International Journal of Computer Applications (0975 – 8887)
Volume 126 – No.7, September 2015

Mitigating Serial Hot Spots on Crime Data using Interpolation Method and Graph Measures

S. Sivaranjani
Assistant Professor
Department of Computer Science and Engineering
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore-641 108
India

S. Sivakumari, PhD
Professor and Head
Department of Computer Science and Engineering
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore-641 108
India

ABSTRACT
Crime detection is the vital and emerging research field in the real world environment which aims to prevent the number of crimes happening in the world. The nature of crime differs in different places based on location, age, religion, habitual characteristics and so on. Mitigating the serial crimes which are identical to each other is the most important scenario to be concerned in the real world. There may be a problem arise while mitigating the hot spots in the different crime locations due to missing values of some important features. Prediction of similar types of crimes also becomes the complex process where the temporal features are scattered. To solve the problem in this work the triangulation based interpolation methodology (TIM) and the graph measures were introduced. The TIM tends to find the missing value among the set of values based on the average level of the most nearer points where the data points are scattered unevenly.And the similarity measures assure the selection of the most nearest neighbour solutions. The similarity measures that are used in this work for predicting the most nearest location with same type of crime behaviour are Distance Measure (DM), Centrality Measure (CM) and Graph Assortativity (GA) measure. The performance evaluations were conducted with the help of spatio temporal data sets where the list of crimes and the location, behaviour are depicted properly. The experimental tests conducted proves that the proposed methodology in this work can mitigate the serial crime hot spots more accurately.

Keywords
Spatio temporal data, Interpolation, Graph distance measures, hot spots.

1. INTRODUCTION
Crime is an immoral act against the rules and regulations defined by the government authorities. Criminology is the study of crimes, criminals, crime victims, theories explaining illegal and deviant behaviour, the social reaction to crime and the effectiveness of anti-crime policies, etc. Crime security and crime forecasting activities are most important concerns for both citizens and government.Crime detection plays a vital role in the criminal justice and enforcement specialist. Nowadays, with the fast progression of technology, a computer-based method for exploring, visualizing, and investigating the occurrences of criminal activity is grown.

In the criminology, Suspect refers to the individual who committed the crime. The victim is the one who is the target of the crime. Crime hot spot is defined as the area with absolutely high rate of crime occurrences. Accurately identifying these hotspots will significantly benefit police force in crime detection and control [1]. There are different hot spot detection techniques like KDE, Kriging, Anselin native Moran's I, Getis-Ord Gi that is shown in Fig. 1

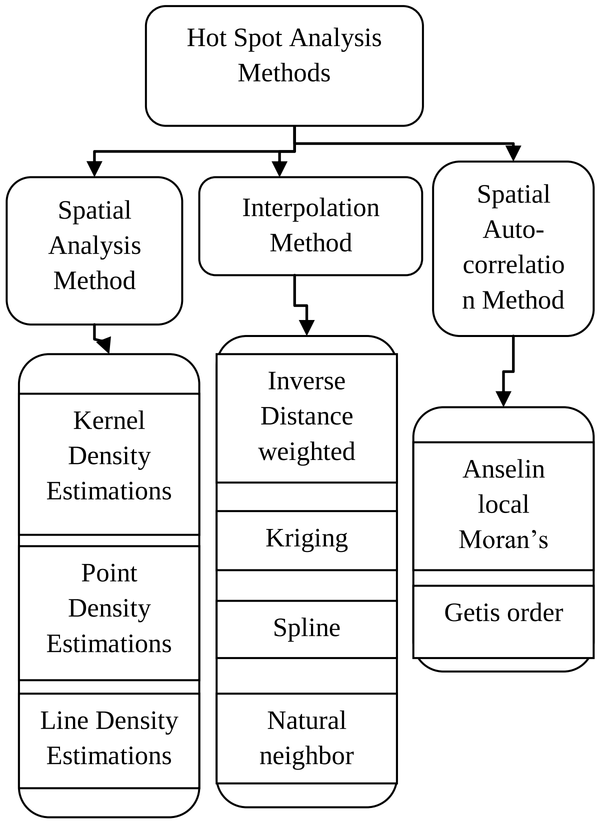

Fig 1: Hot Spot Detection methods

Cluster of crime features refers to a geographical cluster of crime, i.e. lot of crimes in a given area at a specified time. The crime investigation can be done efficiently only by analyzing and searching for number of evidences related to the crime scene.

Such clusters are often visually represented employing a geo-spatial plot of the crime overlayed on the map of the police jurisdiction. However, once cluster topic is discussed from a data-mining viewpoint, it have a tendency to confer with similar sorts of crime within the given geographic area of interest. Such clusters are helpful in characteristic against the law pattern. These crimes could involve single suspect or is also committed by a crowd of suspects.

Geographical information system (GIS) is the widely used tool for facilitating and exploration of the spatial distribution of crime. The elemental strength of GIS over ancient crime analytical tools and ways is that the ability to check, analyse and justify the criminal activity in a very spatial context. Environmental Systems analysis Institute (ESRI), California (1990),outlined GIS as an organized assortment of constituent, software system and personnel to efficiently capture, store and update, manipulate, analyse and show all types of geographically documented data.

Crime mapping[3] is employed by criminologists to map, visualize, and analyse crime incident patterns. Consequently, trendy GIS software system permits enforcement agencies to provide additional versatile electronic maps by combining their crime databases of reportable crime locations with digitized maps of the target areas. Additionally it enabled them to spot crime hot spots together with different trends and patterns

The main contribution of this work is to predict and mitigate the hot spot of serial crime in the geographical area which are most similar to each other. This is done by introducing the triangulation based interpolation method which is used to predict the unknown values from the known parameters by finding the average values of them. Then the similarity measures are used to group the most data's in terms of the most nearest points that are combined together.

