Journal of Computer Science and Electrical Engineering
ISSN: 2663-1946
DOI: https://doi.org/10.61784/jcsee3038

PREDICTING AND ANALYZING THEFT CRIME THROUGH TEMPORAL AND SPATIAL DIMENSIONS

SuZhen Luo[1], ZhiSong Wu[2], LiNing Yuan[3*]
[1]*Ministry of Public Sports, Guangxi Police College, Nanning 53028, Guangxi, China.*
[2]*School of Public Policy and Management, Guangxi Police College, Nanning 53028, Guangxi, China.*
[3]*School of Information Technology, Guangxi Police College, Nanning 53028, Guangxi, China.*
Corresponding Author: LiNing Yuan, Email: yuanlining@gxjcxy.edu.cn

Abstract: The data on theft crimes exhibits characteristics such as dynamism, correlation, and uncertainty in both temporal and spatial dimensions. The factors influencing the occurrence of these crimes are complex and include various elements such as population density, education levels, poverty rates, employment status, and climate conditions. The volume and diversity of this data often pose challenges for traditional situational awareness technologies, which rely on criminological theories and case analyses, making it difficult to meet the actual needs of public security agencies. Consequently, crime data mining algorithms based on machine learning and deep learning have gradually become mainstream. This article analyzes the temporal and spatial characteristics of theft crimes, utilizes the Prophet model to predict future incidents, and employs kernel density estimation functions to identify spatial hotspots of theft crimes.
Keywords: Theft crime; Prophet model; Kernel density estimation; Spatial hotspots

1 INTRODUCTION

The crime of theft, a prevalent form of criminal activity, significantly impacts public safety and the quality of life for residents. In recent years, the rapid advancement of technologies such as big data, artificial intelligence, and geographic information systems (GIS), crime prediction and hotspot analysis increasingly important in criminological research. By conducting a thorough analysis of the temporal and spatial distribution of theft crimes, it is possible to enhance the efficiency of police resource allocation and provide a scientific foundation for developing targeted prevention and control measures. This article aims to explore the latest advancements in the temporal prediction and spatial hotspot analysis of theft crimes and to propose recommendations for improving existing methods in conjunction with relevant technological tools.

In recent years, scholars have made significant advancements in the field of crime prediction. Chainey et al. [1] achieved accurate predictions of theft crimes by introducing spatiotemporal pattern recognition technology. Travaini et al. [2] employed machine learning algorithms to analyze various factors influencing crime occurrence, thereby enhancing the accuracy of prediction models. Meanwhile, Jenga et al. [3] proposed a deep learning-based crime prediction model that substantially improved the predictive capability regarding the temporal distribution of theft crimes. In the realm of spatial hotspot analysis, Johnson et al. [4] utilized GIS technology to identify high-crime areas in urban environments and proposed corresponding early warning mechanisms. Mondal et al. [5] discovered the clustering effect of theft crimes in specific regions through spatial autocorrelation analysis, further validating the existence of crime hotspots.

Despite the advancements made in crime prediction and hotspot analysis, several pressing issues remain to be addressed [6~10]. First, the diversity and complexity of data necessitate that researchers thoroughly consider various influencing factors when constructing models, including socioeconomic conditions, population density, and public security investment. Second, effectively integrating prediction results with actual police operations continues to pose a significant challenge. Finally, the variations in crime characteristics and patterns across different regions underscore the importance of localizing and adapting models, which should be a key focus for future research.

Through a systematic study of temporal prediction and spatial hotspot analysis of theft crimes, this article aims to provide valuable insights for researchers in related fields and to offer theoretical support and technical guidance for social security management efforts. In the time dimension, the Prophet model [11] is employed to decompose theft crimes into a three-part structure consisting of trend, seasonal, and event components, thereby facilitating interpretable modeling of crime time series patterns. The trend component utilizes a piecewise linear function with adaptive change point detection, which effectively captures the long-term trajectory of crime rates influenced by exogenous variables. The seasonal component quantifies the inherent periodic characteristics of criminal activities through Fourier series expansion, while the event component enables the model to assess the impact of specific public security measures on crime suppression using custom functions. In the spatial dimension, kernel density estimation functions [12] are applied to visualize the spatial distribution characteristics of crime, providing a quantitative foundation for identifying crime hotspots and formulating prevention strategies.

2 MODELS AND ALGORITHMS

2.1 Prophet Model

The Prophet model is an open-source tool developed by Facebook for time series forecasting. The core of the algorithm includes an additive model and Bayesian inference. The additive part is typically represented as:

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t \tag{1}$$

where, $g(t)$ represents the trend component, which indicates the long-term growth or decline in the time series; $s(t)$ represents the seasonal component, which refers to the periodic fluctuations in the time series (usually related to seasonal, monthly, and other periodic time characteristics); $h(t)$ represents the holiday effect, which refers to the impact of holidays or special events on the forecast results in the time series; ε_t represents noise, which refers to the random fluctuations or disturbances in the time series that cannot be explained by the above components.

In practical applications, time series data may exhibit significant trend changes, such as a sudden increase or decrease in criminal behavior at specific time points. In the Prophet model, these time points where significant trend changes occur are referred to as "change points. Bayesian inference is utilized to automatically detect these change points in the data, allowing for adjustments to the trend. Furthermore, the Prophet model employs Bayesian inference to estimate model parameters, including trend components, seasonal components, and holiday components. For instance, it uses variational inference methods to sample the posterior distribution of these parameters. The incorporation of Bayesian methods enhances the Prophet model's capacity to quantify uncertainty, thereby making time series forecasting results more robust.

2.2 Kernel Density Estimation

The kernel density estimation (KDE) function estimates the probability density surrounding each data point through a smoothing process, generating smooth and continuous density values for each point. This quantifies the probability density distribution of the data in two-dimensional space, which can be intuitively visualized in the form of a heatmap. For theft crime data, the spatial information can be represented as a tuple (x_i, y_i), where x_i and y_i correspond to latitude and longitude, respectively, and i represents any data point in the dataset. In this case, the estimation function of the kernel density estimation can be expressed as:

$$\hat{f}(x, y) = \frac{1}{nh_x h_y} \sum_{i=1}^{n} \left(\frac{x - x_i}{h_x}, \frac{y - y_i}{h_y} \right) \tag{2}$$

where, n is the sample size, h_x and h_y are the bandwidth parameters that control the smoothness in the x and y directions, and K represents the two-dimensional Gaussian kernel function:

$$K(u, v) = \frac{1}{2\pi} e^{-\frac{u^2 + v^2}{2}} \tag{3}$$

where, $u = \frac{x - x_i}{h_x}$ and $v = \frac{y - y_i}{h_y}$ represent the influence of data points (x_i, y_i) on the target point (x, y).

Generate a density estimation map using the kernel density estimation function, where the density value of each point is represented by the intensity of color in a heatmap. The depth of color indicates the magnitude of the density value. Areas with a high concentration of data points and elevated density typically appear as darker regions (hotspot areas), while areas that are sparse or devoid of data generally appear as lighter regions (low-risk areas).

3 EXPERIMENTS AND RESULTS

In the relevant tasks, the open-source Chicago Crime dataset was utilized to predict the frequency of theft crimes over time and to conduct spatial hotspot analysis. For the time prediction task, theft crime data from 2012 to 2015 was employed to train the Prophet model, which achieved daily and monthly predictions for theft crimes in 2016. The experimental results of the time series prediction are presented in Figure 1. In the spatial hotspot analysis task, theft crime data from the first seven days of January through June 2016 was used to train a kernel density estimation function, resulting in a visualization of theft crime spatial hotspots. The experimental results of the spatial hotspot analysis are displayed in Figure 2.

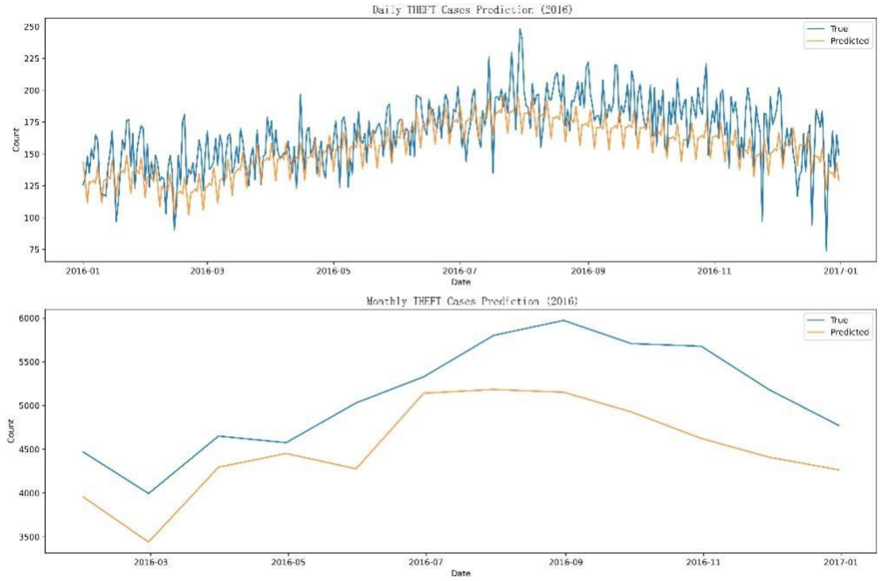
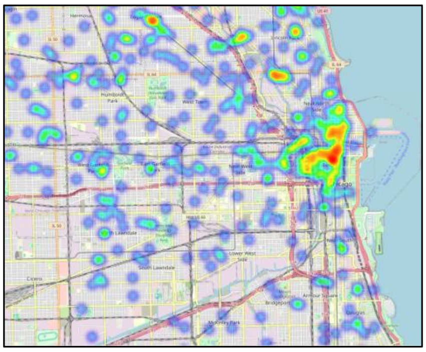
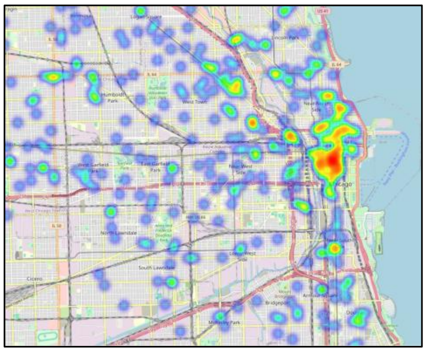


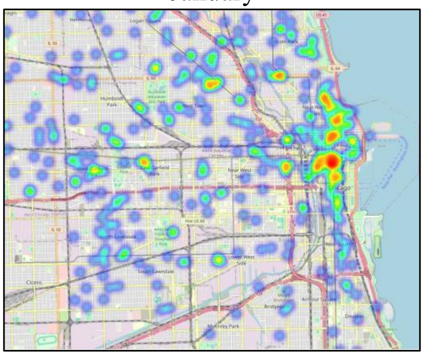
Figure 1 The Experimental Results of Theft Crime Time Series Prediction



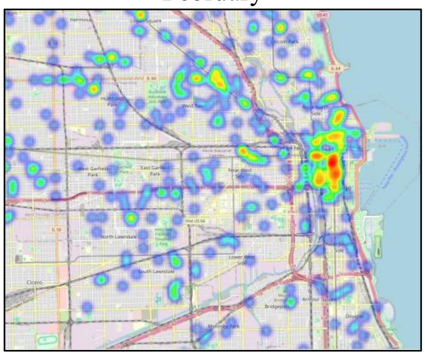
January



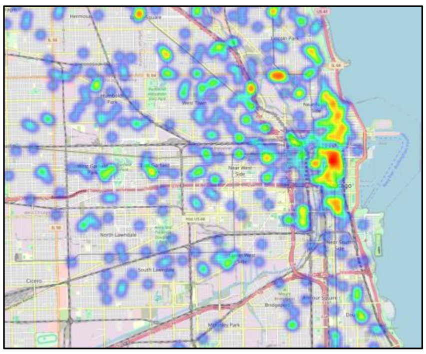
February



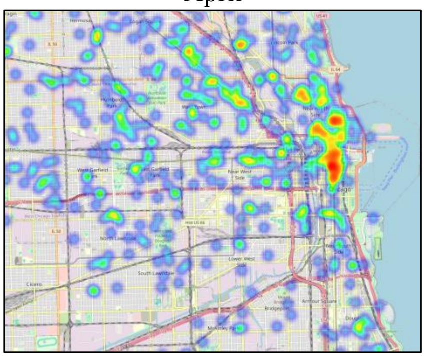
March



April



May



June

Figure 2 Visualization of Hotspots for Theft Crimes (2016)

From a daily trend perspective, theft crimes exhibit periodic fluctuations throughout the week, with peaks generally occurring on weekends and lower numbers reported on Mondays and Tuesdays. From a monthly trend perspective, theft crimes demonstrate significant seasonal variations over the course of the year, with peak periods concentrated from August to November, followed by a notable decline from December to January of the subsequent year. In terms of spatial hotspots, theft crimes are typically concentrated in commercial districts and densely populated residential areas; for instance, bustling shopping areas often experience high foot traffic, making them attractive targets for theft. By employing the Prophet model to analyze the monthly and daily cycles of theft crimes, and integrating it with kernel density estimation to create a crime heatmap, a "time-space" two-dimensional situational awareness framework is established. Theoretically, time series forecasting validates the temporal fluctuation characteristics of crime opportunities as outlined in routine activity theory, revealing the potential influence of time variables on criminal behavior. Additionally, spatial density distribution supports the stability hypothesis of crime hotspots in criminology, identifying the fields that attract crime in high-risk areas.

4 CONCLUSION

Using KDE for theft crime prediction combines the advantages of time series forecasting with spatial data analysis. The Prophet model analyzes historical crime data to accurately forecast future crime trends, enabling law enforcement agencies to allocate resources effectively and implement preventive measures proactively. Meanwhile, the kernel density function leverages spatial data to identify areas with high crime concentrations, highlighting regions at elevated risk for criminal activity. The integration of these two methodologies not only provides insights into the temporal patterns of crime occurrence but also precisely identifies spatial distributions, thereby offering a foundation for developing more targeted crime prevention strategies. This approach demonstrates high predictive accuracy, enhances the efficiency of public safety management, reduces crime rates, and possesses significant social relevance and practical value.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

FUNDING

This work was supported by the Social Science Fund of Guangxi under Grant 23FTQ005

REFERENCES

[1] Chainey S, Tompson L, Uhlig S. The utility of hotspot mapping for predicting spatial patterns of crime. Security journal, 2008, 21(1): 4-28.
[2] Travaini G V, Pacchioni F, Bellumore S, et al. Machine learning and criminal justice: A systematic review of advanced methodology for recidivism risk prediction. International journal of environmental research and public health, 2022, 19(17): 10594.
[3] Jenga K, Catal C, Kar G. Machine learning in crime prediction. Journal of Ambient Intelligence and Humanized Computing, 2023, 14(3): 2887-2913.
[4] Johnson S D, Summers L, Pease K. Offender as forager? A direct test of the boost account of victimization. Journal of Quantitative Criminology, 2009, 25: 181-200.
[5] Mondal S, Singh D, Kumar R. Crime hotspot detection using statistical and geospatial methods: a case study of Pune City, Maharashtra, India. GeoJournal, 2022, 87(6): 5287-5303.
[6] Ratcliffe J. Crime mapping: Spatial and temporal challenges. Handbook of quantitative criminology, 2010: 5-24.
[7] Predictive policing and artificial intelligence. Routledge, Taylor & Francis Group, 2021.
[8] Craglia M, Haining R, Wiles P. A comparative evaluation of approaches to urban crime pattern analysis. Urban Studies, 2000, 37(4): 711-729.
[9] Mityagin S A, Ivanov S V, Boukhanovsky A V. Multi-factorial predictive modelling of drug addiction for large urban areas. 2014 IEEE 8th International Conference on Application of Information and Communication Technologies (AICT). IEEE, 2014: 1-7.
[10] Lin H. Unraveling Crime Dynamics: A Spatiotemporal Analysis of Crime Hotspots. Kent State University, 2024.
[11] Taylor S J, Letham B. Forecasting at scale. The American Statistician, 2018, 72(1): 37-45.
[12] Chen Y C. A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology, 2017, 1(1): 161-187.