IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 26-30
www.iosrjournals.org

An Effective Data Mining Techniques For Analyzing Crime Patterns

V. Vishnupriya[1], M.C.A., M. Valarmathi[2], M.C.A., M.Phil.,
Research Scholar, Computer Science, Vivekanandha college for women, Tiruchengode, India[1].
Head of the Department, Computer Science, Vivekanandha college for women, Tiruchengode, India[2].

Abstract: *Crime prevention and detection become an important trend in crime and a very challenging to solve crimes. The crime data previously stored from various sources have a tendency to increase steadily. To solve the problems, data mining techniques employ many learning algorithms to extort hidden knowledge from huge volume of data. Data mining is data analyzing techniques to find patterns and trends in crimes. In this propose paper clustering is a data analyzing technique in unsupervised type. This technique is used to divide the same data into the same group and the different data into the other group. For the simple and effective clustering techniques, there are several algorithms such as K-means clustering. This approach is supervised learning scheme that used to dispense objects to one of many pre-determined categories. The algorithms of categorization have been widely applied to the numerous problems that include many various applications. Crime are characterized which change over time and increase continuously. The changing and increasing of crime direct to the issues of understanding the crime behavior, crime predicting, precise detection and managing large volumes of data obtained from various sources.*

Keywords: *Data Mining, KNN Classification, Crime Pattern Analysis, E-Knn Classification.*

I. Introduction

Data mining (sometimes called data or knowledge discovery) is the process of analyze data from different perspectives and summarizing it into useful information - information to facilitate can be used to increase revenue, cuts costs or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many special dimensions or angles, categorize it, and review the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is an interdisciplinary subfield of computer science. It is the computational method of discovering patterns in large data sets relating methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use. Aside from the raw analysis step, it involves database and data management aspects, data pre-processing, model and inference considerations, interestingness metrics, complexity considerations, post-processing of discovered structures, visualization, and online updating. Data mining is the analysis step of the "knowledge discovery in databases" process, or KDD.

The previous studies focus in mining the crime data from the crime database for that the KNN clustering is used for clustering the data. The values are classified by using the crime assessment. Crime values are taken from the crime database. To identify crime falls under category, each crime data is provided with its parametric value. Only clustering of crime data is made and the crime is not been split as per the crime ratio. The proposed system gives the brief reviews of investigates on various implementations of data mining and the strategies to resolve the crimes by using data mining techniques. It also discusses research gaps and confronts in the area of crime data mining. Crime prevention and detection become an imperative trend in crime and a very challenging to solve crimes.

The proposed system provides security for the crime data during outsourcing. Clustering and Classification is made on information. While classifying the data, the watermark content is used. The watermark content is used for verifying the classification data. Based on clustering and classification, the data can be classified and kept secure. Both Clustering and Classification is made on crime data. Data is secure by apply water mark content on data. The crime is been split as per the crime ratio. The knowledge results obtained from data mining processes are used to assist in decision making and to solve the problems.

II. Literature Survey

Arit Thammano [1] describes the most popular clustering algorithm because of its efficiency and superior performance. However, the performance of K-means algorithm depends heavily on the selection of

initial centroids. This paper proposes an extension to the original K-means algorithm enabling it to solve classification problems. First, the entropy concept is employed to adapt the traditional K-means algorithm to be used as a classification technique. Then, to improve the performance of K-means algorithm, a new scheme to select the initial cluster centers is proposed. The proposed models are tested on seven benchmark data sets from the UCI machine learning repository. Data classification is one of the fundamental problems in data mining. Classification, as described, is a process of finding a model that describes and distinguishes data classes, for the purpose of being able to use the model to predict the class of objects which class label is unknown. There are many classification techniques that have been used thus far such as Decision tree, Neural networks, Support vector machines, and Bayesian networks. This paper focuses on a type of classification model that is based on K-means clustering algorithm. K-means is the most popular clustering algorithm. It is very efficient and very easy to implement. Besides being used as a clustering technique, K-means has also been adapted for data classification.

Ying zhao, George karypis [2] describe a fast and high-quality document clustering algorithms play an important role in providing intuitive navigation and browsing mechanisms by organizing large amounts of information into a small number of meaningful clusters. In particular, clustering algorithms that build meaningful hierarchies out of large document collections are ideal tools for their interactive visualization and exploration as they provide data-views that are consistent, predictable, and at different levels of granularity. This paper focuses on document clustering algorithms that build such hierarchical solutions and (i) presents a comprehensive study of partition and agglomerative algorithms that use different criterion functions and merging schemes. (ii) presents a new class of clustering algorithms called constrained agglomerative algorithms, which combine features from both partition and agglomerative approaches that allows them to reduce the early-stage errors made by agglomerative methods and hence improve the quality of clustering solutions.

Chun-Nan Hsu, Han-Shen Huang , Bo-Hou Yang [3] describe the Expectation-Maximization (EM) algorithm is one of the most popular algorithms for data mining from incomplete data. However, when applied to large data sets with a large proportion of missing data, the EM algorithm may converge slowly. The triple jump extrapolation method can effectively accelerate the EM algorithm by substantially reducing the number of iterations required for EM to converge.

III. Crimepattern Analysis

The data mining is data analyzing techniques that used to analyze crime data beforehand stored from various resources to find patterns and trends in crimes. In additional, it can be applied to enlarge efficiency in solving the crimes quicker and also can be applied to automatically advise the crimes. Crime preclusion and revealing become an important trend in crime and a very challenging to solve crimes. Several studies have discovered various techniques to solve the crimes that used too many applications. Such studies can help to speed up the process of solving crime and help the huge data are very difficult and complex.

Objective

- Crime prevention and detection become an important trend in crime and a very challenging to solve crimes.
- The data used for analysis require the accuracy and sufficiency.
- This proposed system focuses on Traffic Violation and Border Control, Violent Crime, the Narcotics, Cyber Crime.
- Issues and challenges on crime are Data Collection and Integration, Crime Pattern, Performance, Visualization.

IV. Data Compressive Paradigm

The experimental application for the crime pattern analysis is designed and implemented with the following modules such are,

4.1. Crime Dataset

Add Crime Profile
View Crime Profile
Crime Observation Details

4.2. Watermark

Add Watermark Content

4.3. Knn For Crime Dataset
Embed Watermark Data in Crime Pattern
Extract Watermark Data in Crime Pattern

K-NEAREST NEIGHBORS ALGORITHM
The k-Nearest Neighbors algorithm (k-NN) is a non-parametric technique utilized for classification and regression. In both cases, the input consists of the k closest training exemplar in the characteristic space.

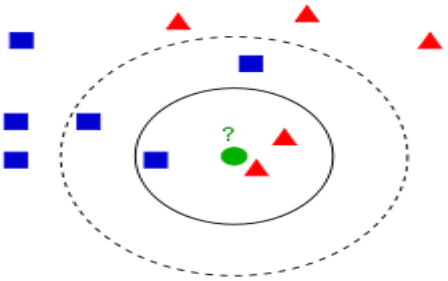

Fig 4.1 *k*-NN classification

In *k*-NN regression, the *k*-NN algorithm is used for estimating continuous variables. One such algorithm uses a weighted average of the *k* nearest neighbors, weighted by the inverse of their distance. This algorithm works as follows:

1. Compute the Euclidean distance from the query example to the labeled examples.
2. Order the labeled examples by increasing distance.

K-NN Algorithm
Input: Crime Data, Watermark Data
Output: Modified Crime Observation Data

1. Add the Crime Profiles (P).
2. Add the Crime Observation Data (O).
3. Enter watermark content (W).
4. Convert the watermark data to bytes and find the length of watermark data (L).
5. Sort the Crime Observation Data (O) Crime wise.
6. I=0
7. For Each Crime's Observation Set in (O)
8. Alter the Observation Data's third value such that OD(3) = 301 + W(I)
9. Change the OD(1) position = OD(1) position + W(I)
10. I=I+1
11. If I>=L Then
12. Break
13. End If
14. Next
15. Output the New Crime Data Set.

Input: Modified Crime Observation Data
Output: Crime Observation Data, Extracted Watermarked Data

1. Select the Crime Data Set (where Watermark Data Embedded) (P).
2. I=0;
3. For Each Crime's Observation Set in (O)
4. W(I) = Observation Data's third value - 301
5. Change the OD(3) value= OD(3) value -301
6. If I=0 Then
7. L= W(I)
8. End If
9. I=I+1
10. If I >L Then

11. Break
12. End If
13. Next
14. Convert the watermark bytes to data.
15. Check the KNN Property.
16. Output Watermark Data.

V. Performances Analysis

The following **Table 5.1** describes experimental result for existing system analysis in patient data set plotted in graph using KNN mining algorithm. In this table is describing the class label surrounded in crime data set in KNN algorithm work.

S. No.	Data Point in X Position	Data Point in Y Position	Class Label
1	75	50	A
2	200	250	B
3	344	300	C
4	180	220	A
5	256	564	A
6	300	200	C
7	200	240	B
8	493	330	B
9	300	340	A
10	300	340	C
11	393	330	B
12	420	480	A
13	93	330	C
14	350	250	B
15	293	330	A
16	564	343	C
17	593	330	B
18	120	120	C
19	420	220	A
20	220	160	B
21	393	430	A

Table 5.1 Class Label- KNN Algorithm without Water Marking

Fig 5.1 [a] and [b] is describing the crime dataset analysis in KNN and KNN with water marking algorithm process. The figure contains data point X position value Y position values and its crime data set. The comparison chart is effective water marking process compare to proposed system.

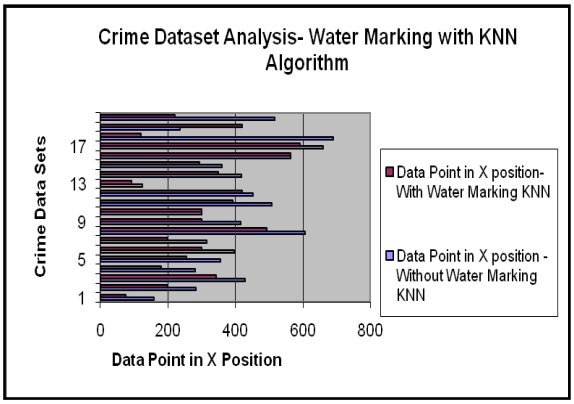

Fig 5.1.a Crime Data Analysis of KNN without Watermarking

The following **Fig 5.2** describes experimental result for comparison of existing and proposed system analysis in crime data set plotted in graph using K-Means and KNN with water marking algorithm. The figure contains the crime id its data point in surrounded data point and its class label details are shown.

