ISSN No. 0976-5697

Volume 8, No. 1, Jan-Feb 2017

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Computational Model for Spatio Temporal Crime Event Prediction

Savita Kumari Sheoran
Department of Computer Science & Applications
Chaudhary Ranbir Singh University, Jind (India)–126102

Romika Yadav*
Department of Computer Science & Applications
Indira Gandhi University, Rewari (India) - 122502

Abstract: With recent socio-economic developments the number of crime offences has risen drastically and the demographic factors are strongly cited as determinant of crime rate. In such scenario crime events usually related with a variety of crime socio-economic opportunity factors and hence the traditional system of criminal records has failed to maintain the required level of intelligence and predictive trends. In such irresolute environment the existing literature on crime prediction suggests to blend crime events with spatial-temporal analysis and use Generalized Linear Model for identification of the patterns within a criminal site selection criterion. Such model may improve quantitative prediction even during the temporal changes of criminals may transpires for certain sites. In this paper we enhance the Generalized Linear Model (GLM) for Crime Site Selection (CSS) and examine it for crime events using regression in association with big data technologies. This novel approach is based on standard deviation and autoregressive vector node has been proposed to find similar crime trends among various crime locations for criminal site selection and subsequently use this information for future crime trends prediction. This analysis presents a more significant insight into the scope and complexion of crime prediction with improved certainty.

Keywords: crime prediction, spatio temporal, criminal site selection, generalized linear model, big data

I. INTRODUCTION

With the recent advents in computation technologies the role of computers for storing and the analysis of data have increased many folds. In present era the security and law enforcement agencies like Police, CBI, and Forensic Department etc. are replacing their traditional methods of investigations and crime prediction with advanced technology based tools to increase efficiency and certainty. Actually, crime dataset is not only a record of crime rather it maintain a link between the crime scenes happened in past and prediction of the crime occurrence in the future based on analysis of historical dataset. The crime not only effect on the individual but it terrifically affect to the people of the whole country. The law enforcement and prosecution agencies are operative on that to secure the country from these crime events. Crime forecasting is process that can level the crime rate change over the years and project those changes for future. Such prediction may helpful in deciding the crime prone sites and types of crime which may happen in near future. Such prediction will help the security agencies to manage their resources and deploy their personnel in more cogent way and hence control the crime in advance.

To contribute towards development of a healthy society is responsibility of every citizen and in general of every human being. The technology must be extended for mankind. Therefore, it is coercion responsibility of scientific and research community to contribute toward dwindling the crime rate. Along with the enforcement agencies the researchers need to develop mechanism to analyze the crime events from the voluminous crime data set. A blend of such technology and expert knowledge will essentially predict the modus operandi of criminals and envisage the crime along with associated information that includes types of crime, probable methods and location of crime.

In recent past Generalized Linear Model (GLM) has been ordained by various experts as one of the best plan to efficiently predict the crime happening. But this model stringently fails to comply with frequent change in crime sites and also have restriction of huge database size of the order of Big Data. Since the estimation of crime is more a spatio-temporal concern and hence analysis must focus on the changing Criminal Site Selection (CSS) of various crime types for different years based upon spatial point pattern. In this paper we explore the question to identify the expected areas where the criminal sites have been changed frequently. In nutshell, this research extrapolates Generalized Linear Model for Generalized Criminal Site Selection in the wake of data size of the scale up to Big Data.

The rest of this paper organized as follows: section II reviews the work done by various technocrats and scientists in the field of crime prediction. Section III provides problem statement for research issues addressed in this paper. Section IV proposes model for Crime Prediction through Crime Site Selection approach. Section V presents results obtained on R platform and discussion them for functional capitulation and section VI concludes the paper and elucidate the scope of future research in this area.

II. RELATED WORK

The Spatio Temporal variables are strong determinant to crime rates. Various security experts and social technocrats have explored this field to depth. Xue Yifei and E. Brown Donald [1] predict future crime locations and times using past crime data. Liu Hua and E. Brown Donald [2] enhancing the model to predict future crime location and times using the new point process transition density model based on the likelihood occurrence of spatio-temporal random events. Brown Donald et al. [3] provides spatial forecast methods for terrorist events to forecast the criminal behavior using spatial model. Primicerio Mario [4] proposed mathematical modeling in modern criminology. H. Huddleston Samuel and E. Brown Donald [5] discussed an improvement in the predictive performance hot spot analysis proposed a multivariate spatial choice model of threat assessment of the criminal events in a given geographical area. Identifying forecasting based features to scaling the features of criminal activities [6,7]. F. Greenberg David [8] provides approach of time series analysis of crime rates in which unemployment relationship is developed with the crime events. H.

