

Crime Pattern Analysis, Visualization And Prediction Using Data Mining

¹Tushar Sonawanev, ²Shirin Shaikh, ³Shaista Shaikh, ⁴Rahul Shinde, ⁵Asif Sayyad

¹ [BE], Computer science, S.R.E.S COE, Maharashtra, India

² [BE], Computer science, S.R.E.S COE, Maharashtra, India

³ [BE], Computer science, S.R.E.S COE, Maharashtra, India

⁴ [BE], Computer science, S.R.E.S COE, Maharashtra, India

⁵ [BE], Computer science, S.R.E.S COE, Maharashtra, India

ABSTRACT

Crime against women these days has become problem of every nation around the globe many countries are trying to curb this problem. Preventive are taken to reduce the increasing number of cases of crime against women. A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analysing and predicting crime and help us prevent the crime to some extent. Crime analysis is an area of vital importance in police department. Study of crime data can help us analyse crime pattern, inter-related clues & important hidden relations between the crimes. That is why data mining can be great aid to analyse, visualize and predict crime using crime data set. Classification and correlation of data set makes it easy to understand similarities & dissimilarities amongst the data objects. We group data objects using clustering technique. Dataset is classified on the basis of some predefined condition. Here grouping is done according to various types of crimes against women taking place in different states and cities of India. Crime mapping will help the administration to plan strategies for prevention of crime, further using data mining technique data can be predicted and visualized in various form in order to provide better understanding of crime patterns.

Keyword: - K-Means , cluster, correlation .

1. INTRODUCTION

India is a vast country with diversified societies. Position of women has been of great importance since ancient times in Indian culture. Unfortunately current scenario depicts a different story. According to National Crime Records Bureau, crime against women has significantly increased in recent years. It has become the most prior to the administration to enforce law & order to reduce this increasing rate of the crime against women. This is where criminology comes into picture . Criminology is scientific study of crime and criminal behaviour in order to detect crime characteristics. Use of data mining techniques can produce important results from crime dataset. The very step in study of crime is crime analysis .Crime analysis is exploring, inter relating and detecting relationships between various crimes and characteristics of crimes .Police department maintains crime data at the record. This data contains huge amount of data set with complex relationships which needs use of data mining techniques in order to be transformed into useful information. The knowledge extracted from the dataset can be a great tool & support to the police department to prevent crimes.

An ideal crime analysis tool should be able to identify crime patterns quickly and in an efficient manner for future crime pattern detection and action. However, in the present scenario, the following major challenges are encountered.

- Increase in the size of crime information that has to be stored and analysed.

- Problem of identifying techniques that can accurately and efficiently analyse these growing volumes of crime data Different methods and structures used for recording crime data.
- The data available is inconsistent and are incomplete thus making the task of formal analysis a far more difficult.
- Investigation of the crime takes longer duration due to complexity of issues

All the above challenges motivated this research work to focus on providing solutions that can enhance the process of crime analysis for identifying and reducing crime in India. The main aim of this research work consist of developing analytical data mining methods that can systematically address the complex problem related to various form of crime. Thus, the main focus is to develop a crime analysis tool that assists the police in

- Detecting crime patterns and perform crime analysis
- Provide information to formulate strategies for crime prevention and reduction
- Identify and analyse common crime patterns to reduce further occurrences of similar incidence
- The present research work proposes the use of an amalgamation of data mining techniques that are linked with a common aim of developing such a crime analysis tool. For this purpose, the following specific objectives were formulated.
- To develop a data cleaning algorithm that cleans the crime dataset, by removing unwanted data to explore and enhance clustering algorithms to identify crime patterns from historical data
- To explore and enhance classification algorithms to predict future crime behaviour based on previous crime trends
- To develop anomalies detection algorithms to identify change in crime patterns.

