



A novel approach for Location Based Crime Analysis Using Data Mining

¹Sk.Wasim Akram, ²P. Bala Viswanadh, ³R. V. Kedar, ⁴P. Siva Sai Teja, ⁵P. Rahul Kumar

¹Asst. professor, ²U.G Student, ³U.G Student, ⁴U.G Student, ⁵U.G Student

¹Department of Computer Science and Engineering,

¹Vasireddy Venkatadri Institute of Technology, Guntur, India

Abstract: In today's era, there is a large amount of increase in the crime rate due to research gap between technologies and the optimal usage of Investigation. Identifying and analyzing the patterns for crime prevention is one of the big challenge. Also, due to some technological limitations, having large amount of data it is difficult to analyze crimes. The goal of this paper is to propose employment of algorithms that works efficiently on large amount of data. This paper is concentrated on crime prevention by concerning various incidents occurred in various states.

Index Terms - Component, formatting, style, styling, insert.

I. INTRODUCTION

In the current scenario there is no separate interface to analyze crime in a particular location for higher officials in the police department. Our main motto is to provide such an interface which gives the crime statistics of a particular location to the higher officials. The development of a cellular network technologies is another main issue which makes mobile phone technology a viable for fighting crime. Nowadays more than 80% percent of population uses mobile phones. Thus, the data from the mobile can be easily cracked by the police. The information may contain the details of victim such as mobile number, location and identification details. This data is most useful for the police to kick off further investigations. Generally, such details are stored in every state and the data is humongous. Such large amount of data is hard to analyze and start the investigation. In some police stations there is no such interface to perform analysis. Our paper gives an idea of our proposed efficient algorithms that help police to analyze the crime aside of having large data.

II. EXTRACTING BEHAVIORAL PATTERNS

^[1]To extract the behavioral patterns from the data that is derived from the formal description of human activity (in particular, the data on locations and calls), it is necessary to have a computational model of participating (interacting) mobiles that simulates their behavior. We must reproduce the reasoning of wireless subscribers to hypothesize on their movements and calls to judge on the possibility of a criminal behavior for a selected group.

Having obtained the behavioral patterns as results of supervised machine learning, it is possible to apply them to location and call data in real time. An algorithm is developed for a particular location of about 100m radius and various Classification, Clustering and Filtering methods are applied.

III. DATA MINING OVERVIEW

IV. Data mining is the extraction of hidden predictive information from a large database. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Data mining is a strong tool because it can provide you with relevant information that you can use to our own advantage.

Basic steps in Data mining:

1. Data Collection
2. Data cleaning
3. Data Analysis
4. Interpretation

We review the Crime Data Mining in two directions ^[2]

1. Crime Types and security concerns
2. Crime Data Mining Approaches used for crime prevention

3.1. Crime Types and security concerns

Crime is defined as “an act or the commission of an act that is forbidden, or the omission of a duty that is commanded by a public law and that makes the offender liable to punishment by that law”.

The following are the different types of crimes

1. Property crime
2. Violent Crime
3. Crime against Women and Child
4. Traffic Violations
5. Terrorist’s attacks and Other Offenses.

3.2. Data Used

The recorded crime data is used for the study (based on mobile and other digital traces) as well as information collected during the investigation process.

The collected data set provides the following:^[6]

1. Location of Crime
2. Street address
3. Time of the crime
4. Type of Crime
5. Geographical information (x, y) coordinates within a range of radius “r” (say 100m).

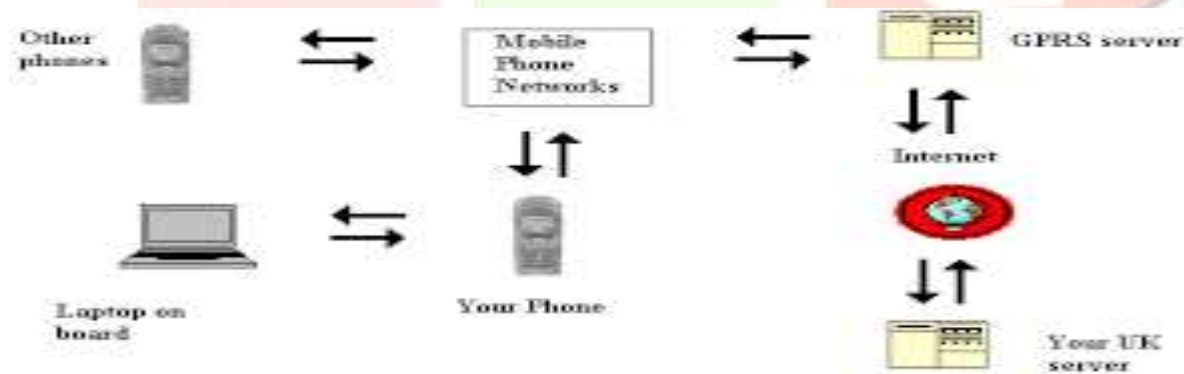
IV. APPROACHES USED FOR CRIME PREVENTION

