International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-16

https://doi.org/10.70917/ijcisim-2025-0249
Article

A Study of Teaching Interaction Strategies and Learning Effect Enhancement Based on Reinforcement Learning in Open Education Platforms

Ling Zhu *

Yulin Open University, Yulin, Shaanxi, 719000, China; zhuling710410@sina.com

Abstract: In this paper, we first constructed an adult learner portrait and preprocessed the data for user behavior extraction. K-prototype and Pearson correlation coefficient based on the hybrid metric of Hamming distance and Euclidean distance are used to analyze the student learning behaviors and visualize them. Then in order to improve the learning effect of open education in adult education, a dialog strategy model based on user profiling and deep reinforcement learning is proposed, and comparative experiments are carried out on two datasets, ReDial and INSPIRED, to verify the effectiveness of the proposed algorithm. Finally, three classes in the open education platform in adult education are selected as research objects, and the interactive platform of teaching dialogue is applied to carry out experimental teaching, and the objective effect is analyzed by using the Spss mean difference test-t-t-test. The results show that compared with the mainstream algorithms, the method proposed in this paper is 0.453 in Recalla@50, 14.2% in Dist-3, 6.2% and 3.9% in BLEU-1 and BLEU-2. Secondly, the teaching platform established in this paper has a positive effect on the learning effect of open education.

Keywords: user profiling; k-prototype clustering; reinforcement learning; dialog interaction

1. Introduction

With the development and maturity of Internet technology, online open education has become a necessary tool for lifelong learning [1]. It is a multi-specialty, multi-directional and no-wall college, which is the best learning platform to meet the desire for further education. Nowadays, network open education has become an important way for people to improve the quality of science and culture, especially for the working population, network open education is a very effective way to improve their skills after work [2-4]. As a new teaching mode, network open education is an important means to meet people's lifelong learning needs [5]. China's online open education started in the mid-1990s and has entered a rapid growth period in recent years. At present, most of China's online open education focuses on basic education and higher education. In basic education [6], there are already more than four thousand schools with integrated campus networks, and some regions are still building educational metropolitan area networks (MANs), and there are more than three hundred online schools of a certain scale. Higher education [7], many colleges and universities have opened the online development of education courses, which has gradually come into people's lives.

But the network open education, also has its own limitations [8]. First of all, the network open education learners are mostly in front of the screen alone to learn, and offline face-to-face form, compared with the lack of direct interaction and feedback between students and teachers [9-10]. The teaching interaction is a crucial factor in learning, which leads to the teacher does not understand the problems that exist in the learning process of students, and is even more unable to make targeted teaching adjustments. Secondly, more and more Internet companies are also involved in the design and development of open education, such as NetEase Open Class, which has quite mature network

technology and strong research and development capabilities, but lack of classroom-related experience, there are certain problems in the design, the lack of teaching content support [11]. Moreover, what the enterprises value is the number of visits to the platform and the profit made in the process. Therefore, it is difficult for these enterprises to create an open education platform that can meet the requirements of lifelong learning for all. Finally, there is the problem of the education model. Currently, the mainstream web development education is simply to record offline courses and move them to the website, which is essentially a simple copy of offline teaching, resulting in the current online advantages of development education not being given full play to, while the problems of teaching interaction and poor learning results are greatly magnified [12-15]. Therefore, ideas and techniques to optimize and solve these problems are proposed for the above limitations, which are of great significance to the development of development education.

Teaching is a social activity that has a formative impact on the learner, in which interaction plays a crucial role. Năznean [16] points out that in the classroom, students learn cultural knowledge, value goals and social norms through interactions and interactions with other people, and recognize and deal with preliminary social relations in their interactions, so that they form the appropriate " social perspective". A number of studies in traditional learning environments have demonstrated the positive effects of instructional interactions on learning outcomes, and it is often regarded as an important external support factor [17-19]. For example, Alvi and Gillies [20] explored the effects of teacher-led instructional interaction on learners' learning outcomes in a traditional classroom setting through a case study, and found that instructional interaction and learning outcomes are interrelated, and that instructional interaction in the classroom can enable learners to share their ideas and strategies, improve their cognitive abilities and learning strategies, and thus promote learners' learning outcomes. Li et al [21] identified student-student interaction (SS), student-instructor interaction (ST), and student-content interaction (SC) as the key factors influencing the learning outcomes of online education, in addition to task value and self-regulated learning playing a part of chain mediation. Learners' self-regulation in instructional interactions includes both cognitive and motivational dimensions, Cho et al [22] conducted a series of studies on instructional interactions in online learning from the perspective of self-regulated learning, which showed that self-regulated learning behaviors occurring in interactions between learners and the content are different from self-regulated learning behaviors in interactions between learners and between learners and instructors.

In open education, the relationship between the content and the learner is much more than that between the message and the receiver, while the content is dependent on the online learning environment, and in terms of the order of interaction, the learner's first time tends to interact with the environment, so scholars prefer to equate interaction with interaction [23-25]. Online learning interaction is a series of interactive behaviors occurring based on the media, representative of the view of Brown et al [26], who believe that the interaction in online learning is a process of co-construction of knowledge between students or between teachers and students through cooperation, communication and negotiation of meaning, which is based on the development of the understanding of interaction, and has been widely recognized by scholars at home and abroad. Quadir et al. [27] developed an open education platform called “Learner's Brief Blog (LDB)” and explored the specific mechanism of teaching interaction affecting the learning effect, however, although the need for interaction in open courses has been emphasized, there is a lack of specific design strategies and evaluation of the effect of interaction on learning in actual courses. Bozkurt et al [28] explored the interaction of the learning process in online education from the perspectives of transactional distance, types of interactions and self-determination theory, and found that most learners tend to lurk or even drop out of the state during the learning process, a result that exposes the drawbacks of large-scale online open learning, and presents teachers with the challenge of facilitating learner interaction.

In the past decades, a large number of research results on instructional interaction strategies in open education have been accumulated. For example, Wang et al [29] crawled the learning data of 71,948 learners in iMOOC, a large-scale open education platform, and identified instructional interaction strategies through text mining techniques, which showed that six interaction strategies, namely, code writing, operation guidance, providing reference, encouragement, specification interpretation, and exchange of opinions, were able to reduce the dropout rate of learners. Alwafi et al [30] explored the effects of different discussion strategies on students' cognitive engagement and interaction levels through content analysis and social network analysis (SNA) with the support of learning analytics techniques, emphasizing that interactions between students were able to establish more cognitive learning connections, and that the use of interaction strategies incorporating learning analytics techniques enabled them to understand their own level of proficiency. Zebua [31] addressed the online education interaction deficiencies by stating that teachers should make interactive adjustments to instructional interaction strategies to facilitate a smooth learning process in new environments, and that teacher competence,

spirituality, and support infrastructure are important directions for improving instructional interaction strategies and are major factors in maximizing instructional interactions. Ong and Quek [32] argued that restricted teacher-student interactions in online teaching may lead to learning outcomes that are poorly, and proposed an effective teaching interaction strategy through the relationship between online learning experience, social needs, and teacher-student interaction. Sun et al [33] found that teacher-student interaction in online education has a positive effect on students' learning effect ($r=0.649$, $p<0.01$), for this reason, they proposed a teacher-student interaction strategy, which is to enhance the psychological atmosphere by creating a favorable students' interaction in the course, which in turn affects learning outcomes.

