Changing Aspects and Impact of Chennai City Crime Hotspot and Cold spot Based on Mining Based Simulation

Dr. M. VijayaKumar¹, Dr. Abdel-Badeeh M. Salem², Dr. S. Karthick³ and Dr. K. Moorthy⁴

¹Professor
K.S.R. College of Engineering, Tiruchengode, India.
Vij370@gmail.com

²Professor of Computer Science
Head of BioMedical Informatics and Knowledge Engineering Research Lab,
Faculty of Computer and Information sciences.
Ain Shams University, Abbasiya, Cairo-Egypt

³Dean
Department of Computer Science Engineering,
S.N.S. College of Technology, Coimbatore, India.

⁴Director (Mgt Studies)
K.S.R. College of Engineering, Tiruchengode, India.

Abstract

In this paper we study the performance evaluation of a forecasting program for Chennai city crimes in micro, me-so analysis of crime hot spots and cold spots. We perform a geographical study of the crime hot spots using one year of crime report data on forecasting crimes from Chennai city. This study is based on kernel density method for a definition of micro, macro analysis of crime hot spots and cold spots, applying spatial clustering simulation for analyzing crime hot spots. Spatial clustering simulation hot spots are good targets for crime forecasting. Spatial
temporal crime hot spots, however, mainly analysis only for past one month. This is the most common method for identifying crime hot spot for short term and medium term crime data and predicting that the resulting crime hot spots will persevere is operative for simulation hot spots. In its place it is necessary to forecast the development of simulation hot spots to avert their crimes. Kernel Density Estimation (KDE) crime hot spots are comparatively easy targets for police department whereas simulation hot spots require forecasting techniques not frequently in use by police department. The research appraisals around a 30 to 40 percent reduction in micro, me-so analysis of crime hot spots and cold spots in Chennai city if the spatial clustering simulation implementation program proposed in this research were applied.

Keywords: crime hotspots, spatial clustering, kernel density estimation, crime forecasting, micro, macro simulation, computational criminology, crime data mining.

1. Introduction

Simulation with computational criminology is promising as a new field that applies computer science and mathematical functions and methods to the study of crime forecasting problems [4]. The difficulty of human activities, social connections and police parameters present unexpected challenges to forecast criminal activities and decide the best promising means to control it. Although, some form of mathematical functions, largely statistics has been used by a crime analyst for their crime data analysis but the crime applications have been affected by the limitations of crime data.

The simulation could be a broad field of computer science that encompasses a variety of approaches that share a group of characteristics. Interactions inside simulation models are very advanced, thus simplicity at the beginning is important for understanding. Models are developed from theory and specified in a computer program where they are ready to accommodate dynamic, non-linear interactions that supported at the time [1].
A formalized computer program provides real credentials for the assumptions of the model and allows simplicity within the analysis project that is essential for duplication and testing of results. These attributes are particularly necessary when one is making an attempt to find the mechanisms through the observed me-so, micro, level patterns are shaped. The simulated experiments involve recognizing and notifying a variety of advanced relationships and then making a model.

1.1 Data mining simulation

Heinonen, Mannila [13] have made a successful data mining technique of attribute oriented induction which is viewed as conceptual clustering. They supply the AOI algorithm for conceptual clustering. It performs two operations (1) Generalizing an attribute, it executes by choosing the closest rows and then selecting the attribute to make the final order. (2) Selecting the attribute for general comparisons, it performs the generalization step of an attribute and causes an update of all the rows promptly.

Spatial data mining and knowledge discovery has emerged as a lively analysis field that focuses on the event of theory, methodology and observe for the extraction of helpful information and information from large and complicated spatial databases. Guo, [6] has proposed spatial information mining continues to be at an awfully early stage and its bounds and potentials are however to be defined. There are opportunities and challenges facing each spatial information mining analysis.

The preliminary results of a crime forecasting model developed in collaboration with the police department of a United States city in the Northeast. At first discuss our approach to architecting datasets from original criminal records. Chung-Hsien Yu, et al [2] have suggested the datasets contain aggregated counts of crime and crime-related events categorized by the police department. The location
and time of these events are embedded in the data. Additional spatial and temporal features are harvested from the raw data set. Second, an ensemble of data mining classification techniques is employed to perform the crime forecasting. We analyze a variety of classification methods to determine which is best for predicting crime “hotspots”.

In this proposed simulation based on spatial clustering, in this study found the data from the output of the CLIQUE optimization method and apply two types of computer simulation like macro, micro, for identifying medium and short time forecasting of crime data. There are a number of reasons for this selection: hot spot identification and analysis, are based on the tactical analysis, police resource allocation and deployment are based on short, medium term forecasting. Additionally this study developed mathematical functions for solving the simulation in a better way.

