Enabling Real Time Crime Intelligence Using Mobile GIS and Prediction Methods

M. Saravanan
Ericsson Research India
Chennai, India
m.saravanan@ericsson.com

Rakhi Thayyil and Shwetha Narayanan
Dept. of Information Technology
SSN College of Engineering
Chennai, India
rakhi3192@gmail.com,shwethanyn@gmail.com

Abstract—A crime investigation is an official effort to uncover information about a crime. In recent years the number of crime cases has been on a rise. The traditional and age-old system of intelligence and criminal record maintenance has failed to live up to the requirements of the existing crime scenario. In this paper we propose a swift response system which can identify the most probable local suspects involved in a crime case, by analyzing the relevant case histories. We have looked into the mobile call detail records of suspects and victims to understand their presence in crime scenario. Records of cell tower near crime scene have been analyzed to track the real perpetrators. With the knowledge of suspect’s journey to crime and about the movements of people in the crime scene, we are able to model the system and to understand the probable suspects involved in the crime. Prediction techniques are used to filter and identify the different types of people present at the crime scene. To solve the case at a rapid pace, we have mapped the current location of the probable suspects using Mobile GIS. We have also evaluated the methods used in this system in comparison with traditional methods.

Keywords—Crime Intelligence; Journey to Crime; Conditional Random Fields; K-Means clustering; Particle Swarm Optimization; Mobile GIS.

I INTRODUCTION

An essential aspect of criminal investigation involves the analysis of the databases held by the investigation department and criminal agencies. The historical features present in this voluminous amount of data on analysis reveals hidden crime patterns and help in assisting human inference in solving a crime. Application of data mining techniques helps in predicting and supporting the investigation of crime, thereby boosting the efforts of the investigation officer to solve the case in a much faster way.

The fundamental part of a criminal investigation is in finding out the probable suspects involved in a crime. This can be further enhanced by linking the crimes which have been committed in the past and by the nature of crime. However, principles of Crime Pattern Theory states that an offender’s direction of travel to a criminal event coincides with paths he or she frequently takes on a routine basis. Thus, although it may appear as though crimes in these areas are haphazardly distributed, the tenets of Crime Pattern Theory suggest that an underlying pattern should be present [1].

With the increase in usage of mobile phones as a means of communication among the people and also among those involved in committing a crime, there is a possibility that large number of crimes can be traced by analyzing the mobile footprints. Here the footprint refers to the individual’s call detail records and Spatio-temporal aspects of the cell tower details. Owing to communication becoming effortless, most of the suspects recently use mobile phones as a means to communicate with their communities on how best to commit a crime. These factors have motivated us to explore the use of mobile phone records in addition to customary crime records as a means to provide a way to solve a crime in an effective and fast manner.

In this study we have developed a novel soft system methodology (SSM) to predict the most probable suspects involved in a crime. This methodology also helps to automate the system that can be applied to various scenarios. Also with the use of Mobile GIS, we can analyze suspect’s journey to crime by analyzing the cell tower records of the crime location. Call detail records of the victim’s phone are analyzed to understand their recent contacts.

II RELATED STUDIES

An intelligent analysis of cybercrime spatial data [7] utilizes the location attributes of the states and union territories in India, including the incidents of various cyber crimes and the number of persons arrested. Geographical Information System (GIS) is used to capture and analyze cyber crime spatial data.

One of the studies involves reviewing the crime data mining techniques and presents four case studies related to an ongoing project [9]. The first step performs Entity extraction followed by clustering techniques. Classification is used to detect email spamming [9]. Social network analysis has been used to analyze criminals’ roles and associations [9].

In an existing study, a method was developed that would reconstruct the most likely routes taken by offenders to their crime locations, within real city networks [4]. By using Dijkstra’s algorithm, and taking into account key locations tied with a criminal event, CriMM [4] was used to analyze the occurrence of crime in crime neutral areas displaying results consistent with Crime Pattern Theory.

We understand from the above papers that most of the existing systems neither provide an end-to-end solution to
solve a crime case nor view the case in multiple dimensions, to arrive at the list of probable suspects. The findings of these insights motivated us to develop a soft system methodology that definitely would help the investigation officer to solve the crime case in a faster manner.

III TOOLS TO BUILD INTELLIGENCE AND PREDICTION IN OUR SYSTEM

Most of the data mining techniques are used as prediction techniques which helps us to accurately and efficiently analyze the large volumes of data. We already found the usage of few of the techniques like clustering, classification, pattern mining, event extraction from textual documents [10], nearest neighbor method etc., [3]. We will give complete overview of tools and methods related to SSM discussed in this study.

A. Feature selection

Feature selection [8] is the process of selecting a subset of terms from the training set and using the selected features for content classification and retrieval. In this study, feature selection is used to identify the crime type, date of occurrence, and the locations involved etc.