2. RELATED WORKS

All The GIS and Remote Sensing methods are used for visualizing the information and analyzing the facts. This could be used to map the Police stations and to spot the Crime Zone as Hot Spot and to statistically analyse the crime [4].

Shyam Varan Nath S [9] implemented data mining methods with the geospatial design of crime which helps detectives in increasing the security for the homeland by improving the productivity of Law enforcement officers. Author applied K-means clustering to identify the crime patterns very efficiently and increased the speed in the process of resolving the crimes.

Devesh Bajpai [10] discussed about the crime network analysis to extract the useful information by using the recent techniques like Social Network Analysis, Entity Extraction, String Comparator etc. The data mining techniques can be applied in the area of crime investigation for decision support, prediction, forecasting etc., The author stated that the usage of artificial intelligence tools for enhanced crime analysis.

Jitendra Kumar, Sripati Mishra, Neeraj Tiwari [15] presented a study in identifying the area which is consisting of major crime named hotspots and the area with minimum crimes. The author applied clustering techniques for the identification of hotspots and safe zone of crime for the data collected from State Crime Record Bureau, Uttar Pradesh.

John David Elijah Sandig *et al.*,[11] offered an online Graphical Information System (GIS) for crime rate and models using frequent pattern analysis. It was a web-based system that contains GIS for robbery, homicide and physical injury incidents within a (Iloilo) City. The frequent happening of the crime and its occurrence time have been identified by using this system. The frequent victims in the community can be identified. The author made use of Google Heat Maps for plotting the crimes

TongWang *et al.*,[12] suggested a pattern detection algorithm called Series Finder to find the common features of all crime patterns and the distinctive aspects of each specific crime pattern which are committed by the same criminal or groups of criminals. A comparison was done between the series finder algorithm and clustering methods for crime analysis.

Kate J. Bowers, Shane D. Johnson and Ken Pease [13] explored the development of a mapping technique that seeks to produce 'prospective' hot-spot maps. Novel map evaluation method was proposed and some standardized metrics that can be derived from maps are compared for its efficiency. The proposed predictive mapping technique was found more advantage than other traditional methods and might ascertain predominantly useful in the shift-by-shift positioning of police personnel.

Arbind Kumar Singh and Manimannan G [14] introduced a novel method of mapping the top level crime in various district/cities of Tamilnadu on the basis of crime factors. Narrative information and crime records are warehoused digitally across individual police departments, allowing the collection of this data to compile a district wise database of crimes they committed in Tamilnadu. The author applied k-mean clustering and Geographical Information System (GIS) analyses as data mining tools to develop the hidden structure present in the data for each year. Finally, the groups were identified as crimes belonging to High Crime Activity (HCA), Intermediate Crime Activity (ICA) and Low Crime Activity (LCA) in that order, which show the behaviour of High Crime Activity cities, Intermediate Crime Activity cities and Low Crime Activity cities. The results of this paper by applying data mining and GIS could potentially be used to identify the hot spot and even prevent the crime happening in future.

3. MITIGATING HOT SPOTS OF SERIAL CRIMES

Crime detection plays a main role in the crime investigation department where the human resources are less to find out the serial crimes which are happening in the scattered manner. The crimes that are happening in the different location have differed in its characteristics based on the user behaviour and culture. However, serial crime would have connection among them regards to the similar characters. Finding and resulting the similar types of crimes happening in different location would help to catch the criminals who are committing crimes. Accuracy of serial crime prediction also becomes the difficult process where the crimes are analysed using the Graphical Information System (GIS). The prediction of serial crimes becomes most complex process where the graph and graph measures are not taken into account.

In this work, similarity connections between the serial crimes that are happening in different location are found by taking similarity values between the data of different locations. However, the similarity finding may be more difficult in case of missing values present in the crime data which are collected from various locations. The accuracy of serial crime will become more difficult process in case of missing values where the similarity cannot be found. This work predicts the serial crimes that are happening in the different locations by contributing following methodologies:

- Triangulation based Interpolation method to find missing values

- Integrating graph measures to find the similarity level

3.1 Triangulation based Interpolation Methodology

Interpolation is the method of victimisation points with known values to propose values at alternate unknown points. It is typically utilized in the sector of precipitation, noise levels, chemical concentrations, elevation, or different spatially-based phenomena. It's the approximate judgment of surface values at the points that are un-sampled the surface values of close points that are recognizable. It is one of the most popular methodologies which are used to find the missing values with the help of the values that are present in the nearer places. ie. It is meant for the prediction of the new data point which is missing from the data set with the help of known discrete data points. According to the literature, the interpolation of data is done by using inverse distance weighting method. This method is used to find the unknown value of particular point by taking the average weight of surrounded known points. The major limitation is that estimates are bounded by the extrema in the sampled values. Additionally, the radial symmetry obscures the effect of linear features such as ridges or valleys. For $n < 1$, the derivative of the interpolated surface is discontinuous at the sampled locations, while for $n > 1$, the surface is flat at these sampled locations where n is the inverse distance weighting power. The value n decides the important of location based on crime density happening there,

This work aims to enhance the interpolation mechanism for locating out the missing values from the best-known values of enclosed points. In this work, triangulation with linear interpolation technique is introduced for interpolation mechanisms which can overcome the drawbacks exist within the previous work. The Triangulation based Interpolation technique is predicated on the optimum Delaney triangulation. This method creates triangles by drawing lines between data points. The real points are connected in such a way that no triangle edges are intersected by different triangles. The result is a patchwork of triangular faces over the extent of the grid. This technique is used as a particular interpolator. Every triangle defines a plane over the grid nodes lying inside the triangle, with the lean and elevation of the triangle determined by the three original data points defining the triangle. All grid nodes inside a given triangle are outlined by the triangular surface. Triangulation with Linear Interpolation works best once data are equally distributed over the grid space. Data sets containing distributed areas lead to distinct triangular aspects on the map.

Delaunay Triangulations is used in this work for predicting the missing values among the set of crime data. This methodology is based on the legal graph method. The legal graph is nothing but the graph with many illegal data points. The Check legal graph theorem to prove the legal graph is given as follows

Algorithm checkLegalTriangle()

Input. A set of crime data point that belongs to one triangle patch

Output. Complete data point with legal graph

1. while T contains an illegal edge $\overline{p_i p_j}$
2. do (* Flip $\overline{p_i p_j}$ *)
3. Let $\overline{p_i p_j p_k}$ and $\overline{p_i p_j p_l}$be the two triangles adjacent to $\overline{p_i p_j}$.
4. Remove $\overline{p_i p_j}$ from T, and add $\overline{p_k p_l}$ instead.
5. return T

The legal triangulation is defined as the triangle with no illegal edge which is used in this work for detecting the set of crimes that belongs to same category. The legal triangulation concept is used to avoid the illegal edges resides in the triangulation, so that legal graph can be obtained. Based on this concept, this work would eliminate the crime that are not belong to cluster and will form the legal graph of data point which would consists of only legal crime data points.

In this algorithm, Pi, Pj, Pk, Pl denotes the data points that belong to the triangulation. This methodology attempts to eliminate the unnecessary edge points which are adjacent to each other.

The above algorithm helps to add or remove the data points that belong to the same triangles based on the adjacent values. The adjacent values are measured and calculated by taking the averaging between the data points of two different triangles.

Consider P is the number of data points that belongs to the different crime areas. The voronoi diagram V can be constructed by sub dividing the plane into the multiple Voronoi cells V (p) for all data point belongs to crime area. Here G is the dual graph of the V where the Delaunay graph DG(P) is defined as the straight line which is the integration of G.

The legal triangulation can be defined from the Delaunay graph as "Consider P be a set of crime data points in the crime areas. A triangulation T of P is legal if and only if T is a Delaunay triangulation."

From this methodology, the missing values can be found by proving the triangulation to be the complete and legal triangle based on which serial crimes can be found.

3.2 Integrating Graph Measures to Find the Similarity Level

Interpolation the effective mapping of crime visualization is not yet reached in the previous work where the graph measures are not taken into consideration for visual mapping process. The different types of graph measures have to be taken in mind when processing a spatial data's in the graph format

Graph measures are the one which are used to predict the different types of crimes that are happening in different location which are depicted in the graph format accurately. Graph measures are used to find the overall depth of the graph in terms of the connectivity present among the different edge nodes. Some of the graph measures that are concentrated in this work are Distance Measures (Graph Distance, Graph Distance Matrix), Centrality Measures (Closeness Centrality, Betweenness Centrality), Graph Associativity, and Vertex Correlation Similarity.

3.2.1 Distance Measures

The Distance measures are the one which is used to define the overall similarity connectivity present among the different vertices of graph where the vertices represent the types of crimes that are located in various places. In this work, two types of distance measures such as Graph Distance method and Graph Distance Matrix method are used to predict the serial crimes which are happening frequently. Graph distance is nothing but the number of edge points that are connecting the different vertices to reach the target point. In this work, shortest path are preferred to find the most nearest location where the similar types of crimes are happening more. Shortest path is nothing but the target connecting point with the less number of edges. Graph Distance [g, s, t] will return the hop count of the smallest path between s and t. The distance is said to be infinity when there is no path between s and t. For a weighted graph, the distance is the minimum of the sum of weights along any path between s and t. Graph Distance Matrix is used to depict the various distances present among the set of data points ie., crimes which are happening in multiple locations. This distance will be calculated by taking the similarity values of crimes happened in two different location. But using this graph distance matrix method, the most similar crimes which are identical in regards to the crimes happened in the different locations can be found. The entries of the distance matrix d_{ij} will indicate the smallest distance from the vertex v_i to vertex v_j. The diagonal entries dij of the distance matrix are always zero. The entry dij is Infinite if there is no path from vertex v_i to vertex v_j. In Graph Distance Matrix [g, d], an entry d_{ij} will be Infinity if there is no path from vertex v_i to vertex v_j in d steps or less. The vertices v_j are assumed to be in the order given by Vertex List[g]. For a weighted graph, the distance is the minimum of the sum of weights along any path from vertex vi to vertex v_j.

