A Hybrid Model of Crime Prediction

Meilin Liu and Tianliang Lu
College of Information Technology and Cyber Security, People’s Public Security University of China, China

Abstract. According to statistics, crimes are not random in spatial and temporal distribution. The key to predicting policing is predicting in advance when and where a crime may occur, and providing a reference for preventive measures of police officers. In this paper, a hybrid model of LSTM and STARMA is established. Crime data is complicated. It can be decomposed into trend components, seasonal components, and random components. The LSTM model is established for the trend components and the seasonal components. The STARMA model is established about random components. It solves the problem that the STARMA cannot be modeled on nonlinear or non-stationary data. Verification results show that the model is effective in crime prediction.

1. Introduction

Based on the analysis of relevant research results, it is found that using time and space data to predict crimes has made a certain development, including space-time anomaly analysis, crime correlation analysis, crime hot spot analysis, crime trend prediction and so on. Since 1880s, London police had tried to use spatial analysis in analysis of criminal cases. By the end of the 20 century, police officer Ross Simon from Simon Frasso university derived a formula about the relationship between crimes and locations, according to the historical crime data. In 2010, Nakaya discussed the space-time cube based on nuclear density estimation and scanning statistics. It made the crime prediction developed from a space or a time perspective to space-time, but it is not perfect in quantitative analysis[1]. In 2014, MS Gerber found that the addition of Twitter data improved crime prediction performance versus a standard approach based on kernel density estimation.[2]. In 2017, BJ Jefferson investigated the implications that predictive crime mapping has for racialized modes of urban policing[3]. G Rosser’ study suggests network-based methods of crime forecasting are likely to outperform grid-based methods[4].

In general, the crime prediction models can be divided into four categories:a. analysis from a single time or space perspective, b. combination of time and space analysis results; c. taking time and space as independent variables to establish traditional modeling; d. space-time series model. The first three types of models can not take space-time autocorrelation and heterogeneity into account when dealing with crime data. Based on this, a hybrid model of STARMA and LSTM is used to predict crimes. Therefore the time and space characteristics can be taken into account when dealing with crime data. Because the accuracy of STARMA model in dealing with non-stationary space-time data is not high, LSTM was used to deal with the trend and seasonal components, and then STARMA model was established for random components. Finally, hybrid space-time crime prediction model based on LSTM and STARMA is established. The public crime data of San Francisco in the United
States was used to test the model and verify the feasibility and accuracy of the model in crime prediction.

2. Space-time model
If sets of multiple time series have an interrelationship in space, they can be cold space-time series. Space-time series is the expansion of time series in space. These time series are space dependent. Because the crime data recorded by the public security stations usually contain the dimensions of time, place, etc., the crime data are usually space-time series.

2.1 Space-time autoregressive moving average (STARMA)
STARMA is proposed by Pfeifer and Deutsch (1980). STARMA model is a space–time model of time series and spatial weight matrix based on the weight of each point in the space. It is essentially a linear structure.

The process of STARMA model establishment includes: space–time data preparation, exploratory space–time analysis, model training, and model validation[5]. Space–time data preparation refers to cleaning and statistics of the data to meet the requirements of modeling. Exploratory space–time analysis refers to calculating ACF to verify the stationary property of time series. If the data is stable, calculating the STACF and STPACF for space–time series to determine the spatial weight matrix and the expression form of STARMA model. The common methods of parameter estimation is least square method. After that, correctness of the model would be checked using the estimated parameters. If the model is found to be wrong, model identification steps should be repeated[6]. After the parameters of the model are determined, regression analysis would be done by using space-time series, spatial weight matrix and random error. Finally the STARMA model is established.

