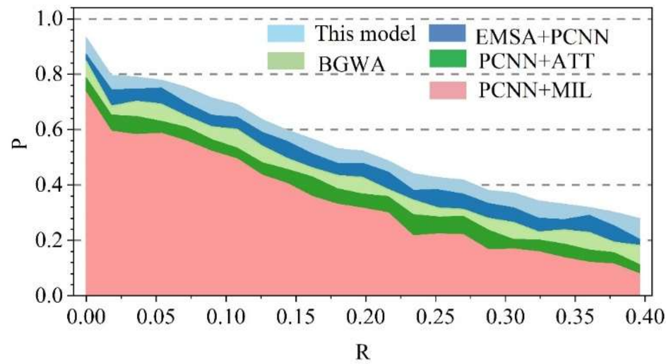

Figure 3: The different models are in the P-R curve of the class data set

Table 2: Comparison of indexes of different relationship extraction models

Model	P@100	P@200	P@300	P@M	AUC
PCNN+MIL	66.5	61.5	57.7	61.7	32.5
PCNN+ATT	68.9	63.5	59.9	63.6	34.6
BGWA	74.0	71.5	68.6	71.4	37.8
EMSA+PCNN	84.1	78.2	72.3	77.6	38.5
This model	84.7	81.4	77.5	80.4	40.9

III. Knowledge mapping of co-curricular courses

III. A. Neo4j graph database

Most of the knowledge graph is based on the structure of the graph, the current knowledge graph storage methods are mainly two: RDF format storage and graph database.RDF format to store data is in the form of ternary groups, such as Google's open Freebase knowledge graph, is in the form of text line by line to store the ternary groups. Through experiments and various forms of use found that the method of graph database to be more more general, and now more typical open source graph database is Neo4j. In this paper, we will use the Neo4j graph database for the construction of knowledge map, the mathematical knowledge points and the relationship between the knowledge points obtained earlier imported into the database, to achieve the construction of the knowledge map.

Neo4j graph database is a non-relational graph database, which is mainly in the form of a graph, to represent the nodes and the relationships between the nodes. Where nodes represent entities and relationships represent relationships between entities. Showing the correlation between data in the form of a graph allows one to visualize the structure of that knowledge system. Like other non-relational databases, Neo4j has efficient query capabilities.

Nodes and their relationships are stored in the Neo4j graph database to express rich entity content.Neo4j is relatively more suitable for storing semi-structured and unstructured data. Compared with some relational databases, there is an absolute advantage.Neo4j database query performance is very high, it reflects the association of entities itself, the creation of nodes corresponding to the construction of the connection.

Neo4j graph databases are widely used in several applications, such as: recommendation engines, community networks and so on. In addition, Neo4j graph databases are used in medical, financial and educational fields to store data and visualize the knowledge graph, which can better manage the associated data. And in the process of large-scale data growth, it can also have better adaptability.

III. B. Knowledge graph visualization

Knowledge mapping focuses on entity identification and relationship identification. In the drawing of knowledge mapping in the subject of co-curricular mathematics, what we need is the mathematical knowledge points and the relationships between them. In this paper, we mainly use the obtained mathematical knowledge points and relationships between knowledge points to import them into Neo4j graphical database to store and draw mathematical knowledge mapping. The process of knowledge mapping is shown in Figure 4:

What is shown in the figure is a part of the mathematical knowledge mapping, which contains some knowledge points and the relationship between them. From the figure, it can be seen that a knowledge point are with more than one knowledge point can establish a relationship between them, indicating that there is a great connection between the various knowledge points, after mapping the relationship between mathematical knowledge points and knowledge points, whether it is a teacher, a student or another researcher, as long as you want to get the relationship between the various knowledge points of mathematics, you can get it through the mapping query. Through the established knowledge map, the whole knowledge structure of mathematics is more expressive.

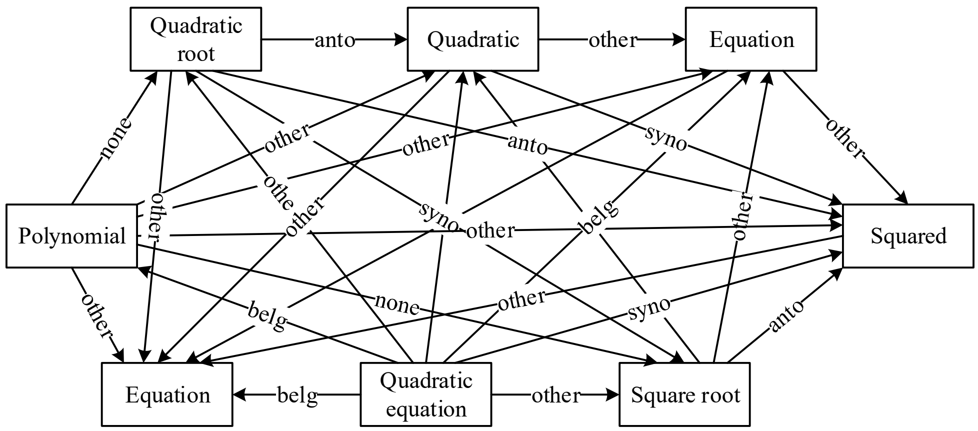

Figure 4: Visual display of some mathematical knowledge graphs

IV. Knowledge graph-based learner profiling and learning recommendation

IV. A. Learner portrait construction based on subject knowledge mapping

Learner portrait construction firstly needs to design the learner portrait model. Secondly, it is necessary to collect learners' learning big data according to the learner portrait model, and carry out data pre-processing such as data cleaning, data transformation, data normalization, etc. on the learning data. Once again, intelligent technology is used to calculate and model the learning data in order to accurately portray the learner's learning situation, and finally, visualization technology is used to present the learner portrait.

Based on the knowledge system, problem system and ability system in subject knowledge mapping, this study collects learners' learning characteristic data, learning behavior data and learning result data, and proposes the process framework for constructing learner portraits in the E-GPPE-C model in four aspects, namely, learner portrait structural design, data collection and preprocessing, learner portrait generation, and learner portrait output.

By analyzing the literature on smart learning and learner portrait, this study designs the learner portrait structure in the E-GPPE-C model from three dimensions: basic feature attributes, learning behavior attributes, and learning result attributes.

Learning data collection and preprocessing are the basis of learner portrait construction, which is described in the following from two aspects: learning data collection and learning data preprocessing.

Learner portrait generation refers to the process of acquiring learners' learning big data by using intelligent technologies, learning terminals and learning systems, and analyzing and calculating the data so as to describe the learners in multiple dimensions, and the specific generation process is shown in Figure 5.

Learner portrait label portrayal based on learning process data mainly refers to the process of labeling learning behavior data such as learning styles, cognitive levels and learning styles. In this study, L1 regularized SVM clustering algorithm is used to filter learners' learning behavior data, generate sparse weight matrix, and achieve feature selection for labeling learning style data.

The learning outcome data mainly come from regional monitoring data, school-level monitoring data, regular teaching and independent practice data, which are acquired mainly from test questions.

It is known that a monitoring data matrix $Y=(y_1,y_2,y_3,...,y_N)$, Y is N*J dimensional, N means there are N students and J means there are J questions. The initial values of each parameter Ks_0,Mr_0,Td_0,Tt_0,Tc_0 are obtained, and Ll_0 is computed as shown in the following equation:

$$Ll_0=\sum_{k=1}^{K} f(Ks_0,Mr_0,Td_0,Tt_0,Tc_0) \tag{9}$$

Taking into account the reliability r and degree d, Ll is calculated and the formula is shown below:

$$Ll = L(Ll_0, r, d) \tag{10}$$

Calculate the similarity sim between Ll and the Subject Knowledge Graph KG with the formula shown below:

$$sim(Ll, KG) = (\sum_{i=1}^{n} | Ll_i - KG_i |^p)^{1/p} \tag{11}$$

If sim>threshold, output knowledge mastery KD, problem mastery QD, ability mastery AD. Combined with the learner's personality characteristics, output the student's personality knowledge map PKS, the calculation formula is shown below:

$$\begin{aligned} PKS &= \sum (KD_i, QD_i, AD_i) \\ &= \{KD_1, KD_2, ..., KD_i\} \cup \{QD_1, QD_2, ..., QD_i\} \cup \{AD_1, AD_2, ..., AD_i\} \end{aligned} \tag{12}$$

In this study, the Echarts tool was used to visualize and present the learner portraits.

