International Journal of Computing Academic Research (IJCAR)
ISSN 2305-9184 Volume 3, Number 2(April 2014), pp. 58-69

http://www.meacse.org/ijcar

Crime Classification Algorithm for Mining Crime Hot Spot and Cold Spot

M. VijayaKumar[1], P.Balamurugan [2], Basim Alhadidi [3] and Hanumanthappa[4]

[1]Department of Computer Applications ,K.S.R.College of Engineering, Tiruchengode, India.
[2]Department of computer science, K.S.R College of Engineering, Tiruchengode, India.
[3]Department of Computer Science,AlBalqa' Applied University, Salt, Jordan.
[4]Department of Computer Science & Applications, Bangalore University,Bangalore, India.

Abstract

Crime may be a behavior deviating from traditional violation of the norms giving peoples' losses and harms. Social, psychological, economic and environmental factors are to be thought-about in crime issues. Crime is tried to be explained by numerous theories from totally different sciences. Social and psychological theories contemplate the foundation causes of crime noticing to factors like social disorganization, temperament disorders and inadequate parenting etc. In this work, we discuss the groundwork results of a crime forecasting model developed in association with the Tamil Nadu police department. First part of the research work to collect the crime records from various police departments and rearrange the crime database. The space and time of these crime incidents are implanted in the data. Additionally spatial and temporal characteristics are yielded from the crime data. The second part of the research performs the crime forecasting based on data mining classification method. In this research analyze various classification methods which are predicting crime hotspots more accurately. The final result of the research model shows the improvement of spatial and temporal crime data to create consistent crime forecasting.

Key Words: spatial-temporal crime data, crime forecasting, classification, crime hotspot and cold spot.

Introduction

In a crime and law enforcement agency, the central focus is on crime, both those reported to the police and those that are not. Thus, the central type of data analyzed is crime and the information surrounding it, such as arrests, offenders, victims, property, evidence, etc. A large range of techniques is contributing for the mining of spatial databases. These techniques give results regarding spatial classification, spatial trend detection and spatial clustering. The aim of spatial data mining is to reveal unknown data. Two issues associated with clustering that have received attention in the past years include: 1. Generalization-based characterization applied to spatial information proposed in; 2. a cluster characterization primarily based on spatial options like lakes, resorts etc.

It is a rapidly evolving space of analysis that is standard in many disciplines as well as statistics (e.g., time series analysis), temporal pattern recognition, temporal databases, optimization, high-performance computing, and parallel computing. The temporal data mining element of the Knowledge Discovery in Temporal Databases method is bothered with the algorithmic means that those temporal patterns are extracted and counted from temporal knowledge. A comprehensive overview of the techniques is presented for the mining of temporal data. Main temporal data mining techniques that are commonly derived by either bottom-up or top-down induction are: temporal classification (basic goal is to predict temporally connected fields in an

exceedingly temporal database primarily based on different fields); and temporal clustering. In temporal analysis, several temporal data mining applications create use of clustering. There are two basic approaches to temporal clustering and analyzing it. One is that the live temporal similarity approach and the other is a temporal optimal partition approach. If the amount of clusters is given, then clustering techniques will be divided into three classes: (i) Metric-distance primarily based technique, (ii) Model-based technique and (iii) Partition-based technique. If the amount of clusters is not given, then we are able to use Non-hierarchical Clustering Algorithms to seek out their k. It is necessary to develop the good crime forecasting to analysis the crime data based on spatial and temporal data mining In this research focus on burglary crime which is identified by the Tamil Nadu city police department. The goal of the research discovers a spatial classification and hotspot identification reliability.