Huddleston Samuel and E. Brown Donald [9] evaluate time series forecasting using discrete event simulation for security application. Liu Hua and E. Brown Donald [10] crime prediction model using a point pattern based density model that is observed the features of past crimes. Fox Jon and Donald E. Brown [11] explore crime prediction using temporal indicator functions with linear models for spatial temporal events prediction.

III. PROBLEM DEFINITION

The present world is a wordplay industrialization, globalization and privatization. The internet technology has envisioned the doctrine of "shrinking universe" into realty. All this have vowed the crime to its ubiquitous presence and an international subject of complex scenery. As envisage from above section, Crime behaviour is a process that aims to develop crime characteristics. This pervasive nature of crime event renders it a more assorted state then it was in the past and substantiated the subject of Criminology from initial aspect. In such complex the existing strategies of crime prediction goes archaic and hence replaced by new techniques. Below table 1, summarize an analysis of strengths and weaknesses of existing crime prediction practices.

Table 1: Strengths and Limitations of Existing Technique of Crime Prediction

Existing Techniques in Crime Prediction	***Strengths of Existing Techniques***	***Limitations of Exiting Techniques***
Generalized Linear Model (GML) for spatial-temporal event prediction [12]	i. Improves event forecasting accuracy. ii. Provide better resource allocation for law enforcement. iii. Provides computationally efficient with temporal indicators. iv. Provides accurate alternative methods kernel density and hierarchical methods.	i. Using SVAR(Structural Vector Autoregressive Models) for modelling CSS (crime site selection) might improve predictive modelling if temporal changes of criminals may consider for certain sites.

Keeping all these factors in mind, this paper will try to calibrate these issues and launch an inclusive plan to mitigate them in real time scenario.

IV. PROPOSED APPROACH

This section present the approach followed in this paper. A computational model for proposed approach is designed simulated it through "R" tool. A step-wise implementation is also mentioned therein.

A. Methodology

Standard deviation is measured for non-time series. Instead of standard deviation measure in stand alone, autoregressive model provides better results of many application areas of time series. In proposed model firstly standard deviation of each variable is calculated to identify the most deviating variable. Deviating variable is predicted through autoregressive model is the key factor among other variables of crime data set. Further clustering technique is applied to find similar crime trends, which has high probability of crime site selection by criminals. To demonstrate the utility of proposed approach brief points of mathematical model and procedure of model are given in subsequent sections.

B. Mathematical Approach

Let $C_1, C_2, C_3, \ldots\ldots, C_p$ are crime events taken from data sources $D_1, D_2, D_3, \ldots\ldots, D_N$ and $std(X)$ be standard deviation as calculated from dataset to identify the most deviating variable in the complete data set. The mapping of data sets would be optimized using big data tools

$$C_p: D_N \rightarrow O_B \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(i)}$$

where O_B is the optimized big data tool
D_N is the crime data set and
C_p is the objective of to predict crime

The time series data component can be calculated as

$$y_t: \gamma_t + \mu_t \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(ii)}$$

where y_t is time series component
γ_t is seasonal times dependent component and
μ_t is irregular time series component (error)

The input is taken as time series then and crime forecasting is carried out through Autoregressive Integrated method. Finally mean clustering is applied to find similar crime trends which are helpful to identify the probable crime sites.

C. Procedure

Following procedure is adapted to calculation simulation results from proposed computational model:

- *Consider time series data as input.*
- *Choose nature of data as Numeric.*
- *Compute deviating value of each variable using standard deviation method.*
- *Find a deviating year using linear model and Autoregressive model.*
- *Find Structural Vectors for Modeling Criminal Site Selection.*
- *Apply clustering technique to finding similar crime trends.*
- *Predict crime trend and deviating value on the similar crime trends obtained in above step.*

V. RESULTS AND DISCUSSIONS

The Haryana is small state in Northern part of India. It has diverse demographics ranging from more industrialized Cyber Hub of Gurugram to high yield agriculture belt. Due to its diverse conditions it is prone to variety of crimes and envisages a panorama of crime events appropriate to our proposed model. The crime data for analysis are taken from the National Crime Record Bureau (NCRB). So it is very difficult to identify the crime expected areas through traditional techniques. Necessarily to provide intelligence events in which the crime data is increased as crime increased in the major cities of India. Increased crime would be handled through big data techniques in which data would work efficiently. The spatial temporal units of Burglary, Kidnapping, Murder, Robbery and Vehicle Thefts are considered for analysis.