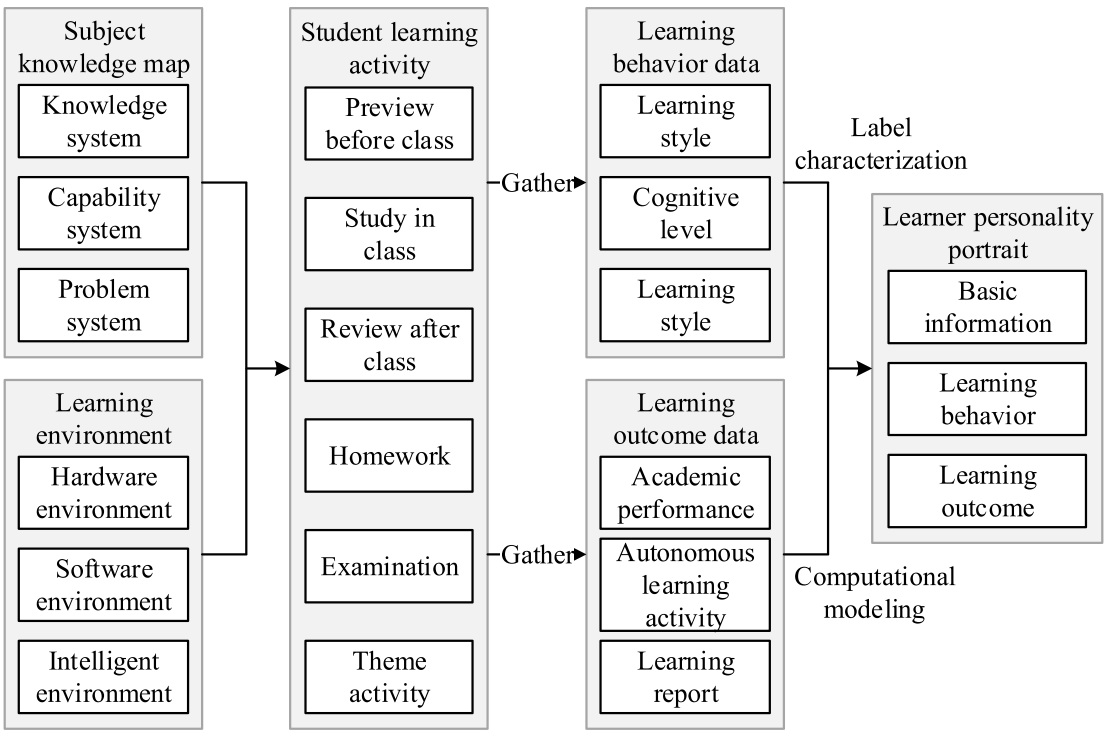

Figure 5: Process of learner portrait generation in the E-PEPPE-C model

IV. B. Learning path recommendation based on subject knowledge mapping

This study carries out learning path recommendation based on subject knowledge mapping and learner portrait, and introduces learner behavioral data into learning path recommendation, and fully considers the intrinsic dependency relationship between learner's cognitive structure and knowledge, in order to improve the accuracy and appropriateness of learning path recommendation.

The learning path recommendation in the E-GPPE-C model mainly includes four stages: (1) current cognitive state annotation, according to the learner learning big data, the Hidden Markov Model [25] is used to annotate the current cognitive state of the group learners on the subject knowledge graph. (2) Target learning state annotation, according to the learning objectives, using the Hidden Markov Model [26] to annotate the group learners' target learning state on the subject knowledge graph. (3) Group learning path planning, according to the current cognitive state and target learning state of group students, use the improved ant colony optimization algorithm [27] to plan the learning path of group learners. (4) Personalized learning path recommendation, according to learner characteristics, using the improved convolutional neural network (CNN) to recommend personalized learning paths for different learners to meet their needs.

In this study, the group learning path is based on learner group portrait and subject knowledge mapping to refine a simple sequence of learning content and learning activities that meets the learning needs of most learners and is applicable to most learners. In this study, an improved ant colony optimization algorithm is used to mine the optimal paths in the candidate set of learning paths, and the ant colony quickly searches for the optimal solution. For this reason, this paper uses the improved ant colony algorithm Diikstra combined with the ant colony algorithm for personalized learning path recommendation.

According to the planned group learning path, the improved convolutional neural network (CNN) is used to train the basic structure of the neural network of personalized learning path and generate the personalized learning path.

There is a problem of gradient vanishing by constituting more network layers. To solve this problem, in this study, based on the LeNet-5 network structure, a normalization layer is added behind the convolutional layer to improve the learning speed of the network. In this study, the improved convolutional neural network consists of an input layer, a hidden layer and an output layer, the input layer is the group learner learning path, the hidden layer includes four parts, namely, convolutional layer, normalization layer, pooling layer and fully connected layer, and the output layer is the personalized learning path.

1) Input layer, the input layer in this study is the group learning path.

2) Convolutional layer, the function of the convolutional layer is to extract features from the input data, and the convolutional kernel calculation formula is shown in (13):

$$\begin{aligned} Z^{l+1}(i,j) &= \left[Z^l \times w^{l+1}\right](i,j)+b \\ &= \sum_{k=1}^{K_l}\sum_{x=1}^{f}\sum_{y=1}^{f}\left[Z_k^l(s_0 i+x, s_0 j+y)w_k^{l+1}(x,y)\right]+b \end{aligned} \tag{13}$$

The summation part of Eq. is equivalent to solving a cross-correlation, where b is the bias, Z^l and Z^{l+1} denote the convolutional inputs and outputs of the $l+1$th layer, $Z(i,j)$ corresponds to the pixels of the feature map, K is the number of channels of the feature map, and f, s_0, p are the parameters of the convolutional layers, which correspond to the convolutional kernel size, convolutional step size and number of filled layers.

3) Normalization layer, the normalization layer is after the convolution layer, before entering the next layer, the input of the previous layer will be normalized, which can speed up the network training, effectively and prevent the gradient disappearance. The formula is shown below:

$$\rho = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{14}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\rho)^2 \tag{15}$$

$$Y_i = \frac{X_i-\tau}{\sqrt{\sigma^2+\mu}} \tag{16}$$

$$Y_i^{'} = \alpha Y_i + \beta \tag{17}$$

Where ρ denotes the mean, X represents the original data, n denotes the number of samples, σ^2 is the variance, μ denotes the constant, and α and β are the learning parameters.

