Application of Credit Card Fraud Detection Based on CS-SVM

Chenglong Li, Ning Ding, Haoyun Dong, and Yiming Zhai

***Abstract*—With the development of e-commerce, credit card fraud is also increasing. At the same time, the way of credit card fraud is also constantly innovating. Support Vector Machine, Logical Regression, Random Forest, Naive Bayes and other algorithms are often used in credit card fraud identification. However, the current fraud detection technology is not accurate, and may cause significant economic losses to cardholders and banks. This paper will introduce an innovative method to optimize the support vector machine by cuckoo search algorithm to improve its ability of identifying credit card fraud. Cuckoo search algorithm improves classification performance by optimizing the parameters of support vector machine kernel function (C, g). The results demonstrate that CS-SVM is superior to SVM in Accuracy, Precision, Recall, F1-score, AUC, and superior to Logistic. Regression, Random Forest, Decision Tree, Naive Bayes, whose accuracy is 98%.**

***Index Terms*—Credit card fraud, fraud detection technique, SVM, CS.**

I. Introduction

Credit card fraud increases as ecommerce becomes more prevalent. [1] According to Robertson [2], global credit card fraud losses increased from $7.6 billion in 2010 to $21.81 billion in 2015. By 2020, global credit card fraud losses are expected to reach $31.67 billion.

However, current fraud detection techniques are far from accurate, and can result in significant financial losses to merchants and card issuers. With the advancement of fraud detection technology, fraudsters are constantly improving the concealment of fraud and avoiding being discovered. Credit card fraud detection methods are divided into two categories: supervised and unsupervised. In the supervised fraud detection method, models are estimated based on samples of fraud and legitimate transactions, and new transactions are classified as fraudulent or legal. In unsupervised fraud detection, outliers or unusual transactions are identified as potential fraudulent transaction cases. Both methods of fraud detection can predict the likelihood of fraud in any given transaction [3]. Support Vector Machines [4], Logistic Regression [5], Random Forest [6], Naive Bayes [7] and other algorithms are often used in credit card fraud detection. Therefore, credit card fraud detection methods need to be continuously innovated to improve the accuracy of fraud detection. Support Vector Machine is a supervised machine learning algorithm for data classification problems. It is widely used in many fields, such as image recognition [8], credit evaluation [9], public safety [10] and so on. Although the support vector machine has achieved good results in credit card fraud detection, the classification performance of the support vector machine will be greatly affected when dealing with high-dimensional noisy input data [11].

The main purpose of this paper is to discuss in depth the performance of support vector machines in credit card fraud detection. Compared to other classifiers, SVM can solve linear and nonlinear binary classification problems, which finds a hyperplane that distinguishes the input data in the support vector. The classification performance of support vector machine is mainly affected by the parameters of kernel function [12]. Support Vector Machine mainly searches for parameters by grid search method. The parameters obtained by this grid search method are not the optimal solution, because it is easy to fall into the local optimal solution. Cuckoo search algorithm combines large step with small step to find the optimal solution. This search method can effectively avoid falling into local optimum. Therefore, the cuckoo search algorithm can improve the classification performance of support vector machine by optimizing the parameters of support vector machine.

The basic framework of the paper is divided into four parts. Section II describes the three data mining techniques employed in this study. Section III discusses the experimental set up and presents our results. Section IV contains a discussion on findings and issues for further research.

II. Modelling

A. *Support Vector Machine (SVM)*

SVM is an excellent machine learning tool for pattern classification and regression that minimizes both prediction error and model complexity. [13] The SVM is based on formalized classification boundaries which are separated by points with different labels, thereby maximizing the boundaries of the closest data points. The classification boundaries defined by the hyperplane will result in different support vectors.

The support vector machine was originally proposed to study the linear separability problem, assuming a training set of size (x_i, y_i) , $i = 1,2,3,\cdots,l$, $x \in R^n$, $y \in \{+1,-1\}$, l is the number of samples, and n is the input dimension. When linearly separable, the optimal classification hyperplane is:

Manuscript received November 18, 2019; revised August 14, 2020.

