# A Micro-Class Teaching Data Retrieval Method of Business English Based on Network Information Classification

Wang Guifang
E-mail: 20072026@zyufl.edu.cn

College of English, Zhejiang Yuexiu University, Shaoxing, Zhejiang 312000, China

*In order to quickly extract the required micro-class teaching data of business English in the fragmented network information environment, a micro-class teaching data retrieval method of business English based on network information classification is offered. This way, it constructs a network information classification method on the basis of the optimised SVM model, and the parameters of the SVM model are optimised and trained by the improved artificial bee colony algorithm. After optimising the function of the SVM method, the method classifies the online teaching information of various business English micro-class teaching; In the classification network information set, targeted retrieval of network teaching resources according to big data techniques is applied to obtain the teaching data with the highest similarity to the user retrieval data by clustering, which completes the targeted retrieval of the microclass teaching data. The experimental results show that the retrieval delay of the proposed method for micro-class teaching data retrieval of business English is less than 1s, the number of correct retrievals is relatively high, and there are few wrong retrieval phenomena.*

*Povzetek: Metoda za hitro iskanje podatkov predavanj poslovnega angleškega jezika temelji na klasifikaciji spletnih informacij z optimiziranim SVM modelom.*

## 1 Introduction

For business English courses, the content is complex. As a key course in the major, it is important to highlight teaching effectiveness [1], [2] Therefore, it needs to try to integrate the two, apply the core function of micro-class instructing to business English instructing, and mine the essence of it. In the new era, the attempt at "micro-class" teaching of business English has its own value. Combined with the current application status of micro-class teaching mode, from the perspective of fragmented content, interactive learning and diversified resources, the specific analysis of teaching value is completed [3], which can provide necessary conditions for teaching innovation. The core resources for implementing and applying micro-class teaching are micro-video courses. The time of each teaching video is usually 6-8 minutes, with clear focus and specific content [4], [5]. The teaching of micro-class makes full use of the convenient features. It supports the development of teaching practice with diversified teaching resources, which has a multifaceted and comprehensive impact on students. The supplement and expansion of resources make the content fuller, more specific and attractive, improving students' comprehensive ability to expand and optimise resources [6]. But in the mean time, in the micro-class teaching style of business English, the diversified teaching resources also pose a challenge to the effect of teaching data retrieval. Quickly extracting the required teaching data

from diversified teaching resources is one of the difficulties in the current micro-class teaching process of business English [7].

In reference [8], Wang et al. used deep learning technology in the cross-modal data retrieval of network information, which improved the retrieval accuracy by about 19% compared with the manual feature method. However, during the teaching of the deep learning network, the parameter initialisation ability needs to be improved, thus affecting the speed of cross-modal data retrieval of network information, which needs to be developed in the coming study work. Reference [9], offered a geospatial data extraction way on the basis of machine learning. This method mainly completed the classification and extraction of geospatial data through machine learning KNN classification. This research provided a reference idea for the study context of this article. However, this method classified data by measuring the distance between different feature values. When the network information data is complex, especially when the sample is unbalanced, this method's classification and extraction effect will also be negatively affected. Reference [10], proposed a zero-sample cross-modal retrieval method based on deep supervised learning, aiming at the problem that category matching and corresponding matching are not considered in the current zero-sample cross-modal retrieval research. This method has a good effect on image retrieval, text and text retrieval image. Still, this method also points out at the end of the

study that the retrieval performance of this method is directly related to the teaching effect of the supervised deep learning network. Based on the above analysis, the key contributions and results of existing research on teaching data retrieval and network information classification are summarized as follows:

| Paper | Key Contributions | Result |
|---|---|---|
| [8] Wang et al | Cross modal data retrieval using deep learning techniques | Compared to manual feature methods, the retrieval accuracy has been improved by approximately 19% |
| [9] Zhu et al | Propose a machine learning based geospatial data extraction method | Complete the classification and extraction of geospatial data through KNN classification method |
| [10] Zeng et al | Propose a zero-sample cross modal retrieval method based on deep supervised learning | Good performance in image retrieval of text and text retrieval of images |

Further analysis of the table content shows that although deep learning techniques have achieved good results in cross modal data retrieval in [8], research has pointed out that the parameter initialization ability needs to be improved during the training phase, which may affect the speed of cross modal data retrieval of network information. In [9], although the machine learning based geospatial data extraction method provides a reference idea, this method measures the distance between different feature values for classification and extraction, which may have a negative impact on the effectiveness in complex network information data or imbalanced samples. In [10], the zero-sample cross modal retrieval method based on deep supervised learning proposed performs well in image retrieval of text and text retrieval of images. However, research has pointed out that its retrieval performance is directly related to the training effect of deep supervised learning networks, implying certain limitations. In summary, the most advanced data retrieval teaching methods still have gaps or limitations in parameter initialization ability, data complexity and sample imbalance issues, as well as the training effectiveness of deep supervised learning networks.

Combined with the issues mentioned above, this topic has a wide range of research value due to the positive role of network information classification methods in improving the retrieval ability of micro-class teaching data in business English. It has become a key research topic for many scholars [11], [12], [13], [14]. Therefore, this paper proposes a micro-class data retrieval method of business English according to network information classification. This method is parted chiefly into two: network information classification and data search of micro-class instructing data of business English. The instructing data search is mainly completed in the classified network information collection. This method's design of the operation link can improve the order and regularity of business English micro-class teaching data, thus improving the data retrieval effect.

# 2 The micro-class data retrieval method of business English

A large number of micro-videos and micro-courses are distributed on the network platform. It is difficult for learners to concentrate on learning. They pay attention to the quantity of learning content and ignore its quality [15], [16], [17], [18], [19], [20] In fragmented learning processes, the various information resources are mixed, the discriminative power of the learner is limited, and the quality of the learned content is not guaranteed. In addition, the fragmented learning time is uncertain, and learners cannot complete the learning in the limited fragmented time. The content of learning is simply and quickly browsed, which is difficult to absorb and incorporate into their own knowledge system. There is no quality and quantity of learning, which reduces the learning quality and efficiency [21], [22], [23]. Therefore, this paper designs a data retrieval method for the business English micro-class.

## 2.1 Establishment of network information classification model based on SVM

Support vector machines (SVM for short) can accurately classify network information [24]. The idea of SVM is on the basis of the optimal classification surface based on linear separability. Its detailed introduction is represented in Fig 1.
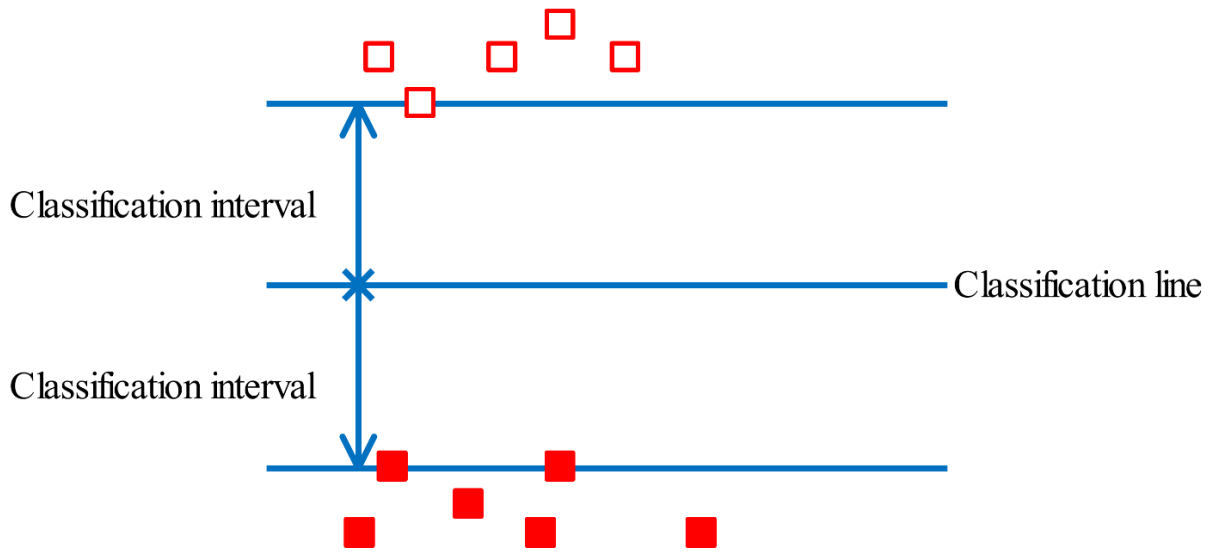
Figure 1: Schematic diagram of optimal classification surface of SVM.

The optimal classification surface is relative to the multi-dimensional space. The optimal classification surface is the optimal classification line, which is to accurately classify the networked micro-class teaching data samples of business English and maximise the interval, that is, the classification interval.

The principle is as follows:

The training set of network information is set as $(a_j, b_j)$, and the following equation is used to describe the linear function $f(a)$ of SVM in high-dimensional feature space:

$$f(a) = \varpi^\top \phi(a) + c \tag{1}$$

Where, $\varpi$ is the weight vector of the network information characteristic, and $c$ is the offset of the network information characteristic; $\phi(a)$ is the input network information sample.

According to the purpose of risk minimisation $\min I(\varpi, d)$, the above equation can be described as a constrained optimisation problem, and its equation is as follows:

$$\min I(\varpi, d) = \frac{\varpi}{2} \varpi^\top + \frac{D}{2} \sum_{j=1}^{m} d_j^2 \tag{2}$$
$$s.t.$$
$$b_j = \varpi^\top \phi(a) + c + d; j = 1,2,3,\ldots,m$$

Where, $I(\varpi, d)$ is the structural risk value of the SVM model, indicating the model's effectiveness; $D$ is the penalty coefficient; $d$ is the classification error; $b_j$ is the network information training sample; $m$ is the total number of network information training samples.

By introducing the Lagrange function and using the Lagrange multiplier, the above-constrained optimisation problem can be converted into an unconstrained optimisation problem in dual space. The equation is as expressed below:

$$L(\varpi, c, d, \beta) = I(\varpi, d)$$
$$- \sum_{j=1}^{m} (\varpi^\top \phi(a_j) + c - b_j) \tag{3}$$

Where $\beta$ is the Lagrange multiplier.

The following equation can be obtained from the KKT condition:

$$\begin{bmatrix} 0 & d1^T \\ d1 & \beta + D^{-1}J \end{bmatrix} \begin{bmatrix} c \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ b \end{bmatrix} \tag{4}$$

Where $J$ is the identity matrix.

In the process of network information classification, the radial basis function is set as the SVM function [25], then in the network information classification method based on the optimised SVM model, the SVM model is:

$$f(a) = \text{sgn} \left[ \sum_{j=1}^{m} b_j \exp \left( -\frac{(\| \quad \|)^2}{2\mu} \right) + c \right] \tag{5}$$

Where $\mu$ is the width coefficient of the radial basis function. $a_j$ and $a_i$ are the $j$-th and $i$-th network training samples, in turn.

According to the classification principle of SVM [26], the learning performance of SVM is determined by $D$ and $\mu$. If the values of these two parameters are too large, there will be over-fitting. Otherwise, there will be under-fitting, so it is necessary to optimise SVM.

## 2.2 Improved artificial bee colony algorithm

There are three kinds of bees in the algorithm: employed bees, on-looker bees and scout bees. These three kinds of bees work and cooperate with each other to find better honey sources. First of all, the employed bees are responsible for searching for new honey sources around the honey source and sharing the honey source location, honey quantity and other information with the on-looker bees after completing the search. Then, according to the

information shared by the employed bees, the on-looker bees choose a certain honey source to continue mining. If the amount of honey in the honey source is more, the probability of being selected is higher. Finally, suppose a honey source has not been updated many times in succession. In that case, the honey source will be abandoned, and the scout bee's bee is responsible for searching for a new honey source randomly to replace the abandoned honey source. It is worth noting that the number of employed bees, the number of on-looker bees and the number of honey sources are the same. There is only one scout bee. The honey source corresponds to the candidate solution of the SVM optimisation problem, which is the parameter $D$ and parameter $\mu$ of the SVM model. The honey amount of honey source corresponds to the fitness value of the SVM optimisation problem.

The classical ABC algorithm uses a random initialisation population to start the iterative search. Suppose that the honey source size representing the candidate solution of the SVM optimisation problem is $Rm$, where the honey source is $Y_j$, then:

$$Y_j = (y_{j,1}, y_{j,2}, \ldots, y_{j,E}) \tag{6}$$

$y_{j,E}$ represents the $j$-th honey source in dimension $E$.

$$y_{j,i} = y_i^{\min} + rand \cdot (y_i^{\max} - y_i^{\min}) \tag{7}$$

Where, $y_{j,i} \in y_{j,E}$, $j = 1,2,\ldots,Rm$, $i = 1,2,\ldots,E$. $y_i^{\max}$ represents the $i$-th dimension upper bound value of SVM model parameters, $y_i^{\min}$ represents the $i$-th dimension lower bound value; $rand$ is a random number evenly distributed between 0 and 1. In the initialisation process, it uses Equation (7) to generate all dimension values of each honey source. After population initialisation, the whole population enters the search phase of employed bees, on-looker bees and scout bees and iterates through these three phases until the algorithm termination condition is reached. The details of the three stages are as follows:

(1) Employed bee's stage

At this stage, each employed bee generates a new honey source $U_j = (u_{j,1}, u_{j,2}, \ldots, u_{j,E})$, $u_{j,i} \in u_{j,E}$ at the corresponding honey source $Y_j$ according to Equation (8), then:

$$u_{j,i} = y_{j,i} + \varphi_{j,i}(y_{j,i} - y_{s,i}) \tag{8}$$

Where, $y_{s,i}$ represents a randomly selected honey source in the population, and $\varphi_{j,i}$ represents a random number evenly distributed between 0 and 1. According to the greedy selection mechanism, when the honey quantity of the candidate honey source $U_j$ is more, that is, the fitness value is better to replace the honey source $Y_j$.

(2) On-looker bees stage

After all the employed bees complete the search, the on-looker bees select a honey source to continue mining according to the information received. The possibility of the $j$-th honey source being chosen is:

$$q_j = \frac{u_{j,i}}{\sum_{i=1}^{Rm} fit_j} \tag{9}$$

Where, $fit_j$ represents the fitness of the $j$-th honey source.

$$fit_j = \begin{cases} \dfrac{1}{1 + \min I(\varpi, d)} & q_j \geq 0 \\ 1 + |\min I(\varpi, d)| & else \end{cases} \tag{10}$$

Where, $\min I(\varpi, d)$ shows the goal function value of the solution, and the higher the fitness is, the greater the probability of honey source selection is.

(3) Scout bee's stage

After the completion of the above two stages, if a honey source has not been updated many times in succession, it means that the honey source has been exhausted. In this case, the honey source will be discarded and replaced by a new honey source, according to Equation (6).

In the traditional artificial bee colony method, the method of employing bees to determine the location of the next honey source search is to use the greedy mechanism to compare the fitness value of the corresponding honey source in the previous two searches. The food source search mode determines whether bees can quickly and accurately find new honey sources. However, the advantages and disadvantages of the position before and after the iteration are not taken into account when searching, and the global optimisation is insufficient. The result is that the search skill of the method is deficient. The main disadvantages are large iteration randomness, slow update speed, easy falling into optimal local solution, etc. To resolve this issue, a global search factor is presented. In each search process, the current honey source information with the best fitness is added to the next location update, then:

$$new - Y_j = Y_j + rand \cdot (Y_{nj} - Y_{hj}) + rand \cdot (Y_{best,j} - Y_j) \tag{11}$$

Where, $Y_{nj}$ and $Y_{hj}$ represent different honey sources; $h$ and $n$ are randomly generated random numbers, and $h$ and $n$ are not equal to each other, and neither is equal to $j$; $rand$ is a random number evenly distributed between 0 and 1; $Y_{best,j}$ represents the honey source with the highest food abundance (fitness value) at present. In the first cycle of artificial bee colony algorithm optimisation, it is the honey source with the highest fitness value among the initial $M$ honey sources.

