Research on Intelligent Mining Methods of Multimedia Teaching Resources in Colleges and Universities in the Age of Big Data

Qi Yue[1], Zhu Xuan[*2]
[1]College of Chemistry, Changchun Normal University, Changchun 130032, China
[2]College of Literature, Changchun Normal University, Changchun130032, China
E-mail: zhuxuan-486@163.com
[*]Corresponding author

To solve the problem of difficult access to college multimedia resources in the era of big data, the present research proposes an intelligent mining method for college multimedia teaching resources. Collect multimedia teaching resource data from the Internet by using focused crawlers with the advantages of subject crawling and URL sorting; Use the methods of removing stop words, word segmentation and word frequency statistics to process the crawled data; Extract features from processed data; The features extracted by clustering analysis are classified. The number of categories is selected as the number of BP neural networks for combination, and the momentum method and learning rate adaptive adjustment strategy are introduced to improve the combined BP neural network. The extracted features are input into the improved combined BP neural network, and the intelligent mining results of university multimedia resources are output. The experimental results indicate that the focused crawler method can efficiently collect multimedia education resources in colleges and universities, and data preprocessing can efficiently reduce data redundancy. The double-feature extraction method can significantly enhance the recall and accuracy of data mining. It can realize the classified mining of multimedia teaching resources in academic centers and display the classified mining results of multimedia teaching resources in various disciplines.

Povzetek: Raziskava uvaja inteligentno metodo rudarjenja multimedijskih učnih virov z uporabo osredotočenih pajkov, obdelave podatkov in izboljšane BP nevronske mreže, kar povečuje učinkovitost in točnost rudarjenja podatkov.

1 Introduction

Big data [1], or huge amount of data, refers to the information that the amount of data involved is so large that it can't be captured, managed, processed and arranged in a reasonable time to help enterprises make more positive business decisions through mainstream software tools [2].

Data mining is a hot issue in artificial intelligence [3], database and other fields [4]. The so-called data mining refers to the extraordinary process of revealing hidden [5], previously unknown and potentially valuable information from a large amount of data in the database. The types of data can be structured [6], semi-structured [7] or even heterogeneous [8]. The object of data mining can be any type of data source. It can be a relational database [9], which is a data source containing structured data; It can also be data warehouse, text [10], multimedia data [11], spatial data, time series data and Web data, which contain semi-structured data or even heterogeneous data. Data mining is a decision support process, which is mainly based on artificial intelligence [12], machine learning, pattern recognition, statistics, database, visualization technology, etc. It analyzes the data of enterprises with high automation, makes inductive reasoning, mines potential patterns from them, and helps decision makers

adjust market strategies, reduce risks and make correct decisions. The process of knowledge discovery consists of the following three stages: data preparation, data mining, result expression and interpretation.

In colleges and universities, multimedia teaching has been realized, but the current multimedia teaching is not very ideal. One of the important factors is that the multimedia teaching resources in colleges and universities are few and single, and it is difficult to find rich multimedia teaching resources on the Internet because there are a lot of data on the Internet.

Many scholars have researched the mining of teaching resources, such as Varde A. Stichude's calculation and estimation of scientific data mining based on classical methods, the automatic implementation of scientists' learning strategies [13], and the integration of clustering complex teaching data into a framework, thus automating scientists' classical learning methods. The knowledge from the existing experimental database can be used as the basis for estimation. Challenges include maintaining domain semantics in clustering, finding the right strategy in classification, achieving a good balance between refinement and simplicity in presenting estimated results according to the target user's needs and getting objective metrics to capture users' subjective preferences. Thus, the mining of teaching data is completed. An educational data

mining system designed by Abu MM et al. to predict and improve the programming ability of college students [14], the system includes two main modules: classification and learning processes. The classification module predicts the current state of students, digs programming data on the Internet, and the learning process module generates corresponding suggestions and feedback to improve the quality of students. Especially for the classification module, a real dataset related to the task is prepared and evaluated to study six key machine learning (ML) algorithms, support vector machines (SVM), decision trees, artificial neural networks, random forests (RF), k-clusters and naive Bayesian classifiers, and use performance metrics and goodness of fit related to accuracy. Complete the excavation of programming teaching resources and evaluate students' programming ability. Zhao G et al. proposed the image network teaching resources retrieval algorithm based on the depth hash algorithm [15], constructed the pixel big data detection model of the multi-view attribute coded image network teaching resources, reconstructed the pixel information collected by the multi-view attribute coded image network teaching resources, and extracted the fuzzy information feature components of the multi-view attribute coded image. The combination of edge contour distribution images is done. The distributed fusion result of the image edge contour of the network-supervised resource view realizes the construction of a view feature parameter set. The gray matrix invariant feature analysis method is used to realize information encoding. The deep hashing algorithm is applied to retrieve teacher resources of the image network with multi-view attributes encoded, and the teacher resource of multi-view attributes is used. The hash coding result of the resource is realized. The encoded image network is used to achieve information reorganization to improve fusion. Finally, the retrieval and mining of teaching resources for images are completed. The above educational resource mining methods all have certain shortcomings, such as long data collection time, insufficient data collection breadth, low accuracy of data mining, long running time of data mining, etc.

Data mining can deal with big data and discover the hidden relationship of data. Therefore, this paper puts forward the research on intelligent mining method of multimedia teaching resources in colleges and universities in the era of big data, using focused crawler to collect multimedia teaching resources in colleges and universities, using word frequency statistics method to process data, then extracting data features, and finally using improved BP neural network to mine data features.

2 Intelligent mining method of multimedia teaching resources in colleges and universities

2.1 Collection of college multimedia teaching resources based on web crawler

Because of the massive data information on the Internet, how to quickly and accurately find the required multimedia teaching information on the Internet has become increasingly important. Search Engine is an online information retrieval tool [16], and a web crawler is a key part of a search engine. The search engine usually includes four parts: information collector (Robot), data indexer (Indexer), query searcher (Server) and result sorter (Ranker).

The specific workflow of the search engine is presented in Figure 1. It includes the following steps:

(1) Finding and obtaining web page information through web crawlers [17];

(2) Organize and process information and create an index database;

(3) Processing and sorting of search results;

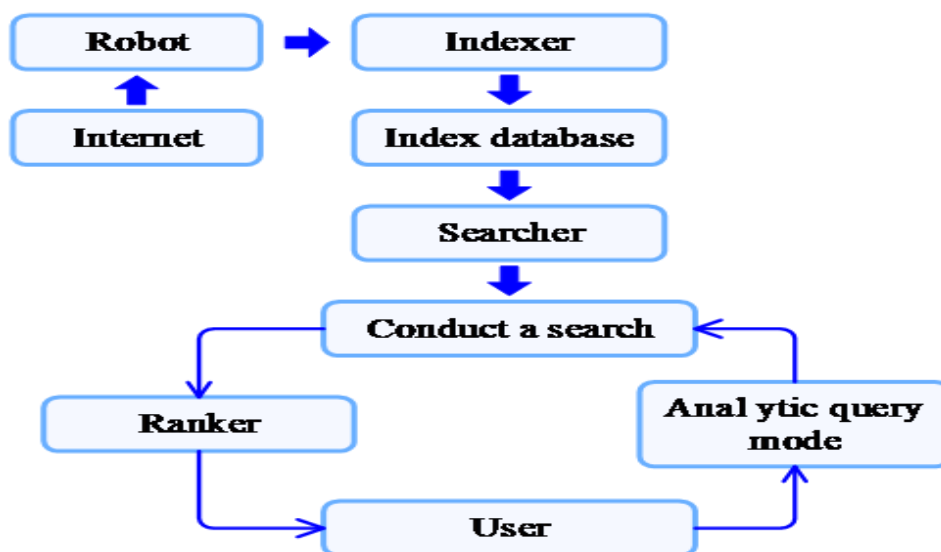(4) The retrieval result is returned.

