Spatial Clustering Simulation on Analysis of Spatial-Temporal Crime Hotspot for Predicting Crime activities

M.Vijaya Kumar
Assistant Professor
K.S.R. College of Engineering
Tiruchengode

Dr .C.Chandrasekar
Associate Professor
Periyar University
Salem

ABSTRACT
This paper presents a spatial-temporal prediction of crime that allows forecasting of the criminal activity behavior in a particular district by using structured crime classification algorithm. The quantity of each crime is understood as the forecasted enhance or reduce the particular moment in time and location of the criminal activity. The proposed algorithm used for forecasting crime is based on one year crime reports. It is proposed a new structured crime classification algorithm which improves the prediction performance on the studied dataset of criminal activity. It execute the following analyses: To find the exact hotspot location and disposition analysis, which shows that it is possible to predict crime location promptly, in a specific space and time, and highest percentage of effectiveness in the prediction of the position of crime. The usage of the said algorithm is to identify the particular crime from number of crimes.

Keywords
Crime hot spots, structured classification, crime Pattern Theory, optimization algorithm, spatial clustering.

1. INTRODUCTION
Public security and crime forecasting activities are some of the most important concerns of both citizens and government. Large amounts of money, human resources, equipment and services are devoted to these activities. In addition, there is a constant concern to justify resources allocated for police. If the police could anticipate, with an acceptable degree of precision, when and where criminal activities of a specific kind are going to take place, it would achieve a double benefit. First, it would be possible to concentrate the necessary logistic activities and resources fighting that specific kind of criminal activity in the geographic area and forecasted time frame, and the comparison between the amount of resources allocated to police forces and the results achieved by them may result in a more adequate basis for planning and distribution of public security.

There are several works dedicated to the study of spatial and temporal verdicts made by criminals, i.e., identifying hotspots where criminal activity is concentrated [1, 2]. A commonly used method is the Spatial and Temporal Analysis of the Crime Program [4], which clusters crime points within ellipses [3]. Surveys additional hotspot methods, the most complicated of which employs the kernel density estimation method [19]. Nevertheless, the main disadvantage of statistical methods is that they do not present additional semantic data for describing the incident under study. In the specific case of crime prediction, this kind of information is highly desirable, as it is needed to support decision making processes and, in general, to prepare preventive and corrective policies. Because of this, we have selected inductive classification methods over statistical ones in order to generate an inductive description of each type of criminal activity studied.

These descriptions by themselves constitute valuable information that provides a general overview of the criminal activity scenario. Further, by using these inductive definitions, it is possible to identify the expected enhance or reduce in specific criminal activities that will most likely occur in specific areas and times. This paper reports, experimental results produced by the proposed algorithm of crime activity forecasting within a specific time period and location using structured crime classification algorithm. In Section 2, it has been presented the details of forecasting exact hotspot identification using Clique optimized clustering algorithm from the 200*200 size grid. The particular implementations give 86% accuracy. Then in Section 3, it is shown that the structured crime classification algorithm design and produce accuracy output of our experiments. It had been performed two analyses: Hotspot identification, and structured crime classification algorithm analysis. For the first analysis, the data used has been from the Chennai city promoter apartment’s data. Within this analysis we perform experiments for spatial and temporal location of crime and expected burglary crime. For the structured crime classification algorithm analysis, data from the Chennai city one year police data had been used. Finally, in Section 4, it has been described our conclusions and discuss future work.

2. PROPOSED SIMULATION
The Proposed simulation description here is part of a police information system designed to prevent and react to crime. The function first step of the Crime records it was collected the number of past year recorded crimes. The second step contains structured crime classification algorithm it was found the particular crime using classification conditions. The input for the third step is the output of structured crime classification algorithm the predicted specific crime made the instance of the KDE grid, after the instances Clique optimized clustering algorithm performs the separations of the clusters. The created suggestions are classified into four types: (1) suggestions on police patrol to avoid the random (2) suggestions to begin specific defensive movements through the crowd media, (3) suggestions to make the new police station in high sensitive area, and (4) general suggestions on the state of the police resource allocations.

The problem of crime prediction requires several different information sources. All of them directly related with public security, but not easily accessible. In this research have chosen three categories crime information on (1) crimes Place, (2) Crime Time, (3) Particular crime (Burglary crime)(4) Density of Population , (5) Number of Promoter Apartments.

2.1 CLIQUE OPTIMIZE ALGORITHM

Clustering is one of the most useful tasks in data mining process which can separate objects of a data set into distinct clusters such that two objects from one cluster are similar to each other, whereas two objects from distinct clusters are not. Spatial clustering means to identify clusters or densely populated areas in a large spatial data set, give out as an important task of spatial data mining.

The research of spatial clustering is very active. The main satisfactory methods of this area include: algorithm based on spatial dataset separation such as k-means, k-medoid, CLARA, etc. algorithm based on hierarchical clustering such as CURE, BIRCH, etc. algorithm based on density such as DBSCAN, and DENCLUE , etc. algorithm based on grid such as SING, wave-cluster and CLIQUE, etc. To estimate the capability of spatial clustering algorithm, some principle is shown as follows: scalability and the ability of discovering cluster of any shape. With these criterions of spatial clustering, proposed CLIQUE Optimization algorithm for finding the subspace clustering of spatial dataset automatically, non-awareness of the order of data input, the ability of finding any shape clusters and the linear scalability with the addition of dataset.

In this paper, the CLIQUE Optimization algorithm follows the major steps.

**Step 1:** divide the space with X, Y axis depending on distributing of input spatial data to avoid demolishing of accepted clusters

**Step 2:** Apply of pruning concept to find Density area and sparse area.

**Step 3:** To find the Threshold Parameter (P); T→minimum number of data point’s density area.

**Algorithm**
The CLIQUE optimization algorithm finds the max out clusters with the “Hill climbing” method and it tries to look for the pair of points between two max out clusters and the cluster which can satisfy the statistically important.

**Step 4:** To identify the Dense Area clusters and Sparse Area using Threshold parameter (T)

In this step, count the number of data points in every area N, recording the value NUMi, calculating the density of Area with the x, y coordinate value and T ≥ NUMi. It is Density of Cluster otherwise it is Sparse area.

**3. RESULTS AND FINDINGS**

Figure 7 is the initial spatial dataset D shows the clustering results of CLIQUE Optimization setting Findings = 7, T = 1. Depicts the partitions found by CLIQUE Optimization algorithm on this spatial dataset. The clustering result of new algorithm with T = 1, (P, Findings) = (1, 7).

Basically, CLIQUE Optimization algorithm is a grid-density clustering algorithm, and CLIQUE is a popular and accepted grid-density algorithm of spatial clustering because of its excellent ability of every aspect. In this experiment, we will compare CLIQUE Optimization algorithm with CLIQUE to approve the more excellent ability of new spatial clustering algorithm.

**4. CRIME CLASSIFICATIONS**

In this chapter, structured crime classification is the relatively easy method of the concept clustering. The method tries to identify more similar object in the data sets same as general clustering goal. The main idea of structured crime classification includes some steps as following. The first step is querying and gathering Hotspot data relevant to the particular crime, the second step is examining each Hotspot of task-related data and the third step is counting the Crime type with more intangible values repeatedly. The more insubstantial value is defined in the generalized hierarchy grid in advance.

In order to define the similarities between crime types, classification on the crime database can be used. Classification is the simplification hierarchy on these crime types. These attributes above can be depicted by classification respectively.

2. Crime based on Time.
3. Crime based on Specific Type (Ex., Burglary).

**4.1. Structured crime classification Algorithm**

**Step 1.** Start
**Step 2.** Collect the crime Record and denote DB.
**Step 3.** Assign DB=|DB| Denote number of crime events.
**Step 4.** Assign $C_i = \text{Similarity}(Obj_1, Obj_2, ..., Obj_n)$ where $Obj_{1..n}$ are Crime Events of the DB.

The objects below study are People’s intelligence recorded in a given geographical area at a specified time. Temporal data includes the date and time components, while the location refers to the area in which the criminal activity was recorded.

**Step 5.** To find the probability of particular crime $P_i = (\text{requested forecast crime, Crime Dataset})$, record the Positive Character, the Negative Character descriptions of the Particular Crime.

**Step 6.** To find the Threshold Value ($T$), $T = \text{(Cluster Density Area - Sparse Area)}$, in hotspots.

**Step 7.** Identify the Classes of crimes types, classes based on the kind of social impact that the criminal activity. Ex: Class1 = (Burglary in Apartments, Shops, Small Houses), Class2 = (Robbery, Injury, Property damage), Class3 = (sex, children abuse, usurp)…, etc.

**Step 8.** $F(C) = \text{class}_i \cup P_i(c_i)$, The properties are subsets of expressive features associated with specific values. For each class, a positive and a negative description are made.