2. Literature Review

This segment reviews the crime hot spot theories that clarify why crime hot spots approaches have become an important method to regulating and how operative they are. It also comprises a review of crime forecasting approaches.

2.1 Crime Hotspot Theories of Crime Mapping

Each level has basic units of analysis which is the data being examined. One will think about units as such as the geographic areas being depicted on maps: points, lines, or polygons. Some criminological theories facilitate making a case for purpose concentrations of crime [12]. Different theories facilitate making a case for linear concentrations of crime or hot spot crime polygons. However, theories of crime are helpful for serving to guide crime and disorder mapping as long as one selects a theory acceptable for the amount of research and action. Crime analysts frequently
assume that crime distributions are clustered and whether or not clusters exist or not, but some are identified from random crime distributions. Testing for clustering is that the initial step in revealing whether or not the information has hot spots of crime. Between these extremes we have near repeats which take place in an area no more than three or four city blocks in extent.

In different viewpoints, the present crime theories vary in several forms [4]. The first and most common theories are placed-based theories. Place-based theory is common in social science, it falls squarely inside the theoretical approach, however the first mechanism in place-based theory structural framework into individual actions. The crime occurs within the specific point; the suitable cell analyzes the address, space, time that is represented in the maps on dots.

There are varieties of crime hotspot theories used for identifying the repeat places, like repeat area hotspot and repeat street hotspot [11]. This type of repeat hotspot theories are mostly used in high crime incident areas. These places may be homes, searching areas and different small locations. During this study, hot spot within the maps are focused in addition, the other places targeting no fear of crimes. These hotspots are represented by dots, so procedure of crime analysis and mapping are based mostly on the dot points.

There are some theories mostly applied to crime maps which contain the original spatial data. It has some limitations. The unique data brings lower effectiveness. Based on this study, preprocess the spatial data and the crime events and then use structured classification algorithm for clustering the crime attributes. Finally events are placed on the map and they identified specific hotspots.

## 2.2 Identifying Crime Hotspots
There are several methods and techniques used to understand and identify crime hotspots, up to date. All the methods are based mostly on statistical technique [6]. Our proposed approach, a preliminary global statistical technique, in that relies on data mining clustering method. The contribution of data mining and global statistics techniques was helpful for identifying the easily crime hot spot in clustering. Crime analysts repeatedly assume that crime distribution is clustered and is the complete spatial uncertainty. Based mostly on this structure, crime classification algorithm is proposed to classify the crime attributes.

Point mapping is a most common approach used for displaying crime patterns in a special type of crime application [14], if these particular individual geographic point objects attributed with data, such as the specific coded data type and time and place are selected easily applying conditions. The selected attributed data are displayed by appropriate symbol representing the class of crime displayed. Point maps are used in general purpose like point density maps, KDE map etc... The existing mapping methods are relatively used to point or places crime data.

Thematic mapping is a special type of mapping used for representing spatial crime data distribution in geographic boundary location. The geographic boundary locations are administrative crime control areas of police control areas like city, block etc. Mapping crime event points are collected in this geographic region. These types of boundary maps are KDE map etc., the quadrate thematic map is completed by the surface smoothing concepts.

2.3 GIS and Importance of “Hot Spots”

Geographic or spatial analyses of crime using GIS have established its meaning in each criminological analysis and criminal justice. In recent years, their use in analyzing crime patterns has been viewed as a very important part of the effort
in law enforcement agencies toward successful and economical crime management [14]. A theme matter within the geographic analysis of crime that has attracted a lot of attention is that the identification of crime “hot spots,” specifically the locations or small areas within which disproportionately sizable amount of the criminal incidents cluster. Researchers and police use the term in many alternative ways in which, some talk to hot spot addresses, others talk to hot spot blocks and some others examine clusters of blocks. Compared to researchers, crime analysts discover for concentrations of individual events that may signify a series of connected crimes. They additionally consider small areas that have an excellent deal of crime or disorder, despite the fact that there is also no common offense. Analysts additionally observe neighborhoods and neighborhood clusters with high crime and disorder levels and take a look at to link these to underlying social conditions.

However there is no common definition of the phrase hot spot of crime survival, the general understanding is that a hot spot is a neighborhood that features a larger than average selection of crime or disorder events, or a neighborhood where individuals have the next than average risk of victimization. This implies the existence of cool spots or areas with but the typical quantity of crime or disorder. It conjointly suggests that some hot spots could also be hotter than others; that is, they vary in how a higher than average they are.

