CRIME RATE PREDICTION USING DATA MINING

Prof. Kaveri Kar
Asst. Professor, CSE Dept.
SSIPMT
Raipur, Chhattisgarh

Rekha Jajodia
CSE Dept.
SSIPMT
Raipur, Chhattisgarh

Dipti Pritwani
CSE Dept.
SSIPMT
Raipur, Chhattisgarh

***ABSTRACT*—**Crime Is One Of The Most Predominant And Alarming Aspects In Our Society And Its Prevention Is A Vital Task. Crime Analysis Is A Systematic Way Of Detecting And Investigating Patterns And Trends In Crime. Thus, It Becomes Necessary To Study Various Reasons, Factors And Relationship Between Different Crimes That Are Occurring And finding The Most Appropriate Methods To Control And Avoid More Crimes. Data Mining Is Used Broadly In Terms Of Analysis, Investigation And Discovery Of Patterns For Occurrence Of Different Crimes. In This Work, Various Clustering Approaches Of Data Mining Are Used To Analyze The Crime Data. The Proposed System Can Predict Regions Which Have High Probability For Crime Rate And Can Forecast Crime Prone Areas. Instead Of Focusing On Causes Of Crime Occurrence Like Criminal Background Of Offender, Political Enmity Etc, It Will Focus Mainly On Crime Factors Of Each Day.

Keywords—crime, prediction, analysis, investigation

I. INTRODUCTION

Crime rate is increasing now-a-days in many countries. In today's world with such higher crime rate and brutal crime happening, there must be some protection against this crime. Here we introduced a system by which crime rate can be reduced. Crime data must feed into the system. We introduced data mining algorithm to predict crime. Data mining algorithm will extract information and patterns from database. System will group crime. Clustering will be done based on places where crime occurred, gang who involved in crime and the timing crime took place.

Data mining consists of five major elements:-

- Extract, transform, and load the data into the data warehouse.
- Store and manage the data in a multidimensional database system.
- Provide data access to business analysts and information technology professionals.
- Analyze the data by application software.
- Present the data in a useful format, such as a graph or table.
- Data mining consists of techniques such as association rule mining, classification, and clustering. This paper focuses on Classification and Clustering to analyze crimes.

II. K-NEAREST NEIGHBOR ALGORITHM

A. ABOUT KNN CLASSIFICATION ALGORITHM

K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition

KNN is a non-parametric; its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point. When we say a technique is non-parametric, it means that it does not make any assumptions on the underlying data distribution

KNN is also a lazy algorithm. This means is that it does not use the training data points to do any generalization.

KNN can be used for classification—the output is a class membership (predicts a class—a discrete value). An object is classified by a majority vote of its neighbors, with the object being assigned to the class most common among its k nearest neighbors. It can also be used for regression—output is the value for the object (predicts continuous values). This value is the average (or median) of the values of its k nearest neighbors.

B. HOW DOES THE KNN ALGORITHM WORK?

In KNN, k is the number of nearest neighbors. The number of neighbors is the core deciding factor. K is generally an odd number if the number of classes is 2. When k=1, then the algorithm is known as the nearest neighbor algorithm. This is the simplest case. Suppose p1 is the point, for which label needs to predict. First, you find the One Closest Point to P1 and Then the Label of the Nearest Point Assigned to P1.

Fig. 1. KNN Classification Algorithm.

ALGORITHM: KNN Classification

Input:

1. A finite set D of points to be classified,
2. A finite set T of points,
3. A function c: T –>{1,...,m},
4. A natural number k.

Output: A function r: D{1,...,m}

Method:

1. Begin
2. For each x in D do
3. Let U <–{}
4. For each t in T add the pair (d(x,t) , c(t)) to U;
5. Sort the pairs in U using the first components;
6. Count the class labels from the first k elements from U;
7. Let r(x) be the class with the highest number of occurrence;
8. End For each
9. Return r
10. End

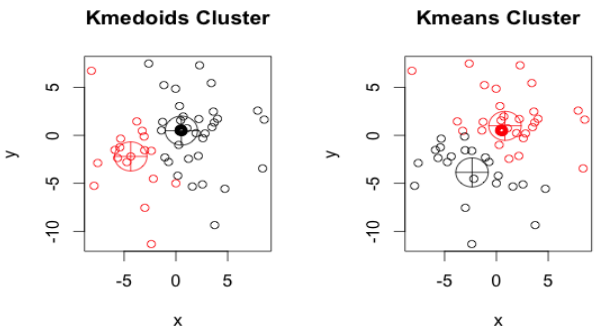
III. K MEDOID ALGORITHM

The K-Medoids Or PAM (Partitioned Around Medoids) Algorithm Is A Clustering Algorithm Reminiscent To The K-Means Algorithm. The K-Medoid Is A Classical Partitioning Technique Of Clustering, Which Clusters The Data Set Of N Objects Into K Clusters, With The Number K Of Clusters Assumed Known A Priori. If Unknown, K Can Be Determined With Methods Such As Silhouette.

A. MEDOID

A Medoid Can Be Defined As The Object Of A Cluster Whose Average Dissimilarity To All The Objects In The Cluster Is Minimal. I.E., It Is A Most Centrally Located Point In The Cluster.

Fig. 2. Difference Between K-Medoid Cluster And K-Means Cluster



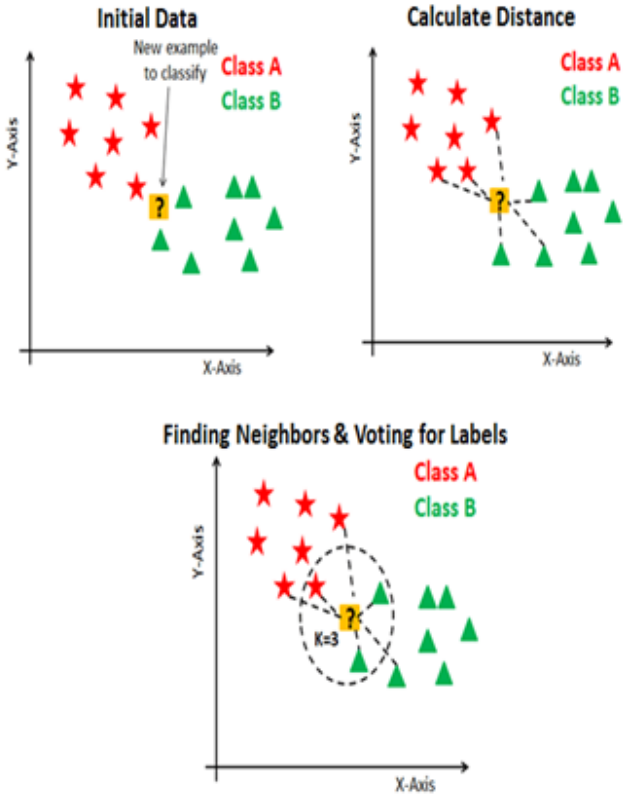


Fig. 3. KNN Explanation

ALGORITHM: K MEDOID ALGORITHM

Input:

k: the number of clusters

D:a data set containing n objects

Output:

A set of k clusters

Method:

1. Arbitrarily choose k objects in D as the initial representative objects or seeds;
2. **Repeat**
3. Assign each remaining object to the cluster with the nearest representative object;
4. Randomly elect a non-representative object O_{random} ;
5. Compute the total cost, S, of swapping representative object O_J, with O_{random};
6. If S<0 then swap O_J with O_{random} to form the new set of k representative objects;
7. **Until** no change;

Fig. 4. K-Medoid Algorithm

IV. DIFFERENCE BETWEEN K-MEANS AND K-MEDOIDS ALGORITHM

Both k-means and k-medoids algorithms are breaking the dataset up into k groups. Also, they are both trying to minimize the distance between points of the same cluster and a particular point which is the center of that cluster.

As we can see in figure 2, the medoids and centroids (of k-means) are slightly different in each group. Also we should note that every time we run these algorithms, because of the random starting points and the nature of the minimization algorithm, we will get slightly different results.

In contrast to the k-means algorithm, k-medoids chooses data points as centers (medoids or exemplars) and can be used with arbitrary distances, while k-means only minimizes the squared Euclidean distances

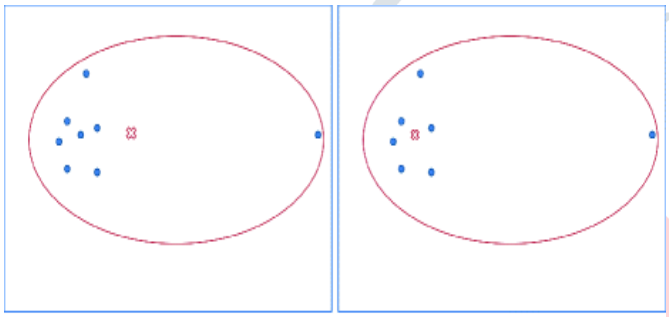


Fig. 5. Difference Between Mean And Medoid

V. CONCLUSIONS

Considering The Methods Proposed For Crime Prediction Shows That The Parameters Such As The Effect Of Outliers In The Data Mining Preprocessing, Quality Of The Training And Testing Data, And The Value Of Features Have Not Been Addressed Before. These Methods Can Be Used To Predict Crime Prone Regions In India On A Particular Day.

From The Clustered Results It Is Easy To Identify Crime Trend Over Years And Can Be Used To Design Precaution Methods For Future.

VI. FUTURE SCOPE

From The Encouraging Results, We Believe That Crime Data Mining 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

VII. REFERENCES

1. Miss. Vrushali Pednekar, Miss. Pratiksha Gadhave, Miss Trupti Mahale(Saraswati College of Engineering, Navi Mumbai), ”Crime Rate Prediction using KNN” International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 6 Issue: 124 – 127 January 2018
2. J. Han, and M. Kamber, “Data mining: concepts and techniques,” Jim Gray, Series Editor Morgan Kaufmann Publishers, August 2000.Https://En.Wikipedia.Org/Wiki/K-Medoids
3. R. B. Santos, Effectiveness of police in reducing crimes and the role of crime analysis, in Crime Analysis with Crime Mapping, ed. R. B. Santos (Sage, 2012), pp. 40–53.
4. ˜J. Agarwal, R. Nagpal, and R. Sehgal, “Crime analysis using k-means clustering,” International Journal of Computer Applications, Vol. 83 No4, December 2013.
5. L. Ding et al., “PrepSearch: an integrated crime detection system,” 2009 IEEE 161-163 ISI 2009, June 8-11, 2009, Richardson, TX, USA.
6. A. Babakura, N. Sulaiman, and M. Yusuf, “Improved method of calssification”