



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

A Survey on Data Analytics, Predictions and Visualization of Crime Patterns in India

Shivam S. Bhirud¹, Shravani S. Swami², Shekhar S. Patil³, Jyoti Sharma⁴,

Shital V. Dabhade⁵

U.G. Student, Department of Computer Engineering, SKNCOE, Pune, Maharashtra, India^{1,2,3,4}

Associate Professor, Department of Computer Engineering, SKNCOE, Pune, Maharashtra, India⁵

ABSTRACT: In India, there has been a drastic pattern of crime observed in the past few years giving rise to a threat to the security of common man. Considering this perilous situation, we aim to study these crime patterns and realise the changes in the overall crime ratio based on the demographic/population ratios which shall ease the process of reaching the future crime-cases prediction. Along with this, a prediction can be made on the maximum criminal activities that need to be controlled and the amount of security measures that need to be taken to decelerate/curb the occurrence of any mild to treacherous level criminal activities.

This paper depicts the survey performed on the prior work done in crime pattern analysis. It includes the study of numerous algorithms, techniques and tools beneficial for visualizations and predictions based on the data acquired from the Indian Government.

KEYWORDS: Data visualization, Prediction, Crime pattern analysis.

I. INTRODUCTION

Every country's strength lies in the power of its common man. So, along with the nation's territorial security, it is crucial to recognize the safety of every individual. India, being the second largest populated country in the world becomes an alarming hub of crime and other illegal acts risking the life of many citizens. Therefore, we observe the urge to meticulously handle it. Here, crime pattern analysis comes into picture. Crime pattern analysis refers to mainly categorizing the obtained crime data and mapping of various patterns to reach significant conclusions.

To analyse the crime patterns our main steps involve the survey of:

- Existing computer aided crime analysis tools.
- Crime data acquisition tools/techniques.
- Visualization tools and techniques.
- Prediction algorithms, techniques and tools.

From the tools studied, we came across some effective tools which collect data and also visualize various patterns in it based on the vulnerability of crime in certain areas of the country. The crime data can be collected from the Indian Government's National Crime Records Bureau repository using several tools whose insights are provided in the further sections of this paper. Different visualization tools with different levels of flexibility are also discussed. Along with these, a study on diverse prediction methods such as k-means, association rule mining, decision tree and many more shall be looked upon.

II. LITERATURE REVIEW

Acknowledging the fact that there is a tremendous increase in the rate of occurrence of crime, we studied various papers which enlighten us on different approaches towards crime analysis and predictions.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

In [1], a different methodology for fetching raw data (criminal records) that needs pre-processing is suggested. Data classification is done by grouping of records based on states/cities, criminal profiles, etc. The explanation of the k-means algorithm for clustering the categories of crime that possess similar characteristics is provided. It is also stated that crime correlation can be performed using Pearson's correlation coefficient and crime predictions can be done with the help of linear regression.

In [2], a detailed study, analysis and distribution of numerous papers/journals published on crime data mining is performed. This paper also provides the classification of the research papers based on the techniques used; technologies used; challenges and issues addressed. An elaboration on the sources of crime data and the types of crimes is also included in this paper. Thus, it gives us brief information of the study done in crime data mining till date.

In [3], pre-processing techniques mainly focusing on handling the missing values for state and the intensity of crime occurred by using traditional KNN with a new distance metric and LVQ (Learning Vector Quantization) is explained. Clustering using k-means and DBScan (Density-Based Spatial Clustering Application with Noise) algorithms are suggested. Prediction of crime trend can be done using C4.5 decision tree algorithm.

In [4], emphasis is laid on the pre-existing tools for crime analysis in India which include: Crime and Criminal Information System (CCIS) which assists in investigations and the authorized officials to create strategies for crime control; Common Integrated Police Application (CIPA) to automate the workflow at police stations, to form CCIS and to enhance the accessibility, transparency and accountability in the working of police dept. The proposed system is an Intelligent Crime Information and Analysis System (ICIAS) which involves a Crime Data Entry Module and a Crime Information Retrieval and Analysis Module to identify crime hotspots and crime zones.

