https://doi.org/10.31449/inf.v46i18.9938

Enhanced Firefly Algorithm with Lévy Flight Strategy for Feature Optimization in LSSVM-Based Online Learner Behavior Recognition

Zhihua Zhang
Xuchang Vocational Technical College, College of Humanities, Xuchang 461000, China
E-mail: xczhangzhihua@163.com

Keywords: online learner behavior recognition, improved firefly algorithm, integrated model, feature extraction, intelligent education

Received: July 1, 2025

With the rapid development of online education, the precise identification of learners' behavioral patterns to enhance the effectiveness of personalized teaching has become an important research topic in the field of smart education. To improve the accuracy of online learning behavior recognition, this study proposes an integrated model based on an optimized least squares support vector machine using an improved firefly algorithm. This method improves the skip search and local convergence capabilities of FA in high-dimensional feature spaces. It introduces a Lévy flight mechanism and a dynamic step-size adjustment strategy. It also combines a bagging ensemble strategy to construct a multi-subclassifier fusion structure. These improvements effectively enhance the model's generalization performance and classification stability. The experimental part is based on the UCI public student learning behavior dataset and real Moodle platform behavior logs for evaluation. The model performance was systematically tested using 50% cross validation and comparative experiments. The results showed that the proposed model achieved accuracies of 97.84% and 97.52% on the training and testing sets, respectively. The minimum classification error rate was 2.42%. In terms of error metrics, the proposed model outperformed others in classification error rate, root mean square error, and cross-entropy loss. Specifically, the classification error rate for problem-solving tasks was 3.15%, the root mean square error was 0.10, and the cross-entropy loss was 0.23. Meanwhile, the model has good resource utilization control, with an average memory usage of less than 450MB and a CPU usage rate of less than 65%. The proposed model demonstrates high accuracy and scalability in recognizing multi-class behaviors. It is suitable for automating the modeling process and deploying intelligent teaching support systems for large-scale learning behavior data on online educational platforms.

Povzetek: Razvit je noi model LFFA-LSSVM, ki z izboljšanim algoritmom kresnice (Firefly) z Lévyjevimi skoki in dinamično stopnjo koraka optimizira izbor značilk ter izboljša razpoznavo vedenja spletnih učencev.

1 Introduction

With the continuous development of information technology and the acceleration of educational digitization, online education has become an important component of higher education reform and lifelong learning systems [1]. Learning behavior recognition not only helps to understand individual learning styles and knowledge mastery levels but also enables precise recommendations and dynamic interventions [2]. However, current online learning data exhibits characteristics such as high dimensionality, diversity, and uneven label distribution, leading to issues such as low recognition accuracy and weak error control capabilities in traditional classification models when applied to behavior recognition tasks. Previous studies have primarily used machine learning or deep learning methods to identify and model online learning behaviors. These methods include temporal behavior analysis based on graph convolutional networks and unsupervised clustering methods based on contextual interactions. These methods have produced some results. Ngo et al. analyzed self-regulation patterns in group collaborative learning based on social shared regulation theory. They revealed the added value of artificial intelligence in education. This study provided theoretical and practical references for personalized learning design, while deepening the understanding of diversity in collaborative learning in higher education [3]. Peng and Fu explored personalized adaptive learning pattern recognition technology in online education. The study constructed learner feature models by mining and analyzing platform learning behavior data. They designed a recognition framework based on cognitive level, learning style, interactive behavior, and social characteristics. Experiments confirmed that this method could effectively identify learners' interactive behavior and social learning patterns [4]. Yan et al. proposed a personalized learning recognition method based on graph convolutional neural networks. The method recognized the emotional state by analyzing the learner's micro-expressions, so as to dynamically adjust the learning path. Experiments showed that the technique not only enhanced the level of personalization in online education, but also optimized the learning experience [5]. Shobana and Kumar developed the intelligent quiz (I-Quiz) assessment system. The system assessed knowledge mastery by analyzing learners' nonverbal behavior, including implicit and explicit features, combined with machine learning techniques. After parameter optimization, the system achieved 85.68% prediction accuracy [6].

Firefly algorithm (FA) is a meta-heuristic optimization method based on group intelligence, which simulates the mechanism of individual synergy through brightness attraction behavior of fireflies in nature [7]. With strong global search capability and simple structure, FA has been widely used in tasks such as feature selection, parameter tuning and combinatorial optimization. Cao et al. proposed an improved visual FA (VFA), which balanced global exploration and local exploitation by introducing a multi-group mechanism and a selective randomization strategy. Experiments proved that the algorithm could effectively improve the performance of the original FA and performed well on CEC2013 test problems with different dimensions [8]. Peng et al. proposed a sliding window FA (SWFA) to optimize the original FA by introducing a sliding window mechanism and an inverse learning strategy. The algorithm used a window archiving and updating mechanism to reduce invalid comparisons, and combined with an adaptive step-size strategy to balance the exploration and exploitation capabilities. Experiments verified that SWFA significantly improved the performance of the algorithm [9]. However, in high-dimensional spaces, FA is prone to fall into local optimality and the search efficiency decreases. Least squares support vector machine (LSSVM) is an improvement of standard support vector machine (SVM). By introducing the squared error loss function and linear system of equations solution, the training efficiency of the model is greatly improved, especially suitable for fast modeling scenarios with small and medium sized data [10]. Lai et al. proposed an innovative framework that combines discriminative feature extraction with sparse feature selection and integrates it into the architecture of SVM. This method had the advantage of fully utilizing the classifier's feature selection ability, thereby optimizing the classification boundary and improving the model's generalization performance. This framework effectively reduced the dimensionality of high-dimensional data, making the extracted low-dimensional features more suitable for SVM classification requirements and improving model performance [11].

In summary, the existing research on online learner behavior recognition suffers from insufficient model generalization ability, high computational complexity, and poor multi-classification performance. In view of this, a study proposes an ensemble learner behavior recognition method based on improved FA. Unlike studies that only use traditional feature selection or static classification strategies, the research achieves two innovations in method design: First, the introduction of the Lévy flight mechanism and the dynamic step size adjustment strategy significantly improves the FA's ability to perform jumping searches and converge accurately in high-dimensional feature spaces. This effectively avoids falling into local optima. The second objective is to develop a framework that incorporates the Bagging ensemble structure. This improves stability and generalization ability in multi-classification scenarios by using a collaborative recognition mechanism involving multiple subclassifiers.