Identifying crime hot spots incorporates a vital sensible suggestion, as a result of there are accumulating items of proof that ‘the additional law enforcement efforts are centered on high- crime places or high-crime time, the additional successful and economical they might be in controlling crime’.

2.4Crime Forecasting
The origin of crime forecasting in year 1998, when the US National Institute of Justice (NIJ) awarded five grants to study crime forecasting for police use as an extension of crime mapping. Gorr, W.L. et al [4] have proposed instead of mapping only recent crimes and assuming that observed patterns would persist, the objective was to forecast crime one period ahead, with results displayed as maps.

Crime detection and forecasting methods require a coarse aggregation of cases (e.g. by month, by square mile), due to both computational considerations and the relatively small number of serious crimes. Kulldorff M[8] has proposed these limitations reduce the spatial and temporal precision with which departments can pinpoint clusters of crime, as well as their ability to rapidly respond to these clusters. The use of expectation-based spatial scan statistic methods originally developed for the bio surveillance domain, which can use a finer aggregation of data and can efficiently search for emerging space time clusters of varying size and duration.

Gorr, W.L. et al [4] have divided the forecasting techniques into two categories in terms of the predicted time period. Crime forecasting includes long-term forecast models for policy planning and applications in broader manner and short-term forecast models for tactical decision making.

The challenges involved in predicting crime rates or the impact of different crime policies are very similar to those in other forecasting domains. A.McCue, C.et al [10] have suggested the classic examples include predicting sales of a product, changes in interest rates, the likelihood of a terrorist attack, or the outcome of political elections. In each of these cases, the inputs needed to generate a reliable forecast may be tilted by a variety of factors, some of which might be begin but troublesome, and some of which may be selfish and opportunistic.
Cohen, J.et al [3] has developed highly reliable methods for forecasting future crime trends and problems, which are the most preferred ways to improve crime prevention and reduction. With the advance of crime forecasting, spatial and temporal predictions of crimes are used to make long and short term planning. In the situation of getting accurate predictions, it is possible to manage security resources efficiently. Police give attention on highlighted areas, target patrols, allocate resources and carry out other police interventions to prevent crimes.

3. Crime data

Crime data of year 2008 have been used in the analysis; Spatial and temporal information related to these incidents were obtained from Chennai Police headquarters. Crime data were recorded by two police stations, in south Chennai and north Chennai. Data includes number, address, occurrence time, location and type. Five types of data are available, which are murder, burglary and auto related crimes and pick pocketing. However, in this study all types of crimes are aggregated to have a higher number of incidents for constructing reliable short and long term forecasting.

All crime related data employed in this research have been got from Chennai 100 Police station centers. The database records the data of case variety, occurring time, case type, occurring location and different transient description of the crime. In Chennai, there have been totally 3,706,34 crime records in 2008. Consistent with the police crime discrimination framework, these recorded crimes could be classified into 75 types. However, as several classes rarely happen here we just focus on those types that frequently occur (e.g. murder, tried murder, wounding and assault, rape, indecent assault and theft of property, fraud burglary, pick-pocketing and stealing vehicles, electro-mobiles, motorcycles, or bikes). When eliminating unqualified information (repeating decision or incomplete and invalid records) and traffic accidents that is
Beyond the scope of this analysis, the full variety of remaining crime records were more than a hundred thousand.