2. REVIEW OF LITERATURE

D.E. Brown constructed a software framework called ReCAP(Regional Crime Analysis Program) for mining data in order to catch professional criminals using data mining and data fusion techniques. In 2009, Li Ding et al.[11] propose an integrated system called PerSearch that takes a given description of a crime, including its location, type, and the physical description of suspects(personal characteristics or vehicles) as input. To detect suspects, the system will process these inputs through four integrated components: geographic profiling ,social network analysis, crime profile, and physical matching. Essentially, geographic profiling determines where" the suspects are, while other components determine the suspects. De Bruin et. al. (2006) introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly .This method also provided a visual clustering of criminal career sand identification of classes of criminals .From the literature study, it could be concluded that crime details increasing to very large quantities running into zota bytes(1024bytes). This in turn is increasing the need for advanced and efficient techniques for analysis. Data mining as an analysis and knowledge discovery tool has immense potential for crime data analysis. As is the case with any other new technology, the requirement of such tool changes, which is further augmented by the new and advanced technologies used by criminals. All these facts confirm that the field is not yet mature and needs further investigations.

3. METHODOLOGY

Following figure shows the methodology used for crime pattern analysis.

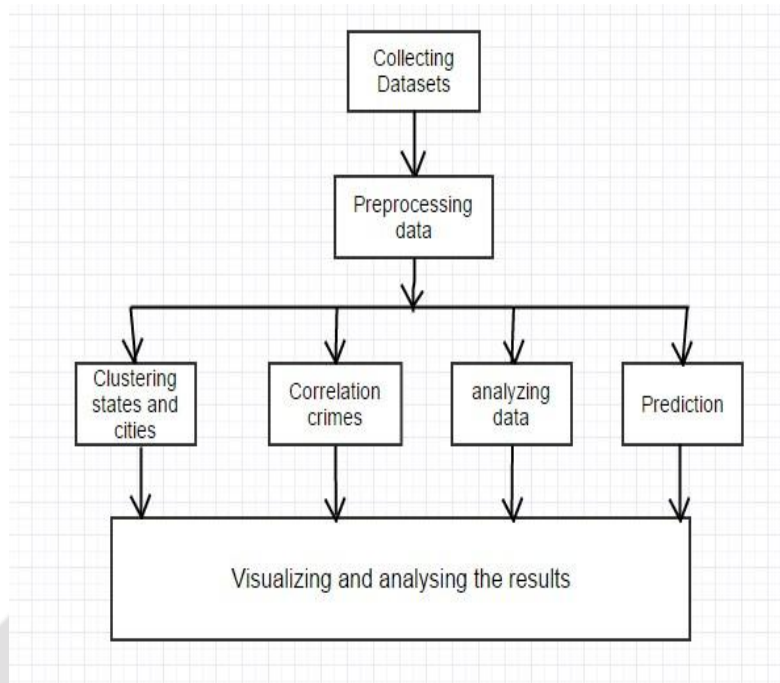


Fig.1 Methodology for crime pattern analysis

3.1. Data Collection

Enormous amount of crime data is collected at the end of year at police records. This data is made available by National Crime Bureau of Records. This data is in the form of number of cases recorded all over the nation throughout the year. The data is in raw form and also contains some wrong as well as missing values. Hence pre-processing of data becomes very necessary in order to bring the data in proper and clean form. Pre-processing of data includes data cleansing and PreProcessing.

3.2. Data Classification.

We classify the data set into various groups based on certain characteristics of the data object here we group crimes according to states & cities. Classification of the crime is done on the basis of different types of crime .K-means algorithm can be used to group data with similar characteristics.