4.1 FIRST PREVENTION APPROACH

Let us consider a case if someone is in danger and he or she is not able to call. He or she may send a missed call to an emergency number or SOS which will activate and would send live location to the family and to the concerned nearest police station asking for help which is useful to track the person and prevent crime before happening. The diagrammatic representation can be illustrated in Fig. as follows:

below Fig, contains:

1. Phone or mobile
2. network connection or internet
3. GPRS server
4. Self mobile server



The above figure depicts Pictorial representation to locate mobile upon getting a missed caller call, the algorithm will start running on a GPRS server and sends the alert call or message to registered mobile numbers. It is performed using a network among the mobiles. This helps to trace the mobile location based on recent activity from the phone.

4.2 SECOND PREVENTION APPROACH

Suppose any criminal activity is reported then the report can be based on any of the following types:

1. Suppose, the criminal name is known, the criminal data can be retrieved from database easily. Police can get other past records at the single click of one command by proper administrator.
2. Using the previous records, criminals list can be data mined and using predictive analysis; list is filtered with similar criminals. This helps to trace the criminal and can be easily caught by direct investigations.
3. Similarly, a criminal is often intended to do crime in similar pattern. So, it may prove useful to make use of this approach to filter criminals which will help police to solve the crime.

V. PROPOSED METHODOLOGY

The main aim of this research work consist of developing analytical data mining methods that can systematically address the complex problem related to various form of crime. The following illustrates the steps which can be followed to detect the Criminal:

1. Location of crime
2. Type of crime
3. State of crime
4. Victim gender
5. Mobile number of criminal
6. Date of crime
7. Time of crime

Below shows SQLqueryfor creating a new table of interest named Details from the huge database table named as CRIME_DB:

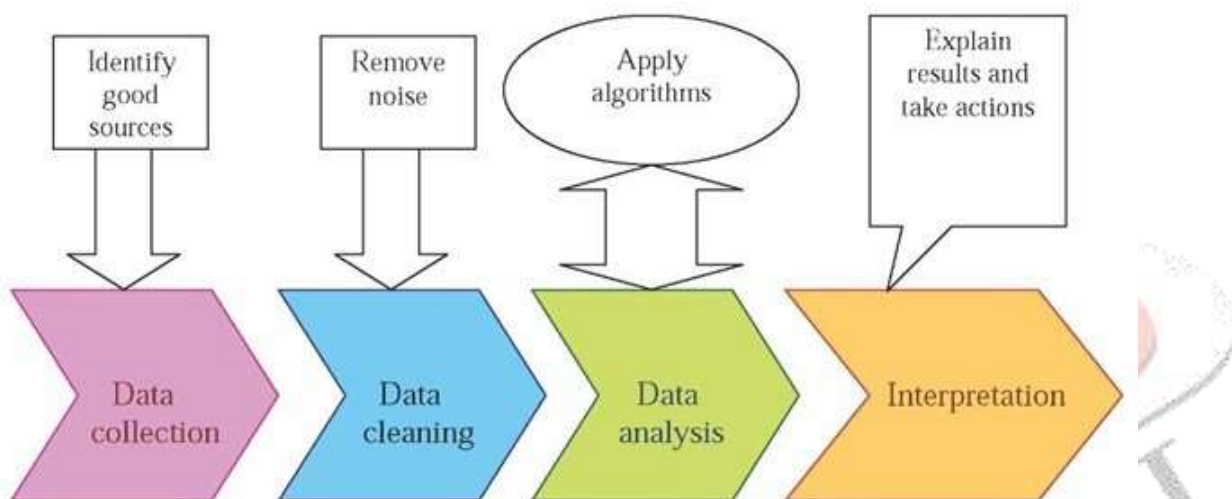
```
CREATE table Details as SELECT mobile, location, active, date FROM CRIME_DB;
```

Run Data mining Algorithms to fetch location of the mobile users and to filternew numbers with unusual behaviors which were recently active in that specific location among the existing old mobile numbers.

```
CREATE table New_Details as SELECT mobile, location, active, date FROM CRIME_DB GROUP By mobile HAVING location='Loc_name';
```

Various algorithms that can be used are:

1. Partition based algorithm (K Mean)
2. Density based algorithm
3. SVM clustering algorithm



The above figure depicts the the methodology of algorithm that work to mine data.

