

Research Article

Mobile Information System of English Teaching Ability Based on Big Data Fuzzy *K*-Means Clustering

Yanfei Miao 

Department of Basic Education, Luxun Academy of Fine Arts, Shenyang 110000, Liaoning, China

Correspondence should be addressed to Yanfei Miao; miaoyanfei@lumei.edu.cn

Received 20 May 2021; Revised 10 June 2021; Accepted 23 June 2021; Published 1 July 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Yanfei Miao. 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 implication of mobile English teaching is that English teachers and students use mobile devices for English teaching and communication at the same time. In order to accurately evaluate language interpretation skills, it is necessary to construct a mobile information system sampling model of the restrictive factors of language interpretation skills. Then, the nonlinear information fusion method is combined with the time series cognition method to make a statistical cognition of language interpretation skills. The parameter of language interpretation skills constraint is a set of nonlinear time series. To this end, this paper studies the language interpretation skills mobile information system, proposes language interpretation skills, and constructs the constraint parameters of the language interpretation skills evaluation and cognition using an indicator cognition model. The quantitative recursive cognition method analyzes the language interpretation ability evaluation model and the entropy feature of language interpretation ability and extracts the constraint feature information. The combination of large-scale data information fusion and *K*-means clustering algorithms provides indexing and integration of index parameters for language interpreting skills. On this basis, the corresponding allocation scheme of teaching resources is formulated to realize the assessment of language interpretation skills. The experimental results of related big data clustering algorithms show that the English teaching method proposed in this paper is highly effective, and the evaluation accuracy and teaching resource utilization rate have been increased by 5% and 6%, respectively.

1. Introduction

The information revolution represented by the Internet has greatly changed people's lives. In recent years, with the continuous expansion of application fields of technologies such as the Internet and cloud computing, a large amount of data has been continuously generated. Today, big data has entered an explosive stage of development. As an important strategic resource, big data is also researched, developed, and implemented in China. As language interpretation continues to deepen, the assessment of language interpretation skills becomes even more important. Among them, the most basic indicator of language interpretation skills is the ability to store and use English vocabulary. These data are also important indicators for assessing the quality of language interpretation. In other words, language learning ability is the ability to use

vocabulary. In the traditional language interpretation skills assessment method, a manual evaluation method is used. The advantage of the manual evaluation method is that the evaluation process is very flexible, but the manual evaluation has a certain subjective awareness. At the same time, in the case of a certain mixing factor, it is easy to cause a large evaluation error; and there are no specific evaluation criteria in the evaluation process, which leads to some differences in the evaluation criteria of the same person.

With the rapid development of the Internet, people's demand for information service systems is increasing. The mobile English teaching ability information service platform can bring convenience to people and improve people's quality of life.

The assessment of language interpretation skills is subject to many constraints. Therefore, it is necessary to

quantitatively test and analyze the level of language interpretation. At present, many schools use the pass rate of the fourth and sixth grade exams as the basis for assessing the overall quality of the school. From the perspective of the administrative management department, ranking schools by simple pass rate is likely to cause comparisons between schools and create inappropriate intangible pressure. In fact, the conditions for running a school are different, students are different, and there is no simple basis for comparison. Starting from the school, many universities unilaterally pursue the pass rate, but do not meet the specific industry qualification standards for university language interpreting. Moreover, the pass rate of the test can only reflect how many students have achieved the passing level of CET-4 or CET-6. This does not reflect the level of compliance of these students with the corresponding requirements of the syllabus. From the student's point of view, the goal is to obtain a certificate, rather than improving the actual use of English. Ignore the improvement of my own language ability, but I forgot the fundamental purpose of learning English. First, the established mean variance model is introduced as a mathematical tool into big data research. Neural networks, decision trees, and other methods are then introduced into quantitative transactions, enabling quantitative transactions to grow rapidly. Research on the evaluation of language interpretation shows that the historical market has certain guidance and vision for the characteristics of future market trends. This makes it possible to use pattern recognition to analyze language interpretation skills.

