International Journal of Computer Applications (0975 – 8887)
Volume 122 – No.8, July 2015

Comparative Evaluation of Crime Incidence using Enhanced Density based Spatial (DBSCAN) Clustering

V.Vinodhini
(Research Scholar), Karpagam University and working as Assistant Professor
Dr.N.G.P Arts and Science College, Coimbatore-48, India

M.Hemalatha, PhD
Department of Computer Science, Karpagam University, Coimbatore, India

ABSTRACT
Criminology is the cream of crimes, crime fatalities, theories of ill-legal and abnormal behavior, social exploration, anti-crime polices, the political terrain of social control. So criminology involves the factual deeds in the streets, police stations, and courts, behind prison bars, board rooms and battlefields. Its practitioners are likely to slot in the orderly appraisal of the effectiveness of criminal justice policies and proposals, as well as the discovery of the source-cultural, economic and global roots of crime, rates of crime and meaning of crime, or the diverse ways of measuring criminal activity and its impact. Criminologists typically accumulate and scrutinize data sets that may be quantified, for example statistical studies on the rise and fall of crime rates, and/or qualitative, for example ethnographic studies on street subcultures and drug use. The research work concentrates in bringing of qualitative and quantitative study of crime rates and their behaviors. Here in this research work Density based spatial clustering is compared and analyzed with an enhanced DBSCAN algorithm, the results are also grouped in order to provide the efficiencies of crime rates. The outlook research work can be resolved by enhancing hybrid models in order to have condensed outlier detection.

Keywords
Criminology, Criminologists, Ethnographic studies, Enhanced DBSCAN, Hybrid model, DBK algorithm.

1. INTRODUCTION
Data mining is a powerful tool that enables criminal investigators that may lack extensive training as data analysts to explore large databases quickly and efficiently. Traditional data mining techniques such as association analysis, classification and prediction, cluster analysis, and outlier analysis to identify patterns in structured data. Newer techniques identify pat- terns from both structured and unstructured data. As with other forms of data mining, crime data mining raises privacy concerns. Nevertheless, researchers have developed various automated data mining techniques for many local law enforcement and national security applications.

Crime Analysis and its Steps

- Amplify in the size of crime in rank
- Predicament of identifying techniques that can precisely and ably explore crime data
- Diverse methods and structures used for footage crime data also the data accessible are contradictory and are curtailed.
- Inquiry takes longer period.

The steps used for crime analyzing are Extraction of patterns, Pattern representation and clustering.

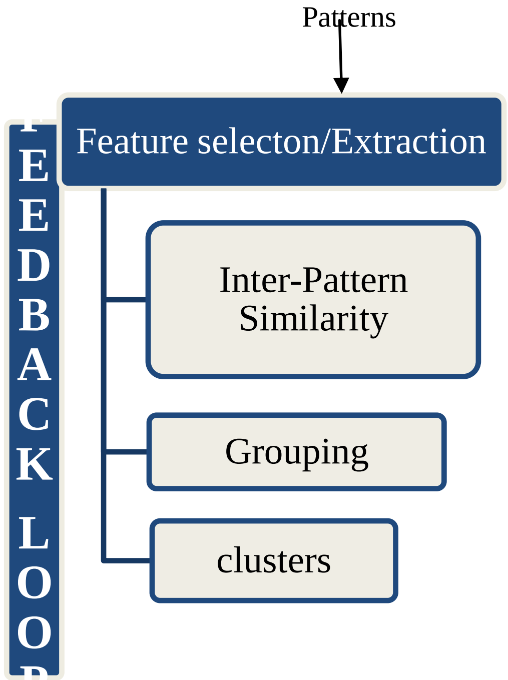

FIG I: Clustering Analysis

Crime cluster analysis is used to

i) Recognize areas with higher incidences of particular types of crime

ii) Crime Patterns

iii) Manage law enforcement resources effectively.

