I J C T A, 8(2), 2015, pp. 499-504

Approaches to Scrutinize Crime Data Using Improved (IKM) K-means Clustering Algorithm

V. Vinodhini[1] and M. Hemalatha[2]

Abstract: Criminology is an area that focuses on the scientific study of crime and criminal behavior and law enforcement. Criminology is a carcass of knowledge regarding crime as a social episode, it includes within its scope of processing, i.e., it includes making of laws (agreement of the will of the public), contravention of laws (disobedience of the requirements of the criminal laws which is created by the public) and taking action related to contravention of laws (response of the society who reacts either positively or negatively when someone commits a crime). Criminology is an area that practice or aims to find crime actions and crime uniqueness. As said earlier data mining plays a crucial task in producing crime intelligence. The research methodology which supports in producing Improved KMeans Clustering algorithm with Performance evaluation and comparison methods. The Performance evaluation which is used to evaluate datasets and performance metrics.

Keywords: Criminology, KMeans, Improved KMeans, Improved DBSCAN, Performance metrics.

I. INTRODUCTION

Criminology, it is a study of the causes of crimes and development of criminals and also it is the study of the origin and development of criminal laws. It is an efficient process of collecting, categorizing, analyzing, and disseminating timely, accurate, and useful information that describes crime patterns, crime trends, and potential suspects.

Today, crime has become a challenging factor. Some of the challenges behind the crime analyzing are

- Increase in size of crime information.
- Problem of identifying techniques that can accurately and efficiently analyze crime data
- Different methods and structures used for recording crime data.
- The data available is inconsistent and are incomplete.
- Investigation of the crime takes longer duration.

The stages used in crime, analyzing include a selection of raw data, preprocessing of target data, transformation of clean data, and mining of transformed data. The mining stage includes classification, clustering, association and rule mining. Here the work is concerted on clustering techniques; the clustering techniques used are KMeans and improved KMeans Clustering algorithm. The next stage is an evaluation of pattern discovery. Pattern discovery plays a vital role, where several techniques are used for this purpose, this research work analyzes the use of clustering techniques for mining knowledge in crime data.

[1] (Research Scholar), Karpagam University and working as Assistant Professor, Dr. N. G. P Arts and Science College, Coimbatore-48, India, *E-mail: vvprof133@gmail.com*

[2] Department of Computer Science, Karpagam University, Coimbatore, India

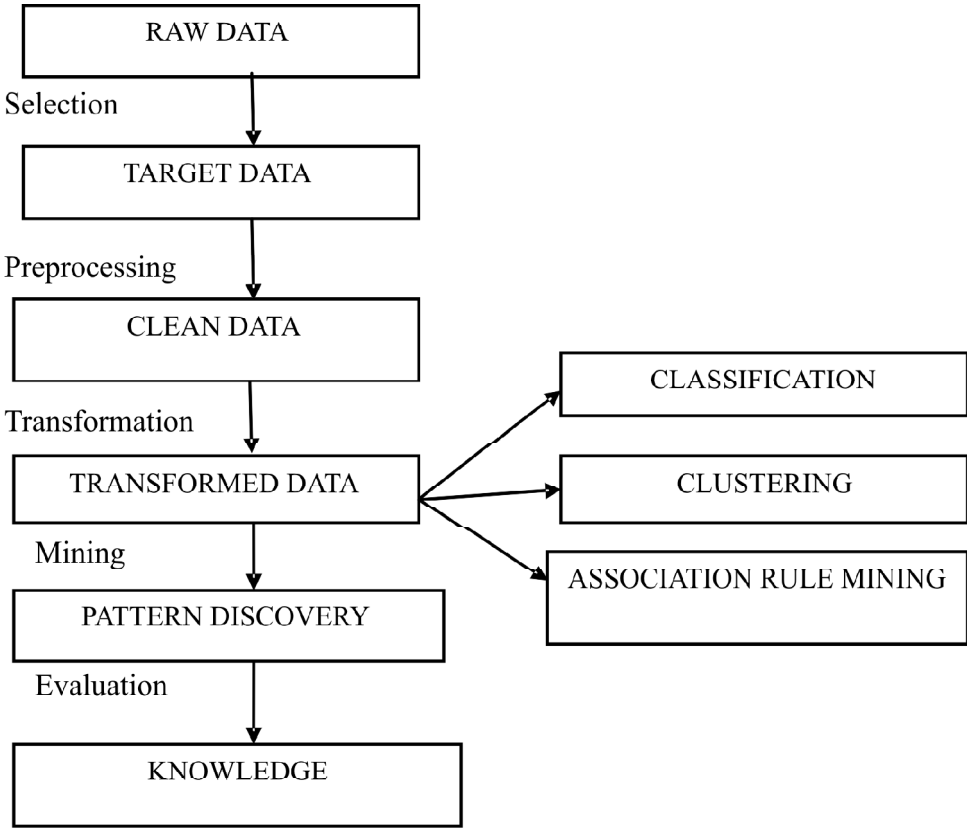

Figure I: Stages in crime analysis

Clustering

Cluster analysis is the business of a collection of patterns (usually represented as a vector of measurements, or a point in a multidimensional space) into clusters based on similarity. In clustering, patterns within a valid cluster are more similar to each other than they are to a pattern belonging to a different cluster.

Cluster analysis is performed in three main steps,

- Feature extraction – Transformation of raw data into useful feature.
- Pattern proximity – Similarity measured using distance metrics.
- Grouping – Cluster formation using clustering algorithms.
- Distance-based clustering-Two or more objects belong to the same cluster if they are close according to a given distance.

Crime cluster analysis is used to identify areas with higher incidences of particular types of crime. Crime patterns Manage law enforcement resources effectively

In this research the two popular techniques used are

- K Means
- Improved KMeans

Crime cluster analysis is used to identify areas with higher incidences of particular types of crime and Crime patterns like serial-rapist or a serial killer. These help to manage law enforcement resources more effectively and solving one of crime results in solving all cases related to the crime pattern. In this research the two popular techniques, K Means and IKM are enhanced to improve the task of crime pattern analysis.

Clustering Algorithms

The various clustering algorithms used for grouping data are categorized as partition based, association based, hierarchical, spectral, density based and grid based clustering. In this study, the partition based and density based clustering algorithms are enhanced and combined.

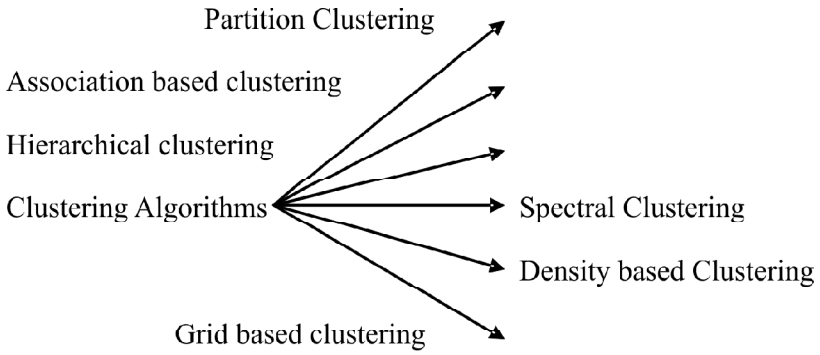

Figure II: Clustering Algorithms

- Given a historical crime database, D, the research problem 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 analyze common crime patterns to reduce further occurrences of similar incidents.
- To develop such a problem the primary objective formulated are,

Primary Objectives are

- To explore and enhance clustering algorithms to identify crime patterns from historical data

Secondary Objectives are

- To enhance Partition based clustering algorithm (KMeans) and Improved KMeans Clustering algorithm (IKM)

  To analyze crime behavior based on previous crime trends.

II. METHODOLOGY

The methodology includes two Segments, the first Segment includes k-means clustering algorithm, the second Segment includes improved KMeans algorithm. Further research work is concentrated on a brief review of,

PHASE 1: Improved K-means clustering algorithm:

Improved K-means clustering algorithm is concentrated by comparison of KMeans clustering algorithm. Accuracy and speed is calculated based on the distance between each cluster, k value prediction.

Accuracy is formulated as

$$Accuracy = \frac{\max\left(\sum_{Cak, L_{sm}} T(C_a, L_s)\right)}{m} x100$$

Where m is the number of data points C_{ak} denotes the k[th] cluster and L_{sm} is the m[th] class. $T(C_{ak}, L_{sm})$ is the number of data points that belong to class m that are assigned to cluster k. Accuracy is computed as the maximum sum of $T(C_{ak}, L_{sm})$ for all pairs of clusters and classes and these pairs have no overlaps. Speed of clustering measured in seconds and is the time taken by the clustering algorithms to partition a given input data set. Table 1 presents the accuracy and speed of the proposed KM model and the conventional IKM algorithms.

Step 1: Optimal prediction of k value.

Step 2: Automatic selection of centroid.

Step 3: Reduce distance calculation.

Step 4: Dimensionality reduction.

New distance metric, is calculated by optimal K value prediction. The Optimal k value prediction is calculated by predicting the centroid position. The centroid position is calculated in order to reduce the distance metric and also lessening in dimensionality. The dimensionality reduced dataset using PCA is D'.

The Distance metric is calculated by using a selection of K value prediction. The distance between each data point and clusters is C_i, $i = 1, \ldots \ldots k$ using distance measure. The closest cluster (C_r) is calculated by selecting the centroid position of each cluster. Each closest cluster is stored in the array of clusters, distance between C_j to closest cluster center is C_r.

Input Cluster dataset d with m data points $\{X_i\ i = 1\ldots\ldots m\}$, clusters number k. Output Cluster dataset $C_1, \ldots, C_k$.

Distance measure is computed for two points X_a, X_b distance measure (dm) is given by

$$DM_{ab} = \min\left(\sum_{s=1}^{p} L(q_s, q_{s+1})\right)$$

Each point is assigned to the cluster whose DM of its center to the point is minimized, Recalculate the center of each cluster, If no points change the categories then stop, Else repeat the initialization procedure.

Recalculate Cluster centers, Compute DM_{new} of x_i to new cluster centers until convergence, If DM_{new} is less than or equal to DM_{ab}, then xi belongs to the same cluster j else Compute DM with every other cluster center and assign x_i to the cluster whose DM is a minimum Set Cluster[i] = j and Set Distance[i] = DM_{new}.

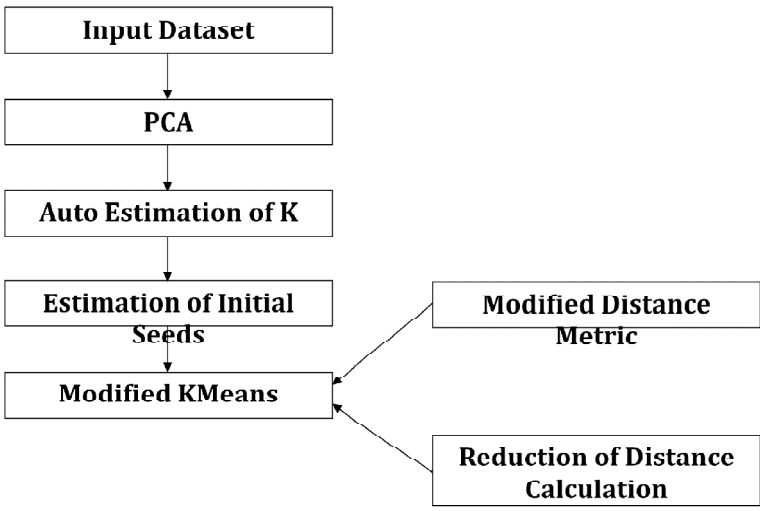

Figure III: Improved K-means algorithm

Calculate dBIC which is the ratio of BIC change, If dBIC (1) <0, then K_T= 1 and calculate inter-cluster ratio, K_T= number of clusters for which the recorded ratio is minimum of all and repeat steps for all KT.

The inter and intra clusters are calculated using inter and intra ratios. Calculate modified inter and intra cluster ratio between cluster C_k and C_{k+1} then Calculate inter-intra ratio.

K_T = Number of clusters for which the dynamic validity index is maximum Optimal k = K_T

III. EXPERIMENTAL RESULTS

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

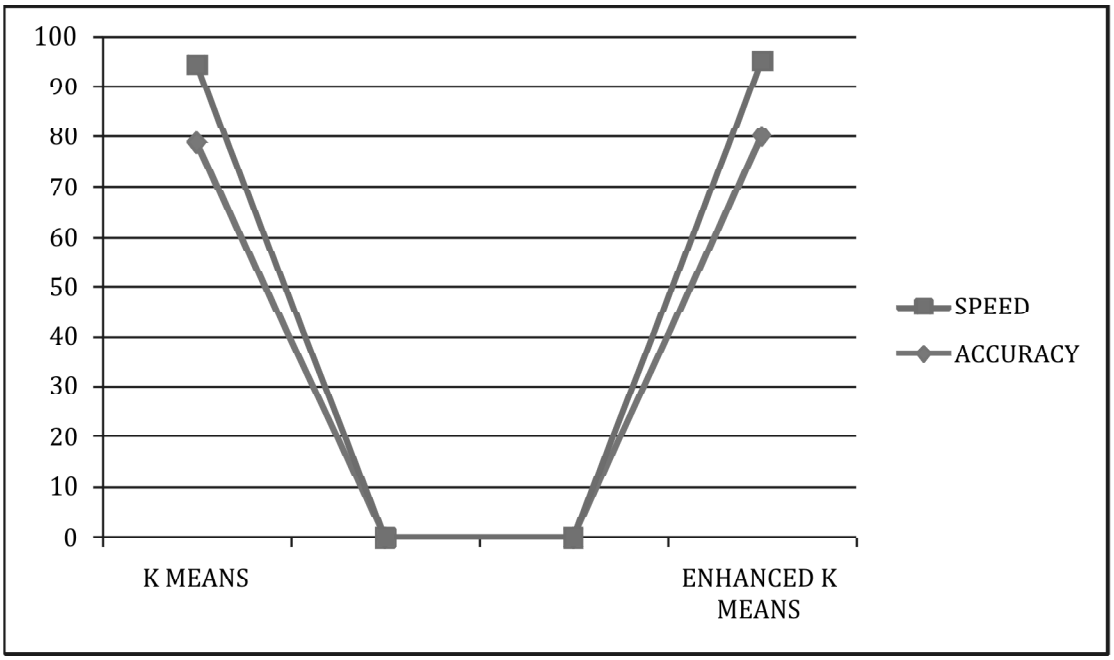

Figure V: Accuracy (%) and Speed (Seconds) of Clustering

The Enhanced Kmeans algorithm showed an efficiency gain of 14.91% and 15.5% respectively over KMeans algorithms. Thus, from the results, it is evident that the proposed model is competent and has enhanced the clustering accuracy and speed of clustering.

Table I
Accuracy (%) and Speed (Seconds) of improved k-means clustering

Algorithm	*Accuracy*	*Speed*
KMeans	78.81	15.57
Enhanced Kmeans	80.16	14.91

IV. CONCLUSION

Data Mining applied in the context of law enforcement, brain analysis and holds the promise of alleviating crime related problem. This study presents a hybrid and ensemble clustering-based algorithm to analyze crime data for identifying crime patterns, crime trends and potential suspects. The proposed algorithm combines improved KMeans and improved KMeans to enhance the clustering process. Experimental results showed that the proposed algorithm is efficient in terms of clustering accuracy and speed of clustering. The result of this data mining could potentially be used to lessen and even prevent crime for the upcoming

years. The results of clustering can be enhanced further through appropriate handling of missing values also the future work can be concentrated in enhanced density based K clustering. Future research is premeditated in this track.

References

[1] David, G. (2006), Globalization and International Security: Have the Rules of the Game Changed?, Annual meeting of the International Studies Association, California, USA, http://www.allacademic.com/meta/ p98627_index.html.

[2] Akpinar, E. and Usul, N. (2005), Geographic Information Systems Technologies in Crime Analysis and Crime Mapping, Pp. 1-12.

[3] Ester, M., Kriegel, H., Sander, J. and Xu, X. (1996), A density-based algorithm for discovering clusters in large spatial databases with noise, Evangelos Simoudis, Jiawei Han, Usama M. Fayyad, Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96). AAAI Press, Pp. 226–231.

[4] Karypis, G., Han, E.H. and Kumar V. (1999), Chameleon: hierarchical clustering using dynamic modeling, IEEE Comput, Vol. 32, Pp. 68–75.

[5] http://www.systemicpeace.org/inscr/inscr.htm, Last Access Date : November, 2013.

[6] Banerjee, A. and Ghosh, J. (2002), On scaling up balanced clustering algorithms, Proceedings of the 2nd SIAM ICDM, 333-349, Arlington, VA.

[7] Chakraborty, S., Nagwani, N.K. and Dey, L. (2011), Performance Comparison of Incremental K-means and Incremental DBSCAN Algorithms, *International Journal of Computer Applications*, Vol. 27, No.11, Pp.14-18.

[8] Erman, J., Arlitt, M. and Mahanti, A. (2006) Traffic Classification Using Clustering Algorithms, Proceedings of the 2006 SIGCOMM workshop on Mining network data, ACM, New York, Pp. 281-286.

[9] CJ Huang, YJ Chen, HM Chen, JJ Jian, SC Tseng… - Applied Soft …, 2014 – Elsevier, ... Data mining; Clustering analysis; Neural networks; Feature selection; Segmentation.

[10] [PDF]A Survey on Crime Data Analysis of Data Mining Using Clustering Techniques A Bharathi, R Shilpa - 2014 - ijarcsms.com

[11] DE Brown, "The regional crime analysis program (RECAP): A Frame work for mining data to ... WA Kosters, J. Laros and JN Kok, "Data mining approaches to criminal career analysis ... Saving...Error saving.TrView as HTMLFewer

[12] [PDF] City Crime Profiling Using Cluster Analysis, P Gera, R Vohra - International Journal of Computer Science & …, 2014 - ijcsit.com.

[13] An improvement of DBSCAN Algorithm to analyze cluster for large datasets, C Dharni, M Bnasal - Innovation and Technology in Education (…, 2013 - ieeexplore.ieee.org)

[14] [C] Enhanced Algorithms to Identify Change in Crime Patterns, A Malathi, SS Baboo - International Journal of Combinatorial …, 2011 - redalyc.org

[15] Evolving Data Mining Algorithms on the Prevailing Crime Trend–An Intelligent Crime Prediction Model, A Malathi, SS Baboo - International Journal of Combinatorial …, 2011 - redalyc.org.