3.2.1 K-Means Algorithm.

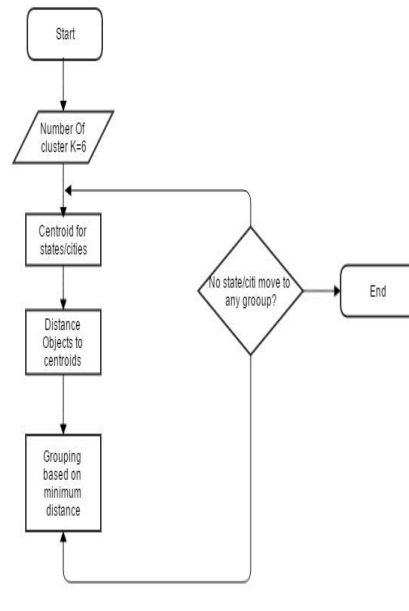


Fig2. K-Mean Algorithm

K-means algorithm mainly used to partition the clusters based on their means. Initially number of objects are grouped and specified as 'k' clusters. The mean value is calculated as the mean distance between the objects. The relocation iterative technique which is used to improve the partitions by moving objects from one group to other. Then number of iterations is done until the convergence occurs. K-means algorithm steps are given as

Input: Number of clusters.

Step1: Arbitrarily choose k objects from a dataset D of N objects as the initial cluster centers.

Step 2: reassign each object which distributed to a cluster based on a cluster center which it is the most similar or the nearer.

Step 3: Update the cluster means, i.e. calculate the mean value of the object for each cluster.

Output: A set of k clusters. K-means algorithm is a base for all other clustering algorithms to find the mean values.

3.3 Correlating Crime

Many crimes are related to other crime or criminal. Finding this correlation can be of great help in finding missing clues. Correlations can be used to help make predictions. If two variables have been known in the past to correlate, then we can assume they will continue to correlate in the future. We can use the value of one variable that is known now to predict the value that the other variable will take on in the future. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations. The form of the definition involves a "product moment", that is, the mean (the first moment about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name.

$$\rho_{X, Y} = \text{COV}(X, Y) / \sigma_X \sigma_Y$$

where: cov is covariance.

and σX is the standard deviation of X . Pearson product-moment correlation coefficient is a measure of the linear correlation or dependence between two variables X and Y , giving a value between $+1$ and -1 , where 1 is called total positive correlation, 0 is no correlation, and -1 is called total negative correlation. Figure 5 explains how correlation between X and Y depending on value of r .

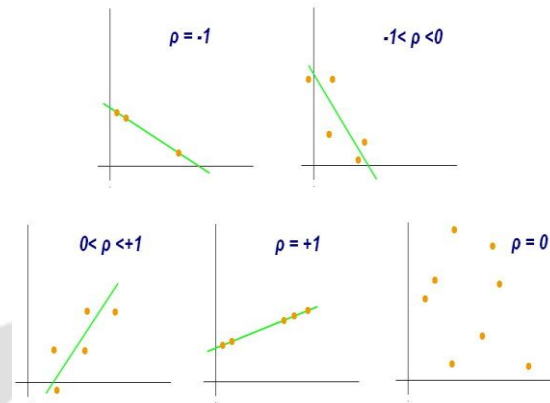


Fig3. Correlation example of person

3.4 Predicting Crime

Prediction of crime is a great aid to the administration in order to curb the crime incidences. Prediction is stating probability of an event in future period time. In this case crime against women can be predicted using linear regression. Prediction about various types of crimes and most probable places of occurrences of crime will be predicted linear regression is an approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variable) denoted X . The case of one explanatory variable is called simple linear regression. In simple linear regression, we predict value of one variable from the value of a second variable. The variable we are predicting is called the criterion variable and is referred to as Y . The variable we are basing our predictions on is called the predictor variable and is referred to as X . When there is only one predictor variable, the prediction method is called simple regression. In simple linear regression, the predictions of Y when plotted as a function of X form a straight line. Line a regression consists of finding the best-fitting straight line through the points. The best-fitting line is called a regression line. The most commonly-used criterion for the best-fitting line is the line that minimizes the sum of the squared errors of prediction.