The study conducts single-point analysis and cluster analysis of adult learning behaviors by constructing learner portraits, on the basis of which K-prototype and Pearson correlation coefficient based on the hybrid metric of Hamming distance and Euclidean distance are used. Teaching interactive dialogue modeling was carried out through deep reinforcement learning, and a dialogue strategy model based on user image and deep reinforcement learning was proposed to enhance the learning effect of open education and realize the teaching interactive dialogue teaching platform. Comparison experiments with other dialog recommendation algorithm models in terms of recommendation, dialog and manual evaluation are conducted on two datasets, ReDial and INSPIRED, to prove the effectiveness of the algorithm. Finally, experimental teaching is conducted to analyze the learning effect of open education using Spss mean difference test-t-test.

2. Adult Learner Portrait Construction in Open Education Platforms

2.1. Adult Learner Data Acquisition and Processing

2.1.1. User data preprocessing module

In the mining and processing process of the entire user image, it is necessary to go through the collection of user information, integration, statute, cleanup, transformation and other stages, each stage is mainly centered on the processing of different objectives, of which the first few phases are the initialization of the data information processing stage, in order to make the entire system of data information to meet the requirements of processing.

In the entire user profile construction process, the first step needs to complete the entire data information preprocessing operations, so as to obtain the content to meet the needs of algorithm processing, specific data information preprocessing operations schematic shown in Figure 1.

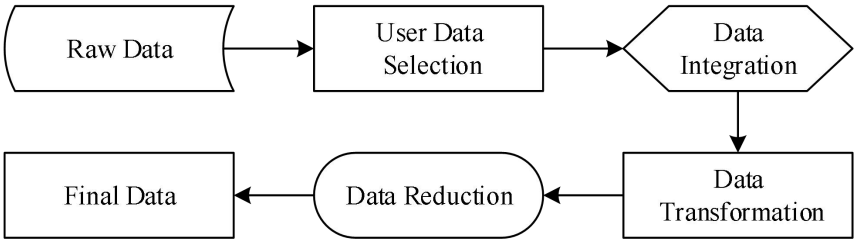

Figure 1. Data preprocessing diagram.

2.1.2. Behavior extraction and analysis module

In the construction of the entire user profile, the main role of the behavioral extraction and analysis module is to determine the user's behavioral preferences, the user behavioral characteristics of the tree model, mainly describes the specific user behavioral preference modeling process, using the model for the tree network model, each time the construction of the model may affect the final recommendation results. Users are categorized into multiple behavioral preference types, and in each behavioral preference type, it may include multiple course types, and each course corresponds to the corresponding weight information. In this tree model of user behavioral characteristics, the first layer is mainly the user as the main body, in determining the user as the main body, the second layer is mainly to analyze the behavioral preferences, and finally complete the construction of the user profile.

2.2. K-prototype based clustering model for adult learners

2.2.1. Distance metrics

The K-prototype measures numerical and categorical features according to the K-means and

K-modes distance calculations Euclidean distance and Hamming distance, respectively, which are combined to form the distance to the prototype [34]. Assume that the set $D=\{X_1,X_2,\cdots,X_n\}$ with n samples and m features, X_i, X_j denote the two samples respectively.

For numerical features, firstly, we need to normalize them and map them to the interval $[0,1]$, and then calculate the distance of numerical features, the Euclidean distance used in K-means is derived from the formula of distance between two points in Euclidean space, and the distance is denoted as:

$$d_1(X_i,X_j)=\sum_{l=1}^{m_r}\left(x_{il}^r-x_{jl}^r\right)^2 \tag{1}$$

For category type features, K-modes are computed using the Hamming distance:

$$d_2(X_i,X_j)=\sum_{l=1}^{m_t}\delta\left(x_{il}^t,x_{jl}^t\right) \tag{2}$$

where $\delta(p,q)=0$ when $p=q$ and $\delta(p,q)=1$ when $p\neq q$. For sample i, x_{il}^r and x_{jl}^r are numerical features, x_{il}^t and x_{jl}^t are categorical features, and m_r and m_t are the number of numerical and categorical features, respectively.

Calculating the dissimilarity between objects of mixed feature types can be done by combining different features into a single dissimilarity matrix, let k be the number of clusters, and $Q_c=\{q_{c1},q_{c2},\cdots,q_{cm}\}$ denote the center of the category chosen for category c, thus , the distance between the data and the center cluster can be expressed as follows:

$$d(X_i,Q_j)=d_1(X_i,Q_j)+\gamma_t d_2(X_i,Q_j) \tag{3}$$

Then the loss function of the K-prototype can be defined as:

$$L=\sum_{c=1}^{k}\left(L_c^r+L_c^t\right)=L^r+L^t \tag{4}$$

L_c^r then represents the total loss of all numerical features in the sample of category c, L_c^t denotes the total loss of all category features, and γ_c is the weight of category features in category c. The γ_c affects the accuracy of clustering. When $\gamma_c=0$, only numerical features are considered, which is equivalent to the K-means method, and when γ_c is larger, the category features take up more weight, and the clustering results are dominated by categorical variables. Therefore, choosing the appropriate γ_c can lead to better clustering. The choice of γ_c is affected by the mean square deviation of the numerical variables, and when the mean square deviation is set to 1, it is better to set γ_c to 0.5~0.7.

2.2.2. Hybrid Clustering K-prototype

The numerical variables are normalized to have a variance of 1, so γ_c is set to 0.5. Specific steps of the K-prototypes algorithm:

Step 1: Randomly select k initial clustering centers $\{c_1,c_2,\cdots,c_k\}$ from the dataset D.

Step 2: Traverse the dataset D, calculate the distance between the sample and each cluster center according to Equation (3), and assign the sample to the category closest to the center.

Step 3: After each sample assignment, update the clustering centers for numerical features and category features. Formula (1) is used to calculate for numerical features and formula (2) is used for category features.

Step 4: Compute the loss function using equations (3) and (4).

