IOP Conference Series: Materials Science and Engineering

PAPER • OPEN ACCESS

A review on applied statistical and artificial intelligence techniques in crime forecasting

IOP Conf. Series: Materials Science and Engineering **551** (2019) 012030 doi:10.1088/1757-899X/551/1/012030

A review on applied statistical and artificial intelligence techniques in crime forecasting

Alif Ridzuan Khairuddin1[1], Razana Alwee[1] and Habibollah Haron[1]

[1]Applied Industrial Analytics Research Group (ALIAS), School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia

E-mail: alifjam1991@gmail.com

Abstract. Crime forecasting is an important component of crime analysis towards providing early information about possible crime occurrences in the future. Different models have been proposed to assess different crime data structures and representations. From the literature study conducted, there are several types of crime forecasting models that have been introduced such as statistical model and artificial intelligence (AI) model. Recent trends indicate that researchers have shifted their interest towards AI model due to its flexibility in handling variations in crime data structures. The study found that AI model is capable of capturing nonlinearity pattern of crime data in which statistical model fails to achieve. Moreover, the structure of crime data is mostly nonlinear. Thus, an AI model is favoured among researchers towards developing a robust crime forecasting model. This paper provides a review on the background, trends, and challenges on applied statistical and AI model in crime forecasting.

1. Introduction

Forecasting is widely applied in estimating the degree of risks or uncertainties in many fields including but not limited to geographical, finance, economic, engineering, security, health, and various interdisciplinary fields [1]. In the real world, crime is a part of society in which the occurrences are unpredictable by law enforcement agencies [2]. Forecasting provides useful information to government and law enforcement agencies in planning for crime prevention measures. The term crime analysis refers to a discipline practiced in the policing community [3]. The analysis of crime data may help relevant stakeholders to understand the behavior of crime trends and forecast future values based on past observations [4]. This is crucial as crime forecasting assists the authorities by providing useful information for enforcing early crime prevention measures [5].

Crime forecasting is a promising solution to tackle crimes which affect society's livelihood and properties [6]. The main objective of crime forecasting is to efficiently and realistically predict and forecast crime patterns towards providing significant insights about possible future crime trends based on past crime data. Crime forecasting is beneficial to the police and authority as the method is capable of providing early information regarding possible crime occurrences in the future. Advantages of crime forecasting include preventing the recurrence of crimes in specific areas or regions by analyzing the pattern of past crimes occurrences, providing insights for allocating appropriate resources within community for better policing coverage, and providing useful information to authority for planning efficient solutions in crime prevention measures. This paper discussed the background, development, trends, and challenges of applied statistical and AI models in crime forecasting. The rest of the paper is organized into the following sections. Section 2 provides literature study in crime forecasting. Section 3 provides recent trends in crime forecasting and its challenges. Finally, Section 4 concludes the paper and highlights future works.

2. Criminology literatures on crime forecasting

In criminology study, researchers attempt to forecast crimes based on two objectives encompassing identifying potential crime hotspots in specific areas or regions and identifying future crime patterns. Different researchers employ different methods in crime forecasting, depending on the nature and types of data. In crime forecasting, the most widely utilized crime data representation is time series data representation.

Time series data is a set of observations with a discrete-time stochastic nature [7]. Time series data is advantageous as the representation is able to extrapolate time series well ahead of time, into the future [8]. In addition, time series based models are useful when there exists limited knowledge on the underlying data-generating process and in the absence of satisfactory explanatory models that relate the estimation variables to explanatory variables [9]. In time series based model, there are two types of analysis including univariate and multivariate analyses. Univariate analysis involves one variable of time series data while multivariate analysis employs multiple time series data in model development. In most works related to crime forecasting, multivariate analysis is favored in developing a crime forecasting model [8, 10-12]. This is attributed to the ability of multivariate analysis in finding cross-correlations between multiple time series data and discovering distinctive patterns of crime data that never occurred in the past [8, 13-14].

In literature, several types of crime forecasting models have been introduced such as statistical model, AI model, and a mixture of both models, known as hybrid model. Examples of statistical model variations include linear regression, exponential smoothing, and autoregressive integrated moving average (ARIMA). On the other hand, examples of AI model variations cover mostly machine learning-based techniques such as artificial neural network (ANN) and support vector regression (SVR). Table 1 lists existing statistical- and artificial intelligence-based crime forecasting models proposed by different researchers.

3. Recent trends in crime forecasting and its challenges

From the literature study conducted, research on crime forecasting is quite limited. However, crime forecasting is not new because the field has been studied in the last 30 years. Even though crime forecasting shows promising results in estimating future crime rates, the method is rarely applied by most countries [5]. This is due to challenges faced by researchers, associated with crime data and model accuracy. In the real world, historical crime data are limited and rather difficult to obtain [8]. Meanwhile, a majority of available crime data contain 'noise' and are sometimes incomplete. This makes it harder for researchers to develop effective crime forecasting models due to the insufficiency of data. Another reason for the lack of interest by most countries in implementing crime forecasting is related to accuracy, which is plague by the nature of the available crime data. This may eventually to inconsistency of accuracy performance and varying results when the method is employed on different crime datasets.

In the last decade, researchers have gradually shifted their research interest from statistical model into AI based model in crime forecasting. This is associated to the incapability of statistical model in handling abrupt changes in varied environments or systems [26]. This incapability negatively affects the accuracy performance of the forecasting model. Further, statistical model is a parametric technique, which assumes the obtained data are stationary and linear [8]. In other words, the crime data must be first transformed into stationary and linear data prior to feeding into the forecasting model. Another reason for the shift of interest into AI is linked to the inability of statistical model to capture the nonlinearity pattern of crime data [27]. In most cases, a high volume of crime data contains a mixture of both linear and nonlinear patterns. Such mixtures consequently contribute to worsening the performance of statistical crime forecasting models.

Table 1. Existing crime forecasting models proposed by researchers

Reference	Crime model	User data description
[2]	• Artificial Neural Network (ANN) • Nonlinear Autoregressive with Exogenous Input	Crime record data from Selangor and Kuala Lumpur, Malaysia in 2013.
[8]	• Particle swarm optimization (PSO) • Support Vector Regression (SVR) • ARIMA	Historical USA crime rate dataset with external economic data from 1960 to 2009.
[10]	• Exponential Smoothing • ARIMA	Weekly burglary counts over a nearly five year period for the City of Pittsburgh.
[11]	• Interpolation method • Kernel density estimation (KDE)	Crime dataset in Arlington, Texas from 2007 to 2008.
[12]	• Artificial Neural Network (ANN) • Support vector machine (SVM)	Crime dataset comprising aggregated counts of crime and crime related events categorized by police department in USA.
[15]	• Autoregressive (AR)	Crime dataset in Chicago, USA from 2001 to 2014.
[16]	• ARIMA	50 weeks of property crime data in China.
[17]	• Linear regression • Adaptive regression • Decision stump algorithms	Crime and communities dataset in UCI machine learning repository.
[18]	• Random walk • Brown's simple exponential smoothing • Holt's two-parameter linear exponential smoothing	Crime data collected from computer aided system and offense report at Pittsburgh from 1991 to 1998.
[19]	• Space-time Autoregressive	USA regional and state rates of violent and property crimes.
[20]	• Eigenvector Spatial Filtering • Generalized Linear Mixed Model	Vehicle burglary crime data in Plano, Texas from 2004 to 2009.
[21]	• Linear Regression	Complaint dataset from Internet Crime Complaint Centre (IC3) from 2000 to 2005.
[22]	• Artificial Neural Network (ANN) • Probabilistic model	Crime related and spatial dataset in Pakistan.
[23]	• Fuzzy Logic • Self-organizing Map (SOM)	Monthly crime data of 20 county police bureaus in Taiwan from 2003 to 2004.
[24]	• Genetic programming	Crime and communities dataset in UCI machine learning repository.
[25]	• Random forest • Support vector machine (SVM) • Correlation analysis	Crime dataset in India from 2001 to 2012.

To overcome such drawbacks, several nonlinear models have been introduced by researchers to capture the nonlinearity of crime time series data through employing AI techniques. This is because AI techniques contain some nonlinear functions which are able to detect nonlinear patterns in data, thus improving forecasting performance [27]. In addition, AI techniques have the capability to execute global searching, rendering higher efficiency in handling complex environment [27]. Due to that, AI models are able to handle complex patterns in real world crime data effectively. Although AI models look promising, several limitations exist. Firstly, the performance of AI models heavily depends on the structure of data. Secondly, certain AI models such as ANN only work well when dealing with huge datasets but observe varied performances when dealing with smaller datasets. Lastly, it also suffers overfitting issue. Overfitting occurs when AI models take into account the noise of data during data training. Noisy crime data prevent the development of robust AI models.

To overcome the problems in statistical and AI models, researchers have integrated both statistical and AI techniques in their proposed crime forecasting models. From the study conducted, by applying both statistical and AI techniques into a crime forecasting model, the model is capable of capturing both linear and nonlinear patterns in crime data [8]. Findings from existing studies indicate that the hybrid model is effective as compared to individual statistical and AI models, as the hybrid model is capable of overcoming individual limitations resulting in better forecast outcomes. In overall, the literature study conducted found that the structure of real world crime data itself poses a major challenge in producing a good forecast. Further, the study observed that there is no universal model that is able to handle all types of crime data well, instead, researchers tend to develop crime forecasting models that fit their crime datasets. In most cases, the structure of crime data in specific case studies may vary. Thus, different approaches have been proposed to forecast different datasets.

