Analysis of Students' Learning Behavior Based on Clustering Algorithms

Yangyijie Chen

The School of Film and Television Art, Hunan Mass Media Vocational and Technical College, Changsha, Hunan 410100, China, chenyangyijie@xmphdssss.cn

Corresponding author: Yangyijie Chen, chenyangyijie@xmphdssss.cn

Abstract. This article is based on a clustering algorithm framework and conducts experimental verification on a student learning behaviour dataset in a computer-aided teaching environment. We have developed a training and testing algorithm framework specifically designed for analyzing student behaviour data, with the aim of exploring student behaviour patterns through clustering algorithms. During the training process, we utilized neural network convolution testing to evaluate the stability and effectiveness of the constructed model. In the analysis of experimental results, we paid special attention to the stability of the model and compared the accuracy and recall of neural time networks (networks used to capture time series data) and classical clustering algorithms in identifying student behaviour patterns. The experimental results show that the algorithm model proposed in this paper exhibits more significant advantages in identifying student behaviour recall compared to traditional clustering methods while maintaining efficient computational speed.

Keywords: Computer-Assisted Instruction; Clustering Algorithm; Learning Behavior; Individualized Learning
DOI: https://doi.org/10.14733/cadaps.2025.S4.1-14

1 INTRODUCTION

With the advent of the big data era, educational informatization has achieved rapid development, and the construction of digital campuses in various universities is steadily advancing and gradually improving. The transformative role of big data in higher education is becoming increasingly prominent, playing its unique advantages in improving the quality of school education management and teaching quality and enhancing the evaluation of educational outcomes [1]. In this context, some scholars take students from LG University as research objects, collect data from various applications on the digital campus platform, and use the K-means algorithm and Apriori algorithm for big data mining and processing [2]. Various application systems, such as the school's one-card system, wireless campus network, academic attendance system, and student management system, have been built one after another, and these application systems generate a large amount of data every day [3]. These abundant data resources provide a solid foundation for big data mining in education. Based on the clustering results, a detailed analysis was conducted on these different types of students, and problems such as low economic ability among students, long online time, and low book borrowing quantity were identified. We used principal component analysis to construct comprehensive quality indicators for students based on their academic course grades and practical innovation activity scores [4]. Through the Apriori association rule algorithm, the school behaviour characteristics of students and their comprehensive qualities are mined to obtain the correlation between student behaviour patterns and their comprehensive qualities. To analyze the relationship between college students' campus behaviour characteristics and learning outcomes, in order to provide empirical evidence for the education management and learning ability improvement of college students. This study provides certain data support for improving the management and decision-making efficiency of schools and enhancing the quality of teaching work. The mining results provide reference and guidance for the formulation of school management policies for students [5].

In the LMS environment, student behaviour data presents high-dimensional, massive, and complex characteristics, which require analysis techniques to have the ability to efficiently process large datasets [6]. The clustering algorithm, as an unsupervised learning method, can automatically group similar student behaviour patterns without the need to define labels or categories in advance. This provides strong support for the detailed division and in-depth analysis of student behaviour [7]. In the clustering process, the algorithm constructs clusters based on the similarity between students (such as the distance of behaviour patterns), so that students within the same cluster have highly similar behaviours, while the differences between different clusters are relatively obvious. These data are transformed into a format that can be used for clustering analysis through preprocessing steps such as data cleaning, normalization, etc. Especially in computer-aided teaching based on clustering algorithms, the analysis of students' learning behaviour not only deepens the understanding of individual differences but also provides a scientific basis for the design of personalized teaching paths [8]. The student behaviour data extracted from the LMS system includes but is not limited to login frequency, learning material access duration, forum participation, homework submission status, online test scores, etc. Use clustering algorithms such as K-means, DBSCAN, and hierarchical clustering to group student groups based on their behavioural characteristics. For example, a cluster may consist of students who frequently log in but have less interaction, suggesting that they may be more inclined towards self-directed learning but lack communication [9]. After clustering is completed, each cluster represents a specific learning behaviour pattern. By further analyzing these patterns, we can discover learning preferences, differences in learning progress, and learning difficulties within the student population. For self-directed learning students, more resources for expansion and self-assessment tools can be provided. For students who require supervision, teacher-student interaction can be strengthened, and clear learning goals and plans can be set; For students who encounter learning disabilities, timely guidance and support should be provided to help them overcome the difficulties. Based on the clustering analysis results, teachers can design personalized teaching strategies for student groups with different learning behaviour patterns [10].