The improved artificial bee colony algorithm can make the search of bees directional and accelerate the convergence speed of the algorithm.

## 2.3 Optimisation process of SVM parameters

When using the improved artificial bee colony algorithm to optimise, SVM parameters are used as the honey source. Three types of bees optimise the honey source according to their own tasks to obtain the best honey

source. The specific process is to use the SVM parameters that need to be optimized as honey sources and use an improved artificial bee colony algorithm to optimize these parameters. The bees here can be divided into three types: hired bees, reconnaissance bees, and reconnaissance bees. Hiring bees to evaluate honey sources based on certain evaluation criteria (such as prediction accuracy or model performance indicators) and selecting the optimal honey source based on the evaluation results; Reconnaissance bees are responsible for searching and discovering new honey sources to increase global search effectiveness; The

reconnaissance bee, on the other hand, has high exploration ability and can jump out of the local optimal solution to avoid the algorithm falling into the local optimal. Through the above design, the improved artificial bee colony algorithm can more accurately find the optimal solution of SVM parameters, thereby improving the performance and prediction accuracy of the SVM model. Then the optimal honey source is used to build the network information classification model based on the optimised SVM model. Fig 2 is the flow chart of the network information classification model based on the optimised SVM model.
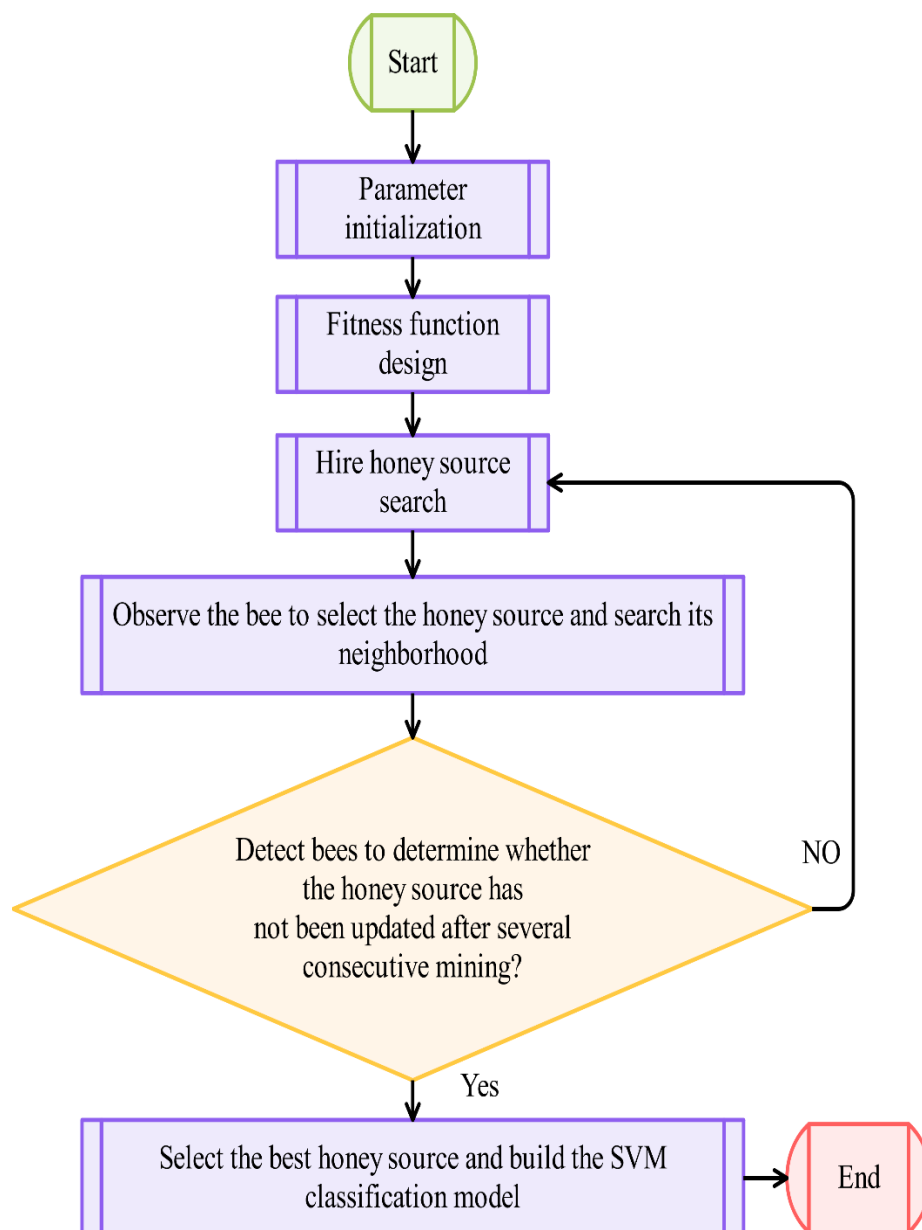


Figure 2: Operation process of network information classification model based on optimised SVM model

(1) The control parameters in the initialisation algorithm mainly include the size of the bee colony, the number of honey sources, the maximum number of honey source cycles, and the maximum number of iterations.

(2) Set the fitness function in the algorithm. The fitness function is calculated using Equation (10).

(3) According to their respective tasks, the three honeybees optimise the honey source, calculate the fitness

value using the fitness function, and optimise all possible solutions found.

(4) According to the set value of the number of iterations, it is able to determine whether the number of cycles of the honey source exceeds the limit. If the number of cycles exceeds the maximum, the newly generated honey source will replace the original honey source. The current honey source is the best searched. It is recorded and determined whether the termination condition is satisfied according to the cyclic condition.

(5) The obtained global optimal honey source, that is, the optimal parameters, is used to construct the SVM model.

(6) After the construction of the SVM model, Equation (5) is used to complete the network information classification.

## 2.4 Directional retrieval method of online teaching resources based on big data technology

Considering the classification of network information resources is completed, big data technology is used to perform directional retrieval of network information resources, and the retrieval process is presented in Fig 3.
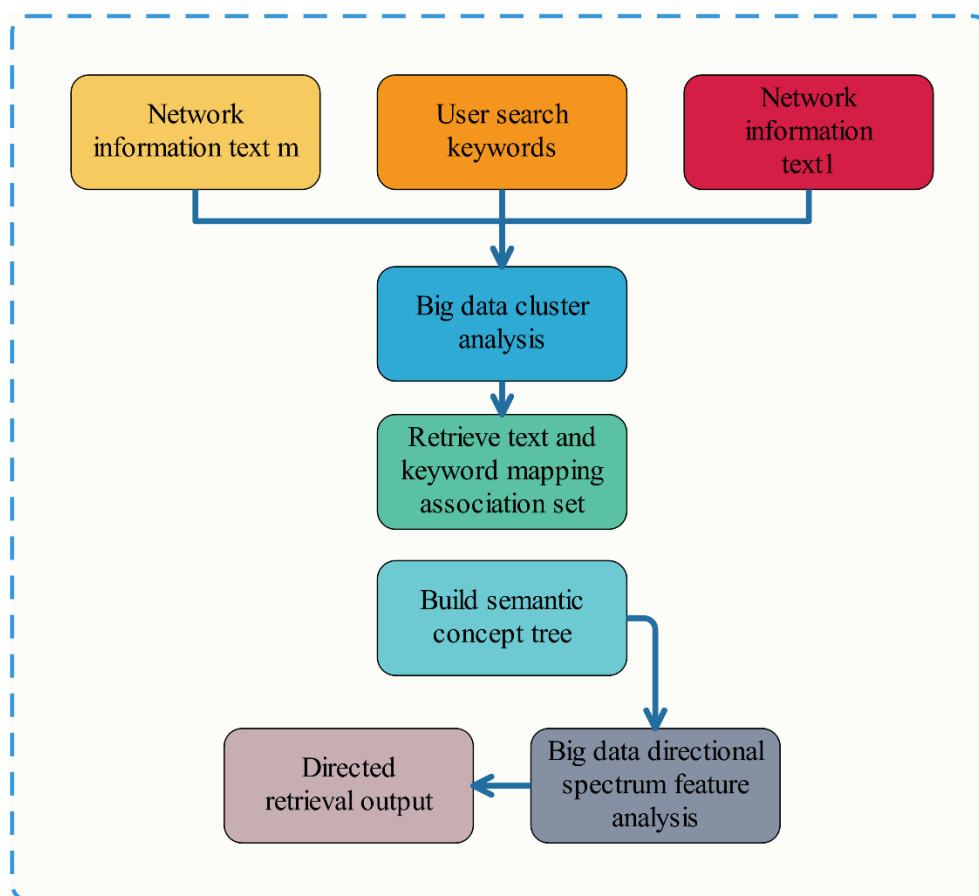


