Research Article

Application of Clustering Algorithm in English Proficiency Evaluation under the Framework of Big Data

Jiling Shang and Chaohui Liang

Department of Basic Teaching, Zhengzhou Railway Vocational and Technical College, Zhengzhou, Henan 451460, China

Correspondence should be addressed to Jiling Shang; 10462@zzrvtc.edu.cn

Received 9 May 2022; Revised 24 June 2022; Accepted 27 June 2022; Published 13 July 2022

Copyright © 2022 Jiling Shang and Chaohui Liang. 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.

With the advancement of science and technology as well as the continual improvement of big data analysis technology, the accuracy of traditional data information classification has declined, making it impossible to assess English ability effectively. A competency evaluation model for college English teaching vacancies is built using this information and the big data architecture. The ability of the big data information model is evaluated and feature information of ability constraints is extracted using the predefined constraint parameter index analysis model. Simultaneously, the K-means clustering algorithm is used to cluster and integrate a series of index parameters of English ability using big data, and the English teaching resource allocation plan is completed in accordance with this, allowing for the scientific evaluation of English teaching ability. The results of the studies show that the clustering method utilized in the context of big data can aid in the evaluation of English competence. In the experiment, four test cycles of English teaching skills were set up, and the effectiveness of the English evaluation techniques described in this paper and two classical cluster evaluation methods were compared and tested. The research shows that using the method described in this paper to evaluate English teaching skills can significantly improve the full utilization of data.

1. Introduction

The concept of big data is abstract, and although “big data” has been widely used, there are still various definitions of it, and there is no universally accepted expression yet. The general description refers to a complex, massive collection of data that cannot be extracted, stored, searched, shared, analyzed, and processed by existing software tools [1]. Education big data refers to big data in the field of education, which not only refers to the specific application of educational data mining, analysis, evaluation, and decision making in the field of education [2] but also refers to the continuous application in the field of education, which in turn helps big data technology to become an independent branch in the field of education, deepening the reform of education in a comprehensive manner and helping to improve the quality of education continuously [3]. There are many kinds of data related to education and teaching in colleges and universities, and usually each department has its own database. Due to the lack of top-level design on data collection and management, there is no constraint and limitation on data cleaning in each database [4]. At present, basically all the academic affairs systems of universities have the function of student evaluation, and the amount of evaluation data stored in the database of student evaluation can be as few as several million or as many as tens of millions. And these data are still rising sharply as time goes by and teaching activities continue to be carried out, and the amount of data is as large as can be imagined [5]. However, such a huge amount of data is left idle in the academic affairs system of many universities, and basically does not play any role. Even if they are used, they are basically applied in a simplistic way as a necessary condition for award evaluation, merit assessment, and title promotion. But this is only the tip of the iceberg for the real meaning and value of these data storage, which is far from playing its proper value [6].
can directly and figuratively grasp their learning progress [7]. Through the big data system, these information and data are recorded, sorted, counted and analyzed, so that teachers can more scientifically and comprehensively grasp students’ learning trends, and students and parents can timely and objectively understand individual development [7]. Through the big data analysis model, it focuses on the cross-sectional and horizontal evaluation of students’ English level [8]. At the same time, it carries out analysis and research, checks deficiencies and fills gaps, formulates personalized learning plans, and teachers implement customized teaching plans to guide and teach knowledge points to students’ English learning, so as to effectively improve students’ English level [9]. From the vertical time dimension, based on the change data accumulated by individual students in the learning process and relying on the horizontal subdivision of investigation points, the data change model of individual students in the learning process can be drawn [10]. By making a score map and comparing with the previous score map, we can realize the visual research on the whole learning process of the student, adjust the learning methods and teaching schemes at any time, strengthen the learning of weak points, and control the learning effect of the student [1]. As mentioned above, students’ learning level evaluation form can be drawn, and students’ learning trajectory can be clearly recorded [11]. If the circumstances allow, learning record files can be created, which can serve as a guide for students’ learning and development while also helping to develop their ability to evaluate, solve problems, and logical thinking [12].

With the qualities of real-time evaluation and technology, it can make decision-making more scientific, dependable, and comprehensive [13]. This paper is split into four sections. The first part focuses on the use of big data in English teaching. The second part discusses the use and development of big data and the K-means clustering algorithm, and the third part summarizes the use of big data and the principle and process of the K-means clustering algorithm using existing common technologies. To accomplish the clustering and integration of index characteristics of English teaching ability, a method based on big data fuzzy k-means clustering and information fusion is proposed in the fourth section. Part 5 tests and analyzes the accuracy of competency assessment of English teaching posts and the performance of relevant data analysis.

