Kumar and Toshniwal *J Big Data* (2016) 3:8
DOI 10.1186/s40537-016-0044-5

Journal of Big Data

RESEARCH Open Access

A novel framework to analyze road accident time series data

Sachin Kumar[1*] and Durga Toshniwal[2]

*Correspondence:
sachinagnihotri16@gmail.com
[1] Centre for Transportation Systems (CTRANS), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India
Full list of author information is available at the end of the article

Abstract

Road accident data analysis plays an important role in identifying key factors associated with road accidents. These associated factors help in taking preventive measures to overcome the road accidents. Various studies have been done on road accident data analysis using traditional statistical techniques and data mining techniques. All these studies focused on identifying key factors associated with road accidents in different countries. Road accident is uncertain and unpredictable events which can occur in any circumstances. Also, road accidents do not have similar impacts in every region of the districts. There are chances that road accident rate is increasing in a certain district but it has some lower impact in other districts. Hence, the more focus on road safety should be on those regions or districts where road accident trend is increasing. Time series analysis is an important area of study which can be helpful in identifying the increasing or decreasing trends in different districts. In this paper, we have proposed a framework to analyze road accident time series data that takes 39 time series data of 39 districts of Gujrat and Uttarakhand state of India. This framework segments the time series data into different clusters. A time series merging algorithm is proposed to find the representative time series (RTS) for each cluster. This RTS is further used for trend analysis of different clusters. The results reveals that road accident trend is going to increase in certain clusters and those districts should be the prime concern to take preventive measure to overcome the road accidents.

Keywords: Road accidents, Time series, Data mining, Clustering, Trend analysis

Background

Road traffic accident (RTA) is one of the important concerns of research as it involves fatality, personal injuries that can lead to full or partial disability and property damage. A report [1] by World Health Organization (WHO) reveals that there are 1.2 million fatal and about four times injured road accidents every year across the world. Road and traffic safety is a term associated with road accidents. The primary focus of road safety is to provide some preventive measures that can be helpful in reducing RTAs.

Road accident data analysis is an important factor that has been succeeds in identifying different factors associated with road accidents [2–5]. Once the associated road accident factors are identified the corresponding actions can be taken to overcome the accident rate and to apply some preventive measures. Road accident data analysis is mainly based on two categories: statistical techniques and data mining techniques. Various studies on

road accident data analysis used traditional statistical techniques [2, 5–9] and data mining techniques [3, 10–14, 18].

Data mining techniques [15, 16] have certain advantages over traditional statistical techniques. Data mining techniques do not require certain assumptions between dependent and independent variables which are required in traditional statistical techniques [7]. Also, data mining techniques is capable of handling large dimensional data whereas statistical techniques have some limitations [3, 13].

The road accidents may have different impact for different type of accidents at different locations [13]. Also, road accidents are also varying district wise and it may happen that certain districts have similar nature of road accidents.

Time series constitute a series of data points collected or sampled at fixed interval. Monthly road accident counts of road accident for a certain period of time also constitute a time series data. The time series data of road accidents is very important to study as it can reveal the future trend of road accidents. This future trend can help in identifying the different regions where the road accidents is tend to increase or decrease so that preventive measures can be taken. This study uses time series data of 39 districts of Gujrat and Uttarakhand states of India. This data is extracted from the road accident data provided by GVK_EMRI [17]. This data consists of 60 monthly counts of road accidents from 5 year duration from Jan-2010 to Dec-2014. It is difficult to analyze all the time series data of 39 districts individually and also based on the assumption that nature and trend of road accidents can be similar in some districts.