In this research the two popular techniques are used

i) Enhanced K Means
ii) Enhanced DBSCAN

Objectives
i) Primary objective-To enhance K Means Clustering algorithm
ii) Secondary objective- To enhance Density based Spatial Clustering of application with noise algorithm.

In this research work enhanced DBSCAN is used because DBSCAN algorithm is deliberate, but can cluster even in the presence of noise also eat plants unswervingly with whole data set. The memory requirement produced by DBSCAN is high; in order to condense the memory requirement of density based algorithm novel partition and merge procedure is used, this can also term as DBK algorithm. DBSCAN can detect random wrought cluster. Since DBSCAN is a densely based cluster the searching is performed on dense region, i.e., on the whole dataset. The searching performed are on the whole dataset, so it contributes to heavy computations.

In order to condense heavy computation merging procedure is used, before merging the clusters are partitioned into smaller sized sets and clustering is applied to this partition. The probing are performed only on to smaller sized sets in order to condense the number of scans and memory requirement. The statute based procedure is used to merge dense region on each partition.

2. METHODOLOGY

The methodology includes two Segments, the first Segment includes DBSCAN algorithm, the second Segment includes Enhanced DBSCAN algorithm. Further research work is concentrated on a brief review of,

PHASE 1: Enhanced DBSCAN algorithm:

Enhanced DBSCAN clustering algorithm uses,

- Automatic selection of starting point for clustering
- Automatic selection of parameters and MinPts values
- Speed Optimization Operations
- Improve Clustering Operation
- Dimensionality Reduction

Conventional DBSCAN Algorithm

Step 1: Begin with a decision on the value of k = number of clusters

Step 2: Put any initial partition that classifies the data into k clusters.

- Take the first k data as single-element clusters
- Assign each of the remaining (N-k) data to the cluster with the nearest centroid. After each assignment, recomputed the centroid of the gaining cluster.

Step 3: Take each data in sequence and compute its distance from the centroid of each of the Clusters. If a sample is not currently in the cluster of the closest centroid, switch this Sample to that cluster and update the centroid of the cluster gaining the new sample and the cluster losing the sample. The distance metric used is Euclidean distance

Step 4: Repeat step 3 until convergence is achieved, that is, until a pass through the training Sample causes no new assignments.

Issues and solutions proposed

- **Issue 1** : Distance Metric
  Solution : Distance Measure_that considers both Local and Global Consistency along with intra and inter distance between data points
- **Issue 2** :Automatic 'k' Value Estimation
  Solution : Enhanced Bayesian Information Criterion using Modified Dynamic Validity Index
- **Issue 3** : Initial Centroid Selection
  Solution : Reverse Neighbor Node (RNN) and Coupling Degree
- **Issue 4** :High Computations
  Solution : Reduction of Distance Calculation Mechanism
- **Issue 5** :High Dimensionality
  Solution : Principal Component Analysis (PCA)

Input: Dataset D, k , Output: C_j (j=1..k), n – Number of data points in D, k is number of clusters

- Compute the distances between objects in D, calculating the minimum distance from each cluster. The distance between clusters should be minimized in order to produce accurate.
- Compute the average distance between all objects, ε.
- Find neighborhood of objects in D, neighborhood cluster distance is calculated.
- Compute Cohesion (x_i) for all $x_i \in D$ (i = 1..n).
- Find x_i whose Cohesion(x_i) is maximum ,Add to C (First Centroid)
- Repeat Steps for the next highest cohesion object x_j, the steps are repeated until higher cohesion is achieved, again centroid position is calculated for each cluster.
- If Coupling($N(x_i)$,$N(x_j)$) < ε, next centroid is found ,Add to C (Next Centroid)
- If |No of centroid| < k, Go to Step 6(Maximum Cohesion (x_i) value) else go to End.
- End

Input: Dataset D with n data points {x_i, i = 1 .. n}, cluster number k. Output: Clustered da taset $C_1, \ldots, C_k$.
The Distance measure is computed using the formula DLG, in order to produce density measure ,

$$\mathrm{DLG}_{ij} = \min\left(\sum_{e=1}^{p} \mathrm{L}(\mathrm{p}_e, p_{e+1}) \right)$$

- Initialization. Randomly choose k data points from the data set to initialize k cluster centers;
- For any two points x_i, x_j, compute DLG (Distance measure)
- Each point is assigned to the cluster whose DLG of its center to the point is minimum; after clustering the centroid position searching and merging using rule based procedure is calculated to have high dimensionality with less noise reduction(DBSCAN).
- Recalculate the center of each cluster;
- If no points change categories then stop. Otherwise, go to output clustered dataset.