Data Architecture

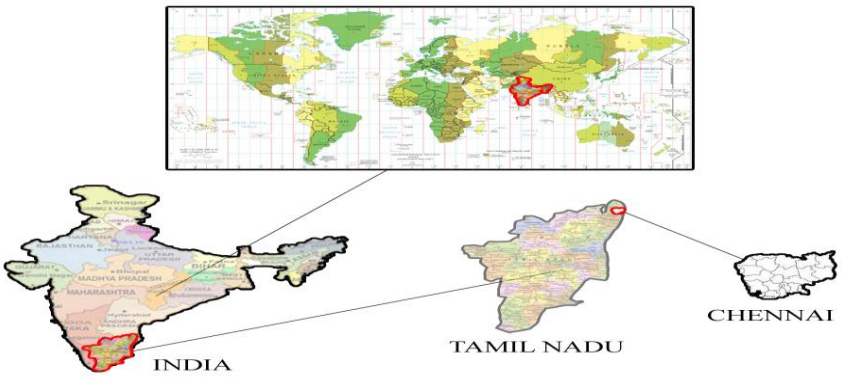

Fig .1 The Study Region

Crime data of year 2008 in the study area are used in the analysis which is illustrated in Fig.1 Spatial and temporal information related to these incidents were obtained from Police headquarters. 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 Tamil Nadu 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 Database, there have been totally 3,706,34 crime records in 2008.

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 plan 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.

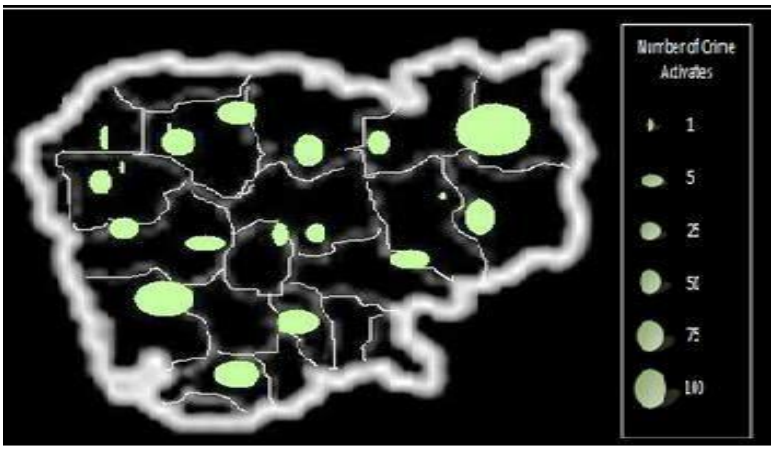

Fig. 2 Criminal Activities in Study Region

Fig. 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.

Literature Survey

Crime Data mining

Crime Data mining can almost exactly be described as mining up information from crime databases. Data that was never explicitly put in is brought to the surface by advanced methods. Without data mining, crime data analysis is partial to a small number of selected aspects. By way of data mining, all potential relations within the crime data are ensured. By using (AI) artificial intelligence, mathematical and statistical methods and visualization, crime relations are discovered that would otherwise be easily ignored. In addition, all crime relations that are found can be applied to make predictions.

Crime Data mining is defined as the classification of interesting structure in data, wherever structure assigns patterns, mathematical or predictive models of the data, in addition to relationships among parts of the data [5]. Crime Data mining in the background of crime and intelligence analysis for national safety is still a juvenile field. Chau et al., 2002 used different technique in crime data mining. Entity extraction has been used to identify person, address, vehicle, and personal properties from police narrative reports automatically.

The general crime data mining tasks can be classified into two main classes: descriptive data mining and predictive data mining. The [7] concisely describes the behavior of datasets and presents interesting general properties of the data. Numerous attempts have used automated techniques to analyze different types of crimes, but without a unifying framework describing how to apply them. In particular, understanding the relationship between analysis capability and crime type characteristics can help investigators more effectively use those techniques to identify trends and patterns, address problem areas, and even predict crimes [8].

Crime Classification on both Temporal and Spatial data mining

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 [9].

In the crime analysis research classification of incidents of murder, rape, and arson from the macro- to micro- level and used several principles that would assist analysts in developing classifications for different crimes. Three forms of classifications highlighted that this might embrace decision kind, crime, and uniform crime reports [10].

