

# Paper on Different Approaches for Crime Prediction system

Varshitha D N

Assistant Professor

Department of Information Science and Engineering  
Vidyavardhaka college of Engineering, Mysuru

Aishwarya P

Department of Information Science and Engineering,  
Vidyavardhaka college of Engineering, Mysuru

K R Dhananjay Gupta

Department of Information Science and Engineering,  
Vidyavardhaka college of Engineering, Mysuru

Vidyashree K P

Assistant Professor

Department of Information Science and Engineering  
Vidyavardhaka college of Engineering, Mysuru

Janya T S

Department of Information Science and Engineering,  
Vidyavardhaka college of Engineering, Mysuru

Sahana R

Department of Information Science and Engineering,  
Vidyavardhaka college of Engineering, Mysuru

**Abstract—** Crime Prediction is a systematic approach for finding the crime patterns and trends. This paper gives different technologies that can be used for building Crime Prediction System. By building Crime Prediction System, it speeds up the process of solving crimes and reduces the rate of crime. We have different techniques which are dependent on the data that are previously reported and recorded and time and location. Crime Prediction system uses recorded data and analyses the data using several analysing techniques and later can predict the patterns and trends of crime using any of the below mentioned approaches.

## I. INTRODUCTION

Crime, in a way, influences organizations and institutions when occurred frequently in a society. Thus, it is necessary to study the factors and relations between different crimes and to find a way to accurately predict and avoid these crimes. Recently law enforcement agencies have been moving towards a more empirical, data driven approach to predictive policing. However, even with new data-driven approaches to predict crime, the fundamental job of crime analysts still remains difficult and often manual; specific patterns of crime are not very easy to find by way of automated tools, whereas larger-scale density-based trends comprised mainly of background crime levels are much easier for data-driven approaches and software to estimate.

Here we will take an interdisciplinary approach between computer science and criminal justice to develop a data mining paradigm that can help solve crimes faster. More specifically, Crime is naturally 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 revention solutions that equivalent to specific places and times. Crime analysis takes past crime data to predict future crime locations and time. Crime prediction for future crime is a process that finds out crime rate change

from one year to the next and projects 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 involves relating past crime trends with factors that will influence the future scope of crime.

This paper constitute of the following techniques for Crime Prediction:

- a) Data mining technique
- b) Crime cast technique
- c) Deep learning technique
- d) Sentimental analysis technique

## II. DATA MINING TECHNIQUES

Data Mining and Data Analyzing techniques are used to find the patterns and trends in crime from the crime data that is previously stored from various sources, this in turn decreases the rate of crime.

There are several techniques in Data Mining for Crime Prediction:

### A. Association Rule Mining

The hidden or sensitive knowledge in unlabeled data is found out by this technique. This technique can also discover the co-occurrences of objects in large datasets.

The two modules are:

- a) Antecedent is called left hand side.
- b) Consequent is called right hand side.

The general form is LHS->RHS. The occurrences of data set is likely to occur in RHS depending upon its occurrence in LHS.

Association Rule Mining has certain limitations: Finding large datasets require more time. To overcome this Apriori algorithm is used, which completes the process faster.

Apriori algorithm can be used for Association Rule Mining in which all item sets are to be scanned. This algorithm is used to prune candidates explored. Even this takes a long time. So, improved Apriori algorithm is developed using compressed database algorithm. This fastens the process.

### B. Classification Rule

The depiction and distinguishment between data classes or notions is classification technique. The input data is grouped into classes. Every class has attribute set and class label. The attribute set consists of a set of input data and class label is the name assigned to it.

The data is of two types training set and testing set. 1) Training set divided randomly which is used to create a classifier which predicts the class of the unknown record. 2) Testing set is used to determine the accuracy of the system.

The different approaches to this method are:

- 1) Decision Tree: This method consists of root, internal nodes and external nodes. Root is the main attribute of the database which is a separate set. Every internal node is different from one another. The leaf node gives the class label which is the result.
- 2) Nearest Neighbour: This method is used to find the similarity between the test set and the train set. If a train set is close to a test set, then the class label of the train set will be assigned to the test set. The limitation of the nearest neighbor method arises when the number of training set is less. To improve this several techniques like k-nearest neighbour algorithm are used. The k-nearest neighbour uses majority votes to decide the class label.
- 3) Neural Network: This method is used to produce the algorithm which has the ability to learn and recognize patterns to obtain knowledge. It consists of input and output and a neuron that shows a connection between input and output. Neurons have a specific weight. When input arrives it is found out by multiplying it with the neuron weight. If the sum of all the neurons is greater than the threshold, then it is considered as the output.

