

Crime Detection and Prevention Using Fuzzy Control Logic

¹Keerthana Nithyakumar, ²Preetha Sridharan, ³Reshma Ramesh, ⁴S.Aarthi

^{1,2,3}III Year UG Students & ⁴Senior Assistant Professor

Department of Computer Science Engineering, Meenakshi Sundararajan Engineering College,
Anna University(affiliated) Chennai, Tamil Nadu, India

Abstract

Now-a-days, crime has increased drastically and crime prevention has become a global issue. Government and community officials are doing their best to prevent crime. This paper aims to provide means to curb the crime rates by implementing fuzzy logic to detect crime patterns. We introduce data mining and clustering algorithms to predict the occurrence of crimes. It is of interest that Fuzzy system analysis can make significant & correct predictions. The most common way to portray the human understanding is with the help of Natural Language Expression, also called fuzzy rule-based system. Fuzzy system is used to identify the trait of an individual. We take five different characteristics like economical status, family background, educational level, alcoholic & drug addict and criminal history.

Keywords

Crime, fuzzy system, natural language expression, rule based system, data mining, K means algorithm, Fuzzy C means

I. INTRODUCTION

Fuzzy set theory exhibits immense potential for effective solving of the uncertainty in the problem. It is a systematic process of gathering, classifying, investigating, and publicizing timely, exact, and useful evidence that describes crime patterns, crime trends, and potential suspects.

The randomness, non-linear nature, and seeming disorder of criminal activity, often makes it difficult to employ traditional prediction tools such as geographic information systems (GIS). Fuzzy is flexible, which will randomly initialize parameters and make them particularly appropriate for criminal activity forecasting. Data mining (DM) analyses (often large) observational data sets to find unsuspected relationships and summarizes the data in novel ways that are understandable and useful for the data owner.

Thus by making use of K-Means algorithm and fuzzy C means, we identify crime patterns that makes it easier to detect and hence, prevent them. We apply technologies in public security to support decision making for situational crime prevention. The expected outcome is a web-based system that supports decisions in which it can provide

intelligent recommendations, so that, we can increase police strategies in an area and make it less attractive to criminals.

II. CRIME ANALYSIS

Crime data mining is receiving increased attention to discover underlying patterns in crime data. The need to act quickly and response to the crime activity happened and identify the interconnection between different crimes. GIS is used to identify the crimes happened in the cities which is older data mining techniques. Real-time solutions can save significant resources and push the capability of law enforcement closer to the pulse of criminal activity. Modern computing systems provide a unique opportunity to study this vast amount of data in ways that were previously not feasible. The volume of data being digitally recorded about crimes, suspicious activities, and suspect records is at an all-time high.

Current manual inspection of crime data by analysts and only limited staff are presents to investigate the crime held, but it is insufficient since there are large number of data present so there should be a simultaneous action needed. Further, complex relationships between various crime attributes can be overlooked and it should be analysed by means of human interruption. Providing automated knowledge discovery tools becomes attractive to enhance and accelerate the efforts of local law enforcement. Similarly, certain crimes may be more probable in locations with higher populations and dense housing. Regional crime patterns can be discovered which enable law enforcement personnel and criminal investigators to address large-scale trends.

The central focus of crime analysis is the study of crime, chaos, and information related to the nature of incidents, criminals, and sufferers. Crime analysts also study other policerelated operational issues, such as insufficient staffs. There are many different types of crime are in crime analysis. The three important kinds of crime analysis are socio-demographic, spatial and temporal. First Goal: Normally after a robbery incident the police will start investigating the case based on the mode of crime. But in aspect of crime analyst this case is investigated based on the previous robbery case which is similarly happened. Second Goal: Based on the crime identified by investigated in first goal it must be prevented by

analysing and responses of those problems.

Two or more crime is discovered by police which is same then it is called crime pattern, because they meet each of the following conditions:

1. Similar type of crime; characters involved; location where it occurred.
2. Victims and Offenders have no relationship between them.
3. Crime occurred within the date range
4. Crime activity happened between the weeks or month.
5. Crime should be analysed and focused with the help of police strategies

Few general trends:

1. Series: Same victims are committing the same group of crime.
2. Hot Prey: Crime occurred by the two or more victims with similar physical properties.
3. Hot Product: Crime occurred by the two or more victim's unique property.
4. Hot Spot: Crime occurred by the two or more victims within the locations nearer to each other.
5. Hot Place: Crime occurred by the two or more victim's at same location
6. Hot Setting: Crime occurred by the two or more victim's which basically related with the place where the crime are occurred.

