

Retraction

Retracted: Construction of English Intelligent Translation Software Framework Based on Data Analysis Algorithm

Mobile Information Systems

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

Copyright © 2023 Mobile Information Systems. 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.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] Z. Zhou, "Construction of English Intelligent Translation Software Framework Based on Data Analysis Algorithm," *Mobile Information Systems*, vol. 2022, Article ID 7773157, 7 pages, 2022.

Research Article

Construction of English Intelligent Translation Software Framework Based on Data Analysis Algorithm

Zhihui Zhou 

School of Foreign Language, Zhanjiang University of Science and Technology, Zhanjiang 524094, Guangdong, China

Correspondence should be addressed to Zhihui Zhou; zhouzhihui@zjkju.edu.cn

Received 13 May 2022; Revised 6 July 2022; Accepted 5 August 2022; Published 31 August 2022

Academic Editor: Mian Ahmad Jan

Copyright © 2022 Zhihui Zhou. 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.

In the era of “Internet + education,” the information technology and learning methods of college students have become inseparable. The rapid development of intelligent translation software can provide convenience for English-Chinese translation and computer English learning and simultaneously improve the quality and efficiency of professional English learning. Under such background, the English teaching of software majors is chosen as the breakthrough to analyze the measures of applying intelligent translation software during the teaching process. Therefore, a data mining algorithm including cluster analysis and BP neural network model is designed. Then, the cluster analysis algorithm is used to classify the data in different forms, which can improve the data utilization efficiency. In addition, K-means algorithm based on feature selection is improved to achieve better performance. In the comparison of translation speed and matching rate, our method is much better than other software.

1. Introduction

In the English learning of college students, translation tools play an important role, which may not be accurate enough in terms of grammar and sentence patterns. In this regard, the translation software should be updated and improved so that it can be better used in college students’ English learning. The rapid development of intelligent translation software under the background of Internet information technology can help solve the problem of language barriers for translators. Software majors use intelligent translation software for English-Chinese translation and computer English learning and for improving the quality and efficiency of professional English learning.

The essence of translation software is machine translation, which is included in the study of machine translation. Machine translation refers to the process of converting one language into another target language through computer-related processing, generally referring to the translation of sentences and full texts between different languages [1].

In the era of Internet + education, translation software, as a kind of machine translation, is quietly changing the way and behavioral habits of college students’ English learning

with the help of information technology. College English teaching has important theoretical and practical significance. Using “translation software” and “college English learning” as keywords to search on CNKI, it is found that the research on translation software in our country at this stage mainly focuses on the translation software itself and the development and design of translation software. In general, the two types of research have different focuses. Some research studies focus on the research on translation software itself, such as the current situation and future development trends of translation software. Human translation and machine translation are compared, focusing on the differences between them; some research studies focus on how to develop and design translation software through different technologies and different angles to enrich translation software in practice. It can be used in different aspects of life [2–5], so as to achieve the ultimate goal of improving the effective utilization rate of translation software. Although the focus of the two types of research is different, they both hold a positive attitude towards the future of translation software, hoping that the future development of translation software will be better and better through continuous improvement and perfection.

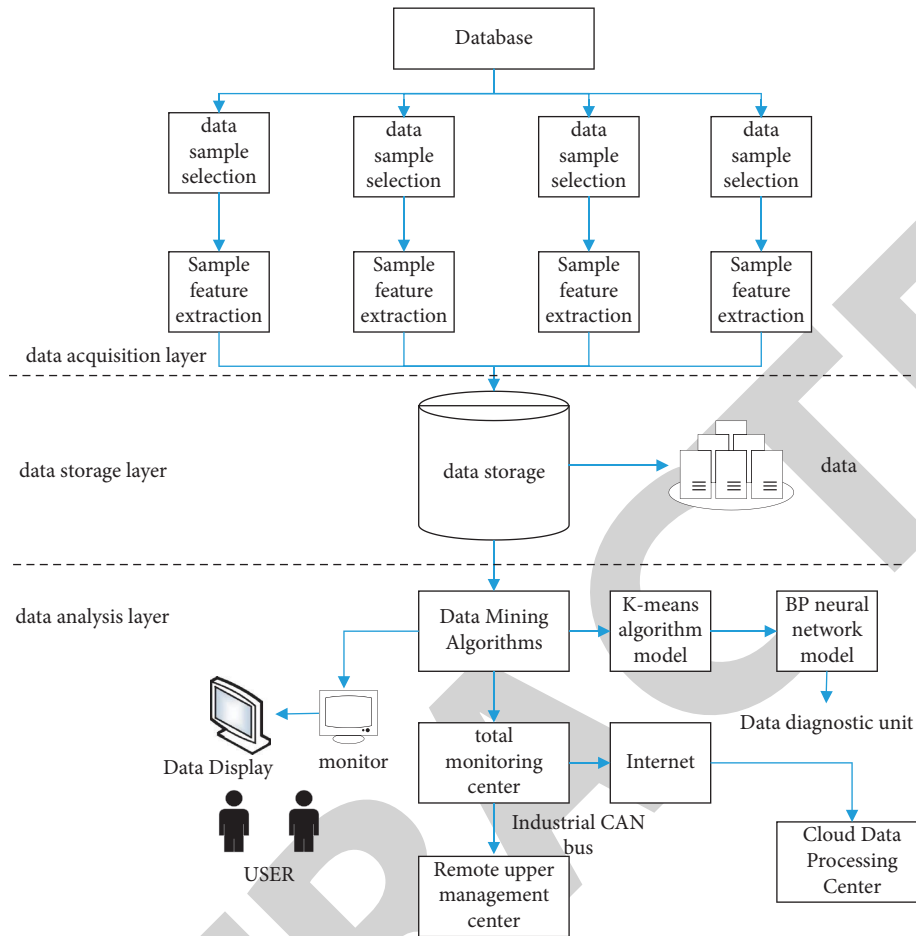


FIGURE 1: Architecture design of big data analysis.

Therefore, it can be seen that under the background of today's Internet+era, in order to make information technology better serve our study and life, further research on translation software is imperative. In addition to the above two types of research, a few scholars have studied and explored the application of translation software and put forward some reference suggestions for reference. Ali et al. [6] used observation and investigation methods to take the undergraduate course of "Technical Translation" of Shanghai Second Polytechnic University as a case study, discussed the application of translation software "Technical Translation" in classroom teaching, and analyzed the problems faced in the practical application of translation software. Reference [3] put forward methods to try to solve problems such as school enterprise cooperation and teacher training. Lei and Shao [4] combined translation software with editing work and discussed how to look at translation software from the perspective of editing, especially the application of translation software.

Through the analysis of the above-mentioned literature, the actual use of translation software, the behavioral characteristics, and tendencies are explored as the research starting point. The corresponding recommendations are rarely studied. Therefore, in this work, the

English intelligent translation software based on data analysis algorithm is designed to provide inspiration and suggestions for college English education [7].

2. Designing Big Data Analysis Architecture

The design is shown in Figure 1.

The data acquisition layer collects a large amount of heterogeneous data from the Internet and converts the collected data into well structured data [8–10].

3. Data Mining Algorithms

An analysis process becomes a class of similar objects (see Figure 2).

3.1. Cluster Analysis Algorithm. By clustering the k sample data points in the space as the center, the information of big data which is closest to different samples is finally classified [11–14].

The cluster analysis algorithm flow is shown in Figure 3.

