Application of Grid Management in Spatio-temporal Prediction of Crime*

Zhang Tianyi, Ran Yibing, Wei Dong

***Abstract*—For traditional data-driven modeling method is not applicable to predict possibility of a crime under the complex inner link problems, put forward a kind of based on grid management concept, a new method of crime prediction system optimization modeling, the method focuses on the grid of geographical features, implementation of crime time, location, weather data, the network platform of data characteristics, such as the existence of the internal correlation analysis. This article in the Chicago area as the research object, by using the BP neural network algorithm modeling, and crime in the pretreatment process data into three-dimensional space-time grid model, establish units within the grid and relations between adjacent grid data, according to the grid position and distance to determine the influence weight, according to the weight integrate the input data set as a model. The experimental results show that compared with the traditional meshless modeling effect, the MAE error of the four types of cases on the 50*50 grid is reduced by about 0.03 on average, and the RMSE error on the 50*50 grid is reduced by about 0.09 on average. The research proves that the selection of appropriate algorithm on the basis of reasonable application of meshless management has a better effect.**

I. Introduction

As an important part of the city management system, the crime management system is a miscellaneous and interrelated system, and the probability prediction of all kinds of cases is an important branch of the crime prediction system, so how to reasonably and effectively realize the management prediction has become a hot research topic. With the deepening of economic reform, the number of various cases remains high. In order to support the scientific prevention and control of crime by the public security organs, it is urgent to use deep learning technology to promote the evolution and transformation of the crime prediction system. In order to comprehensively integrate and effectively utilize the precious crime resources, various crime prediction models emerge at the historic moment.

Case events forecast system of the main research content is based on the experience knowledge of investigators, using the history data, and involve the perpetrators of crime and the victim information as input parameters, such as training model to imitate human decision strategy, scientific analysis of crime situation changes, will predict the crime quantity changes and trends in applied to the daily work of the Ministry of Public Security, the auxiliary police awareness ahead of jurisdiction over a period of time in the future crime dynamic, timely detect the case and arrange for the deployment of police resources. In 2018, Sherry Towers I et al. [1] used Bootstrapping and linear regression method to analyze retrospective data and predict crime incidence, proving that all types of crime rates are dependent on special time points in a year, such as holidays and working days. Youngmi Lee et al. [2] used the mean equation of Poisson Inarch model to construct a regression analysis method including exogenous variables, and selected the least square estimation method to test the optimal function. The results showed that the casecases in Chicago followed the chi-square distribution without causal relationship. Renjie Liao, etc. [3] proposed an adaptive adjustment algorithm, applied to the Bayesian Learning (Bayesian Learning) framework of continuous crime prediction model, the model of each factor using discrete distance decay function to create a geographic profile [4], shows in a given geographical position could be a crime under the probability distribution of the location, so as to assist the public security organ to capture the intention of the perpetrator, and locate the next field of the scene of the crime. However, with the deepening of the research, researchers directly use the academic public crime data as the model input, and the lack of analysis of the internal relationship between cases has become more and more prominent, which has become one of the key problems limiting the improvement of performance. Researchers usually ignore the crime data, on the other hand, in the height of the global geographic area uneven distribution features, lead to "zero crime" in the input characteristics of proportion is great, and difficult to identify and extract crime characteristics, make the model prediction results sparse is serious, is the space and time scales of crime prediction model of the main bottlenecks. Therefore, it is urgent to strengthen the scientific management of case data.

Therefore, this paper draws on the operation mechanism of grid service management in the urban community governance model to propose a new optimization modeling method of crime prediction system based on grid management concept, which can alleviate the problem of insufficient correlation analysis between cases and help improve the model performance.

The main contributions of this paper are as follows:

a) Grid management methods used in crime events forecast model, on the one hand, innovative in using is different from

*Resrach supported by National Nature Science Foundation under Grant 61871020, Beijing Municipal University High-level Innovative Team Building Project under Grant IDHT20190506 and Beijing Municipal Education Commission Science and Technology Program Key Project under Grant KZ201810016019..

Zhang Tianyi Author is with School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China (corresponding author to provide phone: 13718187970; e-mail: zhangtianyi0126@126.com).

