IJCSMS International Journal of Computer Science and Management Studies, Vol. 12, Issue 03, Sept 2012
ISSN (Online): 2231-5268
www.ijcsms.com

CLUSTERED SPATIAL DATA MINING ASSOCIATION RULE TO EXPLORE LARGE VOLUMES OF RURAL CRIME

Thangavelu A[1], Sathyaraj S. R[2], Syed Firdose[3], Krishnakeerthi Ch[4] and Balasubramanian S[5]

[1]**Department of Environmental Science, Central University of Kerala, Kasaragod, Kerala**
thangavelgis@gmail.com
[2]**DRDO-BU CLS, Bharathiar University, Coimbatore**
[3]**Department of Computer Science and Technology, GRIET, Hyderabad**
[4]**Department of Computer Science and Technology, GITAM University, Hyderabad**
[5]**JSS University, Mysore, Karnataka**

Abstract

In this paper which interest to integrate a large volume of data sets into useful information by adopting a various information techniques in the latest technology in world. The method approaches of Single variate Association Rule for Area to Crime based on the knowledge discovery techniques such as, clustering and association-rule mining. Data mining provides a clear finding to prevent from crime with associated to another crime occurrence with the naked observation on correlation between one crime to another crime. This tool is an autonomous pattern detector to reveal plausible cause-effect associations between layers of point and area data. We present TATA algorithm with an exploratory analysis for the effectively explore geo-referenced data. We hope this will lead to new approaches in the exploration of large volumes of spatio-temporal data in crime model.

Keywords: *Algorithm, Association rule, Crime data, Data mining, GIS.*

1. Introduction

Data mining is a powerful tool that can be applied to any industries per business. Recently it can be criminal justice in developed countries. Instead of other sophisticated techniques data mining in used popularly because, it need not require drained IT personnel/ specialized personnel/ advanced training in statistics [1]. Data mining is also called as data or knowledge exposure. Data mining software is one of a number of analytical tools for analyzing data. Data mining is the process of analyzing large number of data from different perspectives and abbreviation it into useful information.

Data mining or knowledge discovery within the GIS environment may provide useful relationships and unique patterns. It allows users to analyze data from many different dimensions or angles, categorize and summarize the relationships identified. Accurately, it is the process of finding correlations or patterns among dozens of fields in large relational databases. The traditional manual data analysis has become insufficient, and methods for efficient computer-based analysis indispensable, such as the technologies developed in the area of Data Mining and knowledge discovery in databases [2].

Data mining, when applied to tactical crime analysis, is a knowledge discovery tool that can be used to review extremely large datasets and incorporate a vast array of variables, far beyond what a single analyst, or even an analytical team or task force, can accurately review. It also does this work in a timely fashion, which is critical to apprehending suspects before they can commit additional crimes. The ability to generate and save queries enables agencies to further automate the analytical process, which ultimately saves time and personnel resources, while further enhancing public safety.

The intelligence and security domain, spatial association rule mining [3] has been proposed to extract cause-effect relations among geographically referenced crime data to identify environmental factors that attract crimes [4]. Moreover, the research on association mining is not limited to association rule mining but covers the extraction of a wide variety of relationships among crime data items. Crime association mining techniques can include incident association mining and entity association mining [5].

There is much difference between hotspot of crime and crime data mining. The hotspot of crime is normally referred as a group of crime in a geographical location/region (hotspot), whereas in data mining terminology, a cluster is a group of similar data points a possible crime pattern. The second one provides subset of cluster which has a one to one correspondence to crime pattern.

The patterns, associations, or relationships among all this data can provide information. The information can be converted into knowledge for behaviour. Several data mining techniques are available to identify patterns developed and applied to discover new and interesting pattern relationships from large spatial data [6,7] Understanding the relationship between analysis capability and the data characteristics, can help investigators to use these techniques more effectively to identify trends, patterns and problem areas. If a dataset has an attributes, each consisting of ‘m’ categories, then there can exist ‘q’ possible combinations, involving k ($k \leq n$) attributes in a combination. The main focus of association rule mining is to generate hypothesis rather than to test them as it is commonly achieved through statistical techniques [8]. Here, the data taken are processed through interdisciplinary approach using computer science and criminal justice to develop a data mining paradigm to solve crime faster.