4) Pooling layer, the function of the pooling layer in this study is to convert the results of individual points in the feature map into statistics of their neighboring regions. The Lp pooling model is used in this study and its basic representation formula is shown below:

$$A_k^l(i,j) = \left[\sum_{x=1}^{f}\sum_{y=1}^{f} A_k^l(s_0 i+x, s_0 j+y)^p\right]^{\frac{1}{p}} \tag{18}$$

5) Fully-connected layer, the fully-connected layer is located in the last part of the hidden layer of the convolutional neural network and transmits signals to other fully-connected layers.

6) Output layer, the output layer in this study uses the classical logistic regression model to generate personalized learning paths.

V. Experimental results of learning path recommendation

V. A. Data set preparation

As science disciplines are rigorously logical and systematic. Compared with liberal arts disciplines, science disciplines pay more attention to the exploration of objective laws and scientific principles, and their knowledge systems are usually clearer, more precise, and have stronger logical connections. This logic and systematicity makes it easier for the knowledge of science disciplines to be abstracted and formalized, and then integrated into the knowledge map. Therefore, in this paper, the knowledge points of the data structure course are selected as the experimental objects for the optimization of local learning path recommendation for co-curricular and heterogeneous course resources.

In order to further compare the models proposed in this paper and verify the reasonableness and effectiveness of the algorithms proposed in this paper, the knowledge points of Mathematical Analysis, Higher Algebra, and Data Structures and Algorithms are selected as the experimental objects of learning path recommendation under different methods, and the methods of comparison are the collaborative filtering-based recommender algorithm (CFR), the association rule-based recommender algorithm (ARR), and the deep learning model-based recommender algorithm (ARR), respectively. Deep Learning Modeling recommendation algorithm (DLRM).

V. B. Experimental results of local learning path recommendation

In order to make the algorithm play a better effect in learning path planning, this paper refers to the commonly used parameter settings of the ant colony algorithm to determine the parameter values, and the maximum number of iterations is 3000, and the iteration results are shown in Fig. 6, which shows that the optimal solution of the problem is found in this model when it is iterated about 870 times.

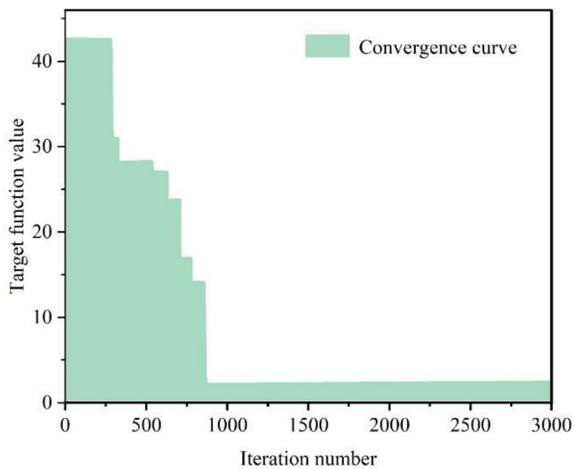

Figure 6: Iteration number

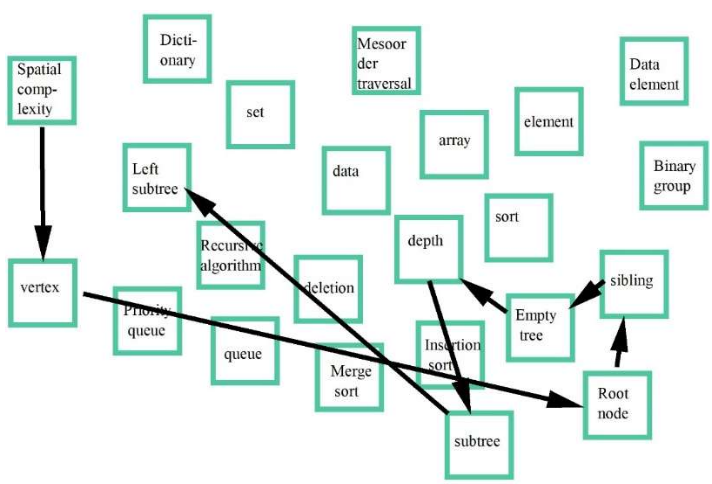

Figure 7: Local learning path diagram

Figure 7 illustrates the local details of the learning path planning process at the beginning stage. In the initial phase of learning path planning, concept nodes of lower difficulty are mainly covered. Although some of the concept nodes are significantly correlated with the application case nodes, these application case nodes are not included in the learning path in the initial stage of path planning. The reason for this is that there is a large difference in difficulty between the two, and the direct introduction of application cases may be beyond the learner's acceptance. The algorithms in this section follow the law of step-by-step education to ensure that the learning paths go from simple to complex, from basic to advanced, and gradually guide learners to master the relevant knowledge.

V. C. Comparative Experiments on Learning Path Recommendation by Different Methods

The extracted knowledge points are numbered to facilitate the entry of data, through the processing of knowledge point numbering, through the relationship between the extraction can be obtained knowledge point knowledge map, the use of software will be processed by the map, “Mathematical Analysis” knowledge point map, “Higher Algebra” knowledge point map and “Data Structures and Algorithms” knowledge point map were shown in Fig. 8 - Fig. 10, respectively.