Ran Yibing Author is with School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China (e-mail: ranyibing@bucea.edu.cn).

Wei Dong Author is with School of Electrical and Information Engineering, Beijing University of Civil Engineering and Architecture, Beijing 100044, China (e-mail: weidong@bucea.edu.cn).

2021 33rd Chinese Control and Decision Conference (CCDC) | 978-1-6654-4089-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/CCDC52312.2021.9602394

the direct use of criminal geographic data to the method of the study area as a whole, to overcome the problems such as insufficient data by using, make the internal contact preserved more complete data, and forecast system for case event modeling method provides a new train of thought.

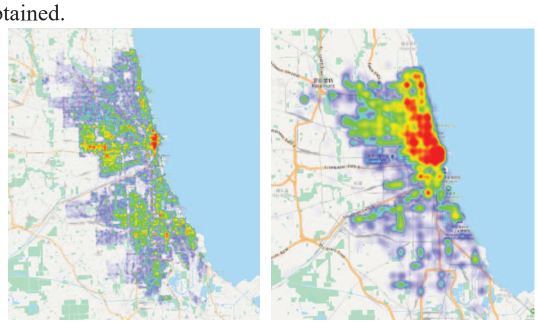
b) Introducing the spatiotemporal grid data model architecture, integrating time, space and crime types and other factors, discussing the geographic features of the grid, simulating the crime environment, building a more comprehensive prediction model, and at the same time improving the utilization rate of crime related data and the prediction accuracy of the model.

c) Assess meshing prediction models with different density degrees in combination with common error measurement standards. The results show that the spatiotemporal crime data prediction model based on grid geographic information management method can effectively realize the crime prediction in the study area. The prediction results of the model are affected by the mesh density, and the prediction effect will be better when a suitable mesh partition scale is selected.

II. Research Ideas and Case Analysis

A. Chicago Area Data Analysis

Chicago is located in the center of the North American continent, is one of the international financial centers, the high prosperity of the back breeds the hotbed of crime, there are a large number of criminal cases, the number of cases far higher than other regions. Therefore, this research adopts the Chicago area for 2016 crime as a data set, the data sets contain crime types, latitude and longitude, date, weather, yelp six hotspots and police stations location data attributes, among them, the yelp as America's largest review sites, the data include the Chicago area restaurants, shopping centers, hotels and tourism hotspot location, as shown in figure 1, the yelp data visualization, hotspot location and the performance of the crime scene to gather high similarity, look for one of forecasting model features were studied. In the data processing stage, missing and remaining data were deleted, and a total of 251,044 pieces of valid data were finally obtained.



(a) Crime data heat map (b) Yelp data heat map

Figure 1. Heat map of raw data

Figure 2 shows the number of events throughout the year all kinds of crime in the region distribution, easy to find theft, beating, criminal damage, attack case number, number of crime than other types of cases, the local public security organ should be attention to its, therefore this research will be focusing on the four types of high-risk cases were studied. Statistical mapping of the data will help you get a big picture of crime in the Chicago area over the course of 2016 and develop research strategies.

