Machine Learning for Enhancing Public Safety in Modern Cities

Eugenio Cesario, University of Calabria

Machine learning offers effective techniques to analyze crime data with spatial and temporal information, providing accurate predictions of criminal activities, with the aim to develop more effective strategies for crime prevention and improve public safety.

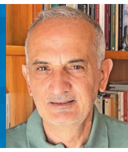
The increasing urbanization occurring over recent years is transforming every aspect of urban society and affecting its sustainable development, bringing both positive transformations (that is, increased opportunities offered in urban areas) and presenting several city management challenges [that is, higher resource consumption (water, electricity), increased traffic, worse air and water quality, public safety services, etc.]. Among the main urban issues, criminal activities are one of the most important social problems in metropolitan areas, because they can severely affect public safety, harm the economy and the sustainable development of a society, as well as reduce the quality of life and well-being of citizens. For this reason, improving strategies to prevent crime effectively has become a crucial issue for policymakers and urban management departments.

MACHINE LEARNING FOR CRIME FORECASTING

Supported by information and communication technology and sensor infrastructures, police departments and public organizations are ever increasing their ability to

Digital Object Identifier 10.1109/MC.2023.3342564
Date of current version: 5 April 2024

 0018-9162/24©2024IEEE



gather detailed data tracking crime events. For this reason, significant volumes of crime data with spatial and temporal information are collected daily and stored in police department databases. The availability of such data offers an opportunity to exploit machine learning and data analytics methodologies for extracting valuable predictive models related to crime events, such as murder cases, child abuse, threats, money laundering, terrorist activities, fraud, etc. The detected data-driven predictive models can empower police departments to optimize their limited resources and design more efficient strategies for preventing criminal events.[1,2,3]

Furthermore, the integration of advanced analytical tools with existing planning software can equip criminal investigators, allowing them to explore large databases efficiently without specialized data science training. The aim of these methods is to develop effective strategies that can actually perform (both spatially and temporally) criminal activity predictions, thus enhancing the efficiency of investigative efforts. Consequently, machine learning-based predictive models should be capable of automatically identifying which areas within a city, commonly referred to as *crime hotspots*,[4] are most impacted by criminal events. Additionally, they should be able to track how crime rates evolve over time in each specific hotspot. This understanding can empower police departments to distribute their resources more efficiently across urban areas, facilitating the strategic deployment of officers to high-risk regions, or reassigning officers from areas expecting a decrease in criminal activities. As a result, machine learning models can enhance crime prevention or enable quick responses to criminal incidents.

From a practical viewpoint, short-term and long-term forecasts can be exploited in different ways. For example, accurate short-term crime forecasts can greatly support police precincts and patrol district activities within small geographic areas. In fact, these forecasts would enable the targeted deployment of patrols to areas with predicted increases in criminal activity, the reassignment of specialized units to regions expecting decreases in crime, and the scheduling of training and vacations during periods of low criminal activity. Similarly, long-term crime forecasts are useful in understanding whether the safety of a particular city area is expected to improve or deteriorate. Furthermore, they can support decisions for an efficient deployment of police officers and other resources, as well as the ability to plan (on the basis of the forecasted risk) for hiring or other resource enhancements across the city area.

RESEARCH AND PROJECTS FOR CRIME PREDICTION

One of the first frameworks for crime data mining has been developed and validated in collaboration with the Tucson and Phoenix Police departments in Arizona.[1] In particular, the system has been exploited in three real-world applications. First, entity extraction methodologies have been used to automatically identify individuals, vehicles, addresses, and personal characteristics from police narrative reports (often containing numerous typos, spelling errors, and grammatical mistakes). The second application concerns the task of detecting identity deception, based on text mining techniques, to discover the actual identity of suspects who have provided false names, fake birth dates, and/or false addresses. The third application has been focused on the detection of subgroups or key members in criminal networks, facilitating the study of interaction patterns among them.

Improving strategies to prevent crime effectively has become a crucial issue for policymakers and urban management departments.

CrimeTensor[5] is an approach used to predict the number of crime incidents across various categories within specific predefined regions. Leveraging tensor learning and spatiotemporal consistency techniques, the framework aims to provide crime predictions considering spatiotemporal categorical correlations in crime events at a fine scale. Crime data are represented as a tensor, and the goal of the approach is to maximize an objective function incorporating spatial, temporal, and categorical information. The prediction process involves the application of tensor decomposition techniques to identify an optimal solution for the defined objective function. To validate the approach, experiments have been conducted using two real-world crime datasets from Xiaogan (China) and New York City (USA).

In 2019, a spatiotemporal crime predictor approach, based on spatial analysis and autoregressive models, has been proposed to perform rolling-horizon crime forecasting.[3] The method exploits density-based clustering techniques to identify high-risk regions, and autoregressive integrated moving average (ARIMA) to perform crime forecasting in each detected region. The approach has been validated on Chicago and New York City crime data, spanning events from 2001 to 2016, showing good forecasting performance on both short-term and long-term prediction horizons.

CrimeTracer[6] is a framework designed for spatial analysis and prediction of crime locations. In particular, the methodology relies on a probabilistic framework that models the spatial behavior of known offenders within their familiar areas, often referred to as *activity spaces*. The framework is applied in the area of Oakland, CA, USA, supporting officers in solving their cases. Also, the system usually reveals that criminals tend to commit opportunistic crimes and recurring violent actions within their activity spaces, rather than venture into unknown territory.

Series Finder[2] is a framework with the aim of automatically detecting crime patterns. In particular the system emulates the instinctive information processing of crime analysts, searching similarities among crime events stored in the databases, to identify the modus operandi and frequent behavioral patterns of a specific offender. The modus operandi is the set of habits that the offender follows and is a type of motif used to characterize the pattern. As more crimes are added to the set, the modus operandi becomes more well-defined, and the pattern should emerge as a frequent pattern of its criminal behavior.

In 2022 an anticrime information support system was proposed, relying on the integration of *k*-means clustering, signal decomposition techniques, and neural networks.[7] The approach aims for the identification of crime urban hotspots and the forecasting of crime trends in each specific area. Interestingly, the application of the system on real cases reveals that the time series of crime occurrences in various regions exhibit only long-term correlations. Similar associations are not detected in the short period, meaning that short-term collaborative law enforcement lacks a theoretical foundation. To overcome this issue, by integrating temporal and spatial factors, the proposed system can assist law enforcement agencies in implementing more precise crime prevention strategies at a micro level. The evaluation of this approach has been conducted on a real-world dataset from Chicago, IL, comprising crime data gathered between 2011 and 2018. A key innovation of this approach is the utilization of bidirectional recurrent neural networks for the forecasting task. The results indicate a strong level of accuracy in short-term predictions, particularly for one-day-ahead forecasts.

Significant volumes of crime data with spatial and temporal information are collected daily and stored in police department databases.

ALGORITHM CHALLENGES IN SPATIOTEMPORAL CRIME DATA ANALYSIS

The application of machine learning techniques in the crime domain has been proven to be an effective and practical approach to discover new insights, relationships, spatial knowledge, regressive models, and predictive patterns from crime data. Although this field of study has made much progress, there is still need for significant enhancements in various crucial aspects of the discovery process. Several challenges persist and need to be addressed, as outlined in the following categories.

Spatiotemporal crime hotspots detection

The incidence of criminal events is not uniformly distributed within a city. In fact, crime trends are strongly affected by the geographic location of the territory (there are low-risk and high-risk areas). For this reason, the detection of crime hotspots from spatiotemporal crime data is a crucial aspect in discovering effective predictive models and developing efficient crime prevention decisions. In the literature, several approaches have been exploited for crime hotspot detection,[4] most of them based on clustering and classification techniques. In particular, density-based clustering algorithms are largely exploited to discover spatial hotspots. However, the density of population, traffic, or events in large cities can vary widely from one area to another, which also makes the incidence of crime events extremely dissimilar in terms of density. For this reason, classic approaches do not work well in identifying multidensity hotspots (that is, different regions having various densities), unless the hotspots are clearly separated by sparse regions.[8] To resolve this issue, multidensity clustering has been studied recently as a more appropriate technique to detect hotspots in large metropolitan areas. Finally, classification approaches (random forest, naive Bayes, J48) and deep learning-based methods have also been used for hotspot detection.[4] However, the application of density-based clustering has been shown to be more popular and effective than classification techniques.

Spatiotemporal crime forecasting

Crime trends can vary with respect to the period of the year, showing seasonal patterns, peaks, dips, etc. that can be very different from hotspot to hotspot. For this reason, an accurate predictive model must be able to automatically detect both which areas in the city are more affected by crime events and how the crime rate of each specific area varies with respect to the temporal period. In the literature, several techniques have been reported for spatiotemporal crime forecasting, including time-series analysis, regression, and deep learning-based techniques. However, several recent research studies highlight that time-series analysis approaches—such as ARMA, ARIMA, and seasonal ARIMA (SARIMA)—achieve better results as compared to other techniques.[4,8] Specifically, SARIMA models show good performance in crime forecasting where data show repeated seasonality patterns and trends; however, their training can be computationally intensive in the presence of nonstationary data. Overall, such approaches are achieving very promising results for future research in crime forecasting.

Accuracy and reliability of data

The quality and reliability of the data to be analyzed represent an important aspect to be considered in crime forecasting research activities. This is due to several factors. The first concerns underreporting: that is, citizens often fail to report a crime, leading to its exclusion from official statistics. For example, a survey conducted by the Malaysian and British police estimates that approximately only 50% of crimes occurring are officially reported to police departments.[4] Another issue within crime data stems from the relatively low accuracy and reliability of classifications made by police officers, sometimes resulting in inaccuracies in the data. Additionally, there are instances where crime records lack important spatial and temporal details, posing further obstacles to a comprehensive analysis. However, the technological advancement in geographical data gathering is enabling the collection of geo-referenced data that are widely used for crime mapping, thus strongly helping crime reporting agencies overcome this issue.

Currently, various publicly accessible data search and exploration platforms provide public access to large spatiotemporal urban data, derived from diverse sources and in different formats. Notably, platforms such as the Chicago Data Portal (https://data.cityofchicago.org/), NYC Open Data (https://opendata.cityofnewyork.us/), and the Los Angeles Open Data (https://data.lacity.org/) are examples of data repositories that make available huge volumes of open data for community access. These resources can be leveraged to enhance the quality of urban services and facilitate data-driven decision-making processes.

Inclusion of sociodemographic data affecting crimes

Data scientists and researchers have identified the inclusion of sociodemographic factors as a crucial aspect for enhancing the precision of crime forecasts. Various reports, such as those published by the U.S. Department of Justice (https://www.ojp.gov/ncjrs/virtual-library/abstracts/demographics-and-criminality-characteristics-crime-america), highlight that the occurrence of crimes is influenced by numerous environmental factors, including (but not limited to): education, employment, gender, average income, marital status, birth rate, substance abuse, religion, and death rate. These factors play a significant role in understanding the dynamics of criminal behavior within a given population. Also, the use of social media is a crucial aspect to be taken into consideration. For this reason, understanding relationships among such factors and their actual influence on crime events is very important for improving crime predictive models, focusing on how best to use sociodemographic data in the areas of crime control, and to effectively target groups and individuals most likely to be involved in criminal activities. C

REFERENCES

1. H. Chen, W. Chung, J. Xu, G. Wang, Y. Qin, and M. Chau, "Crime data mining: A general framework and some examples," *Computer*, vol. 37, no. 4, pp. 50–56, Apr. 2004, doi: 10.1109/MC.2004.1297301.
2. T. Wang, C. Rudin, D. Wagner, and R. Sevieri, "Learning to detect patterns of crime," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECML PKDD)*, 2013, pp. 515–530, doi: 10.1007/978-3-642-40994-3_33.
3. C. Catlett, E. Cesario, D. Talia, and A. Vinci, "Spatio-temporal crime predictions in smart cities: A data-driven approach and experiments," *Pervasive Mobile Comput.*, vol. 53, pp. 62–74, Feb. 2019, doi: 10.1016/J.PMCJ.2019.01.003.
4. U. M. Butt, S. Letchmunan, F. H. Hassan, M. Ali, A. Baqir, and H. H. R. Sherazi, "Spatio-temporal crime hotspot detection and prediction: A systematic literature review," *IEEE Access*, vol. 8, pp. 166,553–166,574, 2020, doi: 10.1109/ACCESS.2020.3022808.
5. W. Liang, Z. Wu, Z. Li, and Y. Ge, "CRIMETENSOR: Fine-scale crime prediction via tensor learning with spatiotemporal consistency," *ACM Trans. Intell. Syst. Technol.*, vol. 13, no. 2, pp. 33–56, 2022, doi: 10.1145/3501807.
6. M. Tayebi, M. Ester, U. Glasser, and P. Brantingham, "CRIMETRACER: Activity space based crime location prediction," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, 2014, pp. 472–480, doi: 10.1109/ASONAM.2014.6921628.
7. Q. Zhu, F. Zhang, S. Liu, and Y. Li, "An anticrime information support system design: Application of K-means-VMD-BiGRU in the city of Chicago," *Inf. Manage.*, vol. 59, no. 5, pp. 1–13, Jul. 2022, doi: 10.1016/j.im.2019.103247.
8. E. Cesario, P. I. Uchubilo, A. Vinci, and X. Zhu, "Multi-density urban hotspots detection in smart cities: A data-driven approach and experiments," *Pervasive Mobile Comput.*, vol. 86, pp. 1–13, Oct. 2022, doi: 10.1016/j.pmcj.2022.101687.

EUGENIO CESARIO is an associate professor of computer engineering at the University of Calabria, 87036 Rende, Italy. Contact him at eugenio.cesario@unical.it.