- Input : Dataset D with x_i (i = 1..n) data pointsOutput : Clusters ($C_1, \ldots, C_k$)
  Apply PCA to D to obtain D', a dimensionality reduced dataset, Automatic estimation of K .
  Automatic estimation of K initial seeds (C_j) .
- Repeat the Steps for each point x_i in D'.
- Calculate distance between each data point x_i and all k cluster centers (c_j, j = 1, .., K) using new distance measure.
- Find the closest center c_j and assign x_i to cluster j.

- Store label of cluster center j along with the distance of x_i to c_j and store them in array Cluster[] and Dist[] respectively,
  Set Cluster[i] = j (j is the nearest cluster)
  Set Dist[i] = DLG_{ij} (distance between x_i to closest cluster center c_j)
- Recalculate Cluster centers
- Compute DLG_{new} of x_i to new cluster centers until convergence

If DLG_{new} is less than or equal to DLG_{ij}, then x_i belongs to the same cluster j ,else
Compute DLG with every other cluster center and assign x_i ,to the cluster whose DLG is Minimum, then
Set Cluster[i] = j and Set Distance[i] = DLG_{new} .

- Output clustered results

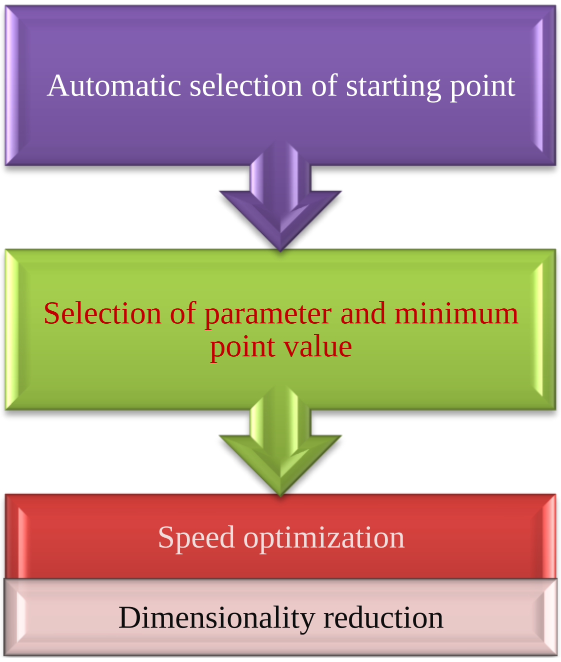

FIG II: Enhanced Density Based Clustering Algorithm

3. EXPERIMENTAL RESULTS

Table 1 presents the accuracy and speed of the proposed Enhanced KMeans model and the conservative improved KMeans algorithms.

TABLE I: Accuracy (%) and Speed (Seconds) of Enhanced DBK Clustering Algorithm

Algorithm	Accuracy	Speed
K Means	78.81	15.57
Enhanced K Means	80.16	14.91
DBSCAN	82.26	18.66
Enhanced DBK	89.22	13.05

By the Proportional analysis it is proved that the DBSCAN algorithm showed an efficiency gain of 18.66% and Enhanced DBK showed the accuracy rate of 89.22% respectively, it is evident that the proposed model is proficient and has enhanced the clustering accuracy and speed of DBSCAN algorithm.

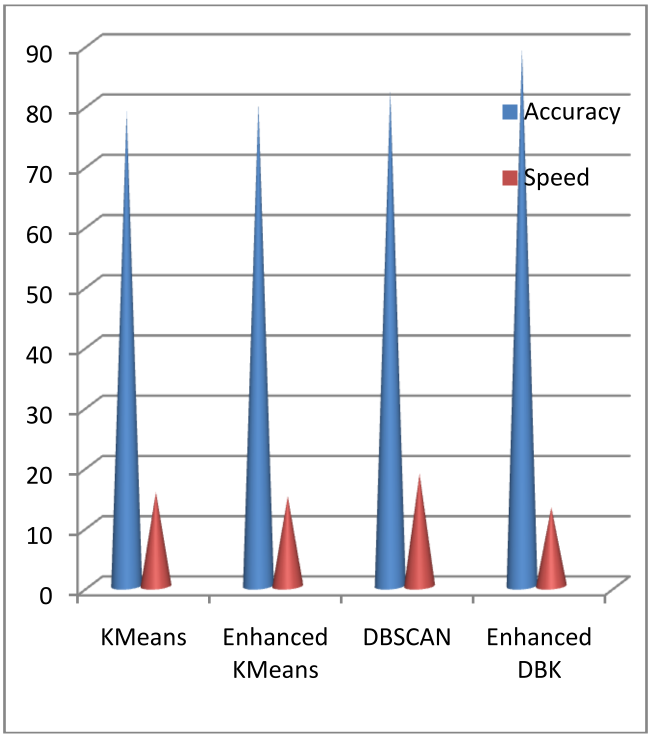

Fig V: Accuracy (%) and Speed (Seconds) of Clustering

4. CONCLUSION

Experimental results showed that the proposed algorithm is efficient in terms of clustering accuracy and speed of clustering. The study gives the comparative analysis of KMeans and Enhanced DBSCAN algorithm. Enhanced DBK algorithm proves to have high accuracy rate even in noisy environment. Prospect work can be concentrated on ensemble model i.e., by combining enhanced KMeans and enhanced DBSCAN a hybrid model called ensemble hybrid model is produced which tolerates more memory requirements . Future research is deliberated in this track.

5. REFERENCES

[1] Chen, Yudong; Zhang, Yi; Hu, Jianming; Li, Xiang (2006). "Traffic Data Analysis Using Kernel PCA and Self-Organizing Map". IEEE Intelligent Vehicles Symposium.

[2] Mining-Intelligence-Fraud-Criminal.../B008KZUJDK, E. and Usul, N. (2005) Geographic

[3] Information Systems Technologies in Crime Analysis and Crime Mapping, Pp.1-12.Norén, G. Niklas; Bate, Andrew; Hopstadius, Johan; Star, Kristina; and Edwards, I. Ralph (2008);

[4] Temporal Pattern Discovery for Trends and Transient Effects: Its Application to Patient Records. ProceedingsoftheFourteenthInternationalConfeLsVegas, NV,pp.963971Keim, Daniel A.; Information Visualization and Visual Data Mining.

[5] Joseph; Data Mining of Online Judicial Records of the Networked US Federal Courts, International

[6] Journal on Social Media: Monitoring, Measurement, Mining, 1:69–83 (2010).Vidisha H. Zodape (MTech. IIIrdsem, student)CSE Department, PIET Nagpur, Maharashtra, India,

[7] International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169 Volume: 3 Issue: 2 178– 183Prof. Leena H. Patil, CSE Department, PIET Nagpur, Maharashtra, India A R "Fast Algorithms for

[8] Mining Association Rules", Sep 12-15 1994, Chile, 487-99, pdf, 1-55860-153-Ms.Monalsaxena "Association rulesMining Using Improved FrequentPattern Tree Algorithm",

[9] International Journal of Computing,Communications and Networking, Volume 2, No.4, October – December 2013. Strategies and methods for detection and preventation delena d.spann

[10] James D.Rately,president of the ACFE, stated in 2012 report-On Occpational fraud and

[11] Abuse. Boba, Rachel (2005). Crime Analysis and Crime Mapping. Sage Publications. pp. 5–6.