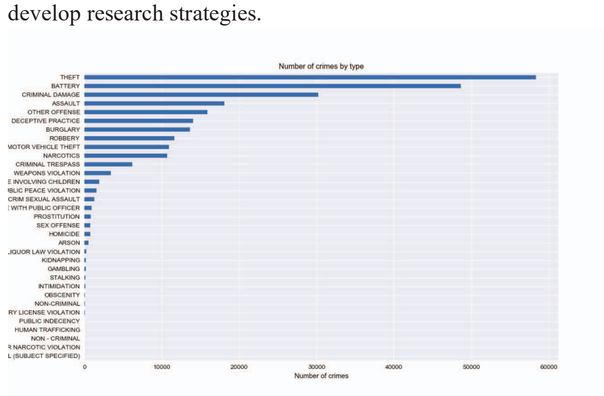


Figure 2. Statistics of the number of various cases in Chicago in 2016

B. Thoughts on Grid Management of Criminal Information

The occurrence of crimes is closely related to the location of police stations, hot spots with high population density and weather anomalies. However, such complex relationship is not linear and the case events are scattered over a wide area, so it is difficult and error to predict the possible occurrence of crimes. Criminologists believe that there are approximately repeated crime patterns in both time and space for criminal events, which differ from the repeated pattern in that the approximate repetition does not require that the characteristics of the second event, such as involving the victim or the location, are exactly the same as those of the previous one [5]. For example, in the case of burglary, a similar repeating phenomenon is shown in that houses located near the location of the burglary are also at higher risk of burglary in the coming days, weeks or months. The grid fine management of the whole region, screening out the spatio-temporal approximately repeated events and testing the aggregation have a positive effect on assisting the public security organs to realize the active discovery of cases, rapid decision making and deployment in advance.

III. Experiment and Analysis

A. Data Preprocessings

Will be covered with Chicago area map of latitude and longitude interval division of grid, such as with two diagonal points of latitude and longitude grid define the grid position, the number of grid for n^2, the last one by one, into the position of the position of crime, the police station in the grid, and USES the gaussian kernel density estimation method analysis gathering for the space of different types of cases, criminal intensity distribution area, to determine the influence of different distance between grid weight; In terms of time, considering the influence of recent repeated events on crime prediction, the data of the previous day and the previous week are taken as the unit of day to predict the probability of various kinds of cases occurring in the next day. The processed data form is shown in Figure 3. In the figure, only the data of different time dimensions in the same spatial grid is intercepted, where cell_range represents the coordinate values of different time dimensions obtained by the method of dividing the grid according to the above equal interval.

cell_range	timestamp
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-08
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-09
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-10
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-11
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-12
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-13
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-14
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-15
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-16
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-17
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-18
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-19
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-20
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-21
((41.4182385935059, -87.524528378), (41.5996124616157, -88.308961))	2016-01-22

Figure 3. Part of the data set after preprocessing

B. *Crime Prediction Model*

In the grid-based space design, the boundaries of Chicago area are first defined on the map, and then the areas are divided into 10*10, 20*20 and 50*50 grids to generate a grid-based urban map [9], and the number of four types of cases in each grid is counted. The sparse density of grid segmentation will lead to the change of data imbalance. The non-crime grid is generally located in the suburbs and mountainous areas with sparse population, as shown in Figure 4[10]. Deleting the empty grid can achieve the purpose of ensuring the accuracy of prediction. As shown in Fig. 5, different types of crimes are displayed in the three-dimensional network framework according to the two dimensions of time and space.

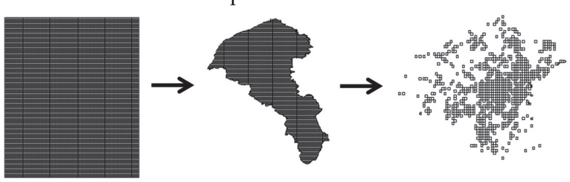


Figure 4. Process diagram of geospatial grid processing

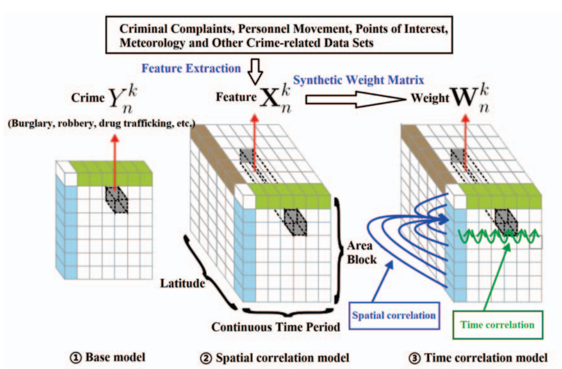


Figure 5. Time and space framework of crime prediction

Kernel Density Estimation is developed from histogram. It avoids subjectively bringing in prior knowledge and only uses the information of the data itself to display and visualize the crime aggregation in the spatial region by generating a smooth Estimation surface. Therefore, for any point x in a crime event in a certain spatial region, let x_i, i= 1,2,...,n is other points of the same type of crime within the scope of the region, and the distance from x_i to x determines the weight of other points in the region. Set the kernel function as $K(x)$, then the probability density of crime at point x is as follows:

$$\hat{f}(x) = \frac{1}{n}\sum_{i=1}^{n} K(\frac{x-x_i}{d}) \quad (1)$$

In the formula, d represents the bandwidth of the kernel density function, $K(x)$ is the kernel function, and n is the number of samples.

