International Journal of Continuing Engineering Education and Life-Long Learning

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

Mining method of students' learning behaviour characteristics in online classroom of colleges and universities based on dense clustering method

Jing Wen

DOI: 10.1504/IJCEELL.2024.10057193

Article History:

Received:	05 August 2021
Last revised:	07 February 2022
Accepted:	09 February 2022
Published online:	03 December 2023

Mining method of students' learning behaviour characteristics in online classroom of colleges and universities based on dense clustering method

Jing Wen

College of Marxism (Law),
Chizhou University,
Chizhou 247000, China
Email: jingwen@mls.sinanet.com

Abstract: Aiming at the problem of poor mining performance of traditional learning behaviour characteristics mining methods, this paper proposes a new learning behaviour characteristics mining method based on dense clustering method. Firstly, students' learning behaviour data from the online classroom software platform is collected. Secondly, the information synthesis parameter processing algorithm combined with ant colony algorithm is used to calculate the sensitivity of characteristics and modify the students' learning behaviour characteristics data. Finally, the dense density clustering method is used to mine the learning behaviour characteristics, and the characteristics mining results are obtained. The simulation results show that this research effectively improves the accuracy and efficiency of characteristics mining, and the accuracy of characteristics matching is always maintained at more than 90%.

Keywords: dense clustering method; learning behaviour; characteristics mining; data revision.

Reference to this paper should be made as follows: Wen, J. (2024) 'Mining method of students' learning behaviour characteristics in online classroom of colleges and universities based on dense clustering method', *Int. J. Continuing Engineering Education and Life-Long Learning*, Vol. 34, No. 1, pp.77–87.

Biographical notes: Jing Wen received her Master's in Labour Economics from Zhongnan University of Economics and Law in 2006. Currently, she is an Associate Professor in the College of Marxism (Law) of Chizhou University. Her research interests include education and management.

1 Introduction

At present, distance learning based on online classroom has been vigorously promoted in the field of education and teaching. Major universities, junior and senior high schools, even primary schools and early childhood can contact online learning, which also verifies the universality of online learning application (Shen, 2019). Although online learning has brought some convenience to teaching, especially in special periods, with the continuous advancement of online learning process, teachers and learners are in a state of separation, and teachers are unable to supervise the learning state, which will lead to some students'

relaxed learning state and unable to complete the online teaching tasks arranged by teachers in time, thus affecting the teaching quality (Wang, 2019). Therefore, the university needs to study students' online learning behaviour, analyse students' online learning characteristics, strengthen the monitoring of all links of online teaching, improve teaching quality, cultivate talents who can create more value for the society, and better serve the development of colleges and universities and local economy (Hu et al., 2020).

Zou et al. (2020) proposed a method of mining students' behaviours in online classroom of universities and colleges based on CNN deep learning model. This method uses CNN learning model to detect students' online learning behaviours, classifies different types of learning behaviours according to the detection results, and finally extracts the characteristics of students' learning behaviours in depth. The results show that this method has good application effect in the detection and mining of students' online learning behaviours, but it repeatedly mines some characteristics data, which lowers both the efficiency and accuracy of characteristics mining. Han et al. (2020) proposed a method of mining Internet behaviour characteristics based on big data analysis. Firstly, crawler technology is used to extract users' web browsing information to form a set of category tags. Then, users' login behaviours are identified, and the tag set is combined with the result of behaviour recognition to form a complete characteristic set. Finally, the user's online behaviours in the set are described, and the description results are processed into characteristics set to realise the extraction and mining of user's online behaviour preferences and characteristics. The results show that the characteristics results obtained by this method are more comprehensive, but it is precisely because of the more characteristics obtained that the characteristics' matching effect is poor, and the characteristics mining takes a long time. Le et al. (2019) proposes to mine learning behaviour based on clustering algorithm. This method takes EDX platform as the data platform, extracts educational big data from it, analyses the basic characteristics of students in the form of data screening and cleaning, and obtains the basic information of students such as gender, age and major. Then, based on the information processing results, the learning behaviour characteristics are obtained by K-means clustering algorithm. The results show that the characteristics mining results obtained by this method can provide data basis for teaching and research, but the overall time-consuming of the mining results is long because the characteristics vectors are not filtered.

Aiming at many problems of the above methods, this paper studies the mining of learning behaviour characteristics based on dense clustering method. The main technical route of excavation is as follows:

1 Collect the learning behaviour characteristic data of students on the online learning platform, including students' login time, downloads of learning resources, examination scores, teachers' task assignment, teaching evaluation and so on.

2 Based on the data collection results, the characteristics sensitivity is calculated by the information synthesis parameter processing algorithm and ant colony algorithm, and the students' learning behaviour characteristics data are modified according to the characteristics sensitivity.

3 Based on the modified characteristics data, the learning behaviour characteristics are mined hierarchically by the dense clustering method to obtain the characteristics mining results.

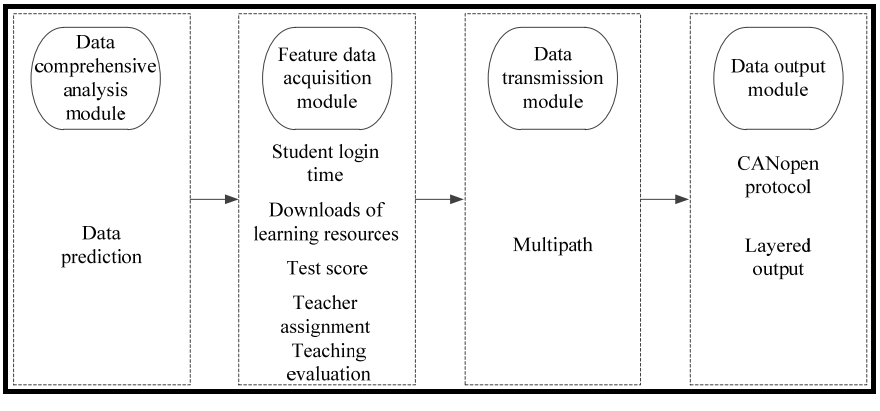
4 In the experimental stage, the accuracy of characteristics mining, the effect of characteristics matching and the efficiency of characteristics mining are selected as the experimental indicators, and the specific experimental results and conclusions are obtained by comparing this method with the traditional methods.

2 Learning behaviour characteristics mining method

2.1 Acquisition of learning behaviour characteristics

In order to better ensure the efficient management of data about students' learning behaviours characteristics in online classroom of universities and colleges, a software platform is designed to collect the learning behaviours characteristic data, and the collected data are pre-processed such as redundant data elimination, duplicate data filtering and so on. The schematic diagram of learning behaviours characteristics acquisition platform is shown in Figure 1.

Figure 1 Schematic diagram of the learning behaviour characteristics collection platform



According to Figure 1, the functional modules of the platform are as follows:

2.1.1 Data comprehensive analysis module

The first layer module of the platform is the data comprehensive analysis module, which is used to analyse the characteristics of all kinds of student information resources on the first layer link and determine the category through the corresponding threshold of characteristic quantity, so as to achieve the effect of data pre-processing of students' learning behaviours in online classroom (Cui et al., 2020).

2.1.2 Characteristics data acquisition module

With the help of cloud computing technology, the data of students' learning behaviours in online classroom are collected, including students' login time, downloads of learning resources, test scores, teachers' task arrangement, teaching evaluation and others.

2.1.3 Data transmission module

The module mainly transmits the dataset in the characteristics data acquisition module to the data output module (Shen et al., 2019). With multiple transmission paths, the platform has high transmission efficiency.

2.1.4 Data output module

Due to the large amount of data in the platform and the real-time and reliability requirements of data transmission, CANopen protocol is used to define the content of data output, and then hierarchical output of data is carried out in the form of multi-path transmission (Mao and Zhou et al., 2019).

Because the learning behaviours characteristics collection platform contains a large number of duplicate data, these data are restrictive for characteristics mining, which is likely to lead to the problem of heavy workload and long mining time. Because there are many types of data in the platform, the pressure of complexity characteristics mining and management is greater (He et al., 2020). Therefore, after completing the design of learning behaviours characteristics collection platform, the core parameter processing method is used to eliminate the repeated data. The information core parameter mainly refers to the association of similar characteristics elements among multiple information, and the degree of association between characteristics elements is mainly determined by the pheromone influence coefficient between different information. The relatively independent characteristics data are reserved to provide data basis for further mining of students' learning behaviours characteristics.

2.2 Characteristic data correction

The influence coefficient between characteristics must be handled well, and the sensitivity of characteristics shall be calculated by using information synthesis parameter processing algorithm combined with ant colony algorithm. The specific formula is as follows:

$$S_{ij} = \sqrt{\sum_{k=1}^{n} x_{ik} - x_{jk}} \tag{1}$$

where x_{ik} represents the fusion difference capability of different characteristics; x_{ik} represents the transfer coefficient of the data characteristics; n is the amount of characteristics data; k is the weight of the data mining coefficient of the difference characteristics.

At the same time, considering that the characteristic data of students' learning behaviours is iteratively mining, its action function can be expressed as:

$$J_k = u^2 \left[P\left(x_{jk}\right) + Q\left(x_{jk}\right) \right] \tag{2}$$

where u^2 represents the weighted value corresponding to the data characteristics transfer coefficient; P and Q represent the correlation coefficient between a certain characteristics and other characteristics. The characteristic data are classified according to the correlation coefficient, and different datasets are obtained (Shao et al., 2020):

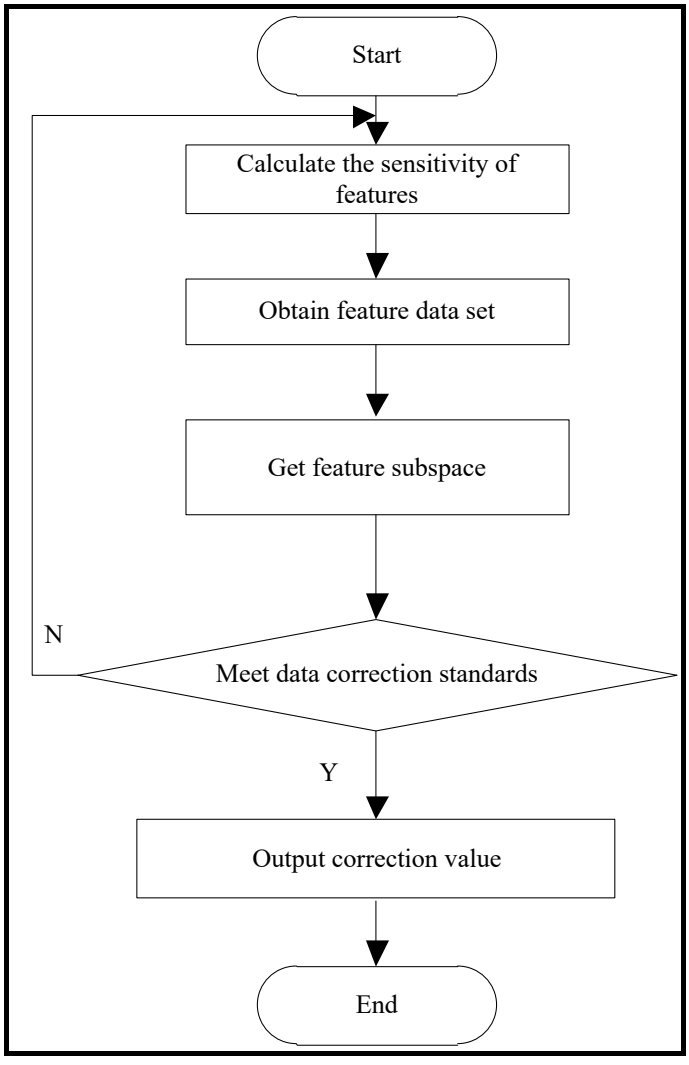
$$\mu = \left\{ f^T x_1, f^T x_2, ..., f^T x_n \right\} \tag{3}$$

where T represents the characteristics subspace.