Step 5: If the new loss function value is less than the set threshold or the number of iterations is greater than the set T, the computation is finished and the clustering center is output, otherwise repeat steps 2, 3 and 4.

2.2.3. Cluster characterization and portrait construction

The selection of the number of clusters is related to the different clusters to get the statistical

characteristics and portrait situation, the profile coefficient combines the degree of cohesion and separation, which can objectively evaluate the clustering effect, therefore, in the by setting up different number of clusters, from which to select the cluster with the best performance of the profile coefficient.

After forming the clusters, the center of each cluster is described, and for the numerical features, the mean and variance of the student group of the cluster are calculated, and the evaluation is carried out in three dimensions, namely, basic attributes, life pattern, and daily consumption, and further fine-grained division is carried out according to the average grades of the compulsory courses in the clusters, so as to make the construction of the portrait of these clusters more comprehensive and specific.

2.3. Data Analysis of Adult Learning Behavior

2.3.1. Single-point analysis of student behavior

The single-point analysis of student behavior will be conducted in five dimensions: teamwork, theory performance, classroom performance, active learning, and extended competence, each based on different data. Teamwork is primarily based on the completion of students' group-type assignments, which is mainly determined based on the number of operational records of group-type assignments captured by the platform as well as the subjective determination of completion by the teacher. Theory grades are determined based on written test scores uploaded to the platform by administrators or teachers. Classroom performance is determined by capturing a combination of the number of valid records of operations performed by students during classroom time periods in the lab building, as well as the completion of attendance. The active learning dimension relies on the time course information of the student's behavioral portrait, and is determined by integrating the IP location information, hourly operation and other data. Expansion ability mainly reflects students' active exploration ability, which is mainly determined by the number of operations of non-essential commands collected and the completion of expansion assignments.The behavioral data required for the five dimensions comes from the automated collection of the platform on the one hand, and relies on the subjective uploads of students or teachers on the other hand. The analysis of single-user latitude attributes is shown in Table 1.

Table 1. Single-user dimension attribute analysis.

dimension	Corresponding Properties
Theoretical Score	Lab report grade, theory exam grade
Classroom performance	Attendance records and personal project sharing counts
Active learning	Total computer usage time and location switching frequency
Expand capabilities	Expand topic score and platform expansion operations
Team work	Team project overall score and team defense score

Based on the above correspondence between user dimensions and attributes, quantitative scores were assigned to each attribute's, and then weighted scores were assigned based on a scale set by the instructor. Each quantified dimension is normalized to the range of [0,20] for weighting and comprehensive quality. The behavioral portrait of the student with an overall quality of 78 and the five-dimensional single-point analysis is shown in Figure 2. It can be seen from the figure that the overall comprehensive ability of the student is not high, mainly due to the relatively weak expansion ability, which can be viewed in the detailed analysis of the expansion ability data, as shown in Figure 3. In the figure, the lowest number of extended operations is 2, the lowest number of effective operations is 17, the highest number of extended operations is 50, and the highest number of effective operations is 200.Meanwhile, for the thinking type of experimental topics, the completion rate of this student is only 15.48%. Therefore, in view of the student's weak quality of expansion ability, the teacher should give personalized guidance and suggest that the student should reasonably allocate time to complete the thinking and expanding type of subject projects.

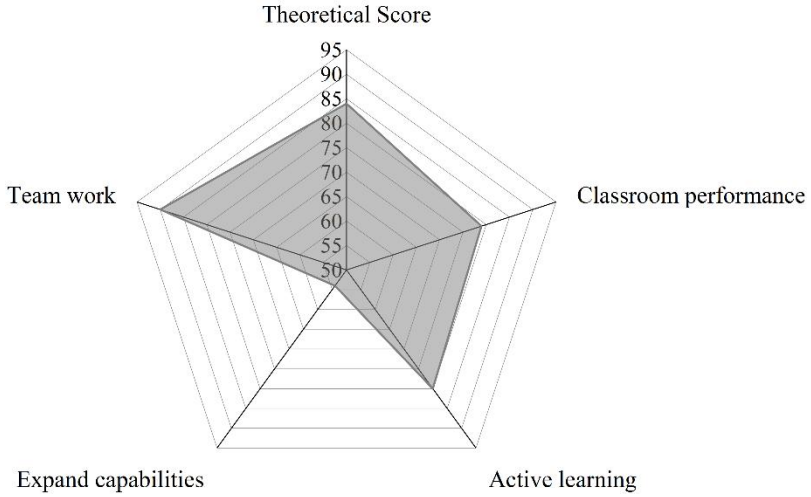

Figure 2. Student behavior single point analysis.

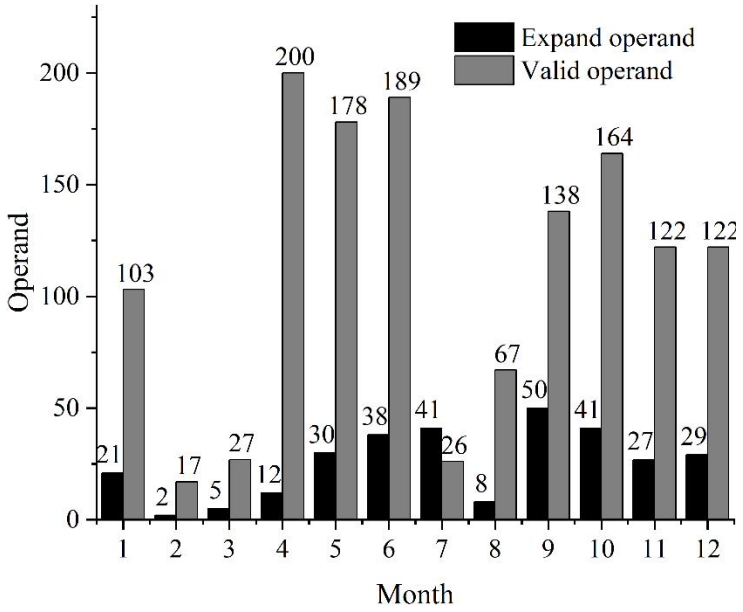

Figure 3. Student expansion capability analysist.

2.3.2. Cluster Analysis of Student Behavior

Student Behavior Cluster Analysis mainly focuses on big data analysis for multiple student groups, and mines effective information from the large amount of experimental behavior data generated by student groups. In this paper, we first carry out the Shapiro-Wilk test on the attribute data to verify the normality of the data, and then use the Pearson correlation coefficient to verify the correlation between the attribute data and the comprehensive score, and finally use the K-prototype algorithm to perform the cluster analysis of the students' experimental behavior.

