



# ACCIDENT PREDICTION BASED ON ACCIDENT TYPES USING SPATIOTEMPORAL CLUSTERING ALGORTIHMS

Dara Anitha Kumari<sup>1</sup>, Dr. A. Govardhan <sup>2</sup>,

<sup>1</sup> Research Scholor, JNTUH, Hyderabad,  
anithakumaridara@gmail.com

<sup>2</sup>Principal, JNTUH college of Engineering,  
Hyderabad,  
govardhan.cse@jntuh.ac.in

June 12, 2018

## Abstract

This paper concentrate on finding and understanding road accidents occurring in specific location and time based on the type of accident ,severity ,road surface, time (day -night),working hours, weather condition. Using spatiotemporal algorithms, techniques, tools, methods help to predict and prevent the future accident. Spatial data mining is the procedure of finding the interesting and previously unknown ,but potentially useful patterns from the spatial database. Major need of this paper improve road safety and accident prediction. Traditional data mining analytical techniques cannot easily locate in new hidden implicit patterns and their relationships . Spatiotemporal algorithms involve deeply and deal with the large dataset.

**Keywords:**Spatiotemporal data mining , Spatial data mining, Clustering algorithm , Road safety, Accident prediction.

## 1 INTRODUCTION

Spatial data mining refers to the extraction of knowledge spatial relationships, or other interesting patterns not explicitly stores in spatial database, spatial data analysis has been a popular approach is to analyze spatial data exploring geographic information. [1][2] Spatial data are data that have a spatial component, that indicates the data are connected to a place on earth , Any data which is directly or indirectly referenced to a location on the surface of the earth are called spatial data mining , it refers to all types of data objects , or elements that are present in graphic space or origin , nature of spatiotemporal data mining is based on location or space and certain time , objectives are spatial, thematic or non spatial ,and temporal. [3][4][5] Some event types are including like earthquake, hurricanes, and road traffic and road accidents. Spatial and temporal helps to identify the features and their relationships even in large dataset. Represent data in spatial data mining has three different data types .

**Raster data type:** Data represent in raster total data will divide into grids ,rows and columns and that intersection point is called as a pixel ,it is having space or location type of data like google earth ,satellite imagery data and digital camera photos.

**Vector data type:** It consists of points, lanes, and polygons.

**Graph data type:** It consists of data as node, edge, path, like GPS (Global Positioning System)it is having a route maps, intersection nodes and graphs.

Spatial data mining task and spatial data mining technique deals with the large datasets then help to understand the data easily visualization and representation. Classification deal for construct the class of the object according to its attribute. Association rules describe the patterns based on spatially related rules from database.[6]An spatial predicates rather than objectives. Spatial clustering is grouping the object from database into meaningful subclasses , objects in one cluster are having more similarities than objects from different clusters . [7]Spatial data mining applications are NASA(National Aeronautics and Space Administration),NIMA(National Imagery and Mapping Agency),NCI(National Cancer Institute ),USDOT(United States

Department of Transportation), it spread more applications domains, including ecology, environmental management ,public safety, transportation, earth science , epidemiology, climatology, meteorology ,biology, crop sciences , forestry , and medicine.

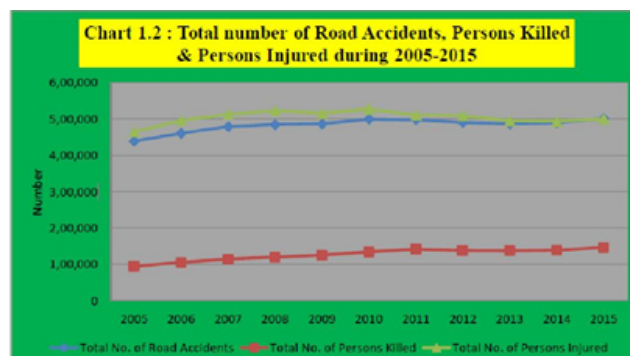
Road accidents ,fatalities, injuries have a great impact an people and society , loose the public asserts, accidents play a major role in damages .India is one of the highest motorization growth rate in the world ,but it is facing dangerous major accidents thus road safety is a severe issue in both national and international . Every four second one death is happening in Indian roads because the driver's behavior ,drunk and drive, rash driving, ignorance of traffic rules .Per day 400 people are dying and 1317 accidents happening in India only.

<b>Number of Road Accidents and Number of persons affected (2005-2015)</b>				
<b>Year</b>	<b>Number of Accidents</b>		<b>Number of persons</b>	
	<b>Total Accidents</b>	<b>Total Fatal</b>	<b>Total Killed</b>	<b>Total Injured</b>
<b>2005</b>	4,39,255	83,491	94,968	4,65,282
<b>2006</b>	4,60,920	93,917	1,05,749	4,96,481
<b>2007</b>	4,79,216	1,01,161	1,14,444	5,13,340
<b>2008</b>	4,84,704	1,06,591	1,19,860	5,23,193
<b>2009</b>	4,86,384	1,10,993	1,25,660	5,15,458
<b>2010</b>	4,99,628	1,19,558	1,34,513	5,27,512
<b>2011</b>	4,97,686	1,21,618	1,42,485	5,11,394
<b>2012</b>	4,90,383	1,23,093	1,38,258	5,09,667
<b>2013</b>	4,86,476	1,22,589	1,37,572	4,94,893
<b>2014</b>	4,89,400	1,25,828	1,39,671	4,93,474
<b>2015</b>	5,01,423	1,31,726	1,46,133	5,00,279

Figure 1: Road Accidents in India 2005-2015

This below the Figure shows road accident deaths effected in major states in India, based on statistic data is giving that 'accidents' happening due to the newest vehicles in most accidents ,mostly from the last ten years .

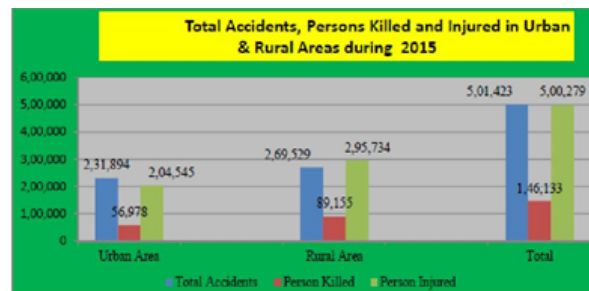
road accident causes injuries, fatalities, sever loses, disabilities, hospitalization, and economical loose . Based on this road accident statistics data help to build so many strategies to prevent



Top* 13 States: Share in Total Number of Persons Killed in Road Accidents (in %) in 2015		
Share of 13 States	83.6	1,22,153
1. Uttar Pradesh	12.1	17,666
2. Tamil Nadu	10.7	15,642
3. Maharashtra	9.0	13,212
4. Kamataka	7.4	10,856
5. Rajasthan	7.2	10,510
6. Madhya Pradesh	6.4	9,314
7. Andhra Pradesh	5.7	8,297
8. Gujarat	5.6	8,119
9. Telangana	4.9	7,110
10. West Bengal	4.3	6,234
11. Bihar	3.7	5,421
12. Punjab	3.3	4,893
13. Haryana	3.3	4,879
*: Top 13 according to their respective shares in 2015		

and the prediction of the road accident ,and improve the road safety, awareness of traffic rules, and build effective proper infrastructure roads.

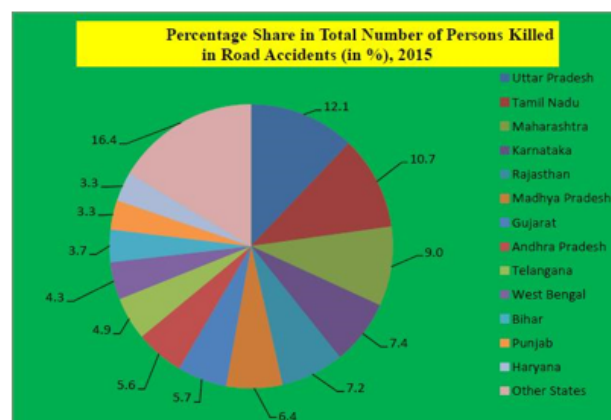
[15][16]Compare with rural and urban areas road accident totally rural areas are more than urban areas in fatal accidents, total accidents ,person killed and injuries, because of road infrastructure ,and not available timely help, treatment. So many people are dying for not getting proper help from nor polices and medical treatment. In spatiotemporal algorithms deal to identify time and space ,and prevent accidents ,helps to improve traffic condition, gives better planning.



Total Accidents, Persons Killed and Injured in Rural & Urban Areas during 2015				
(in Numbers)				
Category	Fatal Accidents	Total Accidents	Person Killed	Person Injured
Urban Area	50,959 (38.7)	2,31,894 (46.2)	56,978(39.0)	2,04,545 (40.9)
Rural Area	80,767 (61.3)	2,69,529 (53.8)	89,155 (61.0)	2,95,734 (59.1)
Total	1,31,726 (100.0)	5,01,423(100.0)	1,46,133(100.0)	5,00,279(100.0)
Note: Figures in parentheses indicate share of the total.				

### Spatial Clustering Methods :

[9]Clustering is an important research analysis in spatial data mining , this is called as unsupervised learning technique, it creates natural class labels ,clustering is a grouping data objects based on the similarities, and characteristics ,and compare with



Road Accident parameters 2014, 2015, and 2016			
Parameters	2014	2015	2016
Total number of Accidents in India	4,89,400	5,01,423	6,31,437
Total number of Accident deaths in India	1,39,671	1,46,133	1,50,784
Total number of Accident injuries in India	4,93,474	5,00,279	4,80,653

Figure 2: Comparison of the last three years An Road Accident Data

other clusters having high dissimilarity with another clusters. [8]The Cluster do not requires predefined data like classification. In this paper we compare and find out they differentiate two cluster algorithms , DBSCAN and Grid Based Algorithms.

#### **Comparison of Two Clustering Algorithms:**

**DBSCAN** :(Density Based Spatial Clustering of applications with Noise) .The data should contain data points , and limited threshold ,collectively data is formed together closely packed together with nearest data points. It is able to work in discovery of arbitrary shapes ,and avoid noise gives the representation of algorithm.

**OPTICS** (Ordering Points to Identify the Clustering Structure):[10]It is difficult to determine and construct a cluster in the real world and large data sets ,in this parameter values are very sensitive, overcome that difficulty in global analysis , parameters we can use OPTICS clustering analysis algorithm, a cluster ordering points in a data set represented as graphically ,it helps to visualize and understand easily ,OPTICS and DBSCAN structures are very similar and also the same time complexity is presented .

**DENCLUE**(Clustering Based on density Distribution Functions)DENCLUE (DENsity-based CLUstEring) to create DENCLUE first give the density estimation ,density estimation is the major issue in density-based algorithm, DENCLUE has invariant against noise ,it can easily takeout and reduce noise .

**Advantages:** [11]No need prior number of clusters in the data like K-means, it can find a cluster even if it is completely surrounding different cluster ,it has a complete notion of noise, it is an robust system t outliers ,from the database it just need minimum points more than two, understand the data very well.

**Disadvantages:** It is not completely deterministic, always depends upon the ordering data points, it treats border points as noise so that it need consist statistical interpretation of density connected points ,if the distance is large densities in dataset it cannot create the cluster ,distance threshold will be difficult

**Grid Based Methods:** [12]This clustering approach used as ,a multi resolution data structure , with the clustering operations mechanism the object space into a more number of cells from grid

structure, this approach method has fast processing time and not depending on the number of objects, it is using two methods.

**STING**(Statistical Information Grid): [12][13] In this spatial area (latitude, longitude, dimensions) is the input object convert to rectangular cells, the STING is a hierarchical structure, having so many advantages here each cell represented as the information of an grid cell, working processing and query independent, it gives efficiency results, but all the cluster boundaries are horizontal or vertical, and it cannot detect diagonal direction. Query-independent approach algorithms and structures are trivial to parallelize, no need to re compute all information when new data comes it is automatically do an incremental update. We can use it for high dimensionality, It can perform clustering as spatial data, and it is a region oriented works efficiently, create an spatial table of location also.

**CLIQUE**(Clustering In QUEst) :[15] It is a bottom -up subspace clustering, it helps to find out the 'data points' objects even in dense space, but accuracy of the cluster is degraded, it works well in lower dimensionality, it may overlap in dense regions.

#### **Issues And Challenges:**

1. Design and development of required spatiotemporal technique and data types and data structures.
2. Find out the major accident prone zone with effective representation and visualization of objects.
3. Resolve Fundamental issues in data handling, analysis and mining.
4. In spatiotemporal calculating space and time is the major issue. Need extension for the algorithms effectively.
5. Developing, representing, geometrical and spatiotemporal data manipulating in spatiotemporal attributes.
6. Need advanced qualitative and extensive approaches on large datasets in spatial data mining, based on time, speed, accuracy, occupy the space and generalization.



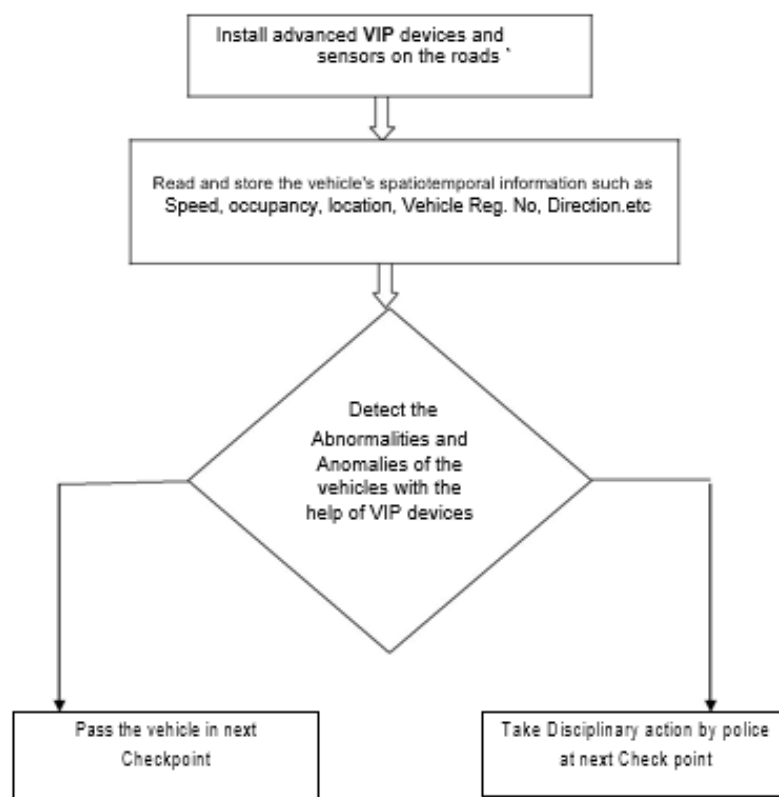


Figure 3: Prevent accidents by prediction based on this flow chart

### **VIP : Video Image Processor.**

It is the process of both hardware and software which extract the desired information from where data provided by an image sensors. Abnormal things like crossing the speed limit, violating traffic rules, Taking wrong directions, risky overtakes, driving without a license, driving without seatbelt, for two wheeler vehicles without helmet, major risky drunken drive, not having pollution certificate, not having vehicle insurance, not having vehicle passing certificate, consider all above thing, then detectors and advanced sensors can identify stores and send that data into the databases.

## **2 CONCLUSION AND FUTURE WORK**

Road accidents and damages are increasing day by day ,preventing the involvement of advanced and highly scalable spatial data mining algorithms ,data is rapidly growing because of using sensor networks and location aware devices storing images ,video, audio files, statistic dynamic data .Spatial data mining is very interesting ,promising and fruitful domain nowadays, with the advanced devices availability ,in this paper discuss about an urgent need to development and prevention of road accident ,decrease the accident, damages, fatalities and deaths supporting spatial data mining techniques like Grid based methods in an cluster moving objects ,getting the knowledge discovery information about advanced tools ,and devices. Grid based method ,STING, and CLIQUE gives the expected results even in moving clusters with huge data sets.

## **References**

- [1] K. Venkateswara Rao ., A .Govardhan .,K .V.Chalapathi Rao .:Spatiotemporal data mining :Issues ,Tasks and Applications. IJCSSES (2012)

- [2] Sachin Kumar ., Durga Toshniwal.: A data mining approach to characterize road accident location. J.Mod .Transport(2016)
- [3] Aravin Kumar ., Aanchal Kakkar .,Rana Majumdar ., Anurag Singh Baghel .:Spatial Data Mining :Recent Trends and Techniques.IEEE(2015)
- [4] Bimal Ghimire .,Shrutilipi Bhattacharjee.,Soumya K .Ghosh .:Analysis of spatial Autocorrelation for Traffic Accident Data based on Spatial Decision Tree.IEEE(2013)
- [5] Dara Anitha Kumari .,Dr.A.Govardhan .:A Comparison Study on Traffic Risk Assessment and Road Safety Analysis Using Spatial Data Mining Techniques.CiiT(2017)
- [6] Emerson D., Nayak R., Weligamage.: Identifying Differences in safe roads and crash prone roads using clustering data mining.
- [7] V.Prasannakumar ., H.Vijith., R.Charutha., N.Geetha.: Spatio-Temporal Clustering of Road Accidents : GIS Based Analysis and Assessment.(2011)
- [8] Hailing Jin ., Baoliang Miao.: The Research Progress of Spatial Data Mining Technique.
- [9] Karolien Geurts ., Isabelle Thomas ., Geert Wets .: Understanding spatial Concentration of road accidents using frequent item set.(2005)
- [10] Jeffrey Erman ., Martin Arlitt ., Anirban Mahanti.: Traffic Classification using Clustering Algorithms.(2006)
- [11] Panos Kalnis., Nikos Mamoulis ., Spiridon Bakiras.: On Discovering Moving Clusters in Spatio-Temporal Data.(2014)
- 12.Amar Mani Aryal ., Sujing Wang.: Discovery of Patterns Using Spatio- temporal data using clustering Techniques.(2017)
- 13.Roger Bivand .: A Review of Spatial Statistical Techniques for location studies (1998)
- [12] YU Zheng ., Yukun Chen ., Quannan li ., Xing Xie ., Wei-Ying Ma .: Understanding Transportation Modes Based on

GPS data for web applications .(2010) 15.Bing Zhang ., Goce Trajcevski .,Feiying Liu.: Clustering Speed in Multi -Lane Traffic Nteworks .(2016)