With the explosive growth of online learning resources, personalized recommendations of educational resources have become a key link in improving teaching quality and learning experience [11]. In the simulation experiment, the learner population was divided multiple times by adjusting clustering parameters such as the number of clusters and distance measurement methods, and indicators such as the Silhouette Coefficient and Calinski Harabasz score (also known as variance ratio criterion) were used to evaluate the clustering effect. The research is not limited to the analysis of basic attributes such as age, gender, and professional background, but also delves into three dimensions: behavioural characteristics such as learning duration, learning path, and interaction frequency, and outcome characteristics such as test scores, homework completion status, and learning progress [12]. Using the Vector Space Model (VSM) to transform the multidimensional behavioural features of learners into vector representations in high-dimensional space, makes it possible to calculate the similarity between learners [13]. This comprehensive analysis model provides a solid foundation for accurately depicting learner profiles, enabling recommendation systems to have a more detailed understanding of each learner's learning habits and effectiveness. By calculating the distance between vectors, it is possible to quantify the behavioural differences between learners, thereby supporting the execution of clustering algorithms. Based on the clustering results, the personalized recommendation algorithm was further optimized. For different learning behaviour groups, some scholars recommend systems that can intelligently match educational resources that meet their learning characteristics and needs. Recommend more extended reading materials and self-assessment questions for learners who prefer self-directed learning, and provide regular learning plans and feedback mechanisms for learners who require supervision [14].

Traditional clustering algorithms often face many challenges when dealing with complex and high-dimensional learning behaviour data. For example, data sparsity, noise and unbalanced distribution may affect the stability of clustering results. clustering algorithm can generate samples that are close to the real data distribution by constructing the competition mechanism between the generator and discriminator, thus effectively enhancing the data set and improving the data quality. The purpose of this article is to explore the analysis method of students' learning behaviour based on a clustering algorithm and try to introduce the clustering algorithm as an auxiliary tool. Specifically, this study first collects and analyzes students' learning behaviour data in the CAI environment, including learning time, learning resource access frequency, interactive behaviour and so on. Then, these data are grouped by a clustering algorithm to identify different learning behaviour patterns. On this basis, the clustering algorithm will be used to generate more learning behaviour samples to enhance the data set and improve the clustering effect. Finally, the clustering results are deeply analyzed to explore the characteristics of different learning behaviour patterns and their influence on the teaching effect. The innovation of this study is mainly reflected in the following aspects:

(a) Combining the clustering algorithm with the clustering algorithm, it is applied to the analysis of students' learning behaviour. This combination not only makes use of the advantages of the clustering algorithm in pattern recognition but also enhances the data set through the clustering algorithm.

(b) Through clustering algorithm, different learning behaviour patterns of students can be identified more finely. These models include traditional study time and resource access frequency, which provide more information for individualized teaching.

(c) The clustering algorithm is used to generate learning behaviour samples with a similar distribution to real data, which effectively solves the challenges of traditional clustering algorithms in dealing with sparse, noisy and unbalanced data.

Through this study, I hope to answer the following key questions: First, is the clustering algorithm effective in the analysis of students' learning behaviour? Can you accurately identify different learning behaviour patterns? Secondly, as an auxiliary tool, can the clustering algorithm further improve the clustering effect? Finally, what are the implications of the identified learning behaviour patterns for teaching practice? How to formulate more effective teaching strategies according to these models?

2 RELATED WORKS

The significant differences in learning styles among different students affect their interaction with digital learning resources and learning outcomes. Moodle, As a powerful learning management system (LMS), provides valuable data sources for capturing and analyzing these learning behaviours. In the CAI environment, a core challenge is how to overcome the uneven participation and learning outcomes caused by differences in students' understanding levels. By recording students' activity logs on the Moodle platform, Wang and Liu [15] gained a deeper understanding of their study habits, interests, and potential challenges they may encounter. By analyzing students' learning styles in detail, it is possible to customize teaching strategies and resources that are more suitable for their personal preferences, thereby enhancing the learning experience and outcomes. Due to mismatched difficulty learning materials, lack of personalized guidance, or insufficient interaction, many students may lose interest and motivation in learning. Moodle logs provide data support for actual learning behaviour.

With the development of educational informatization, various learning behaviours of students have generated a large amount of data. Exploring and utilizing the value hidden in educational data through machine learning algorithms has become a research focus for scholars in order to make the data play its due role. Wang et al. [16] extracted various required features, classified positive and negative samples, and formed three tables: the high-quality graduate table, the determined career table, and the postgraduate entrance examination employment table. Based on the specific situation of graduates' destinations such as postgraduate entrance examination, employment, entrepreneurship, etc., suggestions are made for the specific implementation of their learning behaviour, providing them with personalized content that meets their needs. Starting from the management of teaching methods and teaching modes, we aim to improve our management level and provide early warning for those at risk in their studies. The emergence of massive data provides education managers with a lot of potential value, which helps to improve their management efficiency. Education managers can plan comprehensively based on students' data information. Using the grid search method to adjust the optimal parameters of the Xgboost algorithm model on three tables: high-quality graduates table, destination determination table, and postgraduate entrance examination employment table, in order to improve the overall performance of the model. And obtain the ranking of feature importance scores for the three tables, identify the factors that affect students' graduation destinations, and determine the accuracy, precision, recall, and predicted graduation destinations of the Xgboost algorithm on these three tables. Using the boosting algorithm, known algorithm, and Bayesian algorithm respectively, the performance evaluation results of accuracy, precision, and recall are compared on the three tables of high-quality graduates, determined career paths, and postgraduate entrance examination employment. A three-dimensional clustered bar chart is created to visually display the comparison results. Weng and Chiu [17] used the Xgboost model algorithm to model three tables: the high-quality graduate table, the determined destination table, and the postgraduate entrance examination employment table. The accuracy, precision, and recall performance evaluations obtained were generally better than the known model algorithm and Bayesian model algorithm.

Secondly, Xie et al. [18] proposed a hybrid learning performance warning model based on SMOTE XGBoost FM. Firstly, solve the problem of imbalanced sample categories in the dataset through SMOTE sampling. Addressing the issue of sparse data that prevents the model from capturing correlations between data. This system helps teachers to timely grasp learners' learning dynamics, detect learners' learning risks in advance, and intervene in a targeted manner. The backend of the system adopts the Diango framework, which is mainly divided into the teacher end, student end, and administrator end. Through technologies such as Bootstrap and Layui, the analysis results and prediction results of the data are displayed in a visual way. The new dataset is used as input for the factorization machine (FM), and iterative training is performed to obtain the optimal model. XGBoost is used for initial training on the sampled dataset, with feature crossover. The leaf nodes where the trained samples are located are encoded to generate high-order features, which are then merged with the original features. The experimental results showed that the accuracy of SMOTE XGBoost FM in blended learning performance warning reached 92.7%, which was 5.7% and 11.7% higher than the single XGBoost and FM models, respectively. Functional and non-functional tests were conducted on the system, and the results showed that the system designed in this paper meets expectations. Finally, a learning achievement warning system was designed and implemented.

In the CCL environment, learning behaviour analysis is key to understanding how students interact with teaching content, tools, and peers. Yuan et al. [19] used the CCL platform to record students' learning behaviour data, including but not limited to learning time, resource access types, forum contributions, and homework completion. These data provide rich materials for subsequent analysis. For example, distinguishing the behavioural characteristics of active and passive learners, or identifying key behavioural sequences that affect learning outcomes. Combining multi-level and multi-class classification algorithms (such as BS, LP, CS, KNN, and their ensemble methods) to classify complex learning behaviour data can more accurately predict students' learning performance or identify specific groups that require intervention. Based on the analysis of learning behaviour, customize personalized learning paths and resource recommendations for each student.

3 THEORETICAL BASIS AND METHOD

3.1 The Foundation and Application of Clustering Algorithm

A clustering algorithm, an unsupervised learning approach, aims to partition data objects into clusters where intra-cluster objects exhibit high similarity while inter-cluster objects display significant differences. Clustering can identify students with similar learning characteristics when analyzing students' learning behaviour.

Several clustering algorithms are commonly used, including K-means, DBSCAN, and hierarchical clustering. The K-means algorithm is popular due to its simplicity and efficiency, achieving cluster division by iteratively optimizing the average distance between cluster objects. However, it is sensitive to initial cluster centre selection and may converge to local optima. DBSCAN, a density-based algorithm, effectively handles noisy datasets and outliers, identifying clusters of arbitrary shapes.

3.2 Basic Principle and Potential Value of CLUSTERING ALGORITHM

The self-bidirectional learning method based on the clustering algorithm proposed in this article is a combination of self-learning methods and bidirectional mapping models. Train the discriminator to distinguish truth from falsehood by using real samples sampled and fake samples generated by the generator. Self-learning refers to the process of comparing the samples generated by a generator not only with real samples but also with previously generated samples during the training of a model. The purpose is to enable the trained generator to generate realistic and diverse samples while significantly improving the speed of anomaly detection. So, the generator will receive two types of learning signals from the discriminator feedback during training, which come from two different training stages. The training method for the first training phase is the same as that of the original clustering algorithm model. Through the discriminator's backpropagation training, the generator is able to generate realistic samples. In this stage, the generator receives the first learning signal from the discriminator and saves the samples generated in this stage. In another training stage, the input of the discriminator becomes the current generated sample, and the previously generated sample, and the discriminator will provide feedback information to the generator based on the difference between their outputs. To further improve the authenticity of the generated samples, the generator receives the second type of learning signal from the discriminator, which is the self-learning signal. Figure 1 shows the basic principle of the clustering algorithm.

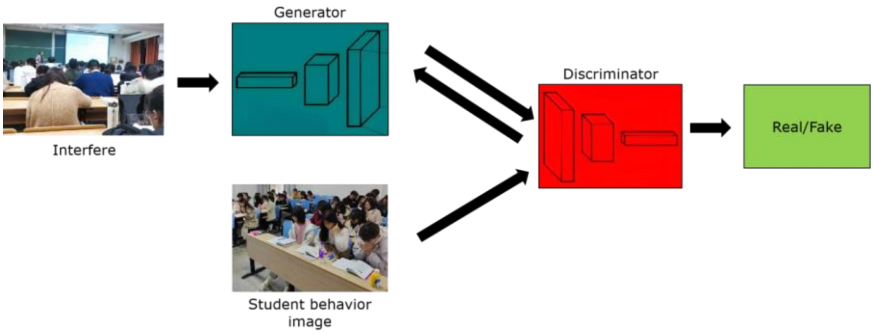


Figure 1: Basic Structure of clustering algorithm.