Back Propagation Neural Networks is a kind of error back propagation neural network, including the signal transmission of forward and reverse transmission two process, if the training error is greater than the expected error, depending on the training error by back propagation algorithm training of the input sample data to get the right to make changes in the other direction, after the output value and value, by error value cumulative repeatedly until you reach the initial goals after stop the iteration.

In this paper, a three-layer BPNN with good nonlinear mapping ability is used to model the four types of cases respectively. The input layer, hidden layer and output layer are all single-layer. The input layer contains nine input variables, namely, the number of neurons in the input layer is 9. The number of neurons in the input layer is 9. The number of neurons in the input layer is 9, and the number of neurons in the input layer is 9. The output layer only contains the probability value representing the possibility of crime, that is, the number of neuron nodes in the output layer is 1. Due to the node number of hidden layer neurons directly affect neural network fitting effect and generalization ability, using the "rule of thirds" algorithm [11] to determine the number of hidden layer nodes interval, in hidden layer construction of BPNN, using the test set test network model of BPNN error and the number of hidden layer nodes, locate the optimal solution through error trend research for 10 hidden layer nodes.

Since the time, place, weather, police station location and network platform data of the crime event are not the same type of parameters, the above indexes must be normalized and the input and output variables are mapped to the interval [0,1] by formula (2).

$$z_i^k = \frac{z_i^k - z_{i\min}}{z_{i\max} - z_{i\min}} \quad (2)$$

Where Z^k_i is the type i of original input data, and Z_{min} and Z_{max} are the maximum and minimum values of the type i of original input data. Substituting the input values of the normalized random training samples into the neural network, the hidden layer output value set Y_j is calculated by formula (3).

$$Y_j = f(\sum_{i=1}^{7} w_{ij}^{(m)} X_i + b_j^{(m)}) = \frac{1}{1+e^{-\sum_{i=1}^{4} w_{ij}^{(m)} X_i - b_j^{(m)}}} \quad (3)$$

In the formula, $w_{ij}^{(m)}$ is the connection weight between the input layer and the hidden layer; $b^{(m)}$ is the simple threshold of the hidden layer and the output layer; where the activation function of the hidden layer $f(x)$. The *logsig* function is selected, the excitation function of the output layer is set as a linear function, and the output value Z of the output layer is calculated using formula (4).

$$Z = \sum_{j=1}^{n} w_j^{(Z)} Y_j + b^{(Z)} \quad (4)$$

When the output values of all training samples are obtained, the mean square error is used to evaluate the training accuracy of the model. The mean square error can be obtained by formula (5).

$$MSE = \frac{1}{h}\sum_{k=1}^{h}(m^k - z^k)^2 \quad (5)$$

In the formula, h is the number of training samples, m^k is the measured equivalent linear crime value of the type k training sample; Z^k is the output value of the output layer of the type k training sample. When the training output value differs greatly from the measured value, the gradient descent method is used to back-propagate the error. The calculation formulas are shown in (6)-(7), until the target accuracy is reached, the training ends.

$$w_{(N+1)} = w_{(N)} - \alpha \frac{\partial MSE}{\partial w} \quad (6)$$

$$b_{(N+1)} = b_{(N)} - \alpha \frac{\partial MSE}{\partial w} \quad (7)$$

In the formula, $w_{(N)}$ and $b_{(N)}$ are the weights and thresholds between each connection layer, respectively; α is the network learning rate.

$$Z_i^k = z_i^k (Z_{i\max} - Z_{i\min}) + Z_{i\min} \quad (8)$$

After the training, the weights and thresholds of each connection layer are used to input the test samples into the neural network to obtain the predicted values of the test samples.

The specific algorithm application process of mesh parameter selection operation is as follows:

(1) The optimal parameter *Hidden_layer_sizes* was selected by cross-validation method, and the weight of the Solver optimized feature was adjusted.

