3rd International Conference and Workshops on Recent Advances and Innovations in Engineering, 22-25 November 2018
(IEEE Conference Record # 43534)

Crime Prediction Using Auto Regression Techniques for Time Series Data

Romika Yadav
Research Scholar
Department of Computer Science & Engineering
Indira Gandhi University Meerpur, Rewari - INDIA
romikasim@gmail.com

Savita Kumari Sheoran
Associate Professor
Department of Computer Science & Engineering
Indira Gandhi University Meerpur, Rewari - INDIA
savita.sheoran@gmail.com

***Abstract*- Crime is undesired anti-social behavior and poses serious threat to society. The civilized societies make everything possible to reduce crime within its regime of influence. Alarming the crime prone areas in advance is one of the best strategies for crime to be ceased to happen. The recent socio-economic developments and proliferation of internet technologies have turned the crime into a global phenomenon. In such scenario the crime data to be dealt is huge in volume, diverse in variety and highly location dependent. Hence the contemporary crime data set is highly spatio-temporal in nature where the traditional system of criminal records has failed to maintain the desired level of intelligence and make a substantial prediction. A blend of 'Big data' tools for data management and Generalized Linear Regression for statistical analysis is used to draw a useable inference from such time series data set. Such enhancement is supportive to detect similar crime trends among various crime locations for criminal site selection. Consequently ARIMA (Auto Regressive Integrated Moving Average) model affords to minimize the error generated in the predictive model. This research paper aims to locate the offender site in advance with more accuracy. We have explored the Auto Regression Techniques to accurately predict the crime with minimum error for such time series data by identifying the relationship among crimes attributes. The experimental result obtained using "R" tool show that our formulation work well for all parameters and improves certainty in prediction.**

***Keywords* - crime location, crime prediction, offenders, ARIMA, crime attributes.**

I. INTRODUCTION

The recent advent in technology has made it Omni-applicable in every wake of life. Now a day, almost every record is in digitalized form and stored in memories as data. These data can be analyzed efficiently using high power computing machines on the robust software platforms. Crime is undesired anti-social behavior but keeping record of crime events is highly desired practice keep check on the further occurrence of such event. By keeping a record of such events a forecasting about the future crime trends can be made out. The law enforcement agencies which have duty to alleviate crime from society analyze the crime records (data) and make a meaning prediction to aware the public, as a preventive measure. The advent of internet technology has networked the whole world in to a single village turning the crime into globalised phenomena. The technology plays a dual edge role in crime prediction. On side it make crime a more complex happening and at the same time offer a computational platform for better analysis and forecasting the crime scene.

Crime forecasting is process of projecting the future crime based on past record. The present day crime data set posses distinguished feature, apart from past, which make it impossible to be handled with traditional data management techniques. The volume, velocity and veracity of crime data set almost or fully fall within the regime of 'Big data'. Similarly the traditional tools of statistical analysis fail drastically to establish a meaning relationship between such data set. The Generalized Linear Regression used for statistical analysis is poor at error estimation. Therefore, both the criminal data management as well as analysis to forecast an accurate and meaning pattern are a big challenge before the academic and research community working in this area.

Since Big Data analytics and ARIMA (Auto Regressive Integrated Moving Average) model work well for data management and to reduce error in prediction respectively for time series data. This research paper work on cardinal principle of exploring Auto Regression Techniques blended with Big Data tools for time series analysis. The Crime Dataset (CD) has a link between the Crime Scenes (CS) and Prediction of the Crime (PC) occurrence in the future i.e. $CD = \{CS, PC\}$. The analysis focuses on the changing Criminal Site Selection (CSS) of various crime types for different years. The simulation is carried out to identify the expected areas where the criminal sites have been changed every year. Here we have considered spatial-temporal patterns of crime events to generate a crime predicted location based on analysis of crime distance and crime time with the method of weighted average and finally identifying the relationship among those attributes which are expected levels of crime. The rest of this paper organized as follows: section II gives details of the work done by various technocrats and scientists in the field of crime prediction. Section III gives problem definition of the Crime Prediction. Section IV gives proposed approach. Section V Consists results and discussions and VI concludes the paper and project future research scopes.