Based on self-learning training methods and certain training techniques, the problem of gradient vanishing and pattern collapse in the original clustering algorithm model can be solved to a certain extent. The discriminator cannot propagate useful feedback information for the generator to update parameters, so in this case, self-learning training methods can be used to train the generator. In addition, the problem of collapse usually occurs in the later stage of training the generator to generate a certain type of sample. Due to the small differences between the generated samples at this point, the discriminator will no longer provide positive feedback to the generator. Thus preventing the generated samples from collapsing into finite patterns. The root cause of the vanishing gradient in the generator is the significant difference between the generated samples and the real samples during the initial training stage. So, when the model learns to generate each batch of samples, it will undergo self-learning training in the final stage.

3.3 Analysis of Student Learning Behavior

Clustering algorithms are used to group the datasets and identify distinct learning behaviour patterns, such as active learners, passive learners, and exploratory learners. Some learning behaviour patterns may be indicative of poor learning outcomes or insufficient motivation. By identifying these patterns and analyzing their underlying causes, targeted teaching suggestions can be provided to teachers, aiding students in overcoming learning challenges and enhancing learning outcomes.

To comprehensively capture the diversity of students' learning behaviour, the author specifically invited 10 independent students as assistants, recording various learning scenes and actions. In this process, 500 rich photos were accumulated. After careful screening and consideration, this study finally selected 200 photos as the core data set. Figure 2 intuitively shows the data images of five typical students' learning behaviors which are analyzed in this article, and they form an important foundation of this study.

Figure 2: Example of students' learning behavior.

Behavior category	*describe*	*Image number*
Active Learners	Students actively participate in discussions and ask or answer questions proactively	1, 5
Passive learners	Students mainly listen or read, with less active	2, 9

	participation	
Exploratory learners	Students try different learning methods or resources and conduct experiments	3, 12
off-task behaviour	Students are distracted during the learning process, such as playing with their phones, chatting, etc	4, 15
Collaborative learners	Students collaborate with others to complete learning tasks, such as group discussions	6, 18

Table 1: Classification of student learning behaviour.

The collected student learning behaviour data was grouped using clustering algorithms, and the five typical learning behaviour patterns mentioned above were identified. These modes not only cover students' active participation and passive acceptance in the learning process, but also include diverse learning states such as exploratory, distracted, and cooperative learning. Figure 2 has visually presented example images of five typical learning behaviours, while Table 1 further refines the specific descriptions and corresponding example image numbers for each behaviour, providing a clear classification basis for subsequent analysis work. In this article, an innovative reverse residual structure is introduced to effectively reduce the information loss caused by nonlinear transformation in the training stage of shallow networks. The ingenious design of this structure enables the input picture of students' behaviour state to retain its original feature information more completely. Subsequently, these image data are propagated forward to the well-designed clustering algorithm for depth feature extraction. In the prediction layer of the clustering algorithm, different levels of candidate frames are accurately matched with the real labelled frames to output the category confidence prediction of each candidate frame and the prediction error of position offset. This series of complex processing flows, as shown in Figure 3, shows the complete realization path of students' learning behaviour state identification based on the improved clustering algorithm model. Through this process, students' learning behaviour can be identified and analyzed more accurately.

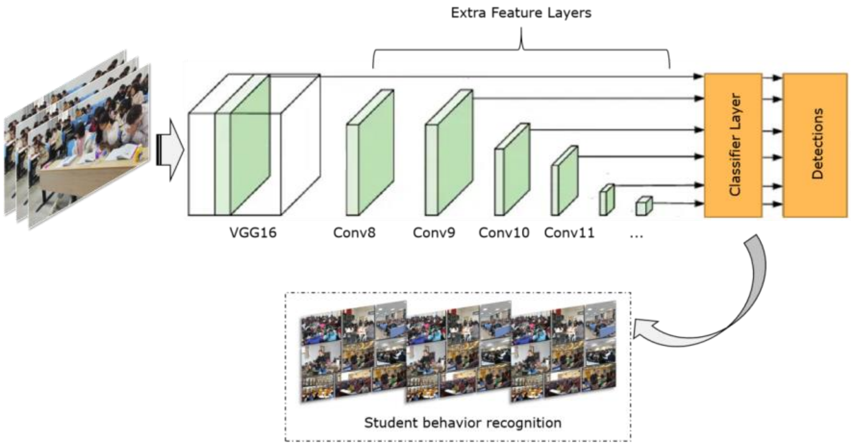


Figure 3: The realization process of students' learning behaviour state recognition.