2 Methods and materials

2.1 Improved FA-based feature extraction for online learner classification

In online learning behavior recognition, extracting discriminative features from high-dimensional data is the key to constructing efficient classification models. Researchers can quickly obtain key information and summarize features by analyzing online learning behavior behavioral data. To reduce data redundancy in learner classification and identification in subsequent studies, the study is based on FA for feature selection and classification of online learners. The traditional FA is shown in Figure 1 [12].

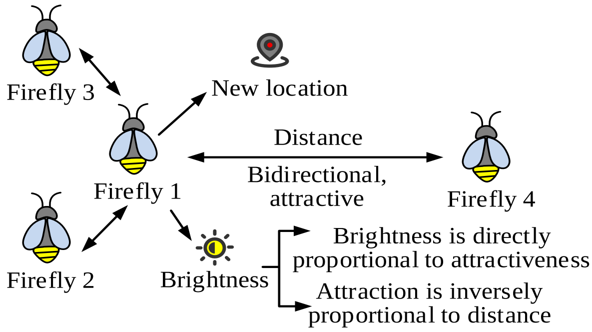

Figure 1: Firefly algorithm (Icon source: https://iconpark.oceanengine.com/official)

In Figure 1, multiple fireflies represent different solutions. They attract and move towards each other based on their brightness. Each firefly is attracted to the brighter individual and moves in its direction, thus continuously approaching the optimal solution. Higher brightness indicates a better fitness value at that location. Although the traditional FA has the ability of global search, it is easy to fall into the local optimum in high-dimensional space, and a Lévy Flingt FA (LFFA) is proposed in the study. It also combines the dynamic attractiveness function to improve the convergence efficiency of the population. The expression of Lévy flight stochastic wandering model is shown in Equation (1).

$$S_n = \sum_{i=1}^{n} X_i = X_1 + ... + X_n \tag{1}$$

In Equation (1), X_i denotes the random move step size constructed by different random distributions. S_n denotes the continuous summation of X_i. The change of X_i obeys the Lévy distribution, which exhibits the characteristic of alternating between short-distance moves and long-distance jumps. This characteristic makes it particularly suitable for multimodal optimization problems. Its stochastic step size formula is shown in Equation (2) [13].

$$S = \frac{u}{|v|^{1/\phi}} \tag{2}$$

In Equation (2), S denotes the Lévy flight step. u and v both denote random variables obeying a normal distribution, controlling the numerator term and tail behavior of the step length, respectively. ϕ is the stability index of the Lévy distribution, which determines the degree of heavy tailing of the distribution. Traditional FA typically uses a fixed step factor that is too small in the early stages, leading to slow convergence, and too large in the later stages, causing oscillation near the optimal solution. For this reason, the study proposes a dynamically adjusted step-size strategy that adapts the step-size factor throughout the iteration process. This improves the convergence accuracy and optimization performance of the algorithm, as shown in Equation (3).

$$\alpha_i = \alpha \frac{d_{i,best}}{d_{\max}} \tag{3}$$

In Equation (3), α_i denotes the optimized step factor. α denotes the step factor of FA. $d_{i,best}$ denotes the spacing of individual d_i from the optimal individual of the population. $d_{\max}$ denotes the maximum distance of d_{best} from other individuals in the population. LFFA utilizes the long-distance jumping property of Lévy flight to enhance the global search ability and avoid falling into local optimum. Meanwhile, dynamically adjusting the step size automatically reduces the search range as the optimal solution is approached, suppressing the oscillation phenomenon. This significantly improves convergence accuracy and optimization search efficiency. The individual update formula of LFFA is shown in Equation (4) [14].

$$x_i^{t+1} = x_i^t + \beta_0 e^{-\gamma d_{ij}^2}(x_j^t - x_i^t) + \alpha \cdot Levy(\lambda) \tag{4}$$

In Equation (4), x_i^t denotes the position of individual i in generation t. β_0 denotes the maximum attraction. i denotes the light intensity attenuation coefficient. $Levy(\lambda)$ denotes the random vector generated by Levy distribution. The LFFA process is shown in Figure 2.

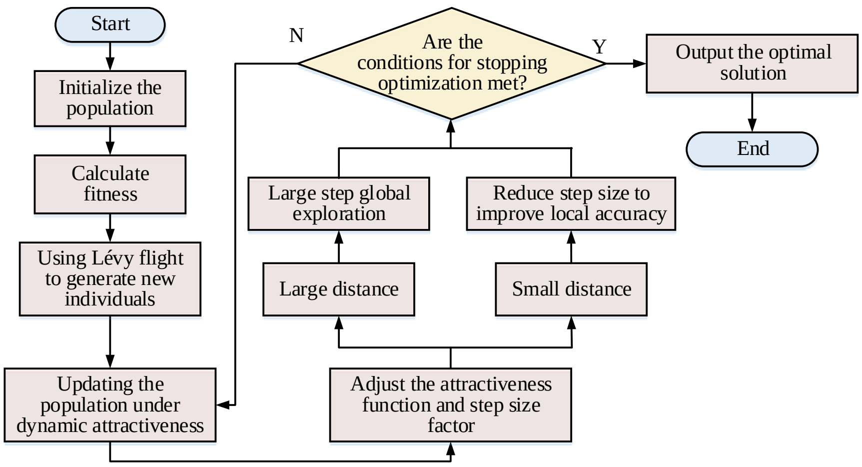

Figure 2: LFFA flowchart

In Figure 2, the algorithm first initializes the population, randomly generates multiple feature subset candidate solutions. The key parameters such as maximum attraction and light intensity attenuation coefficient are also set. In the iterative optimization stage, the algorithm introduces the Lévy flight perturbation mechanism through the introduction of Lévy. It utilizes its long-tailed distribution characteristics to realize the jumping movement of candidate solutions, which effectively enhances the global search capability. Meanwhile, a dynamic position update strategy is adopted to adaptively adjust the attractiveness function and step size factor based on the individual spacing. A larger step size is used for global exploration when the distance is large. When the distance is small, the step size is reduced to improve local development accuracy. The optimal feature subset is output after iteration to convergence. The feature selection removes redundant features by correlation algorithm and filters the optimal feature subset to enhance the model generalization ability and reduce the computational cost. The processing flow is shown in Figure 3 [15].

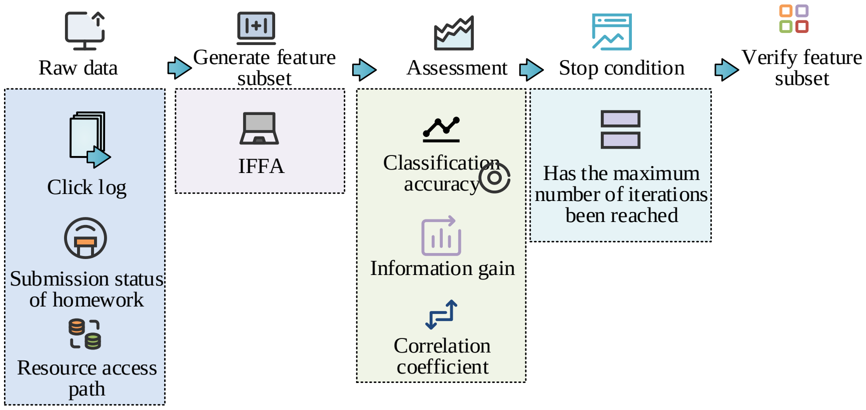

Figure 3: Feature selection processing flow (Icon source: https://iconpark.oceanengine.com/official)

In Figure 3, the raw data is first processed using LFFA to generate a subset of candidate features by programmed computation. Subsequently, the subset is evaluated to determine its advantages and disadvantages under the model performance or objective function. If the current feature subset does not satisfy the set stopping conditions, such as insufficient performance enhancement or the number of iterations reaches the upper limit, then continue to return to the original data to generate a new subset and repeat the evaluation process. If the stopping conditions are met, the final validation stage is entered and the optimal feature subset is output as the result.

2.2 Integrated learning recognition model based on LFFA-LSSVM

After feature extraction and screening through LFFA, the study constructs an LSSVM classifier based on LFFA optimization to further improve the model's ability to recognize complex online learning behaviors. This model is called the LFFA-LSSVM. As the core classifier of the integrated learning framework, this model can effectively combine highly discriminative features extracted in the previous stage to accurately recognize different types of online learners. First, a linear function is used to fit the modeling of the sample data in the high-dimensional feature space, as shown in Equation (5).

$$y = w^T \delta(x) + b \tag{5}$$

In Equation (5), y denotes the output of the model. w^T denotes the weight vector. $\delta(x)$ denotes the nonlinear mapping function. $\delta(x)$ is the bias term. The LSSVM transforms the original SVM by replacing inequality constraints with equality constraints, introducing error variables, and adding regularization constraints. This transformation turns the SVM into a system of linear equations that can be solved, significantly improving computational efficiency. The optimization objective is shown in Equation (6) [16].

$$\min_{w,b,e} J(w,e) = \frac{1}{2}\|w\|^2 + \frac{\eta}{2}\sum_{i=1}^{N} e_i^2 \tag{6}$$

In Equation (6), $J(w,e)$ is the objective function of LSSVM. e_i denotes the error term of the i th sample. η is the regularization parameter. N is the total number of training samples. The Lagrange multiplier method is used to optimize the solution as shown in Equation (7) [17].

$$L(w,b,e,\alpha) = \frac{1}{2}\|w\|^2 + \frac{\eta}{2}\sum_{i=1}^{N} e_i^2 - \sum_{i=1}^{N} \chi_i [w^T \delta(x_i) + b + e_i - y_i] \tag{7}$$

In Equation (7), $L(w,b,e,\alpha)$ denotes the Lagrangian function, which contains the original objective function and the constraint terms of the equation, and serves as the basis for solving the KKT condition. χ_i is the Lagrange multiplier. The regression prediction model obtained from the solution is shown in Equation (8).

$$f(x) = \sum_{i=1}^{N} \chi_i K(x_i, x) + b \tag{8}$$

In Equation (8), $K(x_i, x)$ denotes the linear kernel. The LSSVM modeling process is shown in Figure 4 [18].

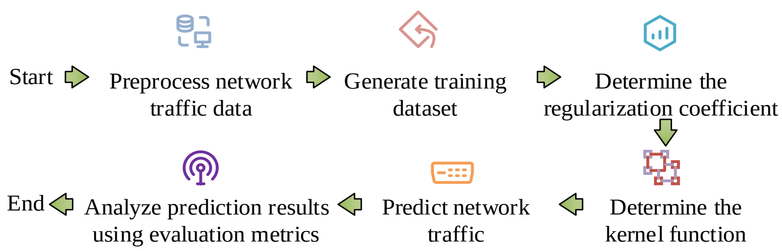

Figure 4: LSSVM model process (Icon source: https://iconpark.oceanengine.com/official)

In Figure 4, the LSSVM model identifies and preprocesses the outliers present in the dataset by collecting the historical data of network traffic to ensure the stability and validity of the data quality. Subsequently, the normalization operation is performed on the processed data to eliminate the interference caused by different feature scales, and the training set is constructed accordingly. In the training stage, appropriate penalty coefficients and kernel function parameters are determined according to the cross-validation results, and then the LSSVM prediction model is built. After the construction is completed, the model is used to predict the future network traffic, and the prediction results are compared and analyzed with the actual observations. Finally, the prediction error is quantified based on evaluation metrics to assess the accuracy and adaptability of the constructed model in the network traffic prediction task. To further improve the recognition stability, the study constructs an integration structure based on the Bagging idea. Multiple LFFA-LSSVM subclassifiers are trained on different feature subspaces and sample subsets. The final recognition results are also output using the weighted voting fusion strategy, as shown in Equation (9).

$$y' = \arg\max_{k} \sum_{i=1}^{T} w_i \cdot II(h_i(x) = k) \tag{9}$$

In Equation (9), y' denotes the final classification decision result. T denotes the total number of subclassifiers. w_i denotes the weight of the i th subclassifier. w_i the predicted output of the i th subclassifier for sample x. k is the classification category. The study establishes a mapping relationship between learning characteristics and academic performance based on key features such as the frequency with which resources are accessed, the number of announcements viewed, participation in discussions, and the number of days absent from school, as shown in Figure 5.