The formula for a regression line is

$Y = aX + b$ where Y is the predicted score, b is the slope of the line, and A is the Y intercept. The slope (b) can be calculated as follows:

$$b = r \frac{S_x}{S_y}$$

And the intercept (A) can be calculated as

$$A = M_Y - bM_X.$$

Where, M_X is the mean of X , M_Y is the mean of Y , S_x is the standard deviation of X , S_y is the standard deviation of Y , and r is the correlation between X and Y .

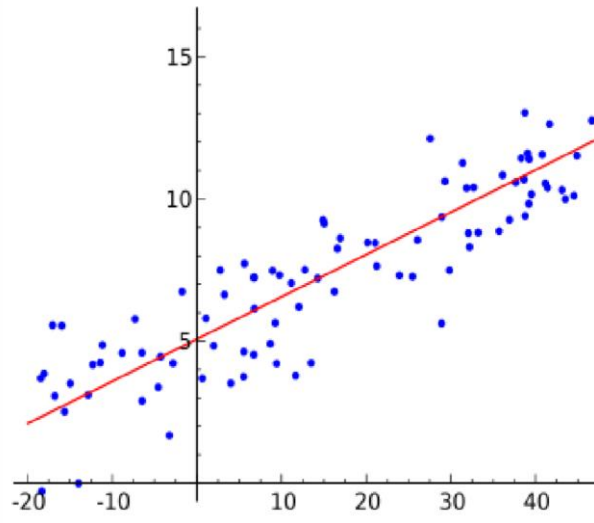


Fig4.Linear regression

4. Result and future scope

Result of this research will be to analyse, correlate and predict the crimes from huge data set available. Results will be in the form of correlation between various crime and location of crime i.e. state/city. Crime can also be correlated on the basis of age group, location of crime & type of crime. Prediction of the crime will be displayed using various diagrams pie charts, heat maps, spikes and graphs.

REFERENCES

- [1] K.S.Arthisree, M.E, A.Jaganraj, Arulmigu Meenakshi, "Identify Crime Detection Using Data Mining Techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Amman College Of Engineering, Thiruvannamalai district, Near Kanchipuram, India,volume 3,Issue 8,August 2013.
- [2] B.Chandra and M. P.Gupta, "Crime Data Mining for Indian Police Information System Manish Gupta", Indian Institute of Technology Delhi, Hauz Khas, New Delhi - 110 016,India.
- [3] D.E. Brown, "The regional crime analysis program (RECAP): A frame work for mining data to catch criminals", In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol. 3, pp. 2848-2853, 1998.
- [4] Ram Singla, Deepak Dembla, Yogesh Chaba and Kevika Singla CITM, "An Optimal KD Model for Crime Pattern Detection Based on Semantic Link Analysis",A Data Mining Tool C,Faridabad.
- [5] Malathi. ,Dr. S. Santhosh Baboo Reader, "An Enhanced Algorithm to Predict a Future Crime using Data Mining", International Journal of Computer Applications (09758887),Volume 21 No.1, May 2011.
- [6] TongWang1, Cynthia Rudin1, Daniel Wagner, and Rich Sevieri, "Learning to detect patterns of crime", JECET, 1:124131, 2012.
- [7] Revatthy Krishnamurthy,J. Satheesh Kumar, "SURVEY OF DATA MINING TECHNIQUES ON CRIME DATA ANALYSIS", International Journal of Data Mining Techniques and Applications Vol 01, Issue 02, December 2012.
- [8] Shyam Varan Nath Florida, "Crime Pattern Detection Using Data Mining", Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT 2006 Workshops) (WI-IATW06).
- [9] Priyanka Gera, Rajan Vohra, "City Crime Profiling Using Cluster Analysis", International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014.
- [10] Malathi. A and Dr. S. Santhosh Baboo, "Enhanced Algorithms to Identify Change in Crime Patterns", International Journal of Combinatorial Optimization Problems and Informatics, Vol. 2, No.3, Sep-Dec 2011, pp. 32-38, ISSN: 200