1. Data Analysis Techniques

Pearson correlation coefficient [35] method is a statistical method to quantitatively predict the correlation between two sample sequences. The magnitude of the Pearson correlation coefficient is determined by the value of r, which reflects the degree of correlation between the two sequential samples. The calculation of Pearson correlation coefficient is shown in equation (5).

$$r = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sqrt{\left(\sum X^2 - \frac{\left(\sum X\right)^2}{N}\right)\left(\sum Y^2 - \frac{\left(\sum Y\right)^2}{N}\right)}} \tag{5}$$

Where: X denotes the set of x coordinates of the sample points to be tested; Y denotes the set of y coordinates of the sample points to be tested; N denotes the total number of samples; and r denotes the correlation coefficient, which takes the value $r \in [-1.0, 1.0]$. In general, $r \in (0.8, 1.0]$ denotes a very strong positive correlation, $r \in (0.6, 0.8]$ denotes a strong positive correlation, and vice versa, a negative correlation.

2. Experimental Analysis

In this paper, 1300 online students' behavioral data of operating system course are collected, their attributes are labeled by creating portraits, correlation validation and cluster analysis are conducted for students' experimental practice ability, and the dimensions of analysis contain the number of effective operations, online hours, after-school hours, and course hours. Some of the original samples are shown in Table 2.

Table 2. Original sample data for cluster analysis.

Student ID	Valid operand	Total duration/minutes	Class time/minutes	Course duration/minutes	Overall score/points
01	268	454	105	362	68
02	480	543	182	374	72
03	651	1049	432	630	82
04	485	776	210	579	83
05	550	815	244	584	84
06	402	270	44	239	80
07	689	1010	272	751	88
08	581	671	222	462	90
09	781	1363	674	702	95
010	320	771	586	498	91

First, four dimensions were extracted from the original samples to carry out W-test normality validation of single-dimension samples respectively, to determine whether the single-sample dataset conforms to the normal distribution; then, the single-sample data that conforms to the normal distribution was verified by Pearson's correlation coefficient with the composite scores, to verify the correlation; and finally, clustering analysis was carried out by using the clustering algorithm. Among them, the normal verification and Pearson correlation coefficient verification are shown in Table 3. From the analysis in Table 3, it can be seen that the W-test normal validation result values w of the four dimensions are significantly greater than 0.05, and the distribution of the sample data are all in line with the normal characteristics. Meanwhile, it can be seen that the Pearson correlation coefficients r of the 3 dimensions of total hours, after-school hours, and course hours with the comprehensive grade of the course are in the range of 0.8~1.0, which proves that there is an extremely strong correlation between these 3 dimensions and the comprehensive grade of the course. And the Pearson correlation coefficients between the effective number of operations and the comprehensive grade are in the range of 0.6~0.8, which proves that there is a strong correlation between the two.

Table 3. One-dimensional sample correlation verification results.

Dimension Name	W-test(w)	Pearson correlation coefficient(r)
Valid operand	0.7439	0.6482
Total duration	0.5163	0.8427
Class time	0.6672	0.8173
Course duration	0.3918	0.8922
Overall score	0.4028	-

Based on the above sample data of 4 dimensions which have some correlation with the composite scores, the next step is to analyze the K-prototype clustering algorithm. k-prototype algorithm is set $k \in (2,5)$ to make a side-by-side comparison to see the classification effect, and analyze it. The optimal number of iterations is adjusted using the sum of error squares as iterations are performed. Figure 4 shows a two-dimensional cut-away display of the data magnification effect for the optimal iteration case of k=4. At the same time, the clustering effect of 1300 students combined with the single-point data of students' behaviors for cluster analysis reveals that each cluster implies the corresponding behavioral information. For example, students with better final grades are generally characterized by a high number of effective operations and high after-school hours; while students with lower final grades are generally characterized by a low number of effective operations and fewer online hours. The 4 categories and behavioral characteristics of students obtained by clustering using the K-prototype algorithm are shown in Table 4, where category 2 accounts for more, indicating that students lack practical knowledge and have poor hands-on skills.

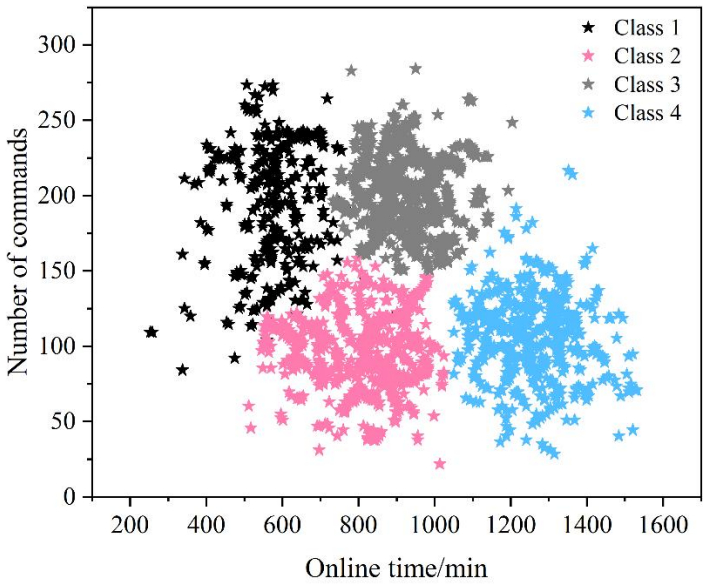

Figure 4. K-prototype (k=4) 2D cross-section.

Table 4. k-means clustering analysis.

Class	Quantity	Category Description	Characteristic
Class 1	137	High number of valid commands and total duration	Strong experimental skills and good overall performance
Class 2	589	The number of valid commands is normal, and the course duration is	The experimental ability is mediocre and the practice outside the class is less
Class 3	494	The number of valid commands is normal, and the post-class duration is	Strong experimental ability and more time for practice after class
Class 4	180	The number of valid commands is low, and the total duration is short.	The practical ability of the experiment is weak, and the time spent on the computer is relatively small

3. Reinforcement learning-based interactive modeling of instructional dialogues under learner profiling

3.1. Interactive dialog strategies for teaching and learning based on deep reinforcement learning

Deep learning can utilize multiple layers of nonlinear computation to achieve data processing, which is an important branch of machine learning, including recurrent neural networks, long and short-term memory networks, etc.. Reinforcement learning is also an important part of machine learning, which is capable of mapping from environment states to system actions. The purpose of reinforcement learning is to maximize the cumulative rewards that intelligences receive when interacting with the environment. Deep reinforcement learning combines deep learning with reinforcement learning and is essentially an

end-to-end perception and decision-making algorithm. The method can be trained to obtain intelligent agents for knowledge construction and learning.