Figure 1: Search engine workflow

However, if search is manually done for multimedia education resources, the work efficiency will be much higher. Therefore, the web crawlers can be used to collect multimedia education resources. In this paper, the focus is on the crawlers known as theme crawlers (or professional crawlers); that is, it is a "theme-oriented" web crawler program. The difference between it and what a crawler (universal crawler) is that a focused crawler is a target theme-driven and selective crawler. When implementing web crawling, college multimedia teaching theme screening is required. It tries to ensure that only the webpage information related to the multimedia teaching theme in colleges and universities is captured. According to the established target theme, the focused crawler selectively accesses the relevant pages on the Web. It pursues the precision of network information rather than the coverage of network resources.

Unlike ordinary web crawlers, the workflow of focused crawlers is more complex. It is essential to filter irrelevant links to the topic according to a specific web page analysis algorithm, retain useful links and put them into the URL queue waiting to be crawled. Then, it will select the next page URL [18] from the queue according to a certain search strategy and repeat the above process until a certain condition of the system is reached. The focused web crawler workflow is shown in Figure 2. In addition, all web pages captured by crawlers will be stored in the system [19], analyzed and filtered to a certain extent, and college multimedia teaching indexes will be established for users to query and retrieve later. For focused crawlers, the analysis results obtained in this process may also give feedback and guidance to the subsequent capture process. The focused crawler structure used in this paper is shown in Figure 3. The focused crawler of this structure first adds the related theme pages to the predefined categories of theme samples according to the predefined college multimedia teaching theme samples to train the college multimedia teaching theme samples. Classifier-based focused crawler includes two most important parts: one is a web page classifier, which is used to learn the features of crawling targets, calculate the relevance of web pages, and filter uncorrelated web pages [20]; The other is a web page selector, which is used to calculate the importance of a web page and dynamically determine the order in which the crawler accesses the web page according to the importance.
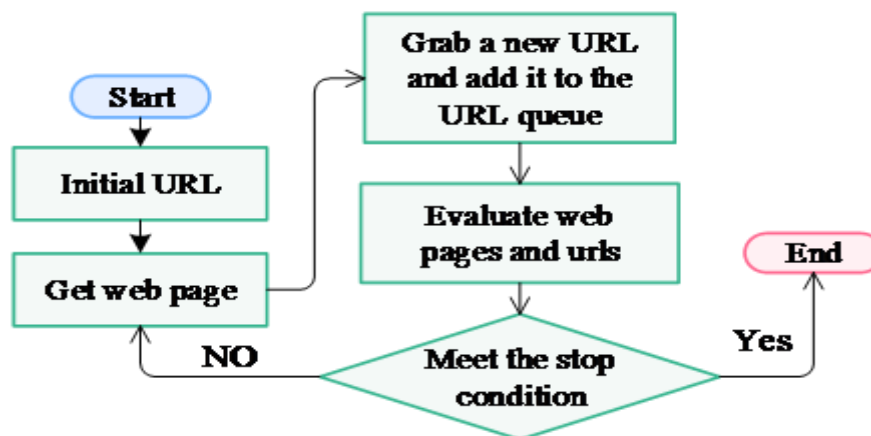

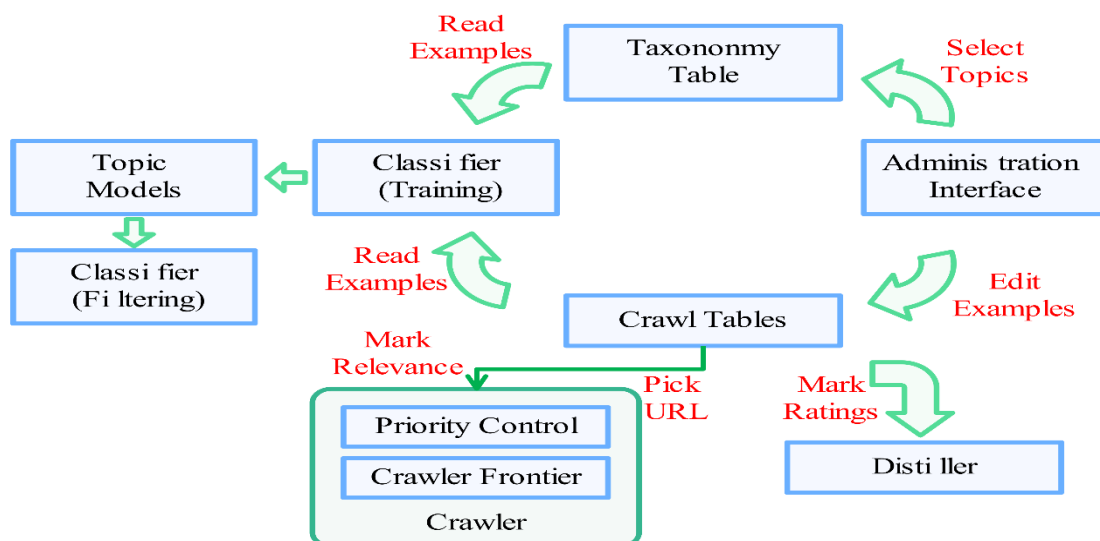Figure 2: focuses on the crawler works

Figure 3: Focuses on the crawler structure

2.2 Data processing of college multimedia teaching resources

Due to the huge scale of multimedia teaching resources in colleges and universities, the collection efficiency of traditional common crawlers is low, but focusing crawlers can obtain the required teaching resources more quickly and accurately, and improve the efficiency of data collection. There are data with low correlation in multimedia education resources in colleges and universities, which will lead to serious data redundancy if they are not processed. Therefore, deleting data with low correlation can reduce redundancy and improve data preprocessing effect. The college multimedia teaching resources collected by the focused web crawler are uniformly mapped into text form and are de disabled and word segmentation processing. On this basis, word frequency statistics are used to complete the processing of college multimedia data.

2.2.1 Deactivation word processing

A complete resource text contains a large number of meaningless words, such as the modal particle "ah", "na", and "ni". These words are collectively called stop words. The removal method is to traverse all texts according to the deactivated thesaurus library and delete all words in the text that appear in the deactivated thesaurus library.

2.2.2 Word segmentation

Word segmentation: the data text of college multimedia teaching resources is divided into thousands of words or phrases. Here, the bidirectional matching method is used for word segmentation. This particular process is as follows.

Step 1: Set the input string to be segmented as D, the output segmentation result is C ;

Step 2: Execute the forward maximum matching algorithm to get the segmentation result C_1. The specific process is as follows:

(1) Initialize dictionary Dic and set maximum segmentation c_{max} ;

(2) Judgment D whether it is empty. If yes, output the word segmentation result C_1; Otherwise, enter the next link;

(3) Comparison c_{max} and D length, specify the smaller value of the two, and record it as F ;

(4) Cut the length from the head substring to F, marked as $\widehat{D}$;

(5) Check the dictionary $\widehat{D}$ whether it is in the dictionary. If not, go to the next step; Otherwise, go to (7);

(6) Set $\widehat{D}$ remove the right word in and judge $\widehat{D}$ whether it is a single word. If yes, go to the next step; Otherwise, go back to (5);

(7) Order $C_1 = C_1 + \widehat{D} + "/"$, $D = D - \widehat{D}$, and go back to (2).

Step 3: Execute the reverse maximum matching algorithm to obtain the segmentation result C_2 ;

Step 4: Judgment C_1 and C_2 whether it is the same. If the same, make $C = C_1$ perhaps $C = C_2$ and jump to step 6; Otherwise, proceed to the next step.

Step 5: Judge C_1 and C_2 whether their lengths are the same. If they are the same, make $C = C_2$, take the segmentation result of the reverse maximum matching algorithm as the result, and skip to step 6; Otherwise, take the short value and assign it to C, skip to step 6.

Step 6: Output segmentation results C.

Therefore, the data segmentation of multimedia teaching resources in colleges and universities is completed.

2.2.3 Statistical processing of word frequency

In the university multimedia teaching resource data text document, the words of each frequency are distributed in a certain rule. The word frequency statistics method uses statistical knowledge to describe the word rules. Zipf's law and cloth's law are two laws that have far-reaching influence on word frequency statistics [21]. For many years, scholars in various fields have conducted in-depth research on the statistical law of word frequency, and many scholars favor this method because of its simplicity and practicality.