3.2.2 Centrality Measure

The Centrality measure [21] is used to find the most important metrics that are residing in the graph. Centrality measure would help the investigators in predicting the important areas where the similar types of crimes happening. In this work, two types of centrality measures are used for predicting the important vertices that are present inside the graph. Those are closeness centrality and betweenness centrality. Closeness centrality is used to indicate the average fraction of distance between one vertex to all other vertices present in the graph. Closeness Centrality will give high centralities to vertices that are at a short average distance to every other reachable vertex. Closeness Centrality for a graph is given by $(1/l_1, 1/l_2,\dots)$ where l_i is the average distance from vertex I to all other vertices connected to i. If d is the distance matrix, then the average distance l_i from vertex I to all connected vertices is given by;

$$(\textstyle\sum_j d_{i,j})/k \qquad (1)$$

where the sum is taken over all finite d_{ij} and k is the number of vertices connected to i. The closeness centrality for isolated vertices is taken to be zero. Closeness Centrality works with undirected graphs, directed graphs, weighted graphs, multi-graphs, and mixed graphs.

Betweenness centrality is an indicator of a node's centrality in a network. It is equal to the number of shortest paths from all vertices to all others that pass through that node. Betweenness Centrality will give high centralities to vertices that are on many shortest paths of other vertex pairs. Betweenness Centrality for a vertex I in a connected graph is given by

$$\textstyle\sum_{s,j\in v^\wedge s\neq iN\neq i} n_{sj}^{j}/n_{sj} \qquad (2)$$

Where

i, j = node value

here n_{sj} is the number of shortest paths from s to i and n^j_{sj} is the number of shortest paths from s to l passing through i. The ratio n^j_{sj}/ n_{sj} is taken to be zero when there is no path from s to t. Betweenness Centrality works with undirected graphs, directed graphs, multi-graphs, and mixed graphs.

3.2.3 Homophily, Assortative mixing, and Similarity

These measures [22] are used to predict the similar types of vertices that are connected with each other in terms of similarity level. ie., it will find out in what way nodes are connected with each other. It will be found by calculating the similarity among the different set of nodes.

3.2.4 Graph Assortativity

The Assortativity or assortative mixing [23] is a preference for a network's nodes to attach to others that are similar in some way. Though the specific measure of similarity may vary, network theorists often examine assortativity in terms of a node's degree. For a graph with m edges and adjacency matrix entries $a_{i,j}$, the assortativity coefficient is given by

$$(\textstyle\sum_{i,j}\left(a_{i,j}-\frac{d_i d_j}{2m}\right)f_{i,j})/(\textstyle\sum_{ij}\left(d_i\delta_{i,j}-\frac{d_i d_j}{2m}\right)f_{i,j}) \qquad (3)$$

Where d_i is the out-degree for the vertex vi and $\delta_{i,j}$ is 1 if there is an edge from v_i to v_j and 0 otherwise. For quantitative data where x_1, x_2….are used, f_{ij} is taken to be x_i x_j. For categorical data where x_1, x_2,…. are used, f_{ij} is taken to be 1 if xi and xj are equal and 0 otherwise. In Graph Assortativity[g], xj is taken to be the vertex out-degree for the vertex vi. In Graph Assortativity [g, "prop"], xi is taken to be Property Value [{g, v_i},"prop"] for the vertex vi. In Graph Assortativity [g, {{$v_{i\,1}$,$v_{i\,2}$,…},…}], vertices in a subset (Vi1, Vi2,…) have the same categorical data xil = xi2=... GraphAssortativity [g, Automatic->{x_1,x_2,…}] takes the vertex list to be VertexList[g]. The option "Data Type"->type can be used to specify the type for the data x1, x2,….. Possible settings are "Quantitative" and "categorical". The option "Normalized"->False can be used to compute the assortativity modularity. For a graph with m edges and adjacency matrix entries ai,j, the assortativity modularity is given by

$$\frac{1}{2m}\textstyle\sum_{ij}\left(a_{i,j}-\frac{d_i d_j}{2m}\right)f_{i,j} \qquad (4)$$

where di is the out-degree for the vertex vi. Graph Assortativity works with undirected graphs, directed graphs, weighted graphs, multi graphs, and mixed graphs.

3.2.5 Vertex Correlation Similarity

It gives the correlation similarity [24] between vertices u and v of the graph g. The vertex correlation similarity is also known as Pearson correlation coefficients. Vertex Correlation Similarity works with undirected graphs, directed graphs, weighted graphs, multi-graphs, and mixed graphs.

By using these above mentioned measures, the relationship exist among the set of vertices nodes can be calculated well which in turn more similar to the same types of crimes. The hot spots of serial crimes which are happening in different location can be found effectively by using these methodologies by calculating the similarity exists among the different methods.

4. EXPERIMENTAL RESULTS

The experimental test for this work has been conducted over the crime data sets which are collected across various locations in north Coimbatore region. The crime data set consists of information about the nature and type of the crimes that can be happened in multiple locations at different times. This work tends to mitigate the hot spots of locations where the number of crimes happened are more in number. And also this hot spot mitigation will give the knowledge about the similar type of crimes that were happened in different locations. This performance evaluation proves the proposed work in this research which is named as interpolation and triangular basis method with the graph measures are compared with the existing work called the Inverse Distance Weighted (IDW) method in terms of performance parameters called the accuracy, precision and recall values. These values are calculated for the different types of crimes which are namely, kidnab, dowry death and dacoity. Kidnap is the process of moving or taking away the persons from one location to another location without their willingness. Dowry deaths are defined as the continuous death of women's due to the harassment by husband. Dacoity is the form of robbery which is done by the group of person involving more than 5 persons are conjointly.This performance evaluation proves that the proposed methodology can work better in case of presence of missing values also. The graph measures are used to improve the performance by averaging the similar location between different nodes of vertices.

The GIS representation of clustered results of crimes which were happened in the various crime locations are depicted as follows:

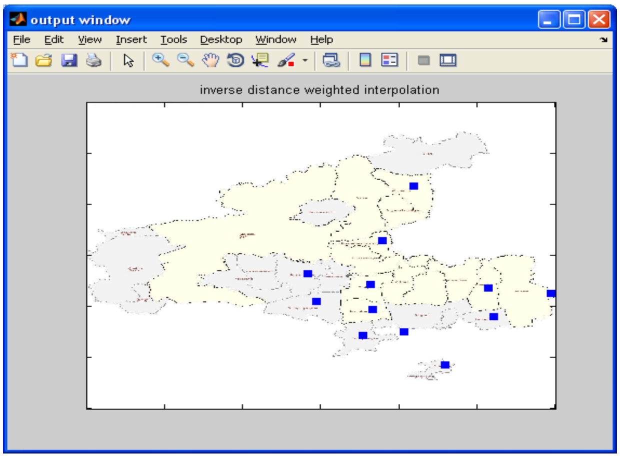

Fig 2: Data lies outside of cluster

The Fig: 2 depict the crimes that falls outside of clusters of crime location. This performance analysis are represented in the graphical represented which is explained in the detailed manner in the following sections.

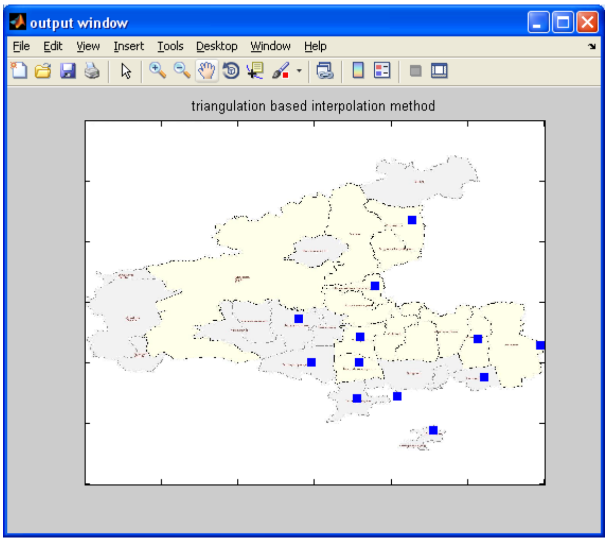

Fig 3: Clustering inside crime location

In the figure:3 depicted above, data points (crime points) are located inside of clusters. That is this diagram depicts the crimes which are happening within the crime location.

4.1 Data Set

Crime data set is used for analysing and predicting the series of crime that are happening in north Coimbatore region, Tamilnadu state, India. The crime data considered in this work consists of attributes like different types of crimes for eg., kidnap, robbery, rapes, murder, cyber crime and gambling etc., latitude and longitude value of hotspots denotes the exact location where the crimes happened. By using these information, the different types of crimes are analysed and finally the similar types of crimes are grouped together to predict the various future crimes.

The Accuracy, Precision and Recall is being calculated for various types of crime by using the formulas as:

4.1.1 Accuracy

Accuracy is defined as the percentage of correct prediction of hot spots which are happened in the different crime locations.

Accuracy is evaluated as,

$$\text{Accuracy} = \frac{(\text{True positive} + \text{True negative})}{(\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative})}$$

4.1.1.1 Precision

Precision value is determined based on the retrieval of information at true positive prediction, false positive.

Precision =True Positive/ (True Positive + False Positive)

4.1.1.2 Recall

The Recall value is determined based on the retrieval of information at true positive prediction, false negative. Recall in this context is also referred to as the True Positive Rate. In that process the fraction of relevant instances that are retrieved.

Recall =True Positive / (True Positive + False Negative)

The kidnap, dowry death and dacoity crime data's are analysed particularly to find out how well the proposed methodology can predict the hot spot where the number of data are missing in number.

The comparison graph for Accuracy is depicted as:

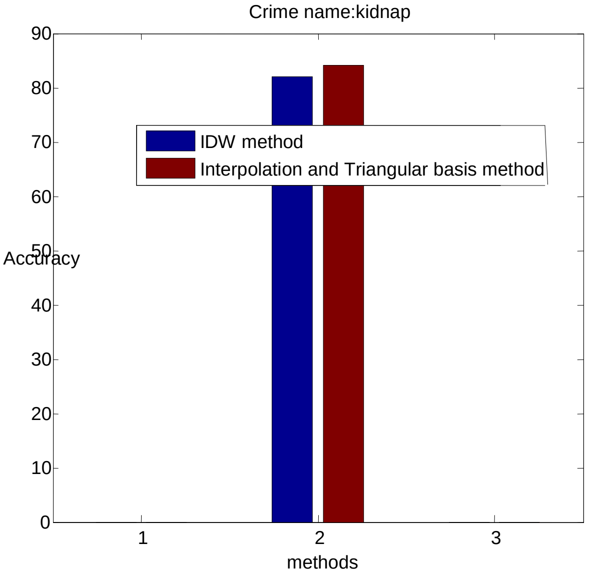

Fig 4: Accuracy comparison

In the above graph, accuracy value is compared where the X axis denotes the methodology and Y axis denotes the Accuracy values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved accuracy where the hot spots are mitigated accurately. From the performance measure values, it can be proved that the proposed methodology gives 20% improvement than the existing method.