The increased crime rate and enormous amount of data being stored in the databases by police personnel/criminal justice, is becoming important to discover knowledge about crime from databases i.e. mining crime databases. By mining data, we refer to a process of nontrivial extraction of implicit, previously unknown, and potentially useful information, such as knowledge rules, constraints, regularities, and so forth [9]. Unlike other databases, crime databases gather information categories from the records. Some data are also added after the matter is known to public. Therefore, as a data miner, the analyst has to deal with all these public versus private data issues so that data mining modelling process does not infringe on these legal boundaries [10].

The present study tents to serve the combination of data mining- Association rule techniques and GIS to understand the pattern of crime in Coimbatore rural police jurisdiction. Association rule data mining algorithm is Triangular Area through Association (TATA).

2. Integration of Data Mining Association Rule with Algorithm

Data mining with algorithm is powerful decision support tools. However, people use them separately for years, data mining with algorithms have been applied; on the other hand, statistical analysis are traditionally developed for two-way relational databases, and have not been generalized to the multi-dimensional data structure. Combining both data mining and algorithms may provide excellent solutions, in data association method. This method integrates data mining and algorithm to detect concept and ideas from algorithm field.

2.1. Application of Data Mining and GIS with Algorithms

In general data mining is defined as the identification of interesting structure here is designated the patterns, the statistical or predictive models and the existing relationship within parts of the data’s for the present study, we use the data mining technique for identifying crime pattern using association rules.

There are unique needs and challenges for representing discovered geographic knowledge in geographic information system. Most digital geographic data bases are at best a very simple representation of geographic knowledge at the level of basic geometric, topological and measurement constraints. Knowledge-based GIS attempts to build higher-level geographic knowledge into digital databases for analysing complex phenomena [11, 12] Geographic knowledge discovery is a potentially rich source for knowledge-based GIS and intelligent spatial analysis. An important research challenge is developing representations of discovered geographic knowledge that are effective for knowledge-based GIS and spatial analysis.

Data mining and GIS with algorithms to a real-world problem: linking criminal incidents, and compare our method with a similarity-based algorithm. Data mining is a collection of “techniques that can be used to find underlying structure and relationship in a large amount of data” [13] and it has been applied to many real-world applications. Various data mining techniques have also been introduced to law enforcement field. [14] (Brown, 1998) is a shared information and decision support system that assists the police departments to analyze and prevent crimes. The system is an integration of three parts: a database management system, a Geographic Information System (GIS), and statistical analysis tool equipment.

2.2. Data Mining in Criminal Investigation

Data mining is a powerful technique with great potential to facilitate criminal investigators focus on the most important information in their crime data. The knowledge discovered from existing data leads to reveal a value-added of its information. Successful data mining methodology depends heavily on the particular choice of techniques used by analyst. [14] (Brown 1998) has constructed a software framework for mining data in order to catch professional criminals. The software can be used to turn data into useful information with two technologies, data fusion and data mining. Data fusion manages fuses and interprets information from multiple sources, and it overcomes confusion from conflicting reports and disorderly or noisy backgrounds. Data mining is concerned with the automatic discovery of patterns and relationships in large databases.

[15] (Abraham et. al. 2006) proposed a method to employ computer log files as history data to search some relationships by using the frequency occurrence of incidents. Then, they analyze the result to produce profiles. The profiles could be used to perceive the behaviour of criminal. It should be noted that the types of crime could be changed over time influenced by the variations of globalization and technology. Therefore, the crime trends [16] introduced a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. This method yields a visual clustering of criminal careers and enables the identification of classes of criminals. They demonstrated the applicability of data mining in the field of criminal career analysis. Four important factors play a role in the analysis of criminal career: crime nature, frequency, duration and severity. They also develop a specific distance measure to combine this profile difference with crime frequency and change of criminal behaviour over time.

2.3. Association Rule and Clustering

Association rule is one of the important tasks of data mining which identifies interesting association and correlation among large data set of items [17]. Association rule is a tool for discovering correlations among massive databases and the concept was introduced for analyzing market basket data to mine customer shopping pattern. The inputs used in the process of generating association rule are taken from a table where corresponding values in each data items have correlation to one another.

For example, a departmental store and medical store keeps a record of daily transactions which represents the items bought during one cash register transaction by a person. The manager of the departmental store generates the summarized report of the daily transactions that includes the information about what type of items have sold at how much quantity. The generated report also includes the information about those products, which are generally purchased together. Suppose a manager made a rule that if a person purchases bread, then he also purchases the butter. As a result, when availability of bread declines, probably the stock of butter also declines. The manager can make this type of check using association rules.

