CHAPTER 4

STRUCTURED CLASSIFICATION ALGORITHM

In this chapter, study region in terms of geographical, social and economic aspects, the information are made public and additionally the methodology of structured crime classification algorithm and CLIQUE optimization algorithm for crime hotspot identification based mostly on the data mining and spatial clustering have been explained.

4.1 DESCRIPTION OF THE STUDY REGION

Figure 4.1 shows union of the foremost influential fast-developing cities in INDIA. Chennai has recently witnessed the fast urbanization specifically, since the 1990 and the city has experienced unprecedented economic growth throughout the period of shaping it as a globalizing metropolis with a wonderful investment surroundings and international competition. Completely different from the other cities in INDIA,

Chennai is the largest industrial and financial center in mainland INDIA and is additionally thought to be one among the best cargo ports. With the floating population accounting for one third of the overall population, Chennai has been described as the “showpiece” of the world’s fastest-growing major economies.
South Chennai is the eighth largest city in Tamil Nadu and North Chennai is the second largest city with a total population of 1125,463 residents as of December 2008. North Chennai is one among the most important and historical fashionable Baltic sea ports.

As has been stated above, Chennai is a very important place for cultural and sport activities. There are numerous cinemas, theatres, swimming pools and parks located within the south and north Chennai areas. In fact, there are many historical places, variety of parks and the Municipality takes part in developing the infrastructural facilities of Chennai.

Also, business centers, searching malls are densely populated within the district that contribute to the economic development of the city. Therefore, Chennai is one among the foremost, very important and active parts of Tamilnadu which attracts offenders to commit crimes.

Total area of Chennai is 178.20 sq. Kms that is mostly residential areas. The percentage of residential areas in Chennai is 55, where only 25% of the area is business oriented. Though the percentage of business area is considerably high, the area is incredibly important and huge when compare with the other districts.

The mixed usage area is 15% with each residential and business area. The remainder includes the education, military, public associations and cultural facilities.

South Chennai and north Chennai police station regions cover many forms of land use space. South Chennai part includes largely business areas like T-Nagar, the heart of Chennai.
Also, public associations, governmental organizations and less residential space are within the boundary of south Chennai police zone. North Chennai police zone consists of largely residential areas and, however, additionally very important business areas within this area particularly on both sides of the most of the streets in Chennai.

### 4.2 DESCRIPTION OF THE STUDY DATA

Crime data of the year 2008 in the study area are used in the analysis which is illustrated in Figure 4.1. 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 are beyond the scope of this analysis, the full variety of remaining crime records were more than a hundred thousand.

4.3 CRIME DATA COLLECTION

The original crime information of 2008 is stored in oracle 9i, that contains a lot of confidential data, like the caller’s name, identification number etc.. As such data is not associated with our investigation. We tend to exclude this “sensitive” information by replacing the item with some special characters in PL/SQL Developer. We intend to make sure SQL sentences to induce the kinds of crime records in keeping with the “case category” field. Those cases that happened during a certain time may also be selected in keeping with the “happening time” field. Every form of records containing x, y coordinates is exported to text format.
Crime mapping refers to the method of conducting spatial analysis inside the variety of activities of crime analysis. So as to look at the spatial patterns, the method of crime mapping is conducted using ArcMap 9. By adding the text format data into ArcMap, using the “Display XY data” tool, the crime data will then be mapped to Chennai administrative map.

![Number of Crime Activates](image)

**Figure 4.2 Criminal Activities in Study Region**

There are 2008 crime activates are recorded in south Chennai and north Chennai police limits. Crime incidents are mapped with graduated symbols. Figure 4.2 shows which different sizes of features represent particular values of variables. To understand the density of criminal activities in the area, the best way is to apply graduated symbols as incidents are overlapping.

### 4.4 STRUCTURED CLASSIFICATION ALGORITHM

In this study, it has been focused on how to incorporate temporal crime data with STEM model (Kelvin Leong 2008). The disadvantage of the
STEM model is a classification of attributes. The proposed study fulfills the seasonal dependency of crime data attributes, additionally the attributes have been classified into different categories based on crime type, crime place, crime time, etc too. The structured classification algorithm collects all the classified temporal crime data and identifies the positive and negative characteristics of the crime data to arrange the crime hotspot and cold spot. The primary objective of this algorithm is to preprocess the crime data into kernel density maps.

There are a variety of crime theories and methods used for recognizing and applying and for identifying hotspot. Some crime theories facilitate to explain the crime hotspots called places based theories. Additionally alternative theories helped for linear concentration of crime, like street theories and neighborhood theories. There are a variety of techniques used for identifying crime hot spots on maps. The most effective technique was a global statistical technique which has mean, median, clustering, correlation technique etc. These theories indicate mapping like dots and colors.

