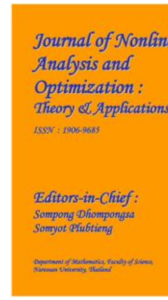
Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 12, No. 1, (2021),
ISSN : **1906-9685**



EVALUATING AND PREVENTING CRIME PATTERNS WITH MACHINE LEARNING ALGORITHMS

[1]Ravishankar Sunkaraboina, [2]Dr Sivaraman, [3]Himagireshwar Rao Matla, [4]Katta Glory
[2]Professor, [1,3]Assistant Professor, [4]UG Student, [1,2,3,4] Department of Computer Science and Engineering, Rishi MS Institute of Engineering and Technlogy for Women, Kukatpally, Hyderabad.

ABSTRACT

Crime prediction is crucial for the creation of policing plans and the execution of measures for crime reduction and control. Machine learning is now the most well-liked prediction method. However, there haven't been many studies that have thoroughly compared different machine learning methods for predicting crime. A large coastal city in southeast China's data on public property crime from 2015 to 2018 is used in this study to assess the predictive skills of several machine learning algorithms. According to results based just on historical crime data, the LSTM model appears to perform better than KNN, random forest, support vector machine, naive Bayes, and convolution neural networks. Therefore, it is recommended that characteristics associated with criminological theories and historical crime data be used for predicting future crime. Not all machine learning techniques are equally effective in crime prediction.
Keywords: naïve bayes, convolution neural networks, KNN and RNN algorithms, crime prediction

I.INTRODUCTION

Spatial and temporal data pertaining to public security have been growing exponentially in recent years. However, not all data have been effectively used to address issues in the real world. To help with crime prevention, several academics have developed crime prediction models. Most of them just used historical crime data to calibrate their forecast algorithms. The prediction of crime hotspots and crime risk are now the two primary topics of crime prediction studies. The crime risk area forecast refers to the relationship between criminal activity and physical environment, both of which originated from the "routine activity theory," based on the important influencing variables of criminal activity. By looking at the historical distribution of crime cases, traditional crime risk estimating approaches often identify crime hotspots and assume that the pattern will persist in the following time periods .For example, considering the proximity of crime places and the aggregation of crime elements, the terrain risk model tends to use crime-related environmental factors and crime history data,and is relatively effective for long-term, stable crime hotspot prediction Many studies have carried out empirical research on crime prediction in different time periods, combining demographic and economic statistics data, land use data, mobile phone data and crime history data. Crime hotspot prediction aims to predictthe likely location of future crime events and hotspots where the future events wouldconcentrate. A commonly used method is kerneldensity estimation. A model that considers temporal or spatial autocorrelations of past events performs better than those that fail to account for the autocorrelation. Recently machine learning algorithms have gainedpopularity. The most popular methods include K-Nearest Neighbor (KNN), random forest algorithm, support vector machine (SVM), neural network
And Bayesian model etc..Some compared the linear methods of crime trend prediction, some compared Bayesian model and BP neural network and others compared the spatiotemporalkernel density method with the random forest method in different periods of crime prediction.

Among these algorithms, KNN is an efficient supervised learning method algorithm SVM is a popular machine learning model because it can not only implement classification and regressiontasks, but also detect outliers. Random forestalgorithm has been proven to have strong non- linear relational data processing ability and high prediction accuracy in multiple fields. Naive Bayes(NB) is a classical classification algorithm, which has only a few parameters and it is not sensitive to missing data. Convolution neural networks (CNN) have strong expansibility,and can enhance its expression ability with avery deep layer to deal with more complex classification problems. Long Short-TermMemory (LSTM) neural network extracts time- series features from features, and has a significant effect on processing data with strong time series trends. This paper will focus on the comparison of the above six machine learning algorithms, and recommend the best performing one to demonstrate the predictive power withand without the use of covariates.