Given an image P of size $M \times N$, the image matrix can be viewed as a one-dimensional vector with the primary order of behaviour, allowing the definition of a one-dimensional marker vector:

$$A = A_1, \ldots, A_i, \ldots, A_{M \times N} \tag{1}$$

Among them, the value range of element A_i in A is $0,1$. Evaluation of the image segmentation effect can be done by calculating the cost function of the marker A:

$$E\ A = \lambda \cdot R\ A + B \tag{2}$$

Among them:

$$R\ A = \sum\nolimits_{i=1}^{M \times N} R_j\ A_i \tag{3}$$

$$B\ A = \sum\nolimits_{i,j} B_{ij} \delta\ A_i, A_j \tag{4}$$

Let p_λ represent a probability density function, $\lambda = \left[\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_M\right]$ represent p_λ parameter vectors of p_λ, and $X_1 = \left[x_t, t = 1, 2, 3, \ldots, T_I\right]$ represent an effective feature of students' learning behaviour.

The clustering algorithm model is trained by using the selected learning behaviour data set, and samples with a similar distribution to the real data are generated. By optimizing the network structure and parameters of the generator and discriminator, the quality and diversity of generated samples are ensured. clustering algorithm-generated samples are incorporated into the original dataset to create an enhanced dataset. The loss functions of the generator and discriminator are crucial for clustering algorithm training. The generator's loss function typically aims to minimize the difference between generated and real samples, while the discriminator's loss function aims to maximize its ability to distinguish between them.

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{5}$$

Where X_{norm} represents the original data, $X_{\min}$ and $X_{\max}$ denote the minimum and maximum values in the data set, respectively.

In order to capture the complexity and temporal characteristics of learning behaviour data more comprehensively and accurately, after in-depth exploration, this study decided to adopt long-term and short-term memory networks (LSTM) as the infrastructure of the generator and discriminator. With its excellent time series modelling ability, LSTM has shown remarkable advantages in processing sequence data, especially for analyzing students' learning behaviour data.

Based on the regression principle of the clustering algorithm for detecting target position and category, the objective function is formulated as the weighted sum of confidence loss $conf$ and position loss loc:

$$L\ x, c, l, g = \frac{1}{N}\ L_{conf}\ x, c + \alpha L_{loc}\ x, l, g \tag{6}$$

Where N denotes the default number of frames matched with the real target frame, c represents the confidence of the prediction frame, and encompasses the position information of the prediction frame; g signifies the location information of the real box; a is a weight parameter set to 1 after cross-validation.

In mathematics, to determine the minimum value of a function, it is essential to first find the rate of change of the function's image. Typically, the maximum value of the function occurs where the rate of change is greatest. Locating the minimum value of the loss function requires identifying the gradient direction. The direction opposite to the gradient indicates the direction towards the minimum value of the loss function.

$$J\ \theta = \min_{\theta}\frac{1}{2}\sum_{i=1}^{m}\left(h_{\theta}\left(x^{\ i}-y^{\ i}\right)\right)^{2} \tag{7}$$

$J\ \theta$ represents the function of minimizing variance θ, which essentially denotes the value of the loss function, indicating the agreement between the model and training data.

To enhance the accuracy of students' behaviour detection, inspired by the residual learning mechanism in residual networks, the attention modulation feature is refined through residual blocks and shortcut operations, yielding the fine modulation feature $X_i^{''}$. The entire process can be formulated as follows:

$$X_i^{''} = R\ X_i^{'}\ + X_i^{''} \tag{8}$$

Here, $R\ \cdot$ represents the residual block, comprising a 3×3 convolution layer and two 1×1 convolution layers. This residual learning mechanism aims to selectively preserve the intricate details of the original features and recalibrate the modulation features, thereby augmenting the expressive power of the features.

Considering that the key points of students' wrists are either flush or positioned below the key points of their necks, the latter is designated as the coordinate origin. Subsequently, all other key points are converted into coordinates. The θ value serves as a threshold to determine whether students have raised their hands:

$$\theta = \frac{Y_h - y_n}{Y_h - y_e} \tag{9}$$

Within this context, Y_h denotes the ordinate of the wrist key point, y_n represents the ordinate of the neck key point, and y_e signifies the ordinate of the elbow key point.

In theory, the adjustment parameters have the capability to approximate any given spatial point, with the origin being just one of the more significant default points. The formula for the L_2 norm penalty term is expressed as follows:

$$\Omega\ \theta = \frac{1}{2n}\left\|W\right\|_2^2 \tag{10}$$

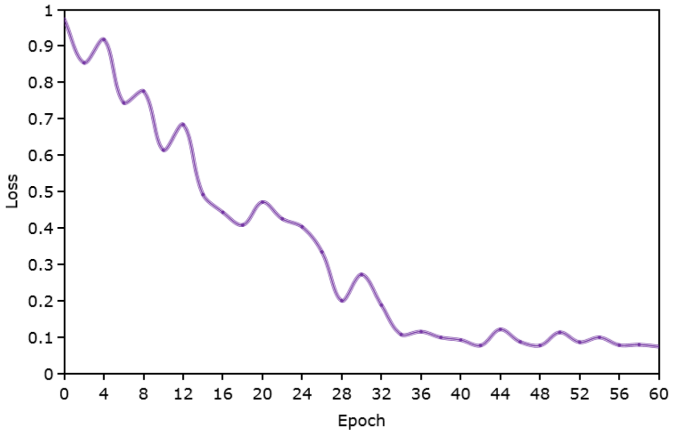
In the formula, n signifies the count of training samples, whereas $2n$ is assigned as the denominator to enhance the resolution of the partial derivative.