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 identify specific hotspots.

4.5 CRIME HOTSPOT THEORIES OF CRIME MAPPING

Crime Hotspot Theory

In different viewpoints, the present crime theories vary in several forms. The first and most common theory is 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 falls into individual actions. The crime occurs within the specific point, the suitable cell analyzes the address, space, time as represented in the maps on dots.

**Different Types of Crime Hotspots**

There are numerous crime hotspot theories used for identifying the repeat places, like repeat area hotspot and repeat street hotspot. 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.

**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. Our proposed approach, a preliminary global statistical technique, relies on data mining clustering method. The contribution of data mining and global statistics techniques has been helpful for identifying the easily crime hot spot in clustering. Crime analysts repeatedly assume that crime distribution is clustered and is a 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, if these particular individual geographic point objects are attributed to the 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. whereas the quadrate thematic map is completed by the surface smoothing concepts.

4.6 LIMITATIONS OF STRUCTURED CLASSIFICATION ALGORITHM

Nowadays most of the researchers use spatial clustering for identifying crime hotspots. The spatial ellipses application technique is a difficult technique for distinguishing crime hotspots that encompasses a long tradition in mapping analysis. Finding category hierarchy of the crime data may be a difficult task for the recorded crime. Additionally, this study has used the method to find the threshold value to crime hotspot and cold spots. This technique purely depends on giving the parameters and complex computing. The suitable parameter can lead to the precise output.

4.7 STRUCTURED CLASSIFICATION METHOD

Structured crime classification technique is a comparatively straightforward technique of the concept spatial clustering. During this classification, it aims at detecting more similar objects within the data sets same as a common spatial clustering goal. The first step of structured crime
classification includes the following steps. The initial step is to collect the relevant data applying queries to the given spatial information. The second step is to seek out the crime class based on the similarity of each crime data on a given spatial crime data set. The third step is to find the probability of the actual crime within the criminal class and finally applying a threshold value to seek out the crime hotspot and crime cold spot. Before that, the study of the crime incident analysis is necessary.

4.8 THE ANALYSIS OF CRIME INCIDENTS

In this study the crime incidents include crime, administrative and public security cases. All types of incidents are recorded in detail like period of committing the crime reaction. Each country follows many different ways to describe and store the crime incidents. However, the basic crime record store, the time of the crime incident, the different types of crime incidents, who committed the crime, how much property was lost, whether the crime occurred in control area or other area, what is the specific report of the crime etc. are the common areas of incident. They come across small difficulties. But along with them some are essential attributes, such as time, place, numerical and crime category attributes.

1. Numerical attributes consist of counters, money involved in the cases etc.

2. Time attributes include time of occurrence of crime, crime organized time, crime resolved time etc.

3. Category attributes include abuses of property, burglary, capture, criminals, group's crime etc.

4. String attributes assume that text records are stored in the police information systems.
Structure crime classification method is used to classify the crime incidents based on the similarity between the crime objects stored in the class. Classification is the simplification hierarchy of these attributes. These crime attributes can respectively be represented by classification in three ways, as shown:

1. Classification of crime place
2. Classification of crime types.
3. Classification of crime time.

Classification of Crime Place

The moment crime case is registered; the case record has to be collected and the similarities between crime places have to be found out. Figure 4.3 shows the similarity among various places.

Classification of Crime Types

The moment crime case is registered the case record has to be collected and the similarities between crime types have to be found out. Figure 4.4 shows the similarity among various types.
Classification of Crime Time

The moment crime case is registered the case record has to be collected and the similarities between crime types have to be found out. Figure 4.5 shows the similarity among various times.

4.9 DESIGN OF PROPOSED ALGORITHM

The existing algorithm almost finds the classifications of crime attributes (Xiang Zhang 2010). But the classification of class attributes has not been not clear as it concentrates on top priority crimes only. In the structured crime classification, to forecast a particular crime of the defined class, to find the probability of the crime occurrence of a particular class crime and find the positive and negative characteristics of the crime where the class of crime is defined by the similarity between the crime events from the database. If the character is positive, KDE map produces a hotspot and if it is negative, it produces a cold spot.
In this method, we assign the count value defined initially as 1. The value is the number of incidents in hotspots. The heuristic method was used to solve the problem. The threshold value plays the major role of this method and if the same incidents occur in a number of times, these incidents are saved directly in the crime hotspot. Additionally, if the number of incidents is greater than the threshold value, after the generalization of an attribute by the classification of the attribute, the crime attribute should be completely up to a hierarchy until the number of crime incidents of the attribute is less than the threshold value. And then the crime attributes with a greater than same value is preferred at the end. Finally, to avoid generalizing, the generalization of all the crime attributes should be terminated once a hot spot is produced.

Let $S$ denote a set of crime incidents. Let $A_i$ be an attribute of crime incidents and $C_i$ be a classification of each crime attribute $A_i$. For two elements $x_1$ and $x_2$ in the tree of $C_i$, if there is a path from $x_1$ to $x_2$, $x_1$ is called the parent of $x_2$. Furthermore, $x_1$ is a generalization of $x_2$. In the structure crime classification algorithm, the notional dissipation is used as an alternative of similar. The dissipation between the events is smaller and the events are more similar. Choose the crime attribute $A_i$ in the crime class $C$. 

Figure 4.5 Classification for Crime Time
Find the similarity of each crime attribute of crime objects if both the objects have same similarity. Join, the two objects have the same crime attribute incident and put into the same class C. And finally, find the F(C) based on the probability of crime incident occurring in the particular class to which it is merged. If F(C) is greater than the positive description, it produces a crime hot spot; otherwise it produces a crime cold spot.

Algorithm:

Begin:
S=DB //Assign The Data
For (each Si in DB)
S.Count=1 //initialize the count
For (each Si)
Apply Purification Attribute Ai by Cn
While (events in DB (S) can be generalized) do
Choose an attribute Ai by crime class Ci
For (each Ci)
C.Ai = Objects (Ci, C.Ai)
C.Ai = Similarity (C.Ai, C.Ai+1) //Similarity of crime attribute Objects
Classi = Probability (C.Ai, Classi)
Threshold T= (Cluster Area - Sparse Area)
For (random Class1, Class2)
If (Class1=Class2)
Delete C2 //join the same events
Probability Pi = (Ai, DB)
Incident Class F(C) = Classi U Pi (Ci)
If F(c) > Positive Description
Produce a hot spot
Else
Produce a cold spot
End while
End
Figure 4.6 Algorithm Flow Design
Step by Step process

Step 1. Start

Step 2. Collect the crime Record and denote DB.

Step 3. Assign DB=|DB| Denote the number of crime events.

![Figure 4.7 Crime DB Denoted in Hotspots](image)

**Figure 4.7 Crime DB Denoted in Hotspots**

Step 4. Assign $C_i =$Similarity (Obj$_1$, Obj$_2$, Obj$_n$) where Obj$_{1,n}$ are Crime Events of the DB.

The following objects are Peoples intelligence recorded in a given geographical area at a specified time. Temporal data include 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 a particular crime $P_i = \text{requested}_f \text{ Dataset crime, Crime Dataset}$, record the Positive Character, the Negative Character descriptions of the Particular Crime.

Figure 4.8 Class Finds in Similarity of Objects

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

Figure 4.9 Classification of Density and Sparse Area
**Step 7.** To Identify the Classes of crime types, classes based on the kind of social impact of the criminal activity. Ex- Class1 = (Burglary in Apartments, Shops, Small Houses), Class2 = (Robbery, Injury, Property damage), Class3 = (sex, children abuse, usurp)…, etc.

<table>
<thead>
<tr>
<th>Time</th>
<th>Month</th>
<th>Place</th>
<th>Class1</th>
<th>Class2</th>
<th>Class3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Period1</td>
<td>May</td>
<td>Adiyar</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Period3</td>
<td>December</td>
<td>Vadapalani</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Period4</td>
<td>January</td>
<td>Kottur</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Period2</td>
<td>November</td>
<td>Amjakarai</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Period1</td>
<td>December</td>
<td>Sadaipet</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**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 have been made.

**Step 9.** If \( F(Class_i) \geq \) positive description

Then

Produce a Hotspot.

Else

Produce a Cold spot.
Step 10. End.

4.10 STRUCTURED CLASSIFICATION ALGORITHM FOR CRIME CLASSIFICATIONS

Data mining is developing and determining data from large amounts of data stored in databases. Data mining is a suitable and progressively more important tool to convert these data into information. There are many methods used in data mining to extract patterns for large amount of spatial databases. To forecast group relationships for data, instances classification is a powerful method used in data mining. Classification is a two way process: viz. construction of predetermined class and classifying features of unknown class objects.

Construction of Predetermined Class

Each object is assumed to belong to a predetermined class, as determined by class label attribute. The training set refers to set of objects used for constructing predetermined class.
Classifying Features of Unknown Class Objects

This is used for finding the accuracy of the algorithm. The known class objects of test sample are compared with classified results from the algorithm.

In data mining a number of classification algorithms are used. In this study the method of decision tree induction is taken. It is for crime mapping in GIS.

Tree Induction Method

In this method tree classification is the knowledge of decision trees from class labeled training objects. A decision tree is a tree structure like the flow chart. In each internal node test in a class of objects, branch represents the outcome of the test and leaf holds the class label.