Zipf's law is described as [22]: given university multimedia teaching resource data text document as d, L represent text d length of (L Large enough), N_{diff} indicates that appears in d total number of different words in, TF_n express d word frequency of Chinese words(n is the number of occurrences of words in the text), TR_n represents and TF_n corresponding word rank, f_n indicates the frequency of words, $f_n = \frac{TE_n}{L}$, then:

$$f_n \times TR_n = K \tag{1}$$

$$TF_n \leq f_n \times L < TF_{n+1} \tag{2}$$

According to Zipf's law, the word frequency is TF_n number of words on the same frequency of $NTIF_n$ For:

$$NTIF_n = \frac{K \cdot L}{TF_n \cdot TF_{n+1}} \tag{3}$$

Formula (3) calculates the number of words with the same frequency $NTIF_n$ is not completely applicable to word frequency TF_n take any value, since it is based on Zipf's law, but Zipf's law cannot well reflect the distribution of words with extremely low word frequency $TF_n = 1,2$. The fluctuation is particularly obvious. Therefore, the maximum value method processes when $TF_n = 1,2$ words with the same time-frequency.

Word frequency is based on the maximum method $TF_n = 1,2$ number of words on the same frequency $NTIF_n$. The expression of is:

$$NTIF = \frac{N_{diff}}{TF_n \times TF_{n+1}}, n = 1,2 \tag{4}$$

Where, N_{diff} appears in the document for d the total number of different words in.

The simultaneous formula (3) and (4) get the number of words with the same frequency NTIF, and the complete expression is:

$$NTIF_n = \begin{cases} \dfrac{K \times L}{TF_n \times TF_{n+1}}, & n > 2 \\[2ex] \dfrac{N_{diff}}{TF_n \times TF_{n+1}}, & n = 1,2 \end{cases} \tag{5}$$

Among them, $K = \dfrac{1}{(\ln N_{diff} + \beta)}$ (β is Euler constant).

The text data preprocessing method of college multimedia teaching resource data based on the statistical rule of word frequency is as follows:

Step 1: Initialize the storage word frequency as $TF_n = 1$, $TF_n = 2$, $TF_n > 2$'s dictionary $dict1$, $dict2$, $dict3$ and corresponding counters recording the number of words with different word frequencies $count1$, $count2$, $count3$, definition word list $TermList$ and counter $word\_count$;

Step 2: Perform word segmentation and record the word frequency for every word;

Step 3: Classify based on various word frequencies and record the number of words for every frequency;

Step 4: Data preprocessing based on word frequency statistics;

Step 5: Select the word frequency with low frequency and low correlation to delete.

Step 6: Output the word sets with different word frequencies, the total number of words corresponding to each set, and the pre-processing list.

Finally, the data processing of multimedia teaching resources in colleges and universities is completed.

2.3 Data feature extraction of college multimedia teaching resources

Because the current methods for extracting text data features have different shortcomings, this paper uses BNS and Odds to extract features of college multimedia teaching resources data after preprocessing [23]. Using these two methods to extract features of college multimedia teaching resources data can not only complement the shortcomings of the other's methods. It can also further improve the accuracy of feature extraction.

The data feature extraction of multimedia teaching resources in colleges and universities can help to extract key information and remove low-frequency vocabulary, so as to obtain accurate and effective features, express the characteristics of multimedia data more fully, improve the accuracy and efficiency of data feature extraction, effectively solve the problems of data complexity and redundancy, improve the accuracy and efficiency of mining, and provide scientific and reliable technical support for intelligent mining of multimedia teaching resources in colleges and universities.

2.3.1 Data feature extraction method of college multimedia teaching resources

(1) BNS method

BNS is a new feature extraction algorithm applicable to the text classification operation of multimedia teaching resource data in colleges and universities. It measures and compares the significance of items concerning category distribution using the probability statistical method, and the formula is:

$$BNS(t) = \left| F^{-1}(t_{pr}) - F^{-1}(f_{pr}) \right|$$

$$t_{pr} = \frac{t_p}{t_p + f_n}$$

$$f_{pr} = \frac{f_p}{f_p + t_n} \tag{6}$$

Where, t indicates an entry, $BNS(t)$ indicates the BNS characteristic value of this term. $F(\ )$ the distribution function representing the standard normal distribution, t_p indicates the number of texts containing entries in the class, f_p refers to the number of texts containing entries outside the class, f_n denotes the number of texts without entries in the class, t_n represents the number of text without entries outside the class. When t_{pr} or f_{pr} is 0, define $F^{-1}(0) = 0.0005$.

BNS algorithm models the features in each college multimedia teaching resource data text with random normal distribution curve, and uses the limit area of the lower end point of the normal curve as a measure of the correlation degree of feature terms to the class. The greater the correlation degree of feature terms to the class, the farther the endpoint of the normal class is from the endpoint of the anti-class (as shown in Figure 4). BNS value is a measure of the difference between the two endpoints.
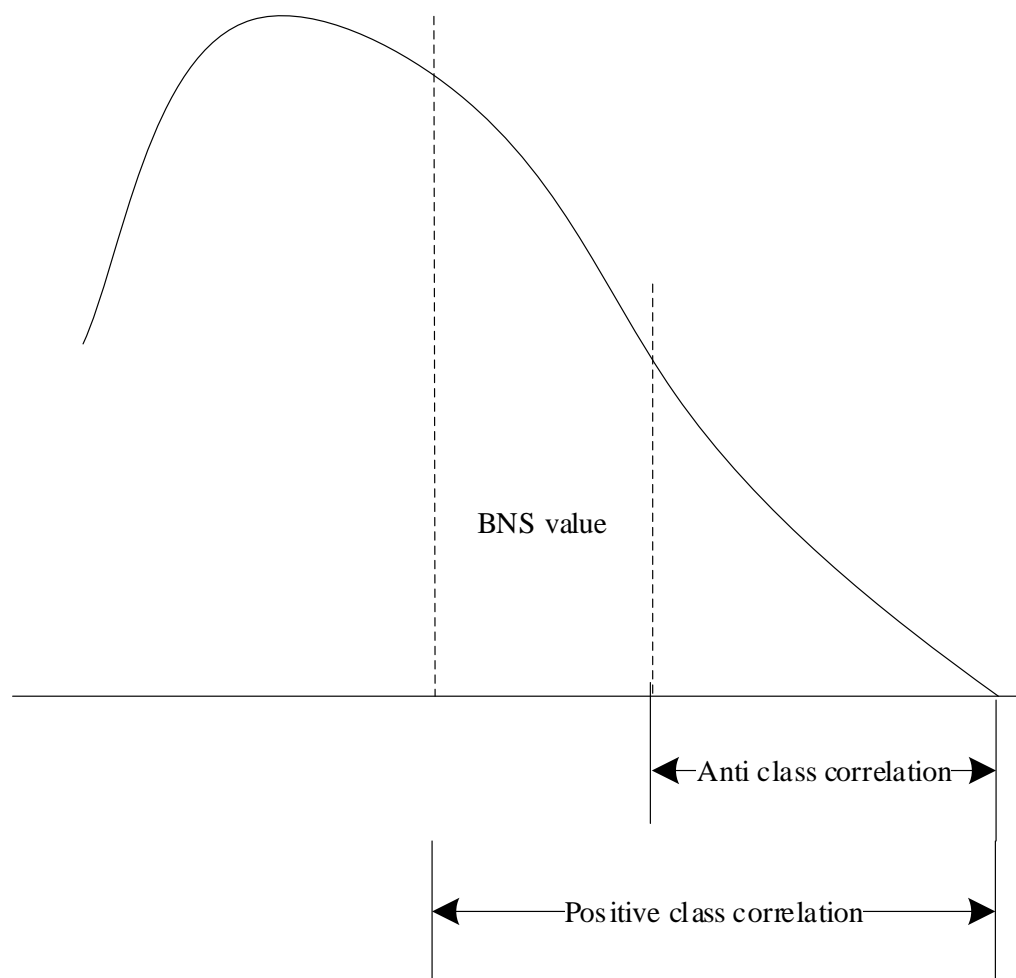
Figure 4: BNS feature extraction algorithm endpoint separation

The BNS algorithm can solve the problem of data set skew of multimedia teaching resources in colleges and universities and performs well in multi-step or combined feature selection, but the accuracy of classification results using the features obtained from the BNS algorithm could be better.

