The Journal of Supercomputing (2021) 77:6505–6524
https://doi.org/10.1007/s11227-020-03580-9

Trend analysis using agglomerative hierarchical clustering approach for time series big data

Subbulakshmi Pasupathi[1] · Vimal Shanmuganathan[2] · Kaliappan Madasamy[3] · Harold Robinson Yesudhas[4] · Mucheol Kim[5]

Accepted: 14 December 2020 / Published online: 2 January 2021

Abstract
Road traffic accidents are a 'global tragedy' that generates unpredictable chunks of data having heterogeneity. To avoid this heterogeneous tragedy, we need to fraternize and categorize the datasets. This can be done with the help of clustering and association rule mining techniques. As the trend of accidents is increasing throughout the year, agglomerative hierarchical clustering approach is proposed for time series big data for trend analysis. This clustering approach segments the time sequence data into different clusters after normalizing the discrete time sequence data. Agglomerative hierarchical clustering takes the objects with similar properties and groups them together to form the group of clusters. The paradigmatic time sequence (PTS) data for each cluster with the help of dynamic time warping are identified that calculate the closest time sequence. The PTS analyzes various zone details and forms a cluster to report the data. This approach is more useful and optimal than the traditional statistical techniques.

Keywords Big data · Agglomerative hierarchical clustering · Paradigmatic time series · Trend analysis

1 Introduction

Road traffic accident (RTA) is an important factor to consider in research as it contains many factors; injuries lead to disability and other personal affections [1]. A report generated by World Health Organization shows that about 2 million accidents are happening throughout worldwide and it may lead to various segment of road safety and injuries [2]. The main focus on road accident is to ensure safety in road and emergency reporting [3]. The time sequence data analysis helps in encouraging the user to predict the emergency service, and accident rate can be predominantly

✉ Mucheol Kim
mucheol.kim@gmail.com

Extended author information available on the last page of the article

reduced in there forth [4]. Data mining technique has been applied to statistically survey the accident rate [5]. The challenges of big data are capture, storage, search, sharing, transfer, analysis, and visualization [6]. These data are in various formats such as structured, unstructured and semi-structured [7]. These data are not be handled by the traditional data processing tools [8]. Here, in this paper we take the huge amount of data generated in the road accidents for finding the factors causing them. This can be handled by the clustering and association rule mining techniques [9]. The trend analysis in PTS is more helpful for finding the factors of the accident since accident ratio increases every year. This can be efficiently identified with the time sequence data in every location. These data are needed to be normalized and then used for the processing with the help of the data mining techniques [10]. The huge amount of road accidents generates the economic cost and the injury to the human. Hence, the safety analysts have to discover the assessment of the traffic accident to illustrate the severity-related consequences. An accident-related analysis is completed from the traffic accident-related input to obtain the needful knowledge to find the solution according to the reports and data stored into the dataset with huge amount of data which are generated.

Time series perform a group of related data points gathered within the specified interval [11]. Monthly basis road accident is measured to store the time series data for revealing the future trend. This will help to identify the dissimilar regions of road accidents for providing the trend analysis [12]. This analysis contains 11 leading cities in India for using the dataset. It is very hard to analyze the time series data of the leading cities independently and also related to the nature and trend of the road accidents in the similar other cities in India [13]. The data mining mechanisms have been involved to construct a framework with cluster-related technology which is able to eliminate the heterogeneity data from the time series. These kinds of techniques have been used to expose the secret elements from the road accident-based information; the trend of the road accidents could be measured within the dissimilar cities with other factors. In this paper, the elaboration of road accident-related information is utilized to discover the trend behind the road accidents from the other cities [14].

The drivers may constantly contribute to minimizing the accidents by strictly following the rules and reduced efficiency gathered from the driving schools. Every driver must be trained to provide the traffic rules and also with the medical checkup effectively to follow the traffic rules. The proposed construction is about to analyze the time series data regarding the road accidents that use the clustering model of data mining technique. This construction of framework deals with the total amount of accidents with the usual statistical methods. The input for this proposed framework has been constructed for the road accidents for dissimilar time series and also the normalization of the particular time series. Additionally, it computes the agglomerative hierarchical clustering approach on the leading cities to discover the paradigmatic time sequence data for every cluster.

The main contribution of the paper is as follows:

- Data preprocessing technique is developed to remove the unwanted noises.
- Dynamic time wrapping discovers the correlation within the data objects for cluster formation.

- Agglomerative hierarchical clustering algorithm is developed the map to reduce framework for forming the time sequence data-based clustering.
- Trend analysis is utilized to predict the factors for finding the road accidents location.

2 Related work

The trend-related information has been utilized in several fields like social media, behavior pattern, the finance sector and the networking that the information is measured within a specific period of time through the series information. Time series data are gathered with the group of information which performs the specific time periods [15]. Several resources are involved to construct a framework that the categorization of the time series information the complexity is the major issue [16]. The time series data has been grouped into their framework that the machine learning concept provides the system with the complexity of learning without any particular programming [17]. The optimization is the main performance parameter that the gathered data and the obtained results [18]. The data gathering has been the major impact for generating the procedure to identify the knowledge from the data [19].

The major functionality for producing the structured data has been measured from the unstructured data that utilizes the human exploration and specific analysis [20]. Another important functionality is to exploit the semi-supervised algorithms to identify the analysis within the enhanced exploration technique [21]. The data analyst may obtain solution for the complex problems by automation in data analysis through the big data and also the data mining concept [22]. The primary dissimilarity is within the user role in modeling-based data exploration that the combined automation-based analysis techniques are utilized to identify the big data [23].

Time series data in the field of network contain the dynamic graph-based attributes to discover the dissimilar temporal patterns [24]. The classification is demonstrated as the mapping information into the pre-defined objects that task is computed through the multi-scale classifier which assigns every group of data into the time series to the particular category [25]. The unique features are classified from the dataset automatically to every time series with time-related pattern recognition system has several similarity measures to produce the data segmentation process [26]. It has been utilized as the input for the following steps that the parameter changes with labeling the particular data [27]. The visual analytics produces the solution for the user demands and the interaction with the data parameters. An efficient clustering procedure has the spatial inconsistency whether minimizing the dimension through the user interface interaction [28].

The agglomerative hierarchical clustering with distance function reduces the computational complexity which utilizes the clustering concept by segregating the dataset into a total amount of clusters until it finishes into a solitary attribute [29]. The optimal cluster member is constructed to maintain the clustering hierarchy; the combined cluster needs to generate the distance matrix and executing the clustering procedure several times despite combined distance function [30]. The single linkage can be appropriated, while unstructured clusters maintain the

outlier complete linkage process through the persistent merge concept [31]. The group of techniques is utilized to analyze the trend analysis in the big data that it does not depend on the functional component which needs no probabilistic knowledge which occurs within a particular dataset, this leads to the safety analyst to maintain the rules of understanding the events and discovers the attributes for preparing the accident-based analysis [32–37].

The existing approaches have been estimated through the total amount of injuries and damages to the properties for making the decision process to generate the detailed data about the accident management but failed to analyze effectively. Most of the existing techniques have been examined the severity of the accident and failed to generate the correlation within other parameters.

3 Proposed work

Data preprocessing in the proposed technique is the main task to generate the data in each data handling methodology that removes the unwanted noise to produce the analysis-based normalization. The time series data requires the improved data preprocessing despite generating the useful for the particular analysis. The dynamic time wrapping computes the correlation within the time sequence information. The similarity computation is utilized the hierarchical clustering technique through Euclidean distance between the similar points for generating the time series data. The hierarchical-based cluster analysis is used to identify the group of similar objects according to the attributes [38]. The data prediction is computed through the agglomerative clustering technique that the complexity has been measured. The PTS value is generated for every cluster through the proposed algorithm, the time sequencing through DTW distance metrics technique that analyzes the accident causing factors along with the distance measurement, and the trend analysis is utilized for predicting the location that causes the accident [39].

3.1 Data preprocessing

Data preprocessing is a prior task to analyze the data in every data handling techniques. The techniques of data preprocessing remove the unwanted noise or other constraints in the network. Here, these time sequence data were preprocessed and an analysis has been normalized [40]. The data transformations have been performed to implement data available for the time series-related analysis. The time series information needs the enhanced data preprocessing in spite of getting the data useful for the specific analysis. The time series information uses the normalization technique for assisting the difficulties in the preprocessing technique [41]. Hence, the analysis is used to get the efficient result for normalizing the time series information based on the time series.

3.1.1 Similarity measure for time sequence

Dynamic time warping (DTW) measures the correlation between two time sequences data objects even if their lengths are not same. DTW is used to minimize the metrics of two time sequence $r_i = \{r_1, r_2, \ldots, r_n\}$ and $f_i = \{f_1, f_2, \ldots, f_m\}$ which are of length x and y, respectively, and has to align r_i and t_j. The dynamic programming establishes a cluster approach using the infinity matrix, and the parameters are computed in Eqs. (1) and (2):

$$\text{mean}(f_i) = 0 \tag{1}$$

$$\text{SD}(f_i) = 1 \tag{2}$$

where $\text{mean}(f_i)$ is the mean value and $\text{SD}(f_i)$ is the standard deviation for producing the normalization time series. The normalization is computed in Eq. (3):

$$\text{Nor}_{f_i} = \sum_{i=1}^{n} \frac{\text{time}_i - \text{mean}(f_i)}{\text{SD}(f_i)} \tag{3}$$

3.1.2 Euclidean distance

The similarity measurement is used to compute the classification using hierarchical clustering methods with Euclidean distance concept, it is demonstrated as the distance within two similar points, it can also find the distance within the time series data with equal length and it is calculated using Eq. (4):

$$\text{Dis}(\alpha, \beta) = \sum_{i=1}^{n} \sqrt{(\alpha_i - \beta_i)^2} \tag{4}$$

where α, β are the time series data with the sequence of n values.

3.1.3 Dynamic time warping

The similarity measure within the two time series is computed using Dynamic Time Warping [42]. The time series is defined as $p_x = \{p_1, p_2, \ldots, p_n\}, q_y = \{q_1, q_2, \ldots, q_m\}$. The dynamic programming functionality is used to discover the similarity distance within time series in δ, the matrix has initialized using Eq. (5)

$$\delta[0,0] = 0, \delta[x,y] = \infty \tag{5}$$

The recursive function for computing the matrix value is in Eq. (6)

$$\delta[x,y] = \text{Dis}(p_x, q_y) + \min\{\delta[x, y-1], \delta[x-1, y], \delta[x-1, y-1]\} \tag{6}$$

3.2 Hierarchy-based cluster analysis

Cluster analysis is done using the similar objects to be grouped in a single forum or groups based on their attributes and properties. Lot of clustering algorithms such as agglomerative clustering is used in statistical data prediction. There may be a variety of algorithms for clustering a time sequence data. Agglomerative hierarchical clustering algorithm is designed to map reduce framework for clustering of time sequence data. The Space and time complexity using agglomerative hierarchical clustering is $P(n_3)$ and the other one is $P(2n)$. The proposed AHCTB is used to predict the road accident data. Table 1 demonstrates the steps for hierarchical agglomerative clustering algorithm.

3.3 Time sequence merging

The time sequencing has been done using the DTW distance metrics algorithm. This can be further used for the trend analysis that generates the factors causing accidents

Table 1 Hierarchical agglomerative clustering algorithm

Agglomerative Hierarchical Clustering Algorithm for mapper & reducer: (Mapper)
Input: Preprocessed Road Accident dataset
Output: PTS for each cluster
Initialization: Set line, month, district as Object Preprocessing: Tokenize dataset 1: Line=get line from the dataset 2: Month=get the Month value from the line 3: District=get the District value from the line Do until reach the Month & District 4: Set district = district value of line If Month Exists 5: Set Month as Key 6: Set month = month value of line End If End 7: Merge month and district value 8: Set month and district value into another variable 9: Map the variable to the reducer. (Reducer) 10: Set sum=0 For each get value from the key 11: Calculate Euclidean Mean for each district = PTS

along with the locations that are subjected to more accidents. AHCTB makes the best effort service in all in establishing the distance between them and is shown in Fig. 1.

The time sequence is analyzed using the time series data with the vector value in n-dimensional space. The standard time sequence is measured using Eq. (7)

$$\gamma_{ij} = \frac{\sum_{k=1}^{n} \text{time}_{ik}}{\sqrt{\sum_{k=1}^{n} \text{time}_{ik}^2}} \tag{7}$$

The time sequence within the vectors time_i and time_j is computed in Eq. (8)

$$\text{Dis}\left(\text{time}_i, \text{time}_j\right) = 1 - \sum_{k=1}^{n} \gamma_{ik} x \gamma_{jk} \tag{8}$$

3.4 Trend analysis

Trend analysis is used to predict the accidents causing factors in different locations. This can be analyzed using various algorithms that predict the accidents in each location. Here, we used least square regression technique for finding the crash rate in each cluster using least square regression in Fig. 2. The text-based non-numeric information with metrics like the distance is used to review the cluster analysis with the Euclidean distance. The linkage illustrates the distance within the group of observations with the pairwise distance within the observations.

The highest linkage clustering is computed in Eq. (9)

$$\max\left\{\text{Dis}(x, y) : x \in p_x, y \in q_y\right\} \tag{9}$$

The lowest linkage clustering is computed in Eq. (9)

$$\min\left\{\text{Dis}(x, y) : x \in p_x, y \in q_y\right\} \tag{10}$$

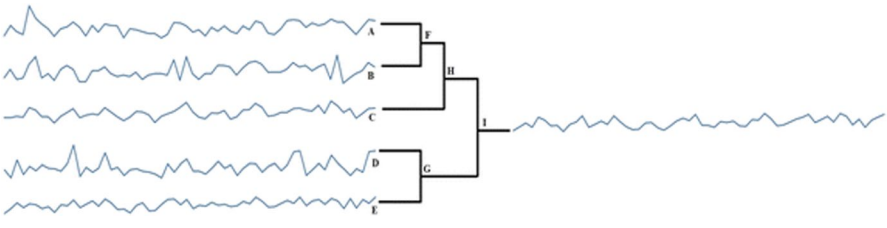

Fig. 1 Working of paradigmatic time sequence merging algorithm

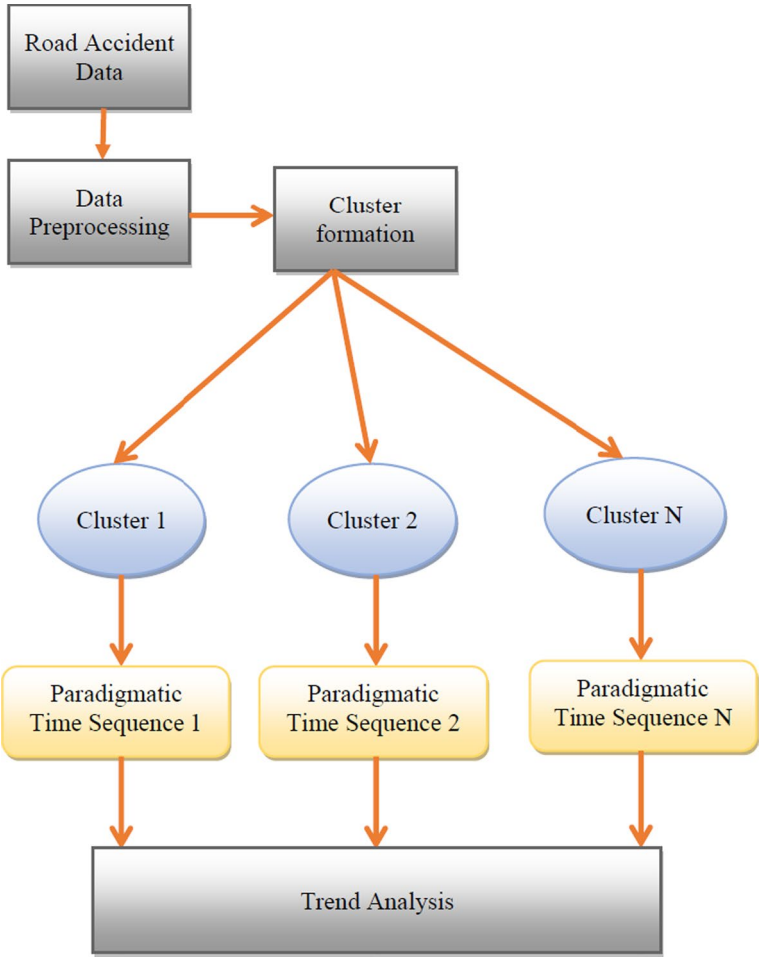

Fig. 2 Proposed system model

4 Performance evaluation

The performance evaluation is performed with MATLAB and HCE 3.5 software for trend analysis using agglomerative hierarchical clustering approach for time series big data.

4.1 Cluster analysis

Firstly, we considered the cluster formation of the heterogeneous dataset which is taken from the kdnuggets.com where the data are based on the accident rates in each district. Here, there are 11 districts—Ahmadabad, Bhavnagar, Gandhinagar,

Dehradun, Gurgaon, Haridwar, Porbandar, Rajkot, Vijayawada, Uttarkashi and Surat which are of heterogeneous in nature. The factors are some of the districts are industrial, metropolitan so in these locations the population will be high. Some other districts are hill stations and pilgrimage centers where the vehicles fall from great heights due to weather conditions such as fog and climate change. In rainy days, the accident rate increases slightly and in dry seasons traffic remains slow.

The hierarchical clustering model uses the cophenetic correlation coefficient using Euclidean distance within the xth and yth values, and it is computed in Eq. (11)

$$\mathrm{Cop}_{\mathrm{cor}} = \frac{\sum_{x,y} \left[p(x,y) - \mathrm{mean}_p\right]\left[q(x,y) - \mathrm{mean}_q\right]}{\sum_{x,y} \left[\left(p(x,y) - \mathrm{mean}_p\right)^2\right]\left[\left(q(x,y) - \mathrm{mean}_q\right)^2\right]} \tag{11}$$

Cophenetic correlation coefficient is measured for the parameters of median, centroid and weighted value, and also, single and complete values are illustrated in Fig. 3 and it denotes that the DTW has the highest amount of coefficient value and Euclidean distance has the lowest amount of coefficient value.

4.2 Trend analysis

Least square-based modeling with regression method has been utilized to produce the trend analysis for every cluster. Least square with producing the optimized solution for the summation of the weighted error has the least value.

4.2.1 Monthly prediction

The clustering of districts makes the time sequence data to be visualized to perform the normal conditions over the trend analysis. In Fig. 4, the accident rate is high because it is a metropolitan city which have high population rate, resulting in high traffic which tends to maximum accident occurrence. Here, in each

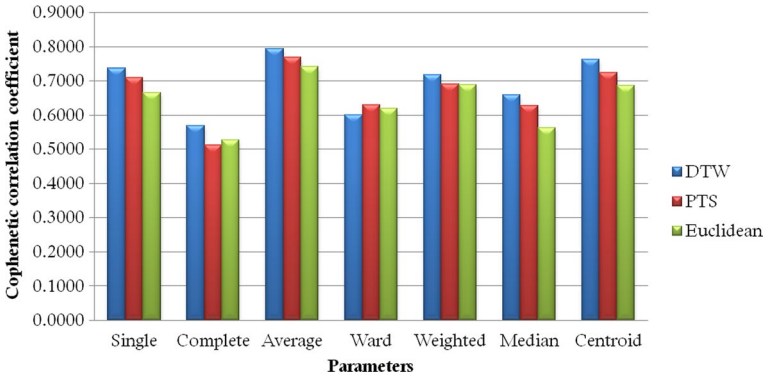

Fig. 3 Cophenetic correlation coefficient

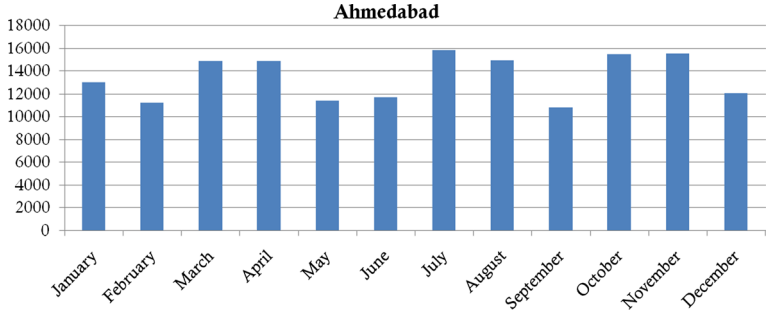

Fig. 4 Accident rates in Ahmedabad city

month the trend rate varies depends on the accident count. The high peak of accident rate is recorded in the months of July and November. The detailed prediction analysis of road accidents with some suitable measurement could be useful to reduce the frequent road accidents.

Figure 5 illustrates the accident rates in Bhavnagar city that the accident rate is high on the month of September, and other months have the moderate rates compared to the other cities.

In Fig. 6, the accident rate occurs at maximum level at most of the months and moderate scale value is reached in other months; the population ratio is high having dense traffic rate and records most sensitive peak values.

Figure 7 demonstrates the accident rates in Gandhinagar city, and the accident rate is high on the month of April compared to the other months.

Figure 8 illustrates the accident rates in Gurgaon city, and the rates are medium in every month.

Figure 9 illustrates the accident rates in Haridwar city, and it shows that the accident rates are low in January, April and July.

In Fig. 10, the accident occurs at certain months as it is a rural area, where there is no high usability of vehicles and less traffic. Here, the accident occur average in every month.

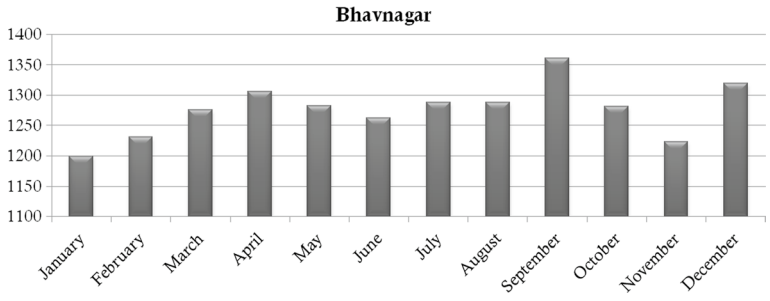

Fig. 5 Accident rates in Bhavnagar city

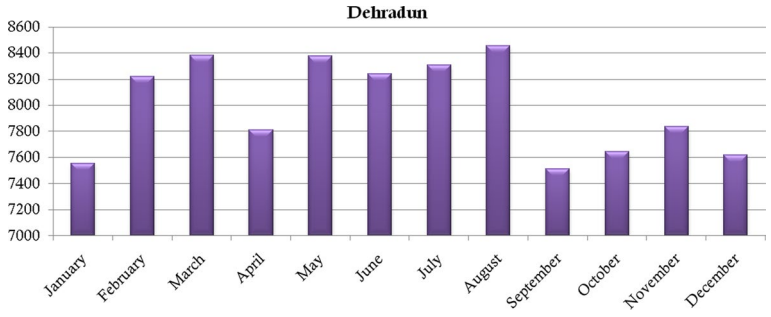

Fig. 6 Accident rates in Dehradun city

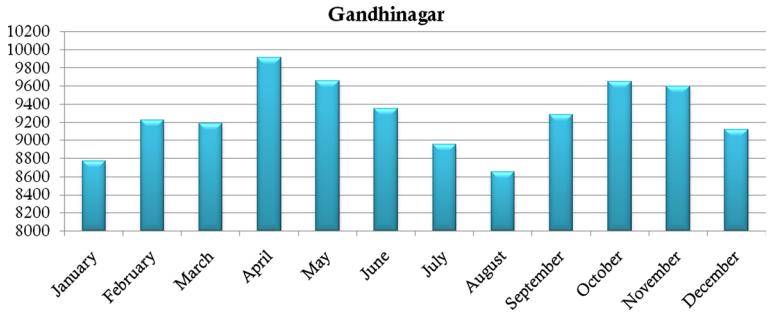

Fig. 7 Accident rates in Gandhinagar city

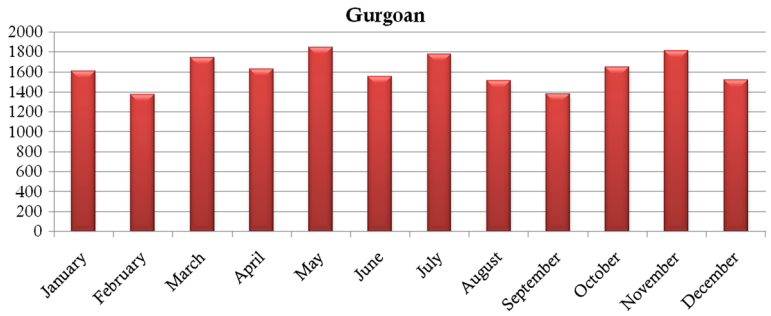

Fig. 8 Accident rates in Gurgaon city

Figure 11 demonstrates the accident rates in Rajkot city, and it shows the accident rate is high on the months of July, August and October because of high population and winter season.

In Fig. 12, there is a high accident rate due to the industrialization of the city. Here, a large number of factories, industries which lead to the transport of goods to and from the cities. The high accident occurs at the month of February and November where lot of accidents and death occurs.

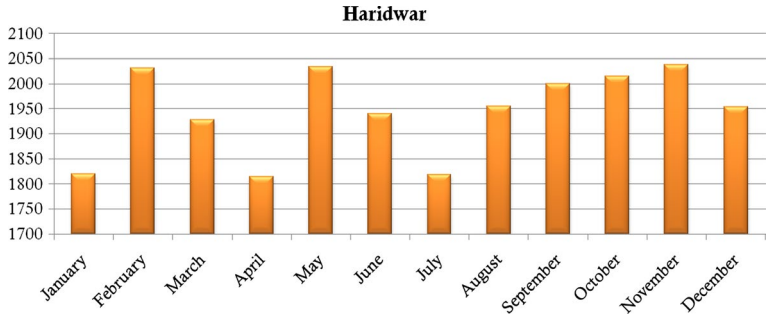

Fig. 9 Accident rates in Haridwar city

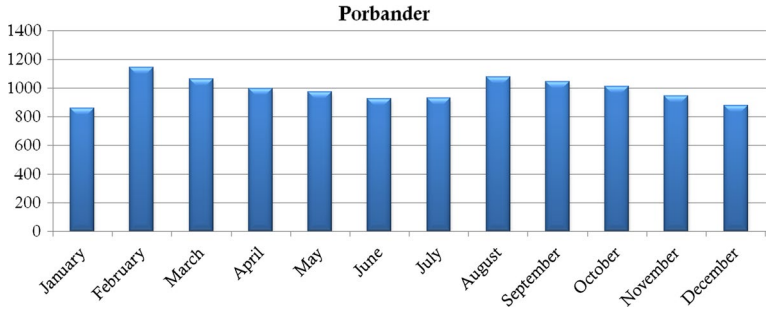

Fig. 10 Accident rates in Porbandar city

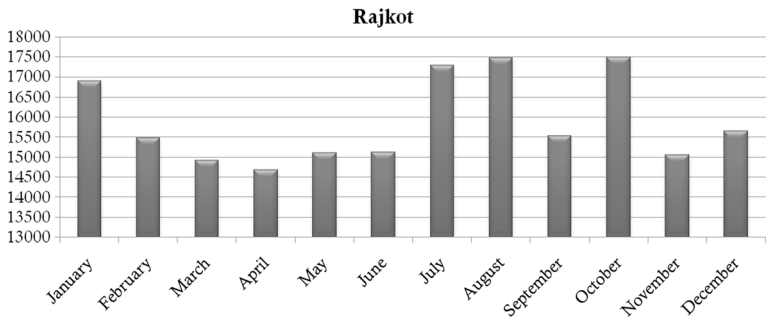

Fig. 11 Accident rates in Rajkot city

In Fig. 13, the accident rates are very low because it is a rural area where the population is not so high as in the urban areas and the traffic rate is low. Here, the accident occurs high at the month of July only. Those accidents may be caused due to the road conditions, climatic changes and also due to the large number of vehicles.

Figure 14 demonstrates the accident rates for Vijayawada city, and it shows that the accident rate is moderate in every month. The population which has a moderate level of accident rate having a slight dense traffic ratio that happens all over the months.

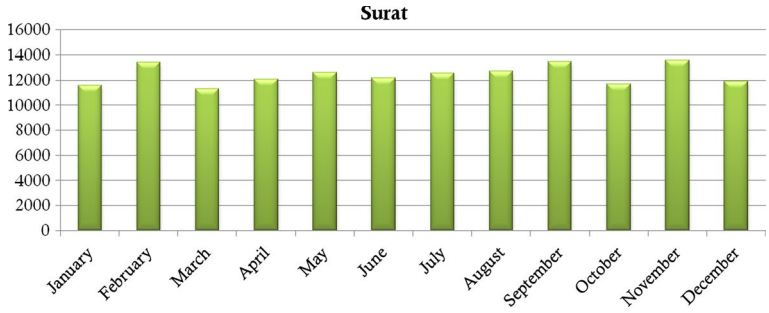

Fig. 12 Accident rates in Surat city

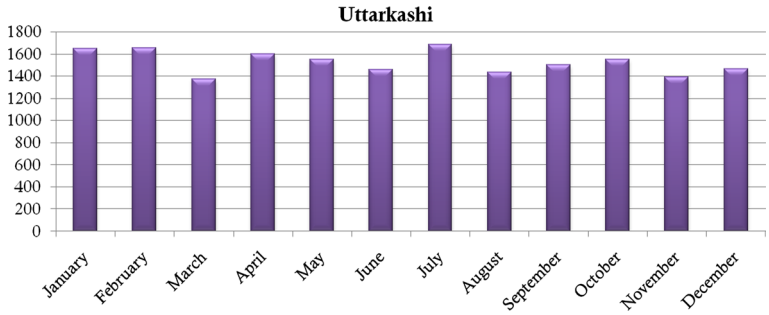

Fig. 13 Accident rates in Uttarkashi city

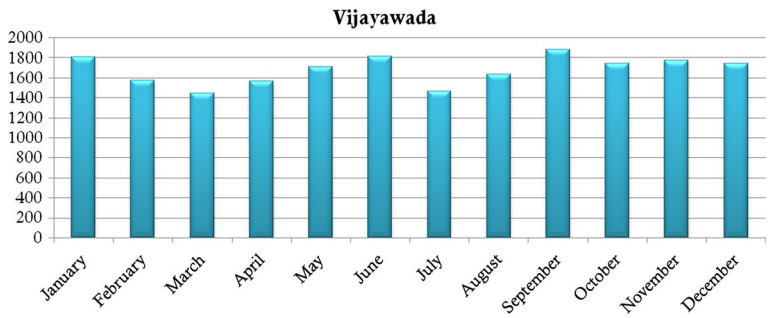

Fig. 14 Accident rates in Vijayawada city

Table 2 illustrates the statistical analysis for the various cities that Porbandar has the minimum value and Rajkot has the maximum value.

Figure 15 demonstrates the accident rates for the one single year for 11 important cities analysis, and it illustrates that the cities of Ahmedabad, Surat, and Rajkot are having the highest amount of accident rates. The cities of Bhavnagar, Gurgaon, Haridwar, Porbandar, Uttarkashi and Vijayawada are having the lowest amount of accident rates. The cities of Dehradun and Gandhinagar are having the medium amount of accident rates.

Table 2 Statistical analysis for the various cities

Districts	Total	Average	Maximum	Minimum
Ahmedabad	154,743	12,895	15,430	10,236
Bhavnagar	15,040	1253	1346	1119
Dehradun	96,371	8031	8495	7443
Gandhinagar	112,773	9398	9957	8791
Gurgaon	18,713	1559	1809	1395
Haridwar	23,228	1936	2136	1811
Porbandar	12,600	1050	1188	932
Rajkot	201,448	16,787	18128	15,092
Surat	154,733	12,894	13,853	11,654
Uttarkashi	18,805	1567	1792	1373
Vijayawada	19,514	1626	1845	1431

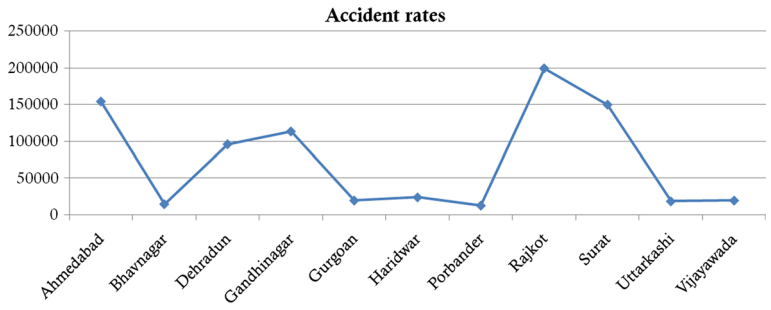

Fig. 15 Accident rates for Indian cities

4.2.2 Weekly analysis

Figure 16 demonstrates the weekly analysis for the road accident rates for the 11 Indian cities in detailed manner.

4.2.3 Analysis for the factors of Accident

Figure 17 illustrates the place of accident analysis, and the percentage is high on state highway and other places are national highway, city road, village road and approach road.

Figure 18 demonstrates type of road user analysis, and the analysis shows that the pedestrian has the highest amount of percentage.

Figure 19 demonstrates the type of accident percentage, and the result shows that the hit and run has the highest amount of accident percentage.

Figure 20 demonstrates the accident rate percentage that winter seasons have higher amount than summer seasons and raining seasons.

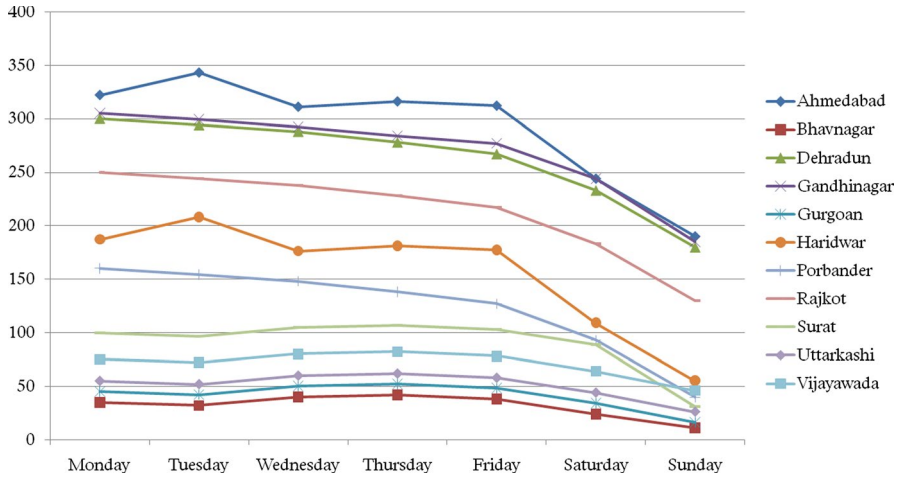

Fig. 16 Weekly analysis

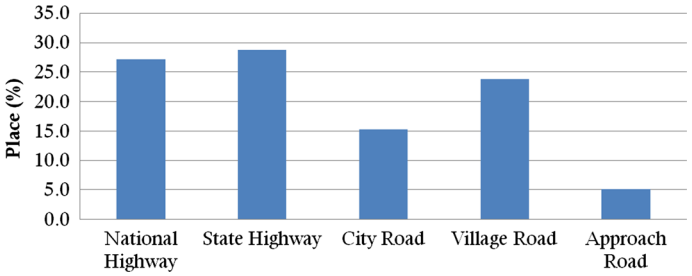

Fig. 17 Place of accident

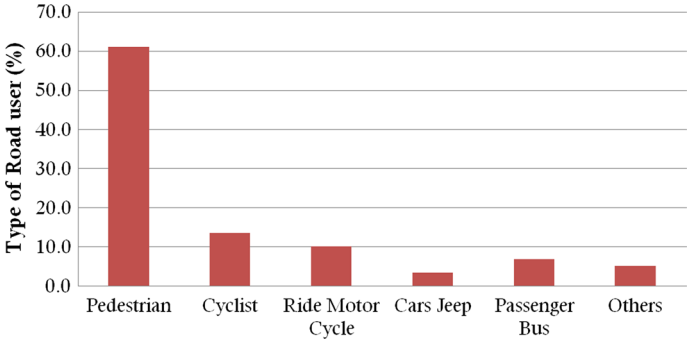

Fig. 18 Type of road user

Figure 21 analyzes the time of accident percentage for the cities, and it is measured that at the time of 12–2 pm the accident rates are high. At the time period of 2–4 pm the accident rate is around 19 percentages. When the time is

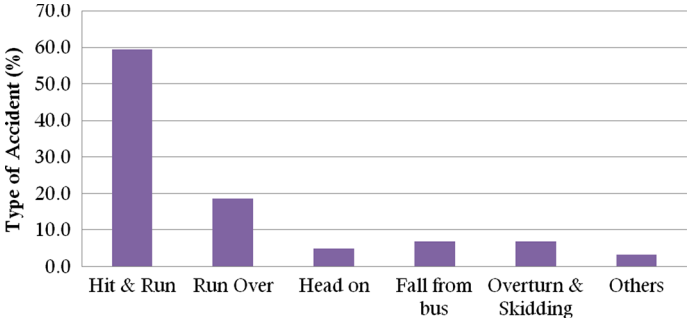

Fig. 19 Type of accident

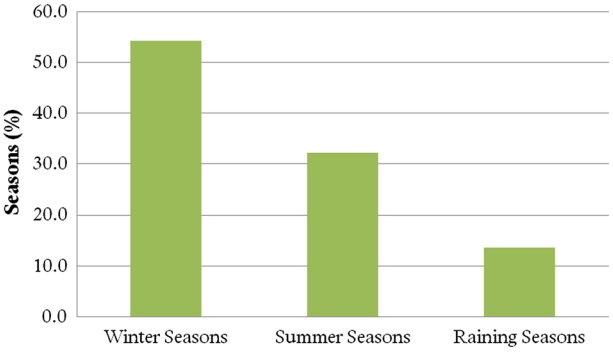

Fig. 20 Seasons

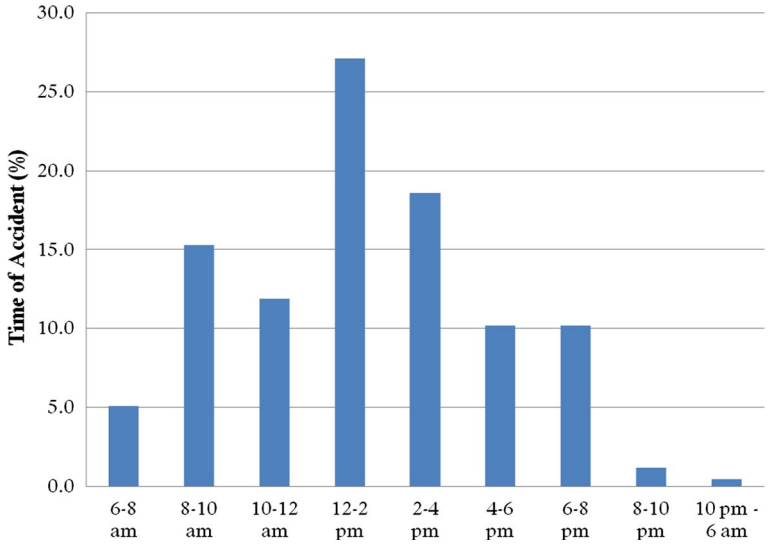

Fig. 21 Time of accident

between 8 and 10 am, the accident rate is 15% and at all the other timings the accident rate percentage is 10 or less.

Figure 22 demonstrates the responsible vehicles for causing the accident, and it shows that the trucks and buses are having the highest amount of 40% responsible vehicles, cars and jeeps involve 30% accidents, tractor involves 12%, two wheelers involve 10% and others involve 9% of accidents.

The state highways have the highest accident percentage of 28.7%, hit and run is the highest type of accident 59.3%, winter seasons have the highest accident of 54.2%, 12–2 pm is the time for most accidents of 27.1% happened. Trucks and buses are involved in 40% of accidents. The performance evaluation shows the trend flow using the paradigmatic time sequence from each cluster by calibrating the mean Euclidean value in a monthly basis. The trend rate represents that the monthly ratio of the metropolitan and urban areas have a high peak value. Here, the high death rate are also at the months where the location having a high traffic rate.

The clustering is the mathematical formation which discovers the particular patterns into the dataset which is inside every cluster discovers the similarity degree. It could be achieved through several algorithms that differ particularly in the notion that generates a cluster to effectively discover them. The cluster analysis is an iterative process to obtain the knowledge discovery through the optimization technique that is used to develop the preprocessing and generates the possible results. The greedy clustering-based hierarchal algorithm as the complexity of $O(n^3)$ that makes of the huge amount of datasets. The exhaustive search-based complexity is $O(n^2)$ and the special case of optimum-based effective technique has the complexity of $O(n^2)$. The prediction-based problems could be solved using the effective greedy approach that utilized the variable to frame the same order with effective size, and the prediction is developed by reducing the huge estimated failures. The performance results proved that the greedy approach has solved several problems with min–max procedures.

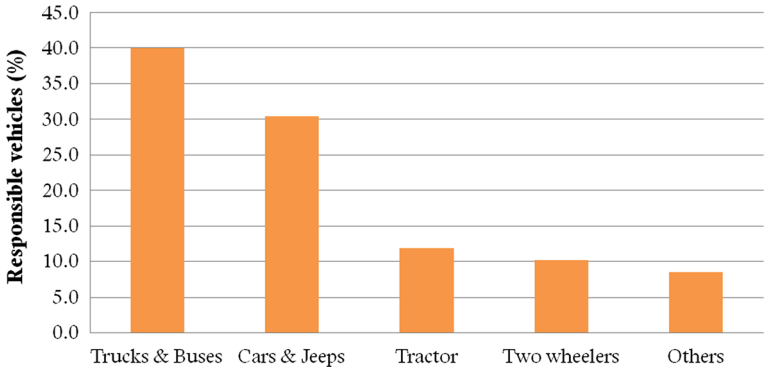

Fig. 22 Responsible vehicles

5 Conclusion

The progress of the system includes preprocessing of heterogeneous data which are classified and clustered based on the district-wise grouping of the road accident rates. Consequently, in each cluster attributes are merged based on the monthly analysis and paradigmatic time sequence is predicted which are then fed for the trend analysis of the accident rate. Here, AHCTB is proposed for the clustering of each district and trend analysis is done to each cluster using PTS. The trend analysis shows the variation of the accident rates in each cluster. It also shows the accidents are prone to high in metropolitan and industrial cities where the population and transportation are very high in nature, whereas in rural areas, the accident rate is low. In future, the proactive approach needs to be constructed by designing the best road safety plans with action plans; the automation system with sensors and cameras should be gathered the driving data of huge crashes. The safety measurement system needs to be developed and increased the technology changes like automated vehicle prediction with smooth traffic system.

Acknowledgements This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1F1A1076976).

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Abellan J, Lopez G, Ona J (2013) Analysis of traffic accident severity using decision rules via decision trees. Expert System Appl 40:6047–6054
2. Kumar S, Toshniwal D (2015) Analyzing road accident data using association rule mining. In: International Conference on Computing, Communication and Security", ICCCS-2015, vol. 20, pp. 30–40
3. Kumar S, Toshniwal D (2016) A novel framework to analyze road accident time series data. J. Big Data 30:5004–5020
4. Wang L, Lu H-P, Zheng Y, Qian Z (2014) Safety analysis for expressway based on Bayesian network: a case study in China. IEEE Commun Mag
5. de Ona J, Mujalli RO, Calvo FJ (2011) Analysis of traffic accident injury severity on Spanish rural highway using Bayesian networks. ScienceDirect Accid Anal Prevent 43:402–411
6. Pakgohar A, Tabrizi RS, Khalili M, Esmaeili A (2011) The role of human factor in incidence and severity of road crashes based on the CART and LR regression: a data mining approach. Procedia Comput Sci 3:764–769. https://doi.org/10.1016/j.procs.2010.12.126
7. Kumar S, Toshniwal D (2016) A data mining approach to characterize road accident locations. J Modern Transp 24(1):62–72
8. Lv Y, Duan Y, Kang W, Li Z, Wang F-Y (2015) Fellow", traffic flow prediction with big data: a deep learning approach. IEEE Trans Intell Transp Syst 16:2
9. de Oña J, López G, Mujalli R, Calvo FJ (2013) Analysis of traffic accidents on rural highways using latent class clustering and Bayesian networks. ScienceDirect Accid Anal Prev 51:1–10
10. Regine S, Simon C, Maurice A (2015) Processing traffic and road accident data in two case studied of road operation assessment. ScienceDirect 6:90–100

11. Kumar S, Toshniwal D (2015) A data mining framework to analyze road accident data. J Big Data 2:26
12. Kee D, Jun GT, Waterson P, Haslam R (2016) A systemic analysis of South Korea Sewol ferry accident—striking a balance between learning and accountability. Appl Ergon 1:1–14
13. Ramos L, Silva L, Santos MY, Pires JM (2015) Detection of road accidents accumulation zones with a visual analytics approach. ScienceDirect 64:969–976
14. Xi J, Zhao Z, Li W, Wang Q (2016) A traffic accident causation analysis method based on AHP-Apriori. ScienceDirect Procedia Eng 137:680–687
15. Fu T-C (2011) A review on time series data mining. Eng Appl Artif Intell 24(1):164181
16. Shanmuganathan V, Yesudhas HR, Khan MS et al (2020) R-CNN and wavelet feature extraction for hand gesture recognition with EMG signals. Neural Comput Appl 32:16723–16736
17. Ilango SS, Vimal S, Kaliappan M et al (2018) Optimization using artificial bee colony based clustering approach for big data. Cluster Comput. https://doi.org/10.1007/s10586-017-1571-3
18. Vo V, Luo J, Vo B (2016) Time series trend analysis based on k-means and support vector machine. Comput Inform 35(1):111127
19. Heaton J (2018) Ian Goodfellow, Yoshua Bengio, and Aaron Courville: deep learning. Genet Program Evolvable Mach 19:305–307 (2018). https://doi.org/10.1007/s10710-017-9314-z
20. Fujiwara T, Li JK, Mubarak M, Ross C, Carothers CD, Ross RB, Ma K-L (2018) A visual analytics system for optimizing the performance of large-scale networks in supercomputing systems. Vis Inform 2(1):98–110. https://doi.org/10.1016/j.visinf.2018.04.010
21. Ramamurthy M, Robinson YH, Vimal S, Suresh A (2020) Auto encoder based dimensionality reduction and classification using convolutional neural networks for hyperspectral images. Microprocess Microsyst 79
22. Keim DA, Munzner T, Rossi F, Verleysen M (2015) Bridging information visualization with machine learning (Dagstuhl seminar 15101). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Wadern, Germany, Dagstuhl Rep 3 5(3)
23. Keim DA, Rossi F, Seidl T, Verleysen M, Wrobel S (2012) Information visualization, visual data mining and machine learning (Dagstuhl seminar 12081). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Wadern, Germany, Dagstuhl Rep 2 2(2)
24. Hadlak S, Schumann H, Cap CH, Wollenberg T (2013) Supporting the visual analysis of dynamic networks by clustering associated temporal attributes. IEEE Trans Vis Comput Graphics 19(12):2267–2276. https://doi.org/10.1109/TVCG.2013.198
25. Xing Z, Pei J, Keogh E (2010) Abrief survey on sequence classication. ACM SIGKDD Explor Newslett 12(1):4048
26. Kalamaras I et al (2018) An interactive visual analytics platform for smart intelligent transportation systems management. IEEE Trans Intell Transp Syst 19(2):487–496. https://doi.org/10.1109/TITS.2017.2727143
27. Steiger M, Bernard J, Mittelstädt S, Lücke-Tieke H, Keim D, May T, Kohlhammer J (2014) Visual analysis of time-series similarities for anomaly detection in sensor networks. Comput Graph Forum 33(3):401410
28. Gopikumar S, Raja S, Robinson YH, Shanmuganathan V, Chang H, Rho S (2020) A method of landfill leachate management using internet of things for sustainable smart city development. Sustain Cities Soc. https://doi.org/10.1016/j.scs.2020.102521
29. Ramamurthy M, Krishnamurthi I, Vimal S, Robinson YH (2020) Deep learning based genome analysis and NGS-RNA LL identification with a novel hybrid model. Biosystems 197
30. Stopar L, Skraba P, Grobelnik M, Mladenic D (2018) Streamstory: exploring multivariate time series on multiple scales. IEEE Trans Vis Comput Graphics 25(4):17881802
31. Sacha D, Kraus M, Bernard J, Behrisch M, Schreck T, Asano Y, Keim DA (2018) SOMFlow: guided exploratory cluster analysis with selforganizing maps and analytic provenance. IEEE Trans Vis Comput Graphics 24(1):120130
32. Xie X, Cai X, Zhou J, Cao N, Wu Y (2019) Asemantic-based method for visualizing large image collections. IEEE Trans Vis Comput Graphics 25(7):23622377
33. Silva PB, Andrade M, Ferreira S (2020) Machine learning applied to road safety modeling: a systematic literature review. J Traffic Transp Eng (Engl Ed) 7(6):2095–7564. https://doi.org/10.1016/j.jtte.2020.07.004
34. Ali M, Jones MW, Xie X, Williams M (2019) TimeCluster: dimension reduction applied to temporal data for visual analytics. Vis Comput 35(6):10131026

35. Liu M, Shi J, Cao K, Zhu J, Liu S (2018) Analyzing the training processes of deep generative models. IEEE Trans Vis Comput Graphics 24(1):7787
36. Senaratne H, Mueller M, Behrisch M, Lalanne F, Bustos-Jiménez J, Schneidewind J, Keim D, Schreck T (2018) Urban mobility analysis with mobile network data: a visual analytics approach. IEEE Trans Intell Transp Syst 19(5):15371546
37. Chen Y, Xu P, Ren L (2018) Sequence synopsis: optimize visual summary of temporal event data. IEEE Trans Vis Comput Graphics 24(1):4555
38. Annamalai S, Udendhran R, Vimal S (2019) An intelligent grid network based on cloud computing infrastructures. Novel Pract Trends Grid Cloud Comput. https://doi.org/10.4018/978-1-5225-9023-1.ch005
39. Annamalai S, Udendhran R, Vimal S (2019) Cloud-based predictive maintenance and machine monitoring for intelligent manufacturing for automobile industry. Novel Pract Trends Grid Cloud Comput. https://doi.org/10.4018/978-1-5225-9023-1.ch006
40. Kumar S, Toshniwal D (2016) Analysis of hourly road accident countsusing hierarchical clustering and Cophenetic correlation coefficient (CPCC). J Big Data 3(13):20–56
41. Vimal S, Suresh A, Subbulakshmi P, Pradeepa S, Kaliappan M (2020) Edge computing-based intrusion detection system for smart cities development using IoT in urban areas. In: Kanagachidambaresan G, Maheswar R, Manikandan V, Ramakrishnan K (eds) Internet of things in smart technologies for sustainable urban development. EAI/Springer innovations in communication and computing. Springer, Cham
42. Vimal S et al (2020) Deep learning-based decision-making with WoT for smart city development. Smart innovation of web of things. CRC Press, Boca Raton, pp 51–62

Affiliations

Subbulakshmi Pasupathi[1] · Vimal Shanmuganathan[2] · Kaliappan Madasamy[3] · Harold Robinson Yesudhas[4] · Mucheol Kim[5]

Subbulakshmi Pasupathi
subbu.psk@gmail.com

Vimal Shanmuganathan
svimalphd@gmail.com

Kaliappan Madasamy
kalsrajan@yahoo.co.in

Harold Robinson Yesudhas
yhrobinphd@gmail.com

[1] School of Computing, Scope, VIT University, Chennai Campus, Tamilnadu, India

[2] Department of IT, National Engineering College, Kovilpatti, Tamil Nadu, India

[3] Department of CSE, Ramco Institute of Technology, Rajapalayam, Tamilnadu, India

[4] School of Information Technology and Engineering, Vellore Institute of Technology, Vellore, India

[5] School of Computer Science and Engineering, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, South Korea