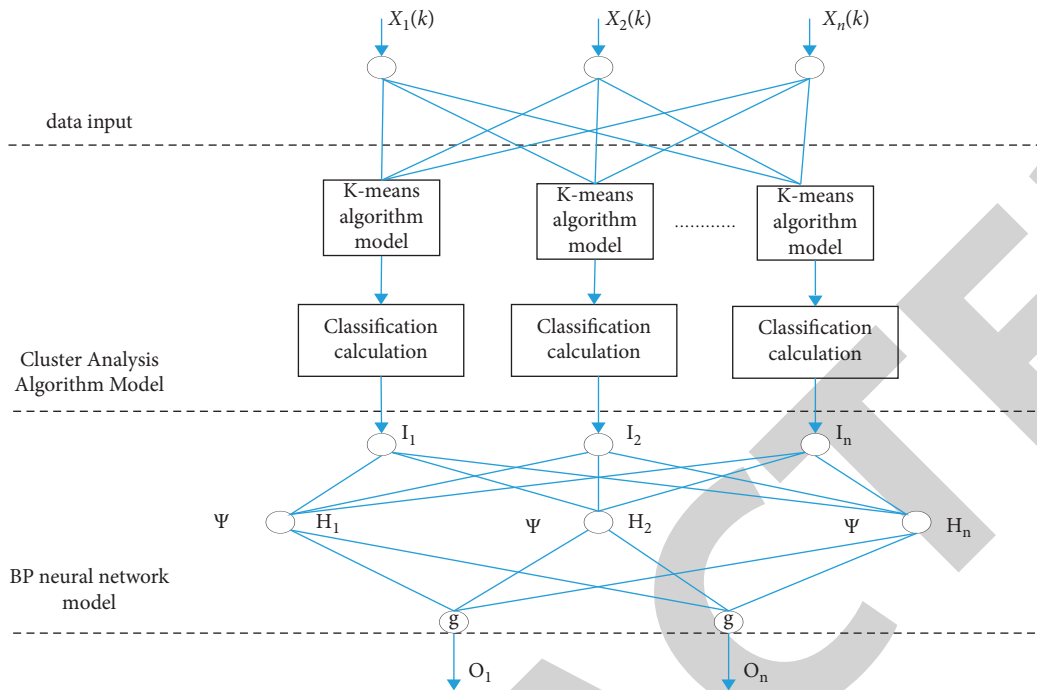


FIGURE 2: Data mining algorithm flow.

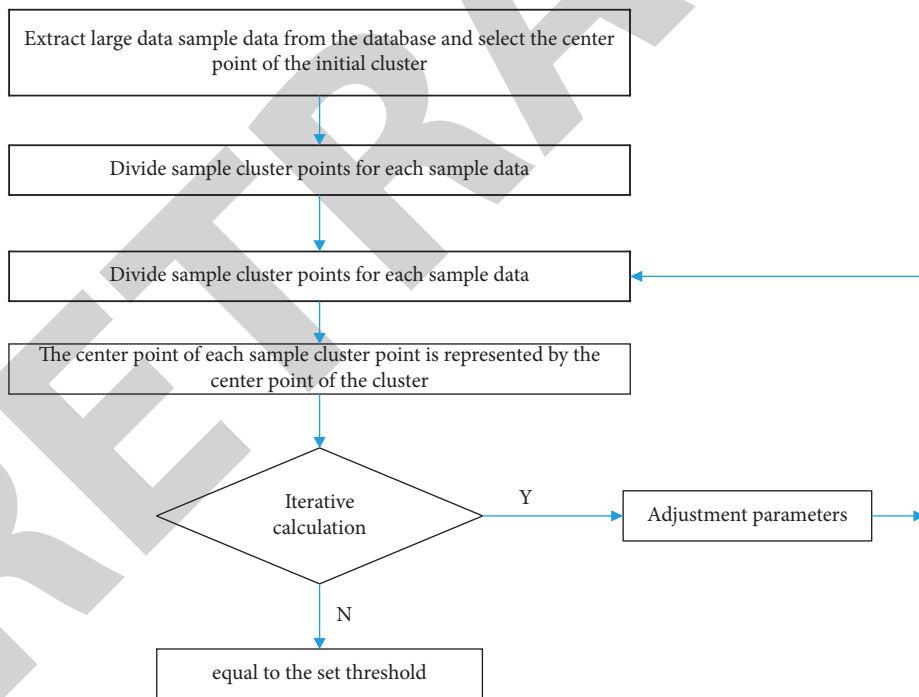


FIGURE 3: Cluster analysis algorithm flow.