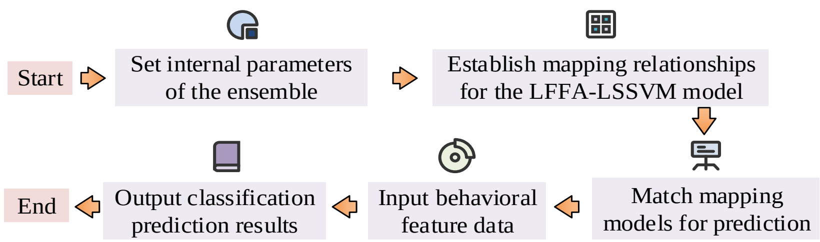

Figure 5: LFFA-LSSVM mapping process (Icon source: https://iconpark.oceanengine.com/official)

In Figure 5, the mapping process starts with model parameter initialization and structure configuration, focusing on constructing the mapping relationship network between each functional module. In the core prediction session, the model transforms the input high-dimensional behavioral feature data into classification output through the feature space mapping mechanism. The system first receives the preprocessed learner behavioral features. Subsequently, pattern matching and classification decisions are made based on the pre-trained mapping relations, and accurate recognition results are finally output. To further enhance the application value of the model in practical teaching systems, a real-time feedback oriented adaptive learning mechanism is added to the LFFA-LSSVM model. This mechanism dynamically adjusts the learning content push strategy based on identified behavior state categories, the current task completion status, and behavior sequence patterns. This creates a perception response closed-loop teaching process. In addition, interpretable artificial intelligence (XAI) mechanisms have been introduced to enhance the usability and controllability of the model in educational settings. These mechanisms are combined with methods such as classification heat mapping to enable teachers and teaching managers to intuitively understand the model's judgment criteria. This assists them in making targeted teaching interventions. The LFFA-LSSVM model process is shown in Figure 6.

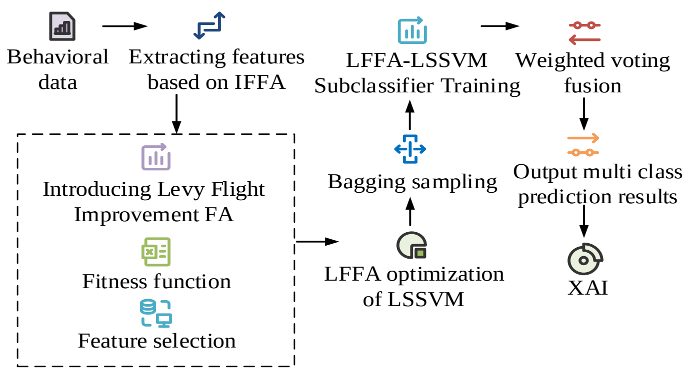

Figure 6: LFFA-LSSVM classification model (Icon source: https://iconpark.oceanengine.com/official)

In Figure 6, firstly, the system collects raw behavior data of learners from online learning platforms, including log information such as access paths, click frequency, and dwell time. Subsequently, the original data is standardized using the feature engineering module to extract representative behavioral feature vectors. Based on this, the LFFA with the Lévy flight strategy and the dynamic step-size mechanism is introduced to perform a global feature subset search and select the most discriminative behavioral features. During the model construction phase, the bagging ensemble strategy is used to combine multiple LSSVM subclassifiers, which are trained using different subsets of features, into a more robust ensemble recognition model. The model outputs behavior category labels for each learner at different stages, such as resource acquisition, problem-solving, learning assessment, or invalid browsing. To enhance the interpretability and educational controllability of the model, the XAI mechanism is further introduced. The recognition results then enter the adaptive feedback module. This module automatically adjusts the difficulty, recommendation strategies, and feedback rhythm of subsequent teaching content. It does so based on the learner's current behavior and task progress. This creates a "recognition feedback optimization" teaching loop.

3 Results

3.1 LFFA-LSSVM model performance testing

To validate the performance of the LFFA-LSSVM model, the study is based on an experimental environment of Windows 10 operating system, Intel Core i7-12700H processor, 16GB RAM, Python 3.9 with MATLAB R2022b. The UCI student performance dataset is selected, which consists of 395 samples and 33 features covering basic information about students, their social backgrounds, and their academic behaviors. Before model training, the mean imputation method is used to fill in missing attribute values, and the minimum maximum normalization method is uniformly applied to all numerical features to map the data to the [0, 1] interval, improving the comparability of various dimensional features. To verify the robustness of the model and the rationality of the parameter configuration, a five-fold cross-validation mechanism is introduced during the training phase. A grid search strategy is also used for hyperparameter optimization. The main parameters are shown in Table 1.

Table 1: Parameter configuration

Parameter	Set value
Maximum number of iterations	300
Initial population size	30
α_i	[0.1, 1.0]
η	{0.01, 0.1, 1, 10}
Kernel function type	Linear kernel
kernel parameter	{0.001, 0.01, 0.1, 1}

In Table 1, the parameter settings are divided into a training set and a testing set in an 8:2 ratio. The LFFA-LSSVM model proposed in the study is tested against Q-learning based on the adaptive logarithmic spiral-Levy flight FA (QL-ADIFA) [19] and improved fuzzy sparse multi-class least squares support vector machine (IF-S-M-LSSVM) [20]. The loss function curve obtained is shown in Figure 7.

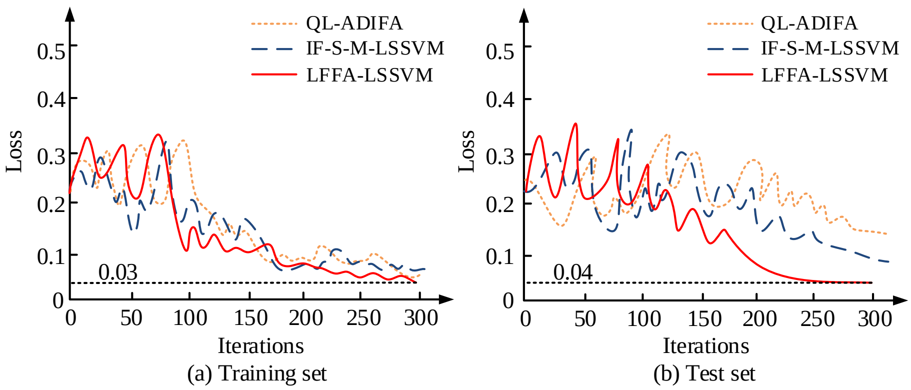

Figure 7: Loss function curves of different models

Figures 7(a)-(b) show the loss trends of the three algorithms on the training and test sets, respectively. In Figure 7(a), LFFA-LSSVM has relatively large fluctuations at the beginning of iterations. However, it starts to decrease rapidly after the 100th iteration and stabilizes around 200 times. Its final training loss is about 0.03, which is significantly lower than that of QL-ADIFA and IFS-M-LSSVM. It indicates that the model has stronger fitting ability and convergence speed in the training stage. In contrast, the IFS-M-LSSVM performs relatively stable though within the first 150 times. However, the overall loss decreases slowly and eventually stays at 0.08, while the loss of QL-ADIFA is higher than 0.1 and has the worst convergence performance. In Figure 7(b), LFFA-LSSVM also shows a strong convergence trend on the test set. Although there is a certain degree of fluctuation during the first iteration, it gradually stabilizes after 200 iterations. The final test loss is about 0.04, which is better than IFS-M-LSSVM and QL-ADIFA. It indicates that LFFA-LSSVM not only has lower training error, but also has better generalization performance and stability, which is suitable for the recognition modeling task under complex behavioral data. The recognition accuracy results of different models on the dataset are shown in Figure 8.