In previous work [10], authors tried to remove the heterogeneity in road accident data using data mining techniques and used a framework using clustering and association rule mining techniques that is capable to remove the heterogeneity from road accident data. Those techniques can certainly reveal the hidden factors behind road accidents; but they can not reveal the trend of the road accidents in different locations. In this manuscript, we are trying to elaborate road accident counts data to identify different districts where the trend of accident is increasing throughout the years. So that more focus would be on these districts to overcome the accident trend. In order to do this, we have proposed a framework to analyze road accident time series data that uses both data mining and traditional statistical techniques. This framework inputs the road accident counts for different time series and then normalizes the time series. Further, it performs agglomerative clustering (AGNES) algorithm on 26 districts of Gujrat and 13 districts of Uttarakhand. Further a time series merging algorithm is proposed to find the representative time series (RTS) for each cluster. Finally, trend analysis is performed on every RTS of different clusters.

Data set

The road accident time series data is formed from the road accident data of Gujrat and Uttarakhand districts obtained from GVK-EMRI for the period of 5 years from Jan-2010 to Dec-2014. The 26 time series data were formed for 26 districts of Gujrat state and 13 time series data were formed for 13 districts of Uttarakhand state. Each time series has 60 monthly counts of road accidents for 5 year duration.

Proposed framework

A framework is developed for trend analysis of road accident data which is illustrated in Fig. 1. The different phases of framework are discussed below:

Data preprocessing

Data preprocessing is one of the important task to be done prior to analyzing data. Data preprocessing leads to data cleaning such as removing noise, handling missing values and applying various data transformations in order to get the data ready for the analysis. Time series data also requires data preprocessing in order to get it normalized prior to the analysis. Normalization of time series prior to analysis assists in handling certain difficulties [19] such as amplitude scaling, noise, offset translation etc. Therefore, preprocessing prior to the analysis helps in getting the correct results. In order to normalize the time series data, z-score normalization method is used. Z-score method normalize a time series $T = \{t_1, t_2, t_3, \ldots, t_n\}$ into a normalized time series $NT = \{Nt_1, Nt_2, Nt_3, \ldots, Nt_n\}$ such that

$$\mu(NT) \approx 0 \text{ and } \sigma(NT) \approx 1$$

where μ(NT) and σ(NT) are the mean and standard deviation respectively of normalized time series NT. The z-score formula for normalizing time series is given by Eq. 1.

$$\mathrm{NT} = \sum_{i=1}^{n} \frac{t1 - \mu(\mathrm{T})}{\sigma(\mathrm{T})} \tag{1}$$

Similarity measure for time series

A variety of popular similarity measures [20] are exist such as Euclidean distance, dynamic time warping (DTW), correlation coefficient and triangle distance metric (TDM) similarity measure for time series data. Similarity measure is very useful and important component in clustering time series data. The outcome of similarity measure is a proximity square matrix of n dimension where n is the number of time series. In this

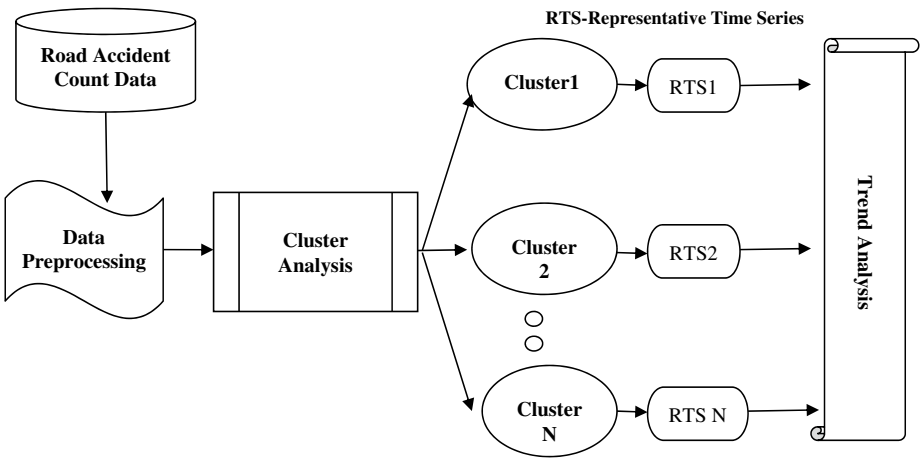

Fig. 1 Proposed framework

study, we consider Euclidean distance, DTW, Pearson correlation coefficient (PCC) and TDM.

Euclidean distance

Euclidean distance is one of the popular and classic similarity measure used in various clustering algorithms such as K-means and hierarchical clustering. Euclidean distance can be defined as the distance between two points or vectors in Euclidean norm. Euclidean distance between two time series of equal length can be computed using Eq. 2 as follows:

$$\text{Dist}(\text{T1}, \text{T2}) = \sum_{i=1}^{n} \sqrt{(T1i - T2i)^2} \tag{2}$$

the above equation calculates distance between two time series T1 and T2 of equal length of time sequence n.

Dynamic time warping

Dynamic time warping is another similarity measure for time series data which can measure distance between two time series objects even if their length is not similar [21]. The advantage of DTW is that in order to minimize the distance between two time series $s_i = \{s_1, s_2, \ldots, s_n\}$ and $t_i = \{t_1, t_2, \ldots, t_m\}$ of length n and m respectively, it optimally align s_i and t_j. The dynamic programming is used to find the similarity distance between two time series in matrix A. Matrix A is initialized to A[0,0] = 0 and A[i, j] = infinity, where i and j is not 0. The distance between two time series objects can be calculated by recursively applying Eq. 3 as given below:

$$\text{A[i, j]} = \text{d(si, ti)} + \min\left\{\text{A[i, j} - 1], \text{ A[i} - 1, \text{ j]}, \text{ A[i} - 1, \text{ j} - 1]\right\} \tag{3}$$

where A is an n × m matrix in which Euclidean distance metric is used to find the distance between s_i and t_i. The value of last cell A[n, m] represents the distance between sequence s and t.

Triangle distance metric

Triangle distance metric [22] is another popular similarity measure for time series data which considers time series object as a vector in n dimensional space. Consider $st_i = \{st_{i1}, st_{i2}, \ldots, st_{in}\}$ be an standardized time series object defined as follows:

$$\text{st}_{ij} = \frac{\sum_{(k=1)}^{n} \sum t_{ik}}{\left(\sqrt{\sum_{k=1}^{n} t_{ik}^2}\right)} \tag{4}$$

The TDM between two vectors ti and tj can be calculated using Eq. 4.

$$\text{d}\left(\text{t}_i, \text{ t}_j\right) = \frac{\sum_{(k=1)}^{n} \sum t_{ik}.t_{kj}}{\left(\sqrt{\sum_{k=1}^{n} t_{ik}^2}\right).\left(\sqrt{\sum_{k=1}^{n} t_{jk}^2}\right)} = 1 - \sum_{(k=1)}^{n} \sum st_{ik}.st_{jk} \tag{5}$$

TDM can be defined as the cosine of a triangle between two vectors and its value ranges between 0 and 2. The more different two time series objects are, the higher the value of TDM is. A low score represents more similar time series objects while higher score represents dissimilar time series objects.

Hierarchical cluster analysis

Cluster analysis mainly concerns with the grouping of data objects into one or more groups. In clustering, the objects with similar properties are assigned to same group while data objects with different properties are allotted to different groups. There are various clustering algorithms such as partition based clustering, density based clustering etc. The choice of clustering algorithm depends on type and nature of data. There are various algorithms for clustering of time series data exist. Our approach used hierarchical clustering algorithm for clustering of time series data. Hierarchical clustering algorithm is of two types—agglomerative hierarchical clustering and divisive hierarchical clustering. The time complexity of first one is $O(n^3)$ and the second one is $O(2^n)$, hence agglomerative clustering is faster than divisive clustering. We used agglomerative hierarchical clustering algorithm (AGNES) in our study. The main requirement for a hierarchical clustering algorithm is the similarity measure which takes important part in clustering process. In this study, we compare AGNES algorithm using seven versions using CPCC with Euclidean, TDM and DTW similarity measures. CPCC can be defined as a measure of the correlation between the Cophenetic distance of two time series objects and the original distance matrix. Table 1 illustrates the results from monthly road accident time series of 26 districts from Gujrat and 13 districts of Uttarakhand of India, which indicates that the performance of average AGNES algorithm is better than others with DTW as a similarity measure.

Time series merging

Cluster analysis results in homogeneous segments of the time series data. Each cluster consists of various time series objects that are similar in nature. Hence there is a need to form a representative time series which can represent the entire time series. To find the time series that can represent the entire cluster, a time series merging algorithm is formed, that takes DTW distance to calculate the closest time series. We took DTW for merging time series objects because of the results obtained from Table 1. Also, an example of time series merging algorithm is given in Fig. 2.

Table 1 CPCC results for different versions of AGNES algorithm

AGNES algorithm	Cophenetic correlation coefficient (CPCC)		
	DTW	TDM	Euclidean
Single	0.7378	0.7101	0.6650
Complete	0.5690	0.5124	0.5266
Average	*0.7949*	*0.7692*	*0.7411*
Ward	0.6003	0.6306	0.6199
Weighted	0.7186	0.6916	0.6886
Median	0.6602	0.6284	0.5627
Centroid	0.7630	0.7243	0.6872

The number in italics shows the highest CPCC value for DTW, TDM and Euclidean distances under average version. Hence, Average AGNES algorithm using DTW has been selected for clustering the data because it has highest CPCC value

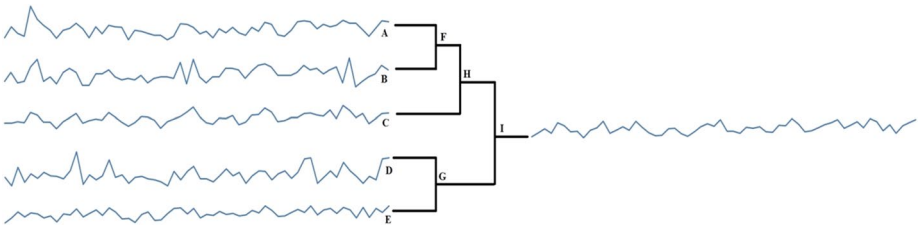

Fig. 2 Working of time series merging algorithm

Algorithm 1: Algorithm to find representative time series for every cluster

Algorithm: Merging Time Series
Input: n number of clusters
Output: A single representative time series for each cluster
Process:
Begin,
For i = 1 to n // i is cluster id
1. Calculate DTW distance between every time series in the cluster i
2. Merge the two closest time series objects (take the average of each data point)
3. Go to 1
4. Repeat until there is one time series object remains
5. Return a single time series to represent the cluster
End

Trend analysis

Least square regression technique [23] is used to fit a trend line on the RTS for each cluster. Least square simply states that it looks for an optimal solution for the overall fit of data such that sum of the squares error (SSE) is least.

Regression line which is also called trend line describes the linear relationship between x and y data points (in our case road accident count is y and month is x). Regression line can be defined as

$$\hat{y} = a + bx \tag{6}$$

where, $\hat{y}$ is the value to be predicted, a is the intercept and b is the slope.

The trend line is used to predict the changes in y with the changes in x. Another factor to validate the accuracy of this trend line is coefficient of determinations (CoD) indicated by r^2 or R^2. CoD is a number between 0 and 1 that determines how well data fit a statistical model. It can be calculated as:

$$r^2 = 1 - \frac{SSE}{SST} \tag{7}$$

where, $SSE = \sum(y_i - \hat{y}_i)^2$ and total sum of squares $(SST) = (y_i - \bar{y}_i)^2$.

Results and discussion

Cluster analysis

Initially, we performed cluster analysis using AGNES algorithm on normalized time series of 13 districts of Uttarakhand and 26 districts of Gujrat state. Cluster analysis provides the homogeneous groups from the data, in which each group contains data with

similar behavior. The cluster analysis was performed on monthly road accidents counts of Gujrat and Uttarakhand state. After clustering on monthly counts of road accidents, we obtained 3 clusters of Gujrat state in which we found 1 cluster with 17 districts and other 2 clusters with 1 and 8 districts only, 4 clusters of Uttarakhand state in which first cluster contains 1 district only and another cluster contains 3 districts and other two clusters contains 3 and 6 districts. Cluster wise distribution of districts is given in Table 2.

Trend analysis

After clustering of states, we formed representative time series for each of the clusters, which collectively represent all the time series in that cluster. In order to validate the representative time series, we compared the trend of each time series in the cluster with the representative time series of that cluster, which is further compared with the merged time series obtained from simple average method. This comparison is given in Table 3 which illustrates that representative time series obtained from our proposed merging algorithms preserves the trend and performs better than simple average method. Further, we performed trend analysis on each representative time series.

We tried to identify the difference between the clusters formed by AGNES algorithm. One of the differences that were found for Gujrat state is that all the districts with industrial areas have similar accident trends and they all are found in same clusters. Other difference is that districts which have more number of villages are in same clusters, and also they have different accident trends. A district "The Dangs" from Gujrat state which is also the smallest district is the only district in its cluster. Similarly, "Haridwar" district which is a famous pilgrimage place in India is also an only cluster in its cluster. Other clusters of Uttarakhand districts can be differentiate on the basis of tourist's locations, Hilly districts and Industrial districts.

A trend line describes the tendency of some events such as road accidents in our case. A trend line using least square regression technique on representative time series obtained using our time series merging algorithm for every cluster of Gujrat and Uttarakhand state has been fitted. A trend line for Gujrat clusters is shown in Fig. 3a–c, while trend line for Uttarakhand clusters is shown in Fig. 4a–d. Figure 3a illustrates that districts of Gujrat state in C1 has slightly negative trend for road accidents, while districts

Table 2 Cluster wise distributions of districts

Gujrat_cluster	
Cluster1	Ahmedabad, Surat, Rajkot, Bansakantha, Junagadh, Panch Mahals, Vadodara, Sabar Kantha
Cluster2	Tapi, Narmada, Porbandar, Kachch, Surendranagar, Gandhi Nagar, Dahod, Amreli, Jamnagar, Anand, Bharuch, Bhavnagar, Kheda, Mahasena, Patan, Navsari, Valsad
Cluster3	The Dangs
Uttarakhand_cluster	
Cluster1	Dehradun, Udhamsinghnagar, Naintal,
Cluster2	Almora, Pauri, Pithoragarh
Cluster3	Tehri, Chamoli, Uttarkashi, Rudraprayag, Bageshwar, Champawat
Cluster4	Haridwar

Table 3 Trend comparison of RTS using proposed algorithm and average merging algorithm

	Gujrat clusters			Uttarakhand cluster			
T.S. Id	**C1**	**C2**	**C3**	**C1**	**C2**	**C3**	**C4**
1	N	P	P	P	P	P	N
2	P	P	–	P	N	N	–
3	N	P	–	P	P	N	–
4	N	P	–	–	–	P	–
5	P	P	–	–	–	P	–
6	N	N	–	–	–	P	–
7	P	P	–	–	–	–	–
8	N	P	–	–	–	–	–
9	–	P	–	–	–	–	–
10	–	P	–	–	–	–	–
11	–	P	–	–	–	–	–
12	–	P	–	–	–	–	–
13	–	P	–	–	–	–	–
14	–	N	–	–	–	–	–
15	–	N	–	–	–	–	–
16	–	N	–	–	–	–	–
17	–	P	–	–	–	–	–
Total P	3	13	1	3	2	4	0
Total N	5	4	0	0	1	2	1
PTMA	N	P	P	P	P	P	N
ATMA	P	N	P	P	P	N	N

