Research Article

Construction of College English Teaching Information Resource Integration Model Based on Fuzzy Clustering Algorithm

Hua Deng

School of Foreign Studies, Liaoning University of International Business and Economics, Dalian 116052, China

Correspondence should be addressed to Hua Deng; denghua@luibe.edu.cn

Received 12 April 2022; Revised 5 May 2022; Accepted 17 May 2022; Published 24 June 2022

Academic Editor: Le Sun

Copyright © 2022 Hua Deng. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The existing college English teaching information resource integration methods have problems of low recall rate, low accuracy rate, and long integration time. Therefore, this paper constructs a college English teaching information resource integration model based on fuzzy clustering algorithm. The ant colony algorithm is used to optimize the topic crawler, the optimized topic crawler is used to collect the college English teaching information resources in the mixed teaching mode, and the weight and naive Bayes algorithm are combined to classify the collected resources. According to the results of resource classification, fuzzy clustering algorithm is used to construct the integration model of college English teaching information resources. The experimental results show that the recall rate of the proposed model is more than 95%, the accuracy rate is more than 94%, and the average resource integration time is only 0.71 s, indicating that the model has relatively reliable performance.

1. Introduction

With the rapid development of information technology, people’s learning environment and learning methods are constantly enriched and improved, and various educational ideas and viewpoints based on the information environment are constantly emerging and innovated. In such a wave, the combination of traditional instructional design and information technology is imperative [1]. Therefore, the reform of college English teaching has been changing to network-based technology-assisted teaching. Multimedia, tablet computer, mobile phone, blackboard platform, and other information tools pour into college English classroom, which has become an effective means to assist teachers’ teaching and enrich college English teaching mode. In this information technology environment, the defects of the traditional college English teaching model have been improved to a certain extent, and the original single plane teaching method has become more three-dimensional [2]. At present, college English teaching mode based on information technology environment can be divided into group learning mode, individual learning mode, and mixed learning mode. It refers to the combination of multiple media and interactive learning modes to provide students with more autonomous and interactive learning modes [3]. In today’s college English teaching, the first and second teaching modes are the most commonly used and common. Their advantages are obvious, but their disadvantages cannot be ignored. In this case, the mixed teaching mode of integrating college English teaching information resources is particularly important. Therefore, how to organize teaching links such as collective learning, autonomous learning, social sharing, and multiple evaluation layer by layer [4], combined with rich college English teaching information resources, and provide students with a more free and effective learning environment is a research topic of college English teaching model reform and an important proposition to effectively promote the development of college English teaching.

Information resources are an important research topic in college English teaching. Reference [5] proposes an information integration model of multimedia integrated learning resources based on cloud storage. The configuration attributes of association rule mining are used to obtain the integrated information of multimedia learning
2. College English Teaching Information Resource Integration Model Based on Fuzzy Clustering Algorithm

College English teaching information resources are a collection of various elements composed of various teaching contents and activities, including curriculum hardware resources, software resources, and educator resources. College English teaching information resources refer to all kinds of resources related to English language phenomenon and language culture carried by network media, as well as various media means such as e-mail and forum. The application of college English teaching information resources enriches the content of English teaching, promotes the diversification of English learning forms, and promotes the realization of English teaching objectives. College English teaching information resources have remarkable characteristics in terms of quantity, structure, connotation, type, and means of transmission, which include (1) various forms, rich content, (2) simple style, strong communication, and (3) convenient communication and collaborative learning.

2.1. Collection of College English Teaching Information Resources Based on Focused Crawler Technology. Focus crawler, also known as theme crawler or professional crawler, is a web crawler “oriented to a specific topic.” The difference between it and the commonly called crawler (general crawler) lies in that the focused crawler is a target-theme-driven and selective crawler, and subject selection is required in the implementation of web crawling [8, 9]. It tries to crawl only web pages that are relevant to the topic. According to the established target theme, focus crawler selectively visits the relevant pages on the web. It pursues the accuracy of network information rather than the coverage of network resources. Compared with the general-purpose crawler, the focus crawler is different in the following aspects:

1. Theme customization for the target capture. Different from general search engines, thematic search engines provide users with specific domain information on the web. So, it collects information to achieve the user’s goal, and theme customization is to let users define their own network information, generally to provide keywords, subject category, initial site, and so on.

2. Filter out the web page information unrelated to the theme. Web crawlers usually grab pages that contain a lot of information unrelated to a user’s customized topic. Therefore, focus crawler is required to perform theme filtering on web pages, remove irrelevant or less relevant content, and screen out web page information with high theme relevance [10].

3. Heuristic search strategies are usually used to evaluate URL links. The focused crawler adopts a heuristic search algorithm to calculate the relevancy of the links contained in the captured web pages according to the subject-oriented words or theme.
Different from web crawler, the work flow of focused crawler is more complex. It is necessary to filter the links irrelevant to the topic according to certain web page analysis algorithm, reserve the useful links, and put them into the URL queue waiting to be captured. Then, it will select the next web page URL from the queue according to a certain search strategy and repeat the above process until it reaches a certain condition of the system, as shown in Figure 1.

In addition, all crawler web pages will be stored by the system, and certain analysis, filtering, and build index, so that users later query and retrieve; for the focused crawler, the analysis results obtained in this process may also provide feedback and guidance for the subsequent grasping process [12]. In order to further improve the crawler effect, ant colony algorithm is used to optimize the focused crawler search. The core problem of focused crawler is how to optimize the priority order of URL queue effectively, so that the page corresponding to the preferentially crawled URL has higher topic relevance. A good focused crawler system must accurately predict the value of URL and pay attention to both short-term and long-term returns. Therefore, the calculation of URL value and selection of crawler strategy are particularly important, which determine the search efficiency and quality of focused crawler [13, 14].

In 1991, Italian scholar Marco Dorigo and other scholars proposed a new simulated evolutionary algorithm inspired by the behavior of ant colony foraging. This is ant colony algorithm, which is a typical example of swarm intelligence. On the basis of ant colony algorithm, this section proposes a search strategy ACO-FC that uses ant colony algorithm to guide the focused crawler. The basic idea is as follows:

In the web pages related to college English teaching information, there are pages $w_i$ and $w_j$ of super college English teaching information resources. If there is a link in $w_i$ to $w_j$, the ant in $w_i$ will decide whether to move from $w_i$ to $w_j$ according to certain conditions. Each link sequence represents a possible ant movement route [15]. Individual ants transmit information through pheromones during movement. Pheromones evaporate over time as the ant crawls. The crawling of ants between pages is divided into multiple cycles. In each cycle, an ant makes a series of moves between web pages until it finds the target resource and returns to the source point. After each crawling cycle, the ant colony updated the pheromone quantity on each path [16].

Assuming $V$ represents the collection of all pages and $E$ represents the collection of paths made up of links, then web pages (links) form a directed figure $G = (V, E)$. The integration structure of college English teaching information resources is shown in Figure 2:

According to Figure 2, the topic relevance value $R(P_k)$ of the page is calculated by the following formula:

$$R(P_k) = (1 - d) + d \left( \frac{\sum_{i=1}^{k} R(T_i)}{C_i} \right).$$  \hspace{1cm} (1)

Among them, $d$ refers to the adjustment factor, which is generally 0.8, $T_i$ refers to the set of pages linked to $P_k$, and $C_i$ refers to the link out degree of $T_i$, that is, the number of links from $T_i$ to other pages.

Suppose $d_{ij}$ represents the distance between pages $w_i$ and $w_j$, as calculated by the following formula:

$$d_{ij} = \frac{C}{\sum_{k=1}^{n} R(P_i)}.$$  \hspace{1cm} (2)

where $C$ is a coordination factor constant and $P_i$ is the set of pages passed from page $i$ to page $j$.

Hypothesis $\Delta \tau_{ij}$ represents the amount of pheromone left on path $e(i, j)$ in this loop. $\Delta \tau_{ij}^k$ represents the pheromone amount left in path $e(i, j)$ by the $k$-th ant in this cycle [17]. Then, the following formula is established as follows:

$$\Delta \tau_{ij}^k = \begin{cases} Q, & \text{if ant passthrough path} ij \text{ during this cycle,} \\ \frac{Q}{L_k}, & \text{if ant passthrough path} ij \text{ during this cycle,} \\ 0, & \text{otherwise,} \end{cases}$$  \hspace{1cm} (3)

where $Q$ is a constant and $L_k$ represents the length of the path taken by the $k$-th ant in this cycle [18], which can be expressed as follows:
where \( n \) represents the number of web nodes that ant \( k \) roams in this cycle.

Set the number of ants as \( m \), then \( b_{i(t)} \) represents the number of ants located at page \( w_j \) at time \( t \), and then \( m = \sum_{i=1}^{n} b_{i(t)} \). \( \tau_{ij}(t) \) represents the amount of pheromone remaining at \( e(i, j) \) at time \( t \). At the initial moment, the pheromone quantity on each path is equal, and set \( \tau_{ij}(1) = c \) (\( c \) is a constant, usually 0). Ant \( k \) decides the next path according to the pheromone intensity of each path in the process of movement. The probability \( P_{ij}^k(t) \) of ant \( k \) moving from position \( i \) to \( j \) at moment \( t \) is expressed by the following formula:

\[
P_{ij}^k(t) = \begin{cases} \left( \frac{\tau_{ij}(t)}{\sum\tau_{ij}(t)} \right)^\alpha \cdot \left( \tau_{ij}(t)^\beta \right) & \text{if } l(i, j) \text{ means there is a link from } w_i \text{ to } w_j \text{ on a given page } P_i. \text{ In order to prevent the ant colony from crawling in the loop and restrict the ants to explore the page incrementing, each ant stores a tabu table to record the links visited. If } w_j \text{ belongs to tabu, the path probability value } P_{ij}^k(t) \text{ from } w_i \text{ to } w_j \text{ is 0, thus preventing ant } k \text{ from exploring the link. At the end of each cycle, the tabu table is emptied [20].} \\
0, \text{ otherwise,}\end{cases}
\]

The imbalance of pheromone intensity \( \tau_{ij}(t) \) will dissipate over time. The pheromone retention coefficient is set as \( \rho (0 < \rho < 1) \), which reflects the persistence of pheromone intensity. And \( 1 - \rho \) is the degree of pheromone loss. When the movement times of each cycle limit are reached, pheromone quantity on each path should be updated according to formula

\[
\tau_{ij}(t+1) = \rho \tau_{ij}(t) + \Delta \tau_{ij}, \quad \tau_{ij} = \sum_{k=1}^{m} \tau_{ij}^k.
\]

However, in order to avoid the problems of “prematurity” and “local convergence” of ant colony algorithm, formulas (7) and (8) were adopted to adjust the pheromone quantity on each path in this paper:

When \( \tau > \tau_{\text{max}} \), the following formula exists:

\[
\tau_{ij}(t+1) = \rho^{-\varphi(n)} \cdot \tau_{ij}(t) + \Delta \tau_{ij},
\]

When \( \tau \leq \tau_{\text{max}} \), the following formula exists:

\[
\tau_{ij}(t+1) = \rho^{1-\varphi(n)} \cdot \tau_{ij}(t) + \Delta \tau_{ij},
\]

where \( \varphi(n) = n/c \) is a function proportional to the number of conversions \( n \), the more the conversions \( n \), the greater the value of \( \varphi(n) \), \( c \) is a constant, so the pheromone quantity can be updated adaptively according to the distribution of solutions, so as to dynamically adjust the pheromone intensity on each path, so that ants are neither over-concentrated nor over-dispersed, thus avoiding prematurity and local convergence, and improving the global search ability [21, 22]. The specific structure of the focused crawler is shown in Figure 3:

According to Figure 3, the implementation process of the focused crawler search strategy based on ant colony algorithm is as follows:

1. Given the initial hyperlink node, other links under it can be obtained to form the web node set of ant crawling.
2. \( m \) ants were placed in \( n \) different nodes to initialize the ant control parameters.
3. Set the initial concentration for each \( e \in E(i, j) \) path.
4. Put the starting node \( v(0) \) of the \( k \)-th ant into tabu.
5. Repeat until tabu is empty, \( n-1 \) times.
6. Calculate the path length of the \( k \)-th ant \( L_k \), and find the average value of the optimal length \( \bar{A} \).
7. For pheromone values on each edge, update them according to formulas (7) and (8).
8. At the end of a cycle, the optimal path is output to realize in-depth mining of college English teaching information resources.

### 2.2 Classification of College English Teaching Information Resources

It is assumed that college English teaching information resource \( X \) is regarded as a \( n \)-dimensional vector \( (w(t_1), w(t_2), \ldots, w(t_n)) \) in the vector space, where \( t_1, t_2, \ldots, t_n \) is \( n \) features of college English teaching information resource, which also represents the keywords after dimensionality reduction, and \( w(t_k) \) is the weight of the \( k \)-th feature in the college English teaching information resource.

As the probability ratio function can comprehensively investigate the situation of positive samples (belonging to this category) and negative samples (not belonging to this category), it has a good performance in the classification process. Therefore, the probability ratio function is selected in this paper, as shown in formula

![Figure 3: Specific structure of focused crawler.](image-url)
where $P(t_k|c_i)$ represents the probability of $t_k$ features appearing in resources belonging to category $c_i$ in the training sample set, while $P(t_k|\overline{c_i})$ represents the probability of $t_k$ features appearing in resources not belonging to category $c_i$. In summary, by calculating $\text{OR}(t_k, c_i)$, the paper can obtain the representative degree of feature item $t_k$ to category $c_i$. The larger $\text{OR}(t_k, c_i)$ is, the more representative it is. The output of probability ratio algorithm is the first $N$ feature words with the largest OR $(t_k, c_i)$ values, which are used as the feature database of training samples to participate in the real-time classification of college English teaching information resources, as shown in formula

$$\text{OR}_{\text{max}}(t_k) = \max_{i=1}^{X} \text{OR}(t_k, c_i).$$  \hfill (10)

However, the method of purely selecting the first $N$ feature words with the largest value of OR $(t_k, c_i)$ as the training sample feature database often has the problem of “indescribable” training samples; that is, some training college English teaching information resources do not contain any selected feature items.

For college English teaching information resource vectors $D$ and $Q$, this paper defined the similarity by using cosine values [23, 24] of vectors, as shown in formula

$$\text{Sim}(D, Q) = \frac{D \cdot Q}{\|D\| \cdot \|Q\|} = \frac{\sum_{i} (a_i \times b_i)}{\sqrt{\sum a_i^2} \times \sqrt{\sum b_i^2}}. \hfill (11)$$

When identifying the college English teaching information resource $d_j$ to be classified, this paper calculates the similarity between it and various types $\text{SCV}(d_j, c_i)$, as shown in formula

$$\text{SCV}(d_j, c_i) = \sum_{d_i \in Tr_k(d_j)} \text{Sim}(d_j, d_i) \cdot \Phi(d_j, c_i), \hfill (12)$$

where $Tr_k(d_j)$ represents the $k$ information sets most similar to $d_j$, and the following formula exists:

$$\Phi(d_j, c_i) = \begin{cases} 1, & \text{When } d_j \text{ belongs to class } c_i, \\ 0, & \text{When } d_j \text{ does not belong to class } c_i. \end{cases} \hfill (13)$$

Naive Bayes algorithm is a kind of module classifier with known prior probability and class-conditional probability. Its basic idea is to calculate the probability that resource $X$ belongs to class $H$ [25]. In Bayesian classifier, it is necessary to build a probabilistic classifier based on modeling word features of different classes. Then, the paper classifies the college English teaching information resources according to the words and the posterior probability of different types of college English teaching information resources. Naive Bayes is easy to implement and calculate, and its Bayes theorem formula is as follows:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}, \hfill (14)$$

where $P(H|X)$ is the posterior probability, representing the posterior probability of $H$ under condition $X$. $P(H)$ is the prior probability, which is the prior probability of hypothesis $H$. A posterior probability $P(H|X)$ contains more information than $P(H)$, and $P(H)$ is independent of $X$. Naive Bayes is a statistical method in the form of probability, which is the estimation of a group of probability parameters. Its purpose is to combine the probability distribution of classification and college English teaching information resources [26]. It is based on Bayes’ rule and relies on simple representation of resource information. The naive Bayes classifier is constructed by selecting the best classification through input vector, and the most possible classification is selected. It computes conditional probabilities for each category $S_i$ of resource $w$, where $f_j$ is the resource feature. The Naive Bayes classifier finds the class that maximizes the formula by selecting the most appropriate class. Assume that all features are independent of each other, as shown in formula

$$S = \arg \max_{i} P(S|f_1, \ldots, f_m) = \frac{\arg \max_{i} P(f_1, \ldots, f_m|S_i)}{P(f_1, \ldots, f)}. \hfill (15)$$

To sum up, by establishing perfect college English teaching information resources, colleges and universities will be able to further strengthen the efficiency of the integration of relevant teaching information resources and realize the scientific management of resources. Therefore, in the process of educational informatization construction, colleges and universities should increase investment, build a college English teaching information resource database as soon as possible, and ensure the resource reserve in the database through wider collection and integration of information resources. On this basis, through the powerful function of the database, realize the scientific and efficient management of information resources, and improve the efficiency of storage, retrieval, analysis, and utilization of relevant information resources. At the same time, the perfect construction of the database can also provide support for the sharing of information resources. Combined with the information exchange and sharing among colleges and universities, it can expand the source of college English teaching information resources and support the effective development of college English teaching.

2.3. Resource Integration Model Construction Based on Fuzzy Clustering Algorithm. Fuzzy C-mean clustering algorithm is a prototype-based clustering algorithm. In the calculation process, it is necessary to repeatedly calculate the distance between samples and each cluster center to determine the category of samples. Usually, different types of data samples need to adopt different types of distance measure functions. Therefore, to find the appropriate distance measure function is one of the core problems of fuzzy C-mean clustering algorithm. In practical problems, for the nonconvex complex data distribution and clustering of high-dimensional data samples, researchers have invested a lot of energy to explore a variety of distance measure functions suitable for fuzzy C-mean clustering algorithm. As a new dimension reduction method tool, manifold learning
algorithm aims to discover low-dimensional manifold structures embedded in high-dimensional data space and provide an effective low-dimensional representation. Currently, it has become a hot research issue in the fields of pattern recognition, machine learning, and data mining.

In this paper, based on the discriminant neighborhood embedding algorithm, a novel discriminant neighborhood embedding (FCM) clustering algorithm based on discriminant neighborhood embedding (FCM-DNE) is proposed.

Set $\{x_i, l_i\}_{i=1}^{N}$ as the dataset of $N$ samples of $c$ categories with class labels, where the characteristic dimension of $x_i \in \mathbb{R}^n, i = 1, \ldots, N$ is $n$, representing the $i$-th sample, and $l_i \in \{1, 2, \ldots, c\}$ represents the category of the $i$-th sample. A discriminant adjacency matrix $F = [F_{ij}]_{N \times N}$ is used to describe the relationship between labeled samples. The value of matrix $F$ is if two labeled samples are not adjacent samples, the value of $F$ between them is given 0; if two samples are adjacent and of the same class, the $F$ value between them is assigned to 1; if two samples are adjacent but not of the same class, the value of $F$ between them is assigned to 0. The expression of discriminating adjacency matrix $F = [F_{ij}]_{N \times N}$ is given as:

\[
F_{ij} = \begin{cases} 
+1, & x_i \text{ and } x_j \text{ belong to the same category,} \\
-1, & x_i \text{ and } x_j \text{ are not of the same kind,} \\
0, & \text{otherwise.}
\end{cases}
\tag{16}
\]

DNE algorithm hopes to find an optimal transformation, so that when each supervised sample point reaches equilibrium under the joint action of local attraction and local repulsion, the sample points within the class are as close as possible, and the sample points between the classes are as separate as possible. If the characteristic dimension of the sample is changed into $m$ dimensions, the optimal linear transformation matrix $P \in \mathbb{R}^{m \times n}$ needs to be found to minimize the following formula:

\[
\Psi(P) = \sum_{i,j} \|P^T x_i - P^T x_j\|^2 F_{ij}.
\tag{17}
\]

Substitute $F_{ij}$ into the above formula, and make $\Delta(P) = \sum_{i,j} \|P^T x_i - P^T x_j\|^2$. In the formula, $x_i$ and $x_j$ belong to adjacent classes, representing the sum of distances between all points of the same class, representing the intra-class compactness. $\delta(P) = \sum_{i,j} \|P^T x_i - P^T x_j\|^2$. In the formula, $x_i$ and $x_j$ belong to adjacent outliers, representing the sum of distances between all points of outliers and representing the divergence between classes. Therefore, cost function $\Psi(P)$ can be expressed as follows:

\[
\Psi(P) = \Delta(P) - \delta(P).
\tag{18}
\]

$\Psi(P)$ in formula (18) shows the difference between the values of compactness within a class and the values of divergence between classes. From the perspective of classification, the smaller the intra-class compactness value is, the better, while the larger the inter-class divergence value is, even if $\Psi(P)$ reaches the minimum transformation matrix $P$ is the optimal transformation matrix. In order to find the minimum value of cost function $\Psi(P)$, it is transformed into the following form:

\[
\Psi(P) = \sum_{i,j} \|P^T x_i - P^T x_j\|^2 F_{ij}.
\]

\[
= 2 \sum_{i,j} (x_i^T P P^T x_i - x_j^T P P^T x_j) F_{ij}
\]

\[
= 2 \sum_{i,j} \text{trace}((P^T x_i x_i^T P - P^T x_j x_j^T P) F_{ij})
\]

\[
= 2 \sum_{i,j} \sum_{i,j} \text{trace}((P^T x_i x_i^T P - P^T x_j x_j^T P) F_{ij})
\]

\[
= 2 \text{trace}(F^{T} X S P - P^{T} X F S^{T} P)
\]

\[
= 2 \text{trace}(F^{T} X (S - F) P)
\]

\[
= 2 \sum_{i=1}^{m} P_i X (S - F) P_i
\]

In formula (19), trace$(\cdot)$ represents the trace of the matrix, $X = [x_1, x_2, \ldots, x_N]$ is the matrix $n \times N$, the annotation sample set, and $P_i$ is the $i$-th column of matrix $P$. $S = [S_{ij}]_{N \times N}$ is a diagonal matrix:

\[
S_{ij} = \begin{cases} 
N \sum_{j=1}^{N} F_{ij}, & i = j, \\
0, & i \neq j.
\end{cases}
\tag{20}
\]

The column vectors of transformation matrix $P$ form the basis of the new space, which maps the input samples of $n$-dimensional features to the new $m$-dimensional feature space. Since the column vectors of transformation matrix $P$ form a basis for a new space, it can be assumed that the column vectors of transformation matrix $P$ are orthonormal to each other; thus, the problem can be transformed into the following constrained optimization problem:

\[
P_{\text{optimal}} = \arg \min_P \sum_{i=1}^{m} P_i^T X (S - F) X^T P_i
\]

\[
\text{s.t. } P_i^T P_i = 1, P_i^T P_j = 0, (i \neq j).
\tag{21}
\]

According to the eigenvalue theorem,

\[
\min \sum_{i=1}^{m} P_i^T X (S - F) X^T P_i = \sum_{i=1}^{d} \lambda_i.
\tag{22}
\]

In formula (22), $\lambda_i$ is the $i$-th eigenvalue of matrix $(X - F)X^T$, and $P_i$ is the eigenvector corresponding to the eigenvalue $\lambda_i$.

Since matrix $(X - F)X^T$ is not semidefinite, its eigenvalues may be greater than, less than, or equal to zero, that is, $\lambda_1 \leq \ldots \leq \lambda_d \leq 0 \leq \lambda_{d+1} \leq \ldots$, and $d$ represents the number of eigenvalues less than zero. When only $d$ eigenvalues less than zero are selected, the objective function reaches the minimum, and the column vectors of transformation matrix $P$ are composed of the eigenvectors corresponding to the eigenvalues of matrix $(X - F)X^T$ less than zero. After obtaining the transformation matrix $P$, the distance between any two samples in the new space is
The objective function is as follows:

\[ J_{\text{LFCM}}(U, S, P, X) = \sum_{i=1}^{c} \sum_{j=1}^{n} u_{ij}^m \| x_j - p_i \|_M^2, \]  

where matrix \( P \) is the eigenvector corresponding to the negative eigenvalue of matrix \( X(S - F)X^T \). Under constraint condition \( \sum_{i=1}^{n} u_{ij}^m = 1 \) for any \( j = 1, 2, \ldots, n \), Lagrange multiplier method is adopted to obtain

\[
\begin{align*}
    p_i &= \frac{\sum_{j=1}^{n} u_{ij}^m x_j}{\sum_{j=1}^{n} u_{ij}^m s_j}, \\
    u_{ij} &= \frac{1}{\sum_{k=1}^{c} \left( \| x_j - p_i \|_M^2 \right)^{1/(m-1)}}.
\end{align*}
\]

Since cluster centers in the integration of college English teaching information resources are generated from samples in the dataset, when cluster center \( v_i \) is determined, \( d_{ij} \) can be extracted from the above manifold distance matrix. For the \( j \)-th cluster center, the integration model of college English teaching information resources is as follows:

\[ J_i = \sum_{j=1}^{n} u_{ij}^m d_{ij}^2 = u_{i1}^m d_{i1}^2 + u_{i2}^m d_{i2}^2 + \cdots + u_{im}^m d_{im}^2. \]  

Given the manifold distance matrix \( D^M \) between \( u_{ij}^m \) and any two samples, then

\[
B_{\text{vec}} = \begin{bmatrix}
  b_{11} & b_{12} & \cdots & b_{1c} \\
  b_{21} & b_{22} & \cdots & b_{2c} \\
  \vdots & \vdots & \ddots & \vdots \\
  b_{n1} & b_{n2} & \cdots & b_{nc}
\end{bmatrix} = \begin{bmatrix} d_{11}^2 & d_{12}^2 & \cdots & d_{1n}^2 \\
  d_{21}^2 & d_{22}^2 & \cdots & d_{2n}^2 \\
  \vdots & \vdots & \ddots & \vdots \\
  d_{n1}^2 & d_{n2}^2 & \cdots & d_{nn}^2
\end{bmatrix} \times \begin{bmatrix} u_{11}^m & u_{12}^m & \cdots & u_{1m}^m \\
  u_{21}^m & u_{22}^m & \cdots & u_{2m}^m \\
  \vdots & \vdots & \ddots & \vdots \\
  u_{n1}^m & u_{n2}^m & \cdots & u_{nm}^m
\end{bmatrix}
\]

Cluster center \( p_j \) of category \( j \) is

\[
p_j = \left\{ x_i \mid i = \arg \min_k B_{kj} \right\}. \]  

According to the newly obtained clustering center, \( d_{ij} \) is extracted from the manifold distance matrix, the partition coefficient \( u_{ij} \) is recalculated, the iterative cycle is carried out until the clustering center does not change, and the integration results of college English teaching information resources are output.
3. Experimental Design

In order to verify the validity of the college English teaching information resource integration model based on fuzzy clustering algorithm designed in this paper, an experimental test is conducted. The overall test scheme is as follows:

(1) In order to ensure the authenticity and reliability of experimental results, experimental parameters need to be designed, as shown in Table 1.

(2) Experimental data

During the experiment, crawler technology was used to capture college English teaching information resources, and the captured resource data were taken as experimental sample data. In order to ensure the authenticity of the experimental results, no processing was done to the experimental data during the experiment.

(3) Evaluation indicators

The recall rate and precision rate of the integration results of college English teaching information resources and the integration time of college English teaching information resources are taken as evaluation indexes. Reference [5] model, reference [6] model, and the model in this paper are used as experimental methods to verify the practical application effects of different methods.

Recall rate refers to the ratio between the amount of information checked out from the database and the total amount. The higher the value is, the more comprehensive the integration result of college English teaching information resources will be.

The precision ratio is an indicator to measure the signal-to-noise ratio of a retrieval system, namely, the percentage of relevant references detected and all references detected. The higher this value is, the more accurate the integration result of college English teaching information resources will be.

The integration time of college English teaching information resources refers to the time consumed to complete the integration of college English teaching information resources. The higher the value is, the higher the integration efficiency is.

First, the recall rates of the college English teaching information resource integration results of reference [5] model, reference [6] model, and this model are compared, and the results are shown in Figure 5.

By analyzing the data in Figure 5, it can be seen that the recall rate of the college English teaching information resource integration results of reference [5] model is between 73% and 87%, and that of reference [6] model is between 75% and 89%. Compared with reference [5] model and reference [6] model, the recall rate of the college English teaching information resource integration results of the model in this paper is always above 95%, indicating that the results of the college English teaching information resource integration method are more comprehensive and better.

Secondly, the accuracy of the college English teaching information resource integration results of reference [5] model, reference [6] model, and this model is compared, and the results are shown in Figure 6.

By analyzing the data in Figure 6, it can be seen that the accuracy of the college English teaching information resource integration results of reference [5] model is between 66% and 90%, and that of reference [6] model is between 57% and 77%. Compared with reference [5] model and reference [6] model, the accuracy rate of the college English teaching information resource integration results of the model in this paper always remains above 94%, indicating that the method has a higher accuracy rate and more accurate integration results.

Finally, reference [5] model, reference [6] model, and this model are compared for the accuracy of the integration results of college English teaching information resources. The results are shown in Table 2.

Analyzing the data in Table 2, the average time for the integration of college English teaching information resources under reference [5] model is 1.85s, and the average time for resource integration under reference [6] model is 1.49s. Compared with the experimental method, the average resource integration time under this model is 0.71s. It is the shortest among the three models, indicating that the method has higher integration efficiency of college English teaching information resources.
English teaching information resources and better practical application effect.

4. Conclusion

Hybrid teaching can complement the advantages of traditional learning mode and network learning mode to achieve better teaching effect. In other words, blended teaching can give full play to the autonomy of teachers and students, and more fully reflect the initiative, enthusiasm, and creativity of students as the main body of the learning process. Therefore, the application of blended teaching in college English teaching has achieved good results. In the process of blended teaching, it is necessary to improve the quality of blended teaching in college English teaching based on a large number of information resources. Therefore, this paper constructs an integration model of college English teaching information resources based on the fuzzy clustering algorithm. The experimental results show that the model of college English teaching information resource integration of recall and precision is higher, to the college English teaching information resource integration time is shorter, and can solve the problems existing in the traditional methods, promote the teaching methods of optimization, thus further improve the teaching quality of mixing, and promote further development in the field of education.

The next step of work can be integrated use of English teaching in colleges and universities information resources, help college English teachers to strengthen the network teaching resources reserves, and provide support for English teaching innovation to explore, through the full use of the information resources at the same time, meet the needs of students’ autonomous learning, and enhance the level of college English teaching and effect. Under the background of education informationization construction, the innovation of network teaching mode supported by information resources should be actively explored to promote the innovation and upgrading of teaching mode and improve the quality of talents training in colleges and universities.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author indicates that there are no conflicts of interest in the study.

Acknowledgments

This work was sponsored in part by the SFLEP National University Foreign Language Teaching and Research Project “Research and Practice of College English Blended Teaching under the Background of Educational Informatization” (2018L00064A) and LUIBE Undergraduate Teaching Reform Research Project “Research and Practice of College English Teaching Model Reform Based on OBE and Information Technology” (2018XJJGYB02).

References


