

Crime Analysis and Prediction Using Fuzzy C-Means Algorithm

B. Sivanagaleela, M.Tech
Computer Science and Engineering
V. R. Siddhartha Engineering College
Vijayawada, Andhrapradesh
Sivanagaleela.b@gmail.com

S. Rajesh, Asst.Professor
Computer Science and Engineering
V.R. Siddhartha Engineering College
Vijayawada, Andhrapradesh
rssingam99@gmail.com

Abstract— Crime analysis is methodological approach for identify the crime areas. The crime areas are mainly based on the crime type these identified crime areas are helpful to reduce the crime rate. This can be very easy to identify the crime areas, based on this process the crime rate can be analyzed. With the increasing of computer systems the crime data analysts can help to the crime investigators to analyze the crime. Based on the clustering and preprocessing extract the crime areas from a structured data. The cause of occurrences of crimes like crime details of person and other factors we are focusing mainly on crime factors of previous years. This system is mainly focus on in which area the crime will occur, does not focus on the identify the criminal. In the existing system naive bayes classification was used In the present system, the fuzzy C- Means algorithm will be use to cluster the crime data for total cognizable crimes such as Kidnapping, murder, Theft, Burglary, cheating, crime against women, robbery and other such crimes.

Keywords— Data preprocessing, clustering, Fuzzy Clustering, Crime Patterns, Crime prediction

I. INTRODUCTION

Crime data are categorized using the crime type, It can be useful to identify the crime areas based on the crime category and very useful to protect the crime areas to decrease the crime rate. Crimes are influenced by organizations and other places occurred frequently in a society. The main use of analyze the crime areas is to identify the crime areas after analyze the clusters, crime occurrences frequently during different years. Crime prediction is used to identify the crime occurred areas and reduce those crimes, this can be very useful to decrease the crime rate. Based on the structured data analyze the crimes.

Crime analysis is an approach it can be very useful to identify the crime areas based on the historical crime data the crime areas are will be identified. In this process analyze the crime problems in all types of categories, statewide the crime patterns are defined based on the geographical terms. Prediction can be based on crime type which may occur next in a location. The crime analysis is mainly used to analyze the crime areas to know in which area the crime occur frequently, using the past data of crime the analysis will be done.

In the crime analysis find the spatial criminal hotspots using a set of past crime records and the datasets of crimes. The need of crime analysis to identify the crime areas and type of crime is ongoing, timely manner, to utilize the crime investigators and it can be proactive for analyzing and rectifying the crime rate. Analyze crime to meet the law enforcement needs of a changing society analyze crime to understand the criminal behaviors.

Crime analysis is used to refer the process of crime occurrences and characteristics of crimes that means in which areas the crime rate is high this can be visualized. Crime analysis is an analytical process for identify the crime. It is used to decrease the crime rate in the areas it is useful for crime enforces to know the crimes trends, patterns and series in ongoing, timely manner. Analyze crime is an approach to solve the crime problems.

Based on these concepts crime analysis and prediction can be done

- a. Crime prediction
 - b. Data mining techniques
 - c. Clustering
 - d. Prediction
- a. The use of crime prediction is to analyzing and decreasing the future crime. Using the old data crime locations are identified.
 - b. The Data mining prediction techniques are useful to enhance the accuracy, performance, speed of predicting the crime. Using the mining techniques crime patterns are analyzed based on the current and old data
 - c. In clustering, the data items are clustered according to their logical relationships or natural groupings and a structure as a whole is generated. There are no pre-defined groups. Each cluster is collection of homogeneous elements.
 - d. Crime Prediction Techniques provide an accurate prediction for the location of the crime. Collecting and managing large volumes of accurate data.

The Fuzzy C-Means is a technique used for clustering of data. Fuzzy clustering is used for real life crime data. The main purpose of this technique is used for the investigators to analyze the crime resources. Each place of crime is identified based on the historical crime data and analyze the crime places. The FCM algorithm is useful for that the data will be clustered the datasets are gathered into n number of group with each data of the dataset is related to every cluster of specific degree. The purpose of the Fuzzy clustering techniques is used for clustering the crime data and identifies the clusters based on the dataset. It is a useful technique to identify various crime patterns and analyze crime data.

Scope

The main scope of the project is to analyze the crime patterns and identify the crime predicted areas based on the crime rate using the crime data. Using the fuzzy clustering obtain the crime patterns to know in which area the crime will occur frequently.