The comparison graph for Precision is depicted as:

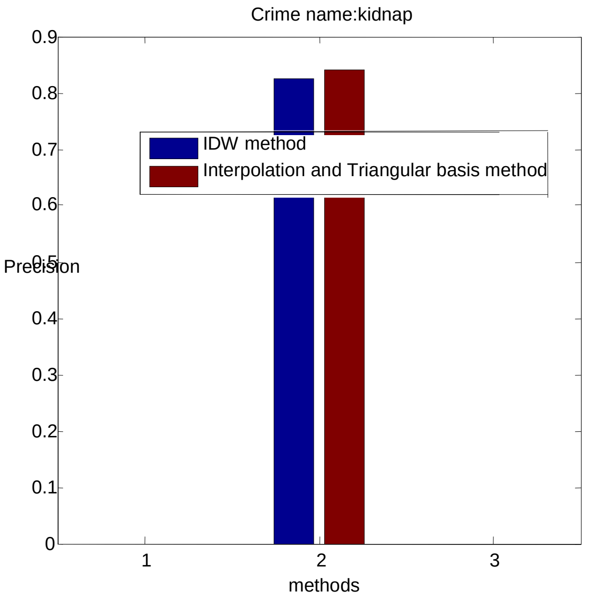

Fig 5: Precision comparison

In the above graph, precision value is compared where the X axis denotes the methodology and Y axis denotes the precision values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved precision rate. From the performance measure values, it can be proved that the proposed methodology gives 8% improvement than the existing method.

The comparison graph for Recall is depicted as:

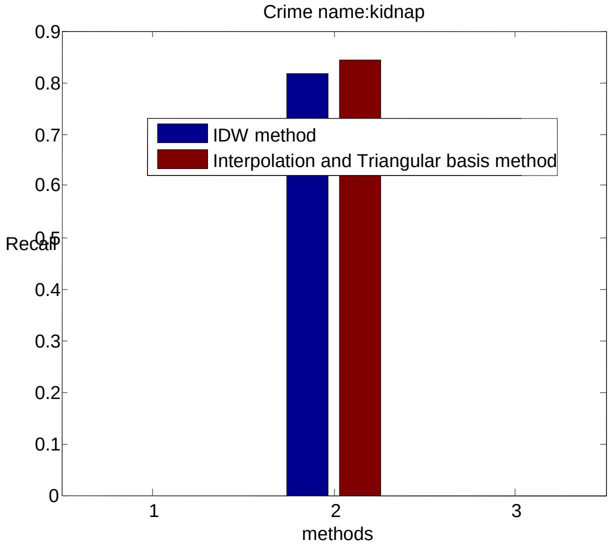

Fig 6: Recall comparison

In the above graph, recall value is compared where the X axis denotes the methodology and Y axis denotes the recall values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved recall rate. From the performance measure values, it can be proved that the proposed methodology gives 25% improvement than the existing method.

The comparison graph for Accuracy is depicted as:

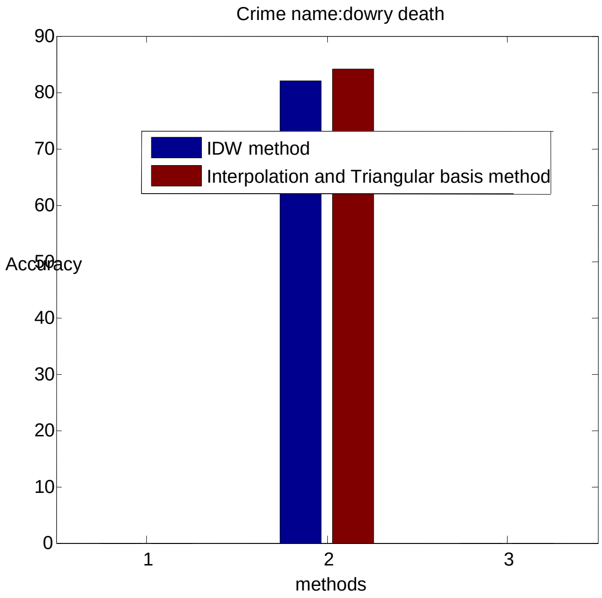

Fig 7: Accuracy comparison

In the above graph, accuracy value is compared where the X axis denotes the methodology and Y axis denotes the Accuracy values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved accuracy where the hot spots are mitigated accurately. From the performance measure values, it can be proved that the proposed methodology gives 5% improvement than the existing method.

The comparison graph for Precision is depicted as:

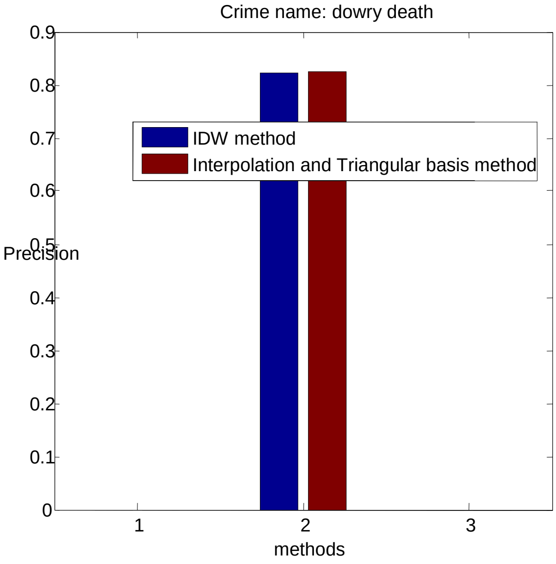

Fig 8: Precision comparison

In the above graph, precision value is compared where the X axis denotes the methodology and Y axis denotes the precision values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved precision rate. From the performance measure values, it can be proved that the proposed methodology gives 6% improvement than the existing method.