III. FUZZY LOGIC

Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic on which the modern computer is based. Fuzzy logic includes 0 and 1 as extreme cases of truth (or "the state of matters" or "fact") but also includes the various states of truth in between. Fuzzy logic seems closer to the way our brains work. We aggregate data and form a number of partial truths which we aggregate further into higher truths which in turn, when certain thresholds are exceeded, cause certain further results such as motor reaction.

A similar kind of process is used in neural networks, expert systems and other artificial intelligence applications. Fuzzy logic is essential to the development of human-like capabilities for AI, sometimes referred to as artificial general intelligence: the representation of generalized human cognitive abilities in software so that, faced with an unfamiliar task, the AI system could find a solution.

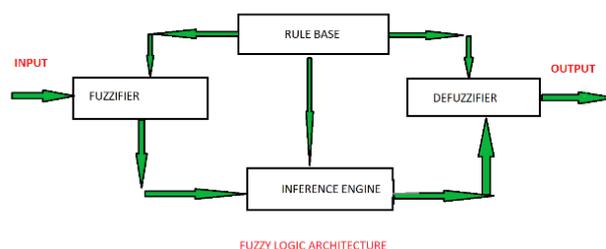
IV. MEMBERSHIP FUNCTION

Membership functions allow you to quantify linguistic term and represent a fuzzy set graphically. A membership function X for a fuzzy set A on the universe of discourse X is defined as $\mu_A: X \rightarrow [0,1]$.

Here, each element of X is mapped to a value between 0 and 1. It is called membership value or degree of membership. It quantifies the degree of membership of the element in X to the fuzzy set A .

- x axis represents the universe of discourse.
- y axis represents the degrees of membership in the $[0, 1]$ interval.

There can be multiple membership functions applicable to fuzzify a numerical value. Simple membership functions are used as use of complex functions does not add more precision in the output.



V. K-MEANS CLUSTERING

In present scenario criminals are becoming technologically sophisticated in committing crime and one challenge faced by intelligence and law enforcement agencies is difficulty in analysing large volume of data involved in crime and terrorist activities therefore agencies need to know technique to catch criminal and remain ahead in the eternal race between the criminals and the law enforcement.

So appropriate field need to be chosen to perform crime analysis and as data mining refers to extracting or mining knowledge from large amounts of data, data mining is used here on high volume crime dataset and knowledge gained from data mining approaches is useful and support police forces.

To perform crime analysis appropriate data mining approach needs to be chosen and as clustering is an approach of data mining which groups a set of objects in such a way that object in the same group are more similar than those in other groups and involved various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them.

In this paper k means clustering technique of data mining used to extract useful information from the high-volume crime dataset and to interpret the data which assist police in identify and analyse crime patterns to reduce further occurrences of similar incidence and provide information to reduce the crime.

K-means is a technique of clustering records into a determined number of unique clusters. The "K" refers to the number of clusters specified. There are various distance measures which determine the cluster to which an observation is to be assigned. This algorithm strives to minimize the measure between the centroid of the cluster and the given observation by iteratively assigning an observation to a cluster and terminate the loop when the smallest distance measure is obtained.

Following is the overview of the algorithm:

1. The sample space is initially partitioned into K clusters and the data elements are randomly assigned to the clusters.

2. Then for each sample:

- A. The distance from the element to the centroid of the cluster is evaluated.

- B. IF the sample is closest to its own cluster THEN leave it ELSE select another cluster.

3. Repeat steps 1 and 2 until no elements are moved from one cluster to another. When step 3 terminates the clusters are stable and each sample is assigned a cluster which results in the lowest possible distance to the centroid of the cluster.

Common distance measures include the Euclidean distance, the Euclidean squared distance and the Manhattan or City distance. We use the Euclidean distance to measure the distance between an element and the centroid. The squared Euclidean measure corresponds to the shortest geometric distance between two points. A faster way of determining the distance is by use of the squared Euclidean distance which calculates the above distance squared. K-means clustering is one of the methods of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean.

Process

1. Initially, the number of clusters must be known let it be k
2. The initial step is the choose a set of K instances as centres of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. This process is iterated. K means algorithm complexity is $O(tkn)$, where n is instances, c is clusters, and t is iterations and relatively efficient. It often terminates at a local optimum. Its disadvantage is applicable only when mean is defined and need to specify c, the number of clusters, in advance. It unable to handle noisy data and outliers and not suitable to discover clusters with non-convex shapes.

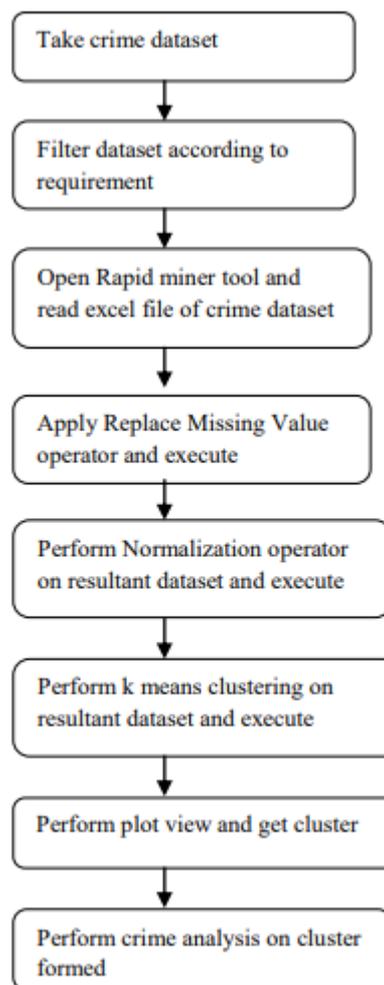


Fig 1: Flow chart of crime analysis

VI. C-MEANS CLUSTERING

Fuzzy clustering means (FCM) is an information bunching strategy in which a dataset is gathered into n bunches with each information point in the dataset having a place with each group to a certain degree. Fuzzy clustering means (FCM), otherwise called fuzzy ISODATA, is an information grouping calculation in which every information point fits in with a bunch to a degree defined by a participation grade.

Bezdek proposed this algorithm in 1973 as an improvement over earlier Hard C-means (HCM) clustering. FCM partition a collection of n vector $X_{j,z} = 1, \dots, n$ into c fuzzy groups, and finds a cluster center in each group such that a cost function of dissimilarity measure is minimized. The major difference between FCM and HCM is that FCM use fuzzy partitioning such that a given data point can belong to several groups with the degree of belongingness specified by membership grades between 0 and 1.

VII. NAÏVE BAYES CLASSIFIER

Naïve Bayes classifier There are various predictive methods that help police departments determine when and where crime will happen before it actually happens. In this analysis, we will focus on a simple predictive task of predicting which gang has committed the crime. For this, we are using the Naïve Bayes [5] approach.

In classification, posterior probability can be interpreted as: “the probability that a particular element belongs to class i given its observed values”. It can be expressed in simple words as follows: Posterior probability = (conditional probability * prior probability) / evidence Let x_i be the feature vector of sample $i, i \in \{1, 2, \dots, n\}$, ω_j be the notation of class $j, j \in \{1, 2, \dots, m\}$, $P(x_i|\omega_j)$ and $P(\omega_j)$ be the probability of observing sample x_i given that it belongs to class ω_j .

Algorithm 1: Pseudocode

- Given training data set D which consists of documents belonging to different class say class A and B
- Calculate the prior probability of class A = number of objects of Class A / total number of objects
Calculate the prior probability of class B = number of objects of class B / total number of objects
- Find n_i , the total number of word frequency of each class.
 n_a = the total number of word frequency of class A
 n_b = the total number of word frequency of class B
- Find conditional probability of keyboard occurrence given a class
 - ✓ $P(\text{word1}/\text{class } A) = \text{wordcount}/n_i(A)$
 - ✓ $P(\text{word1}/\text{class } B) = \text{wordcount}/n_i(B)$
 - ✓ $P(\text{word2}/\text{class } A) = \text{wordcount}/n_i(A)$
 - ✓ $P(\text{word2}/\text{class } B) = \text{wordcount}/n_i(B)$
 - ...
 - ...
 - ✓ $P(\text{wordn}/\text{class } B) = \text{wordcount}/n_i(B)$
- Avoid zero frequency problems by applying uniform distribution
- Classify a new document C based on the probability $P(C/W)$
 - ✓ Find $P(A/W) = P(A) * P(\text{word1}/\text{class } A) * P(\text{word2}/\text{class } A) \dots P(\text{wordn}/\text{class } A)$
 - ✓ Find $P(B/W) = P(B) * P(\text{word1}/\text{class } B) * P(\text{word2}/\text{class } B) \dots P(\text{wordn}/\text{class } B)$
- Assign document to class that has higher probability
-

VIII. CLUSTER ANALYSIS

This involves tracking crime rate changes from one year to the next and used data mining to project those changes into the future. Here we consider homicide crime and plot it with year and analysis variation in graph on cluster formed.

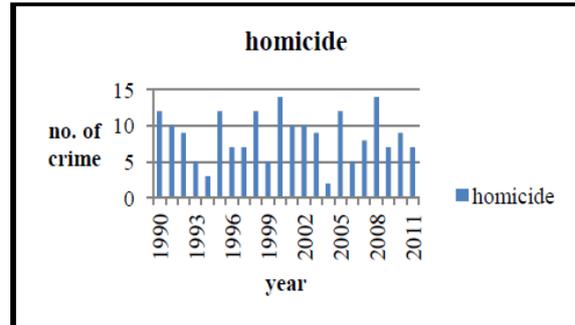


Fig 2: Homicide is minimum in 2004 and maximum and same in 2000 & 2008

From Fig 2 it can be seen that in year 2004 number of homicide crime committed is minimum as compared to in year 2008 where maximum number of homicide crime committed.

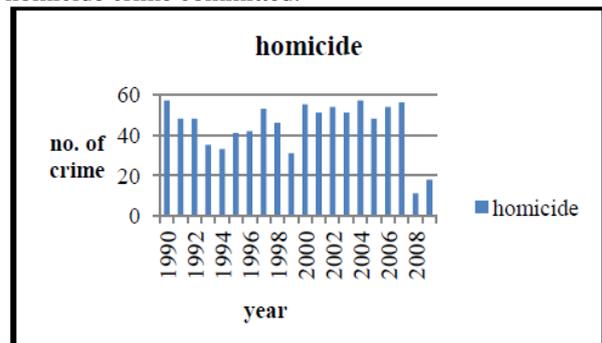


Fig 3: Homicide is minimum in 2008 and maximum in 1990 & 2004.

From Fig 3 it can be seen that in year 2008 number of homicide crime committed is minimum as compared to in year 1990 and 2000 where maximum number of homicide crime committed.

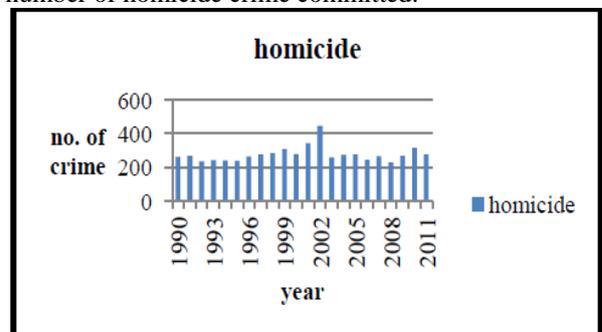


Fig 4: Homicide is minimum in 1992 and maximum in 2002

From Fig 5 it can be seen that in year 2011 number of homicide crime committed is minimum as compared to in year 2003 where maximum number of homicide crime committed.

IX. CONCLUSION

This project focuses on crime analysis by implementing clustering algorithm on crime dataset using k and c means clustering algorithms and here we do crime analysis by considering crime homicide and plotting it with respect to year and got into conclusion that homicide is decreasing from 1990 to 2011 .From the clustered results it is easy to identify crime trend over years and can be used to design precaution methods for future.

X. FUTURE SCOPE

From the encouraging results, we believe that crime datamining has a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis. Visual and intuitive criminal and intelligence investigation techniques can be developed for crime pattern. As we have applied clustering technique of data mining for crime analysis, we can also perform other techniques of data mining such as classification. Also, we can perform analysis on various dataset such as enterprise survey dataset, poverty dataset, aid effectiveness dataset, etc.

REFERENCES

- [1]. Risk assessment of crime prevention system based on Fuzzy logic and entropy theory- Research Gate
- [2]. Fuzzy based ARM clustering-IJSER