Figure 1: Area of Haryana chosen to analyze and forecasts crime trends

In crime data, the main objective is to cluster the crime locations such as states and districts with similar crime trends over series of data. Since all the crimes do not have equal weightage for example vehicle theft is more deviating and influenced crime over murder, robbery, burglary and kidnapping. Figure 2 depicted that vehicle theft having high standard deviation shows that the data is widely spread and need to analyses over crime sites.

Most probable is to predict the crime events based on the deviating variable i.e. Vehicle Theft. In figure 3 depicted forecasting of crime, in which predicted values are shown in blue dots. The historical values are also highlighted through the black line on the basis of those values the predicted values is observed. This observation number of vehicle thefts from the years 2010 to 2015 and the predicted years is 2016 and 2017.

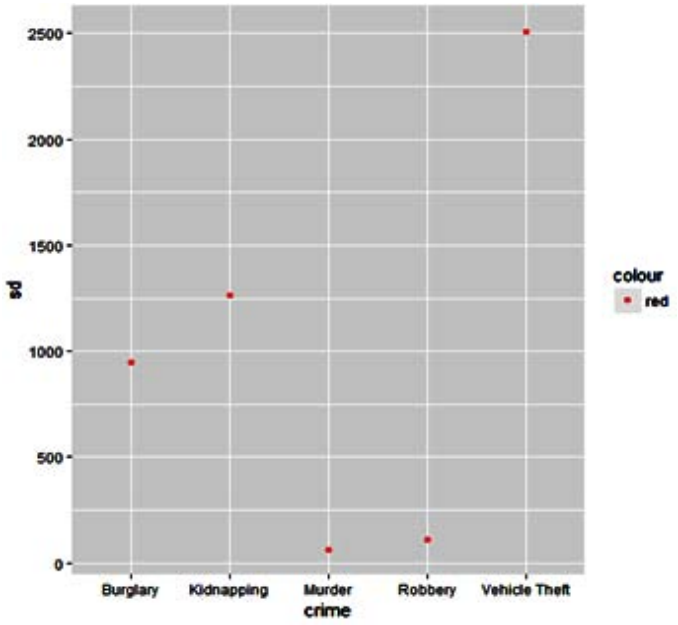

Figure 2: Variables that deviating from each other with their deviated value

Indian Crime data is provided by the NCRB for prediction crime trends. In figure 4 depicted that 21 districts of Haryana of India for year 2013 under important crime i.e. Vehicle Theft. In figure 2 the deviating crime is identified which is vehicle theft. On the basis of that auto regressive model is applied to forecast future crimes of next two years. The next forecasted values is shown in figure 3 in which average of forecasted value of crime is taken for the all 21 districts to find the most criminal sites in the next years. In figure 4 shows that the number of crime which are greater than the average number of crime are shown in blue point. The highlighted blue districts are Gurgaon, Faridabad and Hisar in which high probability of happening of crime.

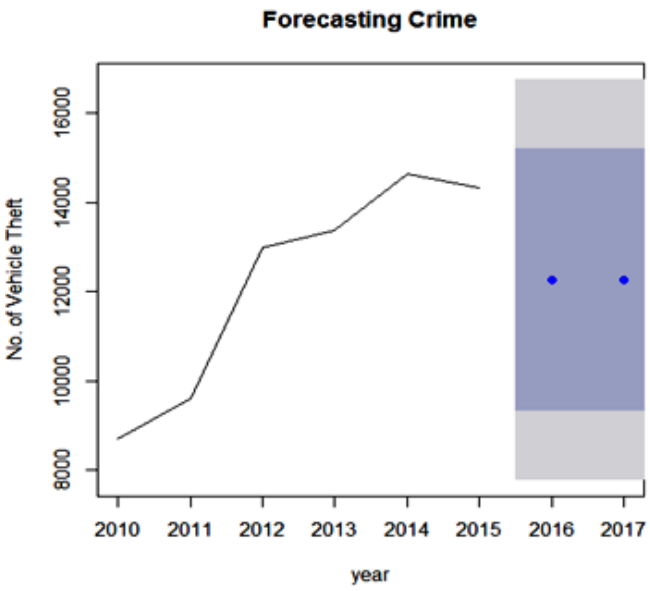

Figure 3: Prediction of Crime Vehicle Theft

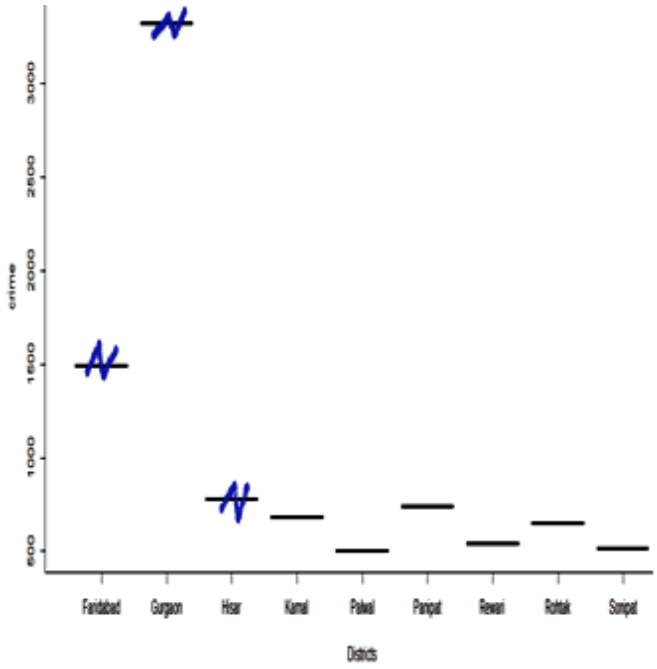

Figure 4 Crime Sites of Various Districts

From the above results and discussions, it is found that proposed methodology gives outstanding results for prediction of crime trends. The proposed approach is useful and easily applied in those types of crime where each dimension has distinguished characteristics.

VI. CONCLUSION AND FUTURE WORK

In this paper, approach for time series data and auto regressive model applied for finding similar crime trends and subsequently crime sites. The results presented in section 5, signify that the Generalized Linear Model (GLM) for Crime Site Selection (CSS) using Big Data just to deliver better results and forecast spatio-temporal crime events with certainty. The Auto regressive model presented in this work suit fit only with the linear data relationship. In order to explore the non-linear regression component of time series data, we are planning to further explore the topic with logistic regression and neural network techniques.

VII. REFERENCES

[1] Xue Yifei and E. Brown Donald, "A Decision Model for Spatial Site Selection by Criminals: A Foundation for Law Enforcement Decision Support", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, pp. 78-85, 2003.

[2] Liu Hua and E. Brown Donald, "New Point Process Transition Density Model for Space–Time Event Prediction", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Applications and Reviews, pp. 310-324, 2004.

[3] Brown Donald et al, "Spatial Forecast Methods for Terrorist Events in Urban Environments", Springer, pp. 426-435, 2004.

[4] Primicerio Mario, "The role of mathematical modelling in modern criminology Comment on "Statistical physics of crime: A review" by M.R. D'Orsogna and M. Perc", Physics of Life Reviews Elsevier, pp. 34-35, 2015.

[5] H. Huddleston Samuel and E. Brown Donald, "A Statistical Threat Assessment", IEEE Transactions on Systems, Man, and Cybernetics-Part C: Systems and Humans, pp. 1307-1315, 2009.

[6] Wu Xiangqian et al, "Offline Text-Independent Writer Identification Based on Scale Invariant Feature Transform", IEEE Transactions on Information Forensics and Security, pp. 526-536, 2014.

[7] R. Scanlon Jacob and S. Gerber Matthew, "Forecasting Violent Extremist Cyber Recruitment", IEEE Transactions on Information Forensics and Security, pp. 2461-2470, 2015.

[8] F. Greenberg David, "Time Series Analysis of Crime Rates", Journal of Quantitative Criminology, pp. 291-327, 2001.

[9] H. Huddleston Samuel and E. Brown Donald, "Using Discrete Event Simulation to Evaluate Time Series Forecasting Methods for Security Applicationss", Proceedings of the 2013 Winter Simulation Conference of IEEE, pp. 2772-2783, 2013.

[10] Liu Hua and E. Brown Donald, "Criminal incident prediction using a point-pattern-based density model", International Journal of Forecasting Elsevier, pp. 603-622, 2003.

[11] Fox Jon and Donald E. Brown, "Using Temporal Indicator Functions with Generalized Linear Models for Spatial-Temporal Event Prediction", New Challenges in Systems Engineering and Architecting Conference on Systems Engineering Research (CSER) Elsevier, pp. 106-111, 2012.

[12] E. Brown Donald and Abbasi Ahmed, "Predictive Analytics", IEEE Intelligent Systems, pp. 6-8, 2015.

[13] E. Brown Donald and Abbasi Ahmed, "Predictive Analytics: Predictive Modeling at the Micro Level", IEEE Intelligent Systems, pp. 6-8, 2015.