Chenglong Li and Yiming zhai were with the School of Criminal Investigation and Forensic Science, People's Public Security University of China, China (e-mail: chenglong_li666@163.com, yiming_zhai@126.com).

Ning Ding is with the School of Criminal Investigation and Forensic Science, People's Public Security University of China, China (Corresponding author; e-mail: dingning_thu@126.com).

Haoyun Dong was with the School of Law and Criminology, People's Public Security University of China, China (e-mail: 541054114@ qq.com).

doi: 10.18178/ijmlc.2021.11.1.1011

$$\omega x + b = 0 \tag{1}$$

At this time, the classification interval is $\frac{2}{\|\omega\|}$, and it is obvious that when $\|\omega\|$ takes the minimum value, the classification interval is the largest. Classification problems can be described as solving the following constrained optimization problems:

$$\begin{cases} \min \frac{\|\omega\|^2}{2} \\ s.t. y_i(\omega x_i + b) - 1 \geq 0, i = 1,2,3,\cdots,l \end{cases} \tag{2}$$

It is worth mentioning that if the majority of samples in the data set are linearly separable, only a few samples (possibly abnormal points) lead to the failure to find the optimal classification hyperplane. For such cases, the usual practice is introduced non-negative slack variables $\xi_i, i = 1,2,3,\cdots,l$, and correct the optimization objectives and constraints, namely:

$$\begin{cases} \min \frac{\|\omega\|^2}{2} + C\sum_{i=1}^{l} \xi_i \\ s.t. \begin{cases} y_i(\omega x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0 \end{cases}, i = 1,2,3,\cdots,l \end{cases} \tag{3}$$

In formula (3), C is a penalty factor, which plays an important role in controlling the degree of penalty of the wrong sample, thus achieving a compromise between the proportion of the wrong sample and the complexity of the algorithm. The larger C, the greater the possibility for misclassification. By solving the above optimization problem by Lagrange multiplier method, the optimal function can be obtained as follows:

$$f(x) = \text{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i (x \cdot x_i) + b\right] \tag{4}$$

In the formula (4), α is a Lagrangian coefficient. When testing the input test sample x, the category of x is determined by formula (4). According to the $\mathrm{K-T}$ condition, the solution to the above optimization problem must satisfy:

$$\alpha_i(y_i(\omega x + \mathrm{b}) - 1) = 0 \tag{5}$$

Therefore, for most samples α_i will take a value of zero, only the support vector machine α_i is not zero, they usually occupy a small proportion of the total sample. In this way, only a small number of support vectors are needed to complete the correct sample classification.

In the case of nonlinear classification problems, the support vector machine maps the samples to a high-dimensional space H by the kernel function K ($x_i \cdot x_j$) , and then classifies the original problem in H. The process and method of finding the optimal classification hyperplane in the high-dimensional feature space is similar to the linear separable SVM case, except that the dot product in the high-dimensional feature space is replaced by the kernel function, thereby greatly reducing the computational complexity. According to the Mercer condition, the corresponding optimal decision function becomes:

$$f(x) = \text{sgn}\left[\sum_{i=1}^{l} y_i \alpha_i K(x \cdot x_i) + b\right] \tag{6}$$

The liner SVM only uses the liner kernel function $\mathrm{K}(x_i, x_j) = x_i x_j$. In the nonlinear SVM, we have many options where Radial basis function is widely used:

$$K(x_i, x_j) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{7}$$

where $\sigma \neq 0$ is the kernel function parameter. In this paper, we choose the Radial basis function (RBF).

B. *Cuckoo Search Algorithm (CS)*

The Cuckoo Search algorithm is a new optimization algorithm proposed by scholars Yang and Deb from University of Cambridge in 2009 [14]. The natural process of the cuckoo nesting parasitization is simulated, the parameters of the problem to be solved are compiled into a nest, and multiple nests constitute a population. Individuals in the population update the population by selecting the bird's nest by Levy flight and discarding the bird's nest with a certain probability. After several iterations, until the optimal solution is obtained.

Fig. 1. Cuckoo.

To simplify the description of the new CS [15], we now use the following four idealization rules:

Each cuckoo bird lays an egg standing for a design solution at a time, and dumps its egg in the nest randomly chosen from hosts.

The best nests with high quality eggs (better solution) will be passed to the next generation.

The number of available host nests is limited to n, and a host bird can recognize the egg of cuckoo bird with a probability $p_a \in [0,1]$.

In this case, it can either throw the egg away or abandon the nest in order to build a completely new nest in a new location.

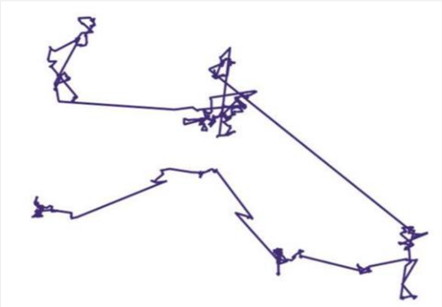

Fig. 2. Levy flight.

The path and location update formula for the cuckoo nest is as follows:

$$x_i^{(t+1)} = x_i^{(t)} + \beta \oplus L(\lambda), i = 1,2,3,\cdots,n \tag{8}$$

In formula (8), $x_i^{(t)}$ represents the position of the i-th bird's nest in the t-th generation; $\oplus$ represents site-to-site multiplication; β represents the step control, which is used to control the step size, and its value obeys the normal state. Distribution; $L(\lambda)$ represents the Levy flight random search path, and $L \sim \zeta = t^{-\lambda}(1 < \lambda \leq 3)$, where ζ represents the random step size obtained by Levy flight. After the position update, the random number $\gamma \in [0,1]$ is compared with p_a. If $\gamma > p_a$, then $x_i^{(t+1)}$ is changed, and vice versa. Finally, the set of bird nest positions $y_i^{(t+1)}$ with excellent test values is retained. At this time, $y_i^{(t+1)}$ is still recorded as $x_i^{(t+1)}$, and refine the formula (8) to get the formula (9):

$$x_i^{t+1} = x_i^t + stepsize \cdot\times (\delta - best), \\ i = 1,2,3,\cdots,n \tag{9}$$

In formula (9), $stepsize$ represents the step size produced by Levy flight, δ represents the position of a certain nest, and $best$ represents the best position in the current nest.

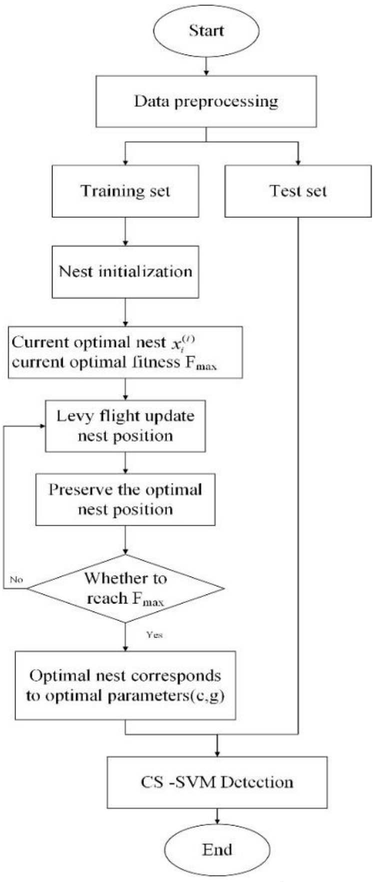
Start
Data preprocessing
Training set
Test set
Nest initialization
Current optimal nest $x_i^{(t)}$
current optimal fitness F_{max}
Levy flight update nest position
Preserve the optimal nest position
Whether to reach F_{max}
No
Yes
Optimal nest corresponds to optimal parameters(c,g)
CS -SVM Detection
End

Fig. 3. Optimization principle.