Intelligence can ultimately obtain the optimal strategy through learning.Q-Learning [36] belongs to one of the basic algorithms in reinforcement learning, which needs to select the action that can obtain the greatest benefit according to the Q value. Deep reinforcement learning for the traditional Q-Learning algorithm in large-scale environments have Q value performance instability and other problems, the experience recovery mechanism and the goal Q network has been improved to avoid the algorithm in the training of the instability that may occur. The error function used in the optimization of the target Q network is shown in Equation (6).

$$L(\theta_i) = E_{s,a,r,s'}\left[(Y_i - Q(s,a \mid \theta_i))^2\right] \tag{6}$$

In Equation (6), s denotes the current state, a denotes the action taken in the current state, r denotes the reward or penalty corresponding to the action a taken in the s state, s' denotes the next state corresponding to the action a taken, i denotes the number of iterations, and $Q(s,a \mid \theta_i)$ denotes the Q function with the parameter θ. The gradient in Eq. (7) is obtained by taking the partial derivatives of the parameter θ.

$$\nabla_{\theta_i} L(\theta_i) = E_{s,a,r,s'}\left[(Y_i - Q(s,a \mid \theta_i))\nabla_{\theta_i} Q(s,a \mid \theta_i)\right] \tag{7}$$

After the introduction of the target Q network, the target Q value remains stable for a certain period of time and the stability of the algorithm is improved. Using the deep Q network model for problem solving in large-scale environments, as the model is iteratively explored, the feedback from different environments enables the intelligent body Agent to learn and mimic human behavior. Based on the deep Q network, the researchers propose the deep dual Q network algorithm to improve the traditional deep Q network. There are two identical Q networks in the deep dual Q network, and in order to eliminate the problem of overestimation of the Q values in the deep Q network, the decoupling of the target Q value and its selection of actions is carried out on its basis. There are two parameters θ and θ' in the deep dual Q network, where the parameter θ is used for the selection of the maximum Q value corresponding to a, and θ' is used for the evaluation of the optimal action for the Q value. The computation of the target Q value for the deep double Q network is shown in Equation (8).

$$Y_i^{\mathrm{DDQN}} = r + \gamma Q(s', \operatorname{argmax}_a Q(s', a \mid \theta_i) \mid \theta_{i'}) \tag{8}$$

In Equation (8), γ denotes the attenuation factor. Practice has shown that the Q value calculation of the deep double Q network is more accurate and has better stability compared to the deep Q network, so the deep double Q network is chosen to train the chatbot conversation model based on deep reinforcement learning in this experiment.

The chatbot conversation model based on the deep bi-Q network mainly includes a user simulator, conversation state tracking, and an intelligent body Agent. The user simulator is mainly used to simulate large-scale human-computer interaction between the user and the intelligent body Agent, and at the same time is able to store the gained experience in the experience pool for the offline training of the intelligent body. The dialog state tracking can collect and update the dialog state to provide a basis for the intelligent body's decision-making. The Intelligent Body Agent, which includes a DRL-based dialog model and a rule-based dialog strategy, receives data information from the dialog state tracking module and outputs an optimal action according to the dialog strategy learned by the Intelligent Body.

3.2. Conversation Strategies Based on User Profiling and Deep Reinforcement Learning

In the dialog strategy model based on user image and deep reinforcement learning, the model first obtains vector representations of user image and user dialog history through one-hot coding and gated neural units respectively, then integrates and splices the two with the current dialog state vector, and jointly inputs them into the action value network to obtain the value of each candidate dialog action; finally, the action with the highest value score is used as the strategy model's output for the control of the

preschool chatbot.

Similar to other deep reinforcement learning-based policy models, the proposed model also uses an experience playback pool to store the collected data samples and is used for training the model afterwards. During the simulated interaction with the user simulator, the squared loss of the minimized TD-error is used as the optimization objective of the policy model, and the parameters of the policy model are learned by back-propagation based on the calculated loss and gradient. The loss function of the model is calculated as in Eq. (9)(10):

$$L(\theta) = E\left[\left(y - Q(s,a;\theta)\right)^2\right] \tag{9}$$

$$y = r + \gamma \max_{a'} Q'(s',a';\theta') \tag{10}$$

where θ and θ' are parameters of the online and target networks, respectively.

3.3. Experimental results and analysis

3.3.1. Data sets

The datasets used for the experiments in this section are ReDial and INSPIRED.In this paper, the two datasets are divided into training, validation and test sets in the ratio of 8:1:1.In this paper, we extract the entities, different attributes and movie dictionaries in each conversation.

3.3.2. Comparative Experimental Settings and Evaluation Indicators

In order to validate the effectiveness of the dialog strategy model based on user profiling and deep reinforcement learning, this paper evaluates the proposed model with recent dialog methods, recommendation methods and dialog recommendation methods on two tasks, dialog and recommendation, and conducts comparative experiments on the above two tasks. Among these methods, for solving the dialog problem there is Transformer, for solving the recommendation problem there is TextCNN, and for solving the dialog recommendation problem there are REDIAL, KBGD, KGSF, RevCore, and UniCRS.

The dialog recommendation system provides recommendation items for the user in the process of interacting with the user in natural language, and when evaluating the dialog recommendation, the methods proposed in this paper need to be evaluated on the recommendation and dialog tasks respectively, and manual evaluation needs to be used for the evaluation of the overall effect. For the recommendation task, this paper adopts Recall as the evaluation index, for the dialog task, this paper adopts Dialog Diversity, Language Gap and Confusion Degree as the evaluation index, and for the manual evaluation, this paper designs Recommendation Consistency and Response Fluency as the evaluation indexes.

3.3.3. Implementation details

In this paper, a simulated online environment is obtained on the training set for training. It is based on Agenda-base, which stores upcoming user actions that are constructed from the information in the user's goals, and generates instant feedback based on the predicted recommendation items. The simulator was tested on user data, and experimental results showed that the simulated online environment had an overall accuracy of 90% for the instant feedback task. This result suggests that the simulator can accurately model real online environments and is therefore able to train and test the models in this paper on them. During training with the simulator, the model parameters were simplified to allow it to reach convergence faster. After training, the parameters of the model were fixed and then offline and online tests were done separately.

The framework adopted in this paper is Pytorch neural network framework and the Adam optimizer is used to train the model of this paper with β_1=0.9 and β_2 =0.999. The learning rate in training is set to 0.001 and the batch size is set to 16. The initial word embeddings are set using word2vec with a dimension of 300. The parameter update index μ was set to 0.8.