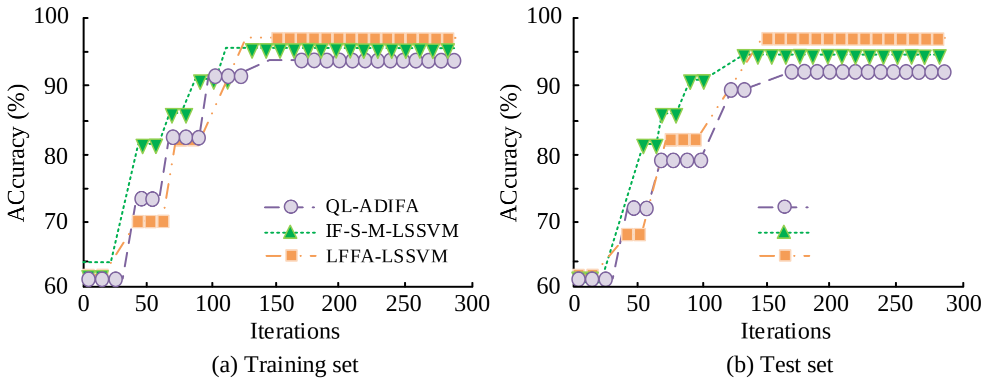

Figure 8: Accuracy variation curves of different models

Figures 8(a)-(b) show the trend of the recognition accuracy of the three algorithms with the number of iterations on the training and test sets, respectively. In Figure 8(a), all three models have low accuracy within the first 50 iterations. Among them, LFFA-LSSVM is slightly slower at the beginning, but rapidly improves after the 70th round. Moreover, it reaches more than 95% in the 150th round and finally stabilizes at about 97.84%. The IFS-M-LSSVM model gradually approaches 95% accuracy after the 100th iteration, with slight fluctuations in some rounds before stabilizing at 95.64%. Although the overall trend of QL-ADIFA is favorable, its final accuracy of 94.85% is slightly lower than the remaining two. In Figure 8(b), the accuracy of LFFA-LSSVM rapidly improves and stays at 97.52% after 150 rounds of iterations with minimal fluctuations, showing good stability and robustness. The test accuracy of IFS-M-LSSVM is about 95.31%. It shows some generalization ability, but there are slight iterative fluctuations. The final accuracy of QL-ADIFA on the test set is 92.68%, which is significantly lower than the remaining two optimization models. It indicates that its generalization performance on the test sample is relatively weak. The precision-recall (PR) curves of different models are shown in Figure 9.

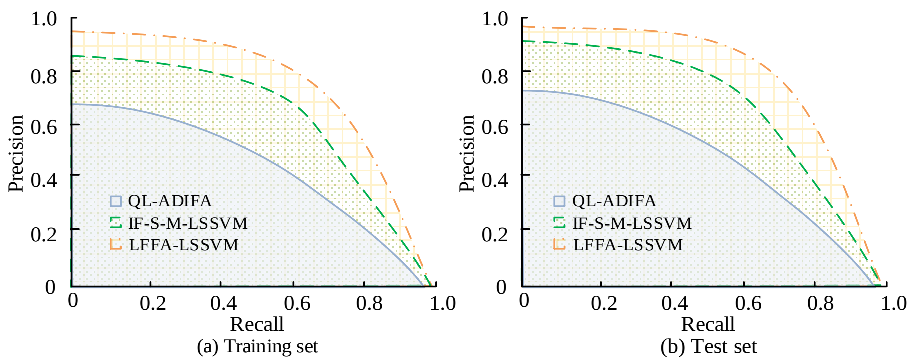

Figure 9: PR curves of different models

Figures 9(a)-(b) show the comparison of the PR curves of the three algorithms on the training and test sets, respectively. In Figure 9(a), LFFA-LSSVM always maintains the highest position. It indicates that it maintains a higher precision rate at all recall levels. The curve approximates a full rectangular shape form, showing very strong recognition stability and very low false positive rate. The IFS-M-LSSVM is located in the middle. Although it is overall better than QL-ADIFA, the drop-in precision rate is more pronounced in the high recall interval segment. QL-ADIFA is at the lowest position, and its precision rate shows a rapid decreasing trend with increasing recall. It shows that it has more serious problems of misclassification and classification instability. In Figure 9(b), the overall trend is consistent with the training set. The LFFA-LSSVM model still maintains superior performance on the test set with the largest curve area. It indicates that the model still has strong precision and better generalization ability while maintaining high recall. The IFS-M-LSSVM still maintains high precision in the low and medium recall intervals. However, there is an obvious decline in the high recall interval, reflecting the lack of ability to recognize some boundary samples. Whereas the PR curve of QL-ADIFA model tends to decline earlier on the test set, indicating its insufficient ability to discriminate unseen samples and poor generalization performance. To further validate the robustness and generalization ability of the LFFA-

LSSVM model, a five-fold cross validation experiment is conducted on the original training set. The accuracy, error rate, root mean square error (RMSE), and cross entropy loss of the model under various fold cross validation are shown in Table 2.

Table 2: Five-fold cross validation experimental results

Evaluation indicators	Training accuracy (%)	Test accuracy (%)	Classification error rate (%)	RMSE	Cross entropy loss
Fold 1	97.65	97.43	2.57	0.11	0.24
Fold 2	97.89	97.61	2.39	0.10	0.22
Fold 3	97.83	97.55	2.45	0.11	0.23
Fold 4	97.76	97.38	2.62	0.12	0.25
Fold 5	97.95	97.64	2.36	0.10	0.22
Average value	97.82	97.52	2.48	0.108	0.232
Standard deviation	±0.11	±0.10	±0.11	±0.008	±0.012

In Table 2, the training accuracy of LFFA-LSSVM remained between 97.65% and 97.95% in the 50%-fold range, with an average of 97.82% and a standard deviation of only ± 0.11%. The average testing accuracy is 97.52%, with a standard deviation of ±0.10%. This indicates that the model generalizes well to unseen samples. With a standard deviation of only ±0.11%, the average classification error rate of 2.48% further indicates that the model performs stably and has a low misjudgment rate under different data partitions. In terms of RMSE, the results of the five experiments are all between 0.10 and 0.12. The average is 0.108, with a standard deviation of ±0.008. This indicates that the model has good fitting accuracy at the probability output level and that the deviation between the predicted results and the true labels is minimal. Additionally, the average value of the cross-entropy loss is 0.232, with a standard deviation of ±0.012. This value remains between 0.22 and 0.25 across all compromises, indicating that the model can accurately approximate the true probability distribution in multi-classification tasks with minimal information loss. Overall, LFFA-LSSVM performs well in accuracy, stability, error control, and information fidelity, demonstrating high robustness and practical application value.