In this paper, the reasonableness and effectiveness of the proposed algorithm are verified based on the experimental simulation of the three comparison methods, comparing the proposed model in this paper. After the experiments, the obtained knowledge mapping data are shown in Table 3 to Table 5 respectively.

For the knowledge point centrality of the Mathematical Analysis atlas it can be observed from the data that the maximum values of the knowledge point centrality of CFR, ARR, DLRM, and the method of this paper are A3, A1, A1, A1, and the minimum values are A4, A7, A3, and A7, respectively.

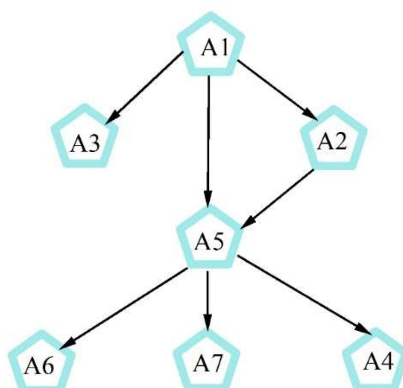

Figure 8: Knowledge map of “Mathematical Analysis”

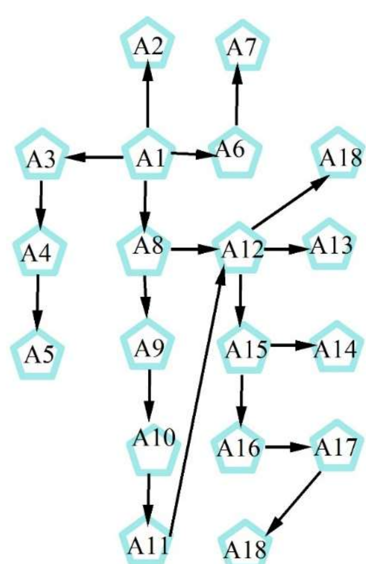

Figure 9: Knowledge Points of “Advanced Algebra”

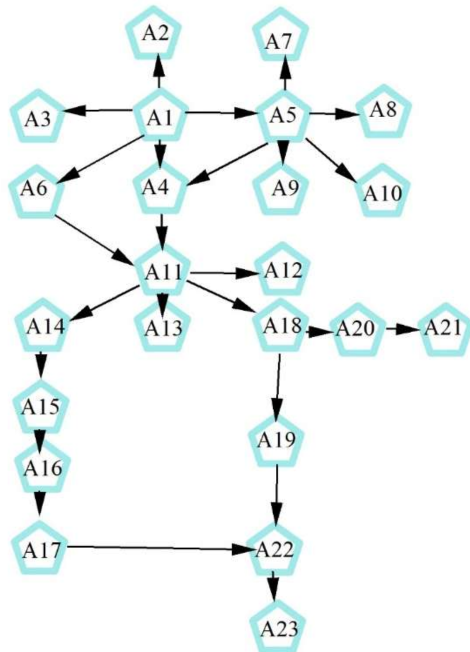

Figure 10: Knowledge Points of “Data Structure and Algorithm”

For Higher Algebra mapping the knowledge point centrality can be observed from the data that the maximum values of knowledge point centrality for CFR, ARR, DLRM as well as this paper are A10, A1, A1, A1 and the minimum values are A18, A18, A16, A18 respectively.

For Data Structures and Algorithms mapping the centrality of knowledge points can be observed from the data that the maximum values of centrality of knowledge points for CFR, ARR, DLRM and this paper are A5, A1, A1, A1 and the minimum values are A23, A8, A21, A23 respectively.

Table 3: The knowledge point center of the mathematical analysis

Serial number	CFR	ARR	DLRM	This method
A1	1.00	2.66	7.57	9.03
A2	0.88	2.01	1.94	5.37
A3	1.44	2.04	1.91	4.13
A4	0.39	0.40	3.28	3.97
A5	1.97	2.13	7.20	8.27
A6	0.58	0.41	3.30	3.95
A7	0.65	0.42	3.32	3.96

The data obtained through the experiment can only get the maximum and minimum of knowledge points can not be analyzed on the trend of knowledge points and the overall situation, so in order to intuitively compare the results of the four centrality of the comparison, this experiment uses the Excel statistical method of line graphs were plotted on three kinds of graphs in different algorithms, and the results obtained were as shown in Fig. 11 - Fig. 13, which can be clearly seen in the center of the node degree.

In the line graphs of the centrality of knowledge points in the mapping of Mathematical Analysis, the trends of DLRM and the method of this paper are more or less the same in terms of the overall trend. For the method of this paper, firstly, knowledge point A1 should be the knowledge point with the largest centrality, and secondly, the centrality of knowledge point A2 is much larger than that of knowledge point A3, so the method of this paper is more practical for this knowledge map.