Advantages of Tree Induction Method

Comparing with other data mining classification techniques, the tree induction method is easy to understand. It can handle both numerical and unconditional data. It is likely to validate an algorithm using geometric tests. They are tough in nature. Therefore, they perform well even if their hypothesis are somewhat violated by the true algorithm from which the data have been generated. Decision trees perform well with large data in a short time.

Function Setting

In this study of the structured classification algorithm, the function setting allows a user to specify the type of problem to be solved. In the function setting, objects contain high-level parameters for solving the
algorithm. In the mining function setting, objects is specified into three categories like specific parameters for classifications, logical data specification and data usage specification.

4.11 PROCESS OF STRUCTURED CRIME CLASSIFICATION ALGORITHMS

<table>
<thead>
<tr>
<th>Data Input</th>
<th>Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police reports (Tamil Nadu police dept.)</td>
<td>Geocoding database</td>
</tr>
<tr>
<td>Geocoding database (GIS mapping dept.)</td>
<td>Database per incident and year</td>
</tr>
<tr>
<td></td>
<td>Tamil Nadu Reporting</td>
</tr>
<tr>
<td></td>
<td>District map</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Geocoding events</th>
<th>crime database</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>KDE maps</td>
<td>crime classification</td>
<td>Class</td>
</tr>
</tbody>
</table>

**Figure 4.11 Process of Structured Crime Classification Algorithm**

Software / tools used:

- ESRI ArcView 3.1
- Microsoft Excel
- Advanced JAVA
- Graphical Information Systems

**Execution Software Specification**

The implementation is done with advanced Java programming language for creating the application and Mysql databases for the storage space. Due to the convenience and simplicity to find out and as it has connected with current analysis, each of the approaches is implemented with
identical data and concept hierarchy primarily based on data. Examples in the application are running on the computer with the following specifications: Intel Pentium four, 2.20 GHz, 644 MHz and 1 GB of RAM. The software we used in this structured classification classified the data and mapped the incidents into the KDE map using geographical information system.

**Datasets and Data Objects Functionality in the Advanced Java**

A physical data specification object specifies the characteristics of the physical data to be used for mining. There are two types of data format used in physical data specification. Data in a multi-record case format is called transactional and single-record case is called nontraditional. This type of data object can be used as input to different tasks.

**An html Imagemapper Tool for Map Creating**

Html ImageMapper, is a mapping tool, developed by Alta 4 Geoinformatik AG. It is a tool to convert arcview maps and data into interactive maps. The program code converts arcview map documents to HTML Image Maps. Html imagemapper has some possibilities of zooming the map, search the data in the map, as the crime objects can be searched and identified in the map easily. Html ImageMapper produced a map containing multiple layer of information. It is operated in client side only and it is not relevant to the server side installation. In this study Html imagemapper was used to investigate the possibilities to convert map data into interactive maps.

**Arcview Tool for Hotspot Analysis and Clustering**

ArcView is a special tool used for crime mapping and crime analysis. In this study, ArcView tool is used for the programming interface with most desktop GIS programs. The ArcView presents additional arithmetic
tools to aid police operations in their crime forecasting efforts. ArcView includes tools to find the standard deviation and center point medium for distribution of crime. The tool provides some special functionality for “crime hotspot” analysis and “crime clustering”.

**Microsoft Office Excel Tool for Crime Mapping**

The application of MS Office in a crime analysis and crime mapping program, should not be under-estimated. In most of the researches, Microsoft Access is a common system for database management. It is mainly used for storage purpose. We can create a number of templates and shortcuts too can be programmed for data entry. For customizing, a user friendly interface can be developed for police personal for entering the crime information.

Microsoft Excel presents a number of arithmetic statistical functions that can be used in a crime analysis and crime mapping program. Data from ArcView can simply be exported to MSExcel to perform an arithmetic analysis and then be imported again into ArcView. Operations that can be performed with Microsoft Excel include the calculation of an analysis to preferred days and times of criminal incidents, the calculation of the comparative change in the number of crime incidents, or the result of development in the number of crime incidents. MSExcel also provides the possibility to visualize the results of arithmetic operation with different types of graphs and charts. For this study MSExcel was the main tool that was used to complete the arithmetic part of this analysis.

**4.11.1 Data Collection and Formation for Input**

To perform a spatial and arithmetic analysis of crime in Chennai city, data have been obtained from different resources. The following resources of crime have been used:
1. Tamil Nadu police department crime record
2. Tamil Nadu police department communications center
3. Tamil Nadu special crime branch records
4. The various crime detection agencies’ records

In this study, Chennai city crime record is used which is maintained by police department headquarters. Within the Chennai city Police Department, all crime reports and records are entered into one common crime database. The crime information is collected from various sources and stored in the MSExcel database. Additionally, the police department maintains the data in map format to easily identify the city limits and the crime zones and for planning the areas for police patrol.

