Location Based Crime Detection Using Data Mining

Dr. Ishwar Baidari and S.P. Sajjan

Abstract--- A huge amount of data set is generated every year on the basis of reporting of crime. This data can prove very useful in analyzing and predicting crime and help us to prevent the crime to some extent. Goal is to detect crime in a particular location using data mining techniques applied on mobile devices that were found during the occurrence of crime. Here we consider that each individual can be tracked by locating their mobile numbers and its digital traces. This approach finds usage not only in fighting against crime, but also effectively replace the current conventional methods, which in turn will save lots of time and money of the Government and other intelligent agencies. There had been an enormous increase in the crime in the recent years. The concern about national security has increased significantly since the 26/11 attacks at Mumbai, India. However, information and technology overload hinders the effective analysis of criminal and terrorist activities.

Keywords--- Crime Detection, Location Based Systems, Data Mining, Digital traces, Database, K Mean.

I. INTRODUCTION

Today’s mobile handheld devices have become a feasible medium for carrying out an extensive variety of actions, involving information exchange. One of the important area in society where mobile technology is gaining usage is crime prevention, crime detection and reporting. The high computational power of smart phones, tablets and PDAs accounts for high demand and usage by the general public. The development of cellular networking technologies is another main issue which makes mobile phone technology a viable mode for fighting crime.

Nowadays, more than 80% percent of population uses mobile phones. This form of telecommunication service is heavily dependent on the accurate determination of the handset locations to promptly switch from one service station to another. Telecommunication servers accumulate huge amount of data that includes the recording of locations of handsets at certain time intervals. Also, the phone numbers of both callers and called addressees are recorded. Such data is a tremendous resource of information that can be used for a wide variety of purposes, including, in particular, targeted advertisement, based on accumulated customer profiles.

Currently, to the best of our knowledge, this kind of data is not used to detect suspicious forms of behavior. A special data mining and reasoning technology is required to automatically process the location data as its manual browsing is unaffordable. The lack of adequate technology leads to disregarding a rich body of data of potential value in filtering out criminal behavior. Crimes might be prevented and networks of criminals groups with peculiar inter-connections identified if it were possible to discover sets of unusual patterns of coordinated movement for groups of mobile phones.

II. EXTRACTING THE BEHAVIOR PATTERNS

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

We review the Crime Data Mining in two directions

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

A. 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 for punishment by that law”. An act of crime encompasses a wide range of activities, ranging from simple violation of civic duties (e.g., illegal parking) to internationally organized crimes (e.g., the 26/11 attacks).

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.

Dr. Ishwar Baidari, Currently Working as a Associate Professor in the Dept. of Computer Science.

S.P. Sajjan, Currently working as Teaching Assistant Professor (Contractual) in the Dept. of Computer Science, Central University, Karnataka.

DOI:10.9756/BIJSESC.8279
B. 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:

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

A. First Prevention Approach

Consider the case if someone is in danger and he or she is not able to call. He or she may send a miss call to an emergency number which will activate and would send constant location to the near ones and to the concerned nearest police station asking for help which may prove useful to track the person and prevent from any happening. The diagrammatic representation can be illustrated in Fig.2 as follows:

Above Fig1, contains:
1) Network Towers
2) Victim Mobile Phone
3) Server running to locate phone.

Fig. 1 depicts Pictorial representation to locate mobile upon getting a missed caller call, the algorithm will start running on a central server and it will send a location request for the mobile phone. After receiving the location, the server will transfer the message to three registered numbers which is stored in the database respective to the caller number with an SOS message along with the current location. The location will get updated after every 3 to 5 minutes and new message will be send. Moreover, the application can be enhanced with the feature of sending message to the nearest police station or to the nearest police control room to take quick action. This will help the police to prevent crime.

B. Second Prevention Approach

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

1) If the criminal name is known, his or her current location can be obtained directly from the database together all other past records at the single click of one command by proper administrator.
2) With the past records, criminals can be data mined and using predictive analysis; list of fewer criminals can be traced out from which the criminal can be easily caught by direct investigations.
3) Again, a criminal is often motivated to do a similar kind of crime based on research. So, it may prove useful to make use of this approach to list out fewer 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 forms of crime. The following illustrates the steps which can be followed to detect the Criminal:

1) Maintain a database of criminals. By collecting a log or historical Crime data of minimum last 2 to 3 months and record the detail which forms our Dataset.

The Dataset must contain:
1) Locations of multiple mobiles
2) IMEI numbers of multiple mobiles
3) How often calls are made from multiple mobile numbers.
4) Common callers of multiple numbers
5) Minimum duration of call number
6) Minimum duration of Received call number
7) GPS Geo Position of Received & call number
8) All other parameters.

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

CREATE table New_Details as SELECT mobile, location, active, date FROM CRIME_DB:

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

2) Run Data mining Algorithms to fetch location of the mobile users and to filter new 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:

a) Partition based algorithm (K Mean)
b) Density based algorithm
c) Hierarchical algorithm
d) Fuzzy C Mean algorithm
e) Expectation Maximization algorithm.

3) From list of different new mobiles (unusual behavior) obtained, will give us the reduced dataset with most probable criminal mobile number details.

Case I: The set of New or less frequently used mobile users amongst the Set of all users who are continuous residents of particular locality.
SELECT * FROM Details INTERSECT SELECT * FROM New_Details;

Case2: The set of mobiles which were found active during last few days during the Crime i.e. 2 to 3 days. The Set of mobiles recently found Active:

SELECT mobile, location from New_Details ORDER BY date Asc;

Case3: Obtaining all mobile details that were intentionally switched off (disabled) after the Occurrence of Crime.