4. K-Means Feature Selection Algorithm

According to the research and status analysis of data mining and clustering algorithm in the paper, the author has consulted a large number of studies and conducted in-depth research on K-means algorithm. Based on the research and use of the K-means algorithm by researchers in recent years, this paper improves the algorithm, and finally, an algorithm based on feature selection is used [15]. The algorithm first

sets the corresponding feature attributes for college students, then filters and cleans irrelevant feature attributes, normalizes the selection of corresponding feature attributes, and then assigns initial values to the cluster centers and updates them continuously.

4.1. Feature Selection Methods. When the collected data object set has too many feature attributes, some invalid or

repeated feature attributes will be mixed, which will increase the complexity of cluster analysis, reduce the performance of the algorithm, and even affect the accuracy of the calculation results. To solve this problem, filtering and cleaning invalid feature attributes is necessary. In the paper, we judge the contribution of a feature attribute to the current cluster by calculating the value of the weight vector of the feature attribute. At the same time, we measure the contribution of a feature to clustering by examining and verifying the difference of feature attributes between data objects of the same category and data objects of different categories [16]. If the feature attributes are obviously different from each other between objects of different categories but are not clearly distinguished from each other between objects of the same data category, that is, the weight of the feature attribute is larger, then the feature attribute has a high degree of contribution to the clustering, with strong feature discrimination ability. When the algorithm starts to calculate, it first randomly sets a data object S_i as the centroid, then divides the data set into categories, and then selects d data objects with a distance S_i from each category of data objects. d data objects of the same category as S_i constitute a new dataset $T(c)$, and objects of different categories from S_i constitute a new set G according to their category (c is updated according to sets $T(c)$ and $G(c)$). The weight vector $W = \{w_1, w_2, \dots, w_q\}$ of the feature attribute; then, the calculation formula of the feature attribute weight is shown in the following formula [17]:

$$W_t^{i+1} = W_t^i - \sum_{x \in T(c)} \frac{\text{diff}(t, S_i, x)}{n * d} + \frac{\sum_{x \in S(S_i)} [q(c)/1 - q(S(S_i)) \sum_{x \in G(c)} \text{diff}(t, S_i, x)]}{(n * d)}. \quad (1)$$

Among them, n is the number of times to extract sample data, and the $\text{diff}(t, S_i, x)$ function represents the difference function of the feature attribute of the data object on t . The calculation method is as follows:

$$\text{diff}(t, S_i, x) = \left| \frac{S_{it} - S_{jt}}{\max_t - \min_t} \right|. \quad (2)$$

This method will balance the d data objects with similar distances to S_i by using the maximum and minimum values in the feature attributes and then multiply the proportion of other data objects in the dataset in all the data objects of different categories from S_i , so the difference between the data object S_i and the category is obtained, so as to evaluate the contribution of the feature attribute to the data object.

4.2. Optimal Selection of Initial Cluster Centers. The Euclidean distance between any two data objects x_i and x_j ($1 < i < j < n$) is defined as follows [6]:

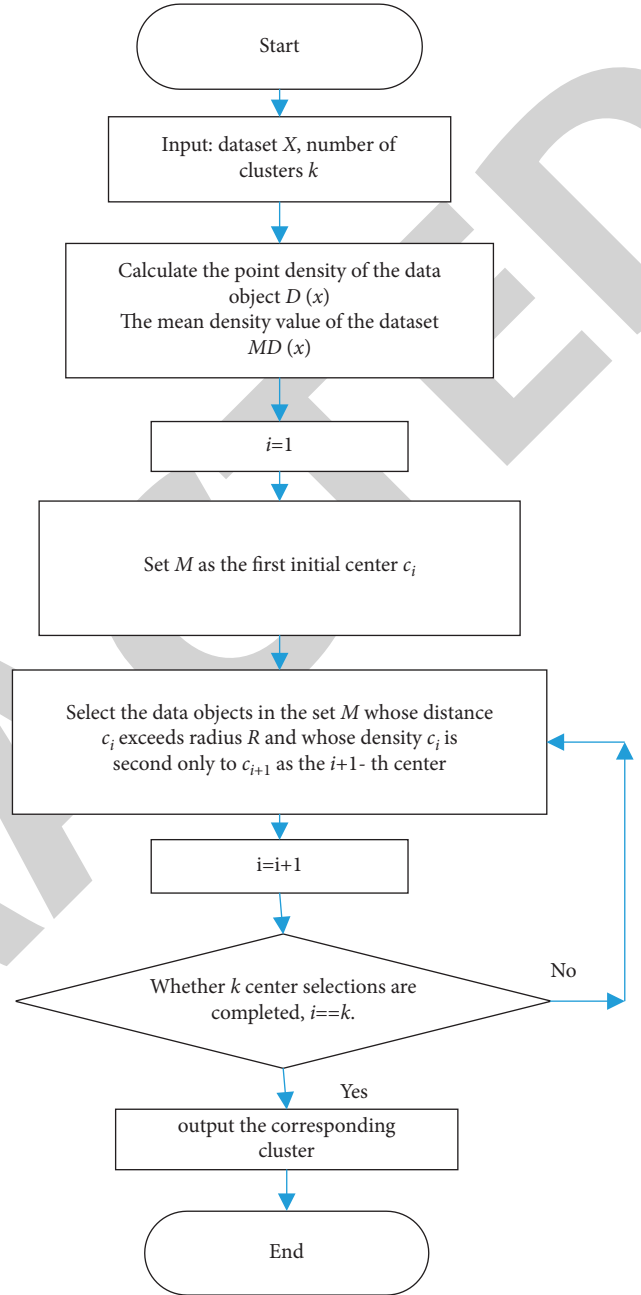


FIGURE 4: Flowchart of initial cluster center selection.

$$d(x_i, x_j) = \sqrt{\sum_{a=1}^p (x_{ia} - x_{ja})^2}, \quad (3)$$

$$\text{density}(x_i) = \frac{\sum_{j=1}^n d(x_i, x_j)}{\sum_{i=1}^n d(x_i, x_j)}. \quad (4)$$

Assume that the neighborhood radius R_i of the data object x_i ($1 \leq i \leq n$) in the dataset can be defined as shown in formula (5): where cR ($0 < cR < 1$) is used as the adjustment coefficient of the neighborhood radius, according to past

TABLE 1: Test parameters.

Parameter name	Numerical size
Internet multilingual words translation volume/characters	200
Translation rate (Kb/s)	10
Internet multilingual words translation volume/characters	400
Translation rate (Kb/s)	10

TABLE 2: Experimental data settings.

Test times/every time	Semantic context translation translation types	Revision parameters
1	3	$0.2 * 10^{-3}$
2	2	$0.1 * 10^{-3}$
3	3	$0.2 * 10^{-3}$
4	3	$0.1 * 10^{-3}$
5	3	$0.2 * 10^{-3}$
6	2	$0.1 * 10^{-3}$

TABLE 3: Comparison results of network multilingual translation sentences speed.

Number of sentences	Reference [4]	Reference [7]	Ours
1 000	2.83	5.24	6.93
2 000	3.04	5.51	7.04
3 000	3.24	5.58	7.68
4 000	3.68	5.65	7.65
5 000	4.18	5.72	8.13
6 000	4.46	5.95	8.52
7 000	4.98	2.24	8.75
8 000	5.03	6.42	8.94
9 000	5.42	6.61	9.23
10 000	5.89	6.74	10.53

experience. It shows that the clustering effect will be better when the value of cR is 0.13 [15].

$$R_i = n^{cR} \times \frac{1}{n} \sum_{i=1}^n e^{-\text{density}(x_i)}. \quad (5)$$

The larger the value of $D(X_i)$, the higher the density of points in the spherical area where the data object is located, as shown in the following formula:

$$D(x_i) = |\{p \mid d(x_i, p) \leq R_i, p \in X\}|. \quad (6)$$

Let $MD(x)$ be the average density of data objects in dataset X , as shown in the following equation:

$$MD(x) = \frac{1}{n} \sum_{x \in X} D(x). \quad (7)$$

The specific steps for selecting and optimizing the initial cluster center are as follows.

Input: the initial dataset X and the specific number of clusters k .

- (1) Calculate the distance between any two data objects in X using formula (5) and then form a distance matrix by these distance values.

- (2) According to the distance density function (6), the neighborhood radius calculation formula (7), and the point density calculation formula (5), obtain the point density $D(X_i)$ corresponding to each data object.

- (3) According to the mean density calculation formula (7), the mean density of the dataset S is calculated as $MD(x)$.

- (4) Calculate the point density of each data object obtained and then compare, and divide all data objects not less than the average density into a set M [14].

- (5) Arrange the data objects in the set M in descending order according to the point density.

- (6) Select a data object whose point density is only less than the data object obtained in step (6).

According to the above process, the corresponding flowchart of the algorithm is shown in Figure 4.

5. Test Analysis

5.1. Setting Test Parameters. The network multilingual timely translation systems in [4, 7] were introduced, and the test parameters were set, as shown in Table 1.

The translation system test experiment needs to pay attention to the randomness of the test object selection. In order to ensure the accuracy of the whole experiment process, it is necessary to strictly limit the conditions of the experimental object (see Table 2).

5.2. Internet Multilingual Translation Speed Test. Taking the number of online multilingual sentences as an independent variable, three translation systems are used to test the speed of online multilingual translation. It can be seen that the network multilingual timely translation system in [4] does not count the network multilingual data in the database in terms of hardware design, so the training data cannot be obtained. The average translation speed during the network multilingual test is 4.275 sentences per second; the

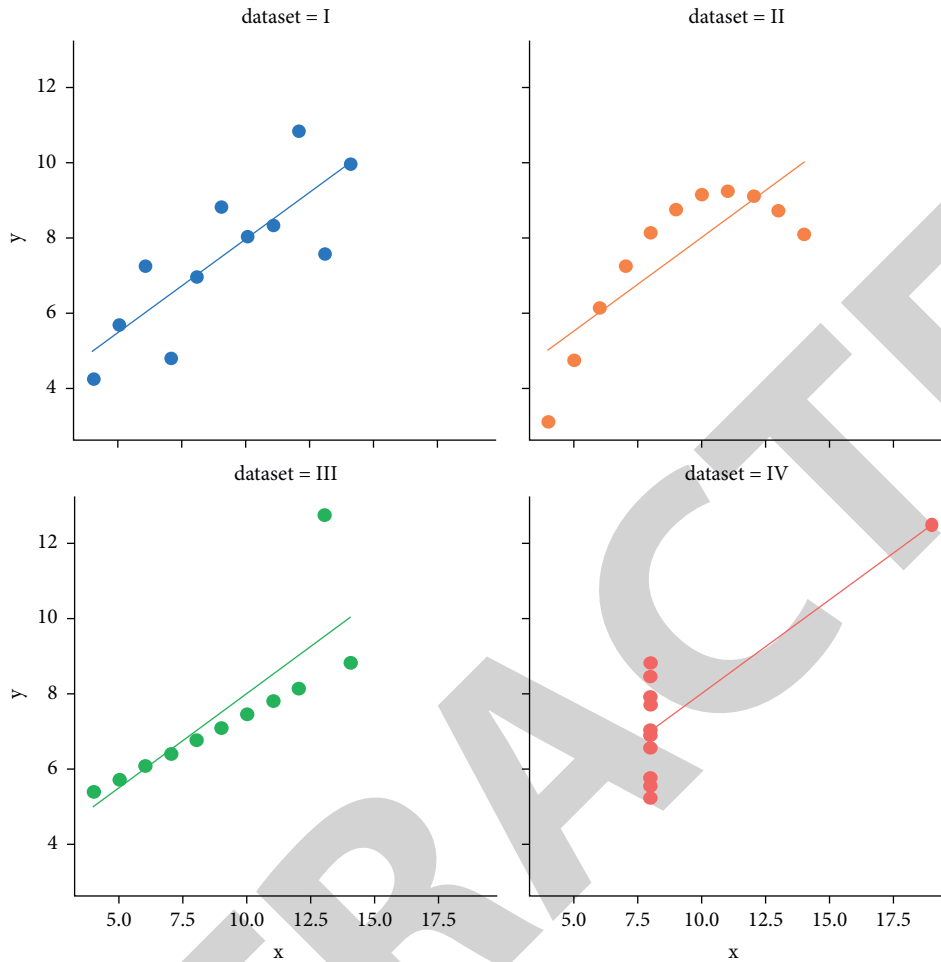


FIGURE 5: Test results of network multilingual matching rate.

performance of the network multilingual timely translation system in [7] is relatively better than that of the network multilingual translation system in [4], but due to the inability to extract the semantic features of network multilanguage, the translation of network multilanguage becomes more complicated. After calculation, the average translation speed in the process of network multilanguage test is 5.566 sentences per second; based on data, the network multilanguage timely translation system of the English intelligent translation software based on the analysis algorithm combines the software and hardware advantages of the above two systems to speed up the translation speed of network multilanguage. After calculation, the average translation speed in the network multilanguage test process is 1 second (see Table 3).

5.3. Network Multilingual Matching Rate Test. The online multilingual timely translation systems in [4, 7] and the English intelligent translation system based on data analysis algorithms are used, respectively. The software's network multilanguage timely translation system tests the matching rate of network multilanguage (see Figure 5).

From the test results in Figure 5, that there are many English intelligent translation software based on the data analysis algorithm, and the matching rate of the language real-time translation system is the highest.

6. Conclusion

A data mining method is proposed to achieve the classification of data. In such method, clustering analysis algorithm is used to increase the recognition accuracy of various types of data. Then, the efficiency and accuracy of data are further improved by using BP network. Furthermore, such method enables users to precisely improve the accuracy of the data from the clustered big data. In addition, the Apriori algorithm can simplify the connection operation part by comparing and deleting unnecessary connection operations. Because of this, the improved Apriori algorithm saves the cost and reduces the interface load of the data mining platform. In actual use, the optimized algorithm can save a lot of computing time, and the performance is much better than the traditional Apriori algorithm. The application of data mining algorithm in the analysis of college students' English learning translation software offers great help to future education. However, the negative effects of security

information leakage and student behavior not being trusted should also be a cause for concern. Therefore, educators should carefully monitor big data algorithms in the process of promoting data mining technology. They cannot blindly trust the evaluation results from intelligent algorithms and ignore the feelings of students. Teachers can use data mining results as auxiliary plans. Scientific evaluation and prevention of negative factors cannot be overlooked in data mining technology.

Data Availability

No data were used to support this study.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding the publication of this article.