**Data Preprocessing**

Data preprocessing is the process of putting the collected data into a data structure. Before geocoding, the provided data have to be sorted out. The data have to be verified, completed and updated.

The crime data, provided by the Chennai city police department has to be organized and classified. The initial information contained everything that was ever entered into the system, as well as data that were not useful for this study. For this study, the data on the crime objects are classified into crime classes and different tables were made for throughout the year 2008, which included the following information: viz. the crime report number, description of the crime incident, the place of happening and the time of happening.

When comparing the crime rates of south Chennai and north Chennai geographic areas, most crimes occur in south Chennai area, based on
the police record. Facts that were taken into consideration were density of population, size of the area and the type of the crime event. Smaller convergences were defined for the consideration south Chennai area. Since crime rates are higher in areas with commercial activity, these areas have been focused on.

4.11.2 Crime Characterization and Type

It is necessary to know the important options of crime and therefore the criminal justice system. Crime could be a complicated concept that may be outlined in each in terms of legal and non-legal logic. From a legal point of view, it refers to breaches of the criminal laws that govern explicit geographic areas (jurisdictions) and are geared toward protecting the lives, properties and rights of the people living in those jurisdictions. From a non legal point of view, it would outline crime as acts that violate socially accepted rules of human moral or ethical behavior.

The collected spatial crime data are used in MySQL database architecture which will have DBMiner structural design. DBMiner incorporating a number of data mining technologies and arithmetic analysis will create multilevel mining rules for classifying the data attributes (Han et al 1992). There are two strong reasons to select the DBMiner, JAVA and MySQL for implementing the structure classification algorithm.

1. Mining new kinds of rules from large databases include multiple level association rules, classification rules, cluster description rules and prediction.

2. SQL-like data mining query language DMQL and Graphical user interfaces have been enhanced for interactive knowledge mining.
The initial step of the implementation in generalizing the crime attributes. In the main crime database DB, S1, S2, ……, Sn is the set of crime incidents. \{A1, A2, A3, ……, An\} are attributes of (S1, S2, S3…. Sn). Cn represents attributes after the generalization is represented.

**Table 4.2 Crime Type and Description**

<table>
<thead>
<tr>
<th>Crime Type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burglary</td>
<td>The illegal entry into a structure to commit an offense or theft.</td>
</tr>
<tr>
<td>Shoplifting</td>
<td>The illegal entry into a shop to commit a theft.</td>
</tr>
<tr>
<td>Auto theft</td>
<td>The crime of stealing an automobile.</td>
</tr>
<tr>
<td>Robbery</td>
<td>Robbery is taking or making an attempt to acquire something of worth from the possession of an individual or persons by force or threat of force or violence and by putting the victim in concern.</td>
</tr>
<tr>
<td>Abuse</td>
<td>The crime of stealing of the cheap things.</td>
</tr>
<tr>
<td>Pickpocket</td>
<td>In crowded places it usually happens. Someone takes things from people's pockets.</td>
</tr>
<tr>
<td>Fraud</td>
<td>Money based crime. It is about fake exchange and getting money</td>
</tr>
</tbody>
</table>
Table 4.3 Crime database 2008

<table>
<thead>
<tr>
<th>Areas</th>
<th>Robbery</th>
<th>Burglary</th>
<th>Fraud</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Personal Property</td>
<td>Business Property</td>
<td>Total</td>
</tr>
<tr>
<td>Adayar</td>
<td>190</td>
<td>55 245</td>
<td>208 415 623</td>
</tr>
<tr>
<td>Ambattur</td>
<td>181</td>
<td>34 215</td>
<td>342 260 602</td>
</tr>
<tr>
<td>Aminjikarai</td>
<td>172</td>
<td>17 189</td>
<td>600 250 850</td>
</tr>
<tr>
<td>Anna Nagar</td>
<td>216</td>
<td>28 244</td>
<td>375 89 464</td>
</tr>
<tr>
<td>Ashok Nagar</td>
<td>164</td>
<td>14 178</td>
<td>275 159 434</td>
</tr>
<tr>
<td>Choolai</td>
<td>156</td>
<td>17 173</td>
<td>171 128 299</td>
</tr>
<tr>
<td>Kolathur</td>
<td>117</td>
<td>18 135</td>
<td>338 114 452</td>
</tr>
<tr>
<td>Kottur</td>
<td>251</td>
<td>35 286</td>
<td>335 149 484</td>
</tr>
<tr>
<td>Mylapore</td>
<td>370</td>
<td>49 419</td>
<td>232 154 386</td>
</tr>
<tr>
<td>Pallavaram</td>
<td>240</td>
<td>40 280</td>
<td>165 116 281</td>
</tr>
<tr>
<td>Porur</td>
<td>239</td>
<td>44 283</td>
<td>210 73 283</td>
</tr>
<tr>
<td>Saidapet</td>
<td>133</td>
<td>15 148</td>
<td>218 102 320</td>
</tr>
<tr>
<td>Velacheri</td>
<td>100</td>
<td>8 108</td>
<td>257 159 416</td>
</tr>
<tr>
<td>Villivakkam</td>
<td>143</td>
<td>21 164</td>
<td>201 105 306</td>
</tr>
<tr>
<td>West Mambalam</td>
<td>102</td>
<td>28 130</td>
<td>218 96 314</td>
</tr>
<tr>
<td>T-Nagar</td>
<td>87</td>
<td>21 108</td>
<td>145 48 193</td>
</tr>
<tr>
<td>Manali</td>
<td>78</td>
<td>13 91</td>
<td>251 123 374</td>
</tr>
<tr>
<td>Meenambakkam</td>
<td>233</td>
<td>45 278</td>
<td>151 57 208</td>
</tr>
</tbody>
</table>
Figure 4.12 shows the result when the program was run by inputting generalization with 1. Figure 4.13 shows the result when the program was run by inputting generalization threshold with a 2. However, the ultimate generalization result did not fulfill the generalization strategy step for threshold management on generalized relations. Where the amount of tuples of a generalized relation within the target category is larger than the generalization threshold worth, additional generalization ought to be performed. In target class generalization result, except crime class attribute, the other three crime attributes which have generalized in concept hierarchy and will be used as selected attribute for further generalization.

![Structured Classification Algorithm](image)

**Figure 4.12 Classification Rule for Classification Algorithm Program for Generalization = 1**

**Generalization 1**

**For** (each crime incident Si)

S.Ai =parents (S.Ai) //generalize crime attribute

**For** (chance S1, S 2)

**If** (S1=S2) Delete S2  //combine the similarity of the same incidents
Based on the generalization 2, the crime classes are defined by the similarity between the crime objects.

**Figure 4.13 Classification Rule for Classification Algorithm Program for Generalization = 2**

**Generalization 2**

Class = combine (Table 4.3: crime objects1, crime object2 …)

In the generalization 2, the crime objects are collected in crime database Table 4.3. The same similarity objects are combined and put into the crime class based on the structure classification algorithm.

The result of the generalization 1 and 2 is shown in Table 4.4. The main database table contains the number of classifications for the particular crime. Table 4.4 shows the classifications of the crime objects combined and formed into a single class. The classes are represented in class1, class 2…, Etc.
Table 4.4 Crime Class Classification based on Generalization 1 and 2

<table>
<thead>
<tr>
<th>AREAS</th>
<th>Robbery</th>
<th>Burglary</th>
<th>Fraud</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CLASS C1</td>
<td>CLASS C2</td>
<td>CLASS C3</td>
</tr>
<tr>
<td>Adayar</td>
<td>245</td>
<td>623</td>
<td>918</td>
</tr>
<tr>
<td>Ambattur</td>
<td>215</td>
<td>602</td>
<td>199</td>
</tr>
<tr>
<td>Aminjikarai</td>
<td>189</td>
<td>850</td>
<td>124</td>
</tr>
<tr>
<td>Anna Naga</td>
<td>244</td>
<td>464</td>
<td>143</td>
</tr>
<tr>
<td>Ashok Nagar</td>
<td>178</td>
<td>434</td>
<td>160</td>
</tr>
<tr>
<td>Choolai</td>
<td>173</td>
<td>299</td>
<td>289</td>
</tr>
<tr>
<td>Kolathur</td>
<td>135</td>
<td>452</td>
<td>187</td>
</tr>
<tr>
<td>Kottur</td>
<td>286</td>
<td>484</td>
<td>379</td>
</tr>
<tr>
<td>Mylapore</td>
<td>419</td>
<td>386</td>
<td>175</td>
</tr>
<tr>
<td>Pallavaram</td>
<td>280</td>
<td>281</td>
<td>125</td>
</tr>
<tr>
<td>Porur</td>
<td>283</td>
<td>283</td>
<td>176</td>
</tr>
<tr>
<td>Saidapet</td>
<td>148</td>
<td>320</td>
<td>293</td>
</tr>
<tr>
<td>Velacheri</td>
<td>108</td>
<td>416</td>
<td>278</td>
</tr>
<tr>
<td>Villivakkam</td>
<td>164</td>
<td>306</td>
<td>156</td>
</tr>
<tr>
<td>West Mambalam</td>
<td>130</td>
<td>314</td>
<td>110</td>
</tr>
<tr>
<td>T-Nagar</td>
<td>108</td>
<td>193</td>
<td>165</td>
</tr>
<tr>
<td>Manali</td>
<td>91</td>
<td>374</td>
<td>241</td>
</tr>
<tr>
<td>Meenambakkam</td>
<td>278</td>
<td>208</td>
<td>117</td>
</tr>
</tbody>
</table>

