Applied Mathematics and Nonlinear Sciences

<https://www.sciendo.com>

Comparison of compression estimations under the penalty functions of different violent crimes on campus through deep learning and linear spatial autoregressive models

Huiping Hu^{1†}, Xinqun Huang¹, Majed Ahmad Suhaim², Hui Zhang¹

¹ Department of English Education, Shangrao Preschool Education College, Shangrao 334000, China

² Department of Information Technology, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Submission Info

Communicated by Juan Luis García Guirao

Received June 17th 2021

Accepted September 24th 2021

Available online December 13th 2021

Abstract

To reduce the probability of violent crimes, the deep learning (DL) technology and linear spatial autoregressive models (ARMs) are utilised to estimate the model parameters through different penalty functions. In addition, under a determinate space, the influences of environmental factors on violent crimes are discussed. By taking campus violence cases as examples, the major influencing factors of violent crimes are found through data analysis. The results show that campus violence cases are usually caused by the complex surrounding environments and persons. Also, campus security measures only cover a small range, and the security management is difficult. In the meantime, due to the younger ages and lack of self-protection awareness, students may easily become the targets of criminals. Therefore, the results have a positive significance for authorities to analyse the crime rates in a determinate area and take preventive measures against violent crimes.

Keywords: deep learning, campus violence, autoregression, penalty function, parameter estimation

AMS 2010 codes: 62J05

1 Introduction

In recent years, violent crimes on campus are not uncommon. These crime cases not only have a bad impact on the majority of teachers and students but also cause a negative guidance to society, which has received widespread concern [1, 2]. According to previous surveys and data, in China, the number of people who have suffered campus violence has exceeded 57%, and the resulting number of people who have violated the laws and regulations has reached 85%. Therefore, campus violence has become a common issue, which has brought negative impacts to the campus; campus violence not only disrupts the normal teaching order but also has a side

[†]Corresponding author.

Email address: srszhk@sina.com

impact on the mental health of students. Hence, it is necessary to take effective preventive measures against campus violent crimes, and provide schools with a stable and harmonious campus environment [3,4].

As a well-developed and widely researched technology in recent years, deep learning (DL) has solved many problems in many fields and has also provided more possibilities for the development of these fields. Especially, in image recognition, feature values can be extracted continuously through convolutional neural networks (CNNs), thereby recognising and classifying images through continuous learning [5,6]. In addition to images, DL can perform sound detection and diagnosis of sound. Even if it is affected by environmental noise during the sound collection process, sound extraction can also be achieved through feature extraction. Once the spatial factor is added, the difficulty of signal extraction will increase. At this moment, machine learning (ML) can improve the classification accuracy to some extent through abstract learning, thereby discovering the signal defects [7–9].

Therefore, considering the huge negative impacts of campus violence on society, as well as the recognition advantages of ML, DL technology and linear spatial autoregressive models (ARMs) are combined to compare the compression estimations under the penalty functions of different violent crimes. In addition, the campus violence cases are taken as examples for in-depth analysis, with the hope of discovering the major influencing factors of campus violence, and thereby fundamentally reducing the crime rate of campus violence and providing a safer environment for the campus.

2 Literature review

The continuous development of computer technology has promoted the development of DL technology, making DL well-applied in more fields; in addition to image recognition, it also achieved good results in speech classification and recognition [10]. Fawaz et al. (2019) pointed out that deep neural networks (DNNs) have brought revolutionary changes to the field of computer vision, especially the emergence of deeper structures. In addition to images, when processing sequence data such as text and audio, DNNs can also achieve the optimal classification and recognition performance. When using DNNs to process time-series classification, it is found that DNNs show good performance and are evaluated on a univariate time-series benchmark and 12 multivariate time-series datasets. The results show that the training of 8,730 DL models in 97 time-series datasets provides more possibilities for the application of DNNs in event sequence classification [11]. Chen et al. (2018) proposed a long-term memory (LTM) neural network, support vector regression machine and extreme value optimisation algorithm for DL time-series prediction nonlinear learning integration method, which is applied to the information mining and development of implicit wind speed time series; on this basis, the predictions are assembled into a nonlinear learning regression top layer composed of SVRM, and an extreme value optimisation algorithm is introduced to optimise the top layer parameters. The results show that the proposed algorithm model has better prediction performance compared with other models [12]. Karim et al. (2017) proposed a time-series classification method for fully convolutional networks (FNNs) with LTM recursive neural network sub-modules. The results show that the proposed model can significantly improve the performance of FNN, the increase in the model size and the minimal preprocessing of the dataset [13].