(2) Odds method

Odds mainly reflects the difference rate between the advantages of positive and negative terms in the text classification of multimedia teaching resources data in academic centers. The formula is:

$$Odds(t) = log \frac{P(W|pos)\big(1 - P(W|neg)\big)}{P(W|neg)\big(1 - P(W|pos)\big)} \qquad (7)$$

Where, t indicates an entry, $Odds(t)$ represents this term $Odds$ characteristic value. $P(W|pos)$ represents intraclass entries W the conditional probability of occurrence, $P(W|neg)$ denotes out of class entries W the conditional probability of occurrence.

$Odds$ its feature extraction algorithm does not treat all classes equally. It only cares about the target class value and recognizes as many positive classes as possible but does not care about anti-classes. It is suitable for binary classifiers. According to Mladenic and Grobelnik, $Odds$ its feature extraction algorithm is conducive to the information repair of other algorithms, so it can effectively supplement the shortcomings of other algorithms in the combination with other algorithms.

2.3.2 Data feature extraction process of college multimedia teaching resources

This paper uses two algorithms to extract and complement the text features of college multimedia teaching resource data. Based on maintaining the accuracy of feature extraction, the problem of college multimedia teaching resource data set skew is solved [24], and finally, the college multimedia teaching resource data text feature set with fewer dimensions is obtained. The method flow is shown in Figure 5.
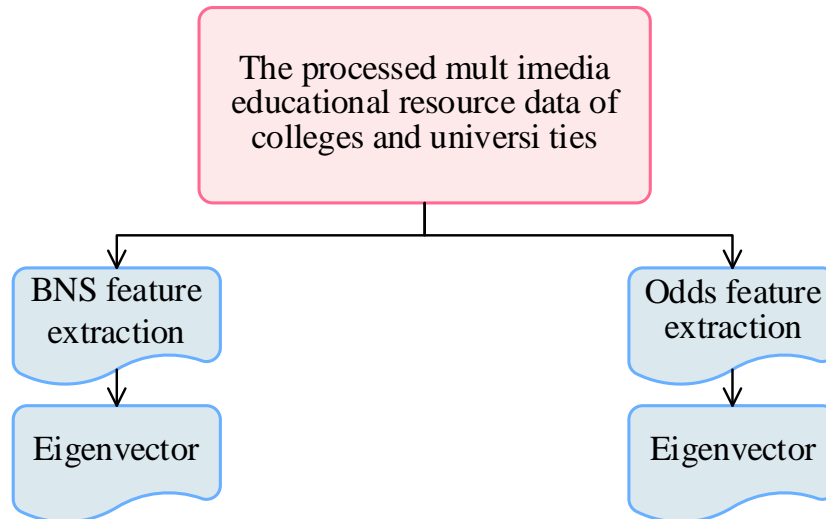
Figure 5: Algorithm flow

The especial steps of the algorithm include the following phases:

Input: Processed college multimedia teaching resource data.

Output: Term vector in the vector space model composed of feature terms.

Description: Text object $t(tNamr, *W, tNumber)$, entry object $w(wName, wNumber. tp, fp, tn, fn, BNS, Odds, BOS)$, $*T$ is a pointer to an array of text objects, $*W$ is a pointer to the term object array, $tName$ is the text name, $w\,Name$ is the content of the entry, $tNumber$ is the number of text objects, $wNumber$ is the number of entries.

Step 1: Initialize $*T$;

Step 2: Write the input college multimedia teaching resource data $*T$, T_n represents the n class, and calculate the tp values and fn value to set the hash table with entries as keywords for each class;

Step 3: Calculate the fp values and tn value;

Step 4: Use the BNS feature extraction algorithm to calculate the BNS feature value and Odds feature value of each term, and calculate the BNS value variance, Odds value variance, BNS value weight, and Odds value weight of each category;

Step 5: Extract the text features of college multimedia teaching resources data according to the calculated feature values and write them into the vector space model.

2.4 Data mining model of college multimedia teaching resources based on improved BP neural network

2.4.1 BP Neural network

BP neural network is a multi-layer feedforward neural network model trained based on the error backpropagation algorithm [25]. This model can learn and store many input-output mode mapping relations without explaining the mathematical formulas describing the mapping relationships in advance. Determine the model structure based on BP neural network, as shown in Figure 6.

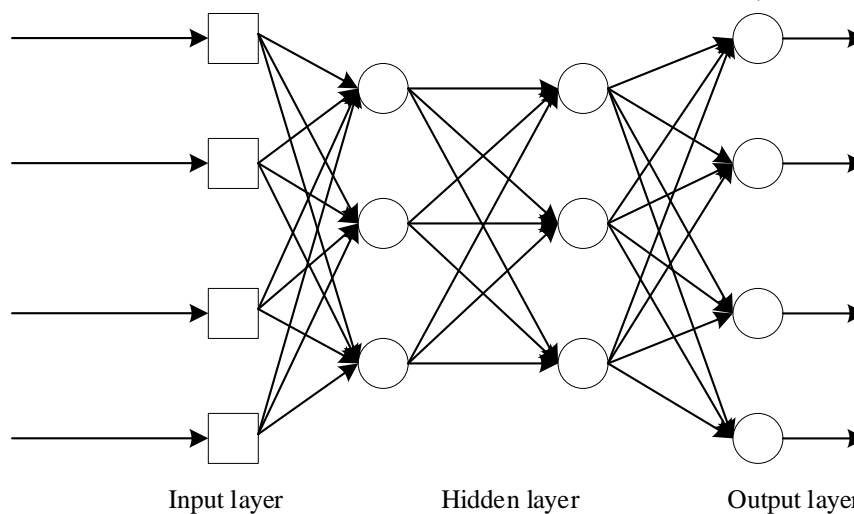Input layer   Hidden layer   Output layer

Figure 6: Model structure of BP neural network

As shown in Figure 6, in the BP neural network model, the final mining results are output through the three-layer architecture of input layer, hidden layer and output layer. The single-layer neuron structure of this model is expressed mathematically, as shown in formula (8):

$$y = f\left[\sum_{x=1}^{n_b} w_x x - \theta\right] \tag{8}$$

In the formula, y represents the output value of the development evaluation model; f represents the activation function of the model; x represents the initial index variable substituted into the model; w_x represents the weight of model neurons; θ represents the model threshold.

In Formula (8), the calculation method of the model activation function is shown in formula (9).

$$f(x) = \frac{1}{1 - e^{-2x}} - 1, (-1 < f(x) < 1) \tag{9}$$

In the formula, e stands for natural constant. The development evaluation model is trained, and the initial learning rate of the model is set to 0.1. Through repeated training, the output error of the model results is controlled between 1 and 5.

On this basis, and dynamically adjusts the network weight and threshold through backpropagation to obtain the minimum value of the sum of squares of errors. The BP algorithm is mainly based on the gradient method to establish the minimum quadratic performance index function and seek the minimum result of the objective function through processing one by one or batch processing. The function expression is:

$$E = \sum_{k=1}^{m} E_k \tag{10}$$

Where, E_k is a local error function

Assumed co ownership N hidden layers, update the E hidden layer i units to output units k and the process of connection weight value of is:

$$W_{i,k}^R = R_{i,k-1}^R + \Delta W_{i,k}^R$$
$$W_{i,k}^R = \alpha \varepsilon_{i,k}^R h_k^{R+1}, (i = 1,2,\cdots,N) \tag{11}$$

Applying the learning algorithm to the data mining of multimedia teaching resources in colleges and universities can effectively find out the hidden mode of multimedia teaching resources in colleges and universities.

2.4.2 Fuzzy clustering

C-means and K-means algorithms are commonly used algorithms in fuzzy clustering [26]. Their common point is that the cluster center is modified through repeated iterative calculation, and the Euclidean distance is used to judge the membership of samples. When a specified condition or threshold is reached, the iteration process ends and the classification is completed. However, the K-means algorithm is highly dependent on the initial cluster center, and the classification results lack stability, so the C-means algorithm is still the mainstream algorithm. In this paper, the C-means fuzzy clustering algorithm is used to cluster the extracted data characteristics of college

multi-media teaching resources and divide the data characteristics of college multimedia teaching resources into c categories, corresponding selection c BP neural networks are combined to complete the mining of college multimedia teaching resources.

Assume a given sample of characteristics of multimedia teaching resources in the academic centers $A = \{x_1, x_2, \cdots, x_n\}$, the number of clusters is c, then the objective function of formula (10) holds.

$$min J = \sum_{i=1}^{c} \sum_{k=1}^{n} u_{ik} d_{ik^2} \tag{10}$$
$$d_{ik} = \|s_k - z_i\|$$

u_{ik} represents the k samples in the i membership in the class, and u_{ik} Formula (11) is satisfied.

$$\sum_{i=1}^{c} u_{ik} = 1, \forall k$$
$$n > \sum_{k=1}^{n} u_{ik} > 0, \forall i \tag{11}$$

Where, S_k represents the k relative position of samples; z_i is for the center of class i; d_{ik} is for k samples to the center distance of the class i.

The purpose of C-means algorithm is to obtain the optimal solution of the objective function, which is restricted by the following two constraints:

$$z_i^{(q)} = \frac{\sum_{k=1}^{n} u_{ik}^{(q)m} x_k}{\sum_{k=1}^{n} u_{ik}^{(q)m}}, (i = 1,2,3,\cdots,c) \tag{12}$$
$$u_{ik}^{(q+1)} = \frac{1}{\sum_{j=1}^{c} \left(\frac{d_{ik}}{d_{jk}}\right)^{\frac{2}{m-1}}}, \forall i, \forall k \tag{13}$$

Where, m is the weighted coefficient of membership.

The basic flow of the algorithm is as follows:

Step 1: Give initial parameters m and c, the value of m is generally taken as 2, and the initial cluster center is calculated z;

Step 2: Use formula (12) and formula (13) to compute the corrected z and u;

Step 3: Give a∈If a more appropriate norm matrix is found $\|U^{(q+1)} - U^{(q)}\| Z < \in$ Stop, otherwise turn to the second step.

Based on the clustering results, the number of clusters is c and cluster center z.

2.4.3 Implementation of data mining of college multimedia teaching resources based on BP neural network

Combined BP neural network generates cluster numbers based on fuzzy clustering c, select c BP networks are combined to complete the data mining of college multimedia teaching resources. All BP networks adopt three-layer structures. These layers include input, hidden and output layers. At the same time, heuristic BP improved algorithm Heuristicbp is used to enhance the efficacy and accuracy of the overall combined BP network. In other words, the momentum method and learning rate adaptive adjustment strategy are used.

The approximate steepest descent method is used by traditional BP algorithm to update the weight and offset values as following equation:

$$\Delta V(k) - \lambda S^m(a^{m-1})G \qquad (14)$$

$$\Delta b(k) = -\lambda S^m \qquad (15)$$

Where, ΔV represents the weight value; Δb represents offset; m represents characteristic attribute

Add momentum coefficient to the above formula γ the momentum improvement formula of backpropagation is calculated by momentum improvement:

$$\Delta V(k) = \gamma \Delta V(k-1)\lambda S^m(a^{m-1})G \qquad (16)$$

$$\Delta b(k) = \gamma \Delta b(k-1)(1-\gamma)aS^m \qquad (17)$$

The improved BP neural network structure is seen in Figure 7.
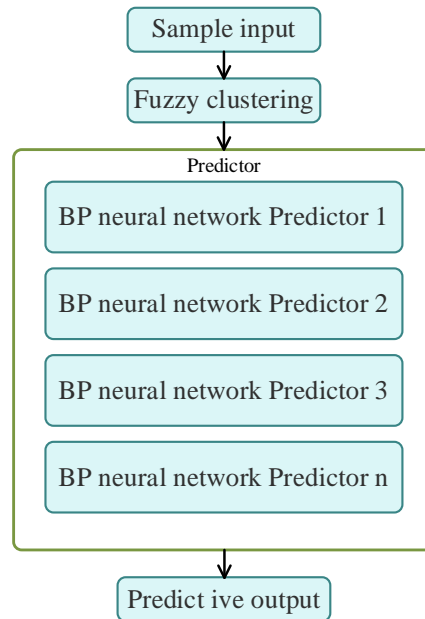
Figure 7: Combined BP neural network

The improved combined BP neural network college multimedia teaching resources data mining process is represented in Figure 8.
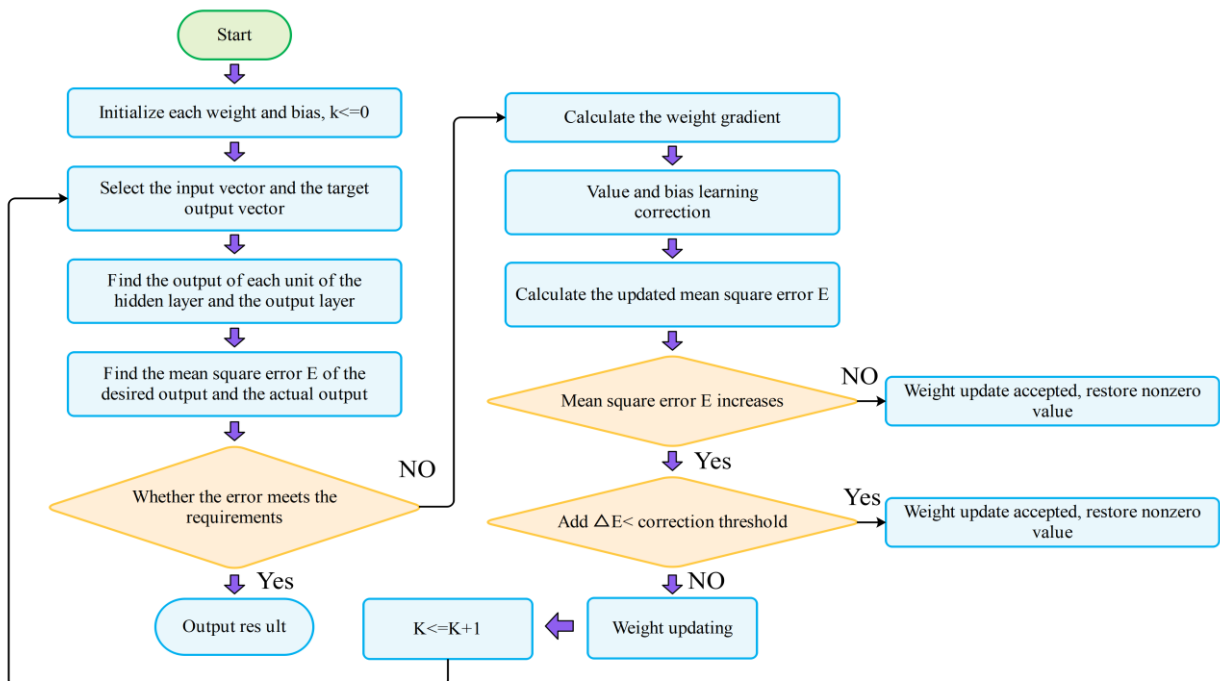
Figure 8: The BP neural network data mining process

Input: untrained BP neural network, characteristics of college multimedia teaching resources, initial learning speed momentum coefficient γ, learning factor and correction threshold.

Output: Mining results of multimedia teaching resources in colleges and universities.

Step 1: Set the weight value and initial offset value, $k \leq 0$;

Step 2: Select an input vector and target output vector;

Step 3: Get the output of each unit of hidden layer and output layer;

Step 4: Get the mean square error between the expected output and the actual output E;

Step 5: Judge whether the error meets the requirements (if it is the output of data mining results); if not, continue to calculate the weight gradient;

Step 6: Complete the weight and bias learning correction;

Step 7: Calculate the updated mean square error E;

Step 8: Determine the mean square error whether increases (if not, the mean value update is accepted, and the non-zero value is restored); if yes, judge whether to increase or correct the threshold value (if yes, the weight update is accepted, and the non-zero value is restored), if not, the weight update is cancelled;

Step 9: When $K \leq K + 1$ select the input vector and target the output vector again.

This completes the mining of efficient multimedia resources,

3 Experimental analyses

3.1 Experimental objects

To conduct intelligent mining of multimedia teaching resources in colleges and universities, the search engine developed by Baidu Company is used to collect data. Baidu Search is a world's leading Chinese search engine. In January 2000, Li Yanhong and Xu Yong founded in Zhongguancun, Beijing, and committed to providing people with "simple and reliable" access to information. In the beginning, Google developed Baidu as an original version to develop and later developed its core technology based on the "Hyperchain analysis. The search service provided.

Baidu has a huge database of Chinese web pages in the world. As of 2010, it has included over 20 billion Chinese web pages. The number of these pages is growing to tens of millions every day. Simultaneously, Baidu servers which are distributed throughout China can directly return the searched information to domestic users from the nearest server, allowing them to enjoy extremely fast search transmission speed.

3.2 Experimental data

A summary of related works is shown in Table 1.

Table 1: Summary of related works

Contrast index	The method in this paper	Reference [13] method	Reference [14] method	Reference [15] method
Correlation data	Intelligent mining performance of multimedia teaching resources in colleges and universities	Intelligent mining performance of multimedia teaching resources in colleges and universities	Intelligent mining performance of multimedia teaching resources in colleges and universities	Intelligent mining performance of multimedia teaching resources in colleges and universities
Method	Input the extracted features into the improved combined BP neural network, and output the intelligent mining results of multimedia resources in colleges and universities.	Based on classical methods, scientific data mining calculates and estimates, and realizes the strategy of scientists' learning automatically.	Classification process and learning process	Using depth hash algorithm to realize the retrieval of multi-view attribute coding image network teaching resources
Research results	Improving the recall and accuracy of data mining can realize classified mining of multimedia teaching resources in colleges and universities.	Strike a good balance between refinement and simplicity.	Using the performance metrics and goodness of fit related to accuracy, the mining of programming teaching resources and the evaluation of students' programming ability are completed	Effectively complete the image retrieval and mining of teaching resources.

To verify the effects of the focused crawler data collection method adopted in the current paper, the data collection amount of college multimedia teaching resource data is compared with ordinary crawlers simultaneously. The results are represented in Figure 9
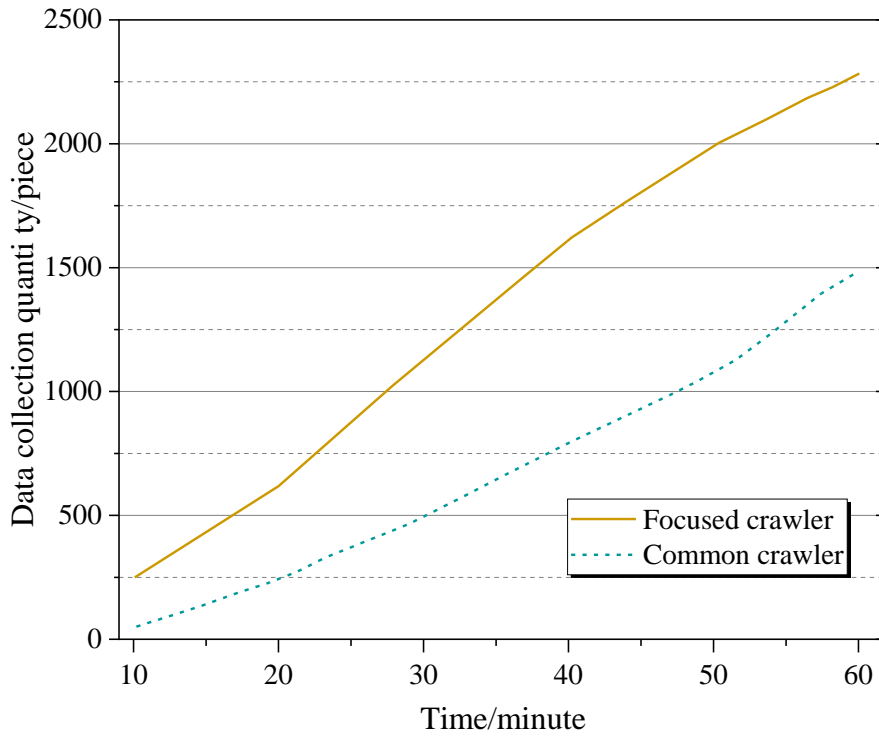
Figure 9: Comparison of data acquisition of two crawlers

As can be seen from Figure 9, the efficiency of collecting multimedia educational resources in colleges and universities by using focused crawler is much higher than that by using ordinary crawler, because focused web crawler not only selects topics when crawling data, but also adopts analysis and evaluation methods, which not only leads to higher data collection amount of efficient

multimedia educational resources but also higher data correlation than ordinary web crawler.

To verify the data processing results of this method, the word frequency statistics of college multimedia teaching resource data collected through web crawlers are done, and the results have been presented in Figure 10.
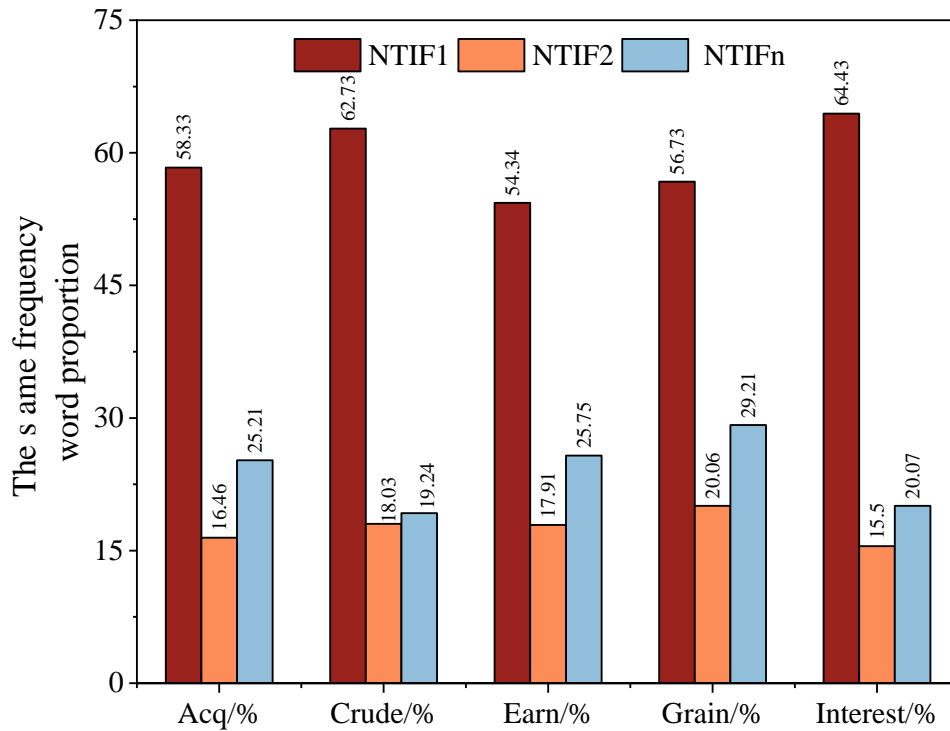
Figure 10: Utes-21578 data set word frequency acquisition results

As seen from Figure 10, after word frequency statistics, $NTIF_1$ The proportion of word frequency is the

highest, more than 50%, and the highest is 64.43%. Its relevance could be higher. Deleting it can significantly

decrease data redundancy and enhance the data preprocessing effect of college multimedia teaching resources.

To prove the data processing effect of this method, the original college multimedia teaching resources and the processed efficient multimedia teaching resources are simultaneously extracted with data features. The data feature extraction results have been provided in Table 2.

Table 2: Feature extraction results of two groups of data

Unprocessed data		Processed data	
Number of BNS features	Number of Odds features	The number of BNS features	The number of Odds features
6424	5276	2486	2057

It can be seen from Table 2 that the original untreated college multimedia teaching data has obtained a large number of features after feature extraction using the method in this paper. Because college multimedia teaching data has yet to be processed, there is a large amount of noise in the data, so there is a large amount of redundancy in the extracted data features, and there are many useless features in the data features. However, the processed college multimedia teaching resources have carried out the operations of data segmentation and word frequency statistics, removed the low-frequency words with low correlation, and the extracted features are efficient and accurate.

To verify the impacts of feature extraction in this method, BNS features and Odds features are input into the improved combined BP neural network, respectively and combined to conduct data mining experiments. The recall rate, accuracy rate and running time are used to evaluate the data mining results using the three features. The data mining results are presented in Table 3

Table 3: Results of mining three kinds of characteristic data

Index	Recall rate/%	Accuracy rate/%	Running time/s
BNS feature	53.82	87.47	637
Odds feature	50.26	79.86	406
BNS+Odds feature	78.67	97.53	512

Table 3 shows that only one feature is used for data mining, and the accuracy and recall rates are lower than the results using two features. Although the operation time of data mining using Odds features is short, the recall and accuracy rates of data mining are sacrificed. Therefore, the dual feature extraction method used in this method is. It is an effective feature extraction method for college multimedia teaching resource mining.

Compare and analyze the training effects of the combined BP neural network before and after the improvement, and analyze the data mining results of this method. The training results are shown in Figure 11.
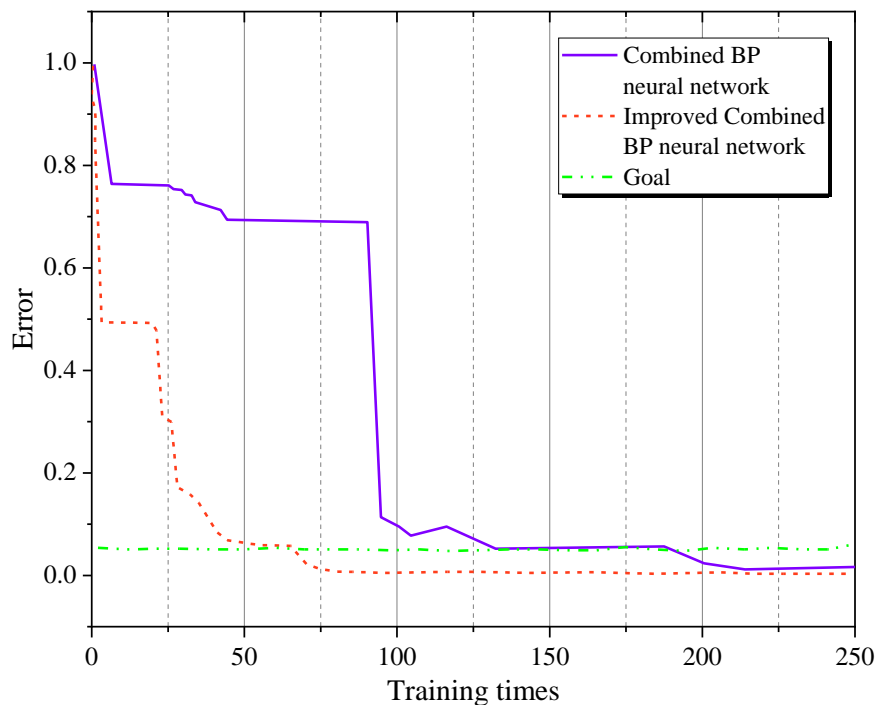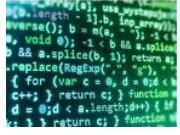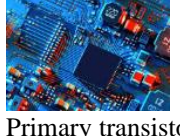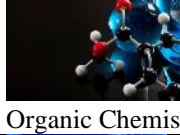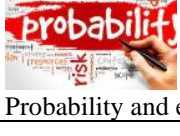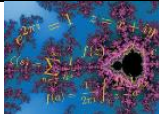

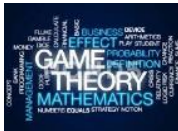Figure 11: Improved BP neural network training results

As can be seen from Figure 11, the error regression speed of the combined BP neural network without improvement is slow, and the result has just reached the target after 190 trainings, and the error rate can be less than the target value after more trainings. On the other hand, the improved combined BP neural network can reach the predetermined target error value only through about 50 trainings, and the error value obtained after continuous training will be far less than the target error value. The error value after training at about 100 is less than the

training result of the combined BP neural network before improvement for more than 200 times. The experimental results show that the improved combined BP neural network has more training advantages and can obtain the data mining results of multimedia educational resources in colleges and universities more quickly.

This method is used to collect college multimedia education resource data and conduct data mining to verify the practical effects of proposed method. The data mining results can be seen in Table 4.

Table 4: Partial data mining results

Index			ID	Multimedia teaching resources
Category	Engineering	Computer Science and Technology	J01	 Parallel programming
		Computer Science and Technology	J02	 Algorithm (II)
		Civil engineering	T01	 Engineering Problems (II)
		Civil engineering	T02	 Civil engineering analysis
		Electrical engineering	E01	 Primary transistor circuit
		Electrical engineering	E02	 Introduction to Electrical Engineering
	Science	Organic chemistry	C01	 Organic Chemistry (I)
		Organic chemistry	C02	 Organic synthesis method
		Applied mathematics	M01	 Probability and engineering applications

		Applied mathematics	M02	 Complex Analysis (II)
		Theoretical physics	P01	 Basis of quantum information
		Theoretical physics	P02	 Intermediate Quantum Mechanics (II)
	Economics	Theoretical economics	WE01	 Game theory
		Theoretical economics	WE02	 Mathematical tools for economists
		Theoretical economics	WE03	 Macroeconomic Theory (II)

It can be seen from Table 4 that this method can effectively mine multimedia teaching resources in academic centers by classification and clearly shows the classification mining results of multimedia teaching resources in various disciplines, and the data mining time is short.

To sum up, studying the intelligent mining method of multimedia teaching resources in colleges and universities can significantly improve the effect and quality of education and teaching by adopting focused crawler to collect resources efficiently, eliminating redundancy and improving data processing efficiency, accurate feature extraction and effective classification methods. The improved combined BP neural network has fast training speed, can obtain mining results faster, and shows the classification results of multimedia teaching resources, which provides an important reference for teaching management and teaching improvement in colleges and universities.

4   Conclusion

The significance of studying the intelligent mining method of multimedia teaching resources in colleges and universities lies in improving the effect and quality of education and teaching. Through intelligent mining, the following aspects can be achieved:

(1) The efficiency of collecting multimedia educational resources in colleges and universities with focused reptiles is much higher than that of collecting multimedia educational resources in colleges and universities with ordinary reptiles;

(2) The correlation degree of this method is low. After deleting it, the data redundancy can be obviously reduced, and a better data preprocessing effect of multimedia teaching resources in colleges and universities can be obtained.

(3) The processed multimedia teaching resources in colleges and universities are operated by word segmentation and word frequency statistics, and the low-frequency words with low correlation are removed, and then the extracted features are efficient and accurate.

(4) The double feature extraction method adopted in this paper is an effective feature extraction method for multimedia teaching resources mining in colleges and universities.

(5) The improved combined BP neural network can reach the predetermined target error value only through about 50 times of training, and the error value obtained after continuous training will be far less than the target error value. The error value after training at about 100 times is less than the training result of the combined BP neural network before improvement for more than 200 times, which can obtain the data mining results of

multimedia educational resources in colleges and universities more quickly.

(6) This method can effectively classify and mine multimedia teaching resources in colleges and universities, and clearly show the results of classified mining of multimedia teaching resources in various disciplines, and the data mining time is short.

In the future research direction, we can further explore the following aspects:

(1) Multimodal data mining: Combining text, images, audio, video and other diverse teaching resources, the association and interaction between different data modes are used in the mining process to improve the accuracy and comprehensiveness of mining results.

(2) Emotional analysis and learning modeling: By digging out students' emotional expressions in the learning process, such as emotion and motivation, an emotional learning model is constructed, and the motivation and influencing factors behind learning are deeply explored to provide more accurate guidance and suggestions for teaching.

(3) Mining of open education resources: With the increase of open education resources, such as MOOC courses and online learning platforms, how to effectively mine these resources to better serve the teaching needs is an important research direction.

(4) Privacy protection and data security: When mining multimedia teaching resources, we should pay attention to students' privacy protection and data security, and formulate appropriate data processing and sharing norms to ensure the safety and reliability of the mining process.

Availability of data and materials

The datasets used in this paper are available from the corresponding author upon request.

Conflicts of interest

The authors declared that they have no conflicts of interest regarding this work.

Authorship contribution statement

Zhu Xuan: Writing-Original draft preparation, Conceptualization, Supervision, Project administration.

Qi Yue: Language review, Methodology, Software

Declarations

Not applicable

References

[1] A. Silik, M. Noori, W. A. Altabey, J. Dang, R. Ghiasi, and Z. Wu, "Optimum wavelet selection for nonparametric analysis toward structural health monitoring for processing big data from sensor network: A comparative study," *Struct*

Health Monit, vol. 21, no. 3, pp. 803–825, 2022. https://doi.org/10.1177/14759217211010261

[2] B. Seidl, R. Schuhmacher, and C. Bueschl, "CPExtract, a software tool for the automated tracer-based pathway specific screening of secondary metabolites in LC-HRMS data," *Anal Chem*, vol. 94, no. 8, pp. 3543–3552, 2022. https://doi.org/10.1021/acs.analchem.1c04530

[3] S. C. Pal, D. Ruidas, A. Saha, A. R. M. T. Islam, and I. Chowdhuri, "Application of novel data-mining technique-based nitrate concentration susceptibility prediction approach for coastal aquifers in India," *J Clean Prod*, vol. 346, p. 131205, 2022. https://doi.org/10.1016/j.jclepro.2022.131205

[4] A. K. Sleiti, "Isobaric Expansion Engines Powered by Low-Grade Heat—Working Fluid Performance and Selection Database for Power and Thermomechanical Refrigeration," *Energy Technology*, vol. 8, no. 11, p. 2000613, 2020. https://doi.org/10.1002/ente.202000613

[5] M. Das, S. K. Ghosh, V. M. Chowdary, P. Mitra, and S. Rijal, "Statistical and Machine Learning Models for Remote Sensing Data Mining—Recent Advancements," *Remote Sensing*, vol. 14, no. 8. MDPI, p. 1906, 2022. https://doi.org/10.3390/rs14081906

[6] B. Robson, S. Boray, and J. Weisman, "Mining real-world high dimensional structured data in medicine and its use in decision support. Some different perspectives on unknowns, interdependency, and distinguishability," *Comput Biol Med*, vol. 141, p. 105118, 2022. https://doi.org/10.1016/j.compbiomed.2021.105118

[7] A. C. Anadiotis *et al.*, "Graph integration of structured, semistructured and unstructured data for data journalism," *Inf Syst*, vol. 104, p. 101846, 2022. https://doi.org/10.1016/j.is.2021.101846

[8] A. Ghose, S. Singh, V. Kulaharia, L. Dokara, S. Maity, and S. Dey, "PySchedCL: Leveraging Concurrency in Heterogeneous Data-Parallel Systems," *IEEE Transactions on Computers*, vol. 71, no. 9, pp. 2234–2247, 2021. https://doi.org/10.1109/TC.2021.3125792

[9] Y. Shin, J. Ahn, and D.-H. Im, "Join optimization for inverted index technique on relational database management systems," *Expert Syst Appl*, vol. 198, p. 116956, 2022. https://doi.org/10.1016/j.eswa.2022.116956

[10] B. Mounica and K. Lavanya, "Real time traffic prediction based on social media text data using deep learning," *Journal of Mobile Multimedia*, pp. 373–392, 2022.

[11] S. Keshary, K. Bekiroglu, S. Seshadhri, and S. Srinivasan, "Multimedia Data-Based Artificial Pancreas for Type 2 Diabetes," *IEEE MultiMedia*, vol. 29, no. 1, pp. 18–27, 2022. https://doi.org/10.1109/MMUL.2022.3154534

[12] D. Shin, "Embodying algorithms, enactive artificial intelligence and the extended cognition:

You can see as much as you know about algorithm," *J Inf Sci*, vol. 49, no. 1, pp. 18–31, 2023. https://doi.org/10.1177/0165551520985495

[13] A. S. Varde, "Computational estimation by scientific data mining with classical methods to automate learning strategies of scientists," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 16, no. 5, pp. 1–52, 2022. https://doi.org/10.1145/3502736

[14] M. A. Marjan, M. P. Uddin, and M. Ibn Afjal, "An Educational Data Mining System For Predicting And Enhancing Tertiary Students' Programming Skill," *Comput J*, vol. 66, no. 5, pp. 1083–1101, 2023. https://doi.org/10.1093/comjnl/bxab214

[15] G. Zhao and J. Ding, "Image Network Teaching Resource Retrieval Algorithm Based on Deep Hash Algorithm," *Sci Program*, vol. 2021, pp. 1–7, 2021. https://doi.org/10.1155/2021/9683908

[16] L. J. Sankpal and S. H. Patil, "Rider-rank algorithm-based feature extraction for Re-ranking the webpages in the search engine," *Comput J*, vol. 63, no. 10, pp. 1479–1489, 2020. https://doi.org/10.1093/comjnl/bxaa032

[17] I. Bifulco, S. Cirillo, C. Esposito, R. Guadagni, and G. Polese, "An intelligent system for focused crawling from Big Data sources," *Expert Syst Appl*, vol. 184, p. 115560, 2021. https://doi.org/10.1016/j.eswa.2021.115560

[18] S. Rajiv and C. Navaneethan, "Keyword weight optimization using gradient strategies in event focused web crawling," *Pattern Recognit Lett*, vol. 142, pp. 3–10, 2021. https://doi.org/10.1016/j.patrec.2020.12.003

[19] K. A. Apoorva and S. Sangeetha, "Analysis of uniform resource locator using boosting algorithms for forensic purpose," *Comput Commun*, vol. 190, pp. 69–77, 2022. https://doi.org/10.1016/j.comcom.2022.04.002

[20] D. Dia, G. Kahn, F. Labernia, Y. Loiseau, and O. Raynaud, "A closed sets based learning classifier for implicit authentication in web browsing," *Discrete Appl Math (1979)*, vol. 273, pp. 65–80, 2020. https://doi.org/10.1016/j.dam.2018.11.016

[21] G. De Marzo, F. S. Labini, and L. Pietronero, "Zipf's law for cosmic structures: How large are the greatest structures in the universe?," *Astron Astrophys*, vol. 651, p. A114, 2021. https://doi.org/10.1051/0004-6361/202141081

[22] M. Huang, H. Ma, C. Ma, P. A. Garber, and P. Fan, "Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: insights into the origin of human language," *Anim Behav*, vol. 160, pp. 145–155, 2020. https://doi.org/10.1016/j.anbehav.2019.11.017

[23] F. Xue and D. Połap, "Detail feature inpainting of art images in online educational videos based on double discrimination network," *Mobile Networks and Applications*, pp. 1–14, 2023. https://doi.org/10.1007/s11036-023-02191-x

[24] S. H. Syed and V. Muralidharan, "Feature extraction using Discrete Wavelet Transform for fault classification of planetary gearbox–A comparative study," *Applied Acoustics*, vol. 188, p. 108572, 2022. https://doi.org/10.1016/j.apacoust.2021.108572

[25] J. Peng and W. Yu, "The algorithm of current prediction based on multi-dimensional Long Short Term Memory networks," *Energy Reports*, vol. 7, pp. 1114–1120, 2021. https://doi.org/10.1016/j.egyr.2021.09.158

[26] Q.-T. Bui *et al.*, "SFCM: a fuzzy clustering algorithm of extracting the shape information of data," *IEEE Transactions on Fuzzy Systems*, vol. 29, no. 1, pp. 75–89, 2020. https://doi.org/10.1109/TFUZZ.2020.3014662