Figure 3: Directional retrieval of network information resources on the basis of big data technology

As presented in Fig 3, big data technology is used to cluster and analyse the classified network information set with user search keywords. Since the user search keywords may belong to multiple categories and the corresponding teaching data may also exist, this time, the network information resources will be used as the vertex in the graph. The vertices in the graph are weighted according to the correlation between network information resources and user search keywords. At this time, an undirected weighted graph will be obtained, thus transforming the problem of micro-class teaching data clustering analysis of business English into the problem of graph division.

The undirected weighted graph can be expressed as:

$$P = \langle \theta, \xi, \rho \rangle \tag{12}$$

Where, $P$ represents the undirected weighted graph used for the retrieval and analysis of business English micro-class data; $\xi$ represents the vertex in the undirected weighted graph, that is, the classified network information sample set; $\xi$ represents the edge weight of the undirected weighted graph, that is, the correlation between the classified network information samples and the search keywords; $\rho$ represents a symmetric matrix. The key to cluster analysis of network information resources by using an undirected weighted graph is the derivation of $\xi$. In fact, it is to calculate the correlation between the classified network information resources set and the retrieval keywords. The calculation equation is:

$$\xi = \frac{g_\delta g_e}{P \max g_\xi} \lg \frac{\lambda}{\chi} \tag{13}$$

Where, $g_\delta$ represents the frequency of user search keywords in the classified network information set; $g_e$ shows the number of network data samples containing user search keywords; $g_\xi$ shows the number of user keywords contained in the classified network information set; $\lambda$ represents the amount of teaching data in the classified network information set; $\chi$ represents the key of user search keywords in the classified network information set, and its value is - 1~1. Equation (13) is used to calculate the correlation between the classified network information sample and the search keywords, sort the network information according to the degree of correlation, and establish the cluster set $L$. According to the correlation between the two, it can evaluate whether there is only one feature class in the classified network information set. If so, it should take it as a subgraph of the undirected weighted graph. After evaluating all the network information samples in the cluster set $L$, $n$ subgraphs will be obtained, thus obtaining the vertex set of the undirected weighted graph. On this basis, the semantic concept tree is established, the teaching data feature classification is implemented according to the attributes of each vertex in the undirected weighted graph, and it is fused. The equation is:

$$\Gamma(\varepsilon_m) = \frac{\xi r(\varepsilon_m)}{G(\varepsilon_m) - r(\varepsilon_m)} \quad (14)$$

Where, $\Gamma(\varepsilon_m)$ represents the result of teaching resource sample fusion, that is, the micro-class teaching data resources of business English that meet the user's retrieval requirements; $r(\varepsilon_m)$ shows the effective possibility of retrieving the network data sample $\varepsilon_m$; $G(\varepsilon_m)$ shows the joint distribution possibility of the network information sample $\varepsilon_m$. The fused network information resources are output as directed retrieval results to achieve business English teaching data retrieval based on network information classification.

## 3 Experimental analysis

To verify the effectiveness of the proposed method, experiments are required. Firstly, the research on business English micro course teaching data retrieval based on network information classification will be integrated as a new module of the education platform. This module interacts with existing platform databases through APIs and provides a user interface for retrieving and browsing business English micro course resources. Then, import the business English micro course teaching data into the database of the education platform, and process and transform it according to the structured characteristics of the existing data to ensure consistency with the existing data. Secondly, develop a data retrieval module for business English micro course teaching, including search interfaces, query algorithms, and result display functions. By defining APIs to achieve data interaction and query request transmission with existing platforms. Finally, design a user-friendly search interface that provides functions such as keyword search, advanced filtering, and sorting. Ensure that the user interface is consistent with the appearance and functionality of existing platforms for seamless use by learners.

The dataset selected for the experiment is sourced from business English micro course teaching data from a certain website. Before the experiment, the experimental data was normalized and the processed values ranged from -1 to 1. The data information of online business English micro course teaching during the experimental process is shown in Table 1.

Table 1: Specific information of network information data

| Business English micro-class data set type | Number/type of information types |
| --- | --- |
| Video frequency | 2 |
| Picture | 2 |
| PPT | 2 |
| Characters | 2 |
| Data | 2 |

The dataset shown in Table 1 contains different types of information, including videos, photos, PPTs, characters, and data, to learn and understand business English courses from different perspectives and levels. At the same time, learners can acquire knowledge through various methods such as watching videos, browsing photos, reading PPTs, reading character texts, and analyzing data files. All kinds of information are well organized and organized, with a certain degree of structure. By learning these materials, learners can improve their abilities in listening, speaking, reading, writing, and other aspects of business English, and prepare for future business scenarios. Therefore, the dataset in Table 1 has characteristics such as diversity, abundant resources, comprehensiveness, structure, and practicality, which can accurately verify the retrieval performance of different methods. The related parameters of the SVM model used in this method are set as follows: the maximum number of iterations is 100. The method in this paper uses the network information classification model based on the optimised SVM model to classify the micro-class teaching data of business English in Table 1. To intuitively reflect the classification effect of this method on the data, it takes the image teaching data in Table 1 as an example. It gives the distribution details of the classification results of the image teaching data before and after the SVM adopts the improved artificial colony algorithm, as shown in Fig 4 and Fig 5.
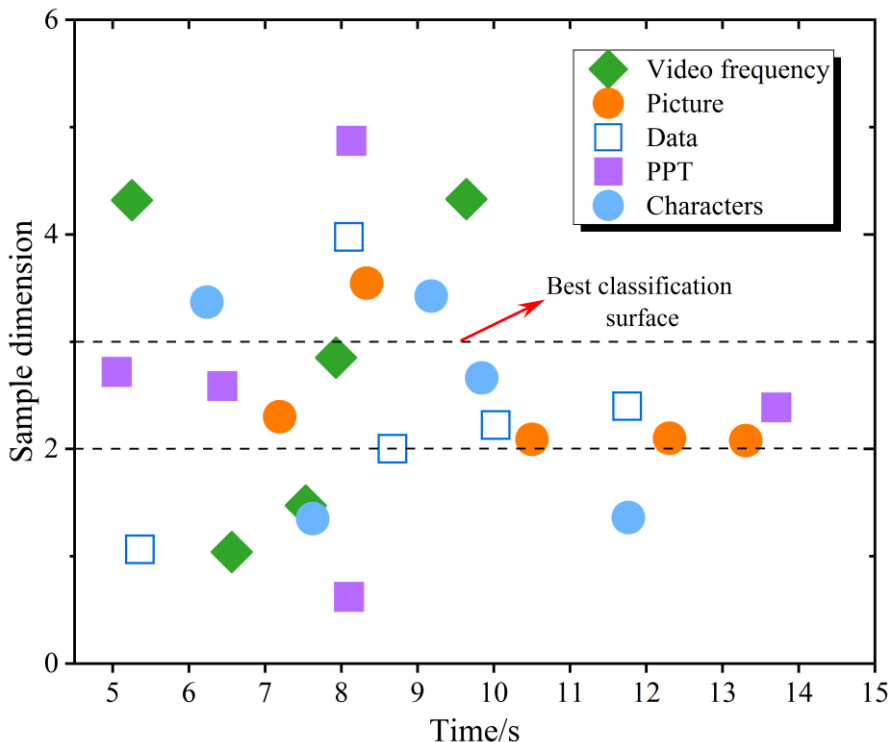
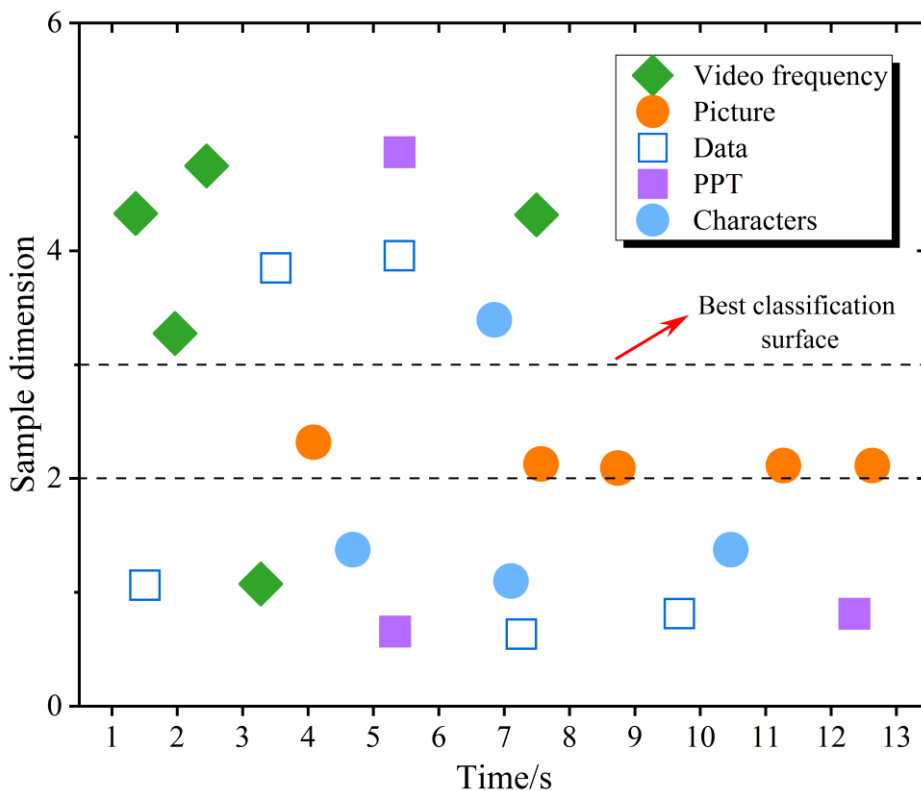Figure 4: Distribution details of classification samples before improved artificial bee colony algorithm