References

- [1] F. B. Saghezchi, A. Radwan, J. Rodriguez, and T. Dagiuklas, "Coalition formation game toward green mobile terminals in heterogeneous wireless networks," *IEEE Wireless Communications*, vol. 20, no. 5, pp. 85–91, 2013.
- [2] B. M. Ester, H. P. Kriegel, J. Sander, and X. Xu, "A Density Based algorithm for discovering clusters in large spatial databases," in *Proceedings of the International Conference on Knowledge Discovery and Data Mining*, Portland Oregon, August 2013.
- [3] F. M. Abd Algalil and S. P. Zambare, "Effects of temperature on the development of Calliphorid fly of forensic importance *Chrysomya megacephala* (Fabricius, 1794)," *Indian Journal of Applied Research*, vol. 5, no. 2, pp. 767–769, 2015.
- [4] H. Lei and W. Shao, "E-commerce English translation in picking robot multilingual system the application of the technology: based on ESP theory," *Agricultural Mechanization Research*, vol. 42, no. 2, pp. 234–237, 2020.
- [5] A. M. Al-Azab, A. A. Zaituon, K. M. Al-Ghamdi, and F. M. A. Al-Galil, "Surveillance of dengue fever vector *Aedes aegypti* in different areas in Jeddah city Saudi Arabia," *Advances in Animal and Veterinary Sciences*, vol. 10, no. 2, pp. 348–353, 2022.
- [6] R. Ali, S. Lee, and T. C. Chung, "Accurate multi-criteria decision making methodology for recommending machine learning algorithm," *Expert Systems with Applications*, vol. 71, pp. 257–278, 2017.
- [7] A. Radwan, M. F. . Domingues, and J. Rodriguez, "Mobile caching-enabled small-cells for delay-tolerant e-Health apps," in *Proceedings of the 2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 103–108, IEEE, Paris, France, May 2017.
- [8] Yu Qian, "Design of interactive English-Chinese translation system based on feature extraction algorithm," *Modern Electronic Technology*, vol. 41, no. 4, pp. 169–171, 2018.
- [9] T. Ren, H. Wang, H. Feng, C. Xu, G. Liu, and P. Ding, "Study on the improved fuzzy clustering algorithm and its application in brain image segmentation," *Applied Soft Computing*, vol. 81, Article ID 105503, 2019.
- [10] S. Palanisamy, B. Thangaraju, O. I. Khalaf, Y. Alotaibi, S. Alghamdi, and F. Alassery, "A novel approach of design and analysis of a hexagonal fractal antenna array (HFAA) for next-generation wireless communication," *Energies*, vol. 14, no. 19, p. 6204, 2021.
- [11] P. Juric, M. Matetic, and M. Brkic, "Data mining of computer game assisted e/m-learning systems in higher education," in *Proceedings of the Formation and Communication Technology, Electronics and Micro Electronics (MIPRO), 2014 37th International Convention on*, pp. 750–755, Opatija, Croatia, May 2014.
- [12] Y. Wang, Li Xiao, Y. Yang, and A. Anwar, "Uyghur-Chinese machine translation system based on paraphrase information Fusion Research," *Computer Engineering*, vol. 45, no. 4, pp. 288–295, 2019.
- [13] A. Fan, S. Bhosale, H. Schwenk et al., "Beyond English-centric multilingual machine translation," *Journal of Machine Learning Research*, vol. 22, no. 107, pp. 1–48, 2021.
- [14] S. Gu and X. Li, "Optimization of computer-aided English translation teaching based on network teaching platform," *Computer-Aided Design and Applications*, vol. 19, no. S1, pp. 151–160, 2021.
- [15] Q. Liu, C. Liu, and Y. Wang, "etc. Integrating external dictionary knowledge in conference scenarios the field of personalized machine translation method," *Journal of Chinese Informatics*, vol. 33, no. 10, pp. 31–37, 2019.
- [16] P. An, Z. Wang, and C. Zhang, "Ensemble unsupervised autoencoders and Gaussian mixture model for cyberattack detection," *Information Processing & Management*, vol. 59, no. 2, Article ID 102844, 2022.
- [17] R. Ali, M. H. Siddiqi, and S. Lee, "Rough set-based approaches for discretization: a compact review," *Artificial Intelligence Review*, vol. 44, no. 2, pp. 235–263, 2015.