Benton (2017) pointed out that because violence and organised crimes will destroy the development of a country, foreign direct investment is reduced, armed conflicts also lower the sovereign credit rating, and the loan business is also affected to some extent. Therefore, analysis of the cross-sectional time series of the crime rate and municipal debt in Mexico has shown that the governments most in need of cost-effective financing are most likely to be charged higher prices due to violent crimes [14]. Eren and Mocan (2017) used administrative data from a state in the southern United States to discuss the impact of juvenile crime punishment in later life. The results show that juvenile criminal punishment has a greater impact on the adulthood of these teenagers and may make them commit more serious crimes; however, the impact on people around them is almost zero [15]. Gerlinger and Hipp (2020) used the negative binomial and logistic regression models to estimate the impact of high-school dropouts and high-performing schools on violent crimes in a county in California, in the United

States. The results show that school dropout is associated with an increase in serious violent robberies, but schools with good grades will not have a significant impact on these types of crimes. Finally, by comparing the juvenile crime model with the juvenile and adult crime model, the viewpoint that reasonable measures should be performed on juvenile delinquency is supported [16].

In sum, in previous explorations, DL technology is mostly applied to the prediction of images, speech and time. At present, campus violence is frequent, and the degree of punishment will also affect the future lives of teenagers who have committed violent crimes to some extent. Therefore, the DL technology and linear spatial ARMs are utilised to estimate the penalty functions of different violent crimes. By taking the campus violence cases as examples, the major influencing factors of crimes are discovered.

3 Method

3.1 Backpropagation neural network

Backpropagation neural network (BPNN) is a feedforward neural network trained according to the error backpropagation algorithm, which has good applications in many fields. BPNN and feedforward neural network are similar in structure; they consist of the input layer, the hidden layer and the output layer. The hidden layer can be a single layer or multiple layers. The difference is that BPNN will also utilise the backpropagation algorithm after the forward propagation ends and obtain the loss function of the difference between the expected output and the actual output of the model. It uses the gradient descent method to adjust the node weights layer-by-layer to improve the fitting degree of the model. This algorithm has an excellent classification effect for those input objects with obvious characteristics. However, when the dimension of the input value is too large or complicated, the convergence speed will also be slowed down, and it is easy for the model to fall into the local minimum. Thus, considering this defect, the hidden layer of BPNN only retains one layer [17–19].

In addition to feature extraction and classification of images, the signal frequency domain features also show good classification performance. In terms of extracting frequency domain features, this algorithm can also exert its advantages as much as possible to reduce the impact of disadvantages; therefore, BPNN is an efficient classifier. The model structure of BPNN is shown in Figure 1:

Fig. 1 Model structure of BPNN. BPNN, backpropagation neural network.

In the input layer of BPNN, each node corresponds to a feature. When outputting through the hidden layer node, the hidden layer node needs to be determined, as shown in Eq. (1):

$$e = \sqrt{m+n} + \alpha, \quad 1 \leq \alpha \leq 10 \quad (1)$$

where: e represents the number of nodes in the hidden layer, n represents the number of nodes in the input layer, m represents the number of nodes in the output layer and α represents a random number. During training, the activation function is also very important. The sigmoid function is expressed as follows [20]:

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

3.2 Deep belief network

Deep belief network (DBN) is a widely used algorithm in DL. It is a probabilistic generative model that establishes the joint distribution between observation data and labels; the discriminant model is only to predict the labels. This kind of network can be used for both supervised and unsupervised learning as a classifier. Among them, the restricted Boltzmann machine (RBM) is a component unit in DBN; the entire DBN is composed of multiple RBM stacks; the structure of DBN is shown in Figure 2 [21, 22]:

Fig. 2 Structure of DBM. RBM, restricted Boltzmann machine.

As shown in Figure 2, in the DBN, there are n RBMs. When the input of the first RBM is the visible layer, the output is the hidden layer 1. The hidden layer 1 is used as the input layer of the next RBM, and the rest may be deduced by analogy. Then, a complete DBN is formed, in which the last RBM output is used as the DBN output. Among them, there is a connection between the visible layer and the hidden layer, but there is no connection between the units in the layer.

3.3 ARM

ARM is a stationary time-series model with a good effect on the prediction of economy, information and natural phenomena. This method does not require much data when predicting, and can complete the prediction task according to its variable sequence. However, ARM will also be limited during applications. For example, it must have autocorrelation, because autocorrelation coefficient is the key. Usually, autocorrelation is used to predict the economic phenomena associated with the previous period. For economic phenomena that are greatly affected by social factors, other variables need to be added, or other methods need to be combined to complete the prediction [23, 24].

Due to the lag in space, the parameter estimation results obtained by the least squares method (LSM) are inconsistent when estimating the parameters of the spatial ARM. Therefore, many scholars have utilised the maximum likelihood estimation method and the generalised matrix estimation method to estimate the parameters. These methods have good verification results in their respective fields. The linear spatial ARM is proposed based on the spatial ARM. Variable selection is an important component of the model and an important link in the analysis of specific data. Generally, the autoregressive variable selection problems are explored specifically for a determinate penalty function. However, due to the large number of penalty functions, different penalty functions have different penalty effects [25, 26].

Assuming that $X^T = (X_1, X_2, \dots, X_n)$ is the regression matrix of $n \times p$, a general spatial ARM can be established, as shown in Eq. (3):

$$Y_n = \rho W_n Y_n + X_n \beta + \varepsilon_n \quad (3)$$

where ρ and β represent fixed parameters, W_n represents the spatial weight matrix of order n , ε_n represents the random error vector of $n \times 1$, the mean is 0, $\sigma^2 I_n$ represents the variance and I_n represents the unit matrix of n orders [27]. When estimating the parameters, the log-likelihood function can be analysed as in Eq. (4):

$$Ln(\theta) = -\frac{n}{2} \log 2\pi - \frac{n}{2} \log \sigma^2 + \log |I_n - \rho W| - \frac{1}{2\sigma^2} \varepsilon_n^T(\theta) \varepsilon_n(\theta) \quad (4)$$

On this basis, the interface likelihood method is used to find the estimates $\hat{\beta}(\rho)$ and $\hat{\sigma}^2(\rho)$ of β and σ^2 , as shown in Eqs (5) and (6):

$$\hat{\beta}(\rho) = (X^T X)^{-1} X^T A(\rho) Y \quad (5)$$

$$\hat{\sigma}^2(\rho) = \frac{1}{n} (A(\rho) Y - X \hat{\beta})^T (A(\rho) Y - X \hat{\beta}) \quad (6)$$

If Eqs (5) and (6) are brought into Eq. (4), Eq. (7) will be obtained:

$$Ln(\rho) = -\frac{n}{2} (1 + \log(2\pi)) - \frac{n}{2} \log \left(\frac{1}{n} (A(\rho) Y)^T M(A(\rho) Y) \right) + \log |A(\rho)| \quad (7)$$

While solving Eq. (7) and obtaining the maximum value, the estimated value $\hat{\rho}$ of the parameter ρ is obtained, which is then brought into Eqs (5) and (6), whose estimates $\hat{\beta}(\hat{\rho})$ and $\hat{\sigma}^2(\hat{\rho})$ are obtained. To further obtain the sparse estimation of the parameters, the penalty function is set as the objective function, as shown in Eq. (8):

$$Q(\beta) = \|A(\hat{\rho}) y - X \beta\|^2 + \sum_{i=1}^p p \lambda_i (|\beta_i|) \quad (8)$$

The basic principle of the penalty function is to compress the smaller parameters to 0 as much as possible, thereby eliminating them finally. Therefore, not only can the parameter be estimated but also the selection of variables can be realised. Therefore, the following three penalty functions are introduced, that is, least absolute shrinkage and selection operator (LASSO), adaptive least absolute shrinkage and selection operator (ALASSO), and smoothly clipped absolute deviation (SCAD). The estimates of these functions are shown in Eqs (9–11):

$$\hat{\beta}_{lasso} = \arg \min \left\| (A(\hat{\rho}) y - X \beta)^T (A(\hat{\rho}) Y - X \beta) \right\|^2 + \lambda \sum_{j=1}^p |\beta_j| \quad (9)$$

$$\hat{\beta}_{alasso} = \arg \min \left\| (A(\hat{\rho}) y - X \beta)^T (A(\hat{\rho}) Y - X \beta) \right\|^2 + \sum_{j=1}^p p \lambda w_j |\beta_j| \quad (10)$$

$$\hat{\beta}_{scad} = \arg \min \left\| (A(\hat{\rho}) y - X \beta)^T (A(\hat{\rho}) Y - X \beta) \right\|^2 + \sum_{j=1}^p p \lambda |\beta_j| \quad (11)$$

Equation (9) is the estimator of LASSO penalty. This method has small variance and fast calculation speed, but it is biased for the parameter estimation. Equation (10) is the estimator of ALASSO penalty, where w represents the weight vector, which can make choices according to different data [28]. Therefore, the obtained estimator has asymptotic normality. Besides, some scholars also found in analysis that because the method is a convex function, it can obtain the global optimal solution. Equation (11) is an estimate of the SCAD penalty, where p_λ can be expressed as Eq. (12):

$$p_\lambda |\theta| = \begin{cases} \lambda |\theta|, & |\lambda| \prec \theta \\ \frac{(a^2-1)\lambda^2 - (|\theta|-a\lambda)^2}{2(a-1)}, & \lambda \prec |\theta| \leq a\lambda \\ \frac{1}{2}(a+1)\lambda^2, & |\theta| \succ a\lambda \end{cases} \quad (12)$$

$$p'_\lambda(\theta) = \lambda (I(\theta \prec \lambda)) + \frac{(a\lambda - \theta) + I(\theta \succ \lambda)}{(a-1)\lambda} \quad (13)$$

4 Experiment

4.1 Model construction and feature extraction

The subsequent orders of the ARMs established by different types of signals are different. At this moment, to ensure that the dimensions of the regression parameter vectors of all signals are the same, the smallest model lag order in a set of ARMs established by all signals in the dataset is used as a criterion for determining the parameter vector dimension.

After obtaining all the regression parameters of the ARM, these models are combined into parameter vectors; then the DBN model is utilised to learn and extract features. Since RBM nodes are binary, it can be seen that each node has only two states: 0 and 1. Besides, the nodes of each layer are not directly connected; therefore, the state probability of each node can be expressed by Eq. (14) to Eq. (17):

$$p(h_j = 1 | v) = \text{sig} \left(c_j + \sum_i v_i w_{ij} \right) \quad (14)$$

$$p(v_i = 1 | h) = \text{sig} \left(b_i + \sum_j h_j w_{ij} \right) \quad (15)$$

$$p(h_j = 0 | v) = 1 - p(h_j = 1 | v) \quad (16)$$

$$p(v_i = 0 | h) = 1 - p(v_i = 1 | h) \quad (17)$$

where v_i and h_j , respectively, represent the state of the i -th node in the visible layer and the j -th node in the hidden layer, and sig represents the activation function. Therefore, the energy function between the visible layer and the hidden layer can be expressed by Eq. (18):

$$E(v, h) = - \sum_i b_i v_i - \sum_j c_j h_j - \sum_{i,j} v_i h_j w_{ij} \quad (18)$$

Therefore, on this basis, the possible probability of the combined allocation of the visible layer and the hidden layer is shown in Eq. (19):

$$p(v, h) = \frac{1}{Z} \sum_h e^{-E(v, h)} = \frac{e^{-E(v, h)}}{\sum_{v, h} e^{-E(v, h)}} \quad (19)$$

After summing all the possible hidden layer node states, the probability distribution of the visible layer unit state vector is obtained, as shown in Eq. (20):

$$p(v) = \frac{1}{z} \sum_h e^{-E(v,h)} = \frac{\sum_h e^{-E(v,h)}}{\sum_{v,h} e^{-E(v,h)}} \quad (20)$$

When training RBM, it is necessary to adjust the weights and offsets, thereby reducing the structural energy and then reaching a stable state. The parameters are adjusted, as shown in Eq. (21):

$$\begin{aligned} \Delta w_{ij} &= \gamma (\langle v_i h_j \rangle_{data} - \langle v_i h_j \rangle_{mod el}) \\ \Delta b_i &= \gamma (\langle v_i \rangle_{data} - \langle v_i \rangle_{mod el}) \\ \Delta c_j &= \gamma (\langle h_j \rangle_{data} - \langle h_j \rangle_{mod el}) \end{aligned} \quad (21)$$

where γ represents the learning rate, $\langle \rangle_{data}$ represents the expected $p(h_j | v)$ of the data distribution and $\langle \rangle_{mod el}$ represents the expected $\sum_v p(v, h_j)$ of the model distribution.

On this basis, an output layer is added to the DBN structure as a classifier. At this moment, the weight and offset of the DBN network need to be used as the feature values of the structure, and the network state is continuously optimised through training [29]. After the DBN is trained, the number of nodes in the output layer is consistent with the number of signal types that need to be identified. This requires the utilisation of supervised learning of the BPNN algorithm to adjust the optimal parameters of each RBM, so as to optimise the entire network. BPNN can adjust the feedback layer-by-layer during the adjustment process; then, it connects the RBM to obtain the optimal parameters.

4.2 Campus violence cases

Here, the campus violence cases are taken as examples. The violent crime cases from 2006 to 2010 in a determinate area are selected, and the simulation analysis is conducted through different punishment functions. Finally, the parameter estimations are compared.

According to statistics, from 2015 to 2019 in this area, the local procuratorate has prosecuted 238 violent crimes in and around the campus, involving 340 people and covering nearly 100 schools, including universities, primary schools and high schools. The data statistics are shown in Figure 3:

Fig. 3 Statistics of campus crime in a determinate area.

As shown in Figure 3, violent incidents occur on and off campus every year. According to the number of incidents and schools, the incidents involve various schools, and the frequency of occurrence is high, which threatens the safety of teachers and students. Since 2016, the number of cases and the number of crimes has declined year by year, which is also a happy fact.

4.3 Simulation parameters setting

In the analysis of violence cases, the model parameters are set. It is assumed that the total number of campus areas studied is R , and they are independent of each other. There are m individuals in each area, and they interact with each other. During the simulation, when different sample sizes are given, the parameters after the space will correspond to different simulation results, where C represents the average number of '0' coefficients which are correctly estimated, I represents the four non-zero coefficient that are erroneously estimated as the average number of '0', and MSE represents the mean square error of the estimator, which is used to measure how close the estimated value is to the true value. In the results given by the simulation, the MSE value is selected to be expanded 100 times for easier comparison.

5 Results and discussion

In the study of campus violence, when different parameter values are applied, the corresponding results are also different. Therefore, when using LASSO, ALASSO and SCAD penalty functions for analysis, different results are obtained. When N is 100, m is 10, R is 10 and ρ is 0.1; the results of the three penalty functions are shown in Figure 4:

Fig. 4 Simulation results when $N = 200$, $R = 10$ and $\rho = 0.1$. (a) The result of ALASSO penalty function; (b) the result of LASSO penalty function; and (c) the result of SCAD penalty function. ALASSO, adaptive least absolute shrinkage and selection operator; LASSO, least absolute shrinkage and selection operator; SCAD, smoothly clipped absolute deviation.

According to the above results, the difference in σ makes the results of the three indicators different. The changes in C and I values are small, and the MSE value magnified by 100 times shows a significant difference. Then, when m is unchanged and different N , R and B values are taken; the results of the three penalty functions are shown in Figures 5–7.

According to the results shown in Figures 4–7, as the sample size and the number of simulations increase, the C value gradually becomes larger, the I value is always 0 and the MSE value is gradually decreasing, which also shows that the accuracy of the simulation results is getting higher. Besides, as the value of σ increases, the value of C becomes smaller, and the value of MSE becomes larger, which means that the larger the variance, the worse the simulation effect will be. Therefore, the variance needs to be controlled. When comparing the three

Fig. 5 Simulation results when $N = 200$, $R = 10$ and $\rho = 0.5$. (a) The result of ALASSO penalty function; (b) the result of LASSO penalty function; and (c) the result of SCAD penalty function. ALASSO, adaptive least absolute shrinkage and selection operator; LASSO, least absolute shrinkage and selection operator; SCAD, smoothly clipped absolute deviation.

Fig. 6 Simulation results when $N = 400$, $R = 30$ and $\rho = 0.1$. (a) The result of ALASSO penalty function; (b) the result of LASSO penalty function; and (c) the result of SCAD penalty function. ALASSO, adaptive least absolute shrinkage and selection operator; LASSO, least absolute shrinkage and selection operator; SCAD, smoothly clipped absolute deviation.

Fig. 7 Simulation results when $N = 400$, $R = 30$ and $\rho = 0.5$. (a) The result of ALASSO penalty function; (b) the result of LASSO penalty function; and (c) the result of SCAD penalty function. ALASSO, adaptive least absolute shrinkage and selection operator; LASSO, least absolute shrinkage and selection operator; SCAD, smoothly clipped absolute deviation.

penalty functions ALASSO, LASSO and SCAD, it is found that the simulation results of ALASSO and SCAD are better than those of LASSO.

When analysing violence on campus, the influencing factors include the identity, education, age, crime area, crime incident and the modus operandi (MO) of the criminals, as well as the age, identity and outcomes of the victims. Therefore, in this determinate area, the population (X1), the proportion of non-school groups (X2), the median income (X3), the percentage of people with a bachelor's degree or above (X4), the proportion of high-school graduates aged 25 years and above (X5), the percentage of children born in unmarried families among all children (X6), the percentage of children living with parents among all children (X7), the percentage of divorced couples (X8) and the percentage of people without full-time jobs (X9) are selected as research targets. The statistical results of these parameters are shown in Table 1:

Table 1 Data parameter estimation of crime influencing factors

Influencing factors	@ @QMLE	LASSO	ALASSO	SCAD
X ₁	−0.2174235	−0.1453376	−0.0861383	−0.2133456
X ₂	0.2461965	0.2118965	0.2431985	0.2265178
X ₃	0.0051896	0	0	0
X ₄	−0.1041869	0	0	0
X ₅	0.1472457	0.00112187	0	0
X ₆	0.14773986	0.16729365	0.1051986	0.18729353
X ₇	−0.2786496	−0.31298654	0	−0.2825
X ₈	0.14085674	0.05428654	0	0
X ₉	0.1751896	0.18532187	0.1507659	0

Based on the above results, the lag coefficients of LASSO, ALASSO and SCAD are estimated to be 0.5801, 0.5811 and 0.5801, respectively. According to these results, among all the factors analysed, the three penalty estimates have some parameters compressed to 0 relative to QMLE, and the signs of non-zero parameters are consistent with the signs of QMLE estimates. Among the above variables, the number of variables that LASSO can compress to 0 is 2, and there are 5 variables that ALASSO and SCAD can compress to 0. Therefore, among the three penalty functions of LASSO, ALASSO, and SCAD, ALASSO and SCAD can compress the variables more efficiently.

For those variables that cannot be compressed regardless of which penalty function is used, attention needs to be paid, indicating that these influencing factors are vital to campus violence. For example, the number of non-school groups around the campus, the proportion of children born in unmarried families, and the proportion of divorced couples are positively correlated with the crime rate, which means that the more non-school groups around the campus, the greater the possibility of violent crime will be; the higher the proportion of children born in unmarried families, the higher the crime rate will be. This is because that when children are born in an unmarried family, they may lack family education and are prone to psychological and personality defects. The higher the proportion of divorced couples, the higher the crime rate will be; the reason is that men are more prone to extreme behaviours than women.

ALASSO, adaptive least absolute shrinkage and selection operator; LASSO, least absolute shrinkage and selection operator; SCAD, smoothly clipped absolute deviation.

In sum, when analysing the influencing factors of violent crimes, it is found that family has a greater impact; in addition, family structure and family relationships will affect the development of children to some extent. Whether from the psychological or personality perspective, this will affect the ideological awareness of children, making them suffer from the negative impacts of the family and society, which, in turn, leads to criminal thoughts. Therefore, family education for children is more important among many educations; any change in the family may become a cause of juvenile delinquency, which will also have a greater impact after the teenagers enter society.

6 Conclusion

Based on the spatial ARMs, BPNN and DBN are utilised to obtain features. By setting different areas, numbers and variance values, the parameters are compressed and estimated. Then, the two penalty functions ALASSO and SCAD can be obtained, which can perform the compression estimation more effectively. Also, it is found that some family associated factors have a greater impact on teenagers and younger adults, which should be valued while considering the crime rate. However, the influencing factors selected in the research process are less. In the future, the influence of social factors will be analysed to further verify the importance of family relations, which will play an important role in preventing juvenile delinquency. The utilisation of DL technology and linear spatial ARMs to explore violent crimes has a positive significance for investigating the crime rate in a determinate area and the surrounding crime situation. To some extent, it can fundamentally prevent the occurrence of violent crimes.

References

- [1] Won S D, Chang E J. The relationship between school violence-related stress and quality of life in school teachers through coping self-efficacy and job satisfaction. *School Mental Health*, 2020, 12(1):pp. 136-144.
- [2] Estévez E, Jiménez T, Segura L. Emotional intelligence and empathy in aggressors and victims of school violence. *Journal of educational psychology*, 2019, 111(3): pp. 488.
- [3] Moore H, Astor R A, Benbenishty R. Role of school-climate in school-based violence among homeless and nonhomeless students: Individual-and school-level analysis. *Child Abuse & Neglect*, 2020, 102: pp. 104378.
- [4] Shaofei Wu. Research on the application of spatial partial differential equation in user oriented information mining, *Alexandria Engineering Journal*, 2020, 59(4): 2193–2199.
- [5] Hemmati F, Alqaradawi M, Gadala M S. Rolling element bearing fault diagnostics using acoustic emission technique and advanced signal processing. *Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology*, 2016, 230(1):pp. 64-77.
- [6] Zhang Y, Liu B, Ji X, et al. Classification of EEG signals based on autoregressive model and wavelet packet decomposition. *Neural Processing Letters*, 2017, 45(2): pp. 365-378.
- [7] Bringmann L F, Ferrer E, Hamaker E L, et al. Modeling nonstationary emotion dynamics in dyads using a time-varying vector-autoregressive model. *Multivariate behavioral research*, 2018, 53(3): pp. 293-314.
- [8] Han L, Zhou M, Jia W, et al. Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model. *Information Sciences* 2019, 476: pp. 491-504.
- [9] Bourque A E, Carrier J F, Filion É, et al. A particle filter motion prediction algorithm based on an autoregressive model for real-time MRI-guided radiotherapy of lung cancer. *Biomedical Physics & Engineering Express*, 2017, 3(3): pp. 035001.
- [10] Zhang S, Yao L, Sun A, et al. Deep learning based recommender system: A survey and new perspectives[J]. *ACM Computing Surveys (CSUR)*, 2019, 52(1): pp. 1-38.
- [11] Fawaz H I, Forestier G, Weber J, et al. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 2019, 33(4): pp. 917-963.
- [12] Chen J, Zeng G Q, Zhou W, et al. Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy conversion and management*, 2018, 165: pp. 681-695.
- [13] Karim F, Majumdar S, Darabi H, et al. LSTM fully convolutional networks for time series classification. *IEEE access*, 2017, 6: pp. 1662-1669.
- [14] Benton A L. Violent crime and capital market punishment: How violent crime affects the supply of debt to municipal Mexico. *Studies in Comparative International Development*, 2017, 52(4): pp. 483-509.
- [15] Eren O, Mocan N. Juvenile punishment, high school graduation and adult Crime: Evidence from idiosyncratic judge harshness. *Review of Economics and Statistics*, 2017: pp. 1-46.
- [16] Gerlinger J, Hipp J R. Schools and neighborhood crime: The effects of dropouts and high-performing schools on juvenile crime. *The Social Science Journal*, 2020: pp. 1-17.
- [17] Shaofei Wu. AN USER INTENTION MINING MODEL BASED ON FRACTAL TIME SERIES PATTERN, FRACTALS-COMPLEX GEOMETRY PATTERNS AND SCALING IN NATURE AND SOCIETY, 2020, 28(8), Article Number 2040017.
- [18] Hu Y, Li J, Hong M, et al. Short term electric load forecasting model and its verification for process industrial enterprises based on hybrid GA-PSO-BPNN algorithm—A case study of papermaking process. *Energy*, 2019, 170: pp. 1215-1227.

- [19] Sun S, Li G, Chen H, et al. A hybrid ICA-BPNN-based FDD strategy for refrigerant charge faults in variable refrigerant flow system. *Applied Thermal Engineering*, 2017, 127: pp. 718-728.
- [20] Kwon H B, Marvel J H, Roh J J. Three-stage performance modeling using DEA–BPNN for better practice benchmarking. *Expert Systems with Applications*, 2017, 71: pp. 429-441.
- [21] Song H, Wen H, Hu L, et al. Secure cooperative transmission with imperfect channel state information based on BPNN. *IEEE Transactions on Vehicular Technology*, 2018, 67(11): pp. 10482-10491.
- [22] Yu L, Zhou R, Tang L, et al. A DBN-based resampling SVM ensemble learning paradigm for credit classification with imbalanced data. *Applied Soft Computing*, 2018, 69: pp. 192-202.
- [23] Taal A, Itard L. Fault Detection and Diagnosis for Indoor Air Quality in DCV systems: Application of 4S3F method and effects of DBN probabilities. *Building and Environment*, 2020: pp. 106-632.
- [24] Candan Ç. Making linear prediction perform like maximum likelihood in Gaussian autoregressive model parameter estimation. *Signal Processing*, 2020, 166: pp. 107-256.
- [25] Shaofei Wu. Design of interactive digital media course teaching information query system, *Information Systems and e-Business Management*, 2020, 18(4): pp. 793–807.
- [26] Han L, Zhou M, Jia W, et al. Intrusion detection model of wireless sensor networks based on game theory and an autoregressive model. *Information Sciences*, 2019, 476: pp. 491-504.
- [27] Jung R C, Tremayne A R. Maximum-Likelihood Estimation in a Special Integer Autoregressive Model. *Econometrics*, 2020, 8(2): pp. 24.
- [28] Andrey Kochergin. Besicovitch cascades and bounded partial quotients. *Applied Mathematics and Nonlinear Sciences*, 2020, 5(2):pp. 267-278.
- [29] Yanan Zhao. Analysis of Trade Effect in Post-Tpp Era: Based on Gravity Model and Gtap Model. *Applied Mathematics and Nonlinear Sciences*, 2020, 5(1):pp. 61-70.