Table 4: The knowledge point center of "higher algebra"

Serial number	CFR	ARR	DLRM	This method
A1	1.36	2.55	5.20	6.05
A2	0.94	0.63	2.63	4.26
A3	1.93	0.7	3.00	4.90
A4	0.97	0.96	2.15	3.22
A5	0.32	0.53	0.98	3.11
A6	1.63	0.88	1.98	4.86
A7	0.33	0.32	1.09	3.41
A8	1.52	1.31	3.46	5.71
A9	0.97	1.09	1.97	5.34
A10	1.99	0.91	1.63	5.31
A11	1.21	1.34	2.38	5.27
A12	1.30	2.64	5.22	4.85
A13	0.65	1.69	0.87	4.32
A14	1.42	1.97	0.77	3.25
A15	2.43	2.30	5.15	5.24
A16	1.23	2.08	0.72	3.24
A17	1.64	2.01	1.97	5.02
A18	0.38	0.29	0.93	2.86

Table 5: The knowledge point center of the data structure and algorithm

Serial number	CFR	ARR	DLRM	This method
A1	1.73	2.30	7.65	8.20
A2	1.01	0.65	3.28	6.26
A3	0.95	0.65	3.28	6.22
A4	1.13	1.91	4.26	7.43
A5	1.90	1.97	7.33	8.20
A6	1.54	1.10	2.22	6.96
A7	0.61	0.30	3.12	6.19
A8	0.62	0.33	3.09	6.20
A9	0.64	0.29	3.07	6.18
A10	0.60	0.31	3.16	6.18
A11	1.70	1.79	4.20	7.11
A12	0.96	0.21	2.47	5.09
A13	1.68	1.06	2.50	5.11
A14	1.18	1.60	3.61	6.10
A15	1.10	0.93	2.13	5.09
A16	0.77	0.95	2.00	4.40
A17	0.70	0.67	1.78	4.13
A18	1.25	1.92	3.85	4.98
A19	1.19	1.48	3.11	4.40
A20	1.12	0.98	1.68	3.20
A21	0.54	0.77	0.86	2.20
A22	0.90	1.27	2.29	1.94
A23	0.29	0.69	1.91	1.88

The line graph of the centrality of knowledge points in the mapping of Advanced Algebra shows that the method of this paper ensures that the centrality of knowledge point A1 is the largest, and the difference between the centrality of knowledge point A16 and that of knowledge point A14 is not much, and at the same time, it is in line with the principle of learning the knowledge points from the shallow to the deep, so that the method of this paper is more practical for this knowledge mapping.

In the line graph of the centrality of knowledge points in the mapping of Data Structures and Algorithms, both ARR and DLRM chose knowledge point A18 as the successor knowledge point of knowledge point A11, but compared to knowledge point A18, knowledge point A14 occupies a longer path in the whole knowledge map, so from the perspective of the whole and the path the centrality of knowledge point A14 is greater than the centrality of knowledge point A18, this paper method knowledge point A6 is smaller than the centrality of knowledge points A1 and A5, and in the selection of subsequent nodes, knowledge point A14 is chosen as the successor node to ensure the depth of learning knowledge.

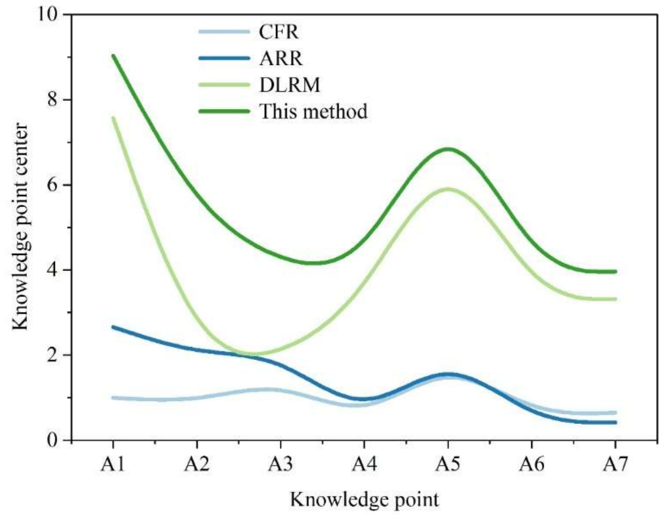

Figure 11: The knowledge point center of the mathematical analysis

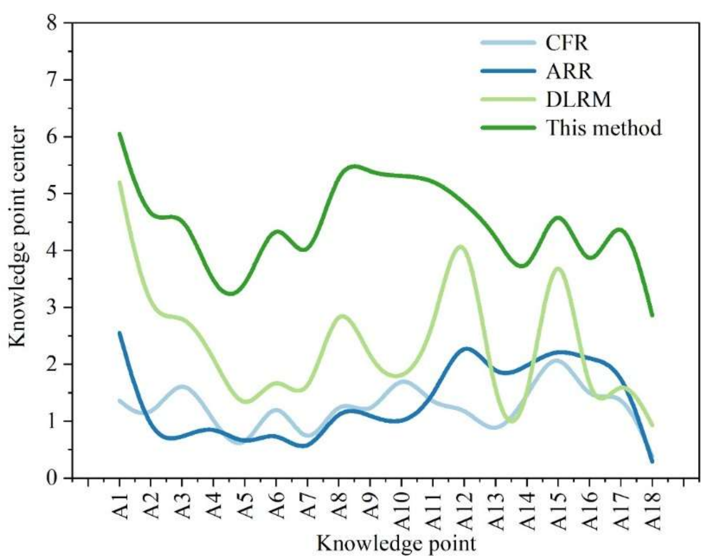

Figure 12: The knowledge point center of "higher algebra"