C. Construction of Fraud Detection Model: CS-SVM

In order to construct an effective SVM model, the parameters of the parameters (C and g) need to be pre-selected. The determination of parameter C requires a trade-off between training error and complexity. The higher the C, the greater the tolerance, and the over-fitting is easy; the smaller C is, the easier it is to fit, the C is too large or too small, and the generalization ability is poor. g is the parameter attached to the RBF function as a kernel function. The larger the g, the smaller the support vector, and the smaller the g value, the more support vectors. The number of support vectors affects the speed of training and prediction. Therefore, the parameters (C and g) have a significant impact on the efficiency and generalization of the SVM. The cuckoo search algorithm has excellent search capabilities, and combines large step with small step to find the optimal solution. we chose the cuckoo search algorithm to optimize the parameters of the SVM. Fig. 3 is the flow chart of CS-SVM principle. The CS-SVM implementation steps are as follows:

Step 1: Data preprocessing, establishing a training set and a test set.

Step 2: Determines the range of values of the SVM parameters c and g, the minimum step size $\mathrm{Step}_{\min}$ of the CS algorithm, the maximum step size $\mathrm{Step}_{\max}$, and the number of iterations N.

Step 3: Set the initial probability parameter p_a to 0.25, randomly generate the positions of n nests, and each nest corresponds to a set of parameters (C, g), calculate the fitness of each set of nest positions corresponding to the training set, and find the best bird nest at present. The position $x_i^{(t)}$ and the best fitness $\mathrm{F}_{\max}$.

Step 4: Retains the position $x_i^{(t)}$ of the optimal nest of the previous generation, and calculates the Levy flight step according to formula (8) and formula (9), and uses Levy flight to update the position of other nests to obtain a new set. Nest position and calculate their fitness F.

Step 5: According to the fitness F, the position of the new bird's nest is compared with the position of the previous generation bird's nest p_{i-1}, and the position of the bird's nest is replaced by a better bird's nest position to obtain a relatively new nest position.

Step 6: Compares the random number γ with p_a, preserves the nests with less probability of discovery in p_t, and updates the nests with higher probability of discovery, calculates the fitness of the new nest, and adapts to the position of the nest in p_t. For comparison, replace the poor position with a better nest position to obtain a new set of better nest position p_t.

Step 7: Finds the optimal nest position $x_i^{(t)}$ in step 6, and determines whether the fitness F satisfies the requirement. If the requirement is met, the search is stopped, and the global best fitness $\mathrm{F}_{\max}$ and its corresponding are output. The optimal nest $x_i^{(t)}$; if the requirements are not met, return to step 4 to continue searching.

Step 8: Performs parameter setting on the SVM according to the optimal parameters C and g corresponding to the optimal nest position $x_i^{(t)}$.

III. Results and Discussion

A. Preparing Data for Models

A publicly available data set can be downloaded from [16]. It included a total of 284,807 transactions made in September 2013 by European cardholders. The data set contains 492 fraud transactions, which is highly imbalanced. Due to the confidentiality issue, each piece of data contains 28 attributes that are privately processed. These 28 attributes are represented by V1, V2, ..., V28 respectively. In this paper, 5094 transaction data were extracted from 284074 transaction data for our research. Randomly scrambled data, and then 70% of the data were selected as training sets, a total of 3566, 30% as test sets, a total of 1528. The programs of CS-SVM algorithm were written by MATLAB R2017.

B. Evaluation Measures

Evaluation metrics include accuracy, precision, recall and F1-score. [17] The Confusion Matrix is an indicator of the results of the evaluation model and is part of the model evaluation. The confusion matrix consists of the following measures:

True Positive (TP): A test result that detects the condition correctly when the condition is present.

True Negative (TN): A test result that does not detect the condition when the condition is absent.