Let a data set I= $\{I_1, I_2, I_3, \ldots\}$ and a database of transaction D=$\{t_1, t_2, t_3, \ldots\}$ where, $t_i=\{I_{i1}, I_{i2}, I_{i3}, I_{i4} \ldots\}$ and $I_{ij} €I$.

	I_1	**I_2**	**..**	**I_n**
t_1	I_{11}	I_{12}	..	I_{1n}
t_2	I_{21}	I_{22}	..	I_{2n}
t_3	I_{31}	I_{32}	..	I_{3n}
..	..	..	..	..
t_r	I_{r1}	I_{r2}	..	I_{rn}

The example quadrate is to identify relations among the items purchased so that the customers could be targeted for marketing specific products.

The association rule consists of an antecedent (X), consequent(Y), measure of the interestingness which is shown in the equation is (1),

$$X \Rightarrow Y \text{ (support\%, confidence \%)} \quad \ldots (1)$$

The antecedent and the consequent are a set of one or more predicates. The support of a rule measures the frequency of collective occurrence of all the antecedent predicates of a rule in the dataset. The support for an association rule X→Y is the percentage of transaction in the database that consists of X ∩ Y.

$$Support = \frac{X \cap Y}{Total\ no.\ of\ cases}$$

The confidence measures the frequency of the occurrence of the consequent given the occurrence of the antecedent. The confidence for an association rule X→Y is the ratio of the number of transactions that contains X∩Y to the number of transactions that contains X.

$$Confidence = \frac{X \cap Y}{\mathrm{Y}}$$

Criteria value filters the data based on the confidence value.

Thus, classical association rule mining algorithms aim to extract association rules with support and confidence greater than user-specified threshold. The process of Association rule has the following sequences [18]:

(i) Understanding of the problem domain and identification of the final goal of the process
(ii) Data collection
(iii) Pre-processing of data is carried out to refine the data
(iv) Encoding of data
(v) Data mining - Selection of association rule, selection of algorithm for data exploration, running the algorithm to generate patterns

Interpretation, presentation and explanation of mined knowledge.

The immense explosion of geo-referenced data calls for efficient discovery of spatial knowledge. The goal of spatial data mining is to automate the extraction of interesting, useful but implicit spatial patterns [19]. Manifestation of spatial information will help the policy makers and police professionals to investigate the occurrence and distribution of crimes. Therefore, a need to integrate GIS with data mining technique arises, as spatial data are large each with a spatial attribute component along with a temporal component in exclusive cases.

2.4. Definitions

The following definitions are made to explain the unique process of association-rule mining using regions of clusters. Let *P* be a data layer storing point data and *R* a real value in [0, 1].

Definition 1: Clusters with Ratio *R* of *P* [denoted *CwR*(*P*)] are the clusters *C* detected by a clustering approach (algorithm) whose normalized sizes (the number of points in a cluster in *C* / the total number of points in *P*) are equal to or greater than *R*.Let *X* and *Y* be sets of layers (for example *X1*= {*P1*} is a one layer expression). The expression *X* will typically identify the *antecedent* while we use *Y* for the *consequent*. In our approach, *clusters_areas*(*X1*) is a set of polygonized clusters (regions of clusters) of a point data layer *X1* (the point-data is converted to area-data). The function *clusters_areas*(*X*) (also sometimes we will denote this as *clusters_areas* (*antecedent*)) is the total area of the regions that result of the intersection of *clusters_areas* (*Xi*), for all *Xi* in *X*. That is, consider the overlay of *clusters_areas* (*Xi*), for all *Xi* in *X* and find the regions that correspond to points *clusters_areas* (*Xi*), for all *Xi* in *X*. The total area of these regions is *clusters_areas*(*X*).

Definition 2: The Clustered Support (*CS*) is the ratio of the area that satisfy both the *antecedent* and the *consequent* to the area of study region *S*. That is, *CS* = (*clusters_areas* (*antecedent*) ⇒ *clusters_areas* (*consequent*)) / *area*(*S*).

Definition 3: The Clustered Confidence (*CC*) for a rule *X* ⇒ *Y* is the conditional probability of areas of *CwR* of the *consequent* given areas of *CwR* of the *antecedent*. That is, *CC* = *clusters_areas* (*X*⇒*Y*) *clusters_areas*(*X*).