The definition here used for STARMA models is the following[5]:

\[ z_i(t) = \sum_{k=1}^{n} \sum_{i=1}^{T} \phi_{ik} w_i^h z_i(t-k) + \sum_{k=1}^{n} \sum_{i=1}^{T} \theta_{ik} w_i^h \epsilon_i(t-k) + \epsilon_i(t) \]  

Where \( z_i(t) | t = 1, 2, \ldots, T; \ i = 1, 2, \ldots, N \) is the time series, \( T \) is the length of time series, \( N \) is the number of spatial units, \( z_i(i = 1, 2, \ldots, N) \) is the spatial variable, \( k \) is the time delay, \( h \) is the space delay, \( p \) is the autocorrelation order, \( q \) is the moving average order, \( w_i^h \) is the \( h \) order spatial weight matrix of the first order, and \( \epsilon_i \) is the random error.

2.2 Long Short-Term Memory (LSTM)
LSTM is an improved recurrent neural network (RNN) proposed by Hochreiter and Schmidhuber (1997). Compared with the traditional RNN, it has stronger learning ability[7]. A complete LSTM includes an input gate, output gate and forget gate. The information in LSTM network can be determined whether it is useful based on the rules. Only information that conforms to algorithm authentication will be left, and inconsistent information will be forgotten through the forget gate[8].

LSTM includes the new input \( x_i \), output \( h_{i-1} \), the input gate \( i \), forget gate \( f \), output gate \( o \). When \( x_i \) is input at \( t \) and \( h_{i-1} \) is the output at \( t - 1 \)[9]:

\[ f_t = \sigma(W_{xf} \cdot x_t + W_{hf} \cdot h_{i-1} + b_f) \]  

\[ i_t = \sigma(W_{xi} \cdot x_t + W_{hi} \cdot h_{i-1} + b_i) \]  

\[ \tilde{C}_t = \tanh(W_{xc} \cdot x_t + W_{hc} \cdot h_{i-1} + b_c) \]  

\[ C_t = f_t \odot C_{i-1} + i \odot \tilde{C}_t \]  

\[ o_t = \sigma(W_{xo} \cdot x_t + W_{ho} \cdot h_{i-1} + b_o) \]
\[ h_t = o_t \odot \tanh(C_t) \quad (7) \]

Where \( \sigma \) represents the activation function of logistic sigmoid. \( W \) and \( b \) represents the weight matrix and the bias. \( C_t \) represents the state of memory unit at time \( t \).

3. A hybrid model of crime prediction

The STARMA model assumes that the space-time series is stationary. Stationary space-time series require data to be stationary in both time and space. In order to ensure the applicability of crime data to STARMA, before constructing STARMA model, it should be tested for stabilization. If the crime data is not a stationary space-time series, data stabilization is in need. The general method of data stabilization is the difference method, but the difference is easy to lose some important information hidden in the data.

The crime prediction model is shown in figure 1. First of all, it is necessary to judge whether the data is stable in both time and space. If the data is not stable, decompose the time series. Time series can be decomposed into trend components, seasonal components, and random components. Random components is generally stationary data. LSTM can fit the trend components and seasonal components, which enhance nonlinear fitting ability for the whole model. If the data is stable, the STARMA model can established directly. Integrating the predicted results with the trend components extracted by the LSTM, the final prediction results. Finally, the model will be tested and evaluated by auto-correlation. If the residual is white noise series, the model fitting is effective.
4. Mathematics experimental verification

The data selected in this section is the official public crime data of San Francisco. It contains all crime data of the region from 2003 to 2017. The attributes of the data set are as follows: IncidntNum, Category, description, Day Of Week, Date, Time, Resolution, PdDistrict, Address, X, Y, Location (Y, X). For the convenience of the study, this study takes assault as an example. San Francisco region is divided into ten PdDistricts. The monthly statistics of the number of assault in each PdDistrict from 2003 to 2017 are calculated to build the crime time series of each police district.

4.1 Decomposition of time series

First, the stationarity of space-time series in time is tested. Take the PdDistricts of Park as an example, the number of crimes have obvious decline between 2006 and 2008. The mean and variance have larger fluctuations over time. So the crime data has no stability in time. Therefore, the data need to be decomposed.

![Figure 2. Decomposition of Park crime data](image)

A LSTM was established about trend and seasonal components of crime data. In this model, hidden size was 40. Number of layers was 1. Time steps was 2. Training steps was 10000. Batch size was 20, and the learning rate was 0.6. Loss for final step is 0.5042956.

4.2 Establishment of STARMA model

4.2.1 Establish the spatial weight matrix

The weight matrix is used to represent the spatial correlation of time series. This paper adopt the center distance method to establish the spatial weight matrix [5]. The reciprocal square of the distance between the location of the PdDistrict represents the weight of the mutual influence between the points. Since there are ten PdDistrict, it is necessary to construct a two-dimensional spatial weight matrix of 10*10. The spatial weight matrix is expressed by the following formula:

$$ W_{ij} = \begin{cases} \frac{1}{d_{ij}^2} & \text{if } d_{ij} < \delta \\ 0 & \text{otherwise} \end{cases} $$

(8)

Where $d_{ij}$ is distance between the PdDistrict $i$ and the police area $j$. Since the location of the police district is the actual administrative division in geography, the police district location remained relatively stable from 2003 to 2017 in the data set studied, and $W_{ij}$ did not change with time, so the value would not change with time. After calculation, the approximate value of $W_{ij}$ is shown in table 1.

| A  | B  | C  | D  | E  | F  | G  | H  | I  | J  |
|----|----|----|----|----|----|----|----|----|----|
| A  | 1.00| 0.02| 0.04| 0.05| 0.02| 0.02| 0.02| 0.02| 0.05| 0.02| 0.03|
| B  | 0.02| 1.00| 0.09| 0.06| 0.12| 0.04| 0.04| 0.01| 0.01| 0.35|
| C  | 0.04| 0.09| 1.00| 0.11| 0.07| 0.03| 0.02| 0.02| 0.01| 0.17|
| D  | 0.05| 0.06| 0.11| 1.00| 0.22| 0.11| 0.06| 0.04| 0.03| 0.17|
| E  | 0.02| 0.12| 0.07| 0.22| 1.00| 0.17| 0.13| 0.03| 0.03| 0.32|
| F  | 0.02| 0.04| 0.03| 0.11| 0.17| 1.00| 0.40| 0.04| 0.08| 0.06|
| G  | 0.02| 0.04| 0.02| 0.06| 0.13| 0.40| 1.00| 0.02| 0.05| 0.05|
4.2.2 Model identification and establishment

Based on the results of STACF and STPACF diagram[3], the parameters of the model are estimated by least square method, and the crime prediction model is established. After the maximum likelihood estimation of the processed crime data, the parameters of STARMA model are shown in table 2:

| Estimate | Std. Error | T. Value | P. Value |
|----------|------------|----------|----------|
| phi10    | -0.249751  | 0.082379 | -3.0317  | 0.002471 ** |
| phi11    | 0.594495   | 0.116446 | 5.1053   | 3.701e-07 ***|
| phi20    | -0.083744  | 0.025544 | -3.2784  | 0.001067 ** |
| theta11  | -0.467086  | 0.095379 | -4.8972  | 1.071e-06 ***|

According to the results in the table, all P values in the significance test were less than 0.01, and the original hypothesis was rejected. Therefore, the parameter estimation is valid. The expression of STARMA model is:

\[ z(t) = -0.249751z(t-1) + 0.594495w^{(1)}z(t-1) - 0.083744z(t-2) - 0.467086w^{(1)}e(t-1) + e(t) \]

4.3 Model test

To obtain the final predicted value, the prediction result of STARMA model and LSTM were added. The fitting diagram of the predicted value and the real value is shown in figure 3 which x-coordinate is time, and the y-coordinate is the number of crimes of each month. In the figure, the dotted line represents the predicted value and the solid line represents the true value.

The residuals in the autocorrelation test diagram are uniformly distributed around 0, and the mean value of autocorrelation coefficient is approximately 0. In the test diagram of the normal distribution of the residuals, the residuals are represented as the standard normal distributions, and the mean value is approximately 0, which shows that the residuals are random errors. Therefore, the prediction model is applicable to the space-time series of crimes in this paper.

The prediction results of the space-time hybrid model and the prediction results of the separate STARMA model were tested for errors, as shown in table 3.

Among them, the mean absolute error decreased by 11.50%, the mean square error decreased by 10.81%, the normal mean square error decreased by 19.77% and the relative square error decreased
by 20.45%. It is shown that the model based on the hybrid model of time and space series has a good performance of time and space prediction.

Table 3. Comparison of model test

|       | STARMA       | The hybrid model | Reduced  |
|-------|--------------|------------------|---------|
| MAE   | 7.69941206   | 6.813822902      | 11.50%  |
| RMSE  | 9.591454044  | 8.554871335      | 10.81%  |
| NMSE  | 0.042059704  | 0.033745709      | 19.77%  |
| RSE   | 1.061751952  | 0.844658452      | 20.45%  |

5. Conclusion

Compared with the traditional STARMA model, the hybrid model has its advantages in crime prediction. The addition of LSTM in the model, on the one hand, deals with the instability of criminal data, making it more suitable for STARMA model; on the other hand, it adds non-linear function to the model and improves the precision of model fitting. This paper predict the changes of number of crimes by month and this method allows us making predictions by week, day, hour or even minute, so as to provide information prediction reference for the public security department in a timely manner.

The model is not perfect, and it needs to further study to improve the model. In future, work can be carried out from the following two points: a. For the determination of space-time lag order in STARMA modeling, more efficient algorithms should be explored to compare all possible order models and select optimal parameters. b. the spatial weight matrix is established according to the center distance method. More data about different locations is in need for calculating more accurate weights.

In this paper, a crime prediction model based on the hybrid model of space-time series data of crimes is established, and its application conditions and accuracy are evaluated, providing theoretical and practical reference for optimizing police force allocation, in order to improve the scientific research and evaluation of police command in actual police situations. The realization of criminal prediction is only half successful to truly reduce the crime rate, and more importantly, the public security organs should take corresponding intervention measures according to the prediction results. The public security organs still have a long way to go in how to establish the criminal prediction result response mechanism to decide when and where to take action measures.

Acknowledgment

This work was supported by the National Key R&D Program of China(2016YFB0801100), the National Natural Science Foundation of China (61602489), and the National Cryptography Development Fund (MMJJ20180108).

References

[1] T. Nakaya, K. Yano, Visualising crime clusters in a space-time cube: an exploratory data-analysis approach using space-time kernel density estimation and scan statistics[J]. T GIS, 2010, 14(3):223–239.
[2] MS Gerber. Predicting crime using Twitter and kernel density estimation, DECIS SUPPORT SYST, 61(1):115-125(2014)
[3] BJ Jefferson. Predictable Policing: Predictive Crime Mapping and Geographies of Policing and Race, ANN ASSOC AM GEOGR:1-16(2017)
[4] G Rosser, T Davies, K J Bowers, et al. Predictive Crime Mapping: Arbitrary Grids or Street Networks, J QUANT CRIMINOL, 33(3): 569-594(2017)
[5] T. Cheng, J. Wang, X. Li. A Hybrid Framework for Space–Time Modeling of Environmental Data, GEOGR ANAL, 43(2):188-210(2011)
[6] P. Pfeifer, A three-stage Iterative Procedure for space-time Modeling Phillip, Technometrics, 22(1): 35-47(1980)

[7] M Sundermeyer, H Ney, Schlyuter R. From feedforward to recurrent LSTM neural networks for language modeling, IEEE-ACM T AUDIO SPE, 23 (3): 517-529(2015)

[8] M Cai, J Liu. Maxout neurons for deep convolutional and LSTM neural networks in speech recognition, SPEECH COMMUN.77 (C): 53-64(2016)

[9] A Graves. Long Short-Term Memory, Supervised Sequence Labelling with Recurrent Neural Networks, Springer Berlin Heidelberg, 1735-1780(2012)