3.3.4. Analysis of results

In order to verify the effect of the dialog strategy model based on user profiling and deep reinforcement learning, this paper carries out control experiments between the proposed model and each comparison model described above, the control experiments are carried out on the ReDial and

INSPIRED datasets, the experimental metrics include Recall, Dist, BLEU, ppl, and the experiments are set up to test the test set with 5 rounds of conversations fixed in the End recommendation, the comparison results of each recommendation indicator fixed at 5 rounds of dialog are shown in Table 5, and the performance of this paper's model on ReDial dataset and INSPIRED dataset are shown in Table 6 and Table 7.

It was observed that the proposed method outperformed other recommendation and dialogue models in all evaluation metrics, with 0.453 in Recalla@50 on the ReDial dataset, an improvement of 14.2% in Dist-3, an improvement of 6.2% and 3.9% in BLEU-1 and BLEU-2, and a reduction of 2.0 in confusion. The comparison reveals that the algorithm based on Reinforcement Learning performs better in all the metrics compared to the traditional methods, indicating that the interactive nature of Reinforcement Learning is more suitable for the scenario of dialog recommendation. Compared with the Redial dataset, the INSPIRED dataset is smaller in size, the action space is significantly simplified, and it performs significantly better in Recall metrics and Dist metrics.

Table 5. Results of a comparative experiment with five rounds of fixed recommendation indicators.

Model	Redial			INSPIRED		
	Recalla@1	Recalla@10	Recalla@50	Recalla@10	Recalla@10	Recalla@50
Popularity	0.008	0.032	0.077	0.004	0.024	0.054
TextCNN	0.008	0.049	0.087	0.005	0.027	0.071
ReDial	0.0015	0.083	0.143	0.005	0.079	0.147
KBRD	0.026	0.077	0.215	0.041	0.096	0.179
KGSF	0.029	0.086	0.257	0.044	0.107	0.231
RevCore	0.032	0.179	0.382	0.047	0.160	0.351
UniCRS	0.037	0.167	0.361	0.073	0.194	0.374
Model in this article	0.066	0.209	0.453	0.114	0.246	0.459

Table 6. Results of the comparative experiment on the Redial dataset.

Model	Redial						
	Dist-2	Dist-3	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ppl
Transformer	0.152	0.253	0.045	0.04	0.005	0.004	19
ReDial	0.208	0.339	0.05	0.009	0.003	0.002	29.4
KBRD	0.244	0.372	0.057	0.02	0.004	0.003	19.3
KGSF	0.283	0.437	0.084	0.021	0.007	0.004	9.5
RevCore	0.406	0.562	0.098	0.027	0.005	0.004	11.2
UniCRS	0.485	0.633	0.206	0.046	0.011	0.007	9.3
Model in this article	0.666	0.764	0.265	0.081	0.034	0.011	7.8

Table 7. Results of the comparative experiment on the nspired dataset.

Model	INSPIRED						
	Dist-2	Dist-3	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ppl
Transformer	0.263	0.654	0.083	0.025	0.009	0.007	15.4
ReDial	0.405	1.229	0.131	0.016	0.007	0.005	25.6
KBRD	0.547	2.022	0.165	0.029	0.014	0.008	20.2
KGSF	0.603	2.524	0.196	0.032	0.017	0.013	11.5
RevCore	0.972	0.561	0.246	0.047	0.026	0.015	12.7
UniCRS	2.044	3.664	0.355	0.066	0.031	0.022	10.4
Model in this article	2.669	4.386	0.447	0.085	0.041	0.028	7.7

In addition, it can be seen that both UniCRS and this paper's model that tries to unify the dialogue and recommendation modules outperform the traditional dialogue recommender system trained separately, which indicates that training the dialogue and recommendation modules separately will create semantic inconsistency problems between the two modules, and by training the two modules jointly, both tasks of recommendation and dialogue can be improved, and it is an effective way to optimize the dialogue recommendation.

Finally, in order to understand the recommendation uniformity and response fluency of the proposed algorithm in this paper, in addition to the above evaluations, volunteers were invited to evaluate the recommendation efficiency and the effectiveness of responding to conversations, called Consistency and

Fluency. 100 randomly selected multiple rounds of conversations from the test set were asked to give ratings by the volunteers. As shown in Table 8, Transformer scores low in Consistency because it tends to generate less informative answers, such as “It's okay” and “Sounds good.” For all four CRS methods, RevCore scored the highest. However, in the Fluency metric, the results of the traditional methods were less satisfactory due to repetitive and short discourse. Finally, this paper's model generated more informative and fluent responses better than all baselines. This suggests that this paper's model has a suitable dialog strategy framework and fine-grained predictions that can learn rich information.

Table 8. Compare experimental results of manual evaluation indicators.

Model	Consistency	Fluency
Transformer	2.3	2.4
ReDial	2	2.8
KBRD	3.1	3.4
KGSF	3.5	3.7
RevCore	3.7	3.9
UniCRS	3.4	3.8
Model in this article	3.9	4

4. Empirical Evidence of Learning Effectiveness of Teaching Interaction Strategies Based on Reinforcement Learning

The study selects three classes in the open education platform for adult education as the object of study, with classes 1 and 2 as the experimental group and class 3 as the control group, class 1 adopts the interactive strategy based on reinforcement learning in this paper, class 2 does not adopt the targeted teaching strategy, and class 3 is in the regular teaching mode.

4.1. Pre-experimental data analysis

Before the experiment, we analyzed the computer scores of the study subjects and the results are shown in Table 9, from the statistical scale, we see that the mean values both class 1 and class 2 have higher scores than the control class.

Table 9. Group statistics.

	Classes	N	Mean	standard deviation	Standard error of the mean
Last semester grades	3	50	81.274	7.4138	6.7442
	1	50	84.473	6.4522	7.7493
Last semester grades	3	50	81.274	7.4138	6.7442
	2	53	82.581	5.7649	9.1936

The results of the independent samples t-test for classes 1 and 2 are shown in Tables 10 and 11. Levene test of variance equation in Spss statistics to test the homogeneity of variance. f-statistic of Levene test for class 3 and class 1 is 1.317, p=0.295>0.05, which indicates that the variance of the two groups is homogeneous and did not reach the level of significance so t-value is the following line of equal variance t=-1.999, df=95, p=0.054<0.05 that indicating that there is a significant difference in the mean scores. In addition, a class 3 and class 2 have a Levene's test f-statistic of 3.276, p=0.089>0.05, indicating that the variances of the two groups are homogeneous and have not reached a significant level, so the t-value is the line above assuming equal variances t=--0.782, df=96, p=0.443>0.05, indicating that there is no significant difference in the mean grades.

Table 10. Independent sample inspection 3 and 1.

