Advances in Vocational and Technical Education (2022)
Clausius Scientific Press, Canada

DOI: 10.23977/avte.2022.040408
ISSN 2523-5834 Vol. 4 Num. 4

The Construction of Higher Vocational English Teaching System in the Era of Big Data

Ying Liu

Liaoning Urban Construction Technical College, Shenyang, Liaoning, China

Keywords: Big data era, English teaching system, K-modes clustering algorithm, teaching ability evaluation

Abstract: In view of the many problems in the college English teaching, the construction of a college English teaching system is mainly through in-depth exploration of the K-modes clustering algorithm, proposes an improved K-modes algorithm, and analyzes the improvement. The application of the latter algorithm in the evaluation of teaching status and ability of teachers. The test data shows that the clustering results obtained by the improved K-modes clustering algorithm are more relevant and more practical, which can help us to improve students' evaluation of teaching.

1. Introduction

Becauseof the development of modern technology, English teaching is becoming intelligent and the teaching quality is significantly improved. Because the traditional teaching has many problems, and it becomes not suitable for teaching now. Therefore, college teachers must be keenly aware of the pressure and challenges in the teaching process in the context of the current big data era, make full use of big data to explore teaching resources, and establish an intelligent teaching platform according to different students' English learning needs to achieve one-to-one. Teaching mode and layered teaching. Efforts to improve the efficiency and level of students learning English. In response to the above problems, a former scholar Li Zhiying studied the English teaching in university. And Deng Yiting proposed a research on college English teaching reform in the era of micromedia. In order to improve the teaching ability, combined with the experience research of the above scholars, it is proposed to build a college English teaching system. According to its shortcomings, the K-modes algorithm is improved, and the improvement is analyzed. Whether the latter algorithm is suitable for current English in university.

2. The necessity of constructing English teaching system with big data

2.1 Opportunities brought to English teaching

Now, earth-shaking changes have taken place in teaching. Education big data comes from this. The main theme is big data. It is mainly used in education data mining, analysis, evaluation, decision-making and other aspects. At the same time, it is also repeatedly used in the education field for independent branches of directional thinking learning. Reform and optimize education in a more

comprehensive way to improve teaching quality. Among them, college English teaching is useful to improve the students' comprehensive level. Educational big data provides development opportunities for English teaching. Computer network technology and cloud platform make English teaching more convenient, and traditional English teaching has been improved, thereby enhancing the diversity and interest of English teaching, and students' learning interest have been significantly improved. Schools can build an English teaching resource database, focusing on recording the energy of students at each stage, sorting out and analyzing it, and putting forward a plan for teaching students. At same time, the development of big data is used to improve the teaching, and the timeliness is enhanced.

2.2 Challenges brought to English teaching

2.2.1 Mass teaching resources weaken the status of traditional teaching materials

With the big data, traditional teaching has been unable to cope with the massive teaching resources. A variety of teaching resources have emerged from the Internet. Traditional teaching materials cannot keep up with the updates of the times, and teaching content is gradually out of date. The current advanced computer network technology and cloud platform provide college students with rich and diverse learning resource data. Teaching methods under big data are more lively and interesting. Students are more willing to obtain learning resources from computer network technology. Traditional textbooks cannot be reused. Thereby reducing the status of traditional teaching materials.

2.2.2 Changing the roles of teachers and students

In the current English classroom teaching, most of the teachers dominate the classroom andstudents passively accept learning in the classroom. Teachers randomly let students to answer questions, and a very small number of students actively answer questions raised by teachers. The long-term use of this classroom teaching model will make teachers feel tired, students will not be able to experience the fun of English teaching, reduce students' desire to learn, and will not enable students to have a relaxed and active classroom environment. Thebig data technology is effectively applied to classroom teaching to solve the boring and boring situation of the current English teaching model. Teachers use micro-classes and flipped teaching methods to transfer the knowledge that needs to be taught to the teaching platform, and students download according to their own needs, so the teaching is not limited by time and space.Thereby enhancing students' interest in learning and the ability of independent learning. Through the big data technology, the role relationship between teachers and students has been changed. Teachers no longer pretend to dominate teaching in the classroom, and students begin to become the importantroleinclass.