B. CRF++

CRF++ [6] is a simple, customizable, and open source implementation of Conditional Random Fields (CRFs) for segmenting/labeling (classifying) sequential data. In this study, CRF++ is trained and applied to identify the features such as Location, Day, Date, Time and the type of crime as to Murder, Kidnap or Robbery.

C. Multi level ranking algorithm

The relevant cases are ranked using a generalized Multi level ranking algorithm. Based on the usage of this algorithm each case is given a weight specifying the relevancy of match to the case. The features used for multi level ranking in the order of precedence are crime scene (location), date of occurrence, day and period of the day.

D. Clustering algorithms

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called cluster) are more similar to each other than to those in other groups [2]. In this paper K-means clustering algorithm [11] is used to cluster the mobile phone users present in the crime location.

E. Particle swarm optimization

In the Particle swarm optimization (PSO) [12] algorithm, each particle belonging to a swarm, has a position and a velocity (their values are randomized initially). The position with the highest fitness score, in each of the iteration, is set to be the entire swarm’s global best (gbest) position, towards which other particles move. In addition, each particle keeps its best position that it has visited, known as the particle’s personal best (pbest). The particle dynamics are governed by the following rules which update particle positions and velocities:

\[ v_i = wv_i + c_1r_1(x_{pbest} - x_i) + c_2r_2(x_{gbest} - x_i) \]  
\[ x_i = x_i + v_i \]

where \( x_i \) is the current position of particle \( i \), \( x_{pbest} \) is the best position attained by particle \( i \), \( x_{gbest} \) is the swarm’s global best position, \( v_i \) is the velocity of particle \( i \), \( w \) is a random inertia weight, \( c_1 \) and \( c_2 \) are spring constants and \( r_1 \) and \( r_2 \) are random numbers between 0 and 1.

In our study the implementation of PSO is slightly modified and applied to the list of suspects based on journey to crime and also to the clusters obtained from K-Means clustering, to find the probable suspects.

IV ENABLING CRIME INTELLIGENCE IN PROPOSED SYSTEM

We have developed a SSM with an intention to provide an end-to-end solution in solving a crime case in an effective and fast manner. Fig 1 gives an overview about our system and its proposed modules. These modules help the Investigation officer to speed up the investigation process by effectively tracking the probable suspects based on their current location.

A. Information retrieval and Feature extraction

The investigation officer keys in a brief description of the crime, containing details about the location, date of occurrence of crime and keywords related to the type of crime as an input. The given input is classified based on the type of crime by applying feature selection technique. Information Retrieval is performed by using Stanford NER [5]. CRF++ is trained to identify the type of crime based on the extracted features and keywords related to the type of crime. The training set for our analysis consists of records of various types of crimes committed across the world which was classified based on Modus Operandi.

Table I shows a description of the features that are extracted from a given input crime case. These features along with the keywords are used to classify a crime case report and store it in the database. The features are selected such that these features can be mapped to any type of crime incident.

B. Retrieval of relevant cases

Based on the features extracted from the given input the relevant case reports are drawn from a database that contains different reports about the different type of crime cases. A consolidated summary of the relevant cases are displayed based on a multi-level ranking algorithm. Each case report is given a ranking based on the algorithm.
TABLE I. DESCRIPTION OF THE FEATURES EXTRACTED

<table>
<thead>
<tr>
<th>FEATURE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Names of people and organisations.</td>
</tr>
<tr>
<td>Location</td>
<td>Location of occurrence of crime.</td>
</tr>
<tr>
<td>Type of crime</td>
<td>Crime types such as Murder, Kidnapping, Robbery etc.</td>
</tr>
<tr>
<td>Date</td>
<td>Date of occurrence of crime.</td>
</tr>
<tr>
<td>Time</td>
<td>Time of occurrence of crime.</td>
</tr>
<tr>
<td>Period of day</td>
<td>Period of the day as to Morning, Afternoon or Evening.</td>
</tr>
<tr>
<td>Day of week</td>
<td>Day of the week when the crime is reported to have occurred.</td>
</tr>
</tbody>
</table>

The generated case summary are then displayed along with the percentage of relevancy to the input crime case. The pseudo code for multi-level ranking algorithm is presented in Fig 2. The retrieval of the relevant cases in a ranked manner helps to find out similarities in modus operandi, crime patterns in that region and nature of the suspects.

C. Mapping of suspects journey to crime

The journey to crime of suspects, who have previously committed crimes in the crime scene are then visualized using Google Maps. Crime Pattern Theory states that an offender’s direction of travel to a criminal event coincides with paths he or she frequently takes on a routine basis. The home and crime locations of all offenders were input from the database containing the information about the suspects and the paths taken by these offenders. The results of the simulation are shown in Fig 3 where links are plotted on Google Map.

D. Analysis of people at the crime scene

The mobile users in the crime scene are then clustered into 3 groups by applying K-Means clustering. The significant features for clustering such as the number of calls made by the caller, caller’s idle time, time and duration of calls etc are arrived at by analyzing the Cell Tower Records of the relevant crime region.

