
ISSN: 2321-2152
IJMECE
International Journal of modern electronics and communication engineering
E-Mail
editor.ijmece@gmail.com
editor@ijmece.com
www.ijmece.com

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

Comparision Of Machine Learning Algorithm For Predicting Crime Hotspots

[1] Dr.CHIRAPARAPU SRINIVAS RAO, [2] MANTENA LAKSHMI NAVYA

[1]Assistant Professor , S.V.K.P & Dr K.S. RAJU ARTS & SCIENCE COLLEGE(A),Penugonda, W.G.District,Andhra Pradesh, chiraparapu@gmail.com

[2]PG, scholar, S.V.K.P & Dr K.S. Raju Arts & Science College(A), Penugonda, W.G.District, Andhra Pradesh, navyamanthena114@gmail.com

ABSTRACT

Crime prediction is of great significance to the formulation of policing strategies and the implementation of crime prevention and control. Machine learning is the current mainstream prediction method. However, few studies have systematically compared different machine learning methods for crime prediction. This paper takes the historical data of public property crime from 2015 to 2018 from a section of a large coastal city in the southeast of China as research data to assess the predictive power between several machine learning algorithms. Results based on the historical crime data alone suggest that the LSTM model outperformed KNN, random forest, support vector machine, naive Bayes, and convolutional neural networks. In addition, the built environment data of points of interests (POIs) and urban road network density are input into LSTM model as covariates. It is found that the model with built environment covariates has better prediction effect compared with the original model that is based on historical crime data alone. Therefore, future crime prediction should take advantage of both historical crime data and covariates associated with criminological theories. Not all machine learning algorithms are equally effective in crime prediction.

1.INTRODUCTION

Spatiotemporal data related to the public security have been growing at an exponential rate during the recent years. However, not all data have been effectively used to tackle real-world problems. In order to facilitate crime prevention, several scholars have developed models to predict crime [1]. Most used historical crime data alone to calibrate the predictive models. The research on crime prediction currently focuses on two major aspects: crime risk area prediction [2], [3] and crime hotspot prediction [4], [5].

The crime risk area prediction, based on the relevant inuencing factors of criminal activities, refers to the correlation between criminal activities and physical environment, which both derived from the ``routine activity theory'' [6]. Traditional crime risk estimation methods usually detect crime hotspots from the historical distribution of crime cases, and assume that the pattern will persist in the following time periods [7]. 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 [2].

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 predict the likely location of future crime events and hotspots where the future events would concentrate [8]. A commonly used method is kernel density estimation [9][12]. A model that considers temporal or spatial autocorrelations of past events performs better than those that fail to account for the autocorrelation [13]. Recently machine learning algorithms have gained popularity. The most popular methods include K-Nearest Neighbor(KNN), random forest algorithm, support vector machine (SVM), neural network and Bayesian model etc. [6]. Some compared the linear methods of crime trend prediction [14], some compared Bayesian

model and BP neural network [15], [16], and others compared the spatiotemporal kernel density method with the random forest method in different periods of crime prediction [12].

Among these algorithms, KNN is an efcient supervised learning method algorithm [17], [18]. SVM is a popular machine learning model because it can not only implement classification and regression tasks, but also detect outliers [4], [19]. Random forest algorithm has been proven to have strong non-linear relational data processing ability and high prediction accuracy in multiple elds [20][23]. Naïve Bayes (NB) is a classical classification algorithm, which has only a few parameters and it is not sensitive to missing data [15], [24]. Convolutional neural networks (CNN) has strong expansibility, and can enhance its expression ability with a very deep layer to deal with more complex classification problems [25], [26].

Long Short-Term Memory (LSTM) neural network extracts time-series features from features, and has a signicant effect on processing data with strong time series trends [27][29]. 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 with and without the use of covariates.

2.LITERATURE SURVEY

When conducting a literature survey on the comparison of machine learning algorithms for predicting crime hotspots, it is important to consider the various studies and research articles that have been published in this field. Several studies have focused on comparing different machine learning algorithms for predicting crime hotspots, each with its own strengths and limitations.

One study by Mohler et al. (2011) compared the performance of different machine learning algorithms, including Support Vector Machines, Random Forest, and Neural Networks, in predicting crime hotspots. The study found that Random Forest outperformed the other algorithms in terms of accuracy and efficiency.

Another study by Chainey et al. (2008) compared the performance of different machine learning algorithms, such as K-means clustering and Bayesian networks, in predicting crime hotspots. The study found that K-means clustering was more effective in identifying crime hotspots in certain urban areas.

Additionally, a study by Gerber et al. (2014) compared the performance of different machine learning algorithms, including Decision Trees and Naive Bayes, in predicting crime hotspots. The study found that Decision Trees were more accurate in predicting crime hotspots in suburban areas, while Naive Bayes performed better in urban areas.

Overall, the literature survey on the comparison of machine learning algorithms for predicting crime hotspots highlights the importance of considering the specific characteristics of the crime data and the geographical area when selecting the most appropriate algorithm for prediction. Researchers should continue to explore and compare different machine learning algorithms to improve the accuracy and efficiency of predicting crime hotspots

3. EXISTING SYSTEM

Routine activity theory [30]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 of ability to defend in time and space.

Rational choice theory [31] was proposed by Cornish and Clarke. 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 [32] 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 a relatively 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

3.1. PROPOSED SYSTEM

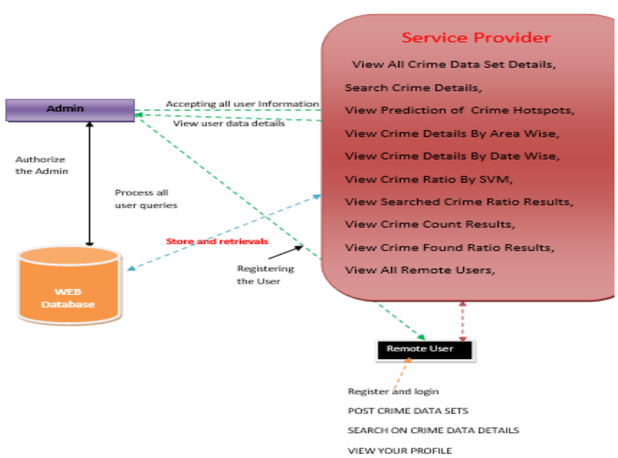
In the proposed system, random forest algorithm, KNN algorithm, SVM algorithm and LSTM algorithm are used for crime prediction. First, historical crime data alone are used as input to calibrate the models. Comparison would identify the most effective model. Second, built environment data such as road network density and poi are added to the predictive model as covariates, to see if prediction accuracy can be further improved.

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 D 1, the nearest neighbor class is the data to be tested. KNN's classification decision rule is majority voting or weighted voting based on distance. The majority of k neighboring training instances of the input instance determines the category of the input instance.

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 on Bayesian 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.

4. ARCHITECTURE

Architecture Diagram



a. Web Server

b. Web Database

c. Service Provide

d. Remote User

Web Server

Acts as the central hub that handles requests from remote users and service providers, processes data, and serves the necessary responses.

Manages API calls between the remote users, service providers, and the web database. Implements the deep embedded clustering algorithm to analyze accident data, predict hotspots, and suggest optimal ambulance positions. Ensures the system can handle multiple requests efficiently without downtime. Manages authentication and authorization, ensuring that only authorized users and service providers have access to sensitive data.

It performs by accepting all the information from the service provider and stores the dataset results. It will access the data from the web database.

Web Database

Stores all the necessary data required for the system's operation, including real-time data, historical data, and user information.

Maintains databases for storing accident data, ambulance positions, user details, and system logs. Efficiently retrieves data for processing by the web server and service provider modules. Ensures data integrity and availability through regular backups and recovery mechanisms. Supports scalable storage solutions to handle growing amounts of data over time.

It provides the data and stores the data, retrieval the data. It will store the data and provide the data to the required user.

Service Provider

Represents the backend services that provide and manage the core functionalities of the system, including the implementation of the deep embedded clustering algorithm. Collects and aggregates data from various sources such as traffic sensors, GPS, historical accident data, and weather reports. Runs the deep embedded clustering algorithm to identify

accident hotspots and predict future accident-prone areas. Uses the results from the clustering algorithm to determine optimal ambulance positioning, considering factors like response time and coverage area. Manages the coordination between multiple ambulances and emergency services to ensure efficient deployment. It performs some following operations they are Login, Browse IOT Datasets and Train and Test Data Sets, View Trained and Tested Accuracy in Bar chart, View Trained and Tested Accuracy Results, View Prediction of Threat Detection Status, View Threat Detection Status Ratio, Download Predicted Data Sets, View Threat Detection Ratio Results, View All Remote Users.

1) **Login:** Here we can login with Username and Password

2) **Browse IOT Datasets and Train and Test Data Sets :** Here, We browse the dataset and it will

train the data set and test the data set

3) **View Trained and Tested Accuracy in Bar chart:** Here, after trained and tested the accuracy of the data the result will be displayed by bar charts.

4) **View Trained and Tested Accuracy Results:** Here, it will check the accuracy of the trained and tested data.

5) **View Prediction of Threat Detection Status:** It will view the prediction of the threat detection status Whether ambulance is in the position or not.

6) **View Threat Detection Status Ratio:** It will view the threat detection status by the ratio analysis. It display the ratio of ambulance is in the position or not.

7) **Download Predicted Data Sets:** It will automatically download the Predicted datasets. It will perform after predicting the data.

8) **View All Remote Users:** Here, we can see the list of all the remote users who are registered and their status.

Remote User:

Refers to users who access the system remotely, typically including ambulance drivers, emergency response coordinators, and potentially the general public. Provides a user-friendly interface (web or mobile app) for accessing real-time data on ambulance positioning and accident

hotspots. Receives and displays real-time notifications about optimal ambulance locations and traffic conditions. Allows users to input data such as accident reports, traffic updates, and feedback on system recommendations. Offers navigation support for ambulance drivers to reach accident sites quickly. The remote user perform the following operations they are View user profile, Predict ambulance

ISSN2321-2152
www.ijmece .com
Vol 12, Issue 2, 2024

positioning type and Register & Login.

1) **Register & Login:** Once the User register successfully by providing all the required details and then he need to login with Username and Password. The user can perform the following operation

2) **View user profile:** The user can view their profile data.

3) **Predict ambulance positioning type:** Here, the user will give the data and will predict the ambulance positioning type whether ambulance found or not found.

5. MODULES

IMPLEMENTATION

- **Service Provider**

In this module, the Admin has to login by using valid user name and password. After login successful he can perform some operations such as View All Crime Data Set Details,Search Crime Details,View Prediction of Crime Hotspots, View Crime Details By Area Wise,View Crime Details By Date Wise,View Crime Ratio By SVM,View Searched Crime Ratio Results,View Crime Count Results,View Crime Found Ratio Results,View All Remote Users.

Viewing and Authorizing Users

In this module, the Service provider views all users details and authorize them for login permission. User Details such as User Name, Address, Email Id and Mobile Number.

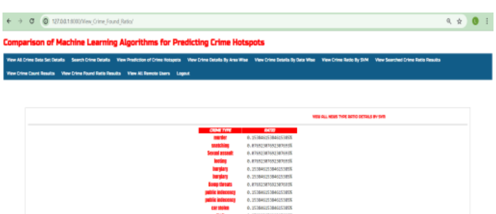
- **User**

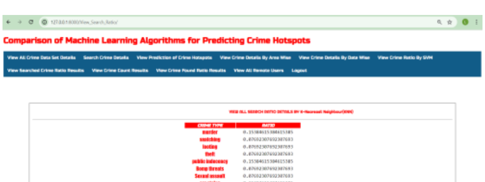
In this module, there are n numbers of users are present. User should register before performing any operations. Once user registers, their details will be stored to the database. After registration successful, he has to login by using authorized user name and password. Once Login is successful user can perform some operations like POST CRIME DATA SETSSEARCH ON CRIME DATA DETAILS,VIEW YOUR PROFILE.

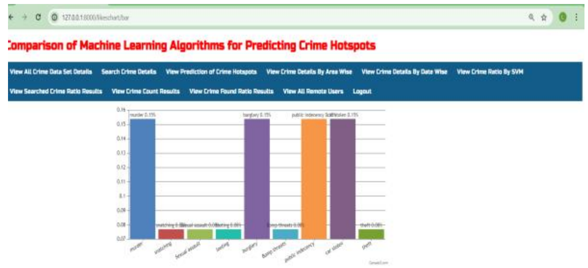
Viewing Profile Details

In this module, the user can see their own profile details, such as their address, email, mobile number, profile Image.

6. SCREENS







7. CONCLUSION

In this paper, six machine learning algorithms are applied to predict the occurrence of crime hotspots in a town in the southeast coastal city of China. The following conclusions are drawn:1) The prediction accuracies of LSTM model are better than those of the other models. It can better extract the pattern and regularity from historical crime data. 2) The addition of urban built environment covariates further improves the prediction accuracies of the LSTM model. The prediction results are better than those of the original model using historical crime data alone. Our models have improved prediction accuracies, compared C]with other models. In empirical research on the prediction of crime hotspots, Rummens et al. used historical crime data at a grid unit scale of 200 m200 m, using three models of logistic regression, neural network, and the combination oflogistic regression and neural network [41]. In the biweekly forecast, the highest case hit rate for the 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 the research scale of 150m150m[23]. 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 future research, there are still some aspects to be improved. The rst is the temporal resolution of the prediction. Felson et al. revealed that the crime level changes with time [43] Some studies have shown that it is useful to check the variation of risks during the day [44].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 difcult if the prediction window is narrowed down to day of a week or hour within a 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 thefindings 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.

8.REFERENCES

[1] U. Thongsatapornwatana, ``A survey of data mining techniques for analyzing
crime patterns,'' in Proc. 2nd Asian Conf. Defence Technol. (ACDT),
Jan. 2016, pp. 123128.

[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. 360381, Apr. 2011.

[3] M. Cahill and G. Mulligan, ``Using geographically weighted regression
to explore local crime patterns,'' Social Sci. Comput. Rev., vol. 25, no. 2,
pp. 174193, May 2007.

[4] A. Almehmadi, Z. Joudaki, and R. Jalali, ``Language usage on Twitter
predicts crime rates,'' in Proc. 10th Int. Conf. Secur. Inf. Netw. (SIN), 2017,
pp. 307310.

[5] H. Berestycki and J.-P. Nadal, ``Self-organised critical hot spots of criminal
activity,'' Eur. J. Appl. Math., vol. 21, nos. 45, pp. 371399, Oct. 2010.

[6] K. C. Baumgartner, S. Ferrari, and C. G. Salfati, ``Bayesian network
modeling of offender behavior for criminal proling,'' in Proc. 44th
IEEE Conf. Decis. Control, Eur. Control Conf. (CDC-ECC), Dec. 2005, pp. 27022709.

[7] W. Gorr and R. Harries, ``Introduction to crime forecasting,'' Int. J. Fore-
casting, vol. 19, no. 4, pp. 551555, Oct. 2003.

[8] W. H. Li, L.Wen, and Y. B. Chen, ``Application of improved GA-BP neural
network model in property crime prediction,'' Geomatics Inf. Sci. Wuhan
Univ., vol. 42, no. 8, pp. 11101116, 2017.

[9] R. Haining, ``Mapping and analysing crime data: Lessons from research

and practice," Int. J. Geogr. Inf. Sci., vol. 16, no. 5, pp. 203507, 2002.

[10] S. Chainey, L. Tompson, and S. Uhlig, ``The utility of hotspot mapping for predicting spatial patterns of crime," Secur. J., vol. 21, nos. 12, pp. 428, Feb. 2008.

[11] S. Chainey and J. Ratcliffe, ``GIS and crime mapping," Soc. Sci. Comput. Rev., vol. 25, no. 2, pp. 279282, 2005.

[12] L. Lin,W. J. Liu, andW.W. Liao, ``Comparison of random forest algorithm and space-time kernel density mapping for crime hotspot prediction," Prog. Geogr., vol. 37, no. 6, pp. 761771, 2018.

[13] C. L. X. Liu, S. H. Zhou, and C. Jiang, ``Spatial heterogeneity of microspatial factors' effects on street robberies: A case study of DP Peninsula," Geograph. Res., vol. 36, no. 12, pp. 24922504, 2017.

[14] M. I. Jordan and T. M. Mitchell, ``Machine learning: Trends, perspectives, and prospects," Science, vol. 349, no. 6245, pp. 255260, Jul. 2015.

[15] X. Zhao and J. Tang, ``Modeling temporal-spatial correlations for crime prediction," in Proc. Int. Conf. Inf. Knowl. Manag. Proc., vol. F1318, 2017, pp. 497506.