The comparison graph for Recall is depicted as

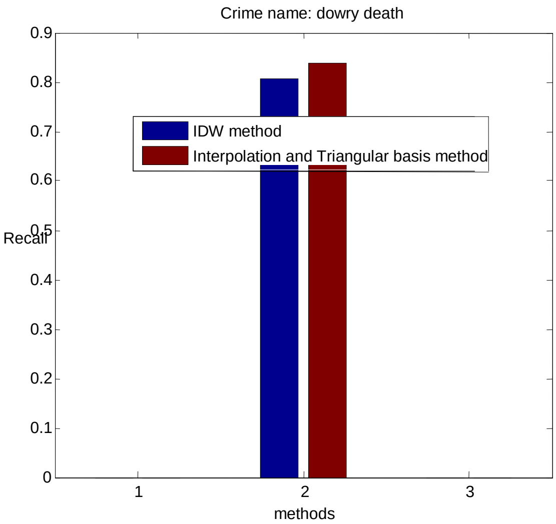

Fig 9: Recall comparison

In the above graph, recall value is compared where the X axis denotes the methodology and Y axis denotes the recall values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved recall rate. From the performance measure values, it can be proved that the proposed methodology gives 30% improvement than the existing method.

The comparison graph for Accuracy is depicted as

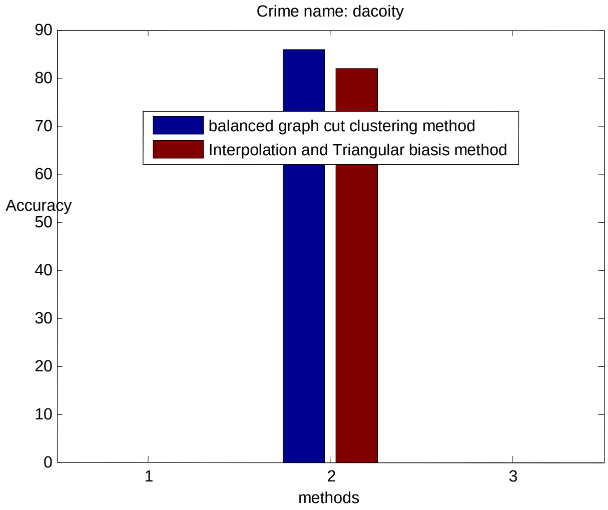

Fig 10: Accuracy comparison

In the above graph, accuracy value is compared where the X axis denotes the methodology and Y axis denotes the Accuracy values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved accuracy where the hot spots are mitigated accurately. From the performance measure values, it can be proved that the proposed methodology gives 12% improvement than the existing method.

The comparison graph for Precision is depicted as

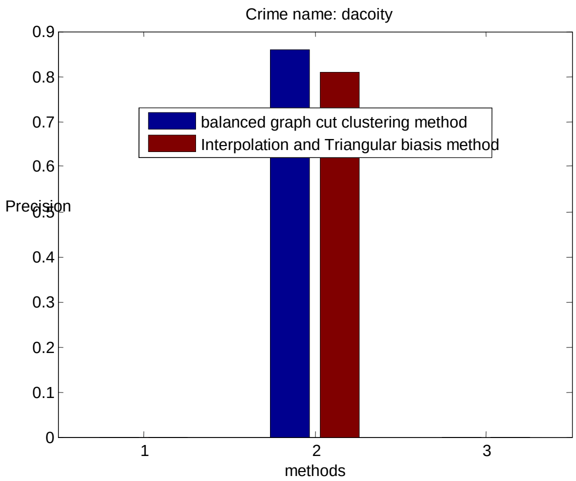

Fig 11: Precision comparison

In the above graph, precision value is compared where the X axis denotes the methodology and Y axis denotes the precision values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved precision rate. From the performance measure values, it can be proved that the proposed methodology gives 35% improvement than the existing method.

The comparison graph for Recall is depicted as

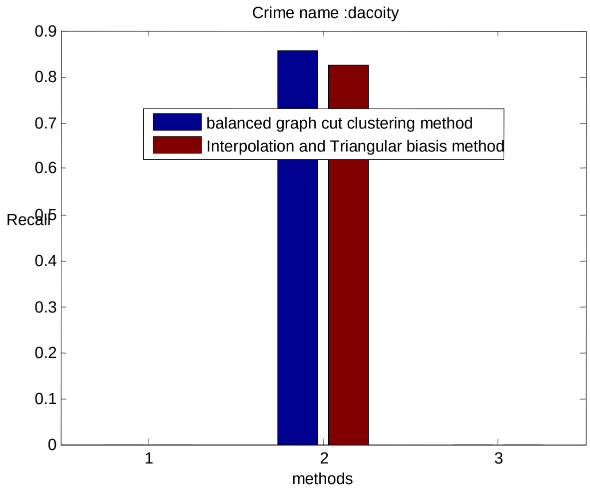

Fig 12: Recall comparison

In the above graph, recall value is compared where the X axis denotes the methodology and Y axis denotes the recall values. From this graph, it can be proved that the proposed methodology provides better result than the existing approach in terms of improved recall rate. From the performance measure values, it can be proved that the proposed methodology gives 40% improvement than the existing method.

5. CONCLUSION

Hot spot mitigation of serial crime plays an important role in the investigation department which will reduce the burden of having more human resources and spending more time in paper work. In this work, triangulation based interpolation method along with the graph measures are introduced which focus to detect the hot spots in terms of similar types of crimes. This methodology can predict the more number of crimes that are identical to each other in terms of improved privacy and accuracy. The experimental tests conducted proves that the proposed methodology can provide the better result than the existing approach in terms of improved accuracy, precision and recall rates. In future, this work can be extended to have classification algorithms which is used to categorize the images based on different types of crimes that varies based on location characteristics.