Table 1: Crime Database 2008

<table>
<thead>
<tr>
<th>Areas</th>
<th>Personal Property</th>
<th>Business Property</th>
<th>Total</th>
<th>Robbery</th>
<th>Burglary in a Dwelling</th>
<th>Burglary in Other Building</th>
<th>Total</th>
<th>Counted per Victim</th>
<th>Other Fraud</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adayar</td>
<td>190</td>
<td>55</td>
<td>245</td>
<td>208</td>
<td>415</td>
<td>623</td>
<td>353</td>
<td>565</td>
<td>918</td>
<td></td>
</tr>
<tr>
<td>Ambattur</td>
<td>181</td>
<td>34</td>
<td>215</td>
<td>342</td>
<td>260</td>
<td>602</td>
<td>35</td>
<td>164</td>
<td>199</td>
<td></td>
</tr>
<tr>
<td>Aminjikarai</td>
<td>172</td>
<td>17</td>
<td>189</td>
<td>600</td>
<td>250</td>
<td>850</td>
<td>4</td>
<td>120</td>
<td>124</td>
<td></td>
</tr>
<tr>
<td>Anna Nagar</td>
<td>216</td>
<td>28</td>
<td>244</td>
<td>375</td>
<td>89</td>
<td>464</td>
<td>25</td>
<td>118</td>
<td>143</td>
<td></td>
</tr>
<tr>
<td>Ashok Nagar</td>
<td>164</td>
<td>14</td>
<td>178</td>
<td>275</td>
<td>159</td>
<td>434</td>
<td>53</td>
<td>107</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>Choolai</td>
<td>156</td>
<td>17</td>
<td>173</td>
<td>171</td>
<td>128</td>
<td>299</td>
<td>70</td>
<td>219</td>
<td>289</td>
<td></td>
</tr>
<tr>
<td>Kolathur</td>
<td>117</td>
<td>18</td>
<td>135</td>
<td>338</td>
<td>114</td>
<td>452</td>
<td>33</td>
<td>154</td>
<td>187</td>
<td></td>
</tr>
<tr>
<td>Kottur</td>
<td>251</td>
<td>35</td>
<td>286</td>
<td>335</td>
<td>149</td>
<td>484</td>
<td>108</td>
<td>271</td>
<td>379</td>
<td></td>
</tr>
<tr>
<td>Mylapore</td>
<td>370</td>
<td>49</td>
<td>419</td>
<td>232</td>
<td>154</td>
<td>386</td>
<td>37</td>
<td>138</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>Pallavaram</td>
<td>240</td>
<td>40</td>
<td>280</td>
<td>165</td>
<td>116</td>
<td>281</td>
<td>45</td>
<td>80</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Porur</td>
<td>239</td>
<td>44</td>
<td>283</td>
<td>210</td>
<td>73</td>
<td>283</td>
<td>25</td>
<td>151</td>
<td>176</td>
<td></td>
</tr>
<tr>
<td>Saidapet</td>
<td>133</td>
<td>15</td>
<td>148</td>
<td>218</td>
<td>102</td>
<td>320</td>
<td>88</td>
<td>205</td>
<td>293</td>
<td></td>
</tr>
<tr>
<td>Velacheri</td>
<td>100</td>
<td>8</td>
<td>108</td>
<td>257</td>
<td>159</td>
<td>416</td>
<td>45</td>
<td>233</td>
<td>278</td>
<td></td>
</tr>
<tr>
<td>Villivakkam</td>
<td>143</td>
<td>21</td>
<td>164</td>
<td>201</td>
<td>105</td>
<td>306</td>
<td>41</td>
<td>115</td>
<td>156</td>
<td></td>
</tr>
<tr>
<td>West Mambalam</td>
<td>102</td>
<td>28</td>
<td>130</td>
<td>218</td>
<td>96</td>
<td>314</td>
<td>23</td>
<td>87</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>T-Nagar</td>
<td>87</td>
<td>21</td>
<td>108</td>
<td>145</td>
<td>48</td>
<td>193</td>
<td>29</td>
<td>136</td>
<td>165</td>
<td></td>
</tr>
<tr>
<td>Manali</td>
<td>78</td>
<td>13</td>
<td>91</td>
<td>251</td>
<td>123</td>
<td>374</td>
<td>57</td>
<td>184</td>
<td>241</td>
<td></td>
</tr>
</tbody>
</table>
4. The Proposed Approach of Spatial clustering simulation

Forecasting validation problems are classified by the extent of the development of research perspective. In our proposed work, spatial clustering simulation classified into the three ways, macro level simulation is applied for long-term forecasting and micro level simulation for short term crime forecasting, me-so level simulation for medium term crime forecasting and tactical decision making. There are a number of reasons for this selection: hot spot identification and analysis, police resource allocation and deployment are based on short, medium, long term forecasting.

4.1 Micro level simulation for short term forecasting

Micro level simulation of crime forecasting analysis includes the common day-to-day analysis of crime by the police department: understanding police reports,
identifying crime patterns in the recorded data, mapping crime hotspot points, identifying and analyzing crime hot spots, etc. The current research of crime hot spots would likely to remain police patrol targets and perchance some of the most important hot spots also require police patrolling. Also, the police department might concentrate on high crime density areas for leading indicator crimes.

4.2 Me-so level simulation for medium term forecasting

This level of simulation is needed for the police department to take the quick decision for the next month activities based on previous month activity. For example the entire last month crime incidents were stored in the record. We apply me-so level simulation for each class of crime data to find the hot spot for last month. Based on the result the police pay more attention to the particular place.

In this simulation we use the time series forecasting used for providing the starting point for evaluation and forecasts of hotspots with the possible crime increase in the following month. There are two primary objectives of me-so level simulation.