		Levene's test for variance			T-test for the mean equation	
		F	Sig.	t	df	Sig.
Last semester grades	Assume equal variance	1.317	0.295	-1.999	95	0.054
	Assume unequal variances			-1.999	91.074	0.054
		T-test for mean equations				
		Mean	Standard error		95% confidence interval for	

		difference		the difference	
				lower limit	superior limit
Last semester grades	Assume equal variance	-2.6734	1.5371	-5.6127	-0.0184
	Assume unequal variances	-2.6734	1.5371	-5.6139	-0.0182

Table 11. Independent sample inspection 3 and 2.

		Levene's test for variance			T-test for the mean equation	
		F	Sig.	t	df	Sig.
Last semester grades	Assume equal variance	3.276	0.089	-0.782	96	0.443
	Assume unequal variances			-0.779	89.241	0.446
		T-test for mean equations				
		Mean difference	Standard error	95% confidence interval for the difference		
				lower limit	superior limit	
Last semester grades	Assume equal variance	-1.1274	1.3527	-3.6918	1.6318	
	Assume unequal variances	-1.1274	1.3509	-3.6729	1.6259	

4.2. Post-experimental data analysis

The computer scores of the study subjects after the experiment were analyzed as shown in Table 12. As far as the mean scores of classes 1 and 3 are concerned, with 50 students in one class, the mean is 61.274 and the standard deviation is 19.4732, the standard error of the mean is equal to the standard deviation divided by N. The means of classes 1 and 2 are greater than the mean of class 3.

Table 12. Group statistics.

	Classes	N	Mean	standard deviation	Standard error of the mean
Last semester grades	3	50	61.274	19.4732	2.5676
	1	50	72.486	14.2143	3.5176
Last semester grades	3	50	61.274	19.4732	2.5676
	2	53	80.568	16.3309	3.2454

Tables 13 and 14 show the results of independent samples t-test for class 1 and class 2. One of the basic assumptions of the test of difference in means is homogeneity of variance, which becomes homogeneity of variance when the variance is the same before the t-test is performed to test the data dispersion. Levene's test for variance equation in Spss statistics to test for homogeneity of variance. the Levene's test F-statistic for class 3 and class 1 is 5.014, $p=0.033<0.05$ indicating that the variances of the two groups are not homogeneous to the level of significance so our t-value is the line below where the variances are not equal $t=-2.732$, $df=82.294$, $p= 0.017<0.05$, indicating a significant difference.

Also, a Levene's test F-statistic of 1.974 for class 3 and class 2, $p=0.204>0.05$, suggests that the two groups are homogeneous in terms of variance and have not reached a significant level, so our t-value is the line above assuming equal variances $t=-5.673$, $df=90$, $p=0.001<0.05$, suggesting a significant difference, indicating a significant difference in the grade point averages.

Table 13. Independent sample inspection 3 and 1.

		Levene's test for variance			t-test for the mean equation		
		F	Sig.	t	df	Sig.	Mean difference
Average scores	Assume equal variance	5.014	0.033	-2.748	92	0.015	-8.7459
	Assume unequal			-2.732	82.294	0.017	-8.7459

	variances						
Ttest for the mean equation							
					Standard error	95% confidence interval for the difference	
						lower limit	superior limit
Last semester grades	Assume equal variance				3.3974	-16.4108	-2.1138
	Assume unequal variances				3.4127	-16.5127	-1.8726

Table 14. Independent sample inspection 3 and 2.

		Levene's test for variance			t-test for the mean equation		
		F	Sig.	t	df	Sig.	Mean difference
Average scores	Assume equal variance	1.974	0.204	-5.673	90	0.001	-19.9137
	Assume unequal variances			-5.416	86.331	0.001	-19. 9137
Ttest for the mean equation							
					Standard error	95% confidence interval for the difference	
						lower limit	superior limit
Last semester grades	Assume equal variance				3.5729	-27.0429	-12.8946
	Assume unequal variances				3.5618	-27.0439	-12.9013

4.3. Experimental conclusions

Through the above experimental data there are some experimental conclusions as shown in Table 15, from the table we can see that class 3 as a control group before and after the experiment with class 2 mean value differences exist, after the experiment the difference increases significantly; and another experimental data, class 3 with class 1 is the case of no significant difference to the existence of a difference in the value of -8.7459, these data show that through the dialogue based on the user image and deep reinforcement learning learning on teaching platform has a positive effect on open education in adult education.

Table 15. experimental situation table.

Class	Class 3 and Class 2	Class 3 and Class 1
Average score difference before experiment	P=0.047<0.05, This indicates a significant difference in the average scores (Difference value=-2.6734)	P=0.043>0.05,It means there is no significant difference in the average score
Average score difference after experiment	P=0.001<0.05,Indicates significant differences,(Difference value=-19.9137)	P=0.0017<0.05,No significant difference was found(Difference value=-8.7459)

5. Conclusion

In this paper, we use the K-prototype method and correlation algorithm based on the hybrid metric of Hamming distance and Euclidean distance to analyze students' learning behaviors, based on which we construct a teaching interaction system based on user profiling and deep reinforcement learning. An experiment was designed to select multiple recommended methods for comparative analysis on the ReDial and INSPIRED datasets. On the ReDial dataset, the score was 0.453 in Recalla@50, and the method proposed in this paper scored the highest in the Fluency index. It has been proved that the improvement of the performance of the interactive teaching system based on user profiling and deep reinforcement learning algorithms. Finally, by combining the objective effect analysis with the learning outcomes of specific students, the data results show that adult education has a positive effect on the learning outcomes of open education on the interactive dialog teaching platform.

Funding
This research was supported by the Research Project on Education and Teaching Reform of Shaanxi Open University in 2023: Research on the Construction of Ideological and Political Education Curriculum in Open Education--A Study and Exploration Based on the Construction of Ideological and Political Education Curriculum in Yulin Open University (No: sxkd2023zx16).

References

1. Kumar, A., Kumar, P., Palvia, S. C. J., & Verma, S. (2017). Online education worldwide: Current status and emerging trends. Journal of Information Technology Case and Application Research, 19(1), 3-9.
2. Rhim, H. C., & Han, H. (2020). Teaching online: foundational concepts of online learning and practical guidelines. Korean journal of medical education, 32(3), 175.
3. Ushanov, A., Morgunova, N., & Petunina, I. (2021). Internet technologies in distance education. International Journal of Emerging Technologies in Learning (iJET), 16(10), 85-95.
4. Harsası, M. (2015). The use of open educational resources in online learning: A study of students' perception. Turkish Online Journal of Distance Education, 16(3), 74-87.
5. Cunha, M. N., Chuchu, T., & Maziriri, E. (2020). Threats, challenges, and opportunities for open universities and massive online open courses in the digital revolution. International Journal of Emerging Technologies in Learning (iJET), 15(12), 191-204.
6. Kisworo, M. W. (2016). Implementing Open Source Platform for Education Quality Enhancement in Primary Education: Indonesia Experience. Turkish Online Journal of Educational Technology-TOJET, 15(3), 80-86.
7. Liu, Z. Y., Lomovtseva, N., & Korobeynikova, E. (2020). Online learning platforms: Reconstructing modern higher education. International Journal of Emerging Technologies in Learning (iJET), 15(13), 4-21.
8. Ohanu, I. B., & Chukwuone, C. A. (2018). Constraints to the use of online platform for teaching and learning technical education in developing countries. Education and Information Technologies, 23(6), 3029-3045.
9. Kassandrinou, A., Angelaki, C., & Mavroidis, I. (2014). Transactional distance among open university students: How does it affect the learning process?. European Journal of Open, Distance and E-Learning, 17(1), 26-42.
10. Abrami, P. C., Bernard, R. M., Bures, E. M., Borokhovski, E., & Tamim, R. M. (2011). Interaction in distance education and online learning: Using evidence and theory to improve practice. Journal of computing in higher education, 23(2), 82-103.
11. Gupta, D. S. K., & Hayath, T. M. (2022). Lack of it Infrastructure for ICT based education as an emerging issue in online education. Technoarete Transactions on Application of Information and Communication Technology (ICT) in Education, 1(3).
12. Johnson, A. M., Jacovina, M. E., Russell, D. G., & Soto, C. M. (2016). Challenges and solutions when using technologies in the classroom. In Adaptive educational technologies for literacy instruction (pp. 13-30). Routledge.
13. Alhih, M., Ossiannilsson, E., & Berigel, M. (2017). Levels of interaction provided by online distance education models. Eurasia Journal of Mathematics, Science and Technology Education, 13(6), 2733-2748.
14. Wang, Y., & Liu, Q. (2020). Effects of online teaching presence on students' interactions and collaborative knowledge construction. Journal of computer assisted learning, 36(3), 370-382.
15. Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. Contemporary educational psychology, 61, 101860.
16. Năznean, A. (2022). EFFECTIVE TEACHER-LEARNER COMMUNICATION AND INTERACTION–A BRIEF LITERATURE. JOURNAL PEDAGOGY, 1, 151-161.
17. Kamran, F., Kanwal, A., Afzal, A., & Rafiq, S. (2023). Impact of interactive teaching methods on students learning outcomes at university level. Journal of Positive School Psychology, 7(7), 89-105.
18. Pianta, R. C. (2017). Teacher-student interactions: Measurement, impacts, improvement, and policy. In Teachers, teaching, and reform (pp. 75-93). Routledge.
19. Akhtar, S., Hussain, M. U. H. A. M. M. A. D., Afzal, M., & Gilani, S. A. (2019). The impact of teacher-student interaction on student motivation and achievement. European Academic Research, 7(2), 1201-1222.
20. Alvi, E., & Gillies, R. M. (2015). Social interactions that support students' self-regulated learning: A case study of one teacher's experiences. International Journal of Educational Research, 72, 14-25.
21. Li, X., Lin, X., Zhang, F., & Tian, Y. (2022). What matters in online education: exploring the impacts of instructional interactions on learning outcomes. Frontiers in Psychology, 12, 792464.
22. Cho, M. H., Demei, S., & Laffey, J. (2010). Relationships between self-regulation and social experiences in asynchronous online learning environments. Journal of Interactive Learning Research, 21(3), 297-316.
23. Cole, M. T., Shelley, D. J., & Swartz, L. B. (2014). Online instruction, e-learning, and student satisfaction: A three year study. International Review of Research in Open and Distributed Learning, 15(6), 111-131.
24. Shu, H., & Gu, X. (2018). Determining the differences between online and face-to-face student–group interactions in a blended learning course. The Internet and higher education, 39, 13-21.
25. Prohorets, E., & Plekhanova, M. (2015). Interaction intensity levels in blended learning environment. Procedia-social and behavioral sciences, 174, 3818-3823.
26. Brown, H. D., & Lee, H. (1994). Teaching by principles: An interactive approach to language pedagogy (Vol. 1, p. 994). Englewood Cliffs, NJ: Prentice Hall Regents.

27. Quadir, B., Yang, J. C., & Chen, N. S. (2022). The effects of interaction types on learning outcomes in a blog-based interactive learning environment. Interactive Learning Environments, 30(2), 293-306.
28. Bozkurt, A., Koutropoulos, A., Singh, L., & Honeychurch, S. (2020). On lurking: Multiple perspectives on lurking within an educational community. The Internet and Higher Education, 44, 100709.
29. Wang, W., Zhao, Y., Wu, Y. J., & Goh, M. (2023). Interaction strategies in online learning: Insights from text analytics on iMOOC. Education and Information Technologies, 28(2), 2145-2172.
30. Alwafi, E. M. (2022). Designing an Online Discussion Strategy with Learning Analytics Feedback on the Level of Cognitive Presence and Student Interaction in an Online Learning Community. Online Learning, 26(1), 80-92.
31. Zebua, R. S. Y. (2020). The Strategy to Build Educative Interaction in Islamic Education on Online Learning Setting. Mudarrisa: Jurnal Kajian Pendidikan Islam, 12(2), 185-202.
32. Ong, S. G. T., & Quek, G. C. L. (2023). Enhancing teacher–student interactions and student online engagement in an online learning environment. Learning environments research, 26(3), 681-707.
33. Sun, H. L., Sun, T., Sha, F. Y., Gu, X. Y., Hou, X. R., Zhu, F. Y., & Fang, P. T. (2022). The influence of teacher–student interaction on the effects of online learning: Based on a serial mediating model. Frontiers in psychology, 13, 779217.
34. Azimah Mohd, Lay Eng Teoh & Hooi Ling Khoo. (2024). Passengers' requests clustering with k-prototype algorithm for the first-mile and last-mile (FMLM) shared-ride taxi service. Multimodal Transportation,3(2),100132-.https://doi.org/10.1016/J.MULTRA.2024.100132.
35. Jun Zhang & Bingqing Lin. (2025). Semi-parametric estimation of Pearson correlation coefficient under additive distortion measurement errors. Communications in Statistics - Theory and Methods,54(18),5806-5829.https://doi.org/10.1080/03610926.2024.2446419.
36. Peide Liu, Hasan Dinçer, Serhat Yüksel & Serkan Eti. (2025). Smart manufacturing investment strategies with renewable energy efficiency using Q-learning algorithm and molecular fuzzy Bayesian networks. Applied Soft Computing,185(PB),114007-114007.https://doi.org/10.1016/J.ASOC.2025.114007.