In [5], an elaborated study of the existing computer aided crime analysis software and tools available in the market are illustrated. The software such as ATAC, CrimeStat, RCAGIS (Regional Crime Analysis Geographic Information System), CrimeConnect, BRAINCEL etc. are available that allow data entry, manipulation, and analysis of the crime patterns using data mining, neural nets and deep learning techniques. Recently, Government of India's model named 'Government to Government (G2G)' model, also called as the Crime & Criminal Information System (CCIS) was designed to create computerized storage, analysis of crime and retrieval of criminal records but till date, in spite of collecting humongous data it fails to perform the crime analysis thus it is somewhat a 'standalone' system.

In [6], the GDP, literacy-rate, police-rate, employment-rate and various crimes such as murder, dacoity and riots and the state in India as location data has been considered to check spatial autocorrelation between various crimes and to compare various attribute clusters and its relation. Heat maps are plotted for various clusters of murder-rate, riot-rate, dacoit-rate, overall crime-rate, police-rate, population density, GDP, etc.

In [7], crime analysis is done efficiently and effectively to identify crime patterns in order to detect and to take appropriate precautionary actions for future crime patterns. Crime information has to be stored and analysed so as to get an accurate analysis of this growing volumes of crime data. Results of data mining could potentially be used to lessen and even prevent crime for the upcoming years. Different structures are used for recording crime data. Clustering techniques like K-means and DBscan algorithms are proposed to get the crime ratio relation between the states or districts.

In [8], ideas that support police detectives to derive solutions for crimes by data mining techniques have been suggested. A study on various classification and clustering techniques like k-means and KNN, with the help of which an attempt to classify categories of crimes at various places has been covered. Finally, reports were generated in the form of geo-spatial plots which clearly depicts the categories of crime in a particular region. As stated, the limitations of this study is, data mining can only help the detectives but cannot replace them.

In [9], a model for crime and criminal data analysis is proposed using simple k-Means algorithm for clustering the data and Apriori Algorithm for data-association. This paper clearly depicts the knowledge extraction from the datasets based on discovering patterns and trends, making forecasts, relationship analysis with most probable explanations, mapping criminal networks and identifying possible suspects. Clustering is done to discover the relation between various types of crimes. Association rules mining is based on generated rules from crime dataset based on frequent occurrence of patterns in crime for taking the preventive actions. The data was collected manually from a police



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 9, September 2017

department in Libya. Analysis of this data was done with the help of WEKA mining software tool and MS-Excel was used for generating the effective visuals.

III. TOOLS AVAILABLE FOR CRIME ANALYSIS

The tools we encountered from our study include analysis tools which take the help of visualizations and data collection tools.

Sr. No.	Software/Tool	Purpose/Function
1.	ATAC	Data entry, manipulation & analysis
2.	CrimeStat	Spatial statistical software uses GIS for analysis using incident location
3.	BRAINCEL	Excel add-in, enhances forecast with neural networks
4.	CrimeConnect	Web-based real time information sharing system
5.	CCIS	Computerised storage of criminal records
6.	CIPA	Automated functions of police station -registration, investigation & prosecution.

Table 3.1: Pre-existing crime analysis tools

The striking fact that came into notice was that, there is no such tool available that beholds both these characteristics together, which if incorporated shall give constructive results with an effective analysis.

IV. ALGORITHMS FOR KNOWLEDGE DISCOVERY IN CRIME DATASETS

Knowledge Discovery in Database (KDD) is traditionally done for extracting fruitful information from the available data. Crime related data can be analysed using:

1. Clustering and classification
2. Association

4.1 Clustering and Classification

The partition clustering methods can be categorized into K-means, AK-mode, Expectation-Maximization, DBScan algorithms etc. The partitioning method constructs 'k' partitions of the data from a given dataset of 'n' objects. After pre-processing, the operational data undergoes various clustering techniques for grouping objects as different clusters.

4.1.1 K-means Algorithm

[10] K-means algorithm is used for the partitioning of the data into clusters based on the characteristics of the data. Data having similar characteristics are clubbed and a cluster is formed.

Steps for K-means algorithm:

Number of clusters is given as input.

Step 1: Randomly choose N objects from the dataset. These N objects are initialised as the cluster center.

Step 2: Each object is reallocated to form a cluster based on the cluster center, to which it is the most nearer or similar.

Step 3: Mean value for the object with respect to each cluster is calculated based on Euclidian distance.

Output is the sets of N clusters.

