

TEMPORAL CLASSIFICATION OF FLOW-PATTERNS IN MID-SIZED URBAN ROAD NETWORK

SaiKiran Annam and BhargabMaitra
Indian Institute of Technology Kharagpur, Kharagpur
bhargab@civil.iitkgp.ernet.in

DebasisBasu
Indian Institute of Technology Bhubaneswar
dbasu@iitbbs.ac.in

ABSTRACT

Traffic flow on an urban road network is temporal in nature. Knowledge of within-day traffic flow variation is important for evaluating alternative measures of traffic flow improvement on urban transportation network. Traditionally, a within-day traffic flow variation is classified as peak and off-peak periods, and is typically interpreted from graphical representation of the traffic flow patterns. But, in case of an urban form, where land-use development of various opportunities is quiet spatially varied, the peak and off-peak periods of traffic flow are not found well-defined in variation profile. In such case, a more scientific approach is required to identify these time periods. In the present work, an attempt has been taken to classify the within-day variation in traffic flow by considering inherent temporally continuous nature of traffic flow data. The classification has been carried out using K-means clustering technique. The work has been demonstrated by considering the multi-modal urban transportation network of Bhubaneswar city, India.

KEYWORDS: Simple moving average method, k-means clustering, traffic flow pattern

1.0 INTRODUCTION

Traffic flow in an urban transportation network is typically a time-series measurement of traffic flow volumes at a fixed spatial point. In the past decades, substantial research works (1),(2) were carried out on modelling of traffic flow prediction based on time series traffic data. Traffic flow often represents a strong periodicity or seasonality (3) such as per day or per week, or even per hour variation (4). The benefits of intelligent transportation system (ITS) cannot even be fully realized without understanding of within-day temporal traffic flow variation. Traditionally, a within-day traffic flow variation is classified as peak and off-peak period, and is typically interpreted from graphical representation of the traffic flow patterns (5). The above approach is deemed simple, but the patterns of traffic flows vary from one study to another and complexity of the graph. Even then, many researchers (6),(7) still follow the traditional method and interpret the peak hour from the traffic flow profile. Therefore, the understanding of temporal variation of traffic flow data often becomes a pre-requisite for planning and management of traffic on urban transportation network (8),(9). Furthermore, knowledge on temporal variations of traffic flow data is having paramount importance for design of traffic control, decision support and traveller information system (10) around the world. In this regard, it must be borne in mind that the change in flow pattern at any location arises due to the varied distribution of activities over time and space of an urban area. In case of an urban form, where land-use development of various opportunities is quiet spatially varied, the peak and off-peak periods of traffic flow are not found well-defined in variation profile. In such case, it is difficult to identify peak and off-peak periods from temporal variation of the traffic flow data. In such situation, it is required to apply various time series modelling approaches such as autoregressive integrated moving average (ARIMA), nonparametric regression, neural networks, and heuristics (11), (12) in order to classify, and then identify peak/off-peak period of traffic flows. The above approach can

primarily be categorized into two groups- a heuristic search and a cluster analysis. Among heuristic search algorithms, various techniques such as genetic algorithm (GA) (13) and artificial immune systems (AIS)-based data analysis algorithm (14) were proposed in literature. The application of this search algorithm was done only for hypothetical arterial networks; and often contained too many adjustable parameters in the algorithm for which they were found difficult to employ (15). The other alternative approach was on classifying temporal variations of traffic flow using clustering techniques (16), (17). In this regard, application of non-hierarchical cluster analysis (15) called K-means clustering have been found popular among many researchers (18), (19), (20). The K-means clustering is simple in its application, which requires reduced data storage (21). However, all these studies using K-means clustering did not explicitly incorporate the time of occurrence of a particular traffic volume value as a dimension in the clustering analysis. Ignoring time factor as a dimension in the cluster analysis often leads to creating cluster of traffic data points of non-consecutive time periods. In order to avoid such problem, the K-means clustering analysis helps to consider the continuous time factor, while clustering the traffic data points.