II. LITERATURE SURVEY:

PRINCIPLES OF THEORETICAL CRIMINOLOGY IN PREDICTION OF CRIME HOTSPOTS: The focus of crime hotspot prediction is to forecast future concentration of criminal events in ageographical space. Theoretical criminology provides the necessary theoretical basis. Specifically, several related criminologicaltheories not only provide guidance for us to understand the important influence of location factors in the formation and aggregation of criminal events, but also provide a basic mechanism for the police to use information of crime hot spots for crime prevention or control.It mainly includes routine activity theory, rational choice theory, and crime patterns theory. These three theories are generally considered as the theoretical basis of situational crime prevention. Routine activity theory was jointly proposed by Cohen and Felson in 1979, and has now been further developed through integration with other theories. This theory believes that the occurrence of most crimes, especially predatory crimes, needs the convergence of the three elements including motivated offenders, suitable targets, and lack ofability to defend in time and space. Rationalchoice theory was proposed by Cornish andClarke. The theory holds that the offender's choices in terms of location, goals, methods be explained by the rational balance of effort, risk and reward. Crime pattern theory integrates the routine activities theory and the rational choice theory, which more closely explains the spatial distribution of criminal events. People form ‘‘cognitive map’’ and ‘‘activity space’’ through daily activities. At the same time, potential offenders also need to use their cognitive maps and choose specific locations for crimes in arelatively familiar space. When committing a crime, the offender tends to avoid those places they don't know but to choose the places where the ‘‘criminal opportunity overlaps with cognitive space’’ based on their rational ability. The reason why these places become crime hotspots is that they have the obvious characteristics of ‘‘producing’’ or ‘‘attracting’’ crime. Therefore, the environmental factors of the places need to be considered besides historical crime data for the prediction of crime hotspots.

BUILT ENVIRONMENT DATA :

At present, a large number of studiesshow that the urban built environment has a significant impact on urban criminal behavior, through the impact of crime opportunities to reduce and prevent crime. In the 2007 Global Habitat Report, it was pointed out that the elements of the built environment have an important impact on the occurrence of criminal acts. Point of interests (POIs) data and road network density data are considered as covariates in the crime prediction model.

1) POI DATA :

The urban infrastructure data POI includes the location information and attribute information of various urban facilities . Cateringfacilities, shopping malls and stores are usually located in places with convenient transportation and large flow of people, gathering a large number of different groups of people to generatethe targets for the criminals, while entertainmentplaces attract criminals . These POIs are selectedas covariates of the prediction model.

2) ROAD NETWORK DENSITY:

The conventional definition of road network density refers to total length of roads divided by the size of an areal unit. The area with a denser road network attracts greater flow of people, including potential victims and criminals. Previous studies have shown that the density of road network has an impact on crime rate, especially in public space.

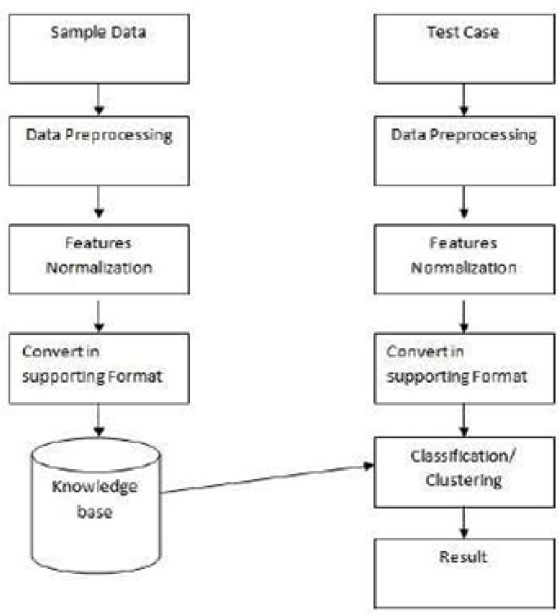
III.EXISTING SYSTEM

The focus of crime hotspot prediction is to forecast future concentration of criminal events in a geographical space. Theoretical criminology provides the necessary theoretical basis. Specifically, several related criminological theories not only provide guidance for us to understand the important influence of location factors in the formation and aggregation of criminal events, but also provide a basicmechanism for the police to use information of crime hotspots for crime prevention or control. It mainly includes routine activity theory, rational choice theory, and crime patterns theory. Datasetsare available in Kaggle.com

IV.PROPOSED SYSTEM

In this paper, random forest algorithm, KNN algorithm, SVM algorithm is used for crime prediction. First, historical crime dataalone are used as input to calibrate the models.Comparison would identify the most effectivemodel. Second, built environment data such asroad network density and poi are added to thepredictive model as covariates, to see ifprediction accuracy can be further improved.

V.SYSTEM ARCHITECTURE:



5.1 IMPLEMENTATION:PREDICTION MODEL:

In this paper, random forest algorithm, KNN algorithm, SVM algorithm and LSTM algorithmare used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the mosteffective model. Second, built environment data such as road

http://doi.org/10.36893/JNAO.2021.V12I1.132-138

network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

A.KNN:

KNN, also known as k-nearest neighbor, takes the feature vector of the instance as the input, calculates the distance between the training set and the new data feature value, and then selects the nearest K classification. If k = 1,the nearest neighbor class is the data to be tested. KNN's classification decision rule is majority voting or weighted voting based ondistance. The majority of k neighboring training instances of the input instance determine the category of the input instance.