PTMA proposed time series merging algorithm; *ATMA* average time series merging algorithm

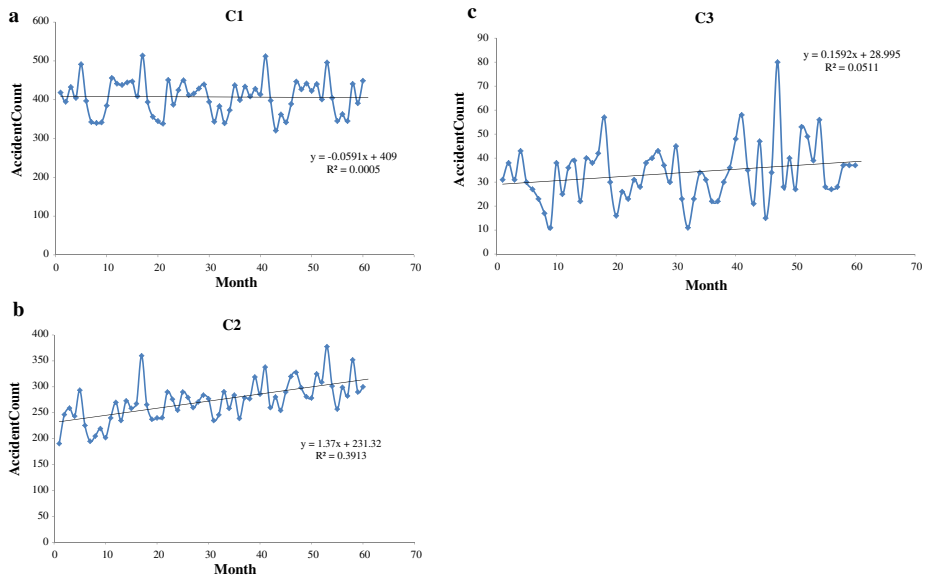

Fig. 3 **a** Monthly analysis of cluster 1 of Gujrat state. **b** Monthly analysis of cluster 2 of Gujrat state. **c** Monthly analysis of cluster 3 of Gujrat state

in C2 has a positive trend for road accidents. The third cluster of Gujrat state which contains only one district has slightly positive trend for road accidents.

Similarly, for Uttarakhand state cluster C1, there is a slight improvement in number of road accidents and trend line shows that this trend will sustain in future time also. The districts in C1 are the industrial districts, which indicates that people are coming to these locations to get the job and hence the population is growing year by year, which results in high traffic and increase in road accidents. Trend for C2 and C3 is almost similar. The districts in C2 and C3 are the hill districts. Most of the road accidents in these areas are vehicle fall from height accidents. It major difference in districts of C2 and C3 is the number of road accidents. C2 has slightly more number of road accidents than C3. The similarity found in road accident nature of C2 and C3 is that in rainy weather the accidents counts increases slightly for both clusters. Also, in summer time, most of the traffic moves to the hill stations. Districts in C2 have more famous hill stations than C3. This may be the reason for more accident counts in C2 than C3. The 4th cluster C4 consists of only one district Haridwar. Although the trend line in Fig. 4d shows slightly decreasing trend for C4 but it has the high peak in the beginning of time series (Jan 2010–Apr 2010). We have identified that reason for this sudden peak which is not available for later years was a famous "KUMBH festival" [24], which is repeated after few years. In this duration, the people from all over India visited Haridwar district for a holy bath in river Ganga [25]. One can assume that a huge traffic coming to the district which certainly results in large number of accidents which has illustrated in Fig. 4d. This festival will again be repeated in Jan 2016–Apr 2016. This means that number of road accidents will definitely be increased for this duration and government should take appropriate preventive measures to overcome the accident rate.