2. Related Work

In the traditional view, evaluating a teacher’s teaching competence and a student’s learning ability is a very subjective and abstract task that must be watched over time. And there are many influencing factors that will trouble the evaluation process, which requires the use of big data for quantitative analysis [14]. Ram et al. collected and organized the long-term accumulated online learning data process of a university, summarized the change trend of education and teaching data, predicted the law of change of e-learning data, and compared it with the traditional regression analysis model. Thus, it reveals the development trajectory of school education and teaching from a deeper level and provides scientific decisions for school development [15]. Chen and Zhao discusses the concept of educational evaluation, evaluation goals, evaluation content, evaluation standards, evaluation methods, and evaluation results under the background of educational big data [16]. Chen studies the traditional clustering algorithm, which needs to give the clustering center and the number of clusters in advance. However, there are many students and their scores are widely distributed. The evaluation of English ability in speaking, listening, reading comprehension, and writing is a high-dimensional problem. Therefore, the cluster center and indeed the number of nodes are difficult to identify. Different cluster centers will make the cluster results different [1]. Furthermore, establishing the centroids is difficult, owing to the lack of a strict theoretical basis for doing so, and the impossibility of deciding the volume of groups only based on people’s subjective consciousness [17]. Pasina et al. and others studied the k-clustering algorithm and applied it to the analysis of instructional analysis results of pedagogical quality evaluation. The clustering of indicators of instructional qualitative assessment results was performed to get different levels of division, and the clustering results were analyzed in detail to find out the implied relationship between indicators and final evaluation results, which changed the shortcomings of the traditional classification based on average score ranking [18]. The educational technology ability of college English teachers is a major reflection of their teaching capability in the informationized environment of English education. How to effectively evaluate this ability and reasonably cluster it is the primary work to promote teachers’ professional development. Yastibas hopes to conduct targeted training and observation of university teachers in the future to promote the effective advancement of pedagogy in education technology [19]. Based on the study of the basic principles of agglomerative and corresponding algorithms, Rahmi focuses on the advantages and disadvantages of several clustering algorithms. Using college English test scores and python language, Rahmi wrote and built a program to apply to students’ test score dataset, gave experimental results, and obtained some interesting experimental result data to study the effectiveness of blended education in Chinese English [20]. Sun uses Python language to implement and improve the k-means algorithm model and establishes the basic model of score analysis on this basis. The relevant attributes of iris data set are analyzed, and then the attribute differences of different classes are found. Then, this difference with the quantity called “weight” in reflected, and the weight is added to the clustering calculation process to realize the improvement of K-means algorithm [21]. Kurada tested the iris data set and the experimental results show that the clustering accuracy of the built model on Iris data set is improved to more than 96%, and the clustering results are relatively stable and the running time is significantly reduced. Finally, the improved k-means algorithm model is used to mine the English score data, which brings objective guiding opinions to the relevant Chinese and English teaching implementers, so as to make the data mining play an auxiliary role in Teaching [22].
In summary, it is evident that in English proficiency assessment, the existing methods of assessment are rather homogeneous. It is mainly to calculate the average scores of students taught by English teachers, and then to rank their scores from the highest to the lowest, and finally to arrive at the evaluation results. Obviously, this assessment method cannot truly reflect the differences in students’ specific English theoretical basis and practical application ability. Traditional evaluation algorithms suffer from issues such as unclear sorting of the large amount of data messages, large errors in analytical models, and limitations. As for the research on English proficiency assessment, the current research results are mostly based on statistical data, rarely combined big data technology, and clustering algorithm for diversified analysis. Therefore, under the big data framework of today’s era, the research of using clustering algorithm to evaluate English ability is of great significance.

3. The K-Means Aggregation Technique in the Framework of Big Data

3.1. K-Means Clustering Algorithm. The word “class” in agglomeration refers to a set that has similarities. Aggregation is the division of a data set in categories, so that the data within the class is the most similar, and the data similarity difference between various classes is as large as possible. Agglomeration analysis is an unsupervised learning that classifies data sets based on similarity. The cluster parsing is a common type of algorithms we use in digital marketing. It is frequently used for studying samples that are not classified but have relevant similarities. It includes three algorithms: K-means, k-center point, and system clustering. Each has its own characteristics and applicable environment. What is k? Unlike KNN, K-means clustering belongs to unsupervised learning. So what is the difference between supervised learning and unsupervised learning? Supervised learning methods must have training sets and test samples. Find a rule in the training set, and use this rule for the test sample. However, unsupervised learning has no training set and only one set of data, so we can find rules in this set of data.

The K-means clustering algorithm originated from signal processing and application, and is currently used more in the field of data mining. The core idea of this algorithm is to cluster objects with n data sets into k categories according to the principle of the closest distance, so that each object can be classified into the cluster corresponding to the mean value closest to it.

3.2. Process of K-Means Clustering Method

3.2.1. Initializing the Clustering Centers. The first step is to select empirically. Based on the specific problem, select k more appropriate examples from the sample set as the preliminary clustering centers. The second step is to use the first k samples as the initial clumping centers. The third step is to divide a random group of all samples into k classes, computing each class’s sample averages, and using the value of the samples’ averages as the initial clustering centers. The fourth step is the density method. Take each sample as the center of the sphere and make a spherical neighborhood with a positive number as the radius. The number of samples falling in the neighborhood is the point density, and the point with the highest density is selected as the first initial cluster center [23]. The next high density points are determined outside the specified distance from the first point to avoid the aggregation of the initial cluster centers. The fifth step is to generate the initial cluster centers of K cluster divisions from the solutions of k − 1 cluster divisions. Firstly, all samples are regarded as a cluster, and the cluster center is the total mean value of the samples. Then, it is determined that the clustering center of the two clustering problems is the total mean value of one clustering problem and the farthest point from it and so on.

3.2.2. Initial Clustering. The first step is to classify the samples into the categories represented by each clustering center according to the principle of proximity. The second step is to take a sample, classify it into the category with its nearest cluster center, recalculate the sample mean, and update the cluster center. Finally, it is necessary to judge whether the division of K cluster centers and all data points has changed. The basic process of k-center algorithm is to select a representative object for each cluster. The remaining objects are assigned to the cluster represented by the nearest representative object according to their distance from each representative object. Then, the nonrepresentative object is used repeatedly to replace the representative object to optimize the clustering quality. If so, repeat the previous steps. If there is no change, the division ends. The process is shown in Figure 1.

4. Research on English Ability Assessment Algorithm

4.1. Big Data Analysis Model for English Teaching Ability Evaluation. Most of the traditional teacher evaluation methods are based on professional titles. Such as junior teachers, intermediate teachers, senior teachers, and other different titles. Teachers overemphasize the subject status of teachers in the teaching process, teachers become the dominant role in the classroom, and students belong to the object status, which makes teachers ignore the actual feelings of students in the classroom. Therefore, teaching evaluation is impossible to implement. However, these evaluation schemes cannot accurately and effectively evaluate the ability of English teachers. In addition, data processing is relatively difficult, and the accuracy of data information classification is low. In this paper, a teaching competence assessment system for college English teachers is optimized and designed under the framework of big data with the guidance of clustering algorithm. The classification accuracy of the data information of the system has been significantly improved, and it can well evaluate the personal ability of English teachers. This paper emphasizes that the cultivation of students’ language skills, language knowledge, emotional
attitudes, learning strategies, and cultural awareness should be taken as the evaluation criteria for senior high school English teachers. The research puts forward the new curriculum standard to promote teachers’ continuous improvement and emphasizes teachers’ analysis and reflection on their teaching behavior. Establish an evaluation system based on teachers’ self-evaluation, with the participation of principals, teachers, students, and parents. To achieve this function, the first step is to use data mining technology to acquire the binding parameters, i.e., integrate resources, collect informational data, and complete information sampling of English job competency binding parameters. A series of fusion and quantitative analysis of the sampled information was conducted to finalize the statistical analysis of English proficiency. The main expression and output of English proficiency assessment in this paper are the parameter indicators and parameters of English postability, including teaching attitude, teaching skills, extracurricular links, expression ability, communication skills, oral English, listening, reading comprehension, vocabulary, writing, and other English teaching skills. On this basis, the message flow model with binding parameters is completed.

\[ y_n = y(u_0 + n\Delta u) = i[a(u_0 + n\Delta u)] + \xi_n. \]  