The trained clustering algorithm model is used to generate more learning behaviour samples to enhance the data set, and the deep features learned by the clustering algorithm are extracted for cluster analysis. The samples generated by the clustering algorithm are fused with the original data to form a richer and more comprehensive data set. The deep features learned by the clustering algorithm are fused with the original features to provide richer and more valuable information for the clustering algorithm. The clustering algorithm is applied to the fused data set again to identify more detailed and accurate learning behaviour patterns. Through in-depth analysis of the clustering results, different learning behaviour patterns are identified, and the characteristics and influencing factors of each pattern are discussed.

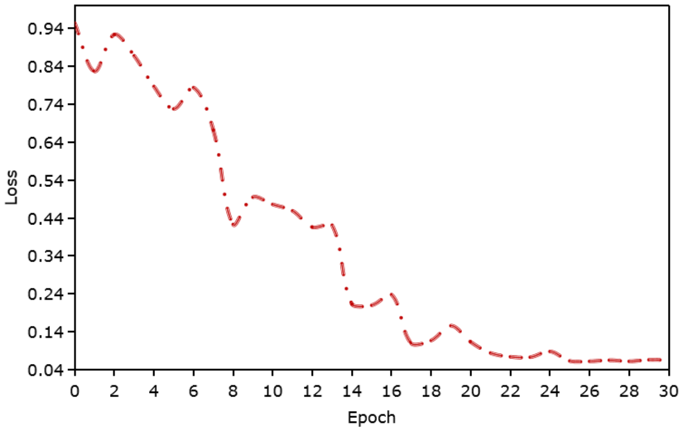
4 STUDENT LEARNING BEHAVIOR ANALYSIS TEST

This study aims to explore the analysis of students' learning behaviour using a clustering algorithm and introduce the clustering algorithm as an auxiliary tool. To verify the effectiveness of the proposed method, experiments are conducted using 200 groups of pictures as datasets, derived from records of students' learning behaviour in a CAI environment, encompassing learning time, resource access frequency, interactive behaviour, and other information. Out of these, 150 groups are utilized as training sets for the clustering algorithm and clustering algorithm model, while the remaining 50 groups serve as test sets to evaluate model performance.

During training, we focus on the trend of generator loss. As depicted in Figure 4, post-training generator losses for the two models are 0.085 and 0.071, respectively, indicating convergence. This demonstrates the clustering algorithm model's stability during training and continuous improvement in generated data quality. Figure 4(a) illustrates the variation of generator loss on the training set, showing a gradual decrease and stabilization with increased training iterations. Similarly, Figure 4(b) showcases the convergence trend in generator loss on the test set.



(a) Training set



(b) Test set

Figure 4: Loss curve of two trainings.

To further verify the performance of the proposed method, a time-consuming test was conducted. As shown in Figure 5, the time consumption of this method in dealing with students' learning behaviour is shorter than that of the traditional clustering algorithm and CNN. In the same hardware environment, the average time taken by this method to process a picture is 30 milliseconds, while the average time taken by traditional clustering algorithm and CNN methods is 60 milliseconds and 45 milliseconds respectively. This result shows that this method has obvious advantages in efficiency and can analyze and identify students' learning behaviour more quickly.

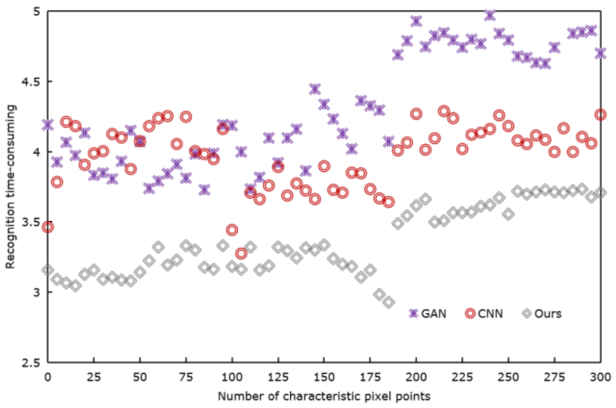


Figure 5: Time-consuming test of image recognition processing.

In addition to time-consuming testing, the accuracy and recall of the method are also assessed. As shown in Figure 6 and Figure 7, this method is obviously superior to the behaviour detection algorithms based on traditional CLUSTERING ALGORITHM and CNN in the accuracy and recall of students' learning behaviour recognition. Specifically, in Figure 7, the accuracy of this method reaches 90%, while the accuracy of traditional CLUSTERING ALGORITHM and CNN methods is 75% and 80%, respectively. The recall of our method also reaches 85%, which is much higher than that of the traditional methods of 70% and 65%. These results show that this method is more reliable in identifying students' learning behaviour.

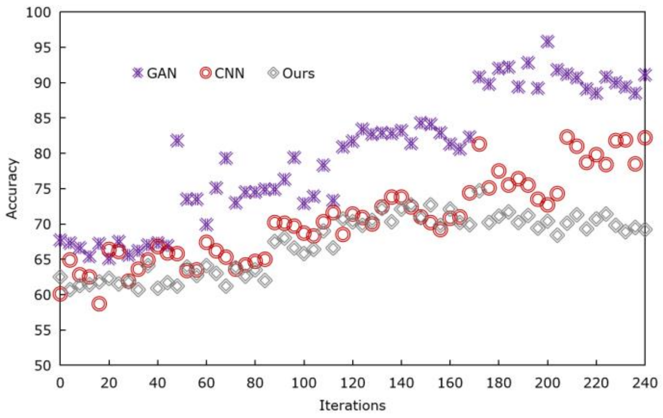


Figure 6: Accuracy of learning behavior recognition.

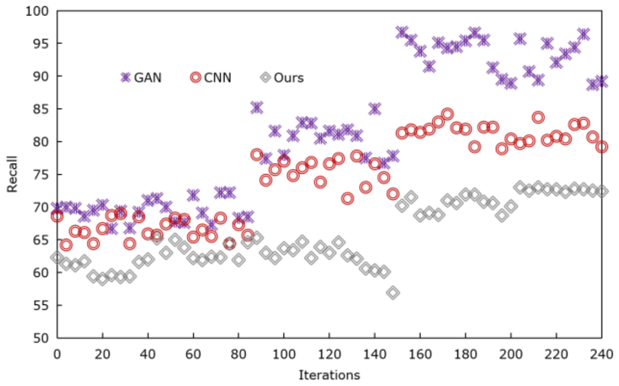


Figure 7: Recall of learning behaviour recognition.

In order to better understand the performance advantages of the proposed method, the experimental results are analyzed quantitatively. This article studies and improves the deep generative model of the Clustering algorithm (CLUSTERING ALGORITHMs), and applies it to anomaly detection tasks in time-series data. An anomaly detection system based on the improved model is designed and implemented, which can be integrated with other operation and maintenance systems and provide round-the-clock anomaly detection services for these systems. The self-bidirectional learning training method proposed in this article has to some extent solved some of the shortcomings of the original CLUSTERING ALGORITHM model, but there is still the phenomenon of unstable performance of the CLUSTERING ALGORITHM model during the training process. Moreover, the currently implemented anomaly detection system is still relatively basic in terms of functional design.

Design a more reasonable distribution similarity calculation method for the self-bidirectional learning training method based on the CLUSTERING ALGORITHM model proposed in this article, so that the generated samples can more accurately capture the differences between the two batches of generated samples during self-comparison, thereby promoting the development of the generated samples from the generator towards more realistic directions and ensuring the performance stability of the model in this training stage. Considering that the current anomaly detection system can only provide anomaly alarm and data transmission functions for operation and maintenance personnel when abnormal data occurs, the subsequent analysis of abnormal data and system fault query work that needs to be done by operation and maintenance personnel is still quite heavy. Therefore, a system fault root cause localization function can be added in future system extensions to locate the cause of abnormal data while detecting normal data.

5 CONCLUSIONS

This article deeply analyzes students' learning behaviour and creatively integrates clustering algorithms as auxiliary tools, aiming to reveal the learning patterns hidden behind the data. The experimental results not only strongly confirm the effectiveness of this innovative method, but also demonstrate its outstanding ability to identify diverse learning behaviour patterns of students. Compared with traditional methods, this method not only jumps to a new height of 90% in accuracy but also achieves significant improvements of 15% and 10% respectively compared to simple clustering algorithm and Convolutional Neural Network methods; Meanwhile, the recall rate has also reached 85%, far exceeding the traditional method's 15% to 20%. This series of impressive data undoubtedly highlights the significant advantages and potential of the method proposed in this article. It is particularly worth mentioning that this method has made a leap in efficiency, significantly reducing the time required to process student learning behaviour data and improving overall work efficiency compared to traditional analysis methods. This feature is crucial for the field of education, as it means that teachers and educational institutions can quickly obtain accurate feedback on student behaviour, thereby adjusting teaching strategies in a timely manner and achieving more personalized teaching. In summary, this article cleverly combines clustering algorithms with CLUSTERING ALGORITHM, which not only opens up new paths for analyzing student learning behaviour, but also significantly improves the accuracy, credibility, and processing speed of the analysis.

Yangyijie Chen, https://orcid.org/0009-0001-9864-0510

REFERENCES

[1] Azhary, S.-A.; Supahar, S.; Kuswanto, K.; Ikhlas, M.; Devi, I.-P.: Relationship between the behaviour of learning and student achievement in physics subject, Jurnal Pendidikan Fisika Indonesia, 16(1), 2020, 1-8. https://doi.org/10.15294/jpfi.v16i1.23096

[2] Cantabella, M.; Martínez, E.-R.; Ayuso, B.; Yáñez, A.-J.; Muñoz, A.: Analysis of student behaviour in learning management systems through a Big Data framework, Future Generation Computer Systems, 90(1), 2019, 262-272. https://doi.org/10.1016/j.future.2018.08.003

[3] Flanagan, B.; Majumdar, R.; Ogata, H.: Early-warning prediction of student performance and engagement in open book assessment by reading behaviour analysis, International Journal of Educational Technology in Higher Education, 19(1), 2022, 1-23. https://doi.org/10.1186/s41239-022-00348-4

[4] Han, J.: Research on personalized recommendation method of educational resources based on learner behaviour analysis, Journal of Circuits, Systems and Computers, 32(05), 2023, 2350079. https://doi.org/10.1142/S0218126623500792

[5] Ikawati, Y.; Rasyid, M.-U.-H.; Winarno, I.: Student behaviour analysis to predict learning styles based Felder Silverman model using ensemble tree method, EMITTER International Journal of Engineering Technology, 9(1), 2021, 92-106. https://doi.org/10.24003/emitter.v9i1.590

[6] Kimura, M.; Tateno, Y.; Matsui, A.; Nakahara, J.: Student leadership behaviour scale (slbs) in leadership education based on experiential learning in university: development, Validation, and Reliability, Information and Technology in Education and Learning, 2(1), 2022, Trans-p003. https://doi.org/10.12937/itel.2.1.Trans.p003

[7] Lechuga, C.-G.; Doroudi, S.: Three algorithms for grouping students: A bridge between personalized tutoring system data and classroom pedagogy, International Journal of Artificial Intelligence in Education, 33(4), 2023, 843-884. https://doi.org/10.1007/s40593-022-00309-y

[8] Leidy, J.: Using Computer Assisted Instruction in an ESL Language Program, IALLT Journal of Language Learning Technologies, 15(1), 2019, 13-24. https://doi.org/10.17161/iallt.v15i1.9074

[9] Li, H.; Ji, J.: Analysis of English listening obstacles based on computer-assisted instruction, Computer-Aided Design and Applications, 18(S4), 2021, 130-140. https://doi.org/10.14733/cadaps.2021.S4.130-140

[10] Lim, M.-H.; Aryadoust, V.: A scientometric review of research trends in computer-assisted language learning (1977–2020), Computer Assisted Language Learning, 35(9), 2022, 2675-2700. https://doi.org/10.1080/09588221.2021.1892768

[11] Michel, M.-C.; Thal, A.; Sparks, A.-D.; Zapanta, P.-E.: Using computer-assisted instruction to increase otolaryngology education during medical school, MedEdPORTAL, 17(1), 2021, 11065. https://doi.org/10.15766/mep_2374-8265.11065

[12] Rosali, L.-J.-D.: Effect of Computer-assisted instruction (cai) on the academic achievement in secondary physics, Open Access Library Journal, 07(5), 2020, 1-11. https://doi.org/10.4236/oalib.1106319

[13] Shin, J.; Chen, F.; Lu, C.; Bulut, O.: Analyzing students' performance in computerized formative assessments to optimize teachers' test administration decisions using deep learning frameworks, Journal of Computers in Education, 9(1), 2022, 71-91. https://doi.org/10.1007/s40692-021-00196-7
[14] Song, Y.; Zhu, M.; Sparks, J.-R.: Exploring the role of process data analysis in understanding student performance and interactive behavior in a game-based argument task, Journal of Educational Computing Research, 61(5), 2023, 1096-1120. https://doi.org/10.1177/07356331221138734
[15] Wang, J.; Liu, F.: Computer-assisted collaborative learning for enhancing students intellectual ability using machine learning techniques, Wireless Personal Communications, 127(3), 2022, 2443-2460. https://doi.org/10.1007/s11277-021-09073-z
[16] Wang, X.; Liu, Q.; Pang, H.; Tan, S.-C.; Lei, J.; Wallace, M.-P.; Li, L.: What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis, Computers & Education, 194(1), 2023, 104703. https://doi.org/10.1016/j.compedu.2022.104703
[17] Weng, X.; Chiu, T.-K.: Instructional design and learning outcomes of intelligent computer assisted language learning: Systematic review in the field, Computers and Education: Artificial Intelligence, 4(1), 2023, 100117. https://doi.org/10.1016/j.caeai.2022.100117
[18] Xie, C.; Cheung, A.-C.; Lau, W.-W.; Slavin, R.-E.: The effects of computer-assisted instruction on mathematics achievement in mainland China: a meta-analysis, International Journal of Educational Research, 102(1), 2020, 101565. https://doi.org/10.1016/j.ijer.2020.101565
[19] Yuan, X.; Bi, H.; Niu, X.: Computer-aided teaching system design method for information design specialty, Computer-Aided Design and Applications, 20(S4), 2023, 44-55. https://doi.org/10.14733/cadaps.2023.S4.44-55