(2) Standardized data processing to eliminate the adverse effects caused by singular sample data;

(3) Input the preprocessed data set D={police station impact factor, hot spot impact factor, daily mean temperature factor, lowest temperature factor, highest temperature factor, rainfall factor, snowfall factor, X crime factor of the previous day, X crime factor of the previous week};

(4) Select N value, and according to different case types, randomly select 30% samples from data set D as training set for training model.

(5) Repeat Step (4), change the size of N value, train the model, and compare the model performance indexes to analyze the influence of meshing on the prediction effect.

C. Evaluation Index

Whether grid management is really suitable for crime prediction system of public security organs and how to choose grid size need to be measured by model evaluation index. The following two indicators were used to evaluate the effect of grid management in the application of temporal and spatial prediction model of crime. Average Absolute Error (MAE) describes the difference between predicted and true values. The smaller the value, the higher the accuracy of the model. Root mean square error (RMSE), also known as standard error, can well reflect the precision of prediction, and is very sensitive to abnormal errors. The two index principles are shown in Equations (9) - (10).

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | \hat{y}_i - y_i | \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2} \quad (10)$$

Among them, y_i is the true value, and $\hat{y}_i$ is the predicted value. $MAE \in [0,+\infty)$, $RMSE \in [0,+\infty)$, $RMSE$=0 when the predicted value is completely consistent with the actual value, which is a perfect model.

D. Mesh Application Effect

TABLE I. BPNN Algorithm Analysis and Prediction Results in Chicago Area

(a) n=10

n=10	*theft*	*criminal damage*	*assault*	*battery*
MAE	0.0525	0.0328	0.0255	0.0426
RMSE	0.1611	0.1054	0.0774	0.1351

(b) n=20

n=20	*theft*	*criminal damage*	*assault*	*battery*
MAE	0.0174	0.0114	0.0084	0.0145
RMSE	0.0674	0.0441	0.0332	0.0555

(c) n=50

n=50	*theft*	*criminal damage*	*assault*	*battery*
MAE	0.0060	0.0037	0.0034	0.0048
RMSE	0.0284	0.0186	0.0142	0.0232

The effect of the crime prediction model based on different grids of sparse density is shown in Table 1. Contrast table 1 (a) and (b) and (c) of the data, four cases of *MAE* and *RMSE* were showing decrease with the increase of n value trend, it is not difficult to found the fine degree of grid has a great influence on the improvement of model accuracy, 50 parts under the condition of four types of model prediction error reaches the minimum case, among them, the mean absolute error is only 0.006, root mean square error is only 0.0284.

BPNN algorithm models the data, and dynamically adjusts the weights of each connection layer under the excitation of external input samples, so that the output of the network is constantly close to the expected value. BPNN performs well in the data environment dealing with multiple input variables. Under the condition of 50 equal mesh, more than 2000 pieces of data are taken to generate the line chart of relative error between the predicted value and the actual value, as shown in Fig. 6, which shows that the predicted value is very close to the actual value. Through the above analysis, it can be seen that the BPNN constructed in this paper has good prediction accuracy and effectiveness.

