Security Analysis in Different Areas using Crime Data

Siddharth Yadav

Student Department of Computer and Science Engineering IMS Engineering College, Ghaziabad, U.P., INDIA

> Shivam Choudhary Student

Department of Computer and Science Engineering IMS Engineering College, Ghaziabad, U.P., INDIA **Umang Kushwaha**

Student Department of Computer and Science Engineering IMS Engineering College, Ghaziabad, U.P.,INDIA

Mrs Sapna Yadav

Professor Department of Computer and Science Engineering IMS Engineering College, Ghaziabad, U.P.,INDIA

Abstract

This paper talks about hotspots and use of k means clustering for crime pattern detection. It first identifies significant attributes in the database. Unlike other papers, it then gives weights to attributes in the data set. The most important attribute (e.g. type of crime) is given the highest priority (weight) as compared to other attributes. We are using this feature of the research paper in our project. The data with missing values are made as test cases. It suggests dividing the database according to respective states, using classification, to make the data easier to analyze. In our project, we are subdividing the data into different types of crime allowing the user to get information of those crimes easily (e.g. percentage of the particular crime in a particular year, the hotspot of that particular crime).

Keywords- Security Analysis

I. INTRODUCTION

Objective is Development of a spatiotemporal crime prediction model based on geographical information systems coupled with spatial statistical methods. In this paper clustering analyses are used to identify hot spots. Cluster analysis aims to collect data into groups according to several algorithms which are Kmeans, Nnh hierarchical, spatiotemporal analysis of crime (STAC), fuzzy, ISODATA, and geographical analysis machine (GAM) clustering techniques. Clusters of STAC do include more homogenous areas than the other methods. STAC is not restricted to include all the observations hence STAC is able to indicate denser crime areas than other methods. This is important in crime prevention for allocating resources effectively. Who is the target of the crime? Majority of the time the victims are recognized and in many cases is the person reporting the crime. Additionally, the crime may have some witnesses. There are other words commonly used such as homicides that Refer to manslaughter or killing someone. Under homicides there can be categories like infant murders, eldercare and killing law officers. For the purposes of our modeling, we will not need to get into the depths of criminal justice but will confine ourselves to the main kinds of crimes.

The Cluster has a special meaning which refers to a special group of crime i.e. a lot of crimes in a given specified regions. These clusters can be represented using a geo plot of the crime described on the map of the police jurisdiction. And densely populated group of crime is used to visually locate the 'hot-spots' of crime, also when we discuss clustering from a data-mining point, we refer to similar kinds of crime in the given crime rate of interest. Such clusters are useful in identifying a crime pattern or a crime spree. Some well-known examples of crime patterns are in the movie Dhoom-2.

II. CRIME REPORTING SYSTEMS

A. Crime Data Mining for Indian Police Information System

This paper is all about India's crime analysis system. It gives ways to enhance the currently existing system in Indian Police System called Crime Criminal Information System (CCIS). It suggests dividing the database according to respective states, using classification, to make the data easier to analyze. In our project, we are subdividing the data into different types of crime allowing the user to get information of those crimes easily (e.g. percentage of the particular crime in a particular year, the hotspot of that particular crime)

B. An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool

The system finds the critical path of the serial killers who are striking over again and again and determines links between their crime activities locations occurred in the past, travel record, and background history, etc. These findings increase the chance of finding these repeat offenders. The formation to integrate information from various crime incidents and also multiple sources and

also discover regular patterns for the structure, organization, operation and information in criminal databases. If a particular criminal uses a pattern of path to commit consecutive crimes, then the next crime location of this serial killer can be predicted from the pattern observed. e.g.: in DHOOM 2, the last location of crime of criminal HRITHIK ROSHAN was predicted by the pattern he formed from his previous crime locations.

C. Evolving Data Mining Algorithms on the Prevailing Crime Trend – An Intelligent Crime Prediction Model A. Malathi and Dr. S. Santosh Baboo

The crime data is divided into days of the week, to observe Spatio temporal distribution of crime. To the clustered results, a classification algorithm was applied to predict the future crime pattern. It enables us to build a model on predicting next records using previous year's data.

III. ALGORITHM TO BE USED

- Time Series analysis of data and forecasting using 5-year moving average
- Clustering for Crime pattern Detection
- Clustering and NN for Criminal Pattern Detection
- Association for Criminal Pattern Detection i.e. to find out the serial killer link etc.
- K-Nearest Neighbor Algorithm for pre-processing of data.
- Time Series Analysis of Crime Data using 5 year Moving Average Algorithm

IV. CRIME ANALYSIS

DBScan Algorithm to do clustering of hot spots i.e. density based clustering based on X and Y coordinates. DBScan Algorithm to create clusters inside the clustering performed on above clusters based on type of crime PAM: To find out the most dangerous crime over the created cluster based on costs Apriori Association Algorithm to detect the pattern of crime over the area and its prediction of risk.

A. Criminal Analysis Process

- Neural Network Algorithm to predict the possible criminals based on inputs to the network
- Associative Criminal Trend Analysis using APRIORI algorithm to detect the pattern of criminals age wise, location wise, street wise etc.

B. Process

1) Pre-Processing

1) Cleaning of the data sets files (csv files) by removing the records having missing attributes or faulty attributes through the use of JAVA code

2) Part A: Hot Spot Detection using X and Y Coordinates

Use of DBScan Algorithm to find out the hot spots i.e. the places with the maximum density of crime activities (any crime) using the X and Y coordinates of the crime activities noted.

- Algorithm used: DBScan
- Threshold used : To make cores : 0.0926, To combine cores: 0.0927
- Sparcity problem: Yes, pruned out sparsed data.
- The best place to look up to find room for improvement is the voluminous raw data that is generated on a regular basis from various sources by applying Big Data Analytics (BDA). BDA refers to the tools and practices that can be used for transforming the raw data into meaningful and crucial information which helps in forming a decision support system for the judiciary and legislature to take steps towards keeping security in check.

V. RESULTS OF CRIME PATTERN ANALYSIS

We show the results in the figure below. The different clusters or the crime patterns are color-coded. For each group, the legend provides the total number of crimes incidents included in the group along with the significant attributes that characterize the group.



Fig. 1:



Fig. 2:



Fig. 3:



Fig. 4:

VI. CONCLUSIONS AND FUTURE DIRECTION

- We successfully determine the hot spots / high density crime areas as shown in the plots.
- Association gives us the percentage risk of a particular crime occurring at a particular area at a particular location at a
 particular time.
- The neural network used is trained over the data sets using supervised learning, then the weights fixed are used to predict the 4th unknown parameter given three known parameter.
- The association used gives us a visualization of the age group of the offenders seen mostly at what time what location and what type of crime.

REFERENCES

- Crime Pattern Detection Using Data Mining: Hsinchun Chen, Wingyan hung, Yi Qin, Michael Chau, Jennifer, Rong Zheng, Homa Atabakhsh, "Crime Data Mining: A General Framework and Some Examples", IEEE Computer Society April 2004.
- [2] An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool: Agrawal R., Gupta A., Sarawagi.
- [3] An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis-A Data Mining Tool : Fayyad, U. Data Mining and
- [4] Evolving Data Mining Algorithms on the Prevailing Crime Trend An Intelligent Crime Prediction Model A. Malathi and Dr. S. Santhosh Baboo : Brown, D.E. (1998) The regional crime analysis program: A frame work for mining data to catch criminals," in Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics.