A Double layered Crime Governance Model Based on BP Neural Network and TF-IDF Algorithm

Zhijie Sun [1], Zehao Su [2], Yuyu Zhang [1], Rui Hou [1, *]

[1] Policing Information Department, Shandong Police College, Jinan, China, 250200

[2] Economic Crime Investigation Department, Shandong Police College, Jinan, China, 250200

* Corresponding Author Email: hr@sdpc.edu.cn

Abstract. The main methods of crime governance are pre prevention and post crackdown. With the development of intelligence and information technology, public security work has also ushered in new opportunities and challenges. This paper used intelligent algorithms and combines crime characteristics with existing big data resources in the public security system to propose a multidimensional facial recognition model based on TF-IDF algorithm, and combines it with a crime warning model based on BP neural network to form a closed crime governance chain of pre crime prevention and post crime investigation. This dual guarantee of crime governance is of great significance in reducing crime incidence, improving case solving accuracy and efficiency, enhancing people's sense of security, and ensuring social security and stability.

Keywords: Double layered crime governance model, Intelligence, BP neural network, TF-IDF algorithm.

1. Introduction

Traditional police work mainly focuses on cracking down, while modern police work must combine prevention and control, prioritize prevention, and transform passive policing into active policing. The current rapid development of intelligent technology can inject intelligent technology into proactive policing models, develop intelligent crime warning systems, automatically identify vulnerable groups, proactively prevent crime, achieve pre warning, accurately combat and effectively manage social instability factors, and ensure social security order.

In the practice of investigating and solving cases, public security organs face some difficulties in facial recognition. One is that the monitoring coverage is wide, and the proportion of facial recognition front-end analysis probes is low. The second reason is that there are more results from facial screening compared to post screening, and the efficiency of manual screening is low. Thirdly, there are abundant big data resources and low utilization of model screening. In the context of informatization and intelligent public security construction, the current public security system has collected rich big data resources [1]. However, in the screening of facial recognition results, modeling thinking is still lacking, and the utilization rate of using big data to achieve facial recognition results is still relatively low.

This paper establishes an intelligent crime warning and multi-dimensional facial recognition system for pre prevention and post disposal, forming a closed crime governance chain for crime prevention and disposal, reducing crime incidence, improving case solving accuracy and efficiency. Using the various behavioral characteristics of former criminal offenders before committing crimes as independent variables and the criminal results as dependent variables, input them into the BP neural network model for training, so that the model can automatically recognize behaviors similar to criminal characteristics, thereby achieving the goal of predicting crimes.

2. Preliminary

2.1. BP neural network

BP (back propagation) neural network is a concept proposed by scientists led by Rumelhart and McClelland in 1986 [2]. BP neural network is a multi-layer backpropagation network that uses algorithms to reverse change parameters by calculating costs. Calculate the cost of each training cycle multiple times, and then change the weight parameters and bias values to minimize the cost while training. BP neural network is a feedforward neural network with error backpropagation, consisting of three layers: input layer, hidden layer, and output layer [3]. The algorithm process is mainly divided into two stages: forward propagation and backward propagation of signals. The forward propagation stage is the process in which data passes through the hidden layer from the input layer and finally reaches the output layer; The backpropagation stage is the process of adjusting the weights and biases from the output layer to the hidden layer, and finally back to the input layer, and then sequentially adjusting the weights and biases from the hidden layer to the output layer, and from the input layer to the hidden layer [4-6]. The combination of the two processes has achieved the goal of improving the accuracy of algorithm training. Artificial neural networks have a certain degree of self-organization and adaptability, as well as non-linear mapping ability and a certain degree of fault tolerance, making them particularly suitable for solving problems with complex internal mechanisms. And after training, neural networks have the ability to apply learning outcomes to new knowledge, which is the advantage of neural network algorithms.

2.2. TF-IDF algorithm

TF-IDF [7-9], also known as word frequency reverse file frequency, is a weighting technique commonly used in information retrieval and text mining to evaluate the importance of a word to this document. The importance of words is related to their frequency in the document, the more times they are used, the more important they are, and vice versa. The main idea of TF-IDF is that the importance of words increases proportionally with the number of times they appear in the file, but at the same time decreases inversely with the frequency of their appearance in the corpus. So, it can be considered that this word has good distinguishing ability and is more suitable for classification.

Drawing inspiration from the TF-IDF algorithm, transfer and apply the TF-IDF algorithm to facial recognition. The more suspected individuals have a criminal record, the greater the likelihood of committing a crime. Therefore, the possibility of being locked as a suspect increase in direct proportion to the number of criminal records, and the characteristics of criminal records are TF like factors. Most people in the city cannot commit crimes, so the possibility of being locked as a suspect is inversely proportional to the frequency of the behavior characteristics of suspects in the total urban population.