In a big data analytics environment, computer software programs face a large number of information processing tasks; and data processing is performed for information update. However, in some software applications, historical parameters need to be retained for subsequent software self-tests. In order to implement the parameter self-holding function, it is necessary to update and calculate data based on the original historical data model. The incremental data analysis calculation model is a common method. Apply the incremental depth calculation model, which automatically builds a new database based on the original database and encodes and supplements the internal information. This series of tasks is done automatically in the network environment. At the same time, the incremental data processing model can also adapt to the information acquisition characteristics in the big data network environment and is more effective in resource acquisition and request docking. The updated database repository does not conflict in automatic information retrieval. Instead, it can coexist in the application software in a fused way. When the program software needs to transfer historical data and updated data, the internal database will automatically match, and the required data will be sent to the program control center through the transfer office. Historical data and updated data under normal conditions are stored independently of each other. It can be seen that the method has strong information timeliness and is greatly improved in the big data analysis environment for automatic updating of models and automatic retrieval of data information.

Aiming at the problem of inaccurate classification of language set information by traditional language

interpretation ability evaluation algorithms, a language interpretation ability mobile information system based on language set fuzzy K -means clustering and resource exchange is proposed. First, establish a constraint parameter analysis model; second, use a quantitative recursive analysis method to evaluate the skills of the language set information model to achieve the skills of constraint feature information entropy feature extraction. Finally, use language set resource exchange and K -means clustering algorithm for fusion. Realize the clustering and integration of the language translation skills index parameters, and formulate the corresponding teaching resource allocation plan to realize the language translation skills mobile information system. The test results show that this method is used in the language translation skills mobile information system and has a good ability to exchange analytical skills resources, so that the interpretation skills can improve the accuracy of the mobile information system and the efficiency of teaching resources. The innovation of this article is that, through combining the theory and literature, it can be seen that the research on mobile learning and teaching design in the mobile learning environment is gradually increasing, which reflects the current teaching practice of the majority of educators. Enthusiasm is high but also feasible. Through research in English teaching practice, it is found that, with the popularization of smart mobile terminals and the development of mobile Internet, it is very easy to construct a mobile learning environment. This can be seen from our previous interviews and questionnaire surveys. What teachers should focus on is how to make full use of the mobile learning environment for teaching design and make it possible to achieve personalized teaching. Based on this, this article combines the current software and hardware characteristics of the latest big data clustering algorithm technology to construct a suitable mobile learning environment design for teaching. Through specific teaching practices and later interviews and questionnaires, it is concluded that English teaching under the environment of big data clustering algorithm is accepted by teachers, students, and parents. It integrates traditional teaching and mobile teaching, making teachers' teaching more focused and making students more interested in English learning.

2. Related Work

KDDCUP99 is an international standard network security data set. In recent years, many security researchers have conducted data mining on the KDDCUP99 data set and proposed many network anomaly traffic detection algorithms [1–3]. A classical K -means clustering algorithm [4] was proposed by Bello-Orgaz et al. However, the clustering result of the K -means algorithm depends on the choice of initial values, and the search based on gradient descent often makes the algorithm fall into local optimum. Defects greatly limit the scope of their application. Timón et al. proposed a fuzzy minimum clustering algorithm executed in parallel [5]. It is called Parallel Fuzzy Minimum (PFM). The experimental results show that the linear acceleration of PFM maintains a good classification quality compared to the

corresponding version. Tu et al. pointed out that the built-in smart grid sensor allocates energy for home users and adjusts power consumption [6] omnidirectional graph route. However, the detection rate of the spectral grouping algorithm is relatively low, and it is mainly used in areas with low accuracy requirements [7]. The results help business leaders and human resource managers formulate clear business strategies. We propose a K -means clustering algorithm based on particle swarm optimization, which uses particle swarm optimization to quickly converge to the global optimal solution. This method improves the stability of the clustering results but does not significantly improve the detection rate.