1. Data collection: This method is used to find out the data that are required to mine and use to analyze. The data must be relevant so that the mining is done efficiently.
2. Data Cleaning: The noise in the data must be removed. The data such as duplicate or incomplete data are defined as noise data. This data effects in the working of algorithm.
3. Analysis: Our paper is suggesting three algorithms that are assumed efficient. They are; K-means, SVM and DBSCAN algorithms. These algorithms are efficient by handling large data.
4. Interpretation : It is the result that can be represented as graphs or other notations by complete working of previous methodologies.

A. Data Collection

Enormous amount of crime data is collected at the end of year at police records. This data is made available by National Crime Bureau of Records. This data is in the form of number of cases recorded all over the nation throughout the year. The data is in raw form and also contains some wrong as well as missing values. Hence preprocessing of data becomes very necessary in order to bring the data in proper and clean form. Preprocessing of data includes data cleansing and Pre-Processing.

B. Data Classification

We classify the data set into various groups based on certain characteristics of the data object here we group crimes according to states & cities. Classification of the crime is done on the basis of different types of crime. K means algorithm can be used to group data with similar characteristics .

C. Data Clustering

For our experimental project we are using three types of algorithms that are assumed to work efficiently: K-means, SVM clustering, DBSCAN. Details of each category of algorithms are given below:

a) *K-Means clustering algorithm* is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed apriori. The main idea is to define k centers, one for each cluster. These centers should be placed in a cunning way because of different location causes different result. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest center. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate k new centroids as barycenter of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new center. A loop has been generated. As a result of this loop we may notice that the k centers change their location step by step until no more changes are done or in other words centers do not move any more. Finally, this algorithm aims at minimizing an objective function know as squared error function given by:

$$J(V) = \sum_{i=1}^c \sum_{j=1}^{c_i} (\|x_i - v_j\|)^2$$

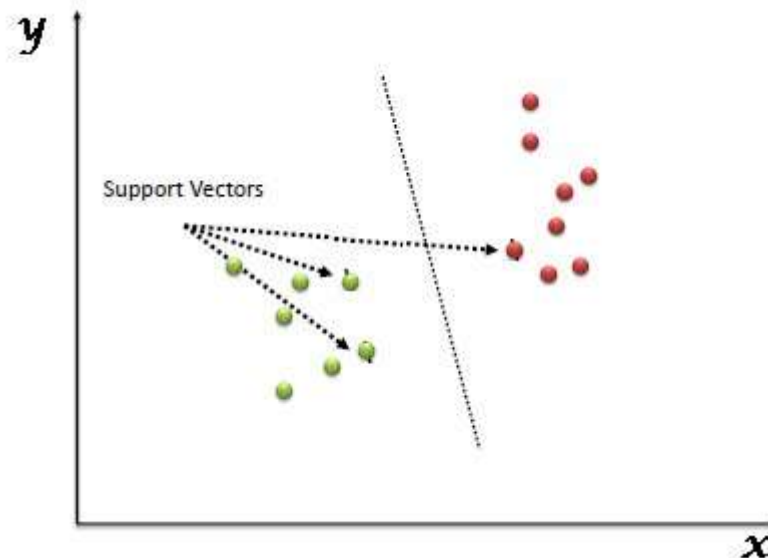
where,

' $\|x_i - v_j\|$ ' is the Euclidean distance between x_i and v_j .

' c_i ' is the number of data points in i th cluster.

' c ' is the number of cluster centers.

b) "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.



above picture depicts the working of SVM clustering algorithm.

Steps involved in algorithm:

1. Define an optimal hyperplane: maximize margin
2. Extend the above definition for non-linearly separable problems: have a penalty term for misclassifications.
3. Map data to high dimensional space where it is easier to classify with linear decision surfaces: reformulate problem so that data is mapped implicitly to this space.

c) *Density based clustering algorithm* has played a vital role in finding non linear shapes structure based on the density. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is most widely used density based algorithm. It uses the concept of **density reachability** and **density connectivity**.

Density Reachability-A point "p" is said to be density reachable from a point "q" if point "p" is within ϵ distance from point "q" and "q" has sufficient number of points in its neighbors which are within distance ϵ .

Density Connectivity- A point "p" and "q" are said to be density connected if there exist a point "r" which has sufficient number of points in its neighbors and both the points "p" and "q" are within the ϵ distance. This is chaining process. So, if "q" is neighbor of "r", "r" is neighbor of "s", "s" is neighbor of "t" which in turn is neighbor of "p" implies that "q" is neighbor of "p".

Algorithmic steps for DBSCAN clustering

Let $X = \{x_1, x_2, x_3, \dots, x_n\}$ be the set of data points. DBSCAN requires two parameters: ϵ (eps) and the minimum number of points required to form a cluster (minPts).

- 1) Start with an arbitrary starting point that has not been visited.
- 2) Extract the neighborhood of this point using ϵ (All points which are within the ϵ distance are neighborhood).
- 3) If there are sufficient neighborhood around this point then clustering process starts and point is marked as visited else this point is labeled as noise (Later this point can become the part of the cluster).
- 4) If a point is found to be a part of the cluster then its ϵ neighborhood is also the part of the cluster and the above procedure from step 2 is repeated for all ϵ neighborhood points. This is repeated until all points in the cluster is determined.
- 5) A new unvisited point is retrieved and processed, leading to the discovery of a further cluster or noise.
- 6) This process continues until all points are marked as visited.

D. Correlating Crime

Many crimes are related to other crime or criminal. Finding this correlation can be of great help in finding missing clues. Correlations can be used to help make predictions. If two variables have been known in the past to correlate, then we can assume they will continue to correlate in the future. We can use the value of one variable that is known now to predict the value that the other variable will take on in the future. Pearson's correlation coefficient between two variables is defined as the covariance of the two variables divided by the product of their standard deviations.

E. Predicting Crime

Prediction of crime is a great aid to the administration in order to curb the crime incidences. Prediction is stating the probability of occurrence of any similar crimes in future period time.

VI. RESULT

The results of this data mining could potentially be used to prevent crime for the forth coming years. From the efficient yielding results, we believe that mining crime data will have a promising future for increasing the effectiveness and efficiency of criminal and intelligence analysis. Many future directions can be explored in this still young domain. Visual and intuitive criminal and intelligence investigation techniques can be developed for any crime pattern.

VII. APPLICATIONS

The above proposed approach is explained in a simpler format but it can be put to use in tracking many different analysis. Firstly to predict the crimes from huge data set available. Secondly, it can replace the exhaustive GPS technology for monitoring the location of vehicles by transport agencies. The approach would prove to be cost effective since it does not require much energy in transmitting large bandwidth signals as in the case of GPS technology and since less bandwidth is used so lesser traffic and hence cheaper means of locating vehicles. Also, as already mentioned, it can be put to use by the police department to keep track of criminals and by using predictive analysis approach, provide a list of criminals to the police to make their hard work easy in a sense to crack the case. Moreover, it can help to prevent from many future crimes to happen since it can help the police to track the victim

VIII. CONCLUSION

This suggested a new way of extracting criminal behavior, mining the data of location-based services. The aim is to provide with the setting up of infrastructure and basic approach to help prevent crime using information technology. We presented the types of crime and how it can be handled by data mining tool. By using this it is very useful to detect crime patterns and analyze crime analysis.

IX. REFERENCES

- [1] A. Malathi, S.S. Babboo and A. Anbarasi, "An intelligent analysis of a city crime data using data mining", In International conference information electronic engineering, Vol. 6, Pp. 130-134, 2011.
- [2] H. Chen, W. Chung, J.J. Xu, G. Wang, Y. Qin and M. Chau, "Crime data mining: a general framework and some examples", Computer, Vol. 37, No. 4, Pp. 50-56, 2004.
- [3] S. Syal and Dr.M. Singh, "Prevent Crime Using Mobile- A Practical Approach", Vol.4, 2014.
- [4] L. McClendon and N. Meghanathan, "Using Machine Learning Algorithms to Analyze Crime Data", Machine Learning and Applications: An International Journal (MLAIJ), Vol, 2.
- [5] J.A. Adeyiga and A.O. Bello, "A Review of Different Clustering Techniques in Criminal Profiling", Vol. 6, No. 4, 2016.
- [6] M. Ramzan Begam, "Survey: Tools and Techniques implemented in Crime Data Sets", IJISSET, Vol. 2, No. 6, 2015.
- [7] Sukanya, "Criminals and Crime Hotspot Detection Using Data Mining algorithms: Clustering and Classification",.
- [8] Vineet Pande, "Crime Detection Using Data Mining", IJERT, Vol. 5, 2016.