SELECT * FROM New_Details WHERE active='False' AND date>26/8/16;

Case4: Categorizing based on call durations:

1) Set of mobiles most frequently used:
   SELECT MAX(duration) FROM New_Details;

2) Set of mobiles less frequently used:
   SELECT MIN(duration) FROM New_Details;

3) Fetching a list of criminals in case criminals name is already registered for a similar type crime and forwarding the list to the police to track the criminal.

Following figure shows the methodology used for crime pattern analysis and detection:

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 data 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

a) Hard Clustering

These methods of clustering are based on the classical setting theory, and require that an object either does or does not belong to a cluster. It means partitioning the data into specified number of mutually exclusive subsets. That is, each document belongs to exactly one cluster; it only assigns a value of 1 or 0.

b) K-Means Algorithm

Mainly used to partition the clusters based on their means. Initially number of objects are grouped and specified as “K” clusters. The mean value is calculated as the mean distance between the objects. The relocation iterative technique which is used to improve the partitions by moving objects from one group to other. Then number of iterations is done until the convergence occurs. Steps are:

   Input: Number of clusters.

   Step 1: Arbitrarily choose k objects from a dataset D of N objects as the initial cluster centers.

   Step 2: Reassign each object which distributed to a cluster based on a cluster center which is the most similar or the nearer.

   Step 3: Update the cluster means, i.e. calculate the mean value of the object for each cluster.

   Output: A set of k clusters.

   K-means algorithm is a base for all other clustering algorithms to find the mean values.

   c) Density-Based

This is based on connectivity and density functions. In density-based clustering, clusters are defined as areas of higher density than the remainder of the data set. Objects in these sparse areas that are required to separate clusters - are usually considered to be noise and border points. The most popular density-based clustering method is DBSCAN. In contrast to many newer methods, it features a well-defined cluster model called "density reach ability". Similar to linkage based clustering; it is based on connecting points within certain distance thresholds. However, it only connects points that satisfy density criterion, in the original variant defined as a minimum number of other objects within this radius.

d) Hierarchical Algorithm

Hierarchical clustering builds a cluster hierarchy (a tree of clusters). Create ahierarchical decomposition of the data set (or objects) based on a criteria and uses distance matrix as clustering criteria. This method does not require the number of clusters as an input, but needs a termination condition. Connectivity based clustering, also known as hierarchical clustering, is based on the core idea of objects being more related to nearby objects than to objects farther away. As such, these algorithms connect “objects” to form “clusters” based on their distance. A cluster can be described largely by the maximum distance needed to connect parts of the cluster. At different distances, different clusters will be formed, which can be represented using a dendrogram, which explains where the common name “hierarchical clustering”.

e) Fuzzy Clustering (Soft Clustering)

Fuzzy clustering also called soft clustering. Allow the objects to belong to several clusters simultaneously, with different degrees of membership. It is more natural than hard
clustering because objects on the boundaries between several classes are not forced to fully belong to one of the classes, but rather are assigned membership degrees between 0 and 1 indicating their partial membership. For example, the clusters or groups that are identified will be overlapping, meaning that one instance may fall into several clusters. In fuzzy clustering, we make a fuzzy partition of the data set. This can be applied to crime analysis because police officers commonly receive descriptions of suspects that are fuzzy in nature. Fuzzy logic stipulates that an element can be a member of a given set in an uncertain manner. Unlike classical mathematical set theory in which an element can only be in two situations, a member or not a member of some set; fuzzy logic generalizes these possibilities and introduces the concept of shades of membership patterns.

f) Fuzzy C-Means

Fuzzy C-Means Clustering (FCM) also known as Fuzzy ISODATA is a clustering technique which is separated from hard k-means that employs hard partitioning. The FCM employs fuzzy partitioning such that a data point can belong to all groups with different membership grades between 0 and 1. FCM is an iterative algorithm. The aim of FCM is to find cluster centers (centroids) that minimize a dissimilarity function. This algorithm works by assigning membership to each data point corresponding to each cluster center on the basis of distance between the cluster center and the data point.

g) Expectation Maximization Algorithm

Expectation Maximization (EM) is a model based approach to solve clustering problem. It is an iterative algorithm that is used in problems where data is incomplete or considered incomplete. EM clusters data, in a manner different from k-means. It is known to be an appropriate optimization algorithm for constructing proper statistical Models of the data.

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 incidence. 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 lessen and even prevent crime for the forthcoming years. 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. Many future directions can be explored in this still-young field. Visual and intuitive criminal and intelligence investigation techniques can be developed for crime pattern

VII. APPLICATIONS

The above defined approach is put in simpler format but it can be put to use in tracking many different things. 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 AND FUTURE ENHANCEMENT

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. It reduces further occurrence of similar incidence. It formulates strategies for crime prevention and reduction. Moreover, other branches of study like usage of predictive analysis, artificial intelligence, and big data approach on large data sets can be put together to enhance the crime prevention system. Also, above approaches can be put in different applications like vehicle location detection in transport industries.

REFERENCES


AUTHORS PROFILE

Dr. Ishwar Baidari, pursued Bachelor of Science from Karnatak University, Dharwad and awarded Ph.D degree from the same University, currently working as an Associate Professor in the Dept.of computer Science. His main research work focuses on Graph Algorithms and Big data Analytics. He has more than 10 years of experience.

S.P. Sajjan, Currently working as teaching assistant professor(Contractual) in the Dept of Computer Scince, Central University, Karnatak. Mr. S P Sajjan obtain his post graduation(M.Sc) from Karnatak University, Dharwad, He has Four year Experience in Teaching and Industry, He has worked three years as a Teaching assistant in Karnatak University, Dharwad.