1. Estimate Earlier Period Performance
2. Preparation of Police Force for the following Month

In this study, the simulation function for analysis of the medium term forecasting based on me-so level simulation is proposed.

\[ F(C, R, S(z), T) \]

Where

\[ C \]----- Crime Classes  
\[ R \]----- Region (or) Area
S(z)—Set of Attributes From the Crime Objects
Ex: S(burglary, sex,
Shop Theft, Road Accidents…. Etc..)

T------ Time

In the proposed month based analysis, simulation function creates a time series data for every month based on the previous month recorded data, for the particular hotspot. The Proposed function has some limitations for a particular crime hotspot region administration level data in crime analysis.

**Month Based analysis**

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

**4.3 Crimes Attribute Ratio (CAR)**

This proposed crime ratio function is used to identify the attribute ratio based on last year total crime classification and particular crime attributes classification. Additionally it was used to identify the crime hot spot and crime cold spot based on last year crime ratio and current year crime ratio.

\[
CAR = \frac{TC}{K}, \quad ------ K= \text{Classification}
\]

If \( CAR_l \geq CAR_p \)

Produce Hot Spot

Else

Produce Cold Spot
5. Results and Discussion

5.1 Micro Simulation Results

In this research, the micro level simulation collects the spatial crime database from Table 1, applied the classification and set the attribute specification for all data. With the result, using the attribute type, it is going to be performed the short term forecasting operation based on micro level simulations, which is used to record the week day’s crime within the hotspot maps. On a day-to-day basis, the density of present hotspot locations is compared with previous hotspot locations.

![Figure 2: Attribute Name and Type](image)

Figure 2: Attribute Name and Type
In this study, micro level simulation identifies the crime incidents and performs matrix operation. In the figure 6.3, it shows the 24X7 crime incidents are converted into 7 x 7 matrices and compares hotspot to the previous day(s) row wise or column wise.

### 5.2 Me-so Simulation Results

This research uses to apply the practical forecasting for the same crime class. In the database table 1, the burglary crime class is taken for experiments in the same space with the different time like month wise.

**Table 2: Crime Data Form 2008 Crime Record with Crime and Hotspot**

<table>
<thead>
<tr>
<th>Months</th>
<th>Crime</th>
<th>Hotspots</th>
</tr>
</thead>
<tbody>
<tr>
<td>January</td>
<td>196</td>
<td>32</td>
</tr>
<tr>
<td>Month</td>
<td>First Number</td>
<td>Second Number</td>
</tr>
<tr>
<td>----------</td>
<td>--------------</td>
<td>---------------</td>
</tr>
<tr>
<td>February</td>
<td>158</td>
<td>28</td>
</tr>
<tr>
<td>March</td>
<td>138</td>
<td>27</td>
</tr>
<tr>
<td>April</td>
<td>100</td>
<td>25</td>
</tr>
<tr>
<td>May</td>
<td>91</td>
<td>21</td>
</tr>
<tr>
<td>June</td>
<td>88</td>
<td>19</td>
</tr>
<tr>
<td>July</td>
<td>82</td>
<td>18</td>
</tr>
<tr>
<td>August</td>
<td>79</td>
<td>16</td>
</tr>
<tr>
<td>September</td>
<td>75</td>
<td>15</td>
</tr>
<tr>
<td>October</td>
<td>72</td>
<td>14</td>
</tr>
<tr>
<td>November</td>
<td>66</td>
<td>14</td>
</tr>
<tr>
<td>December</td>
<td>64</td>
<td>12</td>
</tr>
</tbody>
</table>

**Simulation Function**

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

<table>
<thead>
<tr>
<th>Label</th>
<th>Expansion</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>Crime Class</td>
</tr>
<tr>
<td>R</td>
<td>Area</td>
</tr>
</tbody>
</table>
In this Figure 4 shows the simulation results of month-wise crime forecasting. The X axis is for mentioning months of a year and Y axis is for number of crimes recorded. In this case, it has been taken 200 recorded crime events. The red line represents the hotspots and the blue line represents the number of crime incidents. From the above graph, it has been noted that the occurrence of crime events is more in the month of January and afterwards decreases gradually based on the previous month analysis of high density hotspot.

6. Conclusion
The goal of this simulation is to find the best technique to create geographical crime forecasts and implement it in Data Detective. The KDE technique was selected as benchmark technique, because it was a very basic technique and it was used by the police to create forecasts. Computer simulation was selected as second technique. This technique works in a totally different way with selective parameters for forecasting.

7. References


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