The multivariate function of values for the evaluation of English teaching job proficiency is shown as \( i(\cdot) \), and the rating error measurement feature is formulated as \( \xi_n \). The correlation fusion approach is used to complete the calculation of the solution vector for English education post-competency evaluation in the high-dimensional feature distribution space, and then the feature training subset for postcompetency evaluation is obtained. \( T_i (i = 1, 2, \ldots, M) \) represents the feature training subset, as shown in the following formula:

\[ \sum_{i=1}^{n} \text{diag}(\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n), \quad \varepsilon_i = \sqrt{\mu_i}, \quad \forall i \neq j. \]  

In English postcompetency evaluation, \( y_{yn} = y y_m(1 - y_n) \) represents the conjoint solution of the statistically informed market model, and it meets the first impression characteristic breakdown requirement.

\[ V = \{ v(t), v(t) \in Y, [v] \leq d, t \in I \} \text{ and} \]

\[ \{I_n\}_{nN} = \{y_{1n}, y_{2n}, \ldots, y_{mn}\}. \]

The data information flow model of English post-competency evaluation is developed according to the statistical measurement value, and \( y_m \) indicates its statistical characteristic distribution sequence. The following formula shows the specific expression:

\[ d_{1y}(\nu) = F\{y(n)\} = 0, \]

\[ d_{1y}(\nu) = F\{y(n)y(n + \nu)\} = m(\nu), \text{ and} \]

\[ d_{k\gamma}(v_1 + v_2 + \ldots + v_{k-1}) = 0, \quad k \geq 3. \]

There is a convergent solution for job competency assessment with the constraint shown in the following equation:

\[ y(\xi) = \ln \varphi_{\xi}(\xi) = -\frac{1}{2}\xi^2. \]  

The number of clusters \( K \) is determined by preclustering. The similarity based on frequency and K-means clustering are used to precluster 1–7 categories, respectively, and the relationship between the sum of squares of minimum error and the number of clusters \( K \) is obtained as shown in Figure 2.

It is obvious from the above figure that the inflection point of the image appears when \( K = 3 \). Therefore, it is more appropriate to gather three categories of teaching evaluation results for students in a semester. However, it is important to note that the \( \nu \)-value is related to the amount of data in the sample data set and the dimensionality in each sample data set. In order to prove this, the same method is used to precluster the teaching evaluation data of all teachers in a semester of \( Y \) school. The preclustering results show that the inflection point of the image appears when \( K = 7 \). Therefore, for different types and scales of problems, the value of \( K \) will be different, which should be treated with caution in specific problems.

4.2. Quantitative Recursive Analysis of Teaching Ability Evaluation. A huge data analysis model of English teaching ability is created, statistical methods for quantitative analysis.
are utilized, and an English teaching ability prediction function model is created. The following is the specific objective function:

\[
\max_{y_b,c,e,q} \sum_{b \in B} \sum_{c \in C} \sum_{e \in E} \sum_{q \in Q} y_{b,c,e,q} Y_q,
\]

(7)

s.t. \[\sum_{b \in B} \sum_{c \in C} \sum_{e \in E} \sum_{q \in Q} y_{b,c,e,q} Y_q c_l(t) \leq L_c(t), \quad c \in C. \]

(8)

When the initial value of disturbance characteristics is known, \( \{ y_j \}_{j=1}^n \), the likelihood distribution functional of English teaching skill prediction and estimation is produced as follows:

\[ v_d(u) = L y_b(u). \]

(9)

Statistical model of English education job competency evaluation in a highly dimensional Eigen distributive domain has a continuous function of \( f : J \times JS^p \rightarrow JS \). After iteration, \( l - 1 \) times, \( l \geq 1 \), and the ability of the position to assess the grayscale sequence is consistent with \( O(l) \leq M \). Specifically, the expression of the big data clustering objective function is as follows:

\[ K_n(V, W) = \sum_{j=1}^n \sum_{i=1}^d y_{ji}^n (e_{ji})^2. \]

(10)

Integration with the K-value improvement approach, \( b_0 \) represents an initial capacity assessment sampling amplitude, and the calibrated scalar time sequence is denoted by \( y_{n-1} \), and \( c_j \) represents the value of oscillating decay of competency assessment, and then the result of extracting quantitative recursive characteristics for competency assessment of educational positions is expressed as follows:

\[ y_o = b_0 + \sum_{j=1}^{N_u} b_j y_{n-1} + \sum_{k=0}^{N_g} c_j n_{u-k}. \]

(11)

4.3. Application of the Improved Clustering Algorithm in the Evaluation of English Teachers’ Teaching Conditions. Since it is difficult for students to quantify the teaching ability of teachers in specific evaluations, basically all teaching evaluation systems use graded teaching evaluation data such as “excellent, good, medium, pass, and fail” rather than convenient teaching evaluation data. Calculated numerical data, which is difficult for some managers who are not computer-related majors to deal with. Traditional evaluation algorithms cannot effectively classify data.