In the Classification method which consists of two parts. First, the user defines the number of classes. To test, whether the number of classes was chosen correctly, a set of training data is selected and the classification is performed on it. Subsequently, classification rules are derived from the training dataset. Next, those rules are applied to the test dataset. Classification is considered as predictive spatial data mining, because we first create a model according to which the whole dataset is analyzed [11].

Temporal classification involves predicting the class of a temporal sequence. This is done by looking at previously classified temporal data, building a predictive model based on that data, and then using it to predict the category of unclassified sequences in the data [12].

That Changes in the way in which police record crime also affects figures. The police station officers counting the rules for the count and classifying of notable offenses recorded by the police forces. The classification was created a number of censes variables and classifies every output area [13].

Structured Classification Algorithm

Introduction

In this paper, it has been focused on how to incorporate temporal crime data with STEM model [1]. The disadvantage of the STEM model is a classification of attributes. The proposed paper fulfills the seasonal dependency of crime data attributes, additionally the attributes is 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, 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 research, 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.

Data mining is developing of 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 method 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.

Limitations of structured classification algorithm

Classification techniques have been established to be very successful for crime analysis and crime forecasting. It might be applied to categorize crime data. The distributed data mining algorithm uses a practical value algorithm to calculate C4.5 algorithm, CART algorithm and naïve Bayesian classification algorithms. The technique was applied to crime attribute classifications. The neural mining approach uses rule-based association rules to mine representative data. The approach discusses the importance of use of existing non-numeric data in crime analysis and crime forecasting.

Nowadays the foremost of the researchers, used spatial clustering for identifying crime hotspots. The spatial ellipses application technique is a difficult technique for distinguishing crime hotspots, encompasses a long tradition in mapping analysis. Find category hierarchy of the crime data may be a difficult task for the recorded crime, additionally this paper 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.

Structured classification method

Structured crime classification technique is a comparatively straightforward technique of the concept spatial clustering. During this classification, it aims to detect 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. Finally apply a threshold value to seek out the crime hotspot and crime cold spot. Before that, the study of the crime incident analysis is necessary.

The analysis of crime incidents

In this paper 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. Every country followed some 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 loose, whether the crime occurred in control area or other area, what is the specific report of the crime etc. They come across small difficulties. But along with them some are essential attributes, such as time, place, numerical and crime category attributes [3].

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 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 places
2. Classification of offense types.
3. Classification of Crime occurring time.

Design of Crime Classification Algorithm

The existing algorithm almost finds the classifications of crime attributes [3]. But the classification of class attributes was not clear as it concentrates on top priority crimes only. In the structured crime classification, to forecast a particular crime of the defined class, 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, 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 CARIME DATA denote a set of crime incidents. Let Ai be an attribute of crime incidents and Classificationi be a classification of each crime attribute Ai. For two elements x1, and x2 in the tree of Classificationi, if there is a path from x1 to x2, x 1 is called the parent of x2. Furthermore, x1 is a generalization of x2. 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 Ai in the crime class C. 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(Crime Class) based on the probability of crime incident occurring in the particular class to

which it is merged. If F(Crime Class) is greater than the positive description, it produces a crime hot spot; otherwise it produces a crime cold spot.

Algorithm:

Begin:
CRIME DATA = DB //Assign The Data
For (each CARIME DATA i in DB)
CARIME DATA.Count=1 //initialize the count
For (each CARIME DATAi)
Apply cleansing Attribute Ai by Cn
While (events in DB (CARIME DATA) can be generalized) **do**
Choose an attribute Ai by crime class Classificationi
For (each Classificationi)
C.Ai = Objects (Classificationi, C.Ai)
C.Ai = Similarity (C.Ai, C.Ai+1) //Similarity of crime attribute Objects
Crime Classi = Probability (C.Ai, Crime Classi)
Threshold T= (Crime Cluster Area- Crime Sparse Area)
For (random Crime Class1, Crime Class2)
If (Crime Class1= Crime Class2)
Delete Crime C2 //join the same events
Probability Pi = (Ai, DB)
Incident Class F(C) = Crime Classi U Pi (Ci)
If F(c) > Positive Description
Produce a hot spot
Else
Produce a cold spot
End while
End

Flow Design of the Algorithm

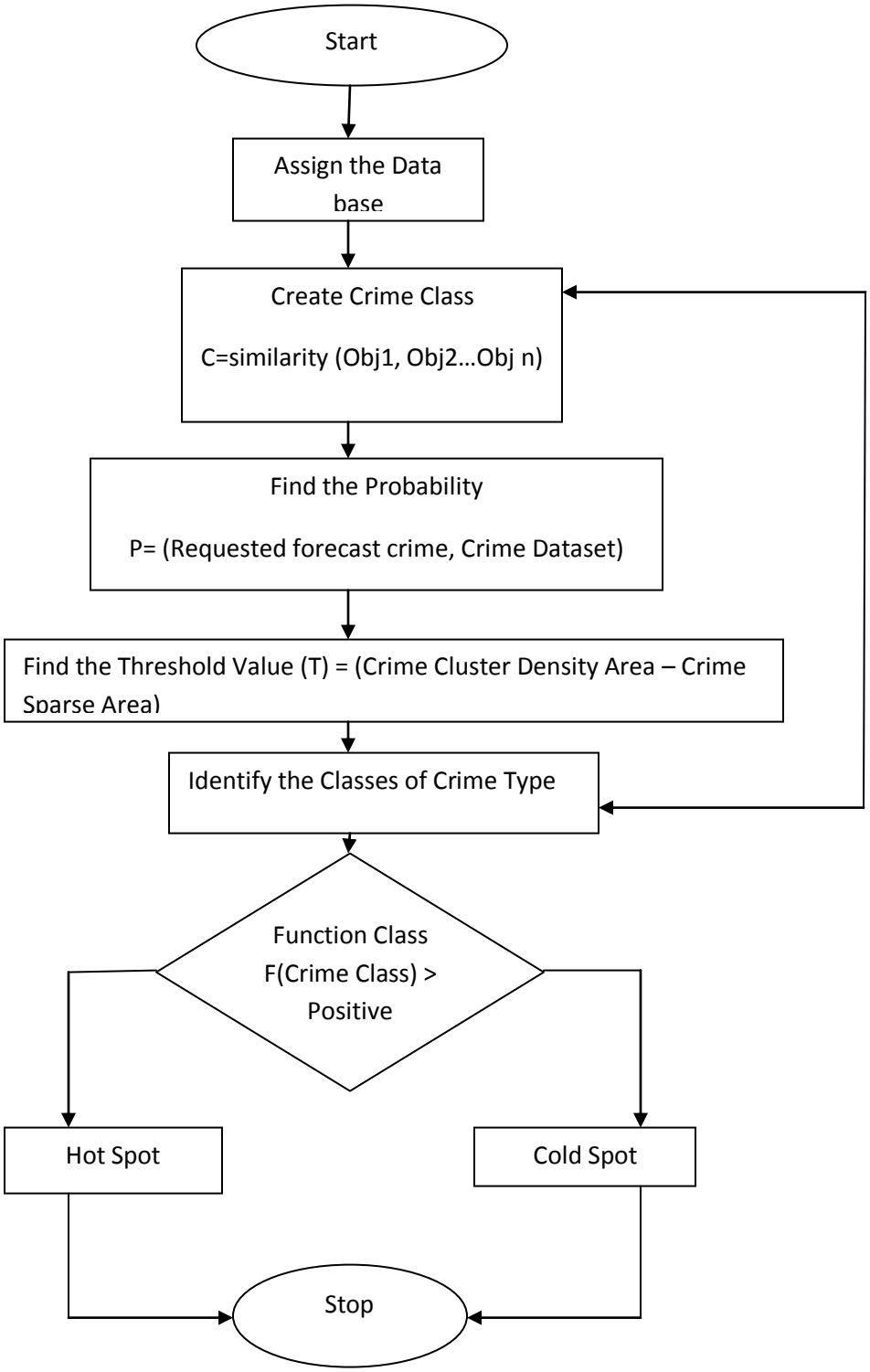

Fig. 3 Flow Design of Crime Classification Algorithm.

Test Results

Time based crime hotspots

From the result of the structure crime classification, it is understood that there are a number of times 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 it would be great if the police department increases the patrol in the midnight hour.

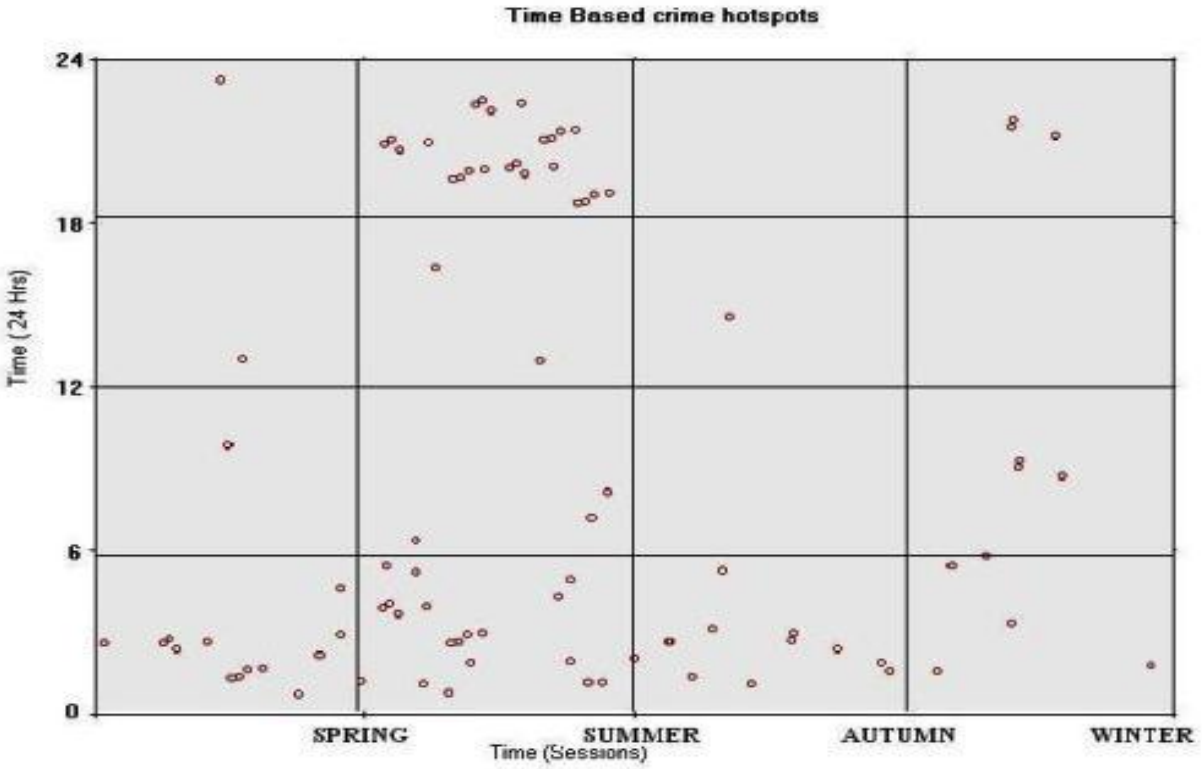

Fig. 4 Time based crime hotspots

Place based crime hotspots

The result of the test shows that there is a group of crime incidents directed at the busiest areas in some time. They were 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.

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 (Crime Class)) > Positive is applied. It is made crime hotspot or else it is represented as a cold spot. To implement this operation, this study used Arcview and spatial analyst with Avenue programming.

Avenue programming could be an Object oriented programming language, utilized in Arcviews. It mainly used for developing Arcview applications. It had been developed by University of Durham information Technology service, 2007. During this programming three different methodologies are used to examine the spatial and temporal data. This study 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 take a look at 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
```

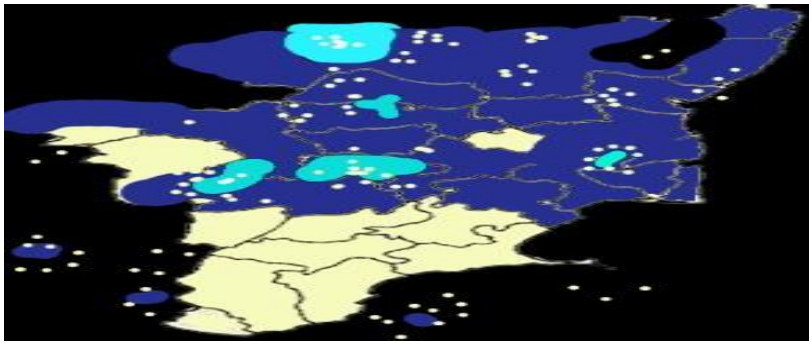

Fig.5 Hotspot and Cold spot

The cyan color represents hotspot and the blue color represents the cold spot.

Conclusion

This paper presented an approach for detecting the crime hot spots by structured classification method. For this purpose some kinds of crime cases were 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.

REFERENCES

[1]Leong, Kelvin, Chan, Stephen,Ng, Vincent and Shiu, Simon., "Introduction of STEM: Space-Time-Event Model for Crime Pattern Analysis", Asian Journal of Information Technology, Medwell Journals, 2008, 7(12),pp. 516-523.

[2]Vijaya Kumar, M. and Chandrasekar, C., 'Spatial Clustering Simulation on Analysis of Spatial-Temporal Crime Hotspot for Predicting Crime activities', International Journal of Computer Applications (IJCA), 2011, 35(3), pp. 36 – 43.

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

[4]Spencer Chainey, Lisa Thompson, Sebastian Uhlig, "The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime", Security Journal, Palgrave Macmillan, 2008, pp. 4-28.
[5]Fayyad, U.M., & Uthurusamy, R. "Evolving data mining into solutions for insights", Communications of the ACM, 2002, 45(8), PP. 28-31.
[6]M.Chau, J.J. Xu, and H. Chen, "Extracting Meaningful Entities from Police Narrative Reports, Proc. Nat'l Conf. Digital Government Research, Digital Government Research Center, 2002, pp. 271-275.
[7]Han J., Data Mining, in Encyclopedia of Distributed Computing,Urban J. and Dasgupta P. (eds.), Kluwer Academic Publishers,1999.
[8]Hsinchun Chen, Wingyan Chung, Jennifer Jie Xu, Gang WangYi Qin and Michael Chau, "Crime Data Mining: A General Frameworkand Some Examples", 2004, IEEE Computer Society, pp. 50-56.
[9]Oskari Heinonen, Heikki Mannila, "Attribute-Oriented Induction and Conceptual Clustering", 1996, University of Helsinki Department of Computer Science Series of Publications. Over the North Pacific a linear response to ENSO? J. Climate, 9, pp. 1468–1478.
[10]Ellis, L. and Walsh, A.(2000). Criminology: A Global Perspective. 1st Ed., Allyn and Bacon, Boston.
[11]Decker, S., & Curry, G. Addressing key features of gang membership: Measuring the involvement of young members. Journal of Criminal Justice, 2000, 28, pp. 473-482.
[12]Srivatsan Laxman and P. S. Sastry, A survey of temporal data mining, in SADHANA, Academy Proceedings in Engineering Sciences, The Indian Academy of Sciences, April 2006, 31, pp. 173–198,
[13]Home Office (2006a), Changes in how police recorded crime, 2006.