The Objectives are,

- Using the Fuzzy C- Means clustering algorithm to predict the crimes. The crime patterns can be analyzed and prevented based on the crime data.
- After collecting the data, preprocessing that data and to use clustering techniques to cluster the data of crime and analyze it.
- To analyze the data and identify the crime patterns, perform the prediction techniques based on the identified crime areas that are obtained after clustering.

II. RELATED WORK

Rasoul Kiani, siamak mahdavi[1] proposed that a crime analysis work based on the clustering and classification techniques. They proposed a model in which the how to predict and analyze the crimes are done through the optimization of the outlier detection these operator parameters which is performed through a Genetic Algorithm. The analysis of crime data and collection of previous crime data will be used to improve the security. Main Objective is classifying crime clusters based on the crime occurrence frequently during different places.

Shiju satyadevan, Devan M.S, Surya Gangadharan.S [2] introduced the how to analyze and prediction of crime using the data mining concepts and techniques and also that they identify the causes of crimes like background of the criminal and other causes and crime occurrences of the each day to identify the crime analysis the association rule mining algorithm is used for to identify the pattern of the crime and how to analyze that crime. This algorithm is also performing the identifying association rules and highlighting the operations of the database. The data is collected and placed in the database for the other process. Based on the unstructured

data they perform the operations. They classify the crime rate using the classification algorithm. Naive Bayes algorithm is used to obtain the model for crime data training like crimes such as Burglary, Arson, and crime against woman. The training means to analyze the inputs such as the testing on crime for unknown input data for testing the accuracy the model data and apply the test data.

Nikhil Dubey, Setu Kumar Chaturvedi [3] proposed the prediction techniques of data mining they are used to identify the accuracy of the crime and performance of the crime speed to control the crime rate. Crime prediction use the old crime data and after analyzing crime data the prediction of crime is identified with location, Data mining technique are very useful to solving crime detection problem. Different types of the data mining techniques are usefull for to identify the crime prone areas. They used the supervise learning technique like support vector machine algorithm. It is used to find a linear plane accuracy that will separate the data provide good accuracy almost all cases. Based on These techniques identify common pattern by comparing current and past crime data and predict the future value. Classically the crime is unpredictable the crime analysis takes the old crime records and obtain the future crime rates. It can be useful for reduce the future crime.

Anant Joshi, Sai sabita, Tanupriya Choudary [4], proposed a model using clustering technique, in clustering, the data items are clustered according to their logical relationships or natural groupings and a structure as a whole is generated. Using K-means clustering data mining approach on a crime data set analyze the regions and obtain crime rates of each type of crimes and cities with high crime rates. Each cluster is the collection of homogeneous elements, which may be exclusive to that group, but the clusters are similar to the each other. In clustering, the data items are clustered according to their logical relationships or natural groupings and a structure.

Vrushali Pednekar, trupti Mahale, Pratiksha Gadhve, Arti Gore [5] they proposed a model using the k- nearest neighbor algorithm, it is the systematic way to obtained and analyze the process of crime patterns and types. The use of an analytical method is to classifying the crime data based on the crimes using these data the analysis of a crime and detection is identified patterns of crimes, these type of crimes occurred at different times, the clusters refers a geographical location of a collection of crime that can be visualized using the geo-spatial plot in the map. The process of this algorithm does not identify not only the important attributes; it is very difficult to identify the mostly occurred crime areas.

III. METHODOLOGY

The detection of the crime types and patterns are used to identify and analyze the crime, this process is used to provide the information to reduce the crime rate and identify the common crime rates. The further occurrences of similar incidence The crime analysis is mainly focused on the crime regions in this process the crime data is used based on the crime data the crime areas are identified, mainly concentrate on crimes such as kidnapping, murder, Theft, Robber etc.

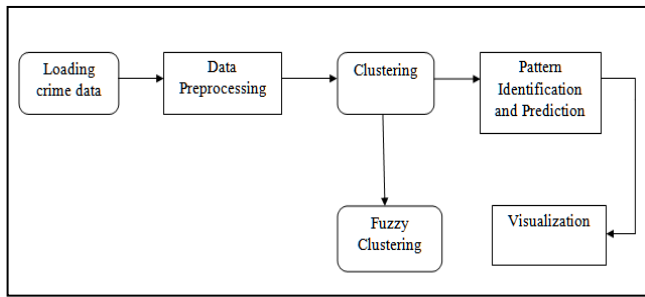


Fig. 1 Crime analysis steps

In the crime analysis process in the step one crime data will be loaded then the noisy data will be removed based on the preprocessing phase. Cluster the data using the clustering technique, the fuzzy c-means is a clustering algorithm is used to cluster the data then identify the crime patterns based on the crime data finally the visualize the crime prone areas.

The system mainly contain four modules

1. Preprocessing the crime data
2. Clustering Crime data
3. Pattern Identification of crime data
4. Predicting the crime areas

In this methodology crime data will be used, the data can be collected from the DataWorld website it maintain the district wise crime data and also consists of various crimes such as Kidnapping, Murder, Theft, Robbery etc.

Preprocessing the crime data

Crime patterns can extract using the crime data. The prediction of the crime data using the different types of existing crime data based on this perception the crime rate is identified. Detection of crimes, Predict the behavior of a criminal or groups of criminals according to their historical data with different attributes. Using this approaches the data is preprocessed it is used to preprocess the irrelevant data and perform the operations and analyze the crime patterns it can be used to identified the crime rate.

Clustering

In the step one the crime data will be loaded then perform the processing step to remove the noisy data in clustering phase the fuzzy clustering will be use, it is an unsupervised method. The clustering algorithm is used in this process like fuzzy, clustering process which is used to cluster the data in to the groups. The cluster analysis used the data variables based only on this data the information is gathered and the variables are described the relationships.

Fuzzy Clustering

Fuzzy clustering is used to assigning the each data point and each cluster is belongs to based on the membership degree, in fuzzy C an object can have a membership value of 0 to 1. Fuzzy is applicable for all types of data like categorical or numerical. In this clustering the values of the cluster data is connected to the each data point it is denoted as degree of the each clusters. The membership values of all the dataset should one. Clustering the raw data and producing the membership

function from the data and creating the fuzzy interface system fuzzy c handles uncertainty condition of the data.

In the fuzzy logic the membership degree is represented as f and degree of member ship function is μ_F , is a well known function and X is defined as the unit interval it is represented as $\mu_F : X \rightarrow [0, 1]$

0 - > It Means that in the given set member values are not included.

1 - > Describe that the member values are included in the set

The fuzzy members are in between the 0 and 1.

Fuzzy C-Means Algorithm

Step 1: Let assume that data points by x_i ($i=1, 2, 3, \dots, N$) these data points are clustered.

Step 2: Let assume the how many clusters are to be find, C , where $2 \leq C \leq N$.

Step 3: Choose the appropriate values of the fuzzy function $f > 1$

Step4: Defined $N \times C \times M$ sized matrix using the membership function U , in randomly, such that $U_{ijm} \in [0, 1]$ for each value of I fix the value of m

Step 5: The cluster centers are calculated v_j for the cluster i for its m^{th} dimension

$$V_j = \frac{(\sum_{i=1}^N (\mu_{ij})^m x_i)}{(\sum_{i=1}^N (\mu_{ij})^m)}$$

Where,

$V_j = j^{\text{th}}$ cluster

C = It represented the clusters

M = Fuzziness index of m cluster center

N = It represent the data points

Step 6: In this step the Euclidean distance is calculated for data point and the cluster with respect to m^{th} dimension

Step 7: Update the fuzzy membership matrix U_{ij} .

$$V_j = \frac{(\sum_{i=1}^N (\mu_{ij})^m x_i)}{(\sum_{i=1}^N (\mu_{ij})^m)}$$

d_{ij} = distance between the each data points

μ_{ij} = the membership degree of data points

Step 8: Repeat step 5 to step 7 until the minimum objective function is specified.

Pattern Identification of crime data

In this step the crime patters are identified. For finding crime pattern that occurs frequently in this phase the crime areas are identified that means in which areas the crime will occur highly. Here corresponding to each location we take the attributes of that place after getting a general crime pattern for a place, when a new case arrives it uses the previous crime pattern then it can be easy to analyze that to know that which area have high crime rate it is identified is easily to investigate crime. The crime analysis is the process of. Using this crime

data set it is very helpful to crime investigative officers to identify the crime rate this process is used to all types of crime categories.