Generalization 3 to find the probability of crime based on Generalization 2, apply the condition probability= (requested crime, crime database DB). In this Generalization 3 also store the positive and negative characteristics of the request crime. It is used in the class to which the particular crime belongs to.
4.12 THE SPATIAL CRIME DATA

Geographic position specifies the situation of a feature or phenomenon by employing a coordinate system, which is used for representing geographical places, crime incident area information, moveable and unmovable things about the crime place, etc.

The Attribute Crime Data

Attribute data talk of the properties of spatial entities that describe the characteristics of the geographic features: like the population of the area, latitude and longitude information of the incident area. Based on this basic geographic information, this study focuses on the positive and negative characteristic information on the crime attribute to be collected. For example let us assume that burglary happened in the house. The police department collects the information of the particular incident and stores in the police system. In this study let us apply the 3rd generalization on this collected data and find the positive and negative characteristics of the stored data.
**Positive Characteristics**

**Table 4.5 Identification of Positive and Negative Characteristics of Crimes Class based on Generalization3 Crime database 2008**

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Area</th>
<th>Positive</th>
<th>Negative</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-Jun</td>
<td>8 pm</td>
<td>Adayar</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12-May</td>
<td>3 am</td>
<td>Ambattur</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18-Dec</td>
<td>11 pm</td>
<td>Aminjikarai</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28-Nov</td>
<td>7.00 am</td>
<td>Anna Naga</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>10-Jan</td>
<td>12.30 pm</td>
<td>Ashok Nagar</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>16-Apr</td>
<td>5 pm</td>
<td>Choolai</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19-Feb</td>
<td>10.30 pm</td>
<td>Kolathur</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>30-Mar</td>
<td>5.45 am</td>
<td>Kottur</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>28-Jun</td>
<td>11.40 pm</td>
<td>Mylapore</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>30-Aug</td>
<td>1.30 am</td>
<td>Pallavaram</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>14-Jul</td>
<td>11 am</td>
<td>Porur</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>19-Sep</td>
<td>3.30 am</td>
<td>Saidapet</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5-Nov</td>
<td>5.00 am</td>
<td>Velacheri</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>30-Oct</td>
<td>8.20 pm</td>
<td>Villivakkam</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>25-Dec</td>
<td>12.20 am</td>
<td>West Mambalam</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>22-Oct</td>
<td>9.00 am</td>
<td>T-Nagam</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19-Oct</td>
<td>11.40 pm</td>
<td>Manali</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23-Aug</td>
<td>6.00 am</td>
<td>Meenambakkam</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

The stored data satisfy the geographic data elements which should be treated either as positive or negative characteristics.
As the result of the generalization 3 shows the burglary in shop the positive satisfies the requirements of the place of occurrence of crime, time of occurrence of crime, incident date, etc. But, the negative shows the type of weapons used in the incident is also stored but it is not needed much for the analysis. In this way the entire recorded crime data are identified as positive and negative. Probability =1, the requested crime contains the positive characters that match with the geographic data elements. Otherwise it is negative character.

The result of the generalization 3 is shown in Table 4.5. The crime database is classified into a month date and time of occurrence of crime and the specific area or location. The two additional columns needed for the classification of crime attributes, are the geographic region or other data. Whenever the positive is 1 it comes under the category of geographic data elements. It is represented in the yellow field in the table 4.5. C1, C2, C3 is represented in crime class. The classes are predefined in the result of generalization 1 and 2. The class fields represent 1, if the happened crime comes under the category of this class.

4.13 TEST RESULTS

Time based Crime Hotspots

From the result of the structure crime classification, it is understood that there are a number of time related incidents, such as a few incidents that happen in the recreation site at the time of weekend holidays, in summer time, assassination in the same site, assault incidents happen in the night time in summer, the vehicle theft that happen in transport site usually in the holidays and working days, … etc. Personal injury cases happened in the night time of summer frequently. So such incidents will be greatly if minimized the police department increases the patrol in the midnight hour.
Place based Crime Hotspots

The result of the test shows that there is a group of crime incidents directed at the busiest areas at some point of time. They have been classified as bus stand, railway station, service site where the people are always gathered at any time, any day, every weekend and the holiday times for every year and so on. The relationship between the different sites and fraud cases are clear that these sites are the places of hot spots. Considering these, the police department must add the deployment of more police forces in these areas.

Class based Crime Hotspots

As a result of the test, crime incidents are classified into a number of crime classes. Each class contains a number of objects and more positive similarities. The crime class is identified as C1, C2, C3… etc. The police forces collect the class and perform the forecasting for the particular crime.

4.14 CREATING CRIME HOTSPOT AND COLD SPOT

After collecting the database from the structured classification algorithm, the result contains the crime Classes and positive and negative fields of data. During this study, the condition for finding crime \(F(C) >\) Positive is applied. It is made as crime hotspot or else it is represented as a cold spot. To implement this operation, this study has used Arcview and spatial analyst with Avenue programming.

Avenue programming could be an Object oriented programming language, utilized in Arcviews. It is mainly used for developing Arcview applications. It was developed by University of Durham information Technology service, 2007. During this programming, three different
Methodologies have been used to examine the spatial and temporal data. The present study has used the method to count the number of nearby crime incidents for every event. The recent spots are created from data from the police crime record.

In the Crime Event Counts, against the crime hot spot is outlined because the space where the number of nearby incidents is high. The primary question in addressing this definition is a way to select “near” events from every incident. One extreme is when “near” is merely for the placement of every event itself, thus each event is a hot spot. The other extreme is once we take all events, with the result that no hotspot space will be outlined.

The basic idea of counting the quantity of events or measuring event density during this methodology is comparable to the approach utilized in the STAC! Program. This was introduced by the Illinois Criminal Justice info Authority, in 1993; however it is totally different the approach to delineate the hot spots. Within the STAC! Program, the user defines a grid size and then divides the areas into this grid. A circle is formed on each node and then the program counts the number of incidents in every circle. The circles within the prime half an hour in terms of incident numbers are then outlined as hot circles. These steps are represented for larger size of grid to induce the second set of hot circles. The combination of the two hot circles can provide an ellipse or the recent spot with the greatest variety of incidents.

The method employed in this study counts the number of events at intervals and the exact distance from every event and then stores this count in that event. Then it uses 3000 feet to represent the scale of three- to five-blocks of an urban space. Once the numbers of surrounding events are counted from every event, the subsequent step is to use spatial interpolation between events so as to form an endless surface showing the relative level of hot spots within the space. In Avenue, the most part of this methodology is written as follows:
for each frompt in theVTab
fromShape=theVTab.ReturnValue(theShapeField,frompt)
totalEvents=0
for each topt in theVTab
toShape=theVTab.ReturnValue(theShapeField,topt)
theDistance=(FromShape.Distance(toShape))
if (theDistance < 3000) then
    totalEvents=1+totalEvents
end
end
end

Figure 4.15 Hotspot and Cold spot

The cyan color represents hotspot and the blue color represents the
cold spot.
4.15 RESULT COMPARISON

The proposed algorithm finds the crime class based on the crime object. There are three important major features measured in this study, which are compared with the existing approaches. Speed is an important feature based on execution time. Accuracy is calculated on the basis of how algorithm flexibility incorporates with the other domains.

Table 4.6 Comparison of Several Features in Classification

<table>
<thead>
<tr>
<th>Feature</th>
<th>NB</th>
<th>ABN</th>
<th>SVM</th>
<th>DT</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Processing Time</td>
<td>1</td>
<td>0.55</td>
<td>1.05</td>
<td>1.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Accuracy</td>
<td>Good in many domains</td>
<td>Good in many domains</td>
<td>Significant</td>
<td>Good in many domains</td>
<td>Significant</td>
</tr>
<tr>
<td>Transparency</td>
<td>No rules</td>
<td>Rules for Single Feature Build only</td>
<td>No rules</td>
<td>Rules</td>
<td>Rules</td>
</tr>
<tr>
<td>Missing value interpretation</td>
<td>Missing value</td>
<td>Missing value</td>
<td>Sparse data</td>
<td>Missing value</td>
<td>Sparse and geographic data</td>
</tr>
</tbody>
</table>

It compares the four algorithms with our proposed algorithm.

1. Naive Bayes (NV)
2. Adaptive Bayes Network (ABN)
3. Support Vector Machine (SVM)
4. Decision Tree (DT)

Compared to all the four algorithms shown in Table 4.6, the proposed algorithm has quick processing time.
This chapter has presented an approach for detecting the crime hotspots by structured classification method. For this purpose, some kinds of crime cases have been analyzed for classification. This method is not as same as the traditional spatial classification. It focuses on preprocessing the crime events before mapping. The advantage of this method is that the precise information of hot spots is more apparent. But many researches about this method will be explored in the future. The classification of crime attribute is the key related to the accuracy of the results.

The output of the structured classification creates different type of hotspots based on Chennai city crime record. It should be used for the police department to identify the particular crimes that occur frequently. Figure 4.16 shows the execution time of structured classification which is better than the other algorithms.