

## *Retraction*

# **Retracted: Application of English Vocabulary Presentation Based on Clustering in College English Teaching**

### **Security and Communication Networks**

Received 13 September 2023; Accepted 13 September 2023; Published 14 September 2023

Copyright © 2023 Security and Communication Networks. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] Q. Yang, "Application of English Vocabulary Presentation Based on Clustering in College English Teaching," *Security and Communication Networks*, vol. 2021, Article ID 9220951, 8 pages, 2021.

## Research Article

# Application of English Vocabulary Presentation Based on Clustering in College English Teaching

Qiujuan Yang 

Weinan Normal University, Weinan 714099, Shaanxi, China

Correspondence should be addressed to Qiujuan Yang; [yangqiujuan@wnu.edu.cn](mailto:yangqiujuan@wnu.edu.cn)

Received 3 November 2021; Accepted 13 December 2021; Published 28 December 2021

Academic Editor: Jian Su

Copyright © 2021 Qiujuan Yang. 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.

As the most basic element in English learning, vocabulary has always been the focus of teaching in college English classes, but the teaching effect is often unsatisfactory. In this paper, the genetic algorithm fitness function design part is integrated with the K-medoids algorithm to form K-GA-medoids, and secondly, it is combined with KNN to form an algorithmic framework for English vocabulary classification. In the classification process, clustering and classification steps are taken to realize the reduction of the training set samples and thus reduce the computational overhead. The experiments show that K-GA-medoids have significantly improved the clustering effect compared with traditional K-medoids, and the combination of K-GA-medoids and KNNs has effectively improved the efficiency of English vocabulary classification compared with the traditional KNN algorithm, while ensuring the classification accuracy. We found that students in college English course consider word memorization as a difficult learning task, and the traditional vocabulary teaching methods are not very effective, and the knowledge of etymology is often little known and rarely covered in classroom lectures. Therefore, the article explores new ideas and strategies for teaching vocabulary in college English from the perspective of etymology.

## 1. Introduction

Vocabulary is the most fundamental element in the learning of almost any language. For second language learners, vocabulary directly determines their language proficiency and ability [1]. College English is arguably the most prevalent subject in Chinese higher education for second language learners. Vocabulary teaching has always been one of the priorities in college English classes, but front-line college English teachers, when surveying students' vocabulary mastery, find that students' vocabulary mastery is not as good as it could be, and that the traditional teaching methods of paraphrasing and explaining usage are not very helpful to students' vocabulary mastery. In recent years, the college English curriculum has been constantly improved and has achieved good results, but the vocabulary teaching still stays in the process of reading and studying chapters with sporadic memorization of words or through rote memorization, instilling vocabulary [2]. As a result, vocabulary teaching in college English classes is either dull and

boring, not very interactive, or students listen to the lectures with lively but not deep impressions; in short, word memory is always a difficult point. If not paid attention to, this will create a hidden danger for the subsequent development of students' comprehensive English literacy [3].

As the old saying goes "It is better to teach a man to fish than to teach him to fish". However, what is "fishing" in the process of teaching vocabulary and how to teach "fishing" instead of "fish" is a question worthy of discussion by all university English teachers and the foreign language community as a whole [4]. To this end, this study will explore the cultural background behind English vocabulary from the perspective of etymology and integrate the knowledge of etymology into the classroom to make each word come alive, so that students can love learning vocabulary and increase their interest in learning.

How to classify English vocabulary more efficiently and accurately is currently a hot topic of inquiry among scholars in the field of computing. With the growing number of messages on the web, the performance demand for English

vocabulary classification tasks is also increasing. For now, the classification methods widely used in English vocabulary classification work are mainly artificial neural networks [5, 6], KNN [7], decision trees [8], support vector machines [9], and plain Bayes [10].

This paper integrates the traditional K-medoids clustering idea with a genetic algorithm to form a new clustering method K-GA-medoids and then combines it with KNN for English vocabulary classification. The algorithmic framework proposed in this paper is proved to be effective in reducing the computational overhead while ensuring the accuracy of the classification results.