3. Model innovation of English teaching evaluation

3.1 Determine the number of clusters K

In the English learning system, the traditional K-modes algorithm calculates the known cluster cluster number *K*, and the calculation effect is not good. In this study, a variety of index clusters were selected to calculate the cluster family number *K* by minimizing the sum of squared error (SSE).

Error square sum formula (1):

$$SSE = \sum_{l=1}^{k} \sum_{x \in L_l} Dist(x, z_l)^2 \quad (1)$$

In the above, k represents the number of clusters, Z_I represents the cluster center of I cluster L_I, and $Dist(x,Z_I)^2$is the similarity between the x and the cluster center Z_I. Different similarity calculation methods have different calculation results. The frequency-based similarity calculation is determined by the basic attributes of the sample, which has certain limitations. Therefore, the frequency-based similarity calculation is applied to the English teaching system.

3.2 Determine the center of the initial clustering algorithm

Relying on the principle of shortest distance and constant updating, the goal is to minimize SSE. Based on the above ideas, the following model is established:

$$\begin{cases} z = \{z_l | z_l \in X, \text{And } z_i \text{ and } z_{i-1} \text{ are different}, \ l \le K\} \\ z_l = \max\limits_{x \in X} DXJSSE(x) \\ DXJSSE(x) = \left(\sum_{j=1, x \in L_j}^{l-1} Dist(x, x_j)^2 + \min\limits_{x \in X} (\sum_{i=1}^{n} Dist(x, x_t)^2 \right) \end{cases} \quad (2)$$

In Model 2, Z represents the sum of the initial cluster centers, Z_I represents the sample data point x when the sum of square errors SSE declines the fastest, where I represents the initial cluster center, and x represents the sample data, n is expressed as a single number of samples. The closest distance calculation method of the traditional K-modes algorithm is to compare and measure the similarity between the test sample data and the cluster center data. This method is generally feasible, but when the attributes of the samples are different,there will be many problems. The study adopts the distance measurement method of co-occurrence rate, and compares the traditional K-modes algorithm measurement to the similarity between the sample data of different attributes and the cluster center data.

4. Test data analysis

Take one of the semester evaluations of School A and School B as an example, and compare and analyze the final calculation results of three K-modes algorithms, mainly from the cluster center, the proportion, the smallest error square sum. The specific criteria are: the cluster center corresponding to the clustering method, the number of cluster samples and calculation results of the actual samples in the cluster calculation, and the sum of the square of the minimum error. It is expressed as the formula (3).

$$\begin{aligned} \text{Correct rate} &= \left(\sum_{i=1}^{k} e_i \right) / n \\ \text{Recall rate} &= \left(\sum_{i=1}^{k} \left(\frac{e_i}{e_i + m_i} \right) \right) / k \end{aligned} \quad (3)$$

In the above, k represents clusters, n is samples, e_i is the number of samples allocated to class i, and m_i is the number of samples that should be allocated to class i but not allocated. According to the above formula, compare the indexes corresponding to the three different clustering algorithms in a certain semester of School A and School B.

Table 1: Comparison of indicators of three different clustering algorithms for teaching evaluation data in a certain semester of a school

algorithm	Cluster center	Number of samples	Number of result samples	Minimum error sum of squares	Correct rate	Recall rate
Randomly determine the initial center 04 matching distance	(4,4,4,5)	867	295	1564	0.925029	0.920513
	(5,5,3,3)		326			
	(3,5,5,3)		181			
SEE determines the initial center AVF calculation distance	(5,4,4,5)		329	947	0.936563	0.940641
	(3,3,3,5)		279			
	(3,5,2,3)		204			
SEE determines the initial center Co-Cccuraneay to calculate the distance	(5,5,4,3)		358	703	0973472	0.970501
	(4,4,3,3)		329			
	(3,3,3,2)		157			

The comparison of the index results of the above three different clustering algorithms shows that the improved K-modes clustering algorithm has a better clustering effect than the first two algorithms.