Figure 5: Distribution details of classification samples after improved artificial bee colony algorithm

Compared with Fig 4 and Fig 5, it can be seen that before the method of this paper classifies the network information of the business English micro-class, the network information samples of the business English micro-class are disordered. There is no obvious boundary between the samples, and then the image data retrieval effect of the business English micro-class will be negatively affected. After classifying the network information of the business English micro-class in the method of this paper, the image network information samples can be accurately classified in the best classification plane, which proves that this method has the

ability to classify the network information of business English micro-class.

Fig 6 shows the training convergence change of the support vector machine in the experiment of the artificial bee colony algorithm before and after the improvement of the artificial bee colony algorithm. According to the parameter optimisation process in Fig 6, the improved artificial bee colony algorithm has a faster convergence speed compared with the traditional artificial bee colony algorithm. It can jump out of the local optimal solution. Improvement is necessary.



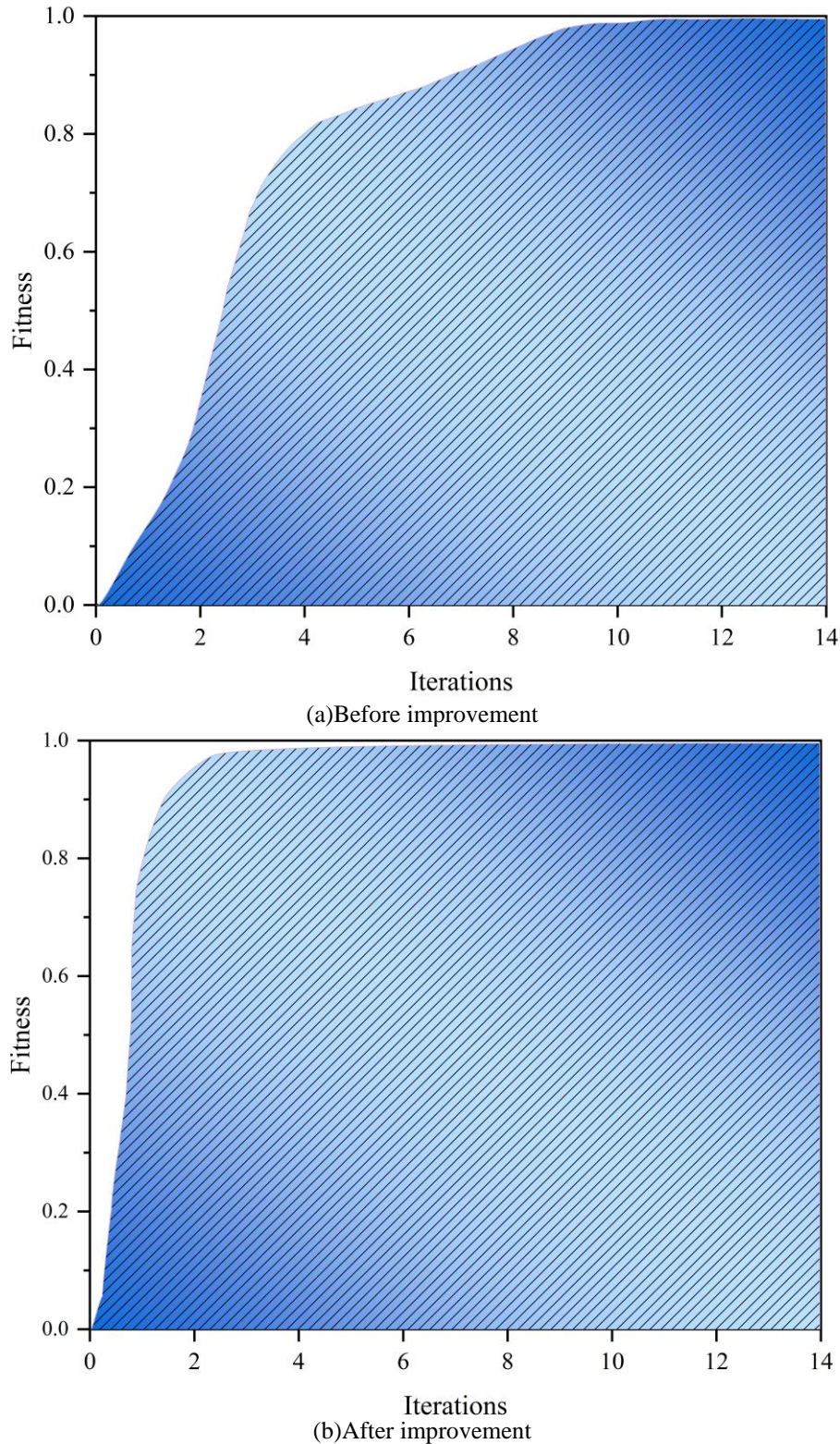(a)Before improvement



(b)After improvement

Figure 6: Change of training convergence of support vector machine

The method in this paper uses the network information classification model based on the optimised SVM model. After classifying the five types of network information data in Table 1, the classification effect is reflected by the Conditional Log-Likelihood Loss ($CLL - loss$) index, which is a classification effect evaluation index. The equation is as follows:

$$CLL - loss = -\sum_{h=1}^{5} \lg\left(\frac{1}{Q(A^{(h)}|y^{(h)})}\right) \quad (15)$$

For the test data sample $y^{(h)}$, when the classification probability of the correct type $A^{(h)}$ is close to 1, the value of $CLL - loss$ is minimal, and if the classification probability of correct type $A^{(h)}$ is close to 0, the value of $CLL - loss$ is maximal. Then the information classification effect of the five micro-class teaching data is shown in Fig 7 after the application of the method in this paper, the method in reference [8], the method in reference [9] and the method in reference [10].
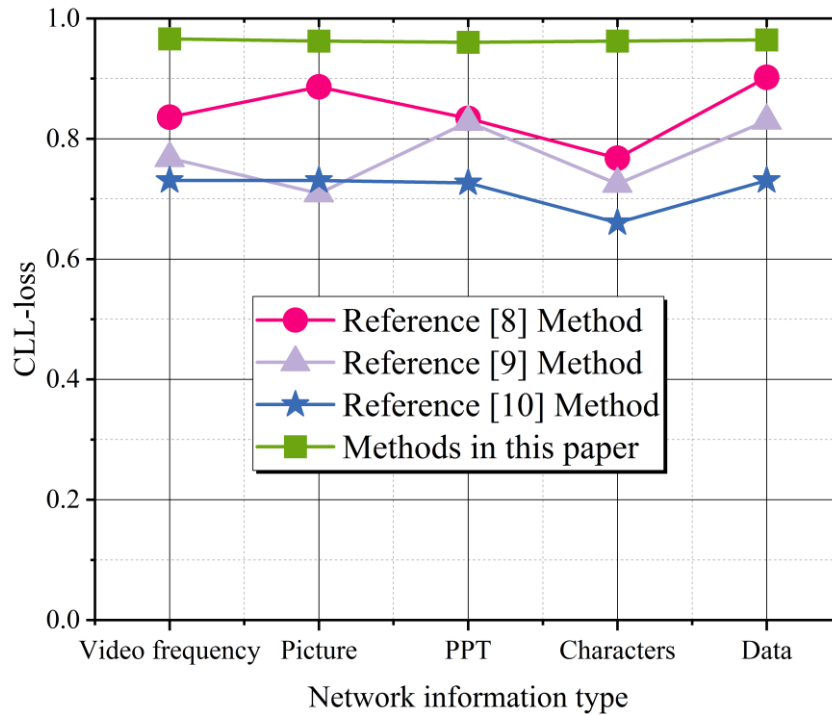


Figure 7: Classification effect of five kinds of network information data

As presented in Fig 7, after applying the method of this article, the $CLL - loss$ value of the five network information classification results is greater than 0.95, which is higher than the maximum $CLL - loss$ value of 0.9 of the classification results by using the three comparison methods. It indicates that the classification probability of correct type $A^{(h)}$ of network information in the method of this paper is close to 1, and the classification accuracy is high.

After classifying the network information of micro-class teaching of business English, the retrieval ability of

the method in this paper to micro-class teaching data is tested. Each group of experiments searches 10 times in total, the number of retrieval files is 100, and the amount of network information resources is gradually increased by 100. To make the retrieval impact of this approach more convincing, it is essential to use the ways of reference [8], reference [9] and reference [10] for comparison and carry out comparative experiments under the same conditions. Table 2 shows the retrieval results of network information resources by four methods.

Table 2: Retrieval results of micro-class teaching network information resources by four methods

| Number of network information samples/piece | Methods in this paper | | The method in reference [8] | | The method in reference [9] | | The method in reference [10] | |
|---|---|---|---|---|---|---|---|---|
| | Number of correct searches/ piece | Number of error retrievals/ piece | Number of correct searches/ piece | Number of error retrievals/ piece | Number of correct searches/ piece | Number of error retrievals/ piece | Number of correct searches/ piece | Number of error retrievals/ piece |
| 100 | 100 | 0 | 98 | 2 | 95 | 5 | 91 | 9 |
| 200 | 199 | 1 | 198 | 2 | 195 | 5 | 192 | 8 |
| 300 | 299 | 1 | 298 | 2 | 295 | 5 | 289 | 11 |
| 400 | 399 | 1 | 398 | 2 | 395 | 5 | 391 | 9 |
| 500 | 499 | 1 | 498 | 2 | 495 | 5 | 492 | 8 |
| 600 | 599 | 1 | 598 | 2 | 595 | 5 | 589 | 11 |
| 700 | 699 | 1 | 698 | 2 | 695 | 5 | 688 | 12 |
| 800 | 799 | 1 | 798 | 2 | 795 | 5 | 786 | 14 |
| 900 | 899 | 1 | 898 | 2 | 895 | 5 | 865 | 35 |
| 1000 | 999 | 1 | 998 | 2 | 995 | 5 | 978 | 22 |

As shown in the data in Table 2, after comparing the methods of reference [8], reference [9] and reference [10], the method in this paper searches the data of business English micro-class teaching, with the rise of the number of network information samples, the maximum number of error searches of micro-class teaching data is only one sample. The number of error searches of reference [8], reference [9] and reference [10] methods is more. This proves that the algorithm in this article has more advantages in the retrieval impact. The number of correct retrieval in this method is relatively high, and there are few wrong retrieval phenomena. In contrast, the number of wrong retrieval in the comparison methods is relatively large.

The above experiments have verified that the approach in this article can retrieve micro-class teaching data of business English. In order to deeply analyse whether there are duplicate and redundant retrieval results in the retrieval results of the algorithm in this article, the experimental setting is a high-precision retrieval index. If the correct results are shown in the first place in the retrieval results of micro-class teaching data of business English, and there is no redundancy in the retrieval results, then the index is higher. In this paper, under the retrieval conditions of different data types, the index test results before and after the retrieval of business English micro-class teaching data are shown in Table 3.

Table 3: Test results of three methods for laser video image high-precision retrieval index

| Business English micro-class teaching data type | Number of retrieved samples/piece | Before use | After use |
|---|---|---|---|
| Video frequency | 10 | 0.87 | 0.98 |
| | 100 | 0.78 | 0.98 |
| Picture | 10 | 0.88 | 0.98 |
| | 100 | 0.79 | 0.98 |
| PPT | 10 | 0.84 | 0.98 |
| | 100 | 0.77 | 0.98 |
| Characters | 10 | 0.91 | 0.98 |
| | 100 | 0.89 | 0.98 |
| Data | 10 | 0.92 | 0.98 |
| | 100 | 0.91 | 0.98 |

As shown in Table 3, under the retrieval conditions of different data types, the results of the $MAP$-index test will not change with the change in the number of samples, and the values of the $MAP$-index are 0.98. This shows that there are no duplicate and redundant retrieval results in the retrieval results of the method in this paper, and the correct results are displayed in the first place.

In the network environment, the retrieval efficiency of micro-class teaching data of business English is also very important. The scale of network information will affect the retrieval efficiency of teaching data to a certain extent. In order to determine whether the method in this paper has an advantage in this respect, it can test the retrieval delay of micro-class teaching resources of business English by the method in this paper, the method in reference [8], the method in reference [9] and the method in reference [10] in the same experimental environment. This delay mainly reflects the time interval of feedback to users after resource retrieval. The test results are shown in Fig 8 -11.
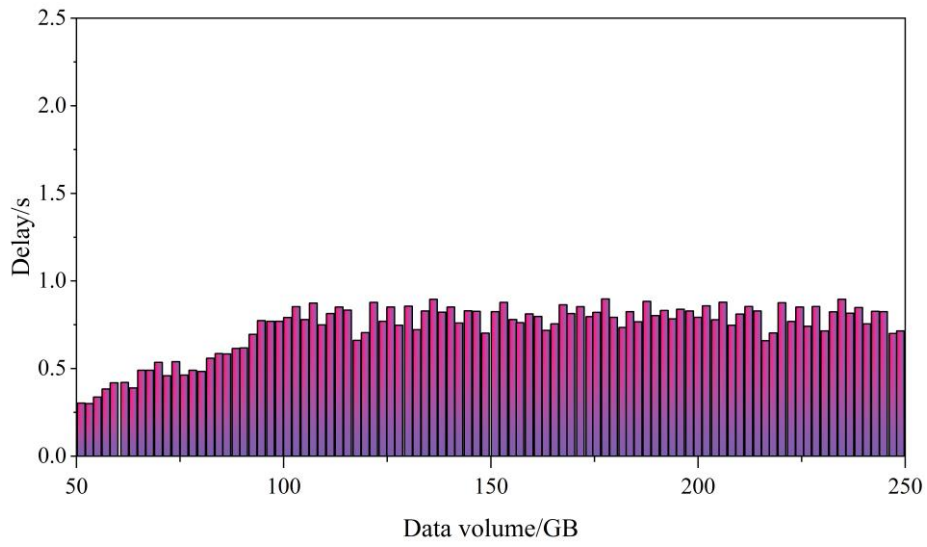
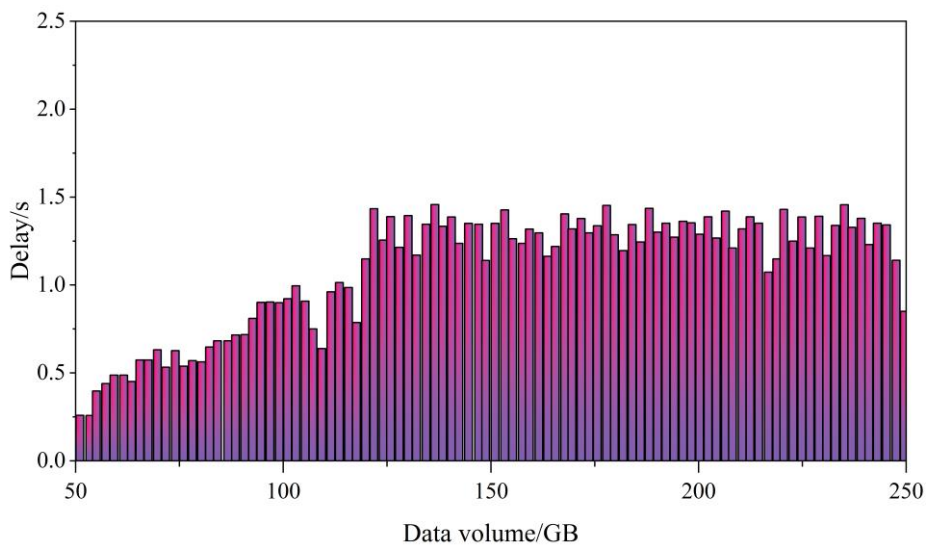Figure 8: The retrieval delay of this method



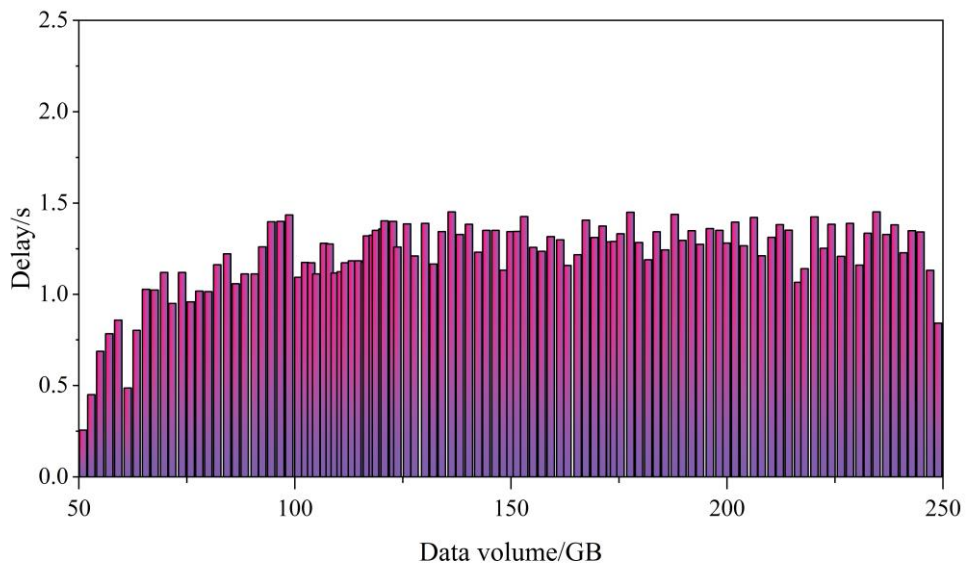Figure 9: Reference [8]method retrieval delay



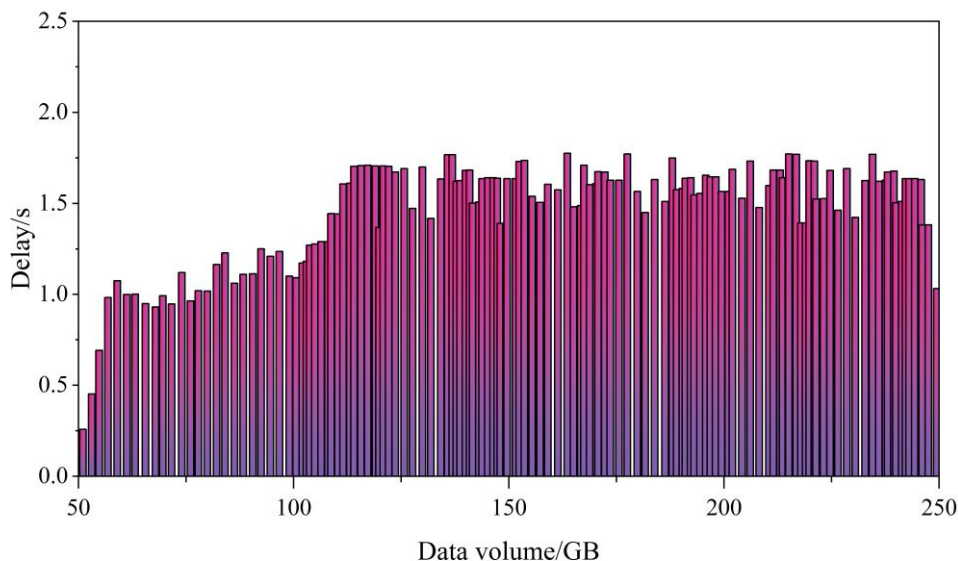Figure 10: Reference [9] method retrieval delay

Figure 11: Reference [10] method retrieval delay

By analysing Figs 8, 9, 10 and 11, it can be seen that the scale of network information will affect the retrieval efficiency of teaching data to a certain extent. Still, the impact on the retrieval efficiency of this article's method is insignificant. With the rise in the number of micro-class teaching data retrieval of business English, the retrieval time delay of this method does not exceed 1s. In contrast, in the same experimental environment, the retrieval delay of business English micro-class teaching resources of the ways in reference [8], reference [9] and reference [10] exceed that of the method in this paper. It can be seen that the method in this paper has advantages in the retrieval efficiency of large-scale business English micro-class data because this method can classify network information before data retrieval, ensure the order of information, and reduce the difficulty and complexity of subsequent retrieval.

Recall rate measures the ratio between the number of relevant documents retrieved by the system and the total number of actual relevant documents. The formula is:

recall rate=number of retrieved relevant documents/total number of real relevant documents. The higher the recall rate, the better the system can find relevant documents, providing more comprehensive information. The F1 score combines two indicators, recall and precision, to comprehensively evaluate the comprehensive performance of classification algorithms or information retrieval systems. Accuracy measures the ratio between the relevant documents retrieved by the system and all retrieved documents. The formula is: accuracy=number of relevant documents retrieved/total number of documents retrieved. The F1 score is the harmonic average of recall and accuracy, used to balance the recall and precision of the system. The formula is: F1 score=2 * (accuracy * recall)/(accuracy+recall). The F1 score ranges from 0 to 1, and the closer the value is to 1, the better the balance between recall and accuracy is achieved by the system. The test results of the methods in this article, reference [8], reference [9], and reference [10] are shown in Table 4.