## 2. Related Work

The study of English etymology has a long history. “Etymology denotes etymology, which comes from the Greek etymology, a discipline devoted to the study of the origin [11], history, and changing meanings of words, the study of words from a historical perspective, a branch of lexicography, and a part of comparative linguistics” [12]. The main elements of etymological strategies are the identification of cognates and etymological meanings. “Understanding etymology helps comprehension and acquisition.” For the current state of English teaching, [13] points out in his article that vocabulary is learned in a random, unimaginative, and isolated way. “The traditional method of teaching vocabulary is for the teacher to give the meaning in Chinese and then give examples. There is usually no need for the teacher to explain the content. Just recite the vocabulary through a word list.” He also believes that etymological separation of words and meanings is an important reason why students find English words difficult to understand, learn, and memorize.

During the teaching process, teachers share the stories behind the vocabulary to help students strengthen their memory of the words. Some scholars have explored the integration of etymology in vocabulary instruction from the perspective of meaning. Al-Ahdal [14] points out that etymology’s can stimulate students’ interest in learning, help them understand vocabulary, as well as improve their vocabulary use. The elaboration of the role of etymology affirms the rational use of etymology in vocabulary teaching, which enables students to better master and use vocabulary. Some scholars have discussed the strategies of vocabulary teaching from the perspective of etymology. Siregar [15] systematically proposes vocabulary teaching strategies. She believes that understanding the etymology of words helps comprehension and acquisition. Etymological strategies he proposed include etymological strategies, comprehensive etymological strategies, and general etymological strategies.

The core idea of English vocabulary classification is to categorize English words to be classified by using known categories of English words, so that they belong to one or several categories of the categories [16]. KNN is widely used in practical classification work because of the advantages of its algorithm such as mature theory and easy to understand. To address the problem that the traditional KNN has huge computational overhead when the sample size of the training

set is large or the sample dimension is high, [17] proposed a convolutional neural network-based English vocabulary classification algorithm CKNN, which first obtains more abstract feature values from short English vocabulary and then performs the classification work related to English vocabulary afterwards. Egorova and Burget [18] proposed a KNN algorithm based on partitioning regions, firstly determining the closeness of the samples to be tested in each region and secondly applying KNN to categorize them accordingly. Kumar et al. [19] proposed a KNN algorithm based on rough sets to reduce the dimensionality of the vectors in the sample space by attribute reduction and then improve the efficiency of categorizing English words. McCurdy et al. [20] proposed a new classification algorithm PCA&KNN, which uses a smaller set of neighbors to perform KNN classification and effectively reduces the complexity of classification computation [21].

## 3. Related Concepts

The main algorithms involved in this paper are genetic algorithm, K-medoids and KNN, which are briefly described as follows.

**3.1. Genetic Algorithm.** The basic framework of the genetic algorithm is shown in Figure 1, which first encodes the representation of individuals based on the actual problem and then performs the random generation of the initial population set to obtain multiple initial feasible solutions. All individuals in the population are given corresponding fitness values to evaluate their degree of excellence [22].

According to the criterion of selecting the best and the worst, the best individuals from the existing population are selected for subsequent mutation and genetic operations. For the subsequently generated populations, the selection, crossover, and mutation operations are performed cyclically, and the algorithm ends when the termination conditions are satisfied, and the approximate optimal solution of the actual problem is obtained [23].

**3.2. K-Medoids Clustering Algorithm.** The K-medoids algorithm always selects the sample point closest to the cluster center as the cluster center when clustering, in order to minimize the influence of various anomalies and outliers on the selection of cluster centers. The algorithm is based on the following idea: firstly, K initial cluster centers are selected, and the remaining samples are grouped into clusters corresponding to their closest cluster centers [24]. Next, noncluster center sample points are selected to replace the current cluster centers, and the current set of cluster centers is updated or not according to the cost of replacement, and so on.