There are data preprocessing methods for data mining in Ramírez-Gallego et al.'s work [8]. On this basis, Carel and Alquier studied the performance of K -means and fuzzy C -means clustering algorithms performed in Hadoop clouds and independent PCs [9]. The experimental results show that, in the cloud composed of 9 computing nodes, about 3.5 times acceleration can be achieved. The training samples of the candidate category are expanded by the category hierarchy, and the training samples of the candidate category are added according to the ancestor node, the child node, and the neighbor node. Hai et al. aimed to reduce the error propagation in the top-down prediction process [10], as well as reducing the height of the category hierarchy tree by eliminating some intermediate nodes in the category hierarchy to improve the performance of the candidate search method. Karami uses the nearest neighbor algorithm to calculate candidate categories [11]. A two-stage classification approach was applied to multilabel classification problems. The count of each partitioned set is mapped to a Hal wavelet tree by using the Haar wavelet transform [12–14]. Laplace noise was added to the nodes of the Hal wavelet tree. Finally, the histogram data to be published is obtained and queried by wavelet inverse transform [15]. The structure of the original histogram is usually reorganized and Laplace noise is added to the count value of each bucket. An instance-based candidate search method needs to be compared to all samples when categorizing. Therefore, the time complexity is relatively large, and overfitting may occur when the training samples are scarce. The main difficulty faced by category-based candidate search methods is how to construct accurate feature weight vectors for each category.

3. Proposed Method

3.1. Fuzzy K -Means Clustering. In the process of human society development, people must clearly distinguish between similar things and different things. At the same time, in mathematical statistics, people often need to have similar shapes or characteristics. Objects are grouped into one class, so there is a way for cluster analysis. This is the process of breaking an abstract or physical collection into different categories with similar attributes. A cluster-analysis-generated cluster is a set of attributes that reflect data objects. These objects are similar to the same cluster object properties and will be associated with other clusters. The object

properties are different from each other. This is an effective multivariate statistical method that guides the research object based on the intrinsic relationship of the research subject without any guidance from the teacher. It is completely classified by the similarity between things. For supervising classification, traditional clustering is to classify each object into a certain category according to its determined attribute characteristics. In this classification, the attributes of the research object are not absolute. In the mathematical method of cluster analysis, the membership degree is 1 or 0. However, in practical applications, people cannot accurately determine the exact attribution of many attributes of an object; that is, most objects have no strict attributes, and the classification problem is not clear. If the membership of 1 and 0 is forced, the general relationship between objects cannot be truly reflected. Therefore, the predecessors combined the concept of ambiguity with the ideal clustering; that is, the theory of fuzzy set is applied to clustering, and the processing object is determined by the membership function. The membership of each object in the classification is 0, and the interval of 1 determines the category. Fuzzy clustering improves the problem that only 0 and 1 memberships are too absolute in traditional clustering. The determined object can be represented by 0 or 1, and the object can be represented by a number in consecutive 0 and 1 intervals. It reflects the uncertainty of object data into categories, solves the reality of uncertain boundary problems between many categories and categories, and can reflect reality more objectively.

The so-called boundary set is clearly defined, and any element in the discussion scope (the domain) belongs to the collection or does not belong to the collection, and there is no intermediate state. Generally, set the domain X , which on A is a collection in X , and $x \in A$ or $x \notin A$ is available for any element $x \in X$. For this general set, A can be described by a binary logic:

$$U_{cj}(x_i) = \begin{cases} 1, & x_i \in c_j, \\ 0, & x_i \notin c_j, \end{cases} \quad i \in \{1, 2, 3, \dots, n\}, j \in \{1, 2, 3, \dots, m\}. \quad (1)$$

In reality, in many cases, regardless of the analysis method used by the sample, it is difficult to determine which set it belongs to. Then there is a set that is not clearly defined; namely,

$$U_{cj}(x_i) \in [0, 1]. \quad (2)$$

After determining the element set and the element vector, the computer will arbitrarily select K elements as the centroid of the first K clusters (i.e., the number of K levels or levels) and calculate the distance from all elements in the set to the K centroids, leaving the elements to be assigned to the nearest cluster. After a round of allocation, the centroid of each cluster will be recalculated. If the centroids of all clusters have not changed, the algorithm ends. Otherwise, update the centroid of each cluster and perform a new round of element assignment process. The centroid of the cluster can be solved according to the following formula:

$$\vec{x}_0 = \frac{1}{|Z|} \sum_{\vec{x}_{1,ez}} \vec{x}_1, \quad (3)$$

where Z represents a cluster and $|Z|$ represents the number of elements in the cluster. After assigning each element in the set to the nearest cluster, the center of each cluster is recalculated according to equation (3). Repeat the above steps until the new round of element assignment is the same as the previous round or the center of the two rounds is unchanged. The algorithm converges, exits the loop, and determines the level of each element.