The optimal recommendation algorithm can be selected by comparing the three methods and the P@20 values in this paper, as shown in Table 6. From the table, it can be seen that the P@20 of the proposed method is 75.45, which is 6.90 higher than that of DLRM, and the recommended learning path algorithm generated in this paper is better than other algorithms.

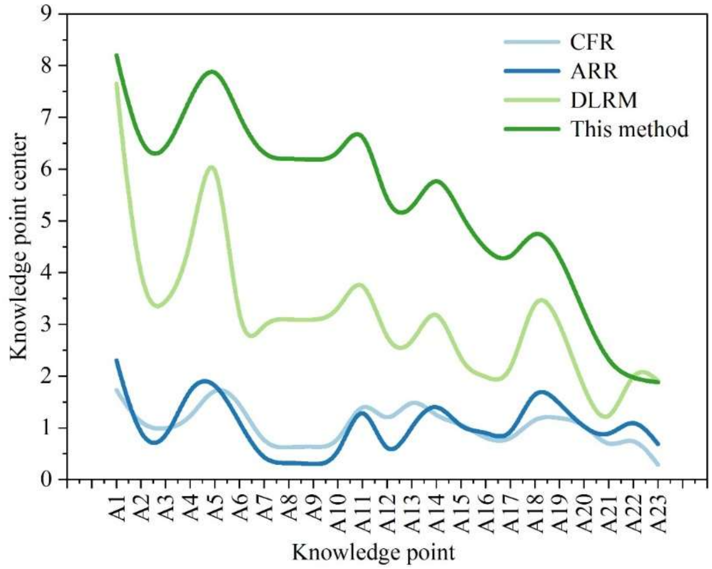

Figure 13: The knowledge point center of the data structure and algorithm

Table 6: P@20 value

Method	P@20
CFR	52.69
ARR	62.69
DLRM	68.55
This method	75.45

VI. Conclusion

This study focuses on the construction of knowledge graph for co-curricular courses to complete the portrayal of learner profiles and the recommendation of learning paths.

(1) In this paper, entity recognition is based on relational multi-classification model, the feature selection of CRF model in specific scenarios is studied, and entity relationship extraction is completed. In order to evaluate the effectiveness of the entity recognition model and entity-relationship extraction model, comparative experiments are conducted with multiple models, and the model in this paper gets better results in the entity recognition and relationship extraction experiments in mathematics subject under the same classroom context, respectively, and the precision, recall and F1 values of the entity recognition model in this paper are improved by 0.97, 0.71, and 0.99 percentage points, and the entity-relationship extraction model's P@ M and AUC values gained 2 and more percentage points compared to the better performing EMSA+PCNN model.

(2) On the basis of the subject knowledge graph constructed in the context of co-curricular heterogeneity in this paper, the method of learner portrait portrayal and learning path recommendation is proposed. The learning path recommended by the method in this paper guarantees the depth of learning knowledge, but also meets the law of gradual progress in education, gradually guiding learners to master the subject knowledge points, and obtains a P@20 value of 75.45, which is better and more reasonable than that recommended by the comparison algorithm.

(3) Although this study has obtained a more considerable effect of the optimized construction of the same-lesson heterogeneous course resources and the path recommendation algorithm design with the support of the knowledge graph, there are still shortcomings, and future research will further optimize the construction and updating mechanism of the knowledge graph to provide the possibility of real-time updating of the dynamic data.

Funding

Project Source: 2024 Beihua University Education and Teaching Reform Research Project. Project Title: Research on the Construction of "Same Course with Different Structures" Course Resources Based on Knowledge Graphs (Project Number: 21xjyb-1241732).

References

[1] Livingstone, K. A. (2019). The Place of Information and Communication Technologies in Curriculum Design and Development. International Journal of Education and Development Using Information and Communication Technology, 15(4), 180-197.

[2] Drozdikova-Zaripova, A. R., & Sabirova, E. G. (2020). Usage of Digital Educational Resources in Teaching Students with Application of" Flipped Classroom" Technology. Contemporary Educational Technology, 12(2).

[3] Al-Hunaiyyan, A., Al-Sharhan, S., Alainati, S., & Al-Duaij, M. (2023). A new competency model for digital instructors: Towards innovative teaching and learning in the digital era. Journal of Data Acquisition and Processing, 38(4), 1167-1186.

[4] Mukul, E., & Büyüközkan, G. (2023). Digital transformation in education: A systematic review of education 4.0. Technological forecasting and social change, 194, 122664.