**3.3. KNN Algorithm.** KNN is a theoretically mature and easy-to-understand classification algorithm whose core idea is that a category of samples to be classified is the same as the

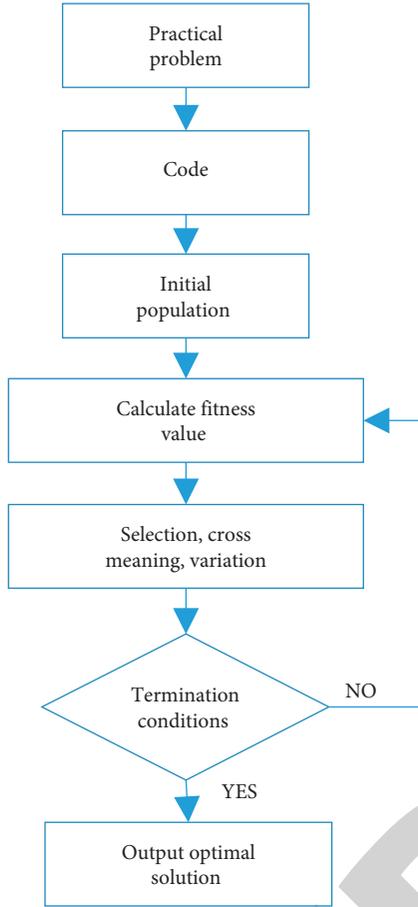


FIGURE 1: Genetic algorithm flow chart.

category with the largest weight in the sample set consisting of the  $K$  nearest samples [25].

#### 4. Introduction to the Algorithm Framework

$K$ -medoids is optimized from  $K$ -means algorithm, which greatly reduces the influence of noise and outliers on the final clustering effect, but it is still sensitive to the selection of initial clustering centers. This framework clusters the training sample set in the first stage of classification and reduces the training set in the classification stage based on the distance between the English words to be classified and the cluster centers, so as to overcome the problem that the classification efficiency of the traditional KNN algorithm will be seriously reduced and lack of effectiveness when the sample size is large problem [26].

**4.1.  $K$ -GA-Medoids Algorithm.**  $K$ -GA-medoids consists of genetic algorithm fused with the core idea of  $K$ -medoids clustering. In the design part of the fitness function of the genetic algorithm, the  $K$ -medoids idea of finding the centroids within clusters is fused with the distance as the core of the fitness function setting, aiming to find the set of cluster centroids closer to the cluster center through genetic iteration as follows:

$$S_i = \frac{1}{(N_i - 1)} \sum_{P \in C_i} |P - M_i|, \quad (1)$$

where  $C_i$  is the  $i$ th cluster of the clustering result,  $N_i$  is the size of the  $i$ th cluster,  $M_i$  denotes the cluster center,  $|P - M_i|$  is the distance from sample  $P$  to cluster center  $M_i$ , and  $S_i$  is the average of the distances from all sample points in the  $i$ th cluster to the cluster center [27].

$$\text{Fit} = \frac{1}{K} \sum_{i=1}^K S_i, \quad (2)$$

The smaller the value, the better the clustering effect, and the relationship between the two is defined as an inverse correlation.

Before the iteration of the algorithm, a certain number of feasible solutions are randomly generated as the initial generation of the population, and the individuals in this generation are judged to be good according to the fitness function, and then the corresponding genetic operations are performed for the individuals in this generation to ensure the continuous evolution and expansion of the population and to provide options for finding better feasible solutions.

- (1) Set parameters: number of clustering centers  $K$ , initial population size  $N$ , selective survival ratio  $P_1$ , the crossover probability  $P_2$ , variation probability  $P_3$ , maximum number of iterations  $T$ , threshold of change in the average fitness of the population  $\epsilon$
- (2) Population initialization: coding the first generation of individuals based on  $K$  values to obtain an initial population of population size  $N$
- (3) Determination of the degree of excellence of the contemporary individuals according to equation (2) and retention of the best individuals so far
- (4) To perform the corresponding genetic operation on the individuals of the population according to the determination result and parameter  $P_1$ ,  $P_2$ , and  $P_3$  and obtain the new generation of the population after one round of iteration
- (5) Determine whether the difference between the average fitness value of the contemporary population and the parent is less than the threshold  $\epsilon$  [28]; if it is, the algorithm terminates and outputs the result; otherwise, turn to step 3

**4.2. Sample Reduction Based on  $K$ -GA-Medoids Clustering.** When using the KNN algorithm to categorize English words, the  $K$ -nearest English words obtained by using the distance formula are normally in the same category or close to the category of the English words to be tested, and the computational overhead can be greatly reduced if the search for the nearest English words is conducted in the sample set of the category or the surrounding categories of the English words to be tested as much as possible. The accuracy of the classification results is inevitably reduced when the reduction of the sample set is large, and the improvement is not obvious when the reduction is small. To address this

problem, we propose a simple reduction method based on K-GA-medoids clustering by considering a compromise between accuracy and computational overhead [29].

The training sample set is defined as  $S$ , containing  $C_1, C_2, C_3, \dots, C_m$  a total of  $m$  categories  $M$  English words.

Calculate the magnitude of the similarity distance between  $d$  and  $O_i$  ( $0 < i \leq n$ ).

Compromise calculation to obtain the sample reduction metric distance  $D_{\text{end}} = (D_{\text{max}} + D_{\text{min}})/2$  for the English word  $d$  to be measured.

Define the reduced training set of the current English vocabulary  $d$  to be tested as  $S_{\text{new}}$ , if the similarity distance between  $d$  and is  $D_i \leq D_{\text{end}}$ , it means that  $d$  is more similar to the English vocabulary in the corresponding cluster of  $O_i$ , and the English vocabulary in it will be included in  $S_{\text{new}}$ ; otherwise, the corresponding cluster will be discarded.

**4.3. Improving the Combination of Clustering and KNN.** Before the formal KNN classification work, the K-GA-medoids algorithm is used to cluster the training sample set to obtain multiple clusters, and in the classification stage, the corresponding sample reduction distance is obtained according to the similarity distance between each English word to be tested and each cluster center as shown in Figure 2. In the classification stage, we obtain the corresponding sample reduction distance according to the similarity distance between each English word to be tested and each cluster center, so as to reduce the training set and the computational overhead [12]. The corresponding algorithm flow and diagram is as follows:

- (1) Read the sample data from the training set and the test set, which are all English words, and perform word separation, deactivation and frequency counting
- (2) Adopt the TF-IDF calculation method to calculate the weight size of each feature word and get the undimensionalized English vocabulary vector
- (3) Adopting the information gain method to reduce the dimensionality of the English vocabulary feature vector and retaining the feature terms with larger gain values
- (4) Performing K-GA-me-Does cluster operation on the English vocabulary vector in the training space to divide it into multiple clusters with higher cohesion
- (5) For each English word  $d$  to be tested in the test space, the training space is reduced using the sample reduction method in Section 3.2 to obtain a new training space corresponding to it  $S_{\text{new}}$
- (6) The KNN algorithm is used to classify the English word  $d$  to be tested, find its  $K$  nearest samples in  $S_{\text{new}}$ , and assign  $d$  to the sample category with the largest proportional weight

## 5. Experimental Results and Analysis

All experiments in this section are conducted on a PC configured with a 2.20 GHz Intel Core i5, 8 GB of RAM, with

Win7 as the operating system, and the algorithm program is written in Python language and programmed in Python3. Implementation: the evaluation metrics involved in the experimental results include precision, recall, and F-value, and the corresponding calculation formulae are defined as follows:

$$\begin{aligned} P &= \frac{TP}{TP + FP}, \\ R &= \frac{TP}{TP + FN}, \\ F &= \frac{P * R * 2}{P + R}, \end{aligned} \quad (3)$$

where TP denotes the number of predicted positive and actual positive, FN denotes the number of predicted negative but actually positive, and FP denotes the number of predicted positive but actually negative.

**5.1. K-GA-Medoids Clustering Effect Analysis.** To verify the effectiveness of K-GA-medoids clustering, this paper will do an experimental comparison of K-medoids and K-GA-medoids on the same data set. The classical English vocabulary clustering collection `mini_newsgroups` is used, which includes a total of 2000 English words in 20 major categories. Before the experiment, the English vocabulary was preprocessed, including deactivation, TF-IDF weight calculation, and feature selection, and then the English vocabulary feature matrix was obtained. Finally, experiments are conducted on K-medoids and K-GA-medoids algorithms for different test set groups. The description of the experimental test sets and the specific experimental results are as shown Table 1.

As shown in Figure 3 that the data obtained from the above experiments are the average values obtained from multiple experiments, and the results show that the K-GA-medoids algorithm has significantly improved the accuracy, recall, and F-value of the clustering effect compared with the general K-medoids algorithm, and the F-value of the five experiments has increased by 2.94% on average, which is better in the clustering of English words. This shows that the new clustering method, K-GA-medoids, which is derived from the integration of GA and K-medoids algorithms, is objective and effective in practical applications.

**5.2. Analysis of the Effect of English Vocabulary Classification Algorithm Framework.** In order to verify the effectiveness of the algorithmic framework of this paper, this section compares it with the traditional KNN algorithm on the same dataset for experiments. Since the number of samples of individual categories in the `mini_newsgroups` dataset is small, which cannot guarantee the diversity and comprehensiveness of the training set in the early stage of the classification work, the experimental dataset in this section adopts five categories of news English vocabulary data crawled by the network, and the test is divided into different

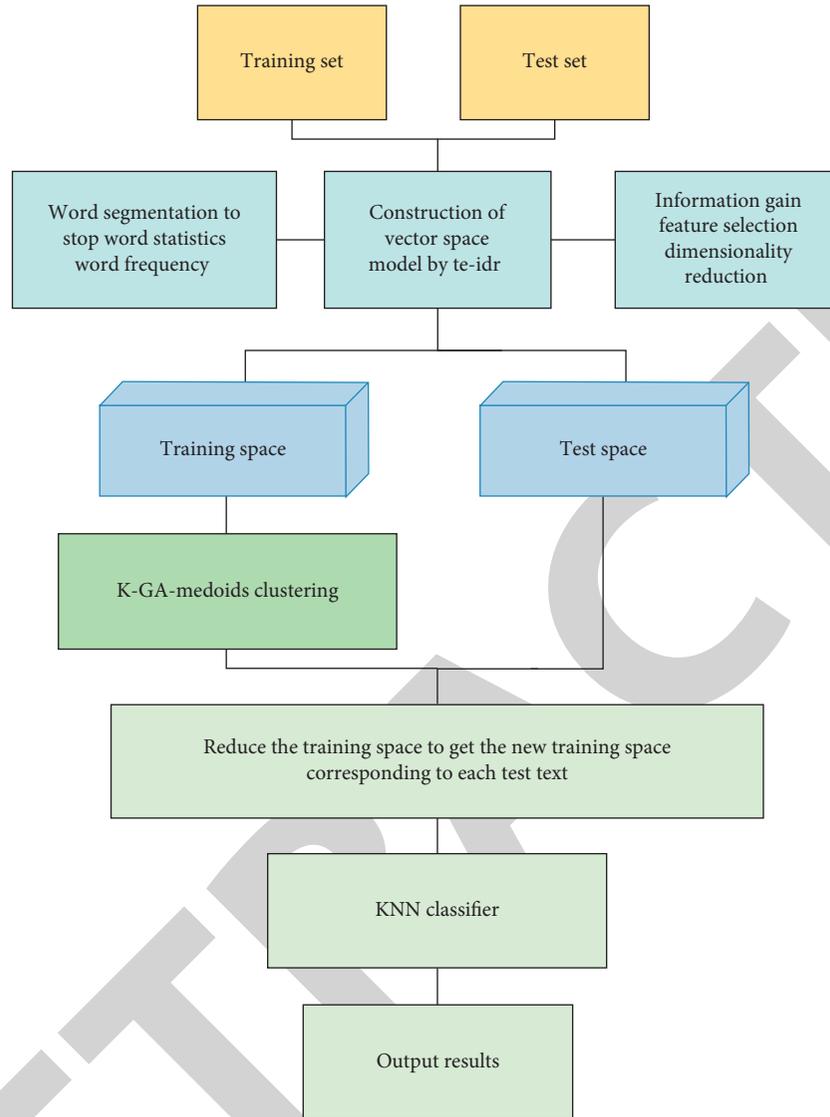


FIGURE 2: Classification algorithm framework flow chart.

TABLE 1: Test set description and experimental data of clustering.

Test set	Category/text number	K-medoids			K-GA-medoids		
		<i>P</i>	<i>R</i>	<i>F</i>	<i>P</i>	<i>R</i>	<i>F</i>
Group 1	2/120	0.5547	0.5525	0.5536	0.6065	0.5667	0.5857
Group 2	4/240	0.4055	0.3983	0.4017	0.4336	0.4071	0.4198
Group 3	6/360	0.4515	0.3828	0.4143	0.4600	0.4286	0.4433
Group 4	6/540	0.4071	0.4037	0.4053	0.4760	0.3850	0.4249
Group 5	8/720	0.3265	0.3107	0.3182	0.3988	0.3394	0.3662

test set groups. The experimental test sets are described, and the specific experimental results are shown in Table 2.

From Figure 4, we can see that, compared with the traditional KNN algorithm, the algorithm framework in this paper has an average increase of 1.72% in the accuracy of *p* value and 7.29% in the classification efficiency of English vocabulary in the five test sets, which indicates that the algorithm framework proposed in this paper has effectively

accelerated the classification efficiency of English vocabulary while ensuring the improvement of the classification accuracy. The scores of different groups are different. We can know that the score of KNN is improved after optimization, and good performance can be obtained without too much adjustment. We usually build the nearest neighbor model quickly, but if the training set is large (a large number of features or a large number of samples), the prediction speed

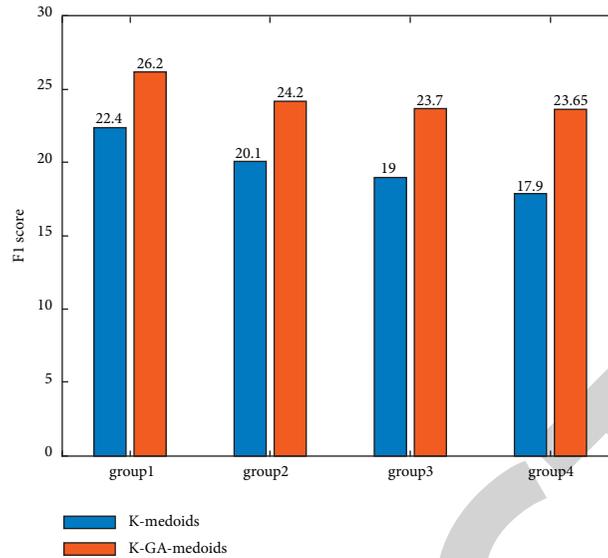
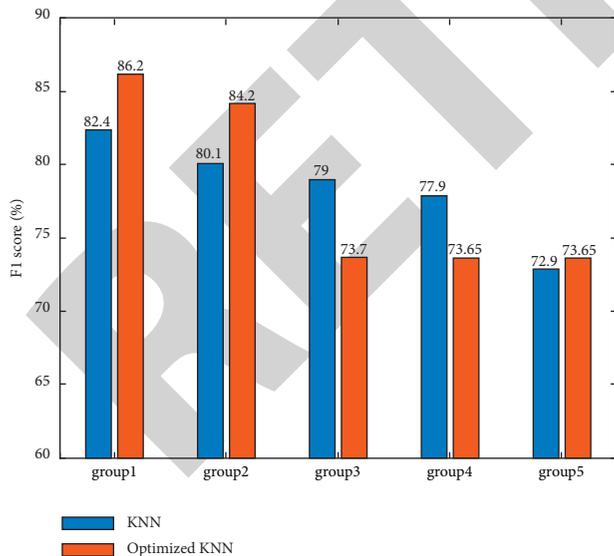


FIGURE 3: Comparison of F-values of clustering effects.

TABLE 2: Test set description and classification experimental data.

Test set	Category/text number	K Value	KNN		Optimized KNN	
			<i>P</i>	<i>T</i> (s)	<i>P</i>	<i>T</i> (s)
Group 1	2/300	7	0.8933	34.82	0.9200	32.65
Group 2	2/600	7	0.8617	66.96	0.8700	57.31
Group 3	3/900	13	0.9111	133.04	0.9827	109.57
Group 4	4/1200	13	0.8358	262.89	0.86257	223.57
Group 5	5/1500	9	0.8746	419.18	0.8787	358.83

FIGURE 4: Comparison of *p* values for classification effects.

may be slow. When using the optimized KNN algorithm, data preprocessing is very important. This algorithm is often ineffective for data sets with many features (hundreds or more), especially for data sets with most values of 0 for most features (so-called sparse data sets).

## 6. Conclusions

In this paper, we propose a new English vocabulary classification algorithm framework by combining a clustering algorithm with KNN to reduce the computational overhead by reasonably reducing the training sample set. At the same time, in order to better improve the clustering effect of the training set of English vocabulary, this paper integrates the genetic algorithm with the K-medoids algorithm to form a new clustering algorithm K-GA-medoids, and the experiments show that its clustering effect is significantly better than the traditional K-medoids, and the classification algorithm framework formed by combining it with KNN is better than the KNN algorithm in terms of classification performance and the classification algorithm framework combined with KNN is better and more efficient than the KNN algorithm. However, the reduction of the sample set inevitably causes the loss of sample information.

So, how to reduce the training set samples more effectively is still a direction we need to explore and study in the future. Using more data makes the effect of model learning better. All data (including teaching interaction process, homework, and wrong questions) learned by each student are fully recorded on the intelligent learning platform. When these daily continuous and massive data are continuously injected into the artificial intelligence system, it can not only help the system better clarify the pedigree of knowledge

points and provide data support for the teaching and research team to optimize the teaching system but also greatly help teachers understand students' learning psychology and mastery of knowledge points, and provide important help for targeted teaching.

### Data Availability

The dataset used in this paper are available from the corresponding author upon request.

### Conflicts of Interest

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

### Acknowledgments

This work was supported by the Research Project on Education and Teaching Reform of Weinan Normal University: Study on Cultivating Patterns for English Normal Students in Local Normal Universities under professional certification (grant no. JG202150).

### References

- [1] X. Wang, "Discussion on application of multimedia teaching in college English vocabulary teaching," *Open Journal of Modern Linguistics*, vol. 6, no. 3, pp. 177–181, 2016.
- [2] M. Ahltoft, M. Skeppstedt, S. Kitajima, A. Henriksson, R. Rzepka, and K. Araki, "Expansion of medical vocabularies using distributional semantics on Japanese patient blogs," *Journal of Biomedical Semantics*, vol. 7, no. 1, p. 58, 2016.
- [3] E. Lee and M. Park, "Student presentation as a means of learning English for upper intermediate to advanced level students," *Journal of Pan-Pacific Association of Applied Linguistics*, vol. 12, no. 1, pp. 47–60, 2008.
- [4] L. Su, J. N. Liu, L. F. Ren, and F. Zhang, "An object classification approach based on randomized visual vocabulary and clustering aggregation," in *Proceedings of the 2nd International Conference on Mechatronics and Control Engineering*, pp. 778–782, Amsterdam, The Netherlands, October 2013.
- [5] Z. Y. Lu and S. H. Fan, "A mixed-method examination of adopting focus-on-form TBLT for children's English vocabulary learning," *English Language Teaching*, vol. 14, no. 2, p. 37, 2021.
- [6] Y. Luo, M. Wang, and Z. Le, "An improved kNN text categorization algorithm based on cluster distribution," *Journal of Computational Information Systems*, vol. 8, no. 3, pp. 1255–1263, 2012.
- [7] N. Pazyura, N. Nychkalo, J. Wang, L. Lukianova, N. Paziura, and N. Muranova, "Use of task-based approach in teaching vocabulary to business English learners at university," *Advanced Education*, vol. 16, no. 16, pp. 98–103, 2020.
- [8] A. Irawan, A. Wilson, and S. Sutrisno, "The implementation of duolingo mobile application in English vocabulary learning," *Scope Journal of English Language Teaching*, vol. 5, no. 1, p. 08, 2020.
- [9] A. Jain, D. K. Tayal, and S. Yadav, "Retrieving web search results using Max–Max soft clustering for Hindi query," *International Journal of System Assurance Engineering & Management*, vol. 7, no. 1, pp. 1–12, 2016.
- [10] M. Lamb and F. E. Arisandy, "The impact of online use of English on motivation to learn," *Computer Assisted Language Learning*, vol. 33, no. 1–2, pp. 85–108, 2020.
- [11] J. Tao, C. Zheng, Z. Lu, and J. C. Liang, "Cluster analysis on Chinese university students' conceptions of English language learning and their online self-regulation," *Australasian Journal of Educational Technology*, vol. 36, no. 2, pp. 105–119, 2020.
- [12] J. Wang, "Research on college english teaching strategies and applications based on big data," in *Proceedings of the 2019 International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI)*, pp. 283–287, IEEE, Taiyuan, China, November 2019.
- [13] W. Zhao, B. Wang, D. Coniam, and B. Xie, "Calibrating the CEFR against the China standards of English for college English vocabulary education in China," *Language Testing in Asia*, vol. 7, no. 1, pp. 1–18, 2017.
- [14] A. Al-Ahdal, "Code mixing in Arabic conversations of college students: a sociolinguistic study of attitudes to switching to English," *Asian ESP Journal*, vol. 16, no. 1, pp. 6–19, 2020.
- [15] I. I. Siregar, *The Effect of Using Clustering Strategy by Random Cards Media on the Students English Vocabulary*, Universitas Muhammadiyah Sumatera Utara, Medan, Indonesia, 2019.
- [16] R. Rukmana, *Increasing the Students' Vocabulary by Using Clustering Technique at the Eighth Grade of SMPN 2 Lembang Kabupaten Pinrang*, IAIN Parepare, Soreang, Indonesia, 2020.
- [17] X. He and A. Godfroid, "Choosing words to teach: a novel method for vocabulary selection and its practical application," *TESOL Quarterly*, vol. 53, no. 2, pp. 348–371, 2019.
- [18] E. Egorova and L. Burget, "Out-of-vocabulary word recovery using fst-based subword unit clustering in a hybrid ASR system," in *Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 5919–5923, IEEE, Calgary, Canada, April 2018.
- [19] J. Kumar, R. Kumar, A. U. Haq, and S. Shafiq, "A non-parametric multi-lingual clustering model for temporal short text," in *Proceedings of the 2020 seventeenth international computer conference on wavelet active media technology and information processing (ICCWAMTIP)*, pp. 58–61, IEEE, Chengdu, China, December 2020.
- [20] K. McCurdy, S. Goldwater, and A. Lopez, "Adaptor grammars for unsupervised paradigm clustering," in *Proceedings of the Eighteenth SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology*, pp. 82–89, Bangkok, Thailand, August 2021.
- [21] L. Kotlerman, I. Dagan, and O. Kurland, "Clustering small-sized collections of short texts," *Information Retrieval Journal*, vol. 21, no. 4, pp. 273–306, 2018.
- [22] D. Wu, C. Zhang, L. Ji, H. Wu, and Y. Wu, "Forest fire recognition based on feature extraction from multi-view images," *Traitement du Signal*, vol. 38, no. 3, pp. 775–783, 2021.
- [23] A. Alahmadi, C. Shank, and A. Foltz, "Vocabulary learning strategies and vocabulary size: insights from educational level and learner styles," *Vocabulary Learning and Instruction*, vol. 7, no. 1, pp. 1–21, 2018.
- [24] S. Naeem and A. Wumaier, "Study and implementing K-mean clustering algorithm on English text and techniques to find the optimal value of K," *International Journal of Computer Applications*, vol. 182, no. 31, pp. 7–14, 2018.
- [25] L. Wang, C. Zhang, Q. Chen, S. Liu, Z. Yang, and H. Li, "A communication strategy of proactive nodes based on loop theorem in wireless sensor networks," in *Proceedings of the 2018 Ninth International Conference on Intelligent Control*

- and Information Processing (ICICIP), pp. 160–167, IEEE, Wanzhou, China, November 2018.
- [26] A. A. Bagherian Poor and E. Serati, “The effect of post-teaching techniques on the English vocabulary recall and retention of Iranian students with learning disability,” *Journal of Teaching Language Skills*, vol. 40, no. 1, pp. 117–152, 2021.
- [27] R. G. Radu, I. M. Rădulescu, C. O. Truică, E. S. Apostol, and M. Mocanu, “Clustering documents using the document to vector model for dimensionality reduction,” in *Proceedings of the 2020 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR)*, pp. 1–6, IEEE, Cluj-Napoca, Romania, May 2020.
- [28] T. Weißer, T. Saßmannshausen, D. Ohrndorf, J. Wagner, and P. Burggrä, “A clustering approach for topic filtering within systematic literature reviews,” *MethodsX*, vol. 7, Article ID 100831, 2020.
- [29] H. Li, D. Zeng, L. Chen, and Q. Chen, “Immune multipath reliable transmission with fault tolerance in wireless sensor networks,” in *Proceedings of the International Conference on Bio-Inspired Computing: Theories and Applications*, pp. 513–517, Springer, Singapore, October 2016.