In view of the above, the current study takes an attempt to classify within-day traffic flow pattern by considering inherent temporally continuous nature of traffic flow data. The classification has been carried out using K-means clustering technique. The overall study follows a two part process. In the first part, the short-term variation in flow pattern is smoothed using a data-smoothing technique such as simple moving average (SMA), and then in the second part, the K-means clustering technique is employed to identify the peak and off-peak time periods of the traffic flow of an urban area. The work has been demonstrated by considering the multi-modal urban transportation network of Bhubaneswar city, India.

2.0 STUDY APPROACH

As mentioned, the objective of this study is demonstrated by classifying the traffic flow data of the Bhubaneswar city using K-means clustering technique. In India, the traffic is heterogeneous in nature and to obtain the traffic in homogeneous units it has to be adjusted using suitable Passenger Car Units (PCU) factors (22). Traffic flow patterns are often found with irregular short-term variation time series data, which pose a difficulty while creating clusters. Thus, it becomes essential to reduce the effect of such variations or noise (3). In order to achieve this, the traffic flow data has been smoothed using a data smoothing technique. Data smoothing technique is widely used for understanding underlying trend, cyclic and seasonal components in a time series data (23). Simple moving average (SMA) technique is one such smoothing technique. In SMA, a set of successive data points (say, l number of data points) are considered on either side of a data point and un-weighted mean value of the data set is taken as the new value of the data point (24). In this case ' l ' is called as the power of SMA. This process is uniformly followed for the entire data set, which results in a new smoothed curve with lesser number of short-term variations (25).

$$X'_i = \frac{X_{i-l} + X_{i-l+1} + \dots + X_i + \dots + X_{i+l-1} + X_{i+l}}{l} \quad \forall i=1,2,3,\dots; l=1,2,3,\dots \quad (1)$$

In the above equation, X'_i is the smoothing average value of a set of l data points taken equally from either side of the original data point X_i .

Once the traffic flow data is smoothed using the SMA method, the flow values have been normalised on a 0-1 scale. This helps bring all the flow volumes of different links in an uniform scale keeping the trend of flow in each link intact. This removes the biasness of the flow value towards higher values. The mean of the normalised flow values of all the arterials are estimated in order to achieve the representative flow pattern of the whole city. After this, clustering technique is employed to classify the traffic flow pattern. Clustering is a technique that attempts to separate a large group of objects into a multiple small groups (also called as clusters) in a manner that the objects belonging to one group are more similar to each other when compared to objects belonging to different groups (26). In this study, a non-hierarchical clustering

technique called K-means clustering is employed to identify the peak and off-peak time periods of the traffic flow in Bhubaneswar city.

The K-means procedure follows a simple recursive algorithm in order to classify a given dataset through a certain number of clusters (assuming number of clusters) fixed a priori. The main idea is to define c number of centroid points i.e. one centroid point for each cluster. The dataset is separated into a predefined ' c ' number of clusters, where total number of data points ' n '. In this method, each cluster is associated with a centroid point and each data point is then assigned to cluster with the nearest centroid. During this process, the distance of the centroid of a cluster to each data point in the dataset is measured, and data points are assigned to the cluster nearest to it. The algorithm aims at minimizing an objective function (as shown in Equation 2), which in this case is a squared error function.

(2)

In the above equation, $X = \{x_1, x_2, \dots, x_n\}$ represents the set of n data points, $V = \{v_1, v_2, \dots, v_c\}$ represents the set of cluster centres for c number of clusters, and c_i is the number of data points in the cluster. The recursive algorithm for k-means clustering technique is shown in the following flowchart (FIGURE 1).

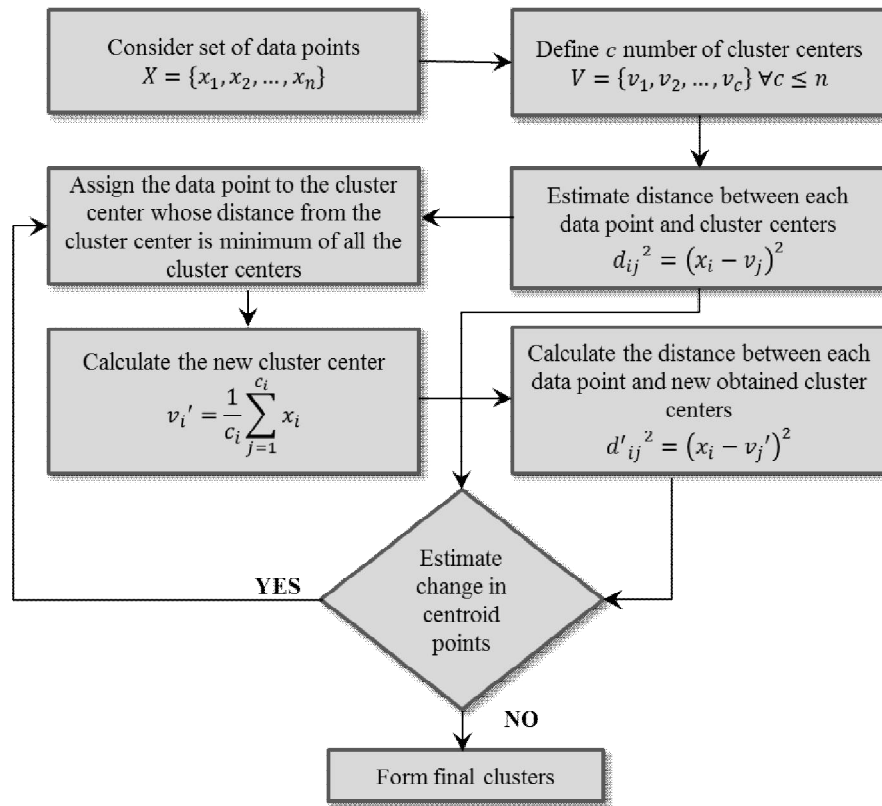


FIGURE 1 Flowchart for the K-means clustering

In order to validate the clustering result, silhouette method is employed for validation of consistency of the estimated clusters (27). The silhouette value is a measure to check similarity of a set of data points to their own cluster as compared to other clusters. A silhouette value of a cluster ranges between -1 and 1. The silhouette value closer to 1 indicates that the data point is well matched to its own cluster, and poorly matched to nearby clusters. This method provides a graphical representation of how well each object lies within its cluster. The silhouette value can be calculated using any distance metric such as the Euclidean

distance or the Manhattan distance. In this study, squared Euclidean distance is considered as the distance metric. The silhouette coefficient of a point 'i' is then given by Equation 3.

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \quad (3)$$

Where S_i is the silhouette coefficient of the point i , a_i = average distance of i to the points in its cluster, b_i = average distance of i to points in neighbouring/second best cluster.

Silhouette value closer to 1 is desirable and negative value indicates $a_i > b_i$ i.e. the point i is closer to a neighbouring cluster than its own cluster. Interpretation of Silhouette value is summarised in **TABLE 1** (28).

TABLE 1 Interpretation of Silhouette Value

Range of Silhouette Value	Interpretation
0.71-1.0	A strong and robust cluster has been formed
0.51-0.70	A reasonable cluster has been found
0.26-0.50	The cluster is weak and could be artificial
-1.0- 0.25	No substantial clustering can be found

The overall study approach is shown in the following schematic diagram (**FIGURE 2**).

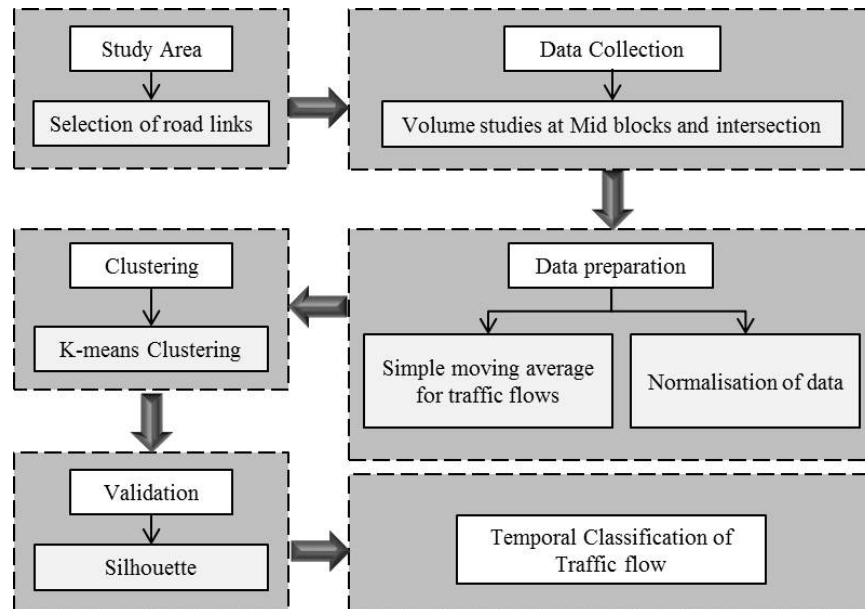


FIGURE 2 Methodological Framework

3.0 STUDY AREA, SURVEY ADMINISTRATION, AND DATABASE DEVELOPMENT

The approach for this study is demonstrated taking the case study of the multi-modal urban transportation network in Bhubaneswar city. The road network within the geographical boundary of the city municipality comprises of 1265 km. of road stretch. The pattern of road network is both organic and grid iron in the old town area and new city area respectively. Roads in the old town area of Bhubaneswar have single carriageway, with limited Right of Ways (RoW) which limits road capacity and scope of expansion. Further, the old town area has a higher percentage of pedestrian and slow moving vehicles which results in

congestion and inhibition in mobility. The new city area (newly developed areas) has planned road network with identifiable hierarchy, and sufficient right-of-way (RoW) for smooth traffic movement. The new city area has a higher percentage of heterogeneous mix of traffic and relatively low pedestrian.

In order to identify the locations for data collection and analysis, firstly, a reconnaissance survey was carried out in the study area to gain an understanding of the different road classes and traffic condition of the city. The study area comprises of 36 kms National Highway, 48 kms. of Arterial roads (major roads which cater the primary traffic formulating the core network pattern of the city), 98 kms. of Sub arterial roads (roads which acts as feeder to arterials), 190 kms. of collector roads (minor roads which acts as feeder for Sub arterials and Arterials in some cases), 893 kms. of local roads (minor roads which connects residential units). A total of 43 different locations/different road segments were selected across the city which included stretches along major arterial roads, sub-arterial roads, collector roads, national highway and also locations in major intersections of the study area. Out of these 43 locations, 38 locations comprised of mid-block sections which includes the five entry and exit links of the city; while the remaining five locations comprises of major intersections in the city (**FIGURE 3**). These locations were selected based on the volume of traffic, connectivity, important bus route and major entry and exit of the city.

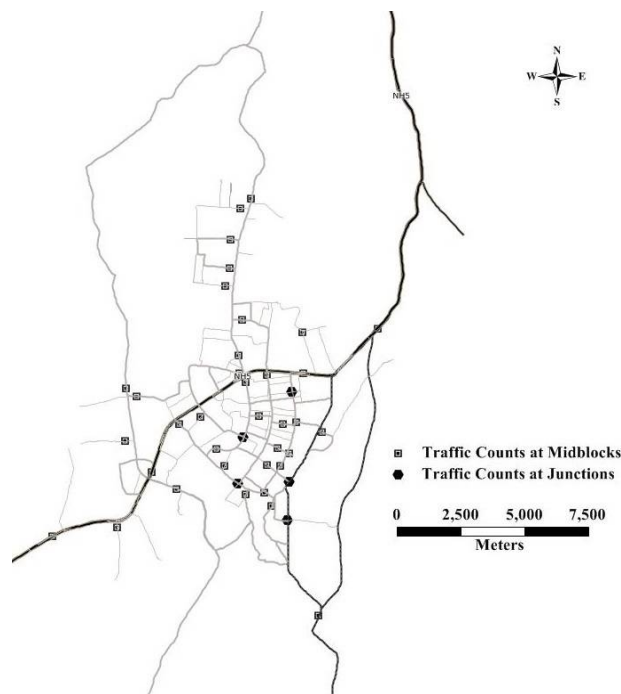


FIGURE 3 Traffic volume survey locations in Bhubaneswar city

In each of the location selected, volume data was required for identifying the peak and off peak durations. Hence, videographic surveys of the locations were conducted for a typical weekday. Videographic surveillance on weekends and holidays were not conducted. Further, data was not collected on days when there is a spike or drop in traffic volume due to special events occurring in the city such as religious festivals, major examinations or public demonstration. Sufficient number of cameras were placed such that the entire location could be covered by the video surveillance. On each of the 43 locations, surveys were carried out for a total duration of 16 hours; from 06:00 a.m. to 10:00 p.m. on typical weekday. The videographic data was then extracted for 10 minute interval. Volume data were further segregated based on location, direction of travel, and vehicular mode.