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

978-1-5386-4525-3/18/$31.00 ©2018 IEEE

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. Luchaup Daniel et al. [12] propose a method to extract the characteristics of offenders from their criminal behavior. They have applied speculative parallel pattern matching system to get the important characteristics of offender. Xiang Yang et al. [13] proposed a system COPLINK to visualize the criminal relationships. Kengpol Athakorn and Neungrit Pakorn [14] using decision support system with efficient prediction tool and risk assessment analysis of criminal activities. Bosse Tibor et al. [15] provide relation between cognitive and biological modelling of criminal behavior based on the biological factors. Kanhangad Vivek et al. [16] proposed the A unified framework for contactless hand verification application system. Xia ZhengYou [17] proposed a technique that choosing the next criminal investigate object in the criminal probabilistic network. Baumgartner K. et al. [18] erect Bayesian network on criminal profiling data to infer the characteristics of unknown offender of the crime scene. But it only suspects predicted correctly single victim homicides. W. Brahana John et al. [19] discussed artificial intelligence crime analysis system initially on rule based then investigating on machine learning and neural network to improve the performance of existing data source. C. Oatleya Giles and W. Ewart Brian [20] use statistical methods for multi-dimensional scaling and data-mining technologies which are based on neural network. Li Sheng-Tun et al. [21] propose intelligent decision-support system using fuzzy self-organizing map and rule extraction for crime prevention. Dahbur kamal and Muscarello Thomas [22] proposed a classification system for serial criminal patterns in the database that provides advantages over the traditional approach of using single clustering to accommodate all the attributes. Thammaboosadee Sotarat et al. [23] provide a multi-stage classifier framework to identifying criminal legal codes.

III. PROBLEM DEFINITION

The internet technology has envisioned the doctrine of "shrinking universe" into realty. It has vowed the crime to its ubiquitous presence and an international subject of complex nature. The pervasive nature of crime event renders it in a more assorted state then it was in the past and substantiated the subject of Criminology from initial aspect. In such complex situation the existing strategies of crime prediction goes archaic and hence replaced by new techniques. Table 1; summarize an analysis of strengths and weaknesses of existing crime prediction practices.

TABLE I. STRENGTHS AND LIMITATIONS OF EXISTING TECHNIQUES 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]	[1]Improves event forecasting accuracy. [2] Provide better resource allocation for law enforcement. [3] Provides computationally efficient with temporal indicators. [4] 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. ii. Consequently location prediction using offender's crime. iii. Finally detection of crime attributes based on crime locations.

The summary in table 1, reflects a research gap in developing error optimal crime prediction model for time series data. 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. MODELING OF COMPUTATIONAL DETAIL

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

A. Methodology

In proposed model first the standard deviation of each variable is calculated to identify the most deviating variable. The deviating variable predicted through autoregressive model is the key factor 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 are given in subsequent sections.

Let C_1, C_2, $C_3, \ldots\ldots\ldots$, 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

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

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

In ARIMA model Y_t can be expressed as a function of past value. When forecast a value past the end of the series, observed series values that aren't yet observed.

Consider the forecast equation Y_t: $\gamma_t + \mu_t + \delta t$ In this equation, Y_t is a linear function of the values of Y at the previous times. Suppose that we have observed n data values and wish to use the observed data and forecast the value of Y_{n+1} and Y_{n+2}, the values of the series at the next two times past the end of the series.

The equation for forecasting the value at time $n+2$ presents a problem. It requires the unobserved value of Y_{n+1}. The solution is to use the forecasted value of (the result of the first equation).

In general, the forecasting procedure, assuming a sample size of Y, is as follows:

The time series data component can be calculated as

y_t: $\gamma_t + \mu_t + \delta t$ ..... (ii)

Where y_t is time series component

γ_t is seasonal time dependent component

μ_t is irregular time series

and δ_t is residuals (error) of time series

The input is taken as time series then and crime forecasting is carried out through Autoregressive Integrated method. Time series parameters associated with their chances of error generate in predictive model. That will help to identify the accuracy level of model.

According to the various psychologist the location of criminals are interconnected and related to the offender's residence. So we build models on the profile of offender's residence and predicting the next crime scenes. This will lead to reduce the workload of police department to investigate serial criminals.

1. *Predict the Residence of the offender*

   If the distance between the point and location of past crime is minimum, the point can be selected as the offender's residence.

$$\min = \sum_{i=1}^{n} \sqrt{(x - x_i)^2 + (y - y_i)^2}$$

   where x >0 , y>0

   (x_i , y_i) – the location of crime scene

   (x , y) - the offender's residence

2. *Predict time of the next crime*

   Time of the crime in the past crime scenes looks like a straight lines as time varies in a series of intervals. So a linear fitting has to make on the time of past crime scenes. Then we predict the time of the next crime is to be happened with the help of linear fitting.

$$P(x_i) = mx_i + c$$

$$P(x_{n+1}) = mx_{n+1} + c$$

   where $P(x_i)$ - Time of crime

   $P(x_{n+1})$ - The predicted crime of time

3. *Predict the location of the next crime with geographical profile*

   The possible location of offender's next crime can be near of the residence of the offender's and location of the last crime scene.

This model determines the relationship between the crime attributes and expected level of crime.

B. 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.*
- *Consequently ARIMA (Auto Regressive Integrated Moving Average) Model parameters provide results associated with their chances of error generate in predictive model.*
- *Find Structural Vectors for Modeling Crime Site.*
- *Predict offender's location by analysis of crime distance and weighted average identifies crime time.*
- *Apply clustering technique to finding similar crime trends.*
- *Detection of crime attributes wise trend*

V. IMPLEMENTATIONS AND RESULTS

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.

The main task in analyzing the crime data is to cluster the crime locations based upon attributes such as states and districts with similar crime trends over series of data. Since all the crimes do not have equal weightage. For instance 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.

Fig. 1. Area of Haryana chosen to analyze and forecasts crime trends

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.

To test predictability of our model various cities of Haryana state in India are considered. The records of vehicle theft have been considered to predict the crime location using the offender's crime based data.

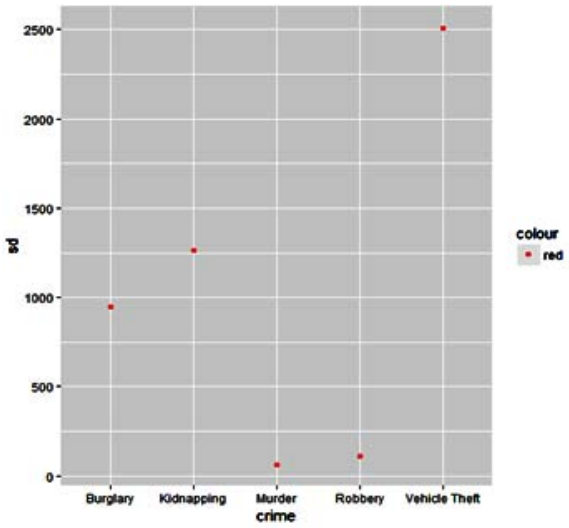

Fig. 2. Variables that deviating from each other with their deviated value

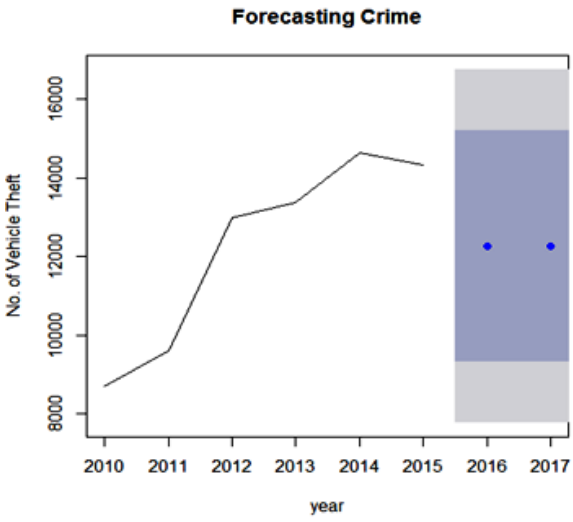

Fig. 3. Prediction of crime vehicle theft

Figure 4 the Red dots connected with blue lines shows the minimum distance of the crime location data using **Model A**. The red dots here are showing the various locations of crime scenes in which crime is happened and having their minimal criminal records between their cities. The blue line connecting each city to showing the offender's pattern to choose crime cities in orderly. So it will show the criminal pattern

Fig. 4. Locations of each crime

In Figure 5 crime time is predicted using **Model B**. The blue dots shows linear model to predict the next crime scene. Here procedure is considered linear fitting model in which points are located along with crime time and crime scenes. Most of the points are lying nearby so clearly shows that criminal activities may be happened surrounded areas of their cities. The next crime location has been identified through the values associated with the fitted values in figure 5.

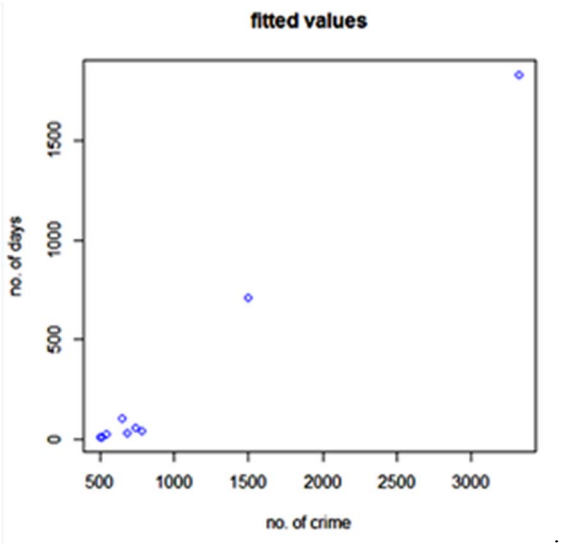

Fig. 5. Fitted values of expected crime

In figure 6 Predicted locations has been identified shown in green color arrow sign. These are expected areas of the offenders to the next crime location using Model A, B & C. The green arrows clearly identified the forecasted locations in which crime may be occurring. So these locations are prone to crime and nearby those locations which are highly chosen offender's for crime.

Finally, this model determines the relationship between the attributes and their expected level of crime. The results of the above method are fitted on the big data analytics 'R' tool in which data is analyzed and generating output accordingly.

Fig. 6. Predicted locations

Figure 7 shows the attribute-wise crime trend covering shopping centers, public offices, hospitals and banks. These attributes are common places where theft occurred frequently and also need to analyze those crime places attributes has relatively number of crimes, which are indicates some of the attributes where chances of crime happening is high. The procedure to get these results clustering technique is applied to find similar crime trends. In which every districts having crime attributes where crime is happened. So these crime attributes values are clustered of various districts of Haryana. Some attributes are having likelihood more number of crimes and rest of having less. Similar crime patterns along with their crime attributes will be shown in the figure 7. Clearly showing in figure 7 the shopping centers and public offices having highest crime places.

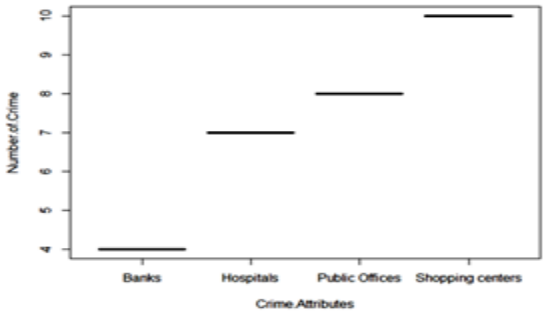

Fig. 7. Attributes wise crime trend

Above discussions envisage that the proposed methodology gives outstanding results for prediction of crime trends. The proposed approach is useful and easy applied in crimes where each dimension has distinguished characteristics.

VI. CONCLUSION AND FUTURE WORK

In this paper studied the auto regressive model applied on time series data to find 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 deliver better results and forecast spatio-temporal crime events with certainty. The Auto regressive model presented in this work fit only with the linear data relationship. In order to further explore the non-linear regression component of time series data, it may be tested with logistic regression and neural network techniques.

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] LuchaupDaniel et al, "Speculative parallel pattern matching", IEEE Transactions on Information Forensics and Seurity, pp. 438-451, 2011.

[13] Xiang Yang et al, "Visualizing criminal relationships: comparison of a hyperbolic tree and a hierarchical list", Decision Support Systems Elsevier, pp. 69-83, 2005.

[14] KengpolAthakorn and NeungritPakorn, "A decision support methodology with risk assessment on prediction of terrorism insurgency distribution range radius and elapsing time: An empirical case study in Thailand", Computers & Industrial Engineering Elsevier, pp. 55-67, 2014.

[15] Bosse Tibor et al, "On the relation between cognitive and biological modelling of criminal behavior", Computers in Human Behavior Elsevier, pp. 1593-1611, 2011.

[16] Kanhangad Vivek et al, "A unified framework for contactless hand verification", IEEE Transactions on Information Forensics and Security, pp. 1014-1027, 2011.

[17] Xia ZhengYou, "Fighting criminals: adaptive inferring and choosing the next investigative objects in the criminal network", Knowledge Based Systems Elsevier, pp. 434-442, 2008.

[18] Baumgartner K. et al, "Constructing bayesian networks for criminal profiling from limited data", Knowledge Based Systems Elsevier, pp. 563-572, 2008.

[19] W. Brahana John et al, "AICAMS: artificial intelligence crime analysis and management system", Knowledge Based Systems Elsevier, pp. 355-361, 1998.

[20] C. Oatleya Giles and W. Ewart Brian, "Crimes analysis software: 'pins in maps', clustering and Bayes net prediction", Expert Systems with Applications Elsevier, pp. 569-588, 2003.

[21] Li Sheng-Tun et al, "An intelligent decision-support model using FSOM and rule extraction for crime prevention", Expert Systems with Applications Elsevier, pp. 7108-7119, 2010.

[22] Dahbur kamal and Muscarello Thomas, "Classification system for serial criminal patterns", Artificial Intelligence and Law, Kluwer Academic Publishers. Printed in the Netherlands, pp. 251-269, 2003.

[23] Thammaboosadee Sotarat et al, "A framework of multi-stage classifier for identifying criminal law sentences", Proceedings of the International Neural Network Society Winter Conference Elsevier, pp. 53-59, 2012.