C*lusters areas* (*X*⇒*Y*) is the area of the regions where points are in a cluster for all layers. Thus, they are in the (vertical) intersection of the layers. But, the set of layers is the union of the layers in *X* and the layers in *Y*.

Definition 4: A Clustered Spatial Association Rule (*CSAR*) is an expression in the form of *X* ⇒*Y* (*CC* %), for *X* ⇒*Y* = *f*. The interpretation is as follows: *CC*% of areas of clusters of *X* intersect with areas of clusters of *Y*.

3. Data Used

The daily recorded crime data is used for the study. The collection of recorded crime data for the year 2003 to 2006 are spatially distributed by fourteen Police boundary as on 2006. The collected data set provides the following:

(i) Location of Crime
(ii) Street address
(iii) Time of the crime
(iv) Type of Crime
(v) Geographical information (x, y) coordinates

4. Methods: Data Mining Algorithms

The methodology adopted in this paper has one type of approaches such as crime to crime. Data mining techniques will provide there is any association among the crimes and it is assumed crime is typically a multi-location event where multiple locations are associated through one crime incident. We adopted one type of algorithm developed by as mentioned earlier TATA, to associate area to crime. The methodology is projected as flow chart shows in Figure 1.

Figure 1. The flow chart shows the methodology of using algorithm in Data mining.

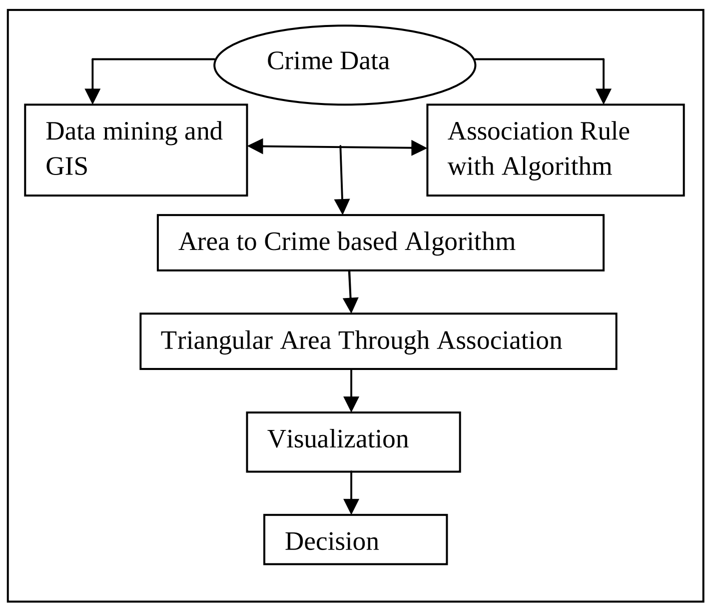


Figure 1. The flow chart shows the methodology of using algorithms in Data mining

There are fourteen police jurisdictions in (14 polygons) where the crimes are spatially spread over. The crime to crime based algorithms analyse the association among different types of theft such as AVT, Grave theft, HB Day, HB Night, Murder for Gain, Murder, Pocket Picking, Robbery and Snatching.

4.1. Data Mining Algorithm Stepwise

The brief steps for the applied algorithms Area to Crime is following:

Triangular Area Through Association (TATA)

Step A-1: All pts are stored in a grid. Polygon is formulated as follows for each map

Step A-2: The points are plotted in the first map related to area

Step A-3: A polygon is constructed for each polygon of Police Jurisdiction. But the constructed polygon will not be helpful to intersect between the respective polygons of Map1 and 2. Therefore fit polygon extent algorithm is constructed. i.e., A Rectangular region bounds the constructed polygon of police jurisdiction

Step A-4: Make a new polygon in memory that consists of X/Y percentage factors for all vertices in the input polygon.

```
Set rectInputPolyExtent = polyinput.Extent
Set polyFactors = New MapObjects2.Polygon
For i = 0 To polyinput.Parts.count - 1
Set ptsFactors = New MapObjects2.Points
For j = 0 To polyinput.Parts.Item(i).count - 1
Set ptFactors = New MapObjects2.Point
ptFactors.Set ((polyinput.Parts.Item(i).Item(j).X -
rectInputPolyExtent.Left) /
rectInputPolyExtent.Width),
_((polyinput.Parts.Item(i).Item(j).Y -
rectInputPolyExtent.Bottom) /
rectInputPolyExtent.Height)
        ptsFactors.Add ptFactors
Next j
        polyFactors.Parts.Add ptsFactors
Next i
```

