International Journal of Continuing Engineering Education and Life-Long Learning

ISSN online: 1741-5055 - ISSN print: 1560-4624
https://www.inderscience.com/ijceell

An intelligent statistical method of real-time status of English teaching assistance resources from the perspective of MOOCs

Lizhen Shi

DOI: 10.1504/IJCEELL.2024.10057192

Article History:

Received:	05 August 2021
Last revised:	07 February 2022
Accepted:	11 February 2022
Published online:	03 December 2023

An intelligent statistical method of real-time status of English teaching assistance resources from the perspective of MOOCs

Lizhen Shi

Zhoukou Vocational and Technical College,
Zhoukou, 466000, China
Email: lizhens@mls.sinanet.com

Abstract: This study takes improving the clustering statistics quality of resources as the expected goal and designs an intelligent real-time statistics method of English teaching auxiliary resources from the perspective of MOOCs. Firstly, the data of English teaching auxiliary resources are screened, and the relationship between the English teaching assistance resource data is determined according to the association rules, and the cosine similarity between the data is judged by the sparse matrix, so as to effectively reduce the calculation error of data status similarity through data preprocessing. Then, the English teaching assistance resource status matrix is constructed to make statistics on the teaching assistance resource data status. According to the text, the calculation error of the proposed method is less than 4%, and the ARI index is always higher than 0.9. The above results also fully prove the validity of the method.

Keywords: MOOC teaching; support vector machine; SVM; sparse matrix; cluster processing; smarter statistics.

Reference to this paper should be made as follows: Shi, L. (2024) 'An intelligent statistical method of real-time status of English teaching assistance resources from the perspective of MOOCs', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 34, No. 1, pp.100–110.

Biographical notes: Lizhen Shi received her BA degree in the Foreign Language Department from Henan University, in 2004. She is currently an Associate Professor in the Zhoukou Vocational and Technical College. Her research interests include teaching method, interpretation and cross-cultural communication.

1 Introduction

With the deepening of education reform, the construction of teaching resources and assistance resources under the perspective of MOOCs has been developed accordingly. Relevant education departments have invested a lot of costs in the construction of teaching and education resources (Park and Mcleod, 2018). Among them, English teaching resources have also obtained great progress.

At present, the number of English teaching resources shows geometric growth. In the process of development, English teaching resources in different application fields have problems of uneven distribution, slow update rate and others, resulting in the uneven

development of English teaching resources (Tang and Yuan, 2018). The practical application of assistance resources is the key factor affecting its development. Application status of the resources includes resource update, historical data replacement, resource upgrade, to be downloaded, downloaded, etc. Through statistics on the status of language teaching resources, the utilisation rate and practicability of different kinds of teaching resources can be accurately judged, so as to develop more targeted teaching resources that are more popular with users. This is also an important impetus to optimise MOOC level (De Bruin et al., 2018). Therefore, it is very important to design an effective intelligent statistical method for real-time status of teaching resources.

Luo et al. (2018) proposed to design a statistical method of course status based on big data platform. This method firstly designs a big data platform composed of facility layer, service layer and customer layer. In this platform, the business logic sequence of English teaching assistance resources statistics is designed, and the functions of authority control and user division are designed. In the English teaching assistance resource module, the real-time status and quantity of resources are counted and integrated. This method can improve the statistical results of resource status in English assistance teaching resource management, but the platform is designed for many subjects, and the statistics of each resource status is small, so it still needs further improvement. Cao et al. (2019) proposed a statistical management method of English teaching resource status based on message analysis. In this method, the English teaching resources storage module and the server supporting the teaching resources are designed, and the artificial intelligence technology is introduced to develop the teaching resources system database and operating environment support, and the effective statistics and management of teaching resources are realised through the message analysis. The system is designed to manage the database of English teaching assistance resources, in which the content of the status of teaching assistance resources is less. This part should be designed in detail and needs further improvement and perfection. Qing (2020) proposed a statistical method of resource status evaluation based on the multidimensional scaling method. Based on the analysis of the ideas and principles of the multidimensional scaling analysis method, this method uses the multidimensional scaling method and cluster analysis to comprehensively evaluate and analyse the application status of the basic teaching assistance resources on the internet, and divides the application status into four levels.

Based on the above analysis, this paper designs an intelligent statistical method of real-time status of English teaching assistance resources from the perspective of MOOCs. The technical route of this paper is as follows:

1 The location of English teaching assistance resources includes a large area of data of a variety of English teaching assistance resources. First of all, the large area plane of the English teaching assistance resource data is divided according to the linear separable condition, and then, according to the minimum interval between the English teaching assistance resource data, the accurate screening of English teaching assistance resource data is realised by the classification characteristics of support vector machine (SVM).

2 According to association rules, the relationship between data of English teaching assistance resources is determined, and the similarity between data of English teaching assistance resources is determined by setting sparse matrix. In the data

preprocessing stage, redundant data are eliminated effectively according to cosine similarity, thus reducing the statistical error of data status.

3 The status quantity of the preprocessed data is clustered according to the state type of the preprocessed data, and the state statistical matrix is obtained by clustering, so as to complete the statistics of the status of the data of educational assistance resources.

2 An intelligent statistical method for real-time status of English teaching assistance resources under MOOCs

The location of English teaching resources contains a large area plane of various English teaching resources data, so after dividing the large area plane of English teaching resources data, the linear classification ability of SVM is used to screen English teaching resources data, and then the sparse matrix is set to process English teaching resources data.

First, the calculation error of data status similarity can be effectively reduced according to cosine similarity. This study completes the status clustering according to the status types of resource data, and then completes the statistics of the status of educational assistance resource data by constructing the status matrix.

2.1 Selection of English teaching assistance resource data in the MOOCs environment

In the context of the continuous expansion and increase of teaching resources and assistance learning resources, in order to use such a huge resource information effectively, it is necessary to obtain its effective state. First of all, it is necessary to collect learning data of different English teaching resources based on different content, and effectively classify each type of English teaching resources, so as to provide more accurate data for the integration and statistics of the status of English teaching resources. Therefore, this paper first uses SVM to screen the data of assistance resources. Advantage of this algorithm is that the required sample data can be collected from a large amount of data under limited conditions (Luo, 2020).

The location of online English teaching assistance resources can be regarded as a large area plane containing a variety of English teaching assistance resources data, which can be divided by the linear separable condition. In this study, the plane containing the data of English teaching assistance resources is expressed in the following form:

$$P(X) = a \times x + b \tag{1}$$

where a represents the location point of English teaching assistance resource data, and b represents the value of bias vector dividing the data plane of English teaching assistance resource.

In this study, the maximum interval between the boundaries is always maintained by dividing the data plane of English teaching assistance resources, so as to minimise the interference encountered in data collection of English teaching assistance resources (Wu, 2020). At this point, the data in the two planes of English teaching assistance resource data after division are as follows:

$$ax_i + b > 0,\ y = 1 \tag{2}$$

$$ax_i + b < 0,\ y = -1 \tag{3}$$

Among them, formula (2) and formula (3) represent linearly separable data of English teaching assistance resources, which can be adjusted by adjusting different resource data positions and offset quantities.

In the process of dividing the data of resources, the sample data of resources obtained from the two hyperplanes are regarded as support vectors in this study. At this time, the classification interval between the English teaching assistance resource data in the two planes can be expressed as:

$$y\left[(a \times x) + b\right] - 1 \geq 0 \tag{4}$$

where margin $= \frac{2}{a}$. At this time, the classification interval between the two planes is the smallest, thus it can be determined that this plane is suitable for data collection in English teaching assistance resource data collection. After determining the dataset of English teaching assistance resources contained in the plane, the process is as follows:

$$R = \sum_{i=1}^{n} \varepsilon_i y v_i \tag{5}$$

where R represents the result value of the collected English teaching assistance resource data, ε_i represents the interference factor in the collection, and v_i represents the proportion coefficient of the collected data.

In above process, with the help of SVM t, this study regards the location of online English teaching assistance resources as a large area plane containing a variety of English teaching assistance resources data, and divides the plane through the linear separable conditions, determine the plane of the minimum interval between English teaching resources data, and thus complete data screening of English teaching resources.

2.2 Preprocesses the data status of English teaching assistance resources

There are many interference noises in the data of English teaching assistance resources obtained by the above process, and the similarity between the data is high, which is easy to affect the statistical effect in the statistics of the real-time status of English teaching assistance resources (Yu, 2018). Therefore, this paper preprocesses the data of English teaching assistance resources in the above planes to provide better data support for subsequent research.

In this section, noise reduction is performed on the data first, extracting resource data through standardised processing:

$$B = \frac{T - i}{q} \tag{6}$$

where T is mean value of English teaching assistance resource data, i is standard deviation value of English teaching assistance resource data, and q is the corresponding basic value of English teaching assistance resource data.

The data that does not conform to standardisation are eliminated through standardised processing, so as to complete the data denoising processing:

$$p(i) = \frac{a(i) - b(i)}{\max\left[\left(a(i), b(i)\right)\right]} \tag{7}$$

where $p(i)$ represents the noise reduction result of English teaching assistance resource data, $a(i)$ represents the noise reduction factor of the data, and $b(i)$ represents the best value of the data.

Further processing is carried out according to the results of the data of English teaching assistance resources after noise reduction. In this process, the similarity between the data of English teaching assistance resources should be determined first, and the data with high similarity should be deleted (Tan, 2020). In this paper, association rules are used to determine the relationship between the data of two kinds of resources. First, the probability of occurrence of similarity between data of auxiliary teaching resources is:

$$G(A \mid B) = \frac{G(AB)}{G(B)} \tag{8}$$

where $G(AB)$ represents the support degree of English teaching assistance resource data, and $G(B)$ represents the confidence degree of English teaching assistance resource data.

Based on the occurrence probability in the data of English teaching assistance resources, the relationship between the support degree and confidence degree of the data is converted into probability (Wang, 2017), namely:

$$conf(A \mid B) = G(A \mid B) = \frac{G(AB)}{P(x)} \tag{9}$$

where $conf(A \mid B)$ represents the number of occurrences of similar points, and $\frac{G(AB)}{P(x)}$ represents the total number of occurrences of similar points.

In the process of intelligent statistics, in addition to the support and confidence of English teaching assistance resource data, it is also an important step to deeply dig the frequent items data. The data sparsity of English teaching resources directly affects the similarity between the data. Therefore, in view of the sparse problem in the data of English teaching assistance resources, the sparse matrix is set to process the data of English teaching assistance resources (Zhang, 2017). The sparsity value of the preprocessing matrix is as follows:

$$S = \frac{F}{F \times I} \tag{10}$$

In formula (10), S is the matrix sparsity of English teaching assistance resource data, F is number of resources, and I is the similarity value of resource data.

The above process determines the similarity points and the degree of sparsity between the data of English teaching assistance resources. When computing resource data similarity, this paper introduces cosine similarity to complete the final processing of data similarity of English teaching assistance resources. In the preprocessing of English teaching assistance resources data, the similarity between the data is regarded as an

N-dimensional space, and each row of data is regarded as a similar row vector (Callies et al., 2018). Assuming that the similarity of data of English teaching assistance resources is expressed as:

$$Same(e, d) = \cos(e, d) = \frac{e, d}{\|e, d\|} \tag{11}$$

where *e* and *d* respectively represent similar scoring vectors of different English teaching assistance resources.

According to the similarity degree of English teaching assistance resource data determined above, data with a higher degree are regarded as redundant data and deleted to realise the preprocessing of English teaching assistance resource data, namely:

$$K(X) = \sum_{i=1}^{n} -f(x_i)\log_2 r(x_i) \tag{12}$$

where $f(x_i)$ represents the maximum amount of redundant data deletion of English teaching assistance resource data, and $r(x_i)$ represents the occurrence frequency of similarity points in sample English teaching assistance resource data.

In the data preprocessing of teaching auxiliary resources, after standardising the English teaching assistance resources data, the data which do not conform to the standardisation are removed to complete the noise reduction of English teaching auxiliary resources data. With the aid of association rules, the relationship between English teaching resource data is determined, and the sparse matrix is set to process resources data. Based on this, with the help of cosine similarity, the final similarity processing of English teaching assistance resource data is completed, which lays a foundation for the follow-up statistics of real-time status of English teaching assistance resources.

2.3 Intelligent statistics of the real-time status of English teaching resources

This study uses the data of teaching assistance resources to make statistics on the real-time status of the resources. From the perspective of user needs, English teaching assistance resources need to meet the needs of users to learn anytime and anywhere. Therefore, according to the status type of English teaching assistance resource data, its status is first processed by clustering. Clustering processing is essentially a process of integrating the data status of English teaching assistance resources into a whole.

The distance between data status of different English teaching assistance resources is set and represented as follows:

$$it\ as: dis = \left(\sum_{i=1}^{n} |d_{ik} - d_{jk}|\right)^2 \tag{13}$$

where d_{ik} and d_{jk} represent the data status types of resources.

In order to judge the actual data amount of teaching auxiliary resources, the Manhattan distance between data status is calculated (Beiler, 2019):

$$d_{ij} = \left(\sum_{i=1}^{n} |d_{ik} - d_{jk}|\right) \tag{14}$$

The distance between data status of English teaching resources is one of the key factors reflecting the effect of statistical control. When the distance between the data status of

different English teaching AIDS resources is set within a reasonable range, it will be more convenient to carry out statistics and control.

In the process of statistics on data resource status, in view of user's satisfaction with the current states of resource status and analysis, if the status of English teaching resources data is favoured by the user, it is the need to consider the status of the user evaluation weight and the preference degree of users to the status of English teaching resources. It is expressed as:

$$L = \frac{|q_i|}{h_i} \sum_{i=1}^{n} \left(w_i d_{ij} \right) \tag{15}$$

where h_i represents the data status space of English teaching assistance resources, q_i represents the current status collection of English teaching assistance resources, and w_i represents the user's integration of the current data status of English teaching assistance resources.

According to the determined Manhattan distance between the data status of English teaching assistance resources and the weight of users' satisfaction with the current status, the statistics of data status of English teaching assistance resources is completed. The statistics set of the status of English teaching assistance resource data is set as follows:

$$D_i = \{d_1, d_2, \ldots, d_n\} \tag{16}$$

Matrix judgement is carried out to the components of the data status of assistance resources in the above formula to determine whether the data status meets the statistical standards, namely:

$$\rho_i = \frac{\sum_{i=1}^{n} D_i}{\sum_{i=1}^{n} n_i} \tag{17}$$

After determining whether the data status of English teaching assistance resources meets the statistical standards, the statistical matrix of teaching assistance resources is designed to complete the statistics of the data status:

$$\begin{Bmatrix} n_{11} n_{12} n_{1n} \\ n_{21} n_{22} n_{2n} \\ \cdots \\ n_{n1} n_{n2} n_{nn} \end{Bmatrix} \times [t\rho_i] \delta = \{C_{mi}\} \tag{18}$$

where C_{mi} represents the statistical factor, t represents the time required for statistics, and δ represents the statistical integrity of the status of English teaching assistance resources.

Since there is a certain deviation in the statistics of the data status of the English teaching assistance resources, it is necessary to correct the obtained status of the English teaching assistance resources through the fuzzy function, and then the following can be obtained:

$$J(A, B) = \mu \sum_{i=1}^{n} C_{mi} + \theta \sum_{j=1}^{m} (1 - t\rho_i) \sigma \tag{19}$$

where μ represents the correction factor, θ represents the equalisation degree of English teaching assistance resources, and σ represents the punishment regularisation term.

Therefore, after screening the data of teaching auxiliary resources, this study determines the relationship between the data of English teaching assistance resources according to association rules, and determines the similarity between the data of English teaching assistance resources by setting sparse matrix. Then, the redundant data are eliminated effectively according to the cosine similarity so as to reduce the statistical error of the data status. Finally, in order to improve the statistical quality of the data, the status quantity of the preprocessed data is clustered according to the status types, and the status statistics matrix is obtained through clustering.

3 Experimental verification

3.1 Experimental scheme

The experiment was carried out with the local area network of a college campus as an example. The data status of the English teaching assistance resource module in the campus was collected, and the collection cycle was set as one week. A total of 100,000 data were collected, among which 10,000 data were used for sample training as the final status statistics. The methods used in the experiment were the method of this paper, the statistics method of online education course status based on big data platform [method of Luo et al. (2018)] and the statistical management method of English teaching resource status based on packet analysis [method of Cao et al. (2019)].

3.2 Experiment indicators

In the experiment, the adjusted rand index (ARI) and the calculation error of English teaching assistance resources were used as the experimental indicators. Among them:

1 The calculation error E of English teaching assistance resources is expressed in the form of percentage, and the calculation method is as follows:

$$E = \frac{N_1}{N_2} \times 100\% \tag{20}$$

where N_1 is the number of error cluster statistics, and N_2 is the total amount of English teaching assistance resources.

2 The adjusted RAND coefficient is mainly used to evaluate the performance of the clustering statistical results, which can reflect the overlap degree of the two classifications. The value range of ARI is [0, 1]. The larger the value is, the higher the statistical quality of resource status is:

$$ARI = \frac{RI - E(RI)}{\max RI - E(RI)} \tag{21}$$

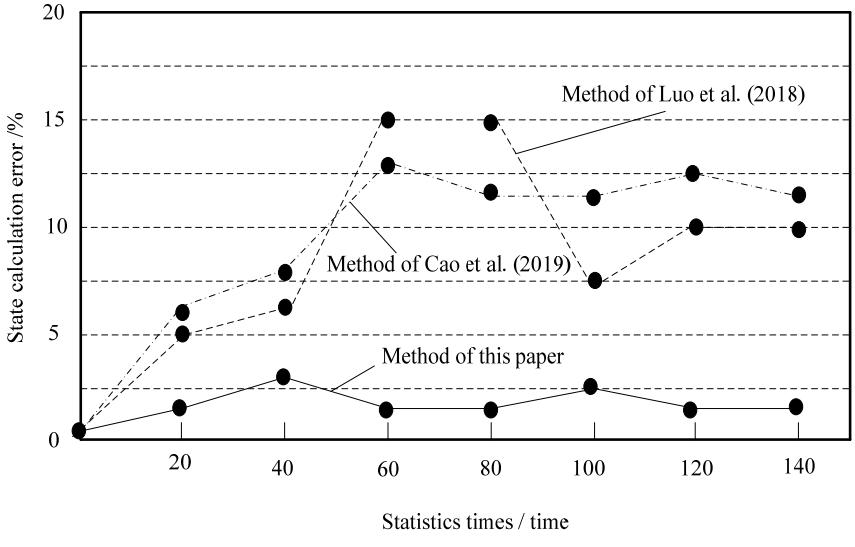
where RI represents the average proportion of correct decisions, $E(RI)$ represents the RAND index, and maxRI represents the largest proportion of correct decisions.

3.3 Results and analysis

3.3.1 Calculation error analysis of English teaching assistance resources

The calculation error of English teaching assistance resources is a key indicator to measure the effectiveness of statistical methods. The calculation error results of different methods of English teaching assistance resources are shown in Figure 1.

Figure 1 Comparison of errors in status calculation of English teaching assistance resources



Through analysis of the experimental results in Figure 1, it can be seen that in the same experimental environment, there is certain difference in the calculation errors of the data status of English teaching assistance resources using different methods. Among them, the calculation error of the method in this paper is always less than 4%, while the calculation error of the other two methods is always higher than that of the method in this paper. The reason for this result is that the method in this paper determines the relationship between resource data by association rules, sets a sparse matrix to process resource data, and completes the preprocessing of British data through cosine similarity, thus reducing the calculation error of the method in this paper.

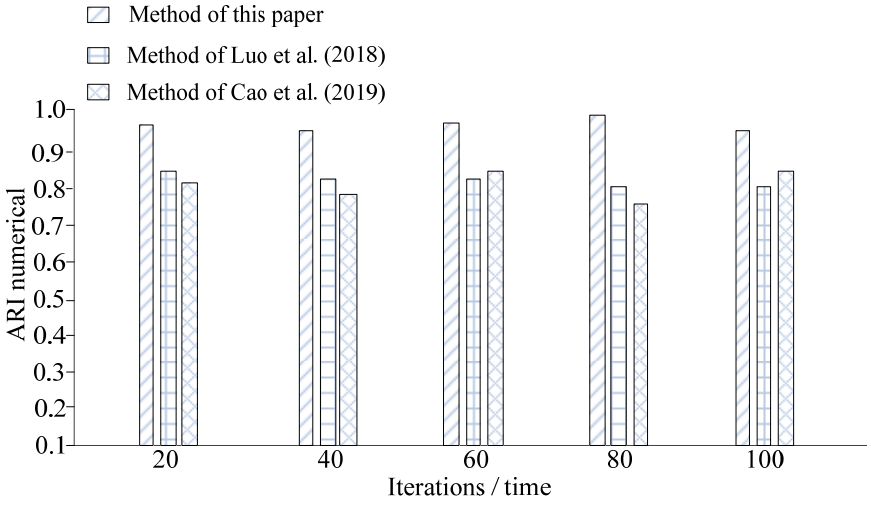
3.3.2 ARI numerical analysis

ARI value in the statistical process of resources is the key factor affecting the statistical effect. The ARI numerical results of different statistical processes are shown in Figure 2.

As can be seen from the data analysis in Figure 2, with the constant change of the number of iterations, the ARI values of the status statistics process using the method in this paper and the two traditional methods also change accordingly. Among them, the ARI value of this method is always higher than 0.9, while the ARI value of the other two methods is within a reasonable range, but always lower than that of this method. This is because the method in this paper performs effective clustering of the obtained status data

and corrects the results by setting the fuzzy correction function, thus improving the clustering statistical performance of the method in this paper.

Figure 2 Comparison of ARI values of different methods (see online version for colours)



4 Conclusions

1 Aiming at the problem that the characteristic calculation error of the status data in the traditional method is relatively high, this paper designs a real-time intelligent statistical method for the status of English teaching assistance resources under the MOOCs perspective. After the classification and screening of English teaching assistance resource data, the calculation error of data status similarity degree is reduced through data preprocessing. Then, the statistical matrix is obtained by clustering the type of resource data status, so as to complete the statistics of the status of teaching assistance resource data.

2 This method has the following advantages: the calculation error of the data status of English teaching assistance resources is less than 4%, and the ARI index is always higher than 0.9, which prove that it can effectively realise effective statistics of the real-time status of English teaching assistance resources.

3 In the following stages, further research will be carried out from the perspective of improving the timeliness of statistics.

References

Beiler, I.R. (2019) 'Negotiating multilingual resources in English writing instruction for recent immigrants to Norway', *TESOL Quarterly*, Vol. 54, No. 2, pp.535–540.

Callies, M., Simsek, T. and Friginal, E. (2018) 'Corpus linguistics for english teachers: new tools, online resources, and classroom activities', *International Journal of Corpus Linguistics*, Vol. 25, No. 2, pp.230–234.

Cao, X., Zhang, P. and Tan, Y. (2019) 'The design and implementation of the usage statistics system of electronic resource based on packet analysis', *Journal of the Library Science of Sichuan*, Vol. 12, No. 3, pp.18–21.

De Bruin, T., Kober, J. and Tuyls, K. (2018) 'Integrating state representation learning into deep reinforcement learning', *IEEE Robotics and Automation Letters*, Vol. 3, No. 3, pp.1394–1401.

Luo, F. (2020) 'Research on Chinese-English translation automatic calibration method based on machine learning', *Automation and Instrumentation*, Vol. 15, No. 8, pp.146–149.

Luo, J., Zeng, D., Pan, Z. and Liu, B. (2018) 'Real-time statistics of access of teaching resource sharing system based on big data platform', *Intelligent Computer and Applications*, Vol. 8, No. 3, pp.148–153+157.

Park, S. and Mcleod, K. (2018) 'Multimedia open educational resources in Mathematics for high school students with learning disabilities', *Journal of Computers in Mathematics and Science Teaching*, Vol. 37, No. 2, pp.131–153.

Qing, F. (2020) 'Evaluation of the development of basic internet resources based on multidimensional scaling', *Statistics and Application*, Vol. 9, No. 5, pp.754–761.

Tan, M. (2020) 'Research on English teaching system based on artificial intelligence and WBIETS wireless network system', *EURASIP Journal on Wireless Communications and Networking*, Vol. 15, No. 1, pp.14–20.

Tang, J. and Yuan, X. (2018) 'The construction and application of digital resources in rural primary school English teaching and learning', *Social Networking*, Vol. 7, No. 2, pp.89–96.

Wang, L. (2017) 'A design of SPOC-based blended college English teaching model', *Heilongjiang Researches on Higher Education*, Vol. 14, No. 10, pp.168–170.

Wu, X. (2020) 'Study on English-Chinese translation system based on human-computer interaction and feature extraction', *Microcomputer Applications*, Vol. 36, No. 4, pp.126–128.

Yu, Q. (2018) 'Design of an interactive English-Chinese translation system based on feature extraction algorithm', *Modern Electronic Technology*, Vol. 41, No. 4, pp.161–163+168.

Zhang, R. (2017) 'On strategy to optimize teaching model of college English in higher vocational colleges based on project-based teaching method', *Vocational and Technical Education*, Vol. 38, No. 14, pp.49–51.