[5] Uygarer, R., & Uzunboylu, H. (2017). An investigation of the digital teaching book compared to traditional books in distance education of teacher education programs. Eurasia Journal of Mathematics, Science and Technology Education, 13(8), 5365-5377.

[6] Al Abri, M., & Dabbagh, N. (2018). Open educational resources: A literature review. Journal of Mason Graduate Research, 6(1), 83-104.

[7] Caiyan, J. (2022). Design of an e-learning resource allocation model from the perspective of educational equity. International Journal of Emerging Technologies in Learning (iJET), 17(3), 50-67.

[8] Chao, Z. (2021). School-based curriculum construction-an important way to build digital learning resources. Frontiers in Educational Research, 4(7).

[9] Zhang, Y. (2025). Study on comprehensive management of online learning resources for digital education and teaching reform. International Journal of Networking and Virtual Organisations, 32(1-4), 36-48.

[10] Wang, E., & Sun, Z. (2025). Research on Personalized Course Resource Recommendation Method Based on GEMRec. Applied Sciences, 15(3), 1075.

[11] Sivtseva, N., Latysheva, S., Songolova, Z., & Solodkova, E. (2020). E-LEARNING RESOURCES AS A MEANS OF ORGANIZING STUDENTS'INDEPENDENT WORK. eLearning & Software for Education, 2.

[12] Vladimirovna, L. N., & Valentinovna, Z. A. (2018). Digital learning resources: Enhancing efficiency within blended higher education. Science for Education Today, 8(6), 121-137.

[13] Behrmann, A., & Bhattacharya, S. (2021, November). Resource Overload in STEM Education and How Students Can Overcome It. In 2021 ASEE Midwest Section Conference.

[14] Bol, L., Esqueda, M. C., Ryan, D., & Kimmel, S. C. (2022). A comparison of academic outcomes in courses taught with open educational resources and publisher content. Educational Researcher, 51(1), 17-26.

[15] Faustino, A., & Kaur, I. (2023). The Strengths and Drawbacks of E-resources in Higher Education. Asian Journal of Advanced Research and Reports, 17(8), 1-9.

[16] Zhang, H., Huang, T., Lv, Z., Liu, S., & Yang, H. (2019). MOOCRC: A highly accurate resource recommendation model for use in MOOC environments. Mobile Networks and Applications, 24, 34-46.

[17] Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., ... & Wahler, A. (2020). Introduction: what is a knowledge graph?. Knowledge graphs: Methodology, tools and selected use cases, 1-10.

[18] Huang, H., & Yang, X. (2024, September). Research and Implementation of Geography Curriculum Resource Construction Method Based on Knowledge Graph Technology. In 2024 14th International Conference on Information Technology in Medicine and Education (ITME) (pp. 804-808). IEEE.

[19] Li, L., Chen, A., Lu, H., & Zou, M. (2024, June). Research on Teaching Resource Aggregation Based on Artificial Intelligence and Knowledge Graph. In International Conference on Artificial Intelligence and Communication Technology (pp. 227-239). Singapore: Springer Nature Singapore.

[20] Niu, Y., Lin, R., & Xue, H. (2023). Research on learning resource recommendation based on knowledge graph and collaborative filtering. Applied Sciences, 13(19), 10933.

[21] Wei, Q., & Yao, X. (2022, January). Personalized recommendation of learning resources based on knowledge graph. In 2022 11th International Conference on Educational and Information Technology (ICEIT) (pp. 46-50). IEEE.

[22] Ma, H., Tang, Y., Zhang, X., Zhu, H., Huang, P., & Zhang, H. (2023). Learning resource recommendation via knowledge graphs and learning style clustering. Journal of Intelligent & Fuzzy Systems, 44(5), 8053-8069.

[23] Christian M. Hafner,Helmut Herwartz & Shu Wang. (2025). Statistical Identification of Independent Shocks with Kernel-based Maximum Likelihood Estimation and an Application to the Global Crude Oil Market. Journal of Business & Economic Statistics,43(2),423-438.

[24] Liu Hongbo,Siti Zobidah Omar,Chen He & Wang Shanshan. (2024). Construction of Image Education Knowledge Map Model Based on Data Mining Technology. Applied Mathematics and Nonlinear Sciences,9(1),

[25] Jianjie Guo,Lin Guo,Wenchao Xu & Haibin Zhang. (2025). Hidden Markov model with Pitman-Yor priors for probabilistic topic model. Communications in Statistics - Theory and Methods,54(9),2791-2805.

[26] Juan Wang & Lida An. (2025). Research on Learning Resource Recommendation Based on Latent Factor Model. Information and Knowledge Management,6(1),

[27] Sofiane Ben Abdi,Abderrazak Debilou,Lemya Guettal & Aicha Guergazi. (2025). Robust trajectory tracking control of a quadrotor under external disturbances and dynamic parameter uncertainties using a hybrid P-PID controller tuned with ant colony optimization. Aerospace Science and Technology,160,110053-110053.