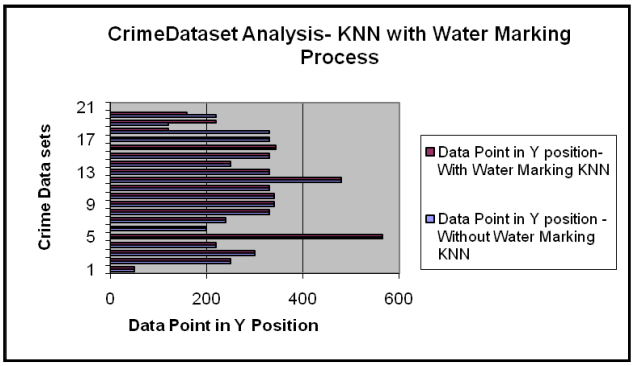

Fig 5.1.b Crime Data Analysis of KNN with Watermarking

VI. Conclusion And Future Enhancement

Crime are characterized which change over time and increase continuously. The changing and increasing of crime lead to the issues of understanding the crime behavior, crime predicting, precise detection, and managing large volumes of data obtained from various sources. Research interests have tried to solve these issues. In the crime investigation procedures, input data is very essential to use in training process and testing process. The training process is used to accomplish the crime model and the testing process is used to validate the algorithm. The issues of crime pattern are concerning with finding and predicting the hidden crime. The proposed methodology provides security for the crime data during outsourcing. Clustering and classification is made on the crime information. While classifying the crime data, watermark content is added for the purpose of defense. The watermark content is used for verifying the classification data. Based on clustering and classification, the data can be classified and kept secured manner. Also the crime data is been split as per the crime ratio.

The proposed algorithm represents the feasible approach to using hubness for improving high-dimensional crime data clustering. And also have it in mind to explore other closely related research directions, including kernel mappings and shared-neighbor clustering for the crime dataset. This would allow us to overcome the foremost drawback of the proposed method detecting only hyperspherical clusters, just as K-Means. In addition, explore methods for using hubs to automatically resolve the number of clusters in the crime data.

Acknowledgements

My heartfelt gratitude goes to my beloved guide Mrs.M.Valarmathi,M.C.A.,M.Phil., Head Of the Department, Department of Computer Science, Vivekanandha College for Women, Tiruchengode, India for dedication and patience in assigning me her valuable advice and efforts during the course of my studies.

References

[1]. J. Han and M. Kamber, Data Mining: Concepts and Techniques, second ed. Morgan Kaufmann, 2006.
[2]. C.C. Aggarwal and P.S. Yu, "Finding Generalized Projected Clusters in High Dimensional Spaces," Proc. 26th ACM SIGMOD Int'l Conf. Management of Data, pp. 70-81, 2000.
[3]. K. Kailing, H.-P. Kriegel, P. Kro ¨ger, and S. Wanka, "Ranking Interesting Subspaces for Clustering High Dimensional Data," Proc. Seventh European Conf. Principles and Practice of Knowledge Discovery in Databases (PKDD), pp. 241-252, 2003.
[4]. K. Kailing, H.-P. Kriegel, and P. Kro ¨ger, "Density-Connected Subspace Clustering for High-Dimensional Data," Proc. Fourth SIAM Int'l Conf. Data Mining (SDM), pp. 246-257, 2004.
[5]. E. Mu ¨ ller, S. Gu ¨nnemann, I. Assent, and T. Seidl, "Evaluating Clustering in Subspace Projections of High Dimensional Data," Proc. VLDB Endowment, vol. 2, pp. 1270-1281, 2009.
[6]. E. Agirre, D. Martı´nez, O.L. de Lacalle, and A. Soroa, "Two Graph-Based Algorithms for State-of-the-Art WSD,"Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP), pp. 585-593, 2006.
[7]. K. Ning, H. Ng, S. Srihari, H. Leong, and A. Nesvizhskii, "Examination of the Relationship between Essential Genes in PPI Network and Hub Proteins in Reverse Nearest Neighbor Topology,"BMC Bioinformatics,vol. 11, pp. 1-14, 2010.
[8]. D. Arthur and S. Vassilvitskii, "K-Means++: The Advantages of Careful Seeding,"Proc. 18th Ann. ACM-SIAM Symp. Discrete Algorithms (SODA),pp. 1027-1035, 2007.
[9]. I.S. Dhillon, Y. Guan, and B. Kulis, "Kernel k-Means: Spectral Clustering and Normalized Cuts,"Proc. 10th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining,pp. 551-556, 2004.
[10]. T.N. Tran, R. Wehrens, and L.M.C. Buydens, "Knn Density-Based Clustering for High Dimensional Multispectral Images,"Proc. Second GRSS/ISPRS Joint Workshop Remote Sensing and Data Fusion over Urban Areas,pp. 147-151, 2003.