Prediction

The crime areas are identified using the crime patterns prediction can be done using the old crime records. In this phase using the Decision tree techniques the crime areas will be predicted. The crime areas are identified using these old crime data the crime patterns prediction can be done based on the crime data. The prediction can be involved it is used to identified the crime rate and compare the each year crime rate based on the data this can be used for reduce the crime rate. Crime prediction is used to visualize the crime area.

IV. IMPLEMENTATION

In this phase based on the crime data the clustering can be done. Using the Indian crime data the crime areas are identified using the fuzzy clustering technique the clusters are identified. Compared to the other clustering techniques fuzzy techniques take less time to cluster the data and identifies the crime areas. Using the fcm clustering algorithm the degree of membership of each cluster is calculated and identifies the clusters using the algorithm steps the crime areas are identified. Finally the crime areas are identified based on the crime data. Which type of crime occurred very frequently is identified. After loading the crime data the noisy data is removed that means duplicate values are deleted and the ready based on the state names the crime areas are identified based on crime types of crimes are identified based on the location. The highly occurred crime areas and low crime rate areas are identified this can be represented in the graphical format.

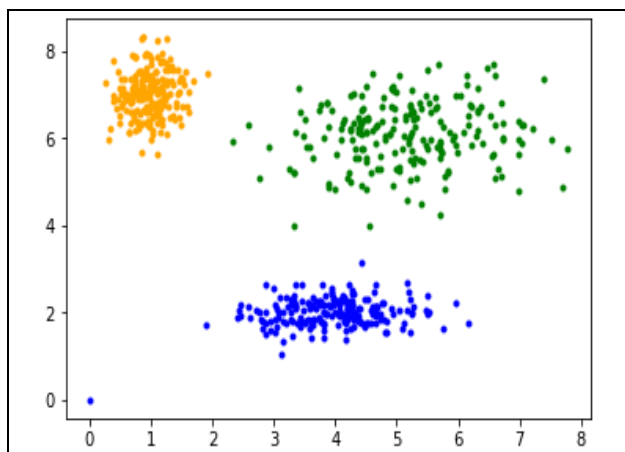


Fig. 2. Clustering of crime data

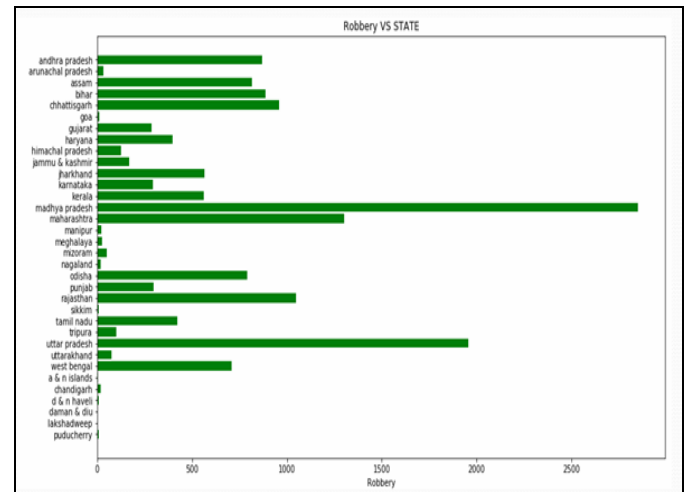


Fig. 3. Robbery rate in India

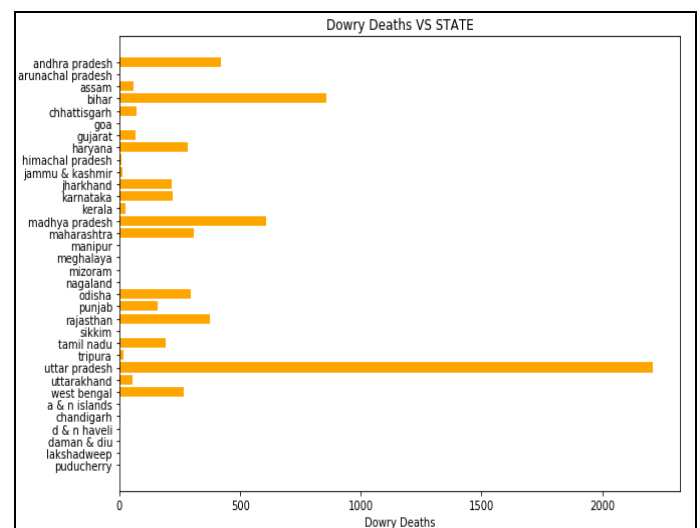


Fig. 4. Dowry Deaths in India

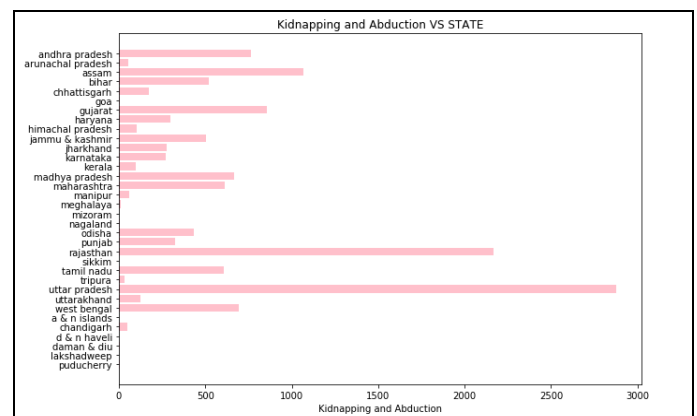


Fig. 5. Kidnapping rate in India.

CONCLUSION

This paper is useful to identify the crime areas based on the clustering technique. Crime patterns are also identified they are not static. We have identified the crime areas and which type of crime occurred very frequently in which place. It is very useful for investigators to solve the crime rate. Based on the fuzzy clustering technique the crime prone areas are identified it takes a less time.

REFERENCES

- [1] Rasoul Kiani, Siamak mahdavi, Amin Keshavarzi, "Analysis and Prediction of Crimes by Clustering and Classification", International Journal of Advanced Research in Artificial Intelligence, Vol. 4, No. 8, pp 11-17, 2015.