B.RANDOM FOREST:

The random forest is a set of tree classifiers

{h(x, βk),k = 1. . . }, in which the meta classifierh(x, βk) is an uncut regression tree constructed by CART algorithm; x is the input vector; βk is an independent random vector with the samedistribution, and the output of the forest is obtained by voting. The randomness of random forest is reflected in two aspects: one is to randomly select the training sample set by using bagging algorithm; the other is to randomly select the split attribute set. Assuming that the training sample has M attributes in total, we specify an attribute number F ≤ M, in each internal node, randomly select F attributes from M attributes as the split attribute set, and takethe best split mode of the f attributes Split the nodes. The multi decision tree is made up of random forest, and the final classification result is determined by the vote of tree classifier.

C.SVM :

SVM, based on statistical learning theory, isa data mining method that can deal with manyproblems such as regression (time series analysis) and pattern recognition (classification problem, discriminant analysis) verysuccessfully. The mechanism of SVM is to find a superior classification hyperplane that meetsthe classification requirements, so that the hyper plane can ensure the classification accuracy and can maximize the blank area on both sides of the hyperplane. In theory, SVMcan realize the optimal classification of linear **D.NB** :

In the field of probability and statistics, Bayesian theory predicts the occurrence probability of an event based on the knowledge of the evidence of an event. In the field of machine learning, the naïve Bayes (NB) classifier is a classification method based onBayesian theory and assuming that each feature is independent of each other. In abstract, NB classifier is based on conditional probability, to solve the probability that a given entity belongs to a certain class.

E.CNN:

CNN uses one-dimensional convolution forsequence prediction, which is the convolution sum of discrete sequences. To convolve the sequence, CNN first finds a sequence with a window size of kernel_size, and perform convolution with the original sequence to obtain a new sequence expression. The convolutional network also includes a pooling operation, which is to filter the features extracted by the convolution to get the most useful characteristics.

F.LSTM:

LSTM is a kind of deep neural network based on RNN. The core of LSTM is to add a special unit (memory module) to learn the current information and to extract the related information and rules between the data, so as to transfer the information. LSTM is more suitablefor deep neural network calculation because of memory module to slow down information loss.Each memory module has three gates, includinginput gate (it), forget gate (ft), and output gate (ot). They are used to selectively memorize the correction parameters of the feedback error function as the gradient decreases. The specific structure is shown in the figure .

http://doi.org/10.36893/JNAO.2021.V12I1.132-138

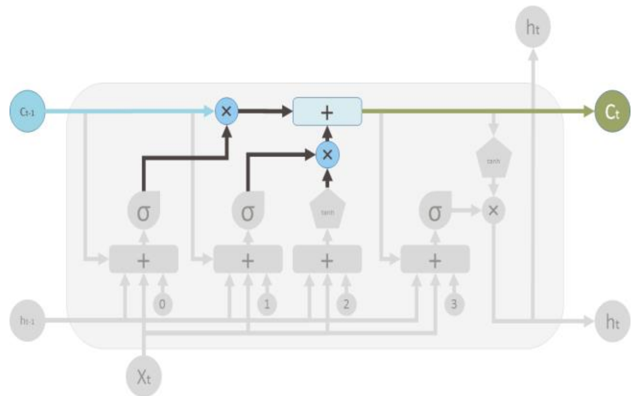


FIGURE 1. The structure chart of LSTM algorithm

In the figure above, LSTM has two state chainsh (hidden layer state) and C (cell state) that are passed over time, only cell state C of RNN is transmitted over time. ht-1 is the value of the current time transmitted from the hidden layer atthe previous time, Xt is the input value at the current time, Ct-1 is the state value of the LSTMmemory cell at the previous time, and Ct is the state value of the memory cell at the currenttime. When ht-1 and Xt pass through the forgetting gate, the information to be discardedis calculated. The value of output to the cell stateis between 0 and 1, 0 means all forgetting, and 1 means all information is reserved. Forgetting gate ft is given by the following equation:

$$ft = \sigma(wf \cdot [ht-1, xt] + bf)$$

where w and b are weight matrix and bias vectorin forgetting gate respectively; σ is activation function Sigmoid. There are two processes for updating new information into a cell. First, the input gate of Sigmoid function is used to calculate the information to be updated, and thena new value kt created by tanh layer is added to the cell state:

$it = \sigma(wi \cdot [ht-1, xt] + bi)$

$kt = \tanh(wk \cdot [ht-1, xt] + bk)$

The results obtained from equation (2) and equation (3) are multiplied and added to the results obtained from the forgetting gate of the previous time cell state value to obtain the current time cell state value, as follows:

$Ct = ft * Ct-1 + ii * kt$ (4)

The final output depends on the cell state. Firstof all, Sigmoid classifies the output results, selects the data to be output, processes the cell state with tanh function, and obtains the state value ht that the hidden layer transfers to the next time. After being processed by sigmoid, ht can obtain the pre output value y at the current time, as shown in equation (5) - equation (7):

$$Ot=\sigma(wO \cdot [ht-1, xt] + bO) ht=Ot*\tanh(Ct)$$

$$y = \sigma(w\,0\,h\,t)$$

VI. Results

http://doi.org/10.36893/JNAO.2021.V12I1.132-138

CONCLUSION

Six In this work, machine learning techniques are utilized to predict the establishment of crime hotspots in a city along the southeast coast of China. These are the conclusions: 1) The LSTM model outperforms the other models in terms of prediction accuracy. It is better at finding patterns and

http://doi.org/10.36893/JNAO.2021.V12I1.132-138

regularities in historical crime data. 2) Including factors linked to the urban built environment improves the LSTM model's ability to predict outcomes. The predictions made using only historical crime data outperform those made using the original model. Compared to other models, our models' forecast accuracy has increased. Rummens et alempirical .'s study on predicting crime hotspots used historical crime data at a grid unit scale of 200 m×200 m, using three models of logistic regression, neural network, and the combination of logistic regression and neural network . In the biweekly forecast, the highest case hit rate forthe two-robbery type is 31.97%, and the highest grid hit rate is 32.95%; Liu et al. Used the random forest model to predict the hot spots in multiple experiments in two weeks under theresearch scale of 150 m × 150 m . The average case hit rate of the model was 52.3%, and the average grid hit rate was 46.6%. The case hit rate of the LSTM model used in this paper was 59.9%, and the average grid hit rate was 57.6%, which was improved compared with the previous research results, For the futureresearch, there are still some aspects to be improved. The first is the temporal resolution of the prediction. Felson et al. revealed that thecrime level changes with time Some studieshave shown that it is useful to check the variation of risks during the day. We chose two weeks as the prediction window. It does not capture the impact of crime changes within a week, let alone the change within a day. The sparsity of data makes the prediction of crime event difficult if the prediction window is narrowed down to day of a week or hour withina day. There is no viable solution to this challenging problem at this time. The second is the spatial resolution of the grid. In this paper, the grid size is 150m $*$ 150m. Future research will assess the impact of changing grid sizes on prediction accuracy. Third, the robustness and generality of the findings of this paper needs to be tested in other study areas. Nonetheless, the findings of this research have proven to be useful in a recent hotspot crime prevention experiment by the local police department at the study size.

REFERENCES

1. U. Thongsatapornwatana, "A survey of datamining techniques for analyzing crime patterns", Proc. 2nd Asian Conf. Defence Technol. (ACDT), pp. 123-128, Jan. 2016.
2. J. M. Caplan, L. W. Kennedy and J. Miller, "Risk terrain modeling: Brokering criminological theory and GIS methods for crime forecasting", Justice Quart., vol. 28, no. 2, pp. 360-381, Apr. 2011.
3. M. Cahill and G. Mulligan, "Using geographically weighted regression to explorelocal crime patterns", Social Sci. Comput. Rev., vol. 25, no. 2, pp. 174-193, May 2007.
4. A. Almehmadi, Z. Joudaki and R. Jalali,"Language usage on Twitter predicts crime rates", Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), pp. 307-310, 2017.
5. H. Berestycki and J.-P. Nadal, "Self-organised critical hot spots of criminal activity", Eur. J. Appl. Math., vol. 21, no. 4, pp. 371-399, Oct. 2010.
6. K. C. Baumgartner, S. Ferrari and C. G. Salfati, "Bayesian network modeling of offender behavior for criminal profiling", Proc. 44th IEEE Conf. Decis. Control Eur. Control Conf. (CDC-ECC), pp. 2702-2709, Dec. 2005.
7. W. Gorr and R. Harries, "Introduction to crime forecasting", Int. J. Forecasting, vol. 19, no. 4, pp. 551-555, Oct. 2003

http://doi.org/10.36893/JNAO.2021.V12I1.132-138