4.0 ANALYSIS, RESULTS, AND DISCUSSION

Preliminary investigation of the traffic flow data revealed that the composition of traffic was dominated by two wheelers followed by private cars. The share of different modes observed during the survey is shown in **FIGURE 4**.

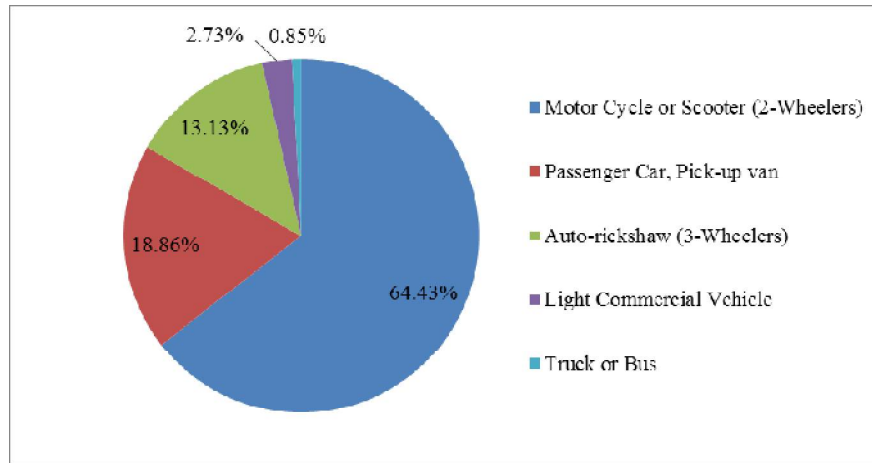


FIGURE 4 Traffic composition in the study area

As the traffic pattern is found to be heterogeneous in nature and consisting of various vehicle modes of different static and dynamic characteristics, it was necessary to convert the mixed traffic flow into homogenous traffic flow using equivalent units known as passenger car units (PCU). In the present study, the following equivalent PCU factors (**TABLE 2**) (22) were considered in order to convert the mixed traffic flow in terms of PCU. The following section discusses the analysis and results from the data collected.

TABLE 2 PCU Factors for Different Vehicle Types

S. No.	Vehicle type	Equivalent PCU factors	
		5%*	>=10%*
1	Motor Cycle or Scooter (2-Wheeled)	0.5	0.75
2	Passenger Car, Pick-up van	1	1
3	Auto-rickshaw (3-Wheeled)	1.2	2
4	Light Commercial Vehicle	1.4	2
5	Truck or Bus	2.2	3.7

* Percentage composition of vehicle type in traffic stream

In order to identify peak and off-peak duration, it is important to obtain the temporal variation of traffic volumes. This temporal variation of traffic flow shows the pattern of peak and off-peak time and duration. Hence, traffic volumes from 43 locations were converted to equivalent PCU values and graphically plotted against the time of the day. As the data is extracted for 10 minute interval, it is possible to observe the changes in traffic volume for both short duration (10-30 minutes) and long duration (30 minutes to 16 hours). Since the objective of the study is to capture the long term variation in traffic flow pattern, it is important to reduce the noise created by short term variations; hence, the flow pattern was smoothed by Simple Moving Average (SMA) technique. SMA was used on the traffic flow volume for all links independently. **FIGURE 5** shows an example of the flow pattern of road link (A6) before and after application of SMA. While performing SMA the mean of three consecutive values are taken and represented corresponding to the central value, hence, the first and last data points are not

represented. Therefore, the flow values from 6:10 a.m. to 9:50 p.m. were represented in the plots of smoothed curve.