In the process of data correction, the data correction standard will be adjusted by constantly changing the characteristics difference coefficient, so as to avoid the problem of misjudgement of local characteristics in data correction and reduce the prediction error of characteristic types (Wang et al., 2019; Yin and Wu, 2020).

The data correction process is shown in Figure 2.

Figure 2 Data correction process



2.3 Learning behaviours characteristics mining based on dense clustering

After data correction, the purpose of improving the accuracy of characteristics mining can be achieved to a certain extent, but to achieve the accurate characteristics matching, further research on the correlation between different types of characteristics is needed

(Zhang et al., 2019). This paper uses the dense clustering algorithm to mine learning behaviours characteristics. This algorithm is a statistical analysis method. It is based on the similarity between data and considers that the similarity coefficient between characteristics data in the same clustering mode is higher than the similarity of characteristics data under different clustering modes. Based on this theory, the dense density clustering algorithm is used for behaviours characteristics mining (Wu and Tian, 2019).

The density reachable relationship is adopted to derive the characteristics with the greatest similarity, and a sample set $E = \{e_1, e_2, ..., e_n\}$ is thus obtained, in this clustering cluster, the behaviours characteristics are layered and the process of mining is as follows:

1 Take a standardised method to standardise the learning behaviours characteristics in the clusters and reduce the impact of different dimension data on the characteristics mining results through this step (Huang et al., 2020).

2 Calculate the side length of the high-density grid unit, generate a dynamic grid according to the calculation result, and then use the mobile grid technology to map the characteristic data and record the relevant information of the characteristic data.

3 In characteristics mining, the boundary removal points are processed to prevent the loss of effective characteristics.

4 Calculate the grid cell density to obtain the density threshold and the core density threshold. The calculation formulas are:

$$v = \left(\frac{1}{\sqrt{g_i - 1}} + \frac{1}{\sqrt{g_j}} \right)^2 \tag{4}$$

$$v(t) = \exp\left(\frac{1}{2} \int_t^{\infty} \left(g_i + g_j \right) \right) du \tag{5}$$

where g_i and g_i respectively represent the lowest density and highest density of the intersecting grid.

5 Search for high-density grid cells in the intersecting grid according to the priority principle to realise the mining of learning behaviours characteristics in the grid (Ge and Jing, 2019) and repeat this operation until all characteristics are mined.

According to the implementation process of the above dense clustering algorithm, the characteristic equation of learning behaviours is obtained:

$$L_x(N) = \frac{L_s(N) + L_n(N)}{L_k(N)} \tag{6}$$

where L_s (N) represents the characteristics value; L_n (N) represents the characteristics vector; L_k (N) represents the characteristics space. A function is set to calculate the inner product of the characteristics space, which can be expressed as:

$$\phi(x) = \sigma^2 I_k \tag{7}$$

where $\phi(x)$ represents the inner product parameter; I_k represents the characteristics subspace. From equation (7), it can be seen that the relationship between different types of characteristics in the characteristics space can be expressed as:

$$\partial_i = \sum_{i=1}^{n} \varphi_i \left(S_{n-1}, S_n \right) \tag{8}$$

where S_{n-1} and S_n both represent the intimacy between characteristics subspaces.

The above equation (8) is substituted into the equation (7), and the characteristics clusters in the characteristics space are calculated:

$$\delta_i^n = \sum_{i=1}^{n} \varphi_i \times \phi(x) \tag{9}$$

Since $\phi(x)$ is a known parameter, there is a quantitative relationship between it and learning behaviours characteristics as shown in the following formula:

$$\phi(x)_k = \frac{(m_i)^2 \times \in (n)}{\chi_k} \tag{10}$$

Under the premise that the meaning of each parameter in the above formula remains unchanged, an evaluation matrix is defined:

$$G = \begin{bmatrix} g_{11} & g_{12} & g_{1n} \\ g_{21} & g_{22} & g_{2n} \\ g_{n1} & g_{n2} & g_{nm} \end{bmatrix} \tag{11}$$

According to the evaluation matrix, the learning behaviours characteristics vector is mined, and the mining result can be expressed as:

$$h(n-1) = a(k) - b(k) \tag{12}$$

where $h(n–1)$ represents the characteristics, vector obtained by mining. Continuously mine the characteristics vectors of usability learning behaviours. In the actual mining process, different learning behaviours data have different characteristics. A local single kernel function is used to mine the learning behaviours data locally and globally, so as to mine the learning behaviours characteristics of college online classroom students.

3 Simulation experiment analysis

In order to prove that this method has certain applicability, after the method design was completed, the simulation experiment analysis link was constructed. The effectiveness of the proposed characteristics mining methods was compared and tested. The test was completed by simulation. Through the numerical comparison with the traditional method, the effectiveness of the proposed method was proved.

3.1 Experimental preparation

Experimental environment: The CPU is Intel Core i7-6700HQ, the memory capacity is 16G, the graphics card is Intel HD Graphics 530, and the hard disk capacity is 512G solid

state drive. And Matlab software was used to process the data generated in the experiment to ensure the accuracy of the experimental results.

The experimental scheme: using characteristic data repetition mining probability, characteristics matching accuracy rate and characteristics mining time consuming as performance evaluation indicators, this research method is compared with (Zou et al., 2020; Han et al., 2020) method.

The probability of repeated mining of characteristics data indicates that the number of characteristics points mined for the same data is repeated with the characteristic's points mined from relevant data. In the experiment, the higher the probability of repeated mining of characteristics data, the more invalid data in the mining results, and the poor the effect of the corresponding mining method. On the contrary, the lower the probability of repeated mining of characteristics data, the more useful the data characteristics mined, and the stronger the reliability of the corresponding characteristics mining method.

3.2 Experimental data

Experimental parameters and sample source: The data in the experiment came from an online course of a certain college. In this college, 1,200 students were selected, and their log-in time, downloads of resource data and test score data in the online course were collected. The factors that affect students' online learning behaviours were analysed. The above data were classified to form different datasets, named dataset 1, dataset 2, dataset 3, dataset 4, dataset 5 and dataset 6. The specific dataset parameters are shown in Table 1.

Table 1 Experimental dataset

Parameter	*Dataset 1*	*Dataset 2*	*Dataset 3*	*Dataset 4*	*Dataset 5*	*Dataset 6*
Student login time/h	60	63	65	58	74	70
Resource downloads/MB	534	649	712	492	781	750
Student test scores/points	74.2	76.1	84.0	63.4	70.1	68.3

Notes: The student login time in the table is the time when students log in at the same time, and the student's test score is the average score.

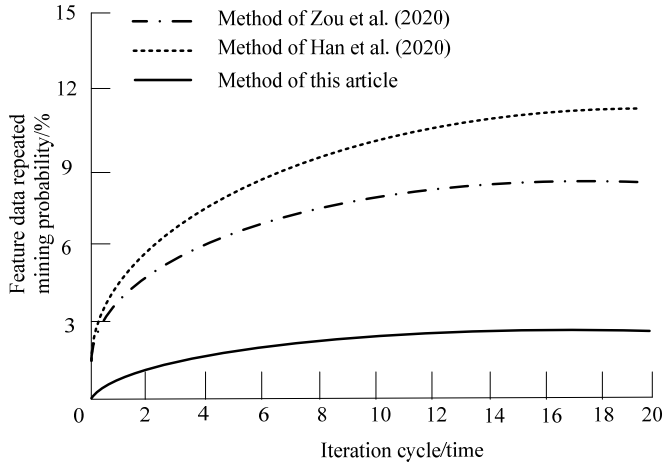
3.3 Result analysis

3.3.1 Repeated mining probability of characteristic data

Taking the repeated mining probability of characteristics data as the test index, the similarities and differences of different methods are compared. The comparison results of repeated mining probability of characteristics data are shown in Figure 3.

According to the analysis of Figure 3, in comparison, the method designed in this paper has a lower probability of repeated mining of characteristics data, which is always kept below 3%. This is because this method first collects and pre-processes the characteristics data before characteristics mining. In this process, it can filter some duplicate data, eliminate redundant data, simplify the complexity of characteristics data duplicate mining, and reduce the probability of characteristics data duplicate mining.

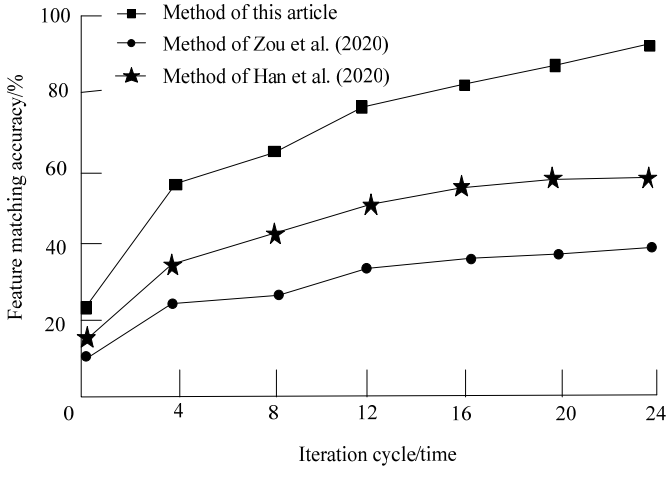
Figure 3 Comparison of probabilities of repeated characteristics data mining



3.3.2 Comparison results and analysis of characteristics matching accuracy

The comparison results of characteristics matching accuracy are shown in Figure 4.

Figure 4 Comparison of characteristics matching accuracy



The characteristics matching accuracy of Zou et al. (2020) method is always lower than 40%, and that of Han et al. (2020) method is always lower than 60%. In contrast, the highest characteristics matching accuracy of the designed method is more than 90%.

3.3.3 Comparison results and analysis of characteristics mining time

The effectiveness of the designed method is further verified, and the time-consuming characteristics mining is used as the test index to compare the similarities and differences

of different methods. The time-consuming comparison results of characteristics mining can be seen in Table 2.

Table 2 characteristics mining time-consuming/s

Dataset	*Designed method*	*Zou et al. (2020) method*	*Han et al. (2020) method*
Dataset 1	5.3	9.3	9.3
Dataset 2	4.4	9.5	9.4
Dataset 3	4.8	9.5	9.1
Dataset 4	5.0	9.6	9.7
Dataset 5	8.3	9.2	9.8
Dataset 6	8.7	9.6	9.8

From the statistical data in Table 2, the characteristics mining time of datasets 1–6 is lower than that of Zou et al. (2020) and Han et al. (2020), which shows that the designed method is more efficient.

4 Conclusions

A method of mining students' learning behaviour characteristics in online classroom of universities and colleges based on the clustering method of thick density is proposed in this paper. This method improves the data transmission efficiency through the learning behaviour characteristics acquisition platform, which is conducive to improving the efficiency of characteristics matching. The clustering algorithm of dense density is used to obtain characteristics clusters, and the characteristics of different characteristics subspaces are processed separately, which is helpful to improve the accuracy of characteristics matching. The probability of repeated characteristics data mining is lower, always kept below 3%, and the highest accuracy of characteristics matching reaches more than 90%, which shows that the method achieves the improvement of characteristics effect.