3.2 Application effect of online learner behavior recognition based on LFFA-LSSVM

To further validate the practicality of the LFFA-LSSVM model in specific learning behavior category recognition, an online learning management system based on the Moodle platform is selected as the experimental environment for the study. The platform supports a variety of learning behaviors acquisition such as course access, homework submission, video learning, discussion and communication. It can completely map six categories of behaviors: information acquisition, knowledge exploration, problem solving, self-regulation, social interaction and learning evaluation. Behavior logs are collected and structured through platform plug-ins, and the classification accuracy and differentiation ability of different models are systematically tested and analyzed. The confusion matrix obtained from the classification test of the six types of learning behaviors is shown in Figure 10.

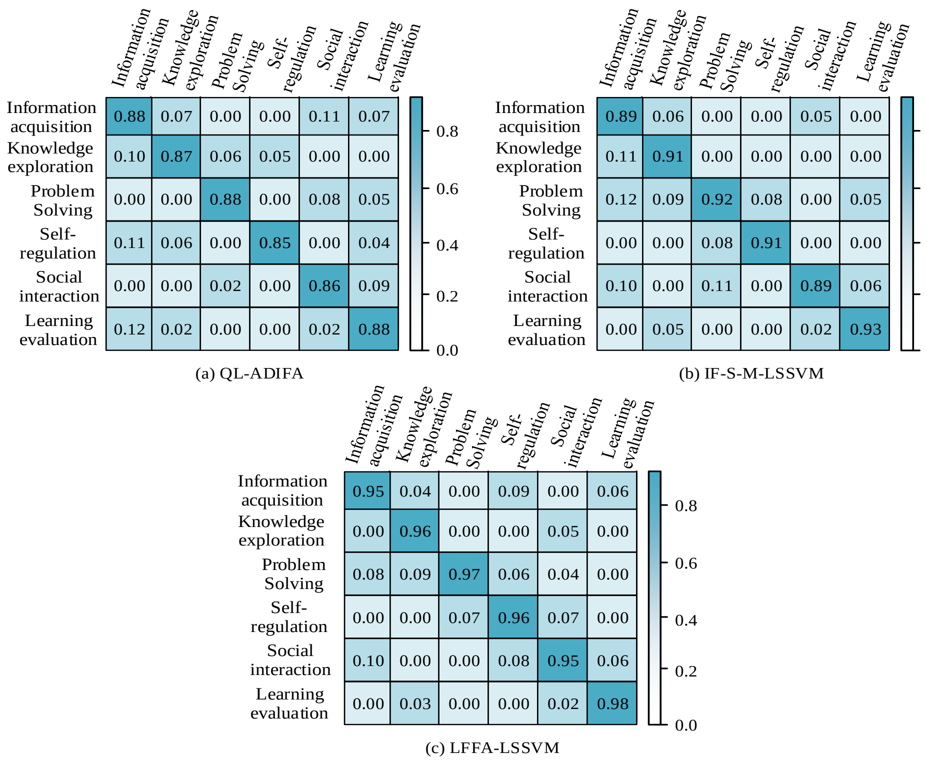

Figure 10: Classification confusion matrix results

Figures 10(a)-(c) show the results of the confusion matrix visualization of the three behavior recognition models in the six categories of online learning behavior classification tasks, respectively. In Figure 10(a), QL-ADIFA is weak in recognition on most categories, especially on problem solving, self-regulation, and learning evaluation behaviors, where the confusion is more serious. For example, the problem-solving category is completely misclassified as other categories with a precision of 0. The precision on self-regulation is 0.85 and it is misclassified into multiple neighboring categories. Overall, the accuracy of 0.88 is achieved in the categories of information acquisition and learning evaluation. The model exhibits obvious imbalance in category discrimination and insufficient recognition stability. In Figure 10(b), the classification ability of IF-S-M-LSSVM is significantly improved compared to QL-ADIFA, with recognition accuracy of more than 0.89 for all categories. The worst is 0.89 for the knowledge exploration category accuracy, and the best is 0.93 for the learning evaluation accuracy. The model achieves high diagonal occupancy in most of the categories, which suggests that it has strong behavioral differentiation to some extent ability. However, there is still a certain degree of confusion between problem solving and self-regulation, especially between similar behavioral semantics, which is still prone to judgmental bias. In Figure 10(c), the LFFA-LSSVM model performs best in the recognition of six categories of learning behaviors. The recognition accuracy of all behaviors reaches more than 0.96. Among them, the problem solving and learning evaluation accuracy is as high as 0.97 and 0.98, respectively. The recognition accuracy of the self-regulation category also improves to 0.96 with very few misclassifications. Except for the diagonal, most of the non-diagonal positions have values close to 0. It indicates that the model not only has a high overall accuracy, but also has a low misclassification rate, a clear distinction between behavioral boundaries, and a strong generalization ability. To comprehensively evaluate the error performance of the model in the multi-category behavior recognition task, the study introduces the classification error rate, RMSE, and cross-entropy loss as error indicators. The classification error rate reflects how accurately the model classifies the final labels. The RMSE measures the deviation between predicted probabilities and true labels. Cross-entropy is more concerned with information loss in multi-category distribution learning. The joint use of multidimensional error metrics helps to comprehensively determine the robustness and stability of the model in classifying complex learned behaviors, as shown in Table 3.