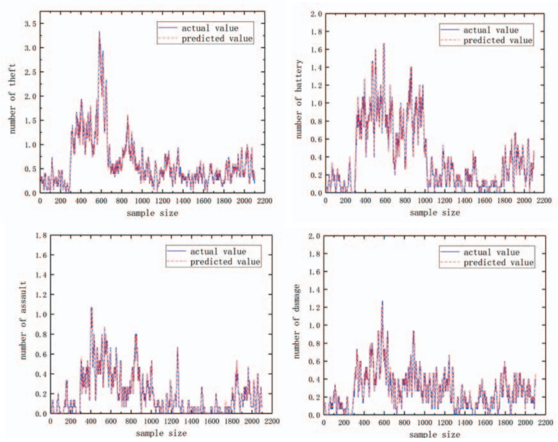


Figure 6. BPNN prediction effect diagram under 50 equal division conditions

IV. CONCLUSION

In general, it can be seen from the mesh data of three different densities that a larger grid unit will blur the accuracy of the point processing method, while a smaller grid unit will destroy the correlation between cases of the same type. Therefore, it is necessary to choose a reasonable grid density to effectively reduce the error of the prediction results and improve the prediction accuracy of the model. In this paper, a method based on grid management is proposed to optimize the performance of the crime prediction model. The model built based on BPNN algorithm can improve the precision of prediction with the help of grid element attributes. In future work, existing research results can be used to optimize the BPNN algorithm. Ouyang Bin et al. [12] proposed a global optimization of the initial connection weight and threshold of BPNN based on particle swarm optimization, which can improve the prediction accuracy and convergence speed of the model without affecting the accuracy. Sun Xia et al. [13] used genetic algorithm to optimize BPNN algorithm to improve the synchronization accuracy of time comparison, and its synchronization accuracy could reach $6.21e^{-11s}$. Zhang Julin [14] used Simulated Annealing algorithm to optimize the prediction model of BPNN. Under the circumstances of low signal-to-noise ratio and small samples, this model can obtain the optimal solution and has high convergence speed and prediction accuracy. At the same time, the Knox [15] method can be introduced to obtain the spatio-temporal critical value and conduct aggregation test on it. The n value in meshing can be selected scientifically to avoid the errors caused by empirical determination of n value, which can further improve the accuracy of the model.

REFERENCES

[1] Sherry Towers, Siqiao Chen, Abish Malik, et al. Factors influencing temporal patterns in crime in a large American city: A predictive analytics perspective. 2018, 13(10).

[2] Youngmi Lee,Sangyeol Lee. On causality test for time series of counts based on poisson ingarch models with application to crime and temperature data[J]. Communications in Statistics - Simulation and Computation, 2019, 48(6).

[3] Liao R , Wang X , Li L , et al. A novel serial crime prediction model based on Bayesian learning theory[C]// International Conference on Machine Learning & Cybernetics. IEEE, 2010.

[4] Ajayakumar J , Shook E . Leveraging parallel spatio-temporal computing for crime analysis in large datasets: analyzing trends in near-repeat phenomenon of crime in cities[J]. International Journal of Geographical Information Science, 2020: 1-25.

[5] 郭雅琦,陈鹏.基于时空临近重复效应的犯罪热点特征及成因分析——以北京市抢劫案件为例[J].地理科学进展, 2020, 39(05): 804-814.

[6] 顾海硕, 陈鹏, 李慧波. 犯罪时空预测方法研究综述与展望[J]. 地球信息科学学报, 2021, 23(01): 43-57.

[7] Rummens A , Hardyns W . Comparison of near-Repeat, Machine Learning and Risk Terrain Modeling for Making Spatiotemporal Predictions of Crime[J]. Applied Spatial Analysis and Policy, 2020, 13(1).

[8] Wang K , Cai Z , Zhu P , et al. Adopting data interpretation on mining fine-grained near-repeat patterns in crimes[J]. Journal of Forensic & Legal Medicine, 2018, 55: 76-86.

[9] Zhao X , Tang J . Modeling Temporal-Spatial Correlations for Crime Prediction[C]// ACM, 2017: 497-506.

[10] Ying-Lung Lin, Meng-Feng Yen, Liang-Chih Yu. Grid-Based Crime Prediction Using Geographical Features. 2018, 7(8).

[11] 王嵘冰,徐红艳,李波,冯勇.BP神经网络隐含层节点数确定方法研究[J].计算机技术与发展, 2018, 28(04): 31-35.

[12] 欧阳斌,陈艳红,邓传军.基于PSO优化BP神经网络的露天矿边坡位移预测模型[J].有色金属, 2020, 72(05): 37-41.

[13] 孙霞,袁媛,张然,王超,杨嘉明. 遗传算法优化BP神经网络在搬运钟时间同步中的应用[A]. 中国卫星导航系统管理办公室学术交流中心.第十一届中国卫星导航年会论文集——S06 时间基准与精密授时[C].中国卫星导航系统管理办公室学术交流中心:中科北斗汇(北京)科技有限公司, 2020: 5.

[14] 张巨林.基于模拟退火优化BP神经网络的变形监测数据预测方法[J/OL].测控技术, 2020: 1-6.

[15] 岳瀚,朱欣焰,呙维,佘冰,高超.Knox时空交互检验空间阈值确定方法[J].武汉大学学报(信息科学版), 2018, 43(11): 1719-1724.