### C. Clustering

It is the grouping of a set of data in such a way that data in the same group(cluster) are very similar to one another than the data that are in the other clusters.

Several techniques in clustering are:

1) K-means algorithm: It partitions the data into k number of clusters in which each data observed is assigned to the nearest centroid. The user gives the specified k number of

centroids. Each cluster must have a centroid. This process will be repeated until all the data is assigned to a cluster.

2) Hierarchical clustering: This method segments the like data into the like group. This is done using several similarity and dissimilarity measures. Each cluster or node consists of child nodes is viewed as a tree.

3) Expectation Maximization: This method is a recursive method in which statistics are used to segment the incomplete data into clusters.

## III. CRIME CAST

Crime Cast is a mathematical simulation technique that finds the crime rate, locations, type of crime, hazards of crime and predicts for future using the previous database.

In this technique, a region with higher crime rate is to be found. These regions are called as Hotspots. The region with the lower crime rate are called as Coldspots. So, Crime cast can be introduced in the hotspots which is the technique of prediction using crime databases by simulating probabilistic model implementation and Artificial Neural Network.

Hotspot region changes frequently. So crime prediction agencies can concentrate more on the Hotspots than on the coldspots.

One of the methods is point pattern density model, by analyzing the previous crime data based on the occurrence of crime in a precise location.

Following methods can be used for the Crime Cast techniques:

### A. Multivariate Time Series Clustering

The flow of data points computed in similar time intervals is used in Multivariate Time Series. This approach is efficient for determining similar crime trends. The technique consists of weightage scheme. This is done because some crimes have more importance than others i.e., more weightage. The distance is measured using Minkowski distance measurement.

### B. Support Vector Machine

This technique consists of a given dataset and predefined level of crimes. K-clustering algorithm is used to select a subset of the crime dataset. A label is put to each data point in the selected set. A crime rate is given and the points whose crime rate is above the given rate are hotspots and the points whose crime rate is below than the given rate are coldspots.

### C. Bayesian Network

A Bayesian Network uses a DAG to represent a set of random variables and their dependencies. It is a probabilistic model. The nodes represent variable, edges represent the conditional dependencies in a Bayesian network. Independent variables are the nodes that are not connected.

### D. Fuzzy Time Series

The initial process had complex matrix operations which caused heavy overheads. Later, a simplified process was processed. This technique can work even when some data is not available.

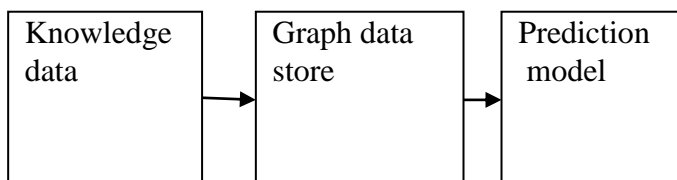
### E. Artificial Neural Network

A neural network has a large number of processing elements which works only on local data called neurons. A crime prediction model was introduced to focus on areas with high crime occurrences. The hotspots are identified using a crime incidence-scanning algorithm.

## IV. DEEP LEARNING TECHNIQUE

Deep Learning is the composition of multiple layers which includes non linear operations such as neural net, propositional formulae, etc. Deep Learning technique of artificial intelligence here is used to build the relationship between crimes and build the model of world with various crimes. Deep Learning uses several algorithm to convert the raw information into higher representations.

The architectural diagram is given below:



With the help of dataset and information visual representation of hotspots is developed. The system can learn about patterns of events and this knowledge can be used for different crime with different time and space.

The Graph Loader loads the graph data store for analysing of events which in turn predicts the time and space of the crime events.

### A. Algorithm

The algorithm has three stages:

- 1) Pre-processing Stage
- 2) Processing Stage
- 3) Post processing Stage

#### 1) Pre-processing Stage

In Pre-processing Stage the data from various resources are collected transformed into clean information. The transformed data will then be stored on graph data store. Then the system will automatically take parsing to understand the type of information, its relationship etc. to provide all possible combination of events.

#### 2) Processing Stage

In Processing Stage, the generated event combinations will be analysed to produce possible configuration. The system decides the most suitable combination with the help of previous data. Therefore producing the set of locations with the set of possible events.

#### 3) Post processing Stage

In the Post processing Stage the set of events are filtered into interesting and important events. It can be found out by using several output stage threshold based filters. Since the contributing factors for the events is different, the probability of those events may also differ.

### B. Limitations

Deep Learning produces inaccurate results when the small dataset is provided. It also gives different probability results because the contributing factors are different in post processing stage.

## V. SENTIMENTAL ANALYSIS

The tweet analysis used to detect crime consists of the following steps:

1. Collecting tweets
2. Cleaning and parsing of data
3. Conducting sentiment analysis on the collected tweets.

The twitter data is collected by using Twitter4J which is an unofficial Java library for the Twitter Application Program Interface. This was used to automate the application so that it could be integrated into Twitter. Tweets are then collected according to the crime related topics within a particular zone of every city. A keyword search strategy was adopted for collection purposes. These keywords are used to recognize the crime-related tweets which include the words such as "gun", "crime", "murder", "kill" and so on similar words. Each and every tweet is then parsed before sentiment analysis is conducted. This analysis is conducted for the very reason to segregate the individual terms in a tweet according to the white-space boundaries and to then convert the tweet into lower case letters and to remove all non-alphanumeric characters from tweets. Then sentiment analysis of tweet is estimated to predict the crime.

The various steps hold this information:

### A. Data collection

One has to create the twitter application through which he can get access tokens, consumer key and secret key in order to collect tweets. This provides the authorization to twitter API calls. Well within here twitter 4j API is used to collect the tweets. Collecting is based on the keyword entered. The data is collected by using the public twitter API called streaming API, it collects the latest tweets based on keyword given. The data that is collected is then pre-processed before sentiment analysis is performed.

### B. Data pre-processing

Once the data is collected, it is pre-processed which is very much necessary in order to remove the whitespaces, punctuations and URLs, to convert all uppercase to lowercase. This is done in order to improve the efficiency. the normalization technique is used to remove all unrelated information from collected twitter data.

### C. Sentiment Analysis

The sentiment analysis is performed on each individual tweet using the lexicon based approach. The vocabulary includes both positive and negative words. Here the individual tweet is split into different words. Each word is then compared with the words in the vocabulary to identify the polarity of tweet. The positivity and negativity of the word is easily identified by this analysis. If the word is positive then its polarity value is 1, if it is negative then its polarity value is -1.

## VI. CONCLUSION

Crime prediction is current trend in the society. It aims at reducing the crime occurrences by predicting the possible crimes that might occur in the future days. Crime prediction and analysis can be performed by various approaches; some of which are discussed in this paper: Data Mining, Crime Cast, Deep Learning, Sentimental Analysis. Each of the above approaches have their own pros and cons. Any one of these proves to be superior in any particular instance.

## REFERENCES

- [1] Shiju Sathyadevan, Devan M, Surya Gangadharan. S, Crime Analysis and Prediction Using Data Mining, 2014 First International Conference on Networks & Soft Computing.
- [2] Abba Babakura, Md Nasir Sulaiman and Mahmud A. Yusuf, Improved Method of Classification Algorithms for Crime Prediction, 2014 International Symposium on Biometrics and Security Technologies (ISBAST).
- [3] Jazeem Azeem, D. John Aravindhar, Hybrid Approach to Crime Prediction using Deep learning, 2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI).
- [4] Anahita Ghazvini, Siti Norul Huda Sheikh Abdullah, Biography Commercial Serial Crime Analysis Using Enhanced Dynamic Neural Network, 2015 Seventh International Conference of Soft Computing and Pattern Recognition (SoCPaR 2015).
- [5] Nafiz Mahmud, Khalid Ibn Zinnah, Yeasin Ar Rahman, Nasim Ahmed, CRIMECAST: A Crime Prediction and Strategy Direction Service, 19th International Conference on Computer and Information Technology, December 18-20, 2016, North South University, Dhaka, Bangladesh.
- [6] Sadhana C S, Sanghareddy B K, Predicting Crime Using Twitter Sentiment, Second National Conference on Emerging Trends In Computer Science And Engineering (NCETCSE-2016), Department Of CSE, BGSIT.
- [7] Nikhil Dubey et al "A Survey Paper on Crime Prediction Technique Using Data Mining", Int. Journal of Engineering Research and Applications, 2014 .
- [8] B. Chandra, Manish Gupta, M.P Gupta: "A Multivariate Time Series Clustering Approach for Crime Trends Prediction" pp. 892-896 IEEE 2008.
- [9] Malathi. A and Dr. S. Santhosh Baboo. Article: an enhanced algorithm to predict a future crime using data mining. International Journal of Computer Applications, 21(1):1-6, May 2011. Published by Foundation of Computer Science.
- [10] Li, G., and Wang, Y.: A privacy-preserving classification method based on singular value decomposition, Arab Journal of Information Technology, vol 9(6), 529-534. (2012).