6. REFERENCES

[1] Hsinchun Chen, Wingyan Chung, Yi Qin, Michael Chau, Jennifer Jie Xu, Gang Wang, Rong Zheng, Homa Atabakhsh, “Crime Data Mining: An Overview and Case Studies”, AI Lab, University of Arizona, proceedings National Conference on Digital Government Research, 2003, available at: http://ai.bpa.arizona.edu/

[2] Smith T.R., Menon, S., Starr, J.E., (1987). Requirement and Principles for the implementation and construction of large scale geographic information system.International Journal of Geographic Information System

[3] Sahu and Peeyush (2011) Effective crime control using GIS gisdevelopment.net.

[4] Guta R., Rajitha K., Basu S. and Mittal S.; Application of GIS in Crime Analysis: A Gateway to Safe City, India Geospatial Forum, 2012

[5] http://aurangabadcitypolice.gov.in/police_juridiction.php , 19-09-2013.

[6] M. Vijaya Kumar and Dr. C. Charasekar; Spatial Statistical Analysis of burglary Crime in Chennai City Promoters Apartments: A Case Study, IJETT, 2011.

[7] http://www.gisresources.com/types-interpolation - methods_3 /; accessed on 29-04-2014.

[8] Jitendra Kumar, Sripati Mishra and Neeraj Tiwari; Identification of Hotspots and Safe ones of Crime in Uttar Pradesh, India: Geo-spatial Analysis Approach; IJRSA, 2012.

[9] Shyam Varan Nath, S.,2007, “ Crime data mining, Advances and innovations in systems, K. Elleithy (ed.),” Computing Sciences and Software Engineering, PP. 405-409

[10] Devesh Bajpai, 2012, “Emerging Trends in Utilization of Data Mining in Criminal Investigation: An Overview”, Journal of Environmental Science, Computer Science and Engineering & Technology, Vol.1, Issue.2, PP.124-131

[11] John David Elijah Sandig, Ruby Mae Somoba, Ma. Beth Concepcion and Bobby D. Gerardo, 2013,” Mining Online GIS for Crime Rate and Models based on Frequent Pattern Analysis”, Proceedings of the World Congress on Engineering and Computer Science, Vol.2, PP.23-27

[12] TongWang , Cynthia Rudin, Daniel Wagner and Rich Sevieri, 2013, “Detecting Patterns of Crime with Series Finder”, Twenty-Seventh AAAI Conference on Artificial Intelligence, PP.140-142.

[13] Kate Bowers,J., Shane Johnson,D., and Ken Pease ,2004,“Prospective Hot-Spotting The Future of Crime Mapping?” Brit J Criminol, vol: 44, PP.641-658

[14] Arbind Kumar Singh., Manimannan, G., 2013,“Detecting Hot Spots on Crime Data Using Data Mining and Geographical Information System” ,Int J of Statistika and Mathematika, ISSN: 2277- 2790 E-ISSN: 2249-8605, Volume 8, Issue 1, PP 05-09.

[15] Jitendra Kumar, Sripati Mishra and Neeraj Tiwari, 2012, “ Identification of Hotspots and Safe Zones of Crime in Uttar Pradesh, India: Geo-spatial Analysis Approach”, Int J Remote Sensing Applications, Vol.2, Issue.1, PP.15-19.

[16] Xiang Zhang et al., 2010, “Detecting and mapping crime hot spots based on improved attribute oriented induce clustering,” IEEE International conference on Geoinformatics, PP.1-5.

[17] Timothy, C., Hart Paul ,A., and Zandbergen,2012,”Effects of Data Quality on Predictive Hotspot Mapping” ,National Institute of Justice ,US Department of Justice,Office of Justice Programs,United States of America.

[18] Kadhim ,B and Swadi Al-Janabi, 2011, “A Proposed Framework for Analyzing Crime Data Set Using Decision Tree and Simple K-Means Mining Algorithms”, Journal of Kufa for Mathematics and Computer, Vol. 1, Issue.3, PP.8-24.

[19] Aniruddha Kshirsagar, Lalit Dole, 2014, “A Review On Data Mining Methods For Identity Crime Detection,” International Journal of Electrical, Electronics and Computer Systems, Vol.2, Issue.1, PP. 51-55.

[20] Grubesic, Tony H., and Alan T. Murray. "Detecting hot spots using cluster analysis and GIS." Proceedings from the Fifth Annual International Crime Mapping Research Conference. Vol. 26. 2001

[21] David A. Bader, Shiva Kintali, Kamesh Madduri, and Milena Mihail, "Approximating Betweenness Centrality" Proceedings of the 5th international conference on Algorithms and models for the web-graph, Pages 124-137, 2007

[22] Aaron Clauset, "Homophily and assortative mixing", Network Analysis and Modeling, 2013

[23] https://en.wikipedia.org/wiki/Assortativity

[24] E. A. Leicht, Petter Holme, and M. E. J. Newman, "Vertex similarity in networks", Disordered Systems and Neural Networks (cond-mat.dis-nn); Data Analysis, Statistics and Probability

[25] M. Vijaya Kumar and Dr. C. Charasekar; Spatial Clustering Simulation on Analysis of Spatial-Temporal Crime Hotspot for Predicting Crime activities, IJCA, 2011.