A fuzzy K-means aggregation based on big data is proposed as an English language competency evaluation model to improve the evaluation of English teachers’ teaching competencies. The least square question is to find a consistent estimate of the resource bound traffic force \( \chi \) for the ELT proficiency evaluation, so as to minimize \( \| Z - Y \chi \| \), where \( \| \cdot \| \) is the \( F \)-norm in the European norm. Entropy features of the characteristic message of ELT proficiency restrictions are extracted as follows:

\[ Q_{loss} = 1 - \frac{1 - q_0}{\sigma} = q_0 + \sigma - 1 = \sum_n Q_{n,\sigma}. \]

(12)

The estimated equation for ELT proficiency was converted to the minimum share of the least square solution as follows:

\[ a(u) = y(u) + jz(u) = b(u) f^{(u)} + n(u). \]

(13)
where \( y(u) \) is the real part and \( z(u) \) is the dummy number portion.

Using the alternative datum technique to randomize the amplitude of ELT proficiency, we obtained \( y^*(k) \). ELT proficiency utilization rate may be represented as follows:

\[
V_{util} = \eta^T.
\]

Data mining technology and fuzzy filling method is used to construct English ability evaluation feature quantity, and the function calculation of resource distribution similarity is completed as follows:

\[
\text{Sim}_1(e_i, e_{ij}) = \frac{\sum_{k=1}^{N} W_{jk} \times W_{lik}}{\sqrt{\sum_{k=1}^{N} W_{jk}^2} \cdot \sqrt{\sum_{k=1}^{N} W_{lik}^2}},
\]

where \( e_i \) is a vector of eigenvectors of the pilot and \( e_{ij} \) is the centroid vectors of K-means aggregation. Exported expression for the integration of teaching resource information is as follows:

\[
Q(x|y) = \frac{Q(y|x)}{Q(y)}.
\]

When clustering and integrating the index parametrization on the basis of quantified recursive characteristics \( (O(j) \mod M) < n \), the corresponding assignment scheme of pedagogical sources and pedagogical competencies is finished, thus optimizing the competency assessment of English teaching positions and raising the precision and usability of the assessment.

5. Result Analysis and Discussion

5.1. Selection and Analysis of Teaching Evaluation Data.

Through big data technology, it allows to keep track of success and achievements of students’ English level in detail, accumulate over time, and more truly describe the characteristics of students’ development and the needs of personal planning. From the perspective of the value orientation of comprehensive quality evaluation, its core value still lies in promoting the development of students. It is only a mean for college selection, not the core goal. Education big data can also be used to find potential problems in class management, so as to give early warning and timely intervention. Education big data can acquire and analyze teaching parameters in a timely and rapid manner, ensuring that more people receive information in a short time and can quickly analyze the teaching. As shown in Figure 3, the school or head teacher can find the problem students that cause serious consequences in time. Give guidance and help in time, and promote the all-round development of students. Instead of waiting until it is irreparable for many years, tell the students that your weakness in that year is partial to the subject.

Comprehensive quality evaluation should enable students to better understand themselves, schools to better understand students and their own services, and the government to better understand schools and their own management. Big data technology is used for evaluation, which can more easily record students’ growth data and make digital portraits for students. As shown in Figure 4, it is a digital portrait made according to big data. It can be seen from Figure 4 that Lucy’s oral and listening skills are weak and her subjects are obvious. Big data can make students’ growth more clearly expressed.

The student evaluation data used for the ELT evaluation data selection were obtained from real data in the university’s executive educational management system. Student evaluations were organized and implemented in the form of online evaluations. Also, to facilitate the students’ completion of the teaching assessment, weightings of the evaluation indicators were generated from the students’ and teachers’ questionnaires and calculated according to the weighted average method. According to the questionnaire survey, several major factors of English teachers’ ability evaluation are as follows: quality, teaching attitude, oral English, teaching skills, extracurricular links, expression ability, communication skills, and others, as shown in Figure 5.

Comprehensive teaching evaluation management consists of two functional modules: online teaching evaluation and teaching evaluation system management. When users access the online teaching evaluation module, the system will provide the corresponding teaching evaluation content according to the user’s role and authority. Integrated teaching and learning assessment management is the core of the system’s operational processing. The process of teaching quality evaluation is carried out in this module. The educational administration manager formulates the teaching evaluation link through the system. After the user logs inside of the website, only the teaching evaluation screen is shown and various evaluations are selected. Based on the various types of user roles, the evaluation is divided into two ways: student evaluation and peer evaluation of teaching. Peer evaluation of teaching is divided into teacher mutual evaluation, expert evaluation, and educational administration supervision evaluation. Student users can sign in and fill out a subjective evaluation of the course based on the course they have taken. The example usage chart of student teaching evaluation is illustrated in Figure 6.

5.2. Cleaning of Students’ Teaching Evaluation Data.

In the activity of students’ teaching evaluation, due to various reasons, it is inevitable that there are some abnormal data. If it is not processed, it will affect the judgment of the actual teaching situation of teachers. Therefore, it is necessary to clean up instructional evaluation data of students and eliminate unfamiliar data in the sample set to ensure the validity and truthfulness of the assessment.

The phenomenon of scientific data distortion often occurs in the classroom. The main reason is closely related to the unclear experimental purpose, improper operation, wrong steps or incomplete analysis of the results. Teachers should pay attention to the problem of abnormal data. By creating learning situations, asking questions to guide students to actively explore, and using a variety of material presentation methods to reduce the frequency of abnormal
Figure 3: An example map of students who were found to be partial to subjects in time.

Figure 4: Big data for digital portrait of English learning process.
Figure 5: The proportion of English teachers’ ability evaluation factors.

Figure 6: The process map of students’ evaluation of English teacher’s ability.
data. Then, help in improving the ability of data collection, collation, and analysis in teaching. The abnormal data in students’ teaching evaluation data is often due to the fact that individual students do not evaluate a teacher objectively and fairly, but with strong personal color. This leads to a significant bias in their evaluations versus those of other students. Therefore, it is necessary to eliminate these anomalies in the data from instructional assessment. According to the results and analysis of abnormal students’ teaching evaluation, 237,924 student teaching evaluation records of 1326 classified sample data files of the college are eliminated, as shown in Figure 7.

5.3. Simulation Experiment Analysis. Simulation analysis using MATLAB. The teaching ability rating has a discussion valve of $E_x = 2$. Let parameters related to the distribution of ELT materials be $\psi_{d_1}^2 = (3/5)$, $\psi_{d_2}^2 = (2/5)$, $\psi_{d_3}^2 = (2/5)$, $\max h_{d_1}(e_2) = (6/5)$, $\max h_{d_2}(e_2) = (3/8)$, $\max h_{d_3}(e_2) = (1/10)$, sampling frequency $g_0 = 600$ Hz, adaptive initial step $\sigma = 0.97$, and the distribution of characteristics of teaching aids has a correlation factor of $C = 1.14$. According to the above parameter settings, the large data distribution time waveform is obtained, which is illustrated in Figure 8.

To determine the capacity to educate English, the above parameters are statistically examined using big data. Four cycles of English teaching skill testing were set up in the experiment, and the effectiveness of the English evaluation technique described in this work, as well as the two classic clustering evaluation methods, were compared and tested. Figure 9 depicts the teaching ability evaluation accuracy and teaching resource utilization rate for the three evaluation techniques. The research shows that adopting the method described in this paper to assess English teaching skills can significantly enhance adequate and data usage.

![Figure 7: Clustering results of English teachers’ teaching status after removing outliers in a semester.](image)

![Figure 8: Big data time domain distribution.](image)
6. Conclusion

Currently, almost all universities have an educational management system with a function for students to evaluate teaching. The amount of evaluation data stored in the student evaluation of teaching database is as little as several million, as many as tens of millions, and these data are still rising sharply with the passage of time and the continuous development of teaching activities. The volume of data is enormous. This study creates an idealized English ability evaluation model based on big data technologies, which is based on systematic analysis and research. The model is based on big data technologies and is targeted at a new generation of students. It performs process-oriented learning analysis and educational service monitoring by collecting and analyzing multidimensional structural and nonstructural data. The anomalous teaching evaluation data is deleted using the enhanced cosine technique, and the data is standardized using the normalizing algorithm, based on preliminary analysis and transformation of teaching evaluation data of students at a specific university. The teaching evaluation data is analyzed using the K-algorithm, and the three models are predicted and simulated using the MATLAB tool. The findings suggest that the models are adaptable.

Data Availability

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

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References