Table 3: Comparative experimental results

Model	Covered area	Classification error rate (%)	RMSE	Cross entropy loss
QL-ADIFA	Information acquisition	12.26	0.26	0.54
	Knowledge exploration	13.16	0.25	0.51
	Problem solving	20.41	0.30	0.61
	Self-regulation	15.49	0.28	0.56

	Social interaction	14.19	0.27	0.53
	Learning evaluation	12.97	0.26	0.52
IF-S-M-LSSVM	Information acquisition	11.59	0.21	0.42
	Knowledge exploration	11.41	0.22	0.40
	Problem solving	8.12	0.17	0.37
	Self-regulation	9.18	0.18	0.39
	Social interaction	11.65	0.21	0.41
	Learning evaluation	7.25	0.16	0.35
LFFA-LSSVM	Information acquisition	5.21	0.13	0.28
	Knowledge exploration	4.46	0.12	0.26
	Problem solving	3.15	0.10	0.23
	Self-regulation	4.51	0.11	0.25
	Social interaction	4.16	0.11	0.25
	Learning evaluation	2.42	0.09	0.22

In Table 3, LFFA-LSSVM outperforms the other two methods in all error metrics, showing the strongest recognition performance and generalization ability. First, in terms of classification error rate, LFFA-LSSVM controls the error rate within 5.5% on all six behavioral categories. Among them, the learning evaluation category is the lowest, only 2.42%, and the problem-solving category is only 3.15%. It indicates that the model has stronger differentiation ability in refined behavioral classification, especially for the easily confused problem solving category with higher discriminative accuracy. Second, in terms of RMSE metrics, the overall prediction bias of LFFA-LSSVM is significantly lower than that of IF-S-M-LSSVM and QL-ADIFA. The RMSE is below 0.13 in all categories. Among them, the learning evaluation category has the lowest RMSE of 0.09, while the overall RMSE of QL-ADIFA is generally in the range of 0.26 to 0.30. It reflects that its prediction results deviate more seriously from the true label and have poorer robustness. Although IF-S-M-LSSVM is improved than QL-ADIFA, it still fails to compress the error effectively on behaviors such as social interaction and self-regulation. Finally, in terms of cross-entropy loss, LFFA-LSSVM also achieves the minimum value, fluctuating between 0.22 and 0.28 on average. The best performance is for the learning evaluation category, which is 0.22. It indicates that the model has the least loss of information under the multi-category distribution and the closest probabilistic prediction of the real category. Whereas, the cross entropy of QL-ADIFA is generally higher than 0.5. It indicates that its distribution learning ability is insufficient and there is large error information. IF-S-M-LSSVM is at an intermediate level, with relatively stable performance but still has room for improvement. The resource consumption of different models in different behavioral categories is shown in Figure 11.

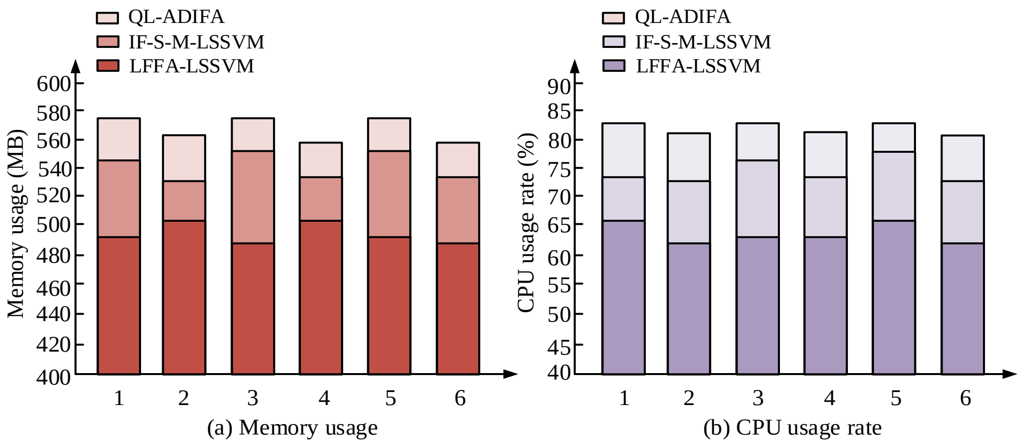

Figure 11: Comparison of resource consumption among different models

Figures 11(a)-(b) show the memory usage and CPU usage of the three models in the six categories of learning behavior recognition tasks, respectively. Among them, 1~6 represent the six behavioral categories of information acquisition, knowledge exploration, problem solving, self-regulation, social interaction, and learning evaluation, respectively. In Figure 11(a), LFFA-LSSVM consistently has the lowest memory footprint among all behavior types, averaging under 450 MB. In contrast, QL-ADIFA occupies more than 500MB in multiple tasks, the highest resource consumption among the three models. The IFS-M-LSSVM is located in the middle. Its memory usage is lower than that of the QL-ADIFA, but slightly higher than that of the LFFA-LSSVM. It occupies close to 480 MB during the problem-solving and self-regulation tasks. This indicates that its structural complexity and resource dependency are still more pronounced. As shown in Figure 11(b), LFFA-LSSVM's CPU utilization is kept between 65%

and 70%. This is significantly lower than the CPU utilization of IF-S-M-LSSVM and QL-ADIFA, which generally exceeds 75%. The CPU utilization of QL-ADIFA is 84.68% and 84.85% for “Problem Solving” and “Social Interaction” tasks, respectively. LFFA-LSSVM remains at 63.25% and 65.12%, respectively. It shows that LFFA-LSSVM not only has superior recognition effect, but also has better parallel computing efficiency and device adaptability. To evaluate the robustness and applicability of the proposed LFFA-LSSVM model on different online learning platforms, the study selects three popular platforms for external validation and prediction testing, including Coursera, Edmodo, and Canvas. In the experimental setup, the model uses LFFA-LSSVM parameters that are trained on Moodle. The model then makes predictions without fine-tuning in order to test its ability to generalize without fine-tuning. The comparison results are shown in Table 4.

Table 4: Comparison of verification results on different platforms

Test indicators	Coursera	Edmodo	Canvas	Average value
Test accuracy (%)	95.83	95.42	94.96	95.40
Classification error rate (%)	4.17	4.58	5.04	4.60
RMSE	0.138	0.142	0.151	0.144
Cross entropy loss	0.265	0.271	0.284	0.273

As shown in Table 4, although there are structural differences in behavior data across different platforms, the proposed model maintains good recognition accuracy on three external platforms. On average, the testing accuracy is 95.40%, indicating that the model has strong cross-platform transfer recognition ability. In addition, the classification error rate and loss function values increase slightly compared to the main platform. This reflects the impact of differences in platform behavior on the model's discrimination boundary. In the future, transfer learning and domain adaptation methods will be combined to further optimize the model's robustness.