3. Experiment

3.1. Feature data selection and data processing

1. Obtain criminal records of former offenders as the training result dataset for crime warning analysis; Compare the facial information of the suspects captured at the scene through a facial recognition system, select a collection of suspected individuals, and organize the data as a dataset for facial analysis.

2. Select crime characteristic data.

(1) Hotel accommodation data characteristics: The hotel accommodation characteristics of suspect may be different from those of ordinary people, because they have no fixed residence or are preparing for crimes, or stay alone for too long, or stay alone frequently. Therefore, collect the hotel accommodation information of former and suspected personnel in the month of the incident, screen the number of people who have stayed in hotels for a similar duration and number of stays as

suspected personnel in the month of the incident, and collect the number of stays in all hotels in the jurisdiction in the month of the incident.

(2) Data characteristics of Internet cafes: The behavior characteristics of suspect going to Internet cafes may be different from that of ordinary people. They may frequently go to and from Internet cafes but the duration of Internet surfing is too short or the duration of Internet surfing is too long. Therefore, collect the internet cafe internet access information of former and suspected personnel in the month of the incident, screen the number of people who have similar internet access hours and times as suspected personnel in the month of the incident, and collect the internet access numbers of all internet cafes in the jurisdiction in the month of the incident.

(3) Railway data characteristics: the railway travel characteristics of suspect may be different from those of ordinary people, and the time interval of railway travel may be too long or too short. Therefore, collect railway travel information of former and suspected personnel in the month of the incident, screen the number of people with similar railway travel time intervals as suspected personnel in the month of the incident, and collect the number of railway travel in the city in the month of the incident.

(4) One code access data collection characteristic: The suspect's one code access data collection characteristics may be different from ordinary people, and may frequently go to various places due to criminal needs or almost do not appear in public places due to fear of exposure. Therefore, collect information on the use of QR codes by individuals with previous criminal records and suspected individuals in the month of the incident, screen the number of people who use the QR code similarly to those of suspected individuals in the month of the incident, and collect the total population in the jurisdiction in the month of the incident.

(5) Characteristics of nucleic acid collection data: The suspect may not do nucleic acid regularly as required due to fear of exposure, or frequently travel across regions due to criminal needs, due to too many nucleic acid tests for epidemic prevention and control. Therefore, collect the number of nucleic acid tests conducted by former and suspected personnel in the month of the incident, screen for the number of people who have similar nucleic acid tests as suspected personnel in the month of the incident, and collect the number of people who have undergone accounting tests in this city in the month of the incident.

(6) Characteristics of criminal record data: Individuals with criminal records are much more likely to commit crimes again than those without criminal records, and the more times they commit crimes, the greater the likelihood of committing crimes again. Therefore, the number of times to query the criminal records of suspected individuals and collect the total number of criminal categories.

3.2. A Crime Warning Model Based on BP Neural Network

The dataset used in this article is from a public security data platform in a certain city, with a total of 3747 original data, including 2417 normal personnel and 1330 suspected criminal records. Each person's information includes basic information data such as hotel accommodation duration, hotel check-in frequency, internet cafe internet usage, internet cafe internet usage, railway travel frequency, travel time interval, one code access data collection, nucleic acid testing frequency, etc.

By inputting the five trajectory characteristics of former convicts, including railways, hotel accommodations, internet cafes, one code access collection, and nucleic acid information collection, as independent variables into the neural network, and outputting whether they committed a crime as the outcome variable, through the training of the BP neural network, a BP neural network with six feature indicators as variables can be obtained.

To ensure the accuracy of the results, approximately 87% of the data was randomly selected as the training set to train the neural network, and then approximately 13% of the data was randomly selected as the testing set to verify the accuracy of the neural network's calculations. The final result of training is shown in Table 1.

Table 1. Model training results

Sample	Test	Prediction		
		0	1	Rate
Train	0	1482	153	90.6%
	1	225	724	76.3%
	Gross	66.1%	33.9%	85.4%
Variant	0	273	21	92.9%
	1	36	107	74.8%
	Gross	70.7%	29.3%	87.0%

According to the accuracy percentage of the test set, it can be seen that the final trained BP neural network achieved an accuracy of 87% in crime prediction.

4. A Multidimensional Face Recognition Model Based on TF-IDF

Based on the facial image of a crime scene, the obtained facial information is compared using a facial recognition system, and a set of comparison data is extracted for data organization as the dataset for facial analysis.

The pre-existing features are TF-like factors, while the trajectory features include five types: railway, hotel accommodation, internet cafe internet access, one code access collection, and nucleic acid information collection, which are IDF-like factors [9]. Therefore, five TF-IDF weights need to be calculated.

To prevent the situation where the TF factor is 0 due to suspected personnel having a history record of 0, the minimum number of history records is set to 1.

$$tf = \frac{1+n_6}{N_6} \tag{1}$$

When the sum of squared differences between a certain characteristic value of a person and a suspected person is less than or equal to 1, it is considered that the person is similar to a certain characteristic of a certain trajectory of a suspected person, and m features need to be met (the number of trajectory features in this paper is 1 or 2).

$$idf = log\frac{N_j}{1+n_j} = log\frac{N_j}{1+\sum_{i=1}^{n} y_i} \ (j = 1,2,...,5; i = 1,2,...,n) \tag{2}$$

Calculate TF-IDF weights [10].

$$w_j = tf \times log\frac{N_j}{1+n_j} = \frac{1+n_6}{N_6} \times log\frac{N_j}{1+\sum_{i=1}^{n} y_i} \ (j = 1,2,...,5; i = 1,2,...,n) \tag{3}$$

The higher the comprehensive score, the more likely the suspected person is to be locked as a suspect. It can be seen from the above figure that the difference between the comprehensive scores of the first two suspects and those of other suspects is large, so the first two suspects are most likely to be suspect.

5. Conclusion

This article applies intelligent models and algorithms to crime prevention and investigation, selecting five trajectory features that are prone to prominent abnormal behaviors, including railways, hotel accommodations, internet cafes, One Code Collection, and nucleic acid information collection, and associating them with the fusion trajectory features of previous offenders. Based on the BP neural network model and TF-IDF algorithm, a crime warning and multi-dimensional facial recognition model is established, forming a closed crime governance chain of pre crime prevention and post crime investigation, and dual governance of crime. Compared to traditional crime governance methods, it has the advantages of new methods, wide dimensions, high accuracy, and high efficiency, and the model parameters can be dynamically updated in real-time, intelligently adapting to changes in the crime situation. This paper finds that the accuracy of crime prediction can be increased to 87% through intelligent algorithms, and the efficiency and accuracy of face recognition can be greatly improved, which provides a strong guarantee for "accurate" targeting of suspect. And with the increase of data volume and feature indicators, accuracy and efficiency can be further improved.

Acknowledgement

The authors gratefully acknowledge Key project of Undergraduate Teaching Reform in Shandong Province: "Research on Talent Training Model of 'Smart Governance +'" (Project number: Z2023060) funds.

References

[1] Wang L, Tang X, Fan J, et al. Prediction of firing accuracy of self-propelled antiaircraft gun in marching fire based on GA-BP neural network [J]. IOP Publishing Ltd, 2023.

[2] Yang M, Huang H, Li S, et al. A BP Neural Network-Based Method for Evaluating the Quality of Creative Education in Minority Regions [J]. Journal of Circuits, Systems and Computers, 2023, 32 (16).

[3] Mendrofa G A, Handari B D, Hertono G F. Ensemble learning model on Artificial Neural Network - Backpropagation (ANN-BP) architecture for coal pillar stability classification [J]. 2024.

[4] Yu J Y, Chen J. Research on Evaluation Model of Entrepreneurship Education Based on BP Neural Network [J]. Journal of Information & Knowledge Management, 2023.

[5] Wu L, Zhang Y, Wei Y, et al.A BP Neural Network-Based GIS-Data-Driven Automated Valuation Framework for Benchmark Land Price [J]. Complexity, 2022, 2022.

[6] Ji Ce Z X. Channel Estimation Algorithm of OFDM System Based on GSA-BP Neural Network[J]. Journal of Northeastern University (Natural Science), 2022, 43 (6): 769 - 775.

[7] Li Q, Su X, Xie X, et al. Accuracy Analysis of Error Compensation in the Ionospheric Model of BDS Broadcasting Based on ABC-BP Neural Network [J]. 2022.

[8] Chai Z L, Zhou C L, Guo R, et al. Research on the Model of Scale Removal under Slurry Impact Based on BP Neural Network [J]. JOM, 2024 (1): 76.

[9] Zhu Y, Wang C. Study on Compensation of Blood Oxygen Sensor's Ambient Light Based on GA-BP Neural Network [J]. IOP Publishing Ltd, 2023.

[10] Wu Y C, Zhang G M. Design of a prediction model based on improved BP neural network and particle swarm optimization for more accurate budget of biogas production [J]. IOP Publishing Ltd, 2023.