Step A-5: Make a new output polygon which proportionally fits the input polygon into the new extent

```
Set polyOutput = New MapObjects2.Polygon
For i = 0 To polyFactors.Parts.count - 1
Set ptsOutput = New MapObjects2.Points
For j = 0 To polyFactors.Parts.Item(i).count - 1
Set ptOutput = New MapObjects2.Point
dFactorX = polyFactors.Parts.Item(i).Item(j).X
dFactorY = polyFactors.Parts.Item(i).Item(j).Y
If bFlipHorizontal Then dFactorX = 1 -
dFactorX
If bFlipVertical Then dFactorY = 1 - dFactorY
ptOutput.Set (rectOutput.Left +
(rectOutput.Width * dFactorX)), _
(rectOutput.Bottom + (rectOutput.Height *
dFactorY))
ptsOutput.Add ptOutput
Next j
    polyOutput.Parts.Add ptsOutput
Next i
Set FitPolygonToExtent = polyOutput
```

Step A-6: Repeat from step 1 to 5 for map 2

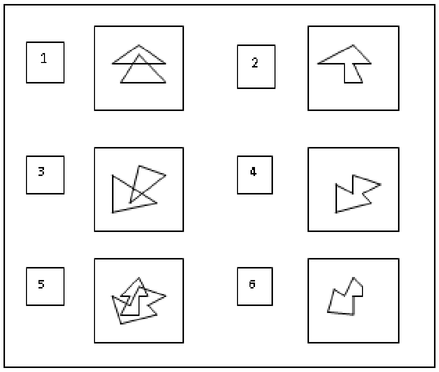
Step A-7: Find the intersect area from map 1 and map 2.

5. Clustered Spatial Association Rules of Algorithms (CSARs)

The algorithm deal with clustered spatial association rules approach approximate shapes of clusters and then polygonize clusters with their boundaries. Thus, it requires an effective clustering and a robust cluster-to-Area transformation to generate accurate association rules. Lee, (2001) [20] proposed an algorithm that extracts cluster boundaries and polygonizes clusters with their extracted boundaries.

The cluster-to-Area transformation does not demand user-supplied parameters to detect shapes of clusters, but derives boundaries of clusters from the distribution of *P*. Points within clusters are not the only contributor to the shape of clusters, but points belonging to other clusters affect the shape of the cluster. It is able to polygonize not only non-convex clusters, but clusters with holes (voids). Now, what are the available is summarized Area-data rather than point-data [21]. This Area-data approximates shapes of clusters of point-data and represents spatial concentrations where most points are aggregated [4]. For these reasons discussed above, we use the boundary based clustering and the cluster boundary extraction approaches for mining associations.

The procedures for the algorithms of clustered spatial association rule process for both categories are described with proxy datasets shown in (Figure 2).



	clusters_Areas	*CS* (%)	*CC* (%)
S	192.22	100.00	NA
dataset I	43.14	22.44	N/A
dataset II	49.96	25.99	N/A
dataset I ⇒dataset II	31.78	16.53	73.67
dataset II ⇒dataset I	31.78	16.53	63.11

Figure 2. The process of single variate association rule approach of algorithms

Figure 2 (1) illustrates a dataset I (the number of points = 6) while Figure 2 (3) presents dataset II (the number of points = 6) with a common study region. Both datasets have two spatial concentrations detected by the boundary-based clustering and those are illustrated in Figure 2 (2) and Figure 2 (4), respectively. Figure 2 (5) illustrates an overlay of regions of clusters of dataset I and those of dataset II. Visual inspection indicates that *clusters_Areas* (dataset I) and *clusters_Areas* (dataset II) intersect. Thus, we may notice that dataset I and dataset II are somehow correlated although we are not able to quantitatively define the association between dataset I and dataset II solely based on visual inspection.

Statistics of dataset I and II in clustered spatial association rule for crime to crime has quantitative analysis and displayed with some numerical indicators in Table and the maps were created and presented in Figure 3. In the map Figure (e.g. Figure 3) the first window of map shows the reference to cluster of crime, second window of map refers to cluster of another type of crime. Third window of map display intersects between *clusters_Areas* (CR-1) AVT and *clusters_Areas* (CR-2) Grave theft using three algorithms mentioned earlier and the advantages is analyzed later.

Statistics of dataset I and II in clustered spatial association rule for area to crime were shown in Tables 2 and the created maps were exposed in Figure 3. The first window of map shows the reference to cluster of crime incidence, second window of map refers to cluster of crime area incidence among the category of location. Third window of map display intersects between *clusters_Areas* incidences of crime and *clusters_Areas* (AVT) category among area of crime (Busstand). The above approaches were performed using the algorithms developed by us and the advantage of each algorithm is analyzed using the crime data mentioned earlier.

1. Statistics of Dataset I and II in Clustered Spatial Association Rule for Area to Crime based on Triangular Area Through Association (TATA) with map

Table 1. Dataset I and II for the Area to Crime using TATA

Dataset I: Refer to AVT
Dataset II: Refer to Grave theft

The Area of study region *Area*(*S*) is 192.22, the total Area of the regions covered by clusters of dataset I is denoted by *clusters_Areas* (dataset I) and its value is 43.14. For the other dataset, *clusters_Areas*(dataset II) is 49.96. The intersection Area of *clusters Areas* (dataset I) and *clusters_Areas* (dataset II) is 31.78 and thus *CS* of dataset I and dataset II is 16.53% (16.53/192.22).
With 8% of *CS*, we are able to derive two association rules. They are as follows:

Rule 1: dataset I⇒ dataset II (73.67% CC),

Rule 2: dataset II ⇒ dataset I (63.11% CC).
From the two rules, it could be inferred that the possibility of AVT (Dataset I) is more feasible in and around the Grave theft (Dataset II).

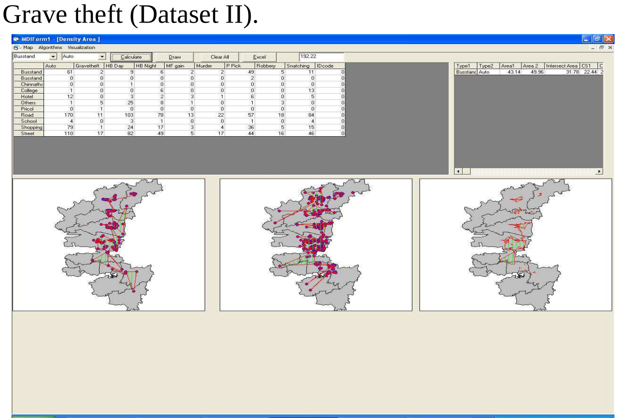


Figure3. Dataset I and II for the Area to Crime using TATA

5.1. CSAR in Real Datasets Result

Achievable *CSAR*s will enhance exponentially as the number of layers under study (in this case, crime types and features) grows. Thus, it is almost impossible for analysts to find interesting associations manually. The area to crime process the real datasets with TATA, view approach automatically generates strong *CSAR*s with user-specified *CC* and *CS*.

5.2. Crime Rules of TATA for Area to Crime

The place to crime TATA -view approach automatically generates strong *CSARs* with user-specified *CC* and *CS*. For instance, *CC* and *CS*. For instance, with 75% *minimum CC* and 12% *minimum CS*, twenty four strong *CSARs* among 72 rules from the Table 1 are bare and accessible in the Table 2.

Table 2. The value of *CS (12%)* and *CC (75%)* of *CSARs* of the real Area to Crime data of TATA

Crime Rules	CS(%)Offence Area ⇒ Crime	CS(%)	CC%
Crime Rule:1	Bus stand ⇒Auto	16.53	63.61
Crime Rule:2	Bus stand ⇒HB day	16.49	50.19
Crime Rule:3	Bus stand ⇒ P Pick	18.24	71.69
Crime Rule:4	Bus stand ⇒ Snatching	13.54	63.59
Crime Rule:5	Road ⇒Auto	22.41	86.23
Crime Rule:6	Road ⇒Grave Theft	13.95	85.80
Crime Rule:7	Road ⇒HB Day	29.34	89.30
Crime Rule:8	Road ⇒HB Night	28.35	90.98
Crime Rule:9	Road ⇒ Murder for gain	20.65	87.43
Crime Rule:10	Road ⇒Murder	25.22	81.70
Crime Rule:11	Road ⇒P Picking	24.88	97.77
Crime Rule:12	Road ⇒Robbery	13.29	88.13
Crime Rule:13	Road ⇒ Snatch	21.29	93.62
Crime Rule:14	Shopping ⇒HB Day	15.48	47.12
Crime Rule:15	Shopping⇒HB Night	13.77	44.17
Crime Rule:16	Shopping ⇒P Pick	12.22	51.73
Crime Rule:17	Street ⇒Auto	18.20	70.04
Crime Rule:18	Street ⇒HB Day	21.90	66.64
Crime Rule:19	Street ⇒HB Night	23.23	74.54
Crime Rule:20	Street ⇒urder for gain	17.20	72.83
Crime Rule:21	Street ⇒Murder	18.48	59.86
Crime Rule:22	Street ⇒P Pick	18.73	73.59
Crime Rule:23	Street ⇒Robbery	12.51	82.96
Crime Rule:24	Street ⇒Snatch	17.92	84.16