4 Conclusion

To address the issues of high dimensionality, redundant features, and imbalanced labels in online learning behavior data, the research constructed an LFFA model that integrated a Lévy flight mechanism and a dynamic step-size strategy. This model was then used to optimize the LSSVM for online learner behavior recognition. The experimental results revealed that the LFFA-LSSVM model demonstrated superior performance in terms of recognition accuracy, error control, and resource consumption. The LFFA-LSSVM model exhibited higher accuracy and lower classification error, reaching 97.84% and 97.52% on the training and test sets, respectively. In the recognition of six typical learning behaviors, the model's error rate was as low as 3.15% and 2.42% in the complex behavior categories of “problem solving” and “learning evaluation”. They were better than the mainstream models such as QL-ADIFA and IF-S-M-LSSVM. In addition, the model also maintained the lowest level of RMSE and cross-entropy loss, showing good robustness and generalization ability. Meanwhile, LFFA-LSSVM also had an advantage in resource consumption, with the average memory consumption controlled below 450MB and the CPU utilization rate not exceeding 70%. This provided a feasible guarantee for the deployment of the model in a real teaching environment. However, the current model primarily uses structured platform behavior logs to construct features and has not yet incorporated multimodal information, such as text content and video interaction. This restricts the ability to deeply understand complex behavioral features. Future research can expand on this foundation by exploring areas such as multimodal behavioral data fusion, cross-platform migration recognition, and personalized learning strategy recommendations to provide more comprehensive and intelligent educational support services. In addition, to enhance the model's adaptability for deployment in real-world teaching scenarios, the study will also focus on the lightweight design of the model. The study will construct a simplified version of the model suitable for small and medium-sized schools or mobile terminals by applying feature dimension compression, submodule pruning, and parameter optimization. This will reduce resource consumption, improve real-time response efficiency, and increase deployment flexibility.

References

[1] Jagadeesh M, Baranidharan B. Facial expression recognition of online learners from real-time videos using a novel deep learning model. Multimedia Systems, 2022, 28(6): 2285-2305. DOI: 10.1007/s00530-022-00957-z.

[2] Du Y, Crespo R G, Martínez O S. Human emotion recognition for enhanced performance evaluation in e-learning. Progress in Artificial Intelligence, 2023, 12(2): 199-211. DOI: 10.1007/s13748-022-00278-2.

[3] Ngo D, Nguyen A, Dang B, Ngo H. Facial expression recognition for examining emotional regulation in synchronous online collaborative learning. International Journal of Artificial Intelligence in Education, 2024, 34(3): 650-669. DOI: 10.1007/s40593-023-00378-7.

[4] Peng P, Fu W. A pattern recognition method of personalized adaptive learning in online education. Mobile Networks and Applications, 2022, 27(3): 1186-1198. DOI: 10.1007/s11036-022-01942-6.

[5] Yan J, Wang N, Wei Y, Han H. Personalized learning pathway generation for online education through image recognition. Traitement du Signal, 2023, 40(6): 2799-2808. DOI: 10.18280/ts.400640

[6] Shobana B T, Kumar G A. I-Quiz: An intelligent assessment tool for Non-Verbal behaviour detection. Computer Systems Science & Engineering, 2022, 40(3): 1007-1021. DOI: 10.32604/csse.2022.019523

[7] Chen M R, Yang L Q, Zeng G Q, Lu K D, Huang Y Y. IFA-EO: An improved firefly algorithm

hybridized with extremal optimization for continuous unconstrained optimization problems. Soft Computing, 2023, 27(6): 2943-2964. DOI: 10.1007/s00500-022-07607-6
[8] Cao L, Ben K, Peng H, Zhang X. Enhancing firefly algorithm with adaptive multi-group mechanism. Applied Intelligence, 2022, 52(9): 9795-9815. DOI:10.1007/s10489-021-02766-9
[9] Peng H, Qian J, Kong F, Fan D, Shao P, Wu Z. Enhancing firefly algorithm with sliding window for continuous optimization problems. Neural Computing and Applications, 2022, 34(16): 13733-13756. DOI: 10.1007/s00521-022-07193-6
[10] Zhou S, Zhang W, Chen L, Xu M. Robust least squares projection twin SVM and its sparse solution. Journal of Systems Engineering and Electronics, 2023, 34(4): 827-838. DOI:10.23919/JSEE.2023.000103
[11] Lai Z, Chen X, Zhang J, Kong H, Wen J. Maximal margin support vector machine for feature representation and classification. IEEE transactions on cybernetics, 2023, 53(10): 6700-6713. DOI:10.1109/TCYB.2022.3232800
[12] Wang C. Emotion recognition of college students' online learning engagement based on deep learning. International Journal of Emerging Technologies in Learning, 2022, 17(6): 110-122. DOI: 10.3991/ijet.v17i06.30019.
[13] Yuan Q, Wang S, Hu M, Zeng L. SLDChOA: a comprehensive and competitive multi-strategy-enhanced chimp algorithm for global optimization and engineering design. The Journal of Supercomputing, 2024, 80(3): 3589-3643. DOI:10.1007/s11227-023-05617-1
[14] Wang J, Ji Y, Wei L, Chen H, Song W W. A dynamic firefly algorithm based on two-way guidance and dimensional mutation. International Journal of Bio-Inspired Computation, 2022, 20(2): 126-137. DOI: 10.1504/IJBIC.2022.126772.
[15] Wang J, Yue K, Duan L. Models and techniques for domain relation extraction: a survey. Journal of Data Science and Intelligent Systems, 2023, 1(2): 65-82. DOI: 10.47852/bonviewJDSIS3202973.
[16] Hazarika B B, Gupta D. Improved twin bounded large margin distribution machines for binary classification. Multimedia Tools and Applications, 2023, 82(9): 13341-13368. DOI: 10.1007/s11042-022-13738-7
[17] Yan X, Lin Z, Lin Z, Vucetic B. A novel exploitative and explorative GWO-SVM algorithm for smart emotion recognition. IEEE Internet of Things Journal, 2023, 10(11): 9999-10011. DOI: 10.1109/JIOT.2023.3235356
[18] Yang J Q, Lin N T, Zhang K, Cui Y, Fu C, Zhang D. Deep learning CNN-APSO-LSSVM hybrid fusion model for feature optimization and gas-bearing prediction. Petroleum Science, 2024, 21(4): 2329-2344. DOI: 10.1016/j.petsci.2024.02.012
[19] Tan S, Zhao S, Wu J. QL-ADIFA: Hybrid optimization using Q-learning and an adaptive logarithmic spiral-levy firefly algorithm. Mathematical Biosciences and Engineering, 2023, 20(8): 13542-13561. DOI: 10.3934/mbe.2023604
[20] Ju H, Yi H. Improved fuzzy sparse multi-class least squares support vector machine. Journal of Intelligent & Fuzzy Systems, 2023, 45(5): 7769-7783. DOI: 10.3233/JIFS-231738