In the experiment, we can know that the ultimate goal of clustering is to make the distance between clusters smaller and the distance between clusters larger . In order to obtain more real and accurate results, showing the advantages anddisadvantages of the three algorithms, the calculation of the distance between the clusters is set to within $D_{内}$, and the distance between clusters is set to between $D_{间}$ to calculate the formula (4):

$$D_{内\mathrm{r}} = \frac{\sum_{x_i \in z_r} \sum_{x_j \in z_r} D(x_i, x_j)}{n_r^2}$$
$$D_{内\mathrm{rt}} = \frac{\sum_{x_i \in z_r} \sum_{x_j \in z_t} D(x_i, x_j)}{n_r n_t} \quad (4)$$

In formula (5), Z represents the r-th cluster, n represents the number of samples within the r-th cluster, and $r \in \{1,2,...,k\}$.

Because the distance measurement methods in the clustering algorithm are different, you can't just use the intra-cluster distance $D_{内}$ and the distance $D_{间}$ between the clusters. The calculation process is too simple to see the advantages and disadvantages of the clustering algorithm, so we use the average distance within the cluster is compared with the average distance between several categories. The smaller the comparison ratio data, the better the clustering effect. The calculation process is shown in formula (5):

$$P = \frac{1}{k}\sum_{r=1}^{k} \frac{D_{内r}}{\frac{1}{k-1}\sum_{r\neq t} D_{间rt}} \quad (5)$$

The final results obtained after calculation by formula (5) are shown in the comparison table of the average distances between classes and classes of three different intra-office algorithms in Table 2.

Table 2: Comparison of the average distance between classes within and between classes of three different clustering algorithms for teaching evaluation data in a certain semester

Randomly determine the initial center 0-1 matching calculation distance				SEE determines the initial center AVF calculation distance			SEE determines the initial center Co-Cccurencey calculates the distance		
	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3	Z_1	Z_2	Z_3
Z_1	0.7241	0.8365	0.8724	0.3197	0.7428	0. 7913	0.2302	0.6974	0. 7102
Z_2		0.5675	0.8102		0.3687	0. 7465		0.2457	0. 6987
Z_3			0. 4578			0. 3 161			0.2154

It can be seen from Table 2 that the smallest ratio data of the improved K-modes clustering algorithm indicates that the clustering effect is the best, which shows that the final calculation results obtained by calculation have stronger relevance. It further proves the improved K-modes clustering algorithm is more practical and can solve the problem of students evaluating teaching in teaching.

4. Conclusion

The final test proved that the clustering results obtained by the improved K-modes clustering algorithm are more relevant and practical, and can improve students' evaluation of teaching in college English teaching, thereby improving teaching quality and teaching efficiency. This proves that this algorithm is feasible in constructing college English evaluation system.

References

[1] Li Zhiying. (2020). Innovation and development of college English teaching in the era of big data. Employment and Security, vol.22, pp:141-142.

[2] Deng Yiting.(2021). Research on College English Teaching Reform in the Micromedia Era. Forest District Teaching, vol.3, pp: 71-74.

[3] Zhang Jiajie, Dong Yan. (2019). The application and practice of new interactive mechanism in college English teaching. Heilongjiang Higher Education Research, vol.37, no.2, pp: 156-160.

[4] Huang Caili. (2019). Analysis on the innovation and development of college English teaching under the background of big data. Talent, vol.8, pp: 7.

[5] Wang Yanxia. (2020). Research on the status quo of the use of metacognitive learning strategies for students and the necessity of constructing a smart teaching system: Taking Lanzhou Institute of Technology as an example. Journal of Luliang Education Institute, vol.37, no.1, pp: 50-51.

[6] Tang Hongmei, Zheng Gang. (2020). Thoughts and countermeasures of teaching management in colleges and universities based on big data. Education Modernization, vol.7, no.48, pp: 5-8.

[7] Zhang Liping.(2020). Research on college English teaching based on the "in-class + mobile learning" model. Liaoning Economic Vocational and Technical College. Journal of Liaoning Institute of Economic Management, vol.4, pp: 95-97.

[8] Xu Huifang, Chen Yuanyan. (2020). Exploration of the construction of the teaching quality evaluation system of local universities under the background of big data. Journal of Mudanjiang Institute of Education, vol.7, pp: 35-36+81.
[9] Fang Li. (2018). Research on College English Teaching under the "Internet+" Thinking Mode. English Square, vol.4, pp: 86-87.
[10] Ma Xing, Wang Nan. (2018). The construction of university teaching quality evaluation system based on big data. Educational Research of Tsinghua University, vol.39, no.2, pp: 38-43.