False Positive (FP):A test result that detects the condition when the condition is absent.

False Negative (FN): A test result that does not detect the condition when the condition is present.

The various evaluation measures are defined as follows:

Accuracy: It is the number of correct predictions made divided by the total number of predictions made.

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+FP+TN} \tag{10}$$

Precision: It is the number of positive predictions divided by the total number of positive class values predicted.

$$\text{Precision} = \frac{TP}{TP+FP} \tag{11}$$

Recall: It is the number of positive predictions divided by the number of positive class values in the test data.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{12}$$

F1-score: The F1-score conveys the balance between the precision and the recall.

$$F1-\text{score} = \frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{13}$$

ROC curve: Receiver operating characteristic curve, referred to as ROC curve, is the horizontal axis of False positive rate, the probability of hitting the vertical axis, and the curve drawn by the tester under different stimulation conditions due to different judgment criteria.

AUC represents the area under the ROC curve, between 0.5 and 1. For a perfect classifier, the value of AUC should be 1. AUC as a numerical value to visually evaluate the quality of the classifier. The larger the AUC value, the better the classification effect.

C. Classification Performance Comparison

Selecting the logarithm of Logistic Regression, Random Forest, Decision Tree, Naive Bayes and SVM for performance comparison, we can find that CS-SVM has remarkable effect in Accuracy, Precision, Recall, F1-score and AUC. Compared with SVM, Accuracy increased by 6%, Precision increased by 5%, Recall increased by 10%, F1-score increased by 3%, and AUC increased by 10%, indicating that CS-SVM optimization is remarkable and has an advantage in identifying credit card fraud. The performance comparison of various classification models is shown in Table I.

TABLE I: MODEL COMPARISON

Models	Accuracy	Precision	Recall	F1-score	AUC
Logistics Regression	94%	93%	92%	93%	0.92
Random Forest	93%	92%	92%	92%	0.92
Decision Tree	92%	92%	91%	91%	0.89
Naive Bayes	94%	94%	91%	93%	0.91
SVM	92%	93%	89%	90%	0.88
CS-SVM	98%	98%	99%	93%	0.98

Fig. 4 is CS-SVM fitness curve. CS-SVM iteratively faster, achieving the best fitness in the 15th generation. Fig. 5 is CS-SVM accuracy. CS-SVM classification accuracy can reach 98.626%. Fig. 6 is CS-SVM confusion matrix. Fig. 6 shows the distribution of four evaluation data of TP, TN, FP, and FN. In summary, CS-SVM is the best performer among the six classifiers.

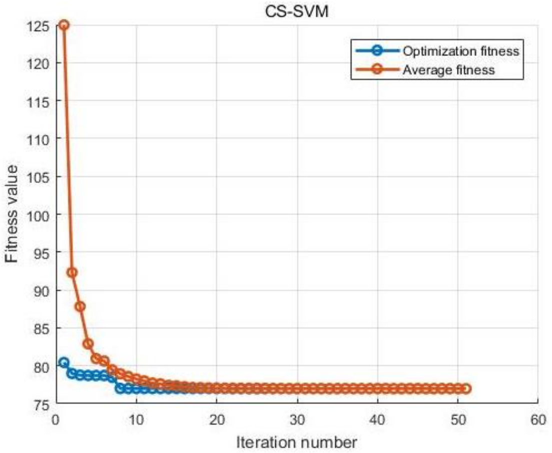

Fig. 4. CS-SVM fitness curve.

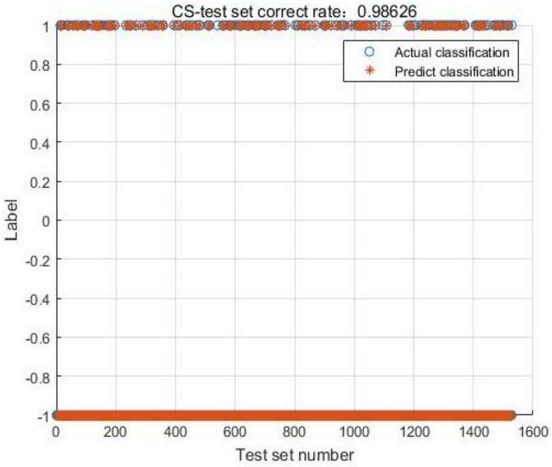

Fig. 5. CS-SVM accuracy.