- [2] Shiju Sathyadevan, Devan M.S, Surya Gangadharan.S, "Crime Analysis and Prediction Using Data Mining Techniques", First International Conference on Networks & Soft Computing (ICNSC)", pp-406-412, IEEE, 2014.
- [3] Nikhil Dubey, Setu Kumar Chaturvedi "Crime Prediction Technique Using Data Mining", International Journal of Engineering Research and Applications, Vol. 4, Issue 3, pp 396-400, March 2014.
- [4] Anant Joshi, Sai sabita, Tanupriya Choudary, "Crime Analysis using K-means Clustering", 3rd International Conference on Computational Intelligence and Networks(CINE), pp-33-39, October 2017.
- [5] Vrushali Pednekar, trupti Mahale, Pratiksha Gadhve, Arti Gore "Crime Rate Prediction using KNN", International Journal on Recent and Innovation Trends in Computing and Communication, Vol 6, Issue 1, January 2018.
- [6] Aishwarya DS, Madhumalathi S, Manisha UA, Sushma K M, Ravikumar V G "Prediction of Crime Pattern and Suspects Using Data mining Techniques", 3rd National Conference on Image Processing, Computing, Communication, Networking and Data Analytics, June 2018.
- [7] Chhaya Chauhan, Smriti Sehgal "Crime Analysis Using Data Mining Techniques and Algorithms", International Conference on Computing, Communication and Automation (ICCCA), IEEE, May 2017.
- [8] Suil Yadav, Meet Timbadia, Ajit Yadav, Rohit Vishwakarma and Nikhilesh Yadav, "Crime Pattern Detection, Analysis & Prediction", International Conference on Electronics, Communication and Aerospace Technology (ICECA), pp225-230, IEEE, 2017.
- [9] Malathi. A, Dr. S. Santhosh Baboo, "An Enhanced Algorithm to Predict a Future Crime using Data Mining" International Journal of Computer Applications (0975 – 8887) Vol. 21–, No.1, May 2011.
- [10] J. Agarwal, R. Nagpal, and R. Sehgal, "Crime analysis using k-means Clustering", International Journal of Computer Applications, Vol. 83 , No4, December 2013.
- [11] J. Mohana Sundaram, Dr. T. Karthikeyan, R. Karthik Raj, "A Survey of Fuzzy Based ARM Clustering on Crime Pattern Discovery" International Journal of Scientific & Engineering Research, Vol 5, Issue 5, May-2014.
- [12] Hardi. M. Patel, Ripal Patel, "Enhance Algorithm To Predict A Crime Using Data Mining" Journal of Emerging Technologies and Innovative Research, Vol. 5, Issue 04, April 2017.