Detecting traffic accident clusters in India with local spatial statistics

Pratyush Mishra, Achyanta Kumar Sarmah

Abstract

Road accidents are a human tragedy. They involve high human suffering and monetary costs in terms of untimely death, injuries and social problems. Road accidents often exhibit discernible spatial and temporal patterns. In general, the dynamics of urban transportation systems are largely the product of interactions among various components of the urban system and of human activities in space and time, particularly transportation and urban land use. Totally avoiding accidents may not be possible for any traffic system. However a scheduled monitoring and management of accidents could be the most effective way to control the factors leading to the occurrences of accidents. This research work proposes to achieve this very goal. Working upon the data set of accidents in cities of different states of India for the period of 2011 to 2015, as made available by the Government of India, this work proposes a dashboard based monitoring and management system for the Regional Transport Offices (RTO). With a two-tiered approach, the proposed system comprises of a relational database for the back-end data service tier and a front-end dashboard based client tier for the RTOs. The system incorporates a national level yearly review mechanism of the accident trends and patterns in various states. The results of review would achieve a minimal number of accidents that would be considered unavoidable by the road traffic agencies. Further on, a 4E (Education, Engineering, Enforcement and Emergency care) based strategy framework for clustering the regions around these presumably minimal number of unavoidable accidents is proposed in order to minimize and manage traffic accidents.

Keywords: Accidents, Traffic control

1. Introduction

Road and traffic accidents are unpredictable incidents and their analysis requires the understanding of the factors involving them. Road and traffic accidents are defined by a group of variables that are largely of discrete nature. The major problem in the analysis of accident data is its heterogeneous nature

Email address: pratyushm@iitkgp.ac.in (Pratyush Mishra, Achyanta Kumar Sarmah)
Corresponding author

Preprint submitted to International Journal of Scientific and Engineering Research December 22, 2018

[1]. Researchers used segementation of the data to diminish this heterogeneity utilizing a few measures, for example, expert knowledge, however there is no certification that this will prompt an optimal segmentation which comprises of homogeneous groups of road accidents [2]. Hence, cluster analysis can assist the segmentation of road accidents.

This paper proposes a dynamically strategic system that is based on the cluster analysis using “K mode” algorithm. A schematic of the system is given in Figure 1.

The paper is organized as follows: In Sect. “Material and methods”, integrated methodolgy adopted and original data source is presented to allow the work to be reproduced by an independent researcher. Next, a description of the “Theory” is presented. In Sect. “Results and system Design”, the results and system design are elaborated . The significane of report is present in Sect. “Discussion and Conclusion”.

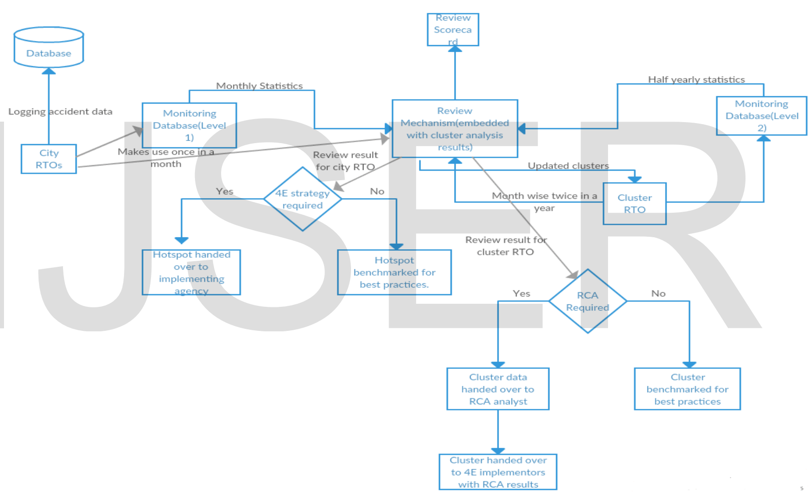

Figure 1:

2. Material and methods

Data Source

The dataset used was published on Open Government Data Portal on Jan 16, 2017. It contains road accident details for 52 cities for 4 years starting from 2011.

Data preprocessing

In this step, we preprocessed the accident data in order to make it appropriate for the analysis. As accident data was not available for the year 2011 in case of few cities, median of the available data was used for that year. We did away with the types of accident, and only considered the total number of accidents, people killed and injured year wise across all the cities.

design tools

The 2-tiered architecture has been used to design the logging and monitoring system. Database tools like entity relationship diagrams and normalization has been used for designing the back-end data service tier. The front-end client tier has been designed as dashboards using system tools like data flow diagram and flowcharts.

Strategy framework

A 4 E's – Education, Engineering, Enforcement and Emergency care. based approach has been taken for designing the strategy framework that would aim at maintaining the minimal number of unavoidable accidents in the various regions.

Education for Safer People

Education is crucial in developing safer people and safer communities; therefore, road safety must be taught in schools as well as through driver trainings and safety campaigns.

- School Education
  - Periodic Road safety courses embedded in school curriculum.
- Training
  - RTO guided basic road safety training for novice drivers.
  - Moreover, advanced training targeting repeated offenders/re-offenders should be considered and should be coupled with crash investigation training for police officers in order to well-document the causes of accidents and collect reliable crash data.
- Campaigns
  - Periodic road safety campaigns by RTOs of hotspots and states.

Engineering for a Safer System

Effective engineering plans take into consideration both infrastructure and technology. A well designed infrastructure reduces road hazards and promotes the safety of road users.

- Infrastructure
  - Improvement of road designs including the strategic allocation of roadside-barriers, creation or improvement of pedestrian walkways and crossings, appropriate implementation of speed management infrastructure, e.g. speed bumps.
- Technology
  - Vehicle-related technologies , for example: Seatbelts, airbags, Advanced Braking Systems (ABS). Compulsory inspection of old vehicles(older that 15 years) in every 2 years.

– Infrastructure-related technologies, for example: Advanced radars and speed cameras, point-to- point systems, Variable Speed Limit (VSL) signs

Enforcing Road Safety Laws

Enforcement is critical in putting education and engineering efforts into effect, in order to achieve a safer system.

- Enforcement could be achieved either in the form of effective sanctions or through policing. Effective sanctions in the form of monetary fines, license suspension etc. followed strictly and in a timely fashion proves to be a good deterrence for drivers with less concerns towards road safety. Policing through technology, either by covert or high visibility patrolling proves to be a powerful deterrent for poor driving behavior.

Multidisciplinary Approach in Emergency Care

Proper and timely emergency care has always proven to be effective in reducing casualties happening from accidents. A multidisciplinary approach for emergency care should involve the following

- Mechanism to give first-aid on the spot of the accident.
- Stringent mechanism for crowd control in the accident site - In many cases it is seen that the gathering crowd becomes reactive in an accident scene. This further worsens the condition of the affected, rather than elevating them.
- Mechanism for located medical care of the affected - People affected in accidents are usually in a state of shock as the event is a totally unforeseen and sudden one. Hence, after the first aid, the affected should not be moved around too much in the name of treatment. Nearer the post first aid treatment location, better are the chances of quicker recovery for the affected.

DATA

As the original data set didn't had sufficient parameters for clustering, we added other components of urban systems and human activities -

- City wise number of total licensed vehicles for a period of 10 years starting from 2004. By this we aim to study the effect of motorization growth rate on accident rates.
- State wise accident data for a period of 10 years starting from 2004. As the city wise accident data was available only for 5 years, state wise data was included to expand scope of analysis.
- National Highways in and around the city were added.

- Medical Infrastructure in the city as on 2011 and 2014 (District hospitals, CHCs, PHCs and Sub- Centers). The motive is to analysis the effect of health infrastructure data on number of fatalities resulting from injuries caused by road accidents.

- State wise health infrastructure as on 2004 and 2014 was added to analyze the growth of medical facilities and its impact on number of fatalities. State wise identified accident hotspot with accident data for the year 2015 was also included as hot-spot analysis is deemed a key issue in traffic safety strategies. The identification of traffic accident hot spots is the first essential step for the appropriate allocation of resources for safety improvements.

The objective of clustering algorithm is to divide the data into different clusters or groups such that the objects within a group are similar to each other whereas objects in different clusters are different from each other. Different clustering techniques like K means, Latent Class Clustering (LCC) have been used in road accident analysis. LCC is a widely used clustering technique which provides several cluster selection criteria to determine the number of clusters. LCC has been widely used for analysis of road accidents to identify clusters in accident data. However if the data contains large number of categorical attributes, LCC can be computationally infeasible and thus K-modes algorithm can be a better option. The problem of selecting K in K-modes algorithm can be overcome by using cluster selection criteria of LCC analysis.

K modes are faster and efficient than LCC in producing locally minimal clustering results. In this paper, we are making use of both K modes clustering and cluster selection criteria of LCC cluster analysis.

The K-modes clustering technique is an enhanced version of traditional k means algorithm. The major extensions to the k means algorithm to k modes algorithm is the distance measure and the clustering process which are explained below:

Distance measure

Given a data set D, the distance between two objects X and Y, where X and Y are described by N categorical variables, can be computed as follows:

$$\text{d(X,Y)}=\sum_{i=1}^{N} \delta(Xi, Yi)$$

where,

$$\delta(Xi, Yi) = \{^{0, Xi=Yi}_{1, Xi \neq Yi}$$

In the above equations, Xi and Yi are the attribute i values in object X and Y. This distance measure is often referred as simple matching dissimilarity measure. More the number of differences in categorical values of X and Y, more is the difference between the two objects. [12]

3. Theory

Dynamics of urban transportation system

The urban transportation system can be thought as responding to the social and economic forces that exit in urban areas.This urban socioeconomic environment is in turn influenced by the characteristics of the transport system.

The role of the system planner may be conceived, in general way, as the direction of his/her efforts to design a system that achieves maximum integration between the system and its environment while ensuring adequate system safety

Charles Raux[13] presented a model that allows us to illustrate and quantify the interaction between public transport supply and demand, through the impact of the parameters of fares, service frequencies and public funding. The approach the model proposed involves constructing a system of models which makes use of simulation techniques derived from systems dynamics and various elementary components obtained from econometric models. The dynamic modelling of urban systems has been widely used in transportation, particularly for aggregate long-term situations, for economic trip forecasting scenarios or for modelling the interaction between transport and land use.

Wang[14] presented a system dynamics approach based on the cause-and-effect analysis and feedback loop structures. The proposed system design model comprises 7 submodels: population, economic development, number of vehicles, environmental influence, travel demand, transport supply, and traffic congestion.

Spatial and temporal patterns of traffic accidents

A major issue with the automobile-oriented infrastructure is traffic accidents. Technology offers a spatial solution to analyzing and critically evaluating traffic accidents on roadways. By studying accidents, solutions can be discovered to make roads safer for automobile-oriented users within an automobile-oriented society.

Numerous factors influence the severity and frequency of traffic accidents. A major assumption is that the higher the interaction of an area located on the road network, the higher the amount and frequency of vehicle accidents. Compared to rural areas, urbanized areas experience higher frequencies of traffic accidents simply because there is more traffic, and more opportunity for multi-vehicle traffic accidents. Black spot analysis identifies hazardous areas on road networks. The purpose is to create safer road systems for users.

Traffic accidents fluctuate daily at different hours per day. There have been studies that analyzes the spatial and temporal patterns of traffic hot spots such as where accidents occur at certain hours throughout the day and days of the week[3]. Several of these studies focus their research on a strict aim of model building and quantifying theory in hopes of gaining accident explainability and predictability. The problem with this is that it is difficult to measure, define, and predict human behavior in conjunction with factors that contribute to the creation of a traffic accident.

Earlier attempts in traffic clustering

Karlaftis and Tarko [4] used cluster analysis to categorize the accident data into different categories and further analyzed cluster results using Negative Binomial (NB) to identify the impact of driver age on road accidents. Ma and Kockelman [5] used clustering as their first step to group the data into different segments and further they used Probit model to identify relationship between different accident characteristics. Poisson models [6] and NB models [7–9] have been used extensively to identify the relationship between traffic accidents and the causative factors. It has been widely recognized that Poisson models outperform the standard regression models in handling the nonnegative, random and discrete features of crash counts [10, 11].

4. Results and system Design

Intial cluster analysis was done both for the cities and states. Initial results, as depicted in appendix, show that cluster of cities are spread across states. Since accidents are location speciifc events at the level of hotspots, city level clustering results are used for system design. Incorporating these findings, the proposed system is designed, as depicted in figure 1. It broadly has the following modules:-

- Logging and Monitoring module :- This module consists of two dashboards, viz: Level 1 dashboards to be used by city RTOs on daily basis to log accident data, and on monthly basis to monitor accident trends in various hotspots of the city; Level 2dashboard to be used by cluster RTOs on halfyearly basis to monitor trends in cluster cities of the two RTOs, the cluster RTO is made dynamic based on the clustering data done after each year. The parametric details of the dashboard could be found in the appendix.
- Database module :- This module consists of a normalised database that shows the following entities- Accident, Hotspot, City, Cluster. An ER-diagram of the database is given in figure 2. The normalised schema for this database can be found in appendix.
- Review module :- This module involves the clustering exercise after every six months, along with the mechanism that will facilitate reviewing of accident trends against some predefined benchmarks.

Intially these benchmarks could be considered as follows :-

- For a hotspot 5% of total accidents occured in the city over the current review period of six months could be considered. Any hotspot having less than 5% of total accidents can be identified for best practices in other hotspots. If a hotspot exhibits more than 10% of total accidents occured in the city, then it would be handed over to 4E strategy implementers. This part of the review mechanism will primarily be used by the city RTOs.
- For a city RTO, 10% of total accidents during the previous review period of six months could be considered. If it happens to be more than 10% of total accidents occured, then the city data would be handed over to RCA(root cause analysis) team along with the current clustering results. The output from the RCA team would be handed over to 4E strategy implementors.

If the city exhibits less than 5% of total accidents occured, in its previous cluster, then the city would be benchmarked for best practices in other current clusters. This part of the review mechanism will be used primarily by the cluster RTOs whose cities belong to the cluster.

The review module will also comprise of a scorecard that will be keeping track of the progress of accidents happening in each of the cities towards a nationally predetermined number of accidents which is considered unavoidable and manageable.

Entities	*Atributes*
Hotspot	{ID, Hotspot Type, City}
City	{ID, Name, Hotspot, State, RTO details}
State	{Name, City, RTO details}
Accident	{ID, Hotspot, Time Stamp, Total Injured, Total Killed, Hospitalization details, Involved Vehicle, License, Cause}

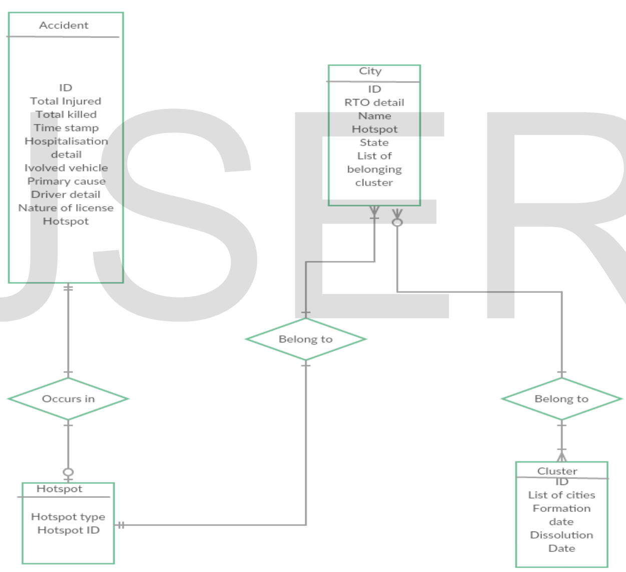

Figure 2:

5. Discussion and Conclusion

India is a signatory of "Brasilia Declaration on Road Safety" through which delegates countries agreed ways to halve road traffic deaths by the end of 2020 – a key milestone within the new Sustainable Development Goal. India is to develop and implement national road safety plans and to adopt and enforce comprehensive legislation, in line with the Global Plan for the Decade of Action for Road Safety. Necessary measures must be taken to enhance road policing strategies and traffic enforcement measures, with a view to reducing road traffic crashes, including by means of promoting integration among traffic enforcement agencies in policing and inspection, as well as collecting road infrastructure and road traffic crashes data.

References

1. Savolainen P, Mannering F, Lord D, Quddus M. The statistical analysis of highway crash-injury severities: a review and assessment of methodological alternatives. Accid Anal Prev. 2011;43:1666–76.

2. Depaire B, Wets G and Vanhoof K. Traffic accident segmentation by means of latent class clustering, accident analysis and prevention, vol. 40. Elsevier; 2008.

3. Discovering Spatial and Temporal Patterns of Traffic Accidents in Stillwater, Oklahoma, Hicks, Alston Paul

4. Karlaftis M, Tarko A. Heterogeneity considerations in accident modeling. Accid Anal Prev. 1998;30(4):425–33.

5. Ma J, Kockelman K. Crash frequency and severity modeling using clustered data from Washington state. In: IEEE Intelligent Transportation Systems Conference. Toronto Canadá; 2006.

6. Jones B, Janssen L, Mannering F. Analysis of the frequency and duration of freeway accidents in Seattle, accident analysis and prevention, vol. 23. Elsevier; 1991.

7. Miaou SP. The relationship between truck accidents and geometric design of road sections–poisson versus negative binomial regressions, accident analysis and prevention, vol. 26. Elsevier; 1994.

8. Poch M, Mannering F. Negative binomial analysis of intersection-accident frequencies. J Transp Eng. 1996;122.

9. Abdel-Aty MA, Radwan AE. Modeling traffic accident occurrence and involvement. Accid Anal Prev Elsevier. 2000;32.

10. Joshua SC, Garber NJ. Estimating truck accident rate and involvements using linear and poisson regression models. Transp Plan Technol. 1990;15.

11. Maher MJ, Summersgill I. A comprehensive methodology for the fitting of predictive accident models. Accid Anal Prev Elsevier. 1996;28.

12. Chaturvedi, Anil & E. Green, Paul & Douglas Caroll, J. (2001). K-modes Clustering. Journal of Classification. 18. 35-55. 10.1007/s00357-001-0004-3.

13. Charles Raux. A systems dynamics model for the urban travel system. AET. European Transport Conference 2003 – ETC 2003, 8 - 10 october 2003, Strasbourg, 2003, London, United Kingdom. AET, 32 p., 2003.

14. Jifeng WANG, Huapu LU ans Hu PENG, System Dynamics Model of Urban Transportation System and Its Application, Journal of Transportation Systems Engineering and Information Technology Volume 8, Issue 3, June 2008, Pages 83-89

Appendix

The brief discussion of cluster of Indian cities on road accidents is presented:

1. Mumbai, Delhi and Chennai formed three different clusters which consists of 20%, 7% and 8.3% respectively of the total accidents respectively.
2. Ahmedabad, Gwalior, Jaipur, Kochi, Kollam, Visakhapatnam, Raipur, Thiruvanthapuram and Vijaywada City comprised of another cluster with combined 15.4% of the total accidents.
3. Bengaluru and Indore formed another cluster with 9.8% of the total accidents.
4. Coimbatore, Ghaziabad, Lucknow, Meerut, Nagpur, Patna, Pune, Rajkot, Surat, Thrissur, Vadodra, Allahabad, Kanpur, Khozikode and Nashik comprised a cluster with combined 15.6% of the total accidents.
5. Asansol-Durgapur, Aurangabad, Dhanbad, Faridabad, Jamshedpur, Jodhpur, Kota, Kannur, Ludhiana, Madurai, Srinagar, Varanasi, Amritsar, Chandigarh and Tiruchirapalli comprised a cluster with combined 6.8% of the total accidents.
6. Bhopal, Hyderabad, Jabalpur, Kolkata and Mallapuram comprised the final cluster with combined 14.6% of the total accidents.

The brief discussion of cluster of Indian state on road accidents is presented:

1. Maharastra and Tamil Nadu formed two different with 14.4% and 12.9% respectively of the total accidents.
2. Arunachal Pradesh Himachal Pradesh, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, Tripura, Uttarakhand, Andaman & Nicobar Islands, Chandigarh, Dadra & Nagar Haveli, Daman & Diu and Lakshadweep comprised of another cluster with combined 1.6% of the total accidents.
3. Karnataka and Madhya Pradesh formed another cluster with 18.7% respectively of the total accidents.
4. Gujarat, Rajasthan and Uttar Pradesh comprised of another cluster with combined 11.6% of the total accidents.
5. Chhattisgarh, Haryana and West Bengal formed another cluster with combined 7.6% of the total accidents.
6. Andhra Pradesh and Kerala formed another cluster with combined 16.1% of the total accidents.
7. Assam, Bihar, Goa, Jammu & Kashmir, Jharkhand, Odisha, Punjab and Delhi formed the final cluster with combined 10.8% of the total accidents.

System design

The cluster analysis results shows that cities form clusters across states. Hence the design of the system need to incorporate a dynamic strategy for reducing road accidents, instead of a static strategy based on cities and states. A schematic of the proposed system is

- A dashboard based monitoring of accidents in hotspots

- The city Road Traffic Office(RTOs) will have a level-1 dashboard that will log the data and monitor the hotspots in the city for accidents. Parameters to be included in the database
  - Location of Accident which will help in identification of hotspots.
  - Time of Occurrence which can be helpful in developing time-slots with highest number of accidents recorded.
  - Type of conflicting vehicle(s) involved in the accident.
  - Nature of license(Regular/ Learner).
  - Primary cause of accident(Driver's fault, vehicle malfunction, damaged road).
  - Further analysis of cause of accident after classification among the above 3 groups.
    - Drivers Fault- exceeding lawful speed, unfastened seat belt, intake of alcohol/drugs, ignoring traffic signs.
    - Vehicle malfunction- Age of vehicle, Defective brakes, Defective steering, Punctured/burst tyres, Bald Tyres, other serious mechanical defect.
    - Damaged Road- Type of road(surfaced, metalled, kutcha, dry, wet), presence of pot holes, road under repair.
- The cluster RTO will have a level-2 dashboard that will monitor the level-1 dashboards of each city in its cluster. Parameters to be included in the database
  - Location and time of accident.
  - Detailed cause of analysis.
  - Nature of license so as the state can focus on better licensing system.

Guidelines for the implementaing agencies based on the 4 E strategy framework

- Public works department of the city shall focus on engineering aspects of road development that encorporates all the infrastructural changes suggested.
- RTOs across the state should have a common database of all the vehicles registered in the state with their age. City RTOs will ensure compulsory inspection of old vehicles in every 2 years and discontinuation of license to those vehicles failing the inspection.
- Department of School Education and Literacy will design the framework for introduction of road safety in school curriculums. City RTOs will act as the nodal agency for compulsory driving training for repeated offenders and suspension of licension until successful completion of the training.

– Multi-agencies coordinations to formalize adequate mechanism that ensures timely response of emergency services in case of any road accident.

Modular View of system design

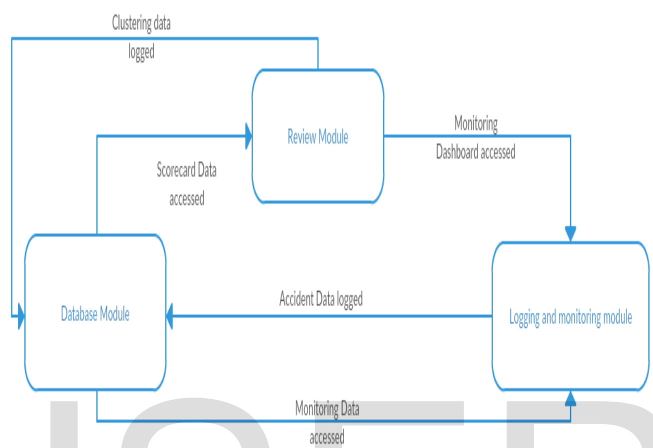

Figure 3: