

A Survey Paper on Crime Prediction Technique Using Data Mining

Nikhil Dubey¹, Setu Kumar Chaturvedi², Ph.D

¹M.Tech. Research scholar Dept. of Computer Science and Engg. Technocrats Institute of Technology, Bhopal

²Professor & Head Dept. of Computer Science and Engineering Technocrats Institute of Technology, Bhopal

ABSTRACT

Crime prediction is an attempt to identify and reducing the future crime. Crime prediction uses past data and after analyzing data, predict the future crime with location and time. In present days serial criminal cases rapidly occur so it is an challenging task to predict future crime accurately with better performance. Data mining technique are very useful to solving Crime detection problem. So the aim of this paper to study various computational techniques used to predict future crime. This paper provides comparative analysis of Data mining Techniques for detection and prediction of future crime.

Keywords- Crime Data, Crime Prediction, Data Mining Technique, Predictive Accuracy.

I. INTRODUCTION

Crime has been a part of society ever since laws were first approved. It is defined as an act committed or omitted in violation of a law forbidding or commanding it and for which punishment is imposed upon conviction.

Crime is classically unpredictable. It is not necessarily random, neither does it take place persistently in space or time. A Good theoretical understanding is needed to provide practical crime prevention solutions that equivalent to specific places and times. Crime analysis takes past crime data to predict future crime locations and time [1]. Crime prediction for future crime is a process that find out crime rate change from one year to the next and project those changes into the future. Crime predictions can be made through both qualitative and quantitative methods. Qualitative approaches to forecasting crime, as environmental scanning, scenario writing, are useful in identifying the future nature of criminal activity. In contrast, perceptible methods are used to predict the future scope of crime and more specifically, crime rates a common method for develop forecasts is to projects annual crime rate trends developed through time series models. This approach also involve relating past crime trends with factors that will influence the future scope of crime [2].

1.1 Standard Crime Prediction Techniques

1.1.1 Centrography

Centrography is the inspection of the descriptive statistics used in measurements of central susceptibility, such as centrality of population or population potential. In Centrography, crimes are assigned x and y coordinate and the center of mass

is measured as follows:

$$x_{center} = \sum_{i=1}^n \frac{x_i}{n}$$
$$y_{center} = \sum_{i=1}^n \frac{y_i}{n}$$

Intuitively, Centrography find the mean x-coordinate and the mean y-coordinate and associate this pair with the criminals residence [3].

1.1.2 Journey to Crime

Journey to crime supports the notion that crimes are likely to occur closer to an offender's home and follow a distance-decay function (DDF) with crimes less likely to occur the further away an offender is from their home base. It is concerned with the 'distance of crime' and that offenders will in general travel limited distances to commit their crime[4].

1.1.3 Routine Activity Theory

According to Cohen and Felson (1979), the union of three elements in time and space are required for a crime to occur: a likely offender, a suitable target and the absence of a capable guardian against crimes [5].

1.1.4 Circle Theory

In the circle method, the distances between crimes are measured and the two most distant crimes are chosen. Then, a extreme circle is drawn so that both of the points are on the great circle. The midpoint of this extreme circle is then the assumed

location of the criminal's residence and the area bounded by the great circle is where the criminal operates. This model is computationally economical and easy to understand. Moreover, it is easy to use and requires very little training in order to master the technique. However it has some drawback.

For example, the area given by this method is often very large and other studies have shown that a smaller area success. Additionally, a few outliers can generate an even larger search area, thereby further slowing the police effort [6].

1.2 Advantages of Crime Prediction Techniques:

The advantages of crime prediction techniques are as follow:

- Visualize criminal network.
- Risk is reduced.
- Increase crime analysts work productivity.

1.3 Disadvantages of Crime Prediction Techniques:

The disadvantages of crime prediction techniques are as follow.

- Unexpected occurrence.
- Takes lot of computational time.

1.4 Challenges of Crime Prediction Techniques:

The challenges of crime prediction techniques are as follow:

- Provide an accurate prediction for the location of the criminal.
- Collecting and managing large volumes of accurate data.
- Provide good performance by combining prior knowledge.
- Fast evolution of crime site data.
- Maintain effective crime analysis resource.

In the next section various data mining techniques for crime prediction are introduced.

II. DATA MINING TECHNIQUES FOR CRIME PREDICTION

The goal of prediction is to forecast the value of an attribute based on values of other attributes. In the prediction techniques a model is first created based on data distribution and then model is used to predict future on unknown value. The basic data mining techniques are introduced which are used to predict crime.

- I. Support Vector Machine
- II. Multivariate Time Series Clustering
- III. Bayesian Network
- IV. Neural Network

V. Fuzzy Time Series

2.1 SUPPORT VECTOR MACHINE

Support Vector Machines were introduced by Vladimir Vapnik and colleagues. The first main paper seems to be (Vapnik, 1995) [7].

Support Vector Machines (SVM) have recently gained prominence in the field of machine learning and pattern classification. Classification is achieved by realizing a linear or non-linear separation surface in the input space.

Hot-Spots Prediction Using Support Vector Machine [8]. This technique used Support vector Machine for Crime prediction and experiment the following approach: for a given percentage of the data and a predefined level of crime rate, we select a subset of the crime dataset to label; and then based on the predefined level of crime rate, specify a class label to each data point in the selected set. The data points which have the crime rate above the predefined rate are positive or members of hotspot class and data points with crime rate below the predefined rate are negative or non-members of hotspot class. Then this labelled data set will be used as the training set in SVM classification. To select a given percentage of the data to be labelled, we use the k-median clustering algorithm. Then, compare the result when the same percentage of the data is selected randomly.

The Technique shows that one class SVM gives good result when appropriate algorithm is choosen. Based on different experiment it is concluding that SVM form appropriate approach to hot spot crime prediction and k-means clustering algorithm for data selection.

2.2 MULTIVARIATE TIME SERIES CLUSTERING:

A multivariate time series is a flow of data points, computed typically at successive points in time spaced at uniform time intervals.

A Multivariate Time Series Clustering Approach for Crime Trends Prediction [9].In this technique, a approach for multivariate time series clustering technique based on dynamic time wrapping and parametric Minkowski model has been proposed to find similar crime trends. Since all types of crime do not have equal weightage, for example, murder will have more weightage over kidnapping and hurt. DTW together with Parametric Minkowski model is used to consider the weightage scheme in the clustering algorithm. In parametric Minkowski model, the distance function is defined by a weighted version of the Minkowski distance measure. The parameters for this model are the weights in different dimensions.

The effectiveness of technique has been illustrated in depth on Indian crime dataset provided by the Indian National Crime Records Bureau. The technique Multivariate Time Series Clustering based on Dynamic Time Wrapping and Minkowski model are efficient for finding similar crime trend and also predict them properly.

2.3 BAYESIAN NETWORK

A Bayesian network, belief network, Bayes network, or probabilistic directed acyclic graphical model is a probabilistic graphical model (a type of demographical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph. Bayesian networks are directed acyclic graphs whose random variables represent nodes in the Bayesian sense: they may be perceptible quantities, untapped variables, unknown parameters or supposition. Edges represent conditional dependencies, nodes that are not connected represent the variables that are conditionally independent of each other. Each node is linked with a probability function that takes as input a particular set of values for the nodes parent variable and relinquish the probability of the variable represented by the node [10].

A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory [11]. This technique introduces a novel serial crime prediction model using Bayesian learning theory. Author mainly studied the factor related to geographic report. For each factor, by using a discrete distance decay function which derives from the crime prediction theory Journey to Crime, create a geographic profile which is a probability distribution of being the next crime site on known geographical locations. The concluding prediction is made by combining all geographic profiles weighted by effect functions which can be adjusted adaptively based on Bayesian learning theory.

By testing this technique on a crime dataset of a serial crime happened in Gansu, China, can successfully capture the offender's intentions and locate the neighborhood of the next crime scene. The working step of this technique are as follows.

- (1) Download the map of Baiyin city from Google.
- (2) Make the populated parts of the map gridded into number of rectangles with the same size which will help a lot to calculate the distribution in (3).
- (3) Make the geographic portrait of each factor selected above by the discrete distance decay function.
- (4) With the first k_{th} ($k \geq 1$) crime sites known, we combine each factor's geographic profile by the dynamic prediction model.

- (5) Set the threshold of accept probability of final geographic profile, P_θ . And plot the area where the probability is greater than threshold.

- (6) Back to step (4) if k is not the maximum number of crime.

The result of this technique is impressive and prediction accuracy is very high. This approach efficiently uses geographical factors of crime and uses them for future prediction.

2.4 ARTIFICIAL NEURAL NETWORK

In 1943, McCulloch and Pitts, gives the first model of artificial neuron. According to Nigrin, A neural network is a circuit composed of a very large number of processing elements that are called Neuron. Each element works only on local information. Furthermore each element operates nonparallelly, thus there is no system clock.

Predicting the geo-temporal variations of crime and disorder [12]. This technique introduces for crime incident prediction by concentrate on geographical areas of concern that outshine traditional policing boundaries. The computerized procedure employ a geographical crime incidence-scanning algorithm to identify clusters with relative high levels of crime hot spots. This collection gives sufficient data for training artificial neural networks (ANNs) capable of modeling trends within them. The approach to ANN specification and assessment is enhanced by application of a novel approach, the Gamma test. The results of this technique are satisfactory using artificial neural network and gamma test provide the facility to predict future crime.

2.5 FUZZY TIME SERIES

The Fuzzy time series forecasting was originally originate by Song & Chissom. They proposed it in a series of papers to forecast student enrolments at the University of Alabama. The Fuzzy time series models developed by Song and Chissom were associated by high computational overheads due to complex matrix operations [13]. To reduce the computational overhead, Chen [14] simplified the process and proposed a simplified model that includes only simple arithmetic operations.

Applicability of Soft computing technique for crime Forecasting A Preliminary Investigation [15] In this technique, Author investigated applicability of Fuzzy Time series technique for crime forecasting. The scope of present technique is limited to probe the suitability of fuzzy time series method to predict the crime and to provide practical computational techniques. In this study, seventeen years actual historic crime data of Delhi city have been used and implemented the Fuzzy Time Series method on three different sets of universe of discourses obtained by partitioning it into five

intervals (Scheme-I), ten intervals (Scheme-II) and twenty intervals (Scheme-III). The computed forecast shows good accuracy. The result of this technique are satisfactory for predicting the future crime. The

approach used in this technique are also worked properly even if some data are not available.

In the next section the observation of the survey are presented.

III. COMPARATIVE ANALYSES:

S.no	Approach	Concept	Predictive Accuracy	Performance	Disadvantage
1	Support Vector Machine[8]	Find a linear hyper plane (decision boundary) that will separate the data.	Provide good accuracy almost all cases.	Less over fitting, robust to noise.	Computationally expensive, thus runs slow
2	Multivariate Time Series Clustering [9]	Based on Minkowski model	Prediction is good	Perform good even when dimensions has different significance	Dimension should have equal weight age.
3	Bayesian Network [11]	Based on bayes Theorem.	Depend on Geographical Factor So Predictive Accuracy is high.	Provide good performance by Combine prior knowledge with observed data.	Totally depend on selection of parameter.
4	Artificial Neural Network[12]	Information processing occurs at many simple Processing elements called neurons.	Prediction accuracy is generally high.	Fast evaluation of the learned target function.	Taking long training time.
5	Fuzzy Time Series[16]	Based on fuzzy logic (truth values between 0..1).	Provide better prediction if some data are not available.	Perform better in time space or state space condition.	Result are effected by various factors.

Table 1: Comparative Analysis of various Data mining Techniques

Table 1 shows the comparative analysis of various data mining technique introduced in this paper. Each technique have different concept for predicting the crime. Column 3 shows the concept of each technique, In the column 4 and 5 comparison is made based on predictive accuracy and performance and in the last column disadvantages of each technique are shown.

IV. CONCLUSION

The Data mining prediction techniques discuss in this paper are capable to enhance the accuracy, performance, speed of predicting the crime. These techniques are effectively identify common pattern by comparing current and past crime data and predict the future value. It's the basic step towards the Crime Prediction have been taken with demonstration and manipulation.

V. FUTURE SCOPE

In this paper, we introduce various data mining technique used in crime prediction these technique are capable for accurately predict the future crime but they also have some disadvantage. In the future work we will try to overcome disadvantages of these technique.

REFERENCES

- [1] Chung-Hsien-Yu, MaxW.Ward, Melissa Morabito, and Wei Ding. "Crime Forecasting Using Data Mining Techniques" 11th International Conference on Data Mining pp. 779-786, IEEE 2011.
- [2] Stephen Schneider, Ph.D. "Predicting Crime: A Review of The Research" Research and Statistics Division 2002.

- [3] J. L. LeBeau, "The Methods and Measure of Centrography and the spatial Dynamics of Rape" Journal of Quantitative Criminology, Vol.3, No.2, pp.125-141, 1987.
- [4] D.Weisburd, W.Bernasco, G.N.Bruinsma, "Determining How Journeys-to-Crime Vary: Measuring Inter- and Intra-Offender Crime Trip Distributions" Springer New York, 2009.
- [5] L. E. Cohen, M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach", American Sociological Review, Vol.44, pp 588-606, 1979.
- [6] D. Canter, P. Larkin, "The environmental range of serial rapists" Journal of Environmental Psychology, pp. 217-227 1993.
- [7] V. N. Vapnik, "Support Vector Network" Machine Learning pp. 273-297, 1995.
- [8] Keivan Kianmehr and Reda Alhajj, "Crime Hot-Spots Prediction Using Support Vector Machine" pp. 952- 960 IEEE 2006.
- [9] B. Chandra, Manish Gupta, M.P Gupta: "A Multivariate Time Series Clustering Approach for Crime Trends Prediction" pp. 892-896 IEEE 2008.
- [10] http://en.wikipedia.org/wiki/Bayesian_network#History.
- [11] Renjie Liao, Xueyao Wang, Lun Li and Zengchang Qinh, "A Novel Serial Crime Prediction Model Based on Bayesian Learning Theory" Ninth International Conference on Machine Learning and Cybernetics, Qingdao, pp. 1757-1762, IEEE 2010.
- [12] Jonathan J. Corcoran, Ian D. Wilson, J. Andrew Ware: "Predicting the geo-temporal variations of crime and disorder" International Institute of Forecasters Elsevier 2003.
- [13] Preetika Saxena et al. Int.J. "Forecasting Enrollments based on Fuzzy Time Series with Higher Forecast Accuracy Rate" Computer Technology & Applications, pp. 957-961, 2012.
- [14] S. M. Chen and Hsu, "A new method to forecasting enrollments using fuzzy time series", International Journal of Applied Science and Engineering, pp. 234-244, 2004.
- [15] Anand Kumar Shrivastav, Dr.Ekata "Applicability of Soft computing technique for Crime Forecasting A Preliminary Investigation " International Journal of Computer Science & Engineering Technology pp 415-421, 2012.