4. Conclusion and future works

Crime forecasting is a huge milestone in crime domain as a mean of preventing offenses from being committed. Forecasting crimes offer the authority with insights of upcoming trends and community with an awareness of occurrences of crimes. An early development of crime forecasting began with an implementation of statistical techniques to forecast crime. Over time, with the introduction of AI technique, researchers have begun to shift away from statistical techniques. Attributed to exponential growth of technology over the last decades, it is now possible to utilize the full potential of AI technique in forecasting crimes. Study found that different researchers have introduced different models to handle different patterns of available crime datasets efficiently. Statistical model is good when handling linear distribution with small datasets. On the other hand, AI model is suitable for handling complex data distributions.

When selecting suitable models to be used, researchers should consider the crime dataset distribution, size, structure, and its representation. In addition, researchers are encouraged to implement multivariate analysis to discover new crime patterns, which could yield greater accuracy for the selected forecasting model. Based on these findings, a good crime forecasting model should possess a low generalization error, low bias trade-off, high variance trade-off, and should be insensitive to outlier. Such features should be sought by researchers in developing a robust crime forecasting model. For future works, the authors will implement several AI techniques to forecast crime rates. The performance of each selected AI technique is then compared and benchmarked. The analysis will then be conducted using a multivariate analysis approach. Through the approach, several external factors such as unemployment or gross domestic product that significantly affects crime rates pattern will be considered towards discovering a new crime pattern with improved model accuracy.

Acknowledgements

This work was supported by Fundamental Research Grant Scheme (FRGS) by Malaysian Ministry of Education (MOE) under grant number R.J130000.7828.4F825 for School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia (UTM).

IOP Conf. Series: Materials Science and Engineering **551** (2019) 012030 doi:10.1088/1757-899X/551/1/012030

References

[1] Jain A and Kumar A M 2007 *App. Soft Comp.* **7** 585-592
[2] Ghazvini A, Nazri M Z B A, Abdullah S N H S, M N Junoh, and Kasim Z A 2015 *Proc.7th Int. Conf. of SoCPaR* p 334-339
[3] Ferreira J, João P and Martins J 2012 *The Elec. J. Info. Sys. Eval.* **15**
[4] Shrivastav A K and Ekata D 2012 *Int. J. of Comp. Sci. and Eng. Tech.* 415-421
[5] Ismail S and Ramli N 2013 *Procedia-Soc. and Behavioral Sci.* **91** 654-660
[6] Huang Y-L, Lin C-T, Yu Y-S, Hsieh W-H and Pai S-M 2015 *Proc. of NCITPA*
[7] Oliveira J F L and Ludermir T B 2015 *Proc. of IEEESMC* p 3867-3872
[8] Alwee R 2014 *Swarm Optimized Support Vector Regression with Autoregressive Integrated Moving Average for Modeling of Crime Rate* PhD, Faculty of Computing, UTM
[9] Khashei M and Bijari M 2011 *App. Soft Comp.* **11** 2664-2675
[10] Huddleston S H, Porter J H and Brown D E 2015 *J. of Business Research* **68** 1810-1818
[11] Hart T and Zandbergen P 2014 *Policing: An Int. J. of Police Strategies and Management* **37** 305-323
[12] Yu C-H, Ward M W, Morabito M and Ding W 2011 *Proc. of 11th IEEE Int. Conf. of ICDMW* p 779-786
[13] Preez J and Witt S F 2003 *Int. J. of Forecasting* **19** 435-451
[14] Han M and Wang Y 2009 *Expert Sys. with App.* **36** 1280-1290
[15] Cesario E, Catlett C and Talia D 2016 *Proc. of IEEE 14th ICPIC, 2nd ICBDICCSTC* p 795-802
[16] Chen P, Yuan H and Shu X 2008 *Proc. of Int. Conf. of FSKD'08* p 627-630
[17] McClendon L and Meghanathan N 2015 *Machine. Learn. and App. : An Int. J. (MLAIJ)* **2**
[18] Gorr W, Olligschlaeger A, and Thompson Y 2003 *Int. J. of Forecasting* **19** 579-594
[19] Santitissadeekorn N, Lloyd D J and Short M B 2015 *Euro. J. of App. Math.* **27** 451-478
[20] Shoesmith G L 2013 *Int. J. of Forecasting* **29** 191-201
[21] Chun Y 2014 *Geo. Analysis* **46** 165-184
[22] Omar A, Kwanbunbumpen A and Alijani D 2007 *Info. Sys. Edu. J.* **15**
[23] Mahmud N, Zinnah K I, Rahman Y A and Ahmed N 2016 *Proc. of 19th ICCIT* p 414-418
[24] Su L-Y 2010 *Comp. and Math. with App.* **59** 737-744
[25] Castelli M, Sormani R, Trujillo L and Popovič A 2017 *J. of Ambient Intel. And Humanized Comp.* **8** 29-36
[26] Baliyan A, Gaurav K and Mishra S K 2015 *Procedia Comp. Sci.* **48** 121-125
[27] Rather A M, Sastry V and Agarwal A 2017 *J. of OPSEARCH* 1-22