The cell tower records are fetched for a period of 3 months prior to the date of crime by considering the spatio-temporal parameters.

K-Means clustering is implemented in this study as follows,
- The features retrieved from the cell tower records are given as input to this module.
- The number of clusters is initialized to 3 for easy grouping and the cluster centroids are randomly selected.
- Then the distance function is used to form 3 clusters, based on their closeness to the cluster centroid.
- The 3 clusters are labelled as Residents, Passerby and Visitors based on the distinguished features.

E. Analysis of call logs of the victim

The call logs of the victim are fetched and analyzed to identify calls from suspects, frequent callers and anonymous calls. The call logs are fetched for a span of 2 days prior to the time of occurrence of crime. Calls from suspects are identified by comparing the mobile numbers present in the call logs with those of suspects available with the investigation department.

F. Identification of probable suspects using Particle Swarm Optimization

Particle Swarm optimization (PSO) is then applied to produce the final list of probable suspects. Each suspect is assigned a weight along three different axes representing his position. The axes mentioned here are related to the following methods such as Suspects journey to crime, Analysis of people’s movement, Analysis of victims call detail records.

The suspects are assigned a position on each axis based on their presence in any of the methods. The sum of weights on all the three axes gives the current position of the suspect. Features such as frequency of calls made to the victim, number of calls made to the victim recently are also used to calculate the position. The suspects rank is then updated using Eqn. (1) which is based on their position. The particles (suspects) are then ranked based on their closeness to the gbest of the swarm.

Eqn. (1) has been used to rank the probable suspects using the redefined features. Here $s_i$ is the suspects position in the list, $v_i$ is the suspects weight factor which measures of the closeness of the suspect to his involvement in the crime, $x_{\text{gbest}}$ is the suspects best position, $x_{\text{gbest}}$ is the best position obtained for the swarm, $w$ is a random inertia weight with value 0.1, $c_1$ and $c_2$ are spring constants with values 1 and $r_1$ and $r_2$ are random numbers with values 1. The final list of probable suspects is displayed in a ranked manner.
G. Tracking the current location of the probable suspects

The current location of the probable suspects is then tracked by using Mobile GIS. The adjacent cell towers are either grouped or analyzed to figure out the suspect’s exact present location. The current location of the probable suspect is then visualized on Google maps.

V EXPERIMENTAL RESULTS AND DISCUSSIONS

To evaluate the performance of the 3 classifiers, we have used a dataset of 139 crime cases which contains 60 Murder, 48 Kidnap and 31 Robbery cases. We have applied 3-fold cross validation method to verify our results on three different classifiers. Filtering and ranking of suspects done with PSO [12] is also evaluated in relation with traditional ranking algorithm.

A. Applying Classifiers for Crime Cases

For categorizing the given crime cases, we have applied 3 different classifiers. The CRF classifiers [8] efficiency is compared with Naïve Bayes and JRip classifiers [3]. The results are summarized in Table II. It was found that CRF showed a better performance for the categorization of crime classes. Probabilistic estimates of Naïve Bayes tend to show poor performance. JRip Rule learner needs more training set to improve its accuracy. The F-measure value for CRF classifier which almost reaches to 1 clearly indicates the usefulness of the algorithm in the proposed SSM. The prediction of correct crime classes is more important to identify correct and relevant suspects in later stages.

B. Particle swarm optimization

The performance of PSO was evaluated by comparing it with a simplified ranking algorithm. The simplified ranking algorithm performs ranking by analyzing the presence of the suspects, found by using the methods such as Suspects journey to crime, Analysis of people’s movement, and Analysis of victims call detail records. That is, based on the weights (repeated occurrences), it can derive the common score to understand the probable suspects.

For simple comparison, we have analyzed 10 different test cases based on accuracy score. The results of comparison are shown in Table III. Even though, a traditional ranking algorithm gives better results, it fails to classify some of the difficult cases. We found that PSO excellently narrows down the suspects from the three different inputs. It acts as a neutralizer to select the probable criminals, who might be present during the crime scene, by filtering the other not relevant members.

TABLE III. COMPARISON OF PSO WITH SIMPLE RANKING ALGORITHM

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Particle swarm optimization</td>
<td>95.6</td>
</tr>
<tr>
<td>Simple ranking algorithm</td>
<td>88.4</td>
</tr>
</tbody>
</table>

VI CONCLUSION

We have successfully modeled the end-to-end SSM to identify the probable suspects involved in a crime by making effective use of the stored case histories and prediction techniques such as text mining and clustering. The obtained results from this model can help the investigation officer to understand the probable suspects and also aid to track the location of the probable suspects. Thus our proposed approach effectively identifies the relevant suspects and helps the officer to solve the crimes at a faster pace compared to the traditional methods in processing the crime cases. We hope to improve this SSM by introducing more relevant localized criminal behavior patterns and crime type features.

REFERENCES