**Step 9.** If $F(\text{Class}_i) \geq \text{positive description}$

Then

Else

Produce a Hotspot.

Produce a Cold spot.
5.2. EXPERIMENTS AND TEST RESULTS

The input data should be entire crime events, and there should be recorded systematically in due time. In the Anna nagar Promoter Apartment area, located in Chennai, area mostly high density of peoples and number of apartments are there. An important part of its task is the collecting of the information from police records, Detective agenesis. Therefore, this area was selected as a case study and test field for both the forecasting crime event reported and the computerized system for behind the conclusion- making development on communal security. For the analysis we used Promoter apartment’s burglary data from the Anna nagar area of Chennai. Inside this analysis this research performed experiments for Classification of the crime events. Experiment used data from the Chennai city police crime dataset.

Table 2. Reported crime from the record

<table>
<thead>
<tr>
<th>Incident Time</th>
<th>Date</th>
<th>Area</th>
<th>Crime type</th>
</tr>
</thead>
<tbody>
<tr>
<td>12.00 am to 6.00 am</td>
<td>May 15</td>
<td>Anna nagar</td>
<td>Burglary</td>
</tr>
<tr>
<td>5:00 PM to 8:00 PM</td>
<td>Aug 19</td>
<td>Choolai</td>
<td>Shop theft</td>
</tr>
<tr>
<td>12:00 AM to 5:00 PM</td>
<td>July 10</td>
<td>Avadi</td>
<td>Sex</td>
</tr>
<tr>
<td>11:00 PM to 3:00 AM</td>
<td>Dec 28</td>
<td>Venanur</td>
<td>Property damage</td>
</tr>
<tr>
<td>11:00 AM to 3:00 PM</td>
<td>Jan 25</td>
<td>Kolthur</td>
<td>Burglary</td>
</tr>
</tbody>
</table>

Table 3. Pre-processed report used as Classification algorithms

As the algorithm calculate the characteristic features and the complementary features of the sample by applying set of data with 140 crime events 73% of the 370 records from the whole record set. These 140 records were: Burglary in Apartments: 91, Home: 12, Isolated Houses: 37. The structured crime classification for the positive description is 82%.

5. SIMULATION RESULTS

In this section, we present our simulation tests with a different dataset and compare against predictions by the Monte Carlo simulation. Additionally, the experiment results using the Spatial- Temporal hot spot and cold spot creation. This measure consists of monthly and yearly measurements of forecast crime events it is based on the recorded police crime data set for the particular position.
To test the proposed forecasting Simulation results, the Chennai city crime dataset was used. This dataset contains maximum of 100000 registered crimes and was made available by the Chennai city Police Department. All crimes were committed within 10 inspection sectors (space- units), over a period of time from January 2005 to December 2008 (time- units). Crimes were organized into 16 crime Classes as shown in Table 3.

By evaluating only records from the last four years (2005 to 2008), a forecast was considered for time unit January 2008, all registered crime Classes and within all 10 inspection sectors. The forecast number of crimes was then matching up to with the real- life police records from that same space- time unit (2,201 crimes during January 2008).

Two limited forecasts of the number \( F \) of crimes belonging to the \( C \) crime Classes to be observed within the \( R \) space unit and during \( T \) time unit are used to estimate the movement observed in the reference dataset. When the movement is known, it is likely to forecast the next value to be observed by appearing at the last observed values. To achieve this forecast, the basic tool to be used is a query to the dataset to count the number of crimes, from a specific Class \( c \) and observed within specific Region \( R \) and set of attributes from the Object \( S(z) \) and time units \( T \), the result of such a function is represented by function \( #(C,R,S(z),T) \).

Specifically, the forecast \( F \) of crimes, belonging to the \( C \) crime-family, to be observed within the \( R \) and set of attributes for the crime objects during time \( T \), is determined by

\[
F(C, R, S(z), T) = \sum_{m=1}^{12} \sum_{i=1}^{k} #(C,R,S(z),T^m) - #(C,R,S(z),T^{m-1})
\]

The function of Month based simulation analysis predicted give up by the analysis of crimes, from the same crime classes, observed within the same Region, set of attributes from the crime objects and time observed the last one year in the reference dataset.

### Cumulative Percentage Crime Ratio (CPR)

Both simulation analysis can be expressed cumulative Percentage of crime ratio.

\[
CPR = \frac{\sum_{i=1}^{k} CPRI - \sum_{i=1}^{k} CPRI_{Year-1}}{\sum_{i=1}^{k} CPRI_{Year}} \cdot 100
\]

### Total Crime Classification

\[
TC_i = \sum_{i=1}^{k} CPRI
\]

### Crime Attribute Ratio

\[
CAR = \frac{T_{ci}}{K_i}
\]

<table>
<thead>
<tr>
<th>Crime Class</th>
<th>Classified Data</th>
<th>Test Data</th>
<th>Monte Carlo Simulation Forecasting Crimes</th>
<th>Our Simulation Forecasting Crimes</th>
<th>Monte Carlo Simulation</th>
<th>Our Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>19000</td>
<td>550</td>
<td>1900</td>
<td>514</td>
<td>15.04</td>
<td>3.7</td>
</tr>
<tr>
<td>Burglary in shops</td>
<td>1000</td>
<td>460</td>
<td>640</td>
<td>370</td>
<td>4.24</td>
<td>5.01</td>
</tr>
<tr>
<td>Burglary in Houses</td>
<td>12007</td>
<td>900</td>
<td>1504</td>
<td>631</td>
<td>9.10</td>
<td>5.38</td>
</tr>
<tr>
<td>Burglary in Villages</td>
<td>6001</td>
<td>820</td>
<td>5000</td>
<td>750</td>
<td>8.28</td>
<td>8.28</td>
</tr>
<tr>
<td>Burglary in Promoter Apartments</td>
<td>13001</td>
<td>1000</td>
<td>5005</td>
<td>640</td>
<td>15.50</td>
<td>3.95</td>
</tr>
<tr>
<td>Total</td>
<td>54009.00</td>
<td>3790.00</td>
<td>14549.00</td>
<td>2885.00</td>
<td>51.16</td>
<td>18.77</td>
</tr>
</tbody>
</table>
Table 5. Preparation from 1/JAN/2005 to 31/DEC/2008, Test month February 2008

<table>
<thead>
<tr>
<th>Crime Class</th>
<th>Classified Data's</th>
<th>Test Data's</th>
<th>Monte Carlo Simulation Forecasting Crimes</th>
<th>Our Simulation Forecasting Crimes</th>
<th>Monte Carlo Simulation</th>
<th>Our Simulation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>23000</td>
<td>650</td>
<td>8000</td>
<td>614</td>
<td>16.04</td>
<td>4.7</td>
</tr>
<tr>
<td>Burglary in shops</td>
<td>2000</td>
<td>360</td>
<td>740</td>
<td>470</td>
<td>5.24</td>
<td>4.81</td>
</tr>
<tr>
<td>Burglary in Houses</td>
<td>13000</td>
<td>1000</td>
<td>14604</td>
<td>731</td>
<td>10.30</td>
<td>4.68</td>
</tr>
<tr>
<td>Burglary in Villas</td>
<td>10000</td>
<td>920</td>
<td>5600</td>
<td>830</td>
<td>9.28</td>
<td>8.82</td>
</tr>
<tr>
<td>Burglary in Promoter Apartments</td>
<td>13000</td>
<td>1100</td>
<td>5105</td>
<td>740</td>
<td>15.50</td>
<td>3.96</td>
</tr>
<tr>
<td>Total</td>
<td>59000.00</td>
<td>4320.00</td>
<td>15840.00</td>
<td>15889.00</td>
<td>56.36</td>
<td>22.77</td>
</tr>
</tbody>
</table>

The result comparison shows Table 5 presented as a reference to the usage of the simulation measure. The results worked with different data set, with the city of Chennai, INDIA and in different time.

These simulation outcomes show that the proposed method has high usefulness, with a Monte Carlo Simulation forecasting all crime events based on space and time units, during January 2008 (with a total of 2,219 Promoter Apartments burglary crimes). This returns those, in common, the proposed forecasting simulation method only decrease small in less than five events of each crime classes. Such accuracy is logically acceptable for computerized crime analysis systems and might embrace a useful tool for development Forecasting police operations.

7. CONCLUSIONS

There are two key advantages of our Simulation model: first, the inductive classification of each class constructed by the learning procedure comprises by itself precious information for describing the criminal circumstances under study. Second, the simulation model was designed for dealing with space and time data, constructing the three level of simulation based on the parameter choices it yields the same results. This means that our simulation model will forecast new criminal events very similar to those before registered in the same space-time framework.

8. REFERENCES

[9] Xiang Zhang; Zhiang Hu; Rong Li; Zheng Zheng; Detecting and mapping crime hot spots based on improved attribute oriented induce clustering Geoinformatics,2010, 18 th International conference,Beijing, pp: 1-5.