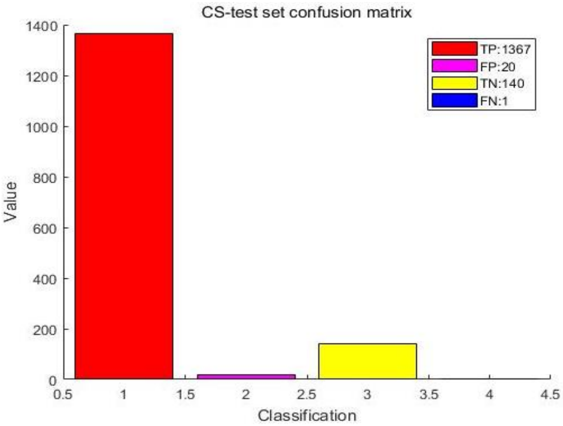

Fig. 6. CS-SVM confusion matrix.

IV. Conclusion

When dealing with high-dimensional, noisy credit card fraud data, ordinary SVM does not present the best classification effect. This paper uses the cuckoo search algorithm to optimize the parameters of SVM to improve the classification performance of SVM. Using logistic regression, random forest, decision tree, naive Bayes, support vector machine to detect credit card fraud data, their accuracy rate is 94%, 93%, 92%, 94%, 92% respectively. However, the accuracy of CS-SVM is 98%, which is the highest. It is superior to SVM in Accuracy, Precision, Recall, F1-score and AUC. It also has advantages, compared with other classification models, such as Logistic Regression, Random Forest, Decision Tree, Naive Bayes, etc. In the future, CS-SVM will be applied to dynamic credit card data monitoring systems to improve the ability to monitor credit card fraud. Future research may focus on the difference in the order of fraud and legitimate transactions before credit cards are taken away [18]. Future research may also examine the fraud differences between different types of fraud, such as the behavior differences between stolen and counterfeit cards.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Contributions

Chenglong Li and Ning Ding conceived the idea of the study; Haoyun Dong interpreted the results; Yiming Zhai analyzed the data; Chenglong Li wrote the paper; all authors discussed the results and revised the manuscript.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (71904194) and National Key R&D Program of China (No. 2018YFC0822800).

References

[1] D. Wang, B. Chen, and J. Chen, “Credit card fraud detection strategies with consumer incentives,” *Omega*, vol. 88, pp. 179-195, 2019.

[2] D. Robertson. (2016). The Nilson report. [Online]. Available: https://www.nilsonreport.com/upload/content_promo/The_Nilson_Report_10-17- 2016.pdf

[3] S. Jha, M. Guillen, and J. C. Westland, “Employing transaction aggregation strategy to detect credit card fraud,” *Expert Systems with Applications*, vol. 39, no. 16, pp. 12650-12657, 2012.

[4] A. Chaudhuri, “Modified fuzzy support vector machine for credit approval classification,” *AI Communications*, vol. 27, no. 2, pp. 189-211, 2014.

[5] Y. Sahin and E. Duman, “Detecting credit card fraud by ANN and logistic regression,” in *Proc. 2011 International Symposium on Innovations in Intelligent Systems and Applications*, 2011, pp. 315-319.

[6] C. Liu, Y. Chan, K. S. H. Alam *et al*., “Financial fraud detection model: Based on random forest,” *International Journal of Economics & Finance*, vol. 7, no. 7, pp. 178-188, 2015.

[7] L. Mukhanov, “Using Bayesian belief networks for credit card fraud detection,” in *Proc. the IASTED International Conference on Artificial Intelligence and Applications*, Insbruck, Austria, 2008, pp. 221-225.

[8] S. Ajina, R. V. Yampolskiy, and N. E. B. Amara, “Evaluation of SVM classification of avatar facial recognition,” in *Proc. 8th International Symposium on Neural Networks*, 2011.

[9] C. J. Fu and Y. P. Yang, “A batch-mode active learning SVM method based on semi-supervised clustering,” *Intelligent Data Analysis*, vol. 19, no. 2, pp. 345-358, 2015.

[10] K. Kianmehr and R. Alhajj, “Effectiveness of support vector machine for crime hot-spots prediction,” *Applied Artificial Intelligence*, vol. 22, no. 5, pp. 433-458, 2008.

[11] M. Zareapoor and P. Shamsolmoali, “Application of credit card fraud detection: Based on bagging ensemble classifier,” *Procedia Computer Science*, vol. 48, pp. 679-685, 2015.

[12] X. Z. Li and J. M. Kong, “Application of GA–SVM method with parameter optimization for landslide development prediction,” *Natural Hazards and Earth System Sciences*, vol. 14, no. 3, pp. 525-533, 2014.

[13] Y. Xue, L. Zhang, B. Wang *et al*., “Nonlinear feature selection using Gaussian kernel SVM-RFE for fault diagnosis,” *Applied Intelligence*, vol. 48, no. 10, pp. 3306-3331, 2018.

[14] X. S. Yang and S. Deb, “Cuckoo search via Lévy flights,” in *Proc. 2009 World Congress on Nature & Biologically Inspired Computing*, IEEE, 2009, pp. 210-214.

[15] K. N. Devi, V. M. Bhaskaran, and G. P. Kumar, “Cuckoo optimized SVM for stock market prediction,” in *Proc. 2015 International Conference on Innovations in Information, Embedded and Communication Systems*, 2015, pp. 1-5.

[16] R. Saia and S. Carta, “Evaluating credit card transactions in the frequency domain for a proactive fraud detection approach,” *SECRYPT*, pp. 335-342, 2017.

[17] K. Randhawa, C. K. Loo, M. Seera *et al*., “Credit card fraud detection using AdaBoost and majority voting,” *IEEE Access*, vol. 6, pp. 14277-14284, 2018.

[18] S. Bhattacharyya, S. Jha, K. Tharakunnel *et al*., “Data mining for credit card fraud: A comparative study,” *Decision Support Systems*, vol. 50, no. 3, pp. 602-613, 2011.

Chenglong Li was born in China in 1995. He got his bachelor in law from National police University of China in 2014 and he continues to pursue a graduate degree at this university now. His research interests include most crime analysis subjects and machine learning. He has published 3 articles in national journals.

Ding Ning received his bachelor's degree and master's degree in engineering from Harbin Institute of Technology in 2009 and 2011, and PhD in engineering from the Department of Automation of Tsinghua University in 2015. Since graduation, he has worked at the School of Criminal Investigation and Forensic Science, People's Public Security University of China. Also, he is a member of Public Security Behavioural Science Lab. His main research interests are crowd evacuation, emergency response, system modeling and simulation. At present, more than his ten international academic papers have been published, of which 8 papers were retrieved by SCI, 1 paper was retrieved by SSCI, and 6 papers were retrieved by EI.

Haoyun Dong was born in Shannxi province. She got her bachelor degree in both Ulster university and National police university of China in 2017, major in economic crime investigation and master degree in university of Glasgow in 2018, major in criminology. She is doing PhD in People's Public Security University of China. Her research is related to the crime in cult and policing. She published 5 articles in national journals.

Yiming Zhai was born in Henan province, China in 1997. He got his bachelor in law from People's Public Security University of China, Beijing, China in 2015 and continues to pursue a graduate degree at this university now. His research interests include most crime analysis subjects and machine learning. He has published an EI article and an SSCI article with his mentor now.