3.2. Analysis of Constraint Parameters Data. For the evaluation of traditional language interpretation skills, the accuracy of the evaluation algorithm data classification is low. Based on this, this paper mainly designed the language interpretation language interpretation skills evaluation system. Accurately assess the assessment of the ability to perform language interpretation. First of all, it is necessary to obtain the constraint parameters, that is, to establish an information sampling model for the language interpretation skills constraint parameters. Based on the information fusion method and the time series analysis method, the statistical analysis of language interpretation skills evaluation is completed. The index parameter that restricts language interpretation skills in language interpretation is essentially a set of nonlinear time series. The evaluation of language interpretation skills is based on the high-dimensional feature distribution space. Combined with the current status quo, the parameter index distribution model is expressed in a specific form. The results of the study analyzed the main indicator parameters of the constraints. Language interpretation includes teaching level, teaching management level, and professional level. On this basis, the construction of the differential equation is the information flow model of the constraint parameters. The multivariate value function of language interpretation skills assessment is represented by h . The evaluation error measurement function is represented by ω_n , and the specific expression is as follows:

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + \omega_n. \quad (4)$$

The calculation of the language interpretation skills evaluation solution vector in the high-dimensional feature distribution space is performed by the related fusion method. Then, the feature training subset of the postcapability assessment is obtained, and the feature training subset represented by $S_i (i = 1, 2, \dots, L)$ needs to satisfy the following conditions:

$$\begin{aligned} \Sigma &= \text{diag}(\delta_1, \delta_2, \dots, \delta_r), \\ \delta_r &= \sqrt{\lambda_r}, \quad \forall i \neq j. \end{aligned} \quad (5)$$

In the evaluation of language interpretation skills, X is used to represent the conjugate solution of the statistical information model, which satisfies the initial value characteristic decomposition conditions:

$$U = \{u(t) | u(t) \in X, \|u\| \leq d, t \in I\}. \quad (6)$$

In the evaluation of language interpretation skills (involving a set of multivariate variables), $x(n)$ represents the distribution order of its statistical characteristics. Based on the statistical measurements, a data flow model for the assessment of language interpretation skills was established. The specific expression is as follows:

$$c_{1x}(\tau) = E\{x(n)x(n+\tau)\} = r(\tau). \quad (7)$$

The teaching level and the distribution level of teaching resources in the evaluation of language interpretation skills satisfy the continuous functional conditions; that is, the postability evaluation has a convergent solution, and the constraint conditions are as follows:

$$\phi_x(\omega) = \ln \varphi_x(\omega) = -\frac{1}{2} \omega^2 \sigma^2. \quad (8)$$

We can build a data flow model for language interpretation skills assessment based on the above-mentioned big data clustering algorithm and constraint parameter data analysis.

3.3. Incremental Calculation in Big Data Feature Learning. In a learning environment where large amounts of data are possible, a detailed evaluation model is created. Historical and up-to-date data must be stored in two separate databases. The incremental calculation and evaluation mode features guarantee the highest calculation speed, and the database partition structure should be as simple as possible. This method can reduce the retrieval task. The first layer of the automatic encoder maps the input layer data f to the hidden layer feature X by the encoding function h .

$$h = f(X) = s_f(W^{(1)}X + b^{(1)}). \quad (9)$$

In the calculation formula, S_f is a nonlinear activation function. The second layer maps the hidden layer feature h to the actual output z of the network by decoding the number of pictures.

$$z = g(h) = S_g(W^{(2)}h + b^{(2)}). \quad (10)$$

In a big data environment, information updates can be done automatically. In the network environment, the parameter resources required for the calculation are obtained and integrated into the structural hierarchy of the evaluation model using the big data operation method. The structure update in big data processing is shown in Figure 1.

The structure update needs to automatically retrieve the update patches under the network platform to repair the loopholes in the original structure. The structural update in the incremental depth calculation evaluation mode is based on environmental threat analysis. Assess the risks in the operating environment where the application software is located. Compare the risk type with the internal structure to determine whether the current structure can meet the information update requirements in the big data environment. Once there are potential risks in the structure, it will be used

as the basis for the next patch retrieval and update, and the structure will be constructed through risk control. When the same system structure is used multiple times, the speed of data analysis and calculation may be reduced. In order to prevent the redundant structure from affecting the system speed, the redundant part is deleted and processed in the automatic update search to ensure completion. The software structure after automatic update is optimized and reasonable. Automatic parameter update can be understood as a learning ability. In order to improve the efficiency of depth calculation and evaluation, feature extraction methods are used to update the parameters in model construction. Within the system structure, the characteristics of the required information are automatically filtered out and compared with the data information obtained in the network environment, and the characteristics are automatically updated after the characteristics are consistent.

4. Experiments

4.1. Evaluation Indicators. The K -means clustering algorithm is an unsupervised learning algorithm that does not artificially mark which clusters each sample point belongs to in the data set used. So how do you judge the quality of each K -means clustering result? In the case where cluster analysis is possible, after the formal representation of the case, the samples that should belong to the same cluster should be relied upon. More closely, the examples for different clusters should be larger. Therefore, using the sum of the distances of each element to each cluster center to represent the metric of each clustering result, we use the sum of squared errors (SSE) metric.

$$SSE = \sum_{i=1}^k \sum_{\vec{x}_n \in c_i} \text{dist}(\vec{x}_{0i}, \vec{x}_n)^2, \quad (11)$$

where k is the total number of clusters. It is at the center of the number of levels. The K -means algorithm needs to give the cluster number k in advance, but, in most cases, the value of k cannot be determined. Usually you can only choose a better k value based on past experience. For K -means clustering, the ultimate goal is to find the sum of the least squares errors so that the model can be better collected. Suppose that, for a data set with n elements, when the number of clusters $k=1$, the SSE of the model reaches the maximum at this time. As the value of k gradually increases, the SSE gradually decreases. When $k=n$, SSE reaches the minimum value. That is, each element in the data set forms a separate cluster with $SSE=0$. In this paper, we set the number of teaching experiments, as well as different English teaching time and English teaching content. The detailed teaching content is shown in Table 1.

4.2. Attribute Analysis of the Evaluation System. The language interpretation skills mobile information system is a complex and huge system involving many factors and attributes. There are various complicated causal relationships among various factors. With the development of system science, management

science, and scientific decision-making theory, a series of system analysis and decision-making theories and methods have emerged. Correspondingly, various decisions and assessments have gradually entered the stage of scientific decision-making and intelligent decision-making based on computer tools based on previous experience decision-making. Therefore, the evaluation of comprehensive language interpretation skills must also use scientific decision-making theories and the most advanced decision-making aids to make the language interpretation skills mobile information system on a scientific and modern road. The function of the language interpretation skills mobile information system is to compare various internal and external reasons for English learning, regardless of the school grade, and then compare the language interpretation situations of several schools. Comparing the overall teaching quality of colleges and universities of different grades, the same grade can more accurately reflect the overall level of the school, which is conducive to reducing school costs and improving teaching standards. Through analysis, the attribute characteristics of interschool language interpretation quality evaluation are extracted. Attributes are internal and external, and when other attributes are ignored, they cannot be considered as unilaterally important. The understanding of things should be multifaceted, specifically the learner's personal factors, environmental factors, and the price paid.

5. Discussion

To test the accuracy of the language interpretation skills mobile information system and the performance of related data analysis, this paper adopts Matlab simulation analysis method to test, adopts statistical analysis method to sample the language interpretation skills evaluation data, and determines the decision threshold. Matlab is a commercial mathematics software produced by MathWorks in the United States. It is used in data analysis, wireless communication, deep learning, image processing and computer vision, signal processing, quantitative finance and risk management, robotics, control systems, and other fields. The capacity assessment is based on $Dx=2$. The sampling frequency of the parameters related to teaching resources and capacity distribution is 500 Hz, the adaptive initial step size is 0.96, and the correlation coefficient of the feature distribution is 1.15. The data reconstruction of the language interpretation postability assessment constraint parameters was obtained, and the time domain waveform map of the big data distribution was obtained as shown in Figure 2.

We conducted a survey on a total of 100 English teachers with different teaching levels at different levels of schools in a second-tier city and performed data clustering and information fusion processing to realize the teaching postability evaluation. The test results of the evaluation indicators are shown in Table 2.

Using spectral clustering, K -means, PSO-based K -means (K-PSO) algorithm, and global K -means (GK) algorithm for comparative experiments 5 times, the average final detection rate and false positive rate were obtained. The detection rate is shown in Figure 3, and the average running time is shown in Figure 4.

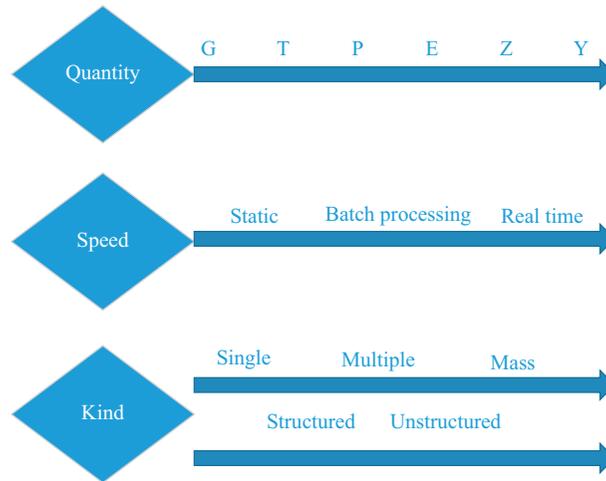


FIGURE 1: Big data processing information structure update mode.

TABLE 1: Setting test data.

Number of experiments	Teaching duration (min)	Teaching content
20	15	English grammar
20	35	Sentence pattern analysis
20	40	Oral practice
20	50	English conversation
20	50	English writing

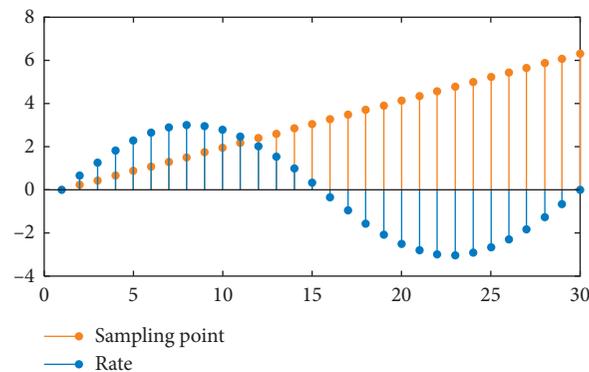


FIGURE 2: The time domain waveform map of the big data distribution.

It can be seen from Figures 3 and 4 that the proposed algorithm has higher detection rate and lower false positive rate. Globally, the running time of the algorithm is more obvious than that of the global K -means algorithm. From the perspective of algorithm operation, due to the different models and methods, there are large errors in the operation results. The error value is determined by the substitution of the measured value and the true value to ensure the validity of the experimental data and eliminate the measurement ratio used by the random error. The ratio varies depending on the number of experiments, but the range of the selection process is between 2.45 and 6.55. Eliminating nonzero errors is the use of extreme value measurements, using time interception methods to set the zero value, effectively avoiding nonzero errors.

It can be seen from Figure 5 that the method of this paper is obviously close to the theoretical value. It can be said that the design method of this paper has the comprehensiveness of evaluation. In order to evaluate the traditional language interpretation skills, the data information of the evaluation algorithm is classified. An evaluation method is proposed. Based on the establishment, the ability to evaluate the big data information model using quantitative recursive analysis method is extracted. The entropy feature of the constraint feature information is combined with the K -means clustering algorithm to cluster and integrate the index parameters of the language interpretation skills. On this basis, a teaching resource allocation plan was developed to achieve a scientific evaluation of language interpretation skills.

TABLE 2: Performance test comparison.

Evaluation cycle	Method of this paper		Method of this paper	
	Evaluation accuracy (%)	Utilization (%)	Evaluation accuracy (%)	Utilization (%)
1	94.61	94.01	85.53	85.12
2	96.23	95.43	87.46	85.43
3	95.83	98.43	86.57	85.46
4	97.49	96.89	86.45	86.74

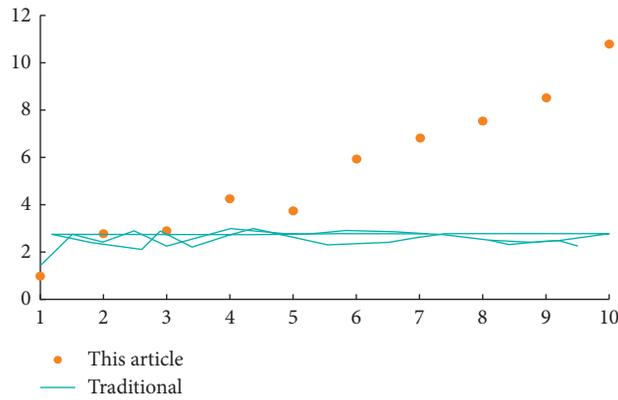


FIGURE 3: Detection rate (%).

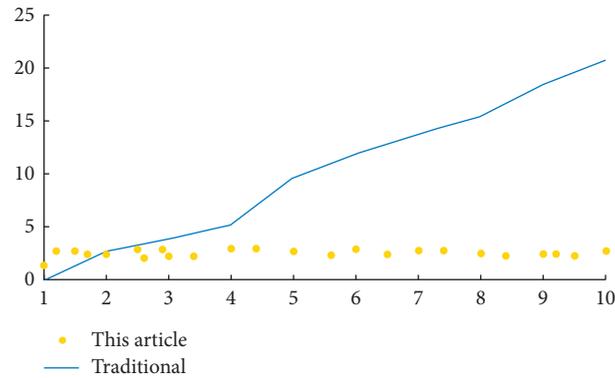


FIGURE 4: Average running time (ms).

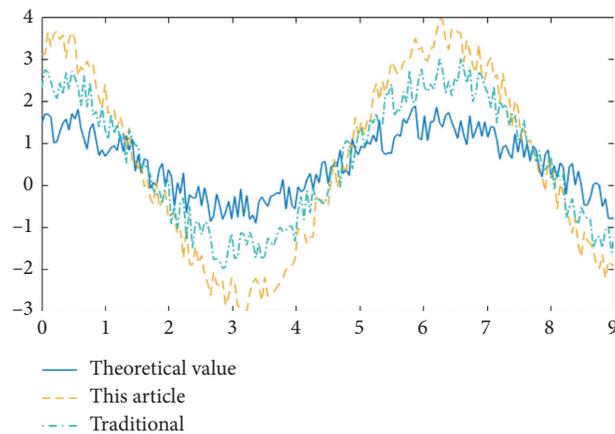


FIGURE 5: Frequency comparison test results.

6. Conclusions

This article focuses on the basic features of big data. Aiming at the heterogeneity of big data, this paper studies the language interpretation skills evaluation algorithm based on big data fuzzy K -means clustering to realize the method of extending language interpretation skills evaluation model from vector space to tensor space. An evaluation method of language interpretation skills is proposed. The main conclusions of the constraint parameter index analysis model for constructing and analyzing language interpretation skills are as follows:

- (1) Fuzzy K -means clustering analysis is also a class of formal logic classification, but the classification information provided is more abundant than traditional clustering and discriminant analysis. This method can be clustered and distinguished. When an unknown sample is returned, a distinction can be made, which is a good pattern recognition method.
- (2) Fuzzy K -means clustering applies the concept of fuzzy to level division. The classification method uses a fuzzy method to measure the relationship between the two factors of each assessment to determine the similarity of the elements and then divide the highly similar assessments into one level and pass the square of the error and the SSE assessment science. The smaller the level, the smaller the value of SSE and the smaller the relationship of all evaluations in each level, and the greater the similarity, the more accurate the classification. The traditional land classification method is to determine the similarity only by the size of the weighted total score.
- (3) The experimental results show that the evaluation algorithm significantly improves the ability and accuracy of information fusion analysis in the process of the language interpretation skills mobile information system.

The mobile English teaching ability information system designed in this article has brought about convenience to the majority of English teachers and students, allowing students to learn English anytime and anywhere and also allowing them to fall in love with English and enjoy the fun of mobile English teaching. But this platform still has many flaws and vulnerabilities that need to be improved.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares no conflicts of interest.

Acknowledgments

This work was supported by Youth Project of Education Department of Liaoning Province "Research on the Online and Offline Blended English Teaching at Art Colleges" (no. WQ201908).

References

- [1] G. E. Guoliang, "Research of English teaching capability evaluation algorithm based on fuzzy K -means clustering of big data," *Modern Electronics Technique*, vol. 2017, no. 20, p. 10, 2017.
- [2] F. Xhafa, A. Bogza, S. Caballé, and L. Barolli, "Apache mahout's k -means vs fuzzy k -means performance evaluation," in *Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative Systems*, pp. 110–116, Ostrawa, Czech Republic, September 2016.
- [3] C. Ioana and H. Daniel, "Termination in a π -calculus with subtyping," *Mathematical Structures in Computer Science*, vol. 26, no. 8, pp. 1395–1432, 2016.
- [4] G. Bello-Organ, J. J. Jung, and D. Camacho, "Social big data: recent achievements and new challenges," *Information Fusion*, vol. 28, pp. 45–59, 2016.
- [5] I. Timón, J. Soto, H. Pérez-Sánchez, and J. M. Cecilia, "Parallel implementation of fuzzy minimal clustering algorithm," *Expert Systems with Applications*, vol. 48, pp. 35–41, 2016.
- [6] C. Tu, X. He, Z. Shuai, and F. Jiang, "Big data issues in smart grid - a review," *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 1099–1107, 2017.
- [7] A. De Mauro, M. Greco, M. Grimaldi, and P. Ritala, "Human resources for big data professions: a systematic classification of job roles and required skill sets," *Information Processing & Management*, vol. 54, no. 5, pp. 807–817, 2018.
- [8] S. Ramírez-Gallego, A. Fernández, S. García, M. Chen, and F. Herrera, "Big data: tutorial and guidelines on information and process fusion for analytics algorithms with MapReduce," *Information Fusion*, vol. 42, pp. 51–61, 2018.
- [9] L. Carel and P. Alquier, "Simultaneous dimension reduction and clustering via the NMF-EM algorithm," *Advances in Data Analysis and Classification*, vol. 15, no. 1, pp. 231–260, 2021.
- [10] M. Hai, Y. Zhang, and H. Li, "A performance comparison of big data processing platform based on parallel clustering algorithms," *Procedia Computer Science*, vol. 139, pp. 127–135, 2018.
- [11] A. Karami, "Application of fuzzy clustering for text data dimensionality reduction," 2019, <https://arxiv.org/abs/1909.10881>.
- [12] J. Heil, V. Häring, B. Marschner, and B. Stumpe, "Advantages of fuzzy k -means over k -means clustering in the classification of diffuse reflectance soil spectra: a case study with West African soils," *Geoderma*, vol. 337, pp. 11–21, 2019.
- [13] K. Ahlawat, A. Chug, and A. P. Singh, "Empirical evaluation of map reduce based hybrid approach for problem of imbalanced classification in big data," *International Journal of Grid and High Performance Computing*, vol. 11, no. 3, pp. 23–45, 2019.
- [14] S. Hussain, Z. F. Muhsin, Y. K. Salal, P. Theodorou, F. Kurtoglu, and G. C. Hazarika, "prediction model on student performance based on internal assessment using deep learning," *International Journal of Emerging Technologies in Learning*, vol. 14, no. 8, pp. 4–22, 2019.
- [15] B. Liu, S. He, D. He, Y. Zhang, and M. Guizani, "A spark-based parallel fuzzy c -means segmentation algorithm for agricultural image big data," *IEEE Access*, vol. 7, pp. 42169–42180, 2019.