The rule says AVT, Grave theft, HB Day, HB Night, Murder, P. Picking, Robbery and Snatching are predominant offences taking places among the ten categories of crime. We know that more than 12% CS of theft, such as;

(1) Bus stand occurs with AVT , HB Day and P. Picking and Snatching
(2) Road occurs together with AVT, Grave theft, HB Day, HB night, Murder for gain, Murder, P. Picking, Robbery and Snatching
(3) Shopping occurs together with HB Day, HB night and P.Picking
(4) Street occurs together with AVT, HB Day, HB night, Murder for gain, Murder, P. Picking, Robbery and Snatching.

and more than 70% CC of theft, are as follows;

1. Who do Bus stand do P. Picking
2. Who does Road do AVT, Grave theft, HB Day, HB night, Murder for gain, Murder, P. Picking, Robbery and Snatching and
3. Who do Street do AVT, HB Night, Murder for gain, P.Picking, Robbery and Snatching

The common crime rules among the three algorithms are given below and it is concluded that the Bus stand, Road, Shopping and Street are the most prominent crimes for the three algorithms.

(i) Bus stand ⇒ P.Picking
(ii) Road ⇒ P.Picking and Snatching
(iii) Shopping ⇒ P.Picking
(iv) Street ⇒ Snatching

This gives us a suggestion of how crimes are associated with each other and the occurrence of one Area may result in another category of crime.

6. Conclusion

We demonstrate our approach with real crime datasets of Coimbatore rural and provide comparisons within the algorithm/techniques. Experimental results reveal the applicability and usefulness of the proposed approach.

Day by day in the control centre of police office, people receive a large number of information about cases received with various approaches. The information has been input into database to form a large amount of case information. These case information has been archived yearly and periodically to form a plenty of historical case resources. By inducing and analyzing these historical cases, people can get some experiences and learn some lessons that can help them solve cases and make decisions in the future. Therefore, in order to assist police departments to solve cases rapidly and make decisions efficiently, we should synthesize and organize these historical data, use proper data mining models to discover the potential and useful knowledge behind the data, and then predict and analyze the important factors in the data including the rate of crime, the constitution of crime population, the crime age structure, the area distribution of crime, the developing tendency of crime, the means and approaches of crime, the hidden areas of criminals and so on. At present all of these have become urgent tasks that need our police office to accomplish in the procedure of data processing.

The Indian Police are uniformly used a software (i.e.) Crime Criminal Information System (CCIS) throughout the country. CCIS is a government to government (G2G) model which aims the storage analysis and retrieval of criminal data similar to GIS packages. The National Crime Record Bureau (NCRB) has developed following e-governance applications Police Station Management System, Prison Statistics, Jail Management Software's, Prosecution Branch System, National Bomb Squad System and Forensic Science Laboratory System, Motor Vehicle Information Counters.

To overcome such feature the combining the powerful tools of GIS to integrate and manipulate spatial data mining methods, spatial data analysis shows great promise for criminology criminal justice, and law enforcement research and practice. The results presented in area to crime review a single, easy and great method for crime analysis. The developed method us to seek inference that one crime out of nine categories of crimes are more predominant and occurs in one out of three areas. Therefore, the result provides that the vigilant squad specialised in the above mentioned categories of crimes should be deployed in these three places for protection of property and prevention of crime. In addition, the association rule mining and GIS technique helps to reveal the spatial pattern of crime and areas in Coimbatore rural jurisdiction of TamilNadu which will identify the Areas that needs further attention by the bureaucrats. In this method help to police department especially for the rural crime areas.

7. Reference

[1] McCue, "Using Data Mining to Predict and Prevent Violent Crimes" (2003), available at: http://www.spss.com/dirvideo/richmond.htm?source=dmpage &zone=rtsidebar.

[2] Frawley, W., Piatetsky-Shapiro, G., Matheus, C. "Knowledge discovery in databases: An overview." In: *Knowledge discovery in databases* (Piatetsky- Shapiro, G., Frawley, W., eds.), The AAAI Press, Menlo Park, CA (1991).

[3] Koperski, K and Han, J Discovery of Spatial Association Rules in Geographic Information Databases. In *Proc. 4th Int. Symp. Advances in Spatial Databases, SSD*, M. J. Egenhofer and J. R. Herring, Eds. Vol. 951. Springer-Verlag, 47{66.Koperski,K and Han, J. 1996. Data mining Methods for the Analysis of Large Geographic Databases.

[4] Estivill-Castro, V and Lee, I. "Data mining Techniques for Autonomous Exploration of Large Volumes of Geo-referenced crime data" *proceedings of the 6th International Conference on Geocomputation*, (2001) http://www.geocomputation.org/2001 /papers/ estivillcastro.pdf.

[5] O'Hara,.C.E.and O'Hara, G.L*Fundamentals of Criminal Investigation* (5th ed.).Springerfield,IL;Charles C.Thomas. (*1980)*

[6] Wang, J., Han, J., Pei, J., CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Item Sets. In: Proceeding of the 2003 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'03), Washington, DC, (2003) pp 236–245.

[7] Han and Kamber. "Data Mining: Concepts and Techniques", The Morgan Kaufmann Series, (2001).

[8] Vinnakota and Lam, "Socioeconomic Inequality of Cancer Mortality in the United States: A Spatial Data Mining Approach", *International Journal of Health Geographics, 5:9.* (2006)

[9] Agrawal, R., Imielinski, T. and Swami, A. Mining Association Rules between Sets of Items in Large Databases", *Proceedings of the 1993 ACM SIGMOD Conference*, (1993), pp "207-216.

[10] Nath, S.V. "Crime Pattern Detection Using Data Mining," in Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, (2006), pp41-44.

[11] Srinivasan, A. and Richards, J. A., Analysis of GIS spatial data using knowledgebased methods. International Journal of Geographical Information Systems, 7:479-500, (1993).

[12] Yuan, M., Use of knowledge acquisition to build wildfire representation in geographic information systems. International Journal of Geographical Information Systems, 11:723-745, (1997).

[13] Kennedy, D.M., Braga, A.A. and Piehl, A.M. The (un) Known Universe: Mapping Gangs and Gang Violence in Boston. In D. Weisburd & T. McEwen (Eds.), *Crime Mapping and Crime Prevention* (pp. 219–262). Monsey, NY: Criminal Justice Press, (1998).

[14] Brown, D.E. “The Regional Crime Analysis Program (ReCAP): A Framework for Mining Data to Catch Criminals”, Proc. of IEEE International Conference on Systems, Man, and Cybernetics, (1998). pp 2848 –2853,

[15] T. Abraham and O. de Vel, "Investigative profiling with computer forensic log data and association rules," in Proceedings of the IEEE International Conference on Data Mining (ICDM'02), pp. 11 – 18.

[16] J.S. de Bruin, T.K. Cocx, W.A. Kosters, J. Laros and J.N. Kok,(2006) “Data mining approaches to criminal career analysis,” in Proceedings of the Sixth International Conference on Data Mining (ICDM’06), (2006) pp. 171-177.

[17] Intan and Rolly. “A Proposal of an Algorithm for Generating Fuzzy Association Rule Mining in Market Based Analysis”, Proceedings of The 3rd International Conference on Computational Intelligence, Robotics and Autonomous System. Issues and an Application, Geographical Analysis 27:4, (2005) pp 286-305.

[18] Kwasnicka, H. and Switalski, K. “Discovery of Association Rules From Medical Data –classical and Evolutionary Approaches”. *Autumn Meeting of Polish Information Processing Society Conference Proceeding*, (2005) pp. 163-177.

[19] Shekhar, S and Chawala S, “Spatial Databases: A Tour”, Prentice Hall, (ISBN 013-017480-7), (2003).

[20] Lee, I. “Polygonization of Point Clusters Through Cluster Boundary Extraction for Spatial Data Mining”, *Technical Report 2000-08*, The University of Newcastle, Australia, (2001).

[21] Estivill-Castro, V. and Lee, I. “AUTOCLUST: Automatic Clustering via Boundary Extraction for Mining Massive Point-Data Sets”, *Proceedings of the 5th International Conference on Geocomputation*, GC049, (2000).