K-means algorithm is a base for all other clustering algorithms to find the mean values.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirccce.com

Vol. 5, Issue 9, September 2017

4.1.2 AK-mode Algorithm

Two important phases involved in AK-mode clustering are the weighing phase and clustering phase. In the weighing phase, the weight of every attribute is calculated using Information Gain Ratio (IGR) value and the greatest value is considered as the decisive factor. The similarity in differences between two data records is considered for calculating the distance between the categorical attributes. The threshold value α is set with the help of the computation result of similarity measures.

[10] Steps for AK-mode algorithm:

Inputs are the dataset, weighted attributes and threshold value.

Output: Cluster result

Step 1: Find the number of clusters and the initial mode of every cluster.

Step 2: Calculate the distance between every mode and its closest mode.

Step 3: Each cluster mode is updated.

Step 4: This process terminates when none of the mode changes, otherwise step 2 is repeated.

4.1.3 Expectation-Maximization Algorithm

Expectation-Maximization Algorithm is an iterative algorithm which is used to find the maximum likelihood estimations of the parameters in every cluster. [10] This algorithm is quite similar to the k-means algorithm. The only difference is that, the EM algorithm considers the weight of each object for clustering instead of the object itself. The weight of the objects is measured in terms of probability distribution. Thus, creates the view of the entire data being a combination of parametric probabilistic distribution. Finding the parameter estimates is done using two steps:

Expectation step: Calculate the probability of the object's belongingness to the cluster.

Maximization step: With the help of probability estimation, re-estimating the parameters, evaluated in the expectation step.

The EM algorithm is fast and easy to implement.

4.2 Association

Crime prevention is another important aspect of the data mining for crime analysis. [9] Police officers can decide upon some of the crime prevention strategies if the detailed study regarding crime can be presented in human language or the language which can be easily understood by humans, however, this is where association rule mining comes into picture. Association rule mining helps in generating the patterns based on the frequent occurrences of crime in a particular area. Thus, it might help the decision makers of our society to take proper preventive measures for preclusion of crime.

[14] Apriori algorithm falls under the category of unsupervised machine learning algorithm and it is typically used for association rule mining.

Steps involved in apriori algorithm are:

Step 1: Apply minimum support to find all the frequent sets with k items in a dataset.

Step 2: Find the frequent set with k+1 item using self-join rule.

Step 3: Check the support for obtaining the frequent items i.e. for itemset x,
 $\text{supp}(x) = (\text{Number of transactions consisting } x) / (\text{Total transactions})$

Step 4: If the set of frequent items is Φ then stop, otherwise repeat step 2.

V. CONCLUSION AND FUTURE WORK

Understanding the severity of attention required towards crime and its trends, measures to decrease the causes and consequences need to be planned which can be made easier by a productive centralised hub of crime data collection, processing, visualization and predictive analysis. This enables a wide variety of professionals to understand the crime patterns in different ways flexibly. As we are aware of the fact that prediction of crime is not a new concept, there are a number of aspects of these predictions that can be brought to the cognizance of the stakeholders of this research. Along with the future years' crime rate prediction, the probability of occurrence of crime in a particular area and time can be evaluated. The study of different age groups of individuals victimized as well as those found guilty under legal offenses can be done. Another aspect that can be considered is the area wise police efficiency which may affect the crime rates



ISSN(Online): 2320-9801
ISSN(Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 9, September 2017

accordingly. These studies shall help in apprehending the different crime zones and the types of crimes they are endangered by, in order to inhibit the necessary controls to make the day to day surroundings secure for all the citizens.

VI. ACKNOWLEDGEMENT

We convey our sincere thanks to Dr. P N. Mahalle, Head of Computer Dept., SKNCOE, Pune, for his valuable support.

REFERENCES

- [1] Tushar Sonawane¹, Shirin Shaikh², Shaista Shaikh³, Rahul Shinde⁴, Asif Sayyad⁵, Crime Pattern Analysis, Visualization and Prediction using Data Mining, IJARIE International Journal of Advanced Research and Innovative Ideas in Education, Vol-1 Issue-4 2015, IJARIE-ISSN(O)-2395-4396 Methodology for crime pattern analysis.
- [2] Sarpreet Kaur¹ and Dr. Williamjeet Singh², Systematic Review of Crime Data Mining, IJARCS International Journal of Advanced Research in Computer Science, Volume 8, No. 5, May-June 2017, ISSN No. 0976-5697.
- [3] Malathi. A 1, Dr. S. Santhosh Baboo 2 and Anbarasi. A 3; 1 Assistant professor Department of Computer Science, Govt Arts College, Coimbatore, India., 2 Readers Department of Computer Science, D. G. Vaishnav College, Chennai, India. 3 Associative Professor Department of Civil Engineering, Kumaraguru college of Technology, Coimbatore, India; Enhanced Algorithm to Predict a future crime using Data Mining, International Journal of Computer Applications (0975 – 8887), Volume 21– No.1, May 2011.
- [4] Neeru Mago, Design and Implementation of Intelligent Crime Information and Analysis System (ICIAS) based on Crime Data Mining, IJARCSSE International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 1, January 2016 ISSN: 2277 128X.
- [5] D.Gnana Rajesh¹ and Dr. M. Punithavalli², A Study on the Existing Computer Aided Crime Analysis Tools to Handle Indian State of Affairs, IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012 ISSN (Online): 1694-0814.
- [6] Ahamed Shafeeq B M¹, Dr. Binu V S², Spatial Patterns of Crimes in India using Data Mining Techniques, Certified International Journal of Engineering and Innovative Technology (IJEIT) Volume 3, Issue 11, ISSN: 2277-3754 ISO 9001:2008, May 2014.
- [7] S. Yamuna, N.Sudha Bhuvaneshwari , 1 M.Phil (CS) Research Scholar School of IT Science, Dr.G.R.Damodaran College of science Coimbatore, 2 Associate Professor MCA, Mphil (cs), (PhD) School of IT Science Dr.G.R.Damodaran College of science Coimbatore, Data mining Techniques to Analyze and Predict Crimes, The International Journal of Engineering And Science (IJES), Volume 1, Issue 2, Pages 243-247 2012, ISSN: 2319 – 1813, ISBN: 2319 – 1805.
- [8] P.DHAKSHINAMOORTHY, T.KALAISELVAN, Assistant Professor, Erode Builder Educational Trust's Group of Institutions, Erode, TamilNadu, CRIME PATTERN DETECTION USING DATA MINING, International Journal of Advanced Research in Computer Science and Applications Vol.1, Issue 1, July 2013, ISSN 2321 -872X.
- [9] Dr. Zakaria Suliman Zubi¹, Ayman Altaher Mahmud², 1Sirte University, Faculty Of Science, Computer Science Department Sirte, P.O Box 727, Libya, 2The Libyan Academy, Information Technology Department, Tripoli, Libya, Using Data Mining Techniques to Analyze Crime patterns in the Libyan National Crime Data, ISBN: 978-960-474-350-6.
- [10] Dr. A. Bharathi¹, R. Shilpa², 1 Professor Dept. of IT Bannari Amman Institute of Technology Sathyamangalam – India, 2 Research Scholar Dept. of IT Bannari Amman Institute of Technology Sathyamangalam – India, A Survey on Crime Data Analysis of Data Mining Using Clustering Techniques, International Journal of Advance Research in Computer Science and Management Studies, Volume 2, Issue 8, August 2014, ISSN: 232 7782.
- [11] Vineet Pande, Viraj Samant, Sindhu Nair, Crime Detection using Data Mining, International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181, Vol. 5 Issue 01, January-2016.
- [12] Lakshmi Iyer¹, Anandi Mani ², Prachi Mishra³, Petia Topalova⁴, 1 Harvard Business School 2 University of Warwick, 3 Research Department, IMF, 4 Research Department, IMF, Political Representation and Crime: Evidence from India November 2010(Work in progress).
- [13] Shiju Sathyadevan¹, Devan M. S², Surya Gangadharan. S³, 1 2 Amrita Center for Cyber Security Amrita Vishwa Vidyapeetham, Amritapuri, Kerala, Amrita Vishwa Vidyapeetham Amritapuri, Kerala, India, 2014 First International Conference on Networks & Soft Computing, 978-1-4799-3486-7 114 2014 IEEE.
- [14] <http://blog.hackerearth.com/beginners-tutorial-ap-rriori-algorithm-data-mining-r-implementation>.