References

Cui, C., Cao, F. and Liang, J. (2020) 'Adaptive undersampling based on density peak clustering', *Pattern Recognition and Artificial Intelligence*, Vol. 33, No. 9, pp.811–819.

Ge, Y. and Jing, G. (2019) 'Human action recognition based on convolution neural network combined with multi-scale method', *Computer Engineering and Applications*, Vol. 55, No. 2, pp.100–103+136.

Han, L., Jiang, G. and Wang, H. (2020) 'Fast mining simulation of online behavioral information characteristics under big data analysis', *Computer Simulation*, Vo. 10, No. 2, pp.158–160.

He, W., Wang, L., Zhu, L. and Mao, G. (2020) 'An analysis on abnormal behavior characteristics of collaborative learning in smart classroom environment', *China Educational Technology*, Vol. 44, No. 6, pp.88–94+129.

Hu, J., Zou, M., Wen, W. and Zhou, Z. (2020) 'Research process of big-data-driven learning analysis technology', *Modern Electronics Technique*, Vo. 43 No. 18, pp.54–58.

Huang, R., Gao, B., Wang, H., Xu, J. and Du, J. (2020) 'Behavioral computing based on perception of teaching process', *E-education Research*, Vo. 41, No. 06, pp.20–26.

Le, H., Fan, Y., Jia, J. and Wang, Q. (2019) 'How do excellent MOOC learners learn-mining learning behavior patterns in MOOC', *China Educational Technology*, Vo. 34, No. 2, pp.72–79.

Mao, G. and Zhou, Y. (2019) 'Study on the computation of metacognitive engagement in learners' habitual behaviors', *E-education Research*, Vol. 40, No. 6, pp.68–75.

Shao, Y., Li, C., Zhao, Q., Cao, Y. and Tian, L. (2020) 'Analysis of classroom behavior characteristics by artificial intelligence to promote teaching reform', *Heilongjiang Animal Science and Veterinary Medicine*, Vol. 52, No. 17, pp.153–158+172–174.

Shen, G. (2019) 'Application research of big data analysis in college wisdom education', *Modern Electronics Technique*, Vo. 42, No. 4, pp.97–100.

Shen, X., Wu, J., Zhang, Y., Li, Y. and Ma, Y. (2019) 'Research on evaluation model of online learning behavior and learning effect of MOOCAP learners', *Distance Education in China*, Vol. 14, No. 7, pp.38–46+93.

Wang, M., Chen, Y. and Yang, Z. (2019) 'A study on the influence mechanism of undergraduates' learning behavior in flipped classroom-taking the course of information retrieval as an example', *Library and Information Service*, Vol. 63, No. 5, pp.119–128.

Wang, Y. (2019) 'Research status and development trend of network learning behavior', *China Adult Education*, Vo. 37, No. 14, pp.3–6.

Wu, F. and Tian, H. (2019) 'Mining meaningful features of learning behavior:research on prediction framework of learning outcomes', *Education Research*, Vo. 25, No. 6, pp.75–82.

Yin, B. and Wu, F. (2020) 'Research on method of modeling learning habits in intelligent learning system', *E-education Research*, Vol. 41, No. 4, pp.55–61.

Zhang, X., Liu, J. and Hao, L. (2019) 'Analysis of teachers' actions using feature dense computation and fusion algorithm', *Laser & Optoelectronics Progress*, Vol. 56, No. 16, pp.163–168.

Zuo, G., Su, X., Wang, H. and Wu, X. (2020) 'Research on college students' classroom behavior detection based on CNN deep learning model', *Intelligent Computer and Applications*, Vo. 10, No. 2, pp.158–160.