Table 4: Test results of recall rate and F1 score

| Method | Recall | F1 Score |
|---|---|---|
| Method of this article | 0.90 | 0.85 |
| Reference [8] Method | 0.82 | 0.78 |
| Reference [9] Method | 0.84 | 0.80 |
| Reference [10] Method | 0.78 | 0.75 |

From Table 4, it can be seen that the method in this paper achieved the best results in terms of recall and F1 score. This method can more accurately retrieve resources related to business English micro courses, providing more comprehensive and high-quality search results. In contrast, the reference [8] method performs slightly worse in terms of recall and F1 score, while the reference [9] method and the reference [10] method also have slightly lower recall and F1 score. Taking into account the comprehensive performance of recall rate and F1 score, this method performs best after optimization and can

provide more accurate and comprehensive results for data retrieval tasks in business English micro course teaching.

Firstly, the research on business English micro course teaching data retrieval based on network information classification will be integrated as a new module of the education platform. This module interacts with existing platform databases through APIs and provides a user interface for retrieving and browsing business English micro course resources. Then, import the business English micro course teaching data into the database of the education platform, and process and transform it

according to the structured characteristics of the existing data to ensure consistency with the existing data. Secondly, develop a data retrieval module for business English micro course teaching, including search interfaces, query algorithms, and result display functions. By defining APIs to achieve data interaction and query request transmission with existing platforms. Finally, design a user-friendly search interface that provides functions such as keyword search, advanced filtering, and sorting. Ensure that the user interface is consistent with the appearance and functionality of existing platforms for seamless use by learners.

## 4 Conclusion

Based on the analysis of the necessity of the research on the data retrieval of business English micro-class, this paper studies the micro-class data retrieval method of business English based on network information classification. This method effectively uses the network information classification model based on the optimised SVM model and accurately classifies the micro-class network information resources of business English based on ensuring the performance of the support vector machine model. This link design can ensure the order and diversity of teaching data in the planned and complex network information resources. The method in this paper is the directed retrieval method of network teaching resources based on big data technology. The teaching data retrieval operation is done in the classified network information classification set to complete the directed retrieval of teaching data. Finally, a comparative experiment is used to verify that the method in this paper has an advantage in dealing with the problem of business English teaching data retrieval among similar retrieval methods.

With the passage of time, the collection of business English micro course teaching data continues to increase, including various forms of data such as video, audio, and text. As the size of the data increases, the system needs to process more data, which leads to an increase in query complexity and dataset size. When facing larger datasets or more complex queries, caching technology is used to store frequently accessed data, query results, or computational results in memory or cache to reduce the latency of subsequent queries.

## Data availability

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation."

## Conflicts of interest

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

## Acknowledgement

## Authorship contribution statement

Wang Guifang: Writing-Original draft preparation, Conceptualization, Supervision, Project administration, Methodology, Software, Validation.

## References

[1] L. Ma, "An immersive context teaching method for college English based on artificial intelligence and machine learning in virtual reality technology," *Mobile Information Systems*, vol. 2021, pp. 1–7, 2021.

[2] Y. Yin, "Microclassroom design based on English embedded grammar compensation teaching," *Math Probl Eng*, vol. 2021, pp. 1–9, 2021.

[3] E. S. Darowski, E. Helder, and N. D. Patson, "Explicit writing instruction in synthesis: Combining in-class discussion and an online tutorial," *Teaching of Psychology*, vol. 49, no. 1, pp. 57–63, 2022.

[4] S. Wang *et al.*, "Research on PBL teaching of immunology based on network teaching platform," *Procedia Comput Sci*, vol. 183, pp. 750–753, 2021.

[5] H. Zhao and L. Guo, "Design of intelligent computer aided network teaching system based on web," *Comput Des Appl*, vol. 19, pp. 12–23, 2021.

[6] T. Jiao, "Mobile English teaching information service platform based on edge computing," *Mobile Information Systems*, vol. 2021, pp. 1–10, 2021.

[7] H. Chen and J. Huang, "Research and application of the interactive English online teaching system based on the internet of things," *Sci Program*, vol. 2021, pp. 1–10, 2021.

[8] Y. Wang, H. Wang, J. Yang, and J. Chen, "Cross-model retrieval with deep learning for business application," in *Journal of Physics: Conference Series*, IOP Publishing, 2021, p. 032035.

[9] F. Ma, T. Sun, L. Liu, and H. Jing, "Detection and diagnosis of chronic kidney disease using deep learning-based heterogeneous modified artificial neural network," *Future Generation Computer Systems*, vol. 111, pp. 17–26, 2020.

[10] X. Xu, J. Tian, K. Lin, H. Lu, J. Shao, and H. T. Shen, "Zero-shot cross-modal retrieval by assembling autoencoder and generative adversarial network," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 17, no. 1s, pp. 1–17, 2021.

[11]     X. Cheng and K. Liu, "Application of multimedia networks in business English teaching in vocational college," *J Healthc Eng*, vol. 2021, 2021.

[12]     D. Jing and X. Jiang, "Optimization of computer-aided English teaching system realized by VB software," *Comput Aided Des Appl*, vol. 19, no. S1, pp. 139–150, 2021.

[13]     Y. Shu, "Experimental data analysis of college English teaching based on computer multimedia technology," *Comput Aided Des Appl*, vol. 17, no. S2, pp. 46–56, 2020.

[14]     L. Wang and Z. Xu, "An English learning system based on mobile edge computing constructs a wireless distance teaching environment," *Mobile Information Systems*, vol. 2021, pp. 1–9, 2021.

[15]     D. Amalia, Igaamo. IGAAMOka, V. Septiani, and M. R. Fazal, "Designing of Mikrokontroler E-Learning Course: Using Arduino and TinkerCad," *Journal of Airport Engineering Technology (JAET)*, vol. 1, no. 1, pp. 8–14, 2020.

[16]     R. M. Baker, M. E. Leonard, and B. H. Milosavljevic, "The sudden switch to online teaching of an upper-level experimental physical chemistry course: challenges and solutions," *J Chem Educ*, vol. 97, no. 9, pp. 3097–3101, 2020.

[17]     L. Liu, "Research on IT English flipped classroom teaching model based on SPOC," *Sci Program*, vol. 2021, pp. 1–9, 2021.

[18]     D. A. Wild, A. Yeung, M. Loedolff, and D. Spagnoli, "Lessons learned by converting a first-year physical chemistry unit into an online course in 2 weeks," *J Chem Educ*, vol. 97, no. 9, pp. 2389–2392, 2020.

[19]     F. Yang, "Design of Traditional Teaching Method of Micro-teaching Based on Blended Learning," in *e-Learning, e-Education, and Online Training: 6th EAI International Conference, eLEOT 2020, Changsha, China, June 20-21, 2020, Proceedings, Part I 6*, Springer, 2020, pp. 159–170.

[20]     F. Zhao, O. I. Fashola, T. I. Olarewaju, and I. Onwumere, "Smart city research: A holistic and state-of-the-art literature review," *Cities*, vol. 119, p. 103406, 2021.

[21]     A. Al-Hasan, "Effects of social network information on online language learning performance: A cross-continental experiment," in *Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries*, IGI Global, 2023, pp. 1574–1591.

[22]     J. Sun, L. Wang, J. Li, F. Li, J. Li, and H. Lu, "Online oil debris monitoring of rotating machinery: A detailed review of more than three decades," *Mech Syst Signal Process*, vol. 149, p. 107341, 2021.

[23]     Z. Xiao *et al.*, "Big data driven vessel trajectory and navigating state prediction with adaptive learning, motion modeling and particle filtering techniques," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 4, pp. 3696–3709, 2020.

[24]     W. Zhang and Z. Wu, "Optimal hybrid framework for carbon price forecasting using time series analysis and least squares support vector machine," *J Forecast*, vol. 41, no. 3, pp. 615–632, 2022.

[25]     M. Sehad and S. Ameur, "A multilayer perceptron and multiclass support vector machine based high accuracy technique for daily rainfall estimation from MSG SEVIRI data," *Advances in Space Research*, vol. 65, no. 4, pp. 1250–1262, 2020.

[26]     X. Yu and H. Wang, "Support vector machine classification model for color fastness to ironing of vat dyes," *Textile Research Journal*, vol. 91, no. 15–16, pp. 1889–1899, 2021.