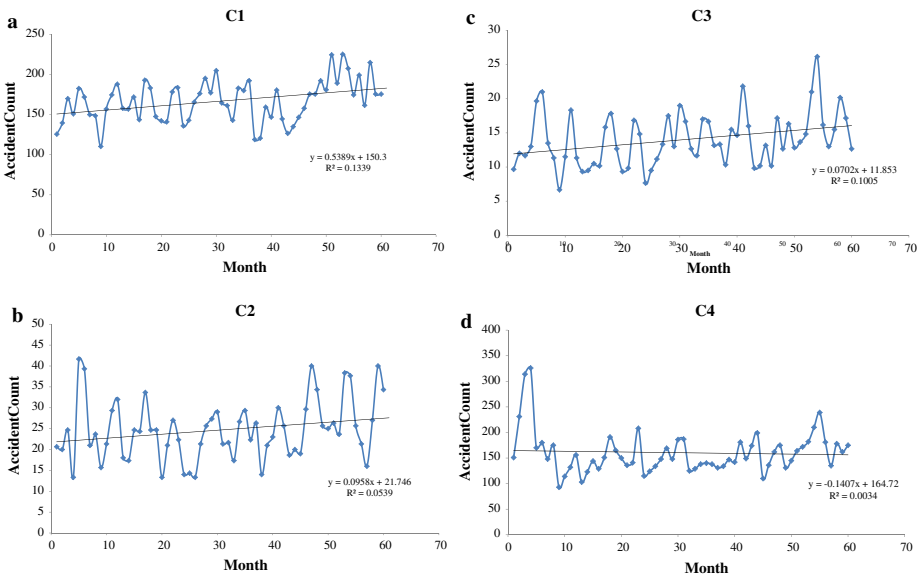

Fig. 4 **a** Monthly analysis of cluster 1 of Uttarakhand state. **b** Monthly analysis of cluster 2 of Uttarakhand state. **c** Monthly analysis of cluster 3 of Uttarakhand state. **d** Monthly analysis of cluster 4 of Uttarakhand state

Conclusion and future work

Road accidents are one of the prime factors for untimely death, partial or full disability and property damage, which is unacceptable in any form. Statistical techniques and data mining techniques both are were used in previous studies on road accident data analysis. One of the important factors in road safety analysis is to identify the certain regions where the trends of road accidents are occurring more than others. Time series analysis plays an important role in trend analysis and identifying whether the trend will increase in future also.

Our study focused on time series formation from the road accident monthly counts and then proposing a framework to analyze this time series data to know the trend of road accidents in different districts of Gujrat and Uttarakhand state of India. The framework normalizes the time series data of 39 districts of Gujrat and Uttarakhand states using z-score normalization. Further, average AGNES algorithm using DTW as a distance measure is applied to cluster the districts of Gujrat and Uttarakhand districts separately. This gives us different clusters for both the states in which districts with similar accident nature are clustered together in one group. As it is difficult and time consuming to analyze every time series of every cluster. A time series merging algorithm is also proposed to merge all the time series and form a representative time series for each cluster. Finally, this representative time series algorithm is analyzed using least square regression method. The trend line is plotted over the time series that fits the data using least square regression method. The trend for each cluster is further illustrates that in some cluster road accident trend is increasing across the years, while in some districts there is an increase in road accidents during some special events in those districts. Our future work will focus on developing novel approach using data mining techniques to analyze the different factors associated with road accidents in those districts where the road accident trend is increasing and providing some preventive measure to overcome the accidents.

Authors' contributions

DT contributed for the underlying idea, helped drafting the manuscript and played a pivotal role guiding and supervising throughout, from initial conception to the final submission of this manuscript. SK developed and implemented the idea, designed the experiments, analyzed the results and wrote the manuscript. All authors read and approved the final manuscript.

Author details

[1] Centre for Transportation Systems (CTRANS), Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India. [2] Computer Science and Engineering Department, Indian Institute of Technology Roorkee, Roorkee, Uttarakhand 247667, India.

Acknowledgements

The authors thankfully acknowledge the GVK-EMRI to provide data for our research.

Competing interests

The authors declare that they have no competing interests.

Received: 10 March 2016 Accepted: 17 May 2016
Published online: 26 May 2016

References

1. World Health Organization (WHO). Global status report on road safety 2013. Supporting a decade of action.
2. Depaire B, Wets G, Vanhoof K. Traffic accident segmentation by means of latent class clustering. Accid Anal Prev. 2008;40(4):1257–66.
3. Abellan J, Lopez G, Ona J. Analysis of traffic accident severity using decision rules via decision trees. Expert Syst Appl. 2013;40:6047–54.
4. Chang LY, Chen WC. Data mining of tree based models to analyze freeway accident frequency. J Saf Res. 2005;36(4):365–75.

5. Abdel-Aty MA, Radwan AE. Modeling traffic accident occurrence and involvement. Accid Anal Prev. 2000;32(5):633–42.
6. Bandyopadhyaya R, Mitra S. Modelling severity level in multi-vehicle collision on indian highways. Proced Soc Behav Sci. 2013;104:1011–9.
7. Chen W, Jovanis P. Method of identifying factors contributing to driver-injury severity in traffic crashes. Transp Res Rec. 2002;1717:1–9.
8. Joshua SC, Garber NJ. Estimating truck accident rate and involvements using linear and poisson regression models. Transp Plan Technol. 1990;15(1):41–58.
9. Poch M, Mannering F. Negative binomial analysis of intersection-accident frequencies. J Transp Eng. 1996;122(2):105–13.
10. Kumar S, Toshniwal D. A data mining framework to analyze road accident data. Journal of Big Data. 2015;2(1):1–18. doi:10.1186/s40537-015-0035-y.
11. Kashani T, Mohaymany AS, Rajbari A. A data mining approach to identify key factors of traffic injury severity. Promet Traffic Transp. 2011;23(1):11–7.
12. Oña JD, López G, Mujalli R, Calvo FJ. Analysis of traffic accidents on rural highways using latent class clustering and bayesian networks. Accid Anal Prev. 2013;51:1–10.
13. Kumar S, Toshniwal D. A data mining approach to characterize road accident locations. J Mod Transp. 2016;24(1):62–72. doi:10.1007/s40534-016-0095-5.
14. Geurts K, Wets G, Brijs T, Vanhoof K. Profiling of high frequency accident locations by use of association rules. Transportation Research Record. 2003. doi:10.3141/1840-14.
15. Tan PN, Steinbach M, Kumar V. Introduction to data mining. Boston: Pearson Addison-Wesley; 2006.
16. Han J, Kamber M. Data mining: concepts and techniques. USA: Morgan Kaufmann Publishers; 2001.
17. http://www.emri.in. Accessed 11 Jan 2016.
18. Kumar S, Toshniwal D. Analyzing Road accident data using association rule mining. International conference on computing, communication and security, ICCCS-2015. 2015. doi:10.1109/CCCS.2015.7374211.
19. Ratanamahatana CA, Jessica L, Dimitrios G, Keogh E, Michail V, Gautam D. Mining time series data. Data mining and knowledge discovery handbook. US: Springer; 2010. p. 1049–77.
20. Warren LT. Clustering of time series data—a survey. Pattern Recogn. 2005;38(11):1857–74.
21. Keogh E, Ratanamahatana CA. Exact indexing of dynamic time warping. Knowl Inf Syst. 2005;7(3):358–86.
22. Zhang X, Jun W, Xuecheng Y, Haiying O, Tingjie L. A novel pattern extraction method for time series classification. Optim Eng. 2009;10(2):253–71.
23. Clement E, Lamberton D, Protter P. An analysis of a least- squares regression algorithm for American option pricing. Financ Stoch. 2002;6:449–72.
24. https://en.wikipedia.org/wiki/Kumbh_Mela. Accessed 28 Jan 2016.
25. https://en.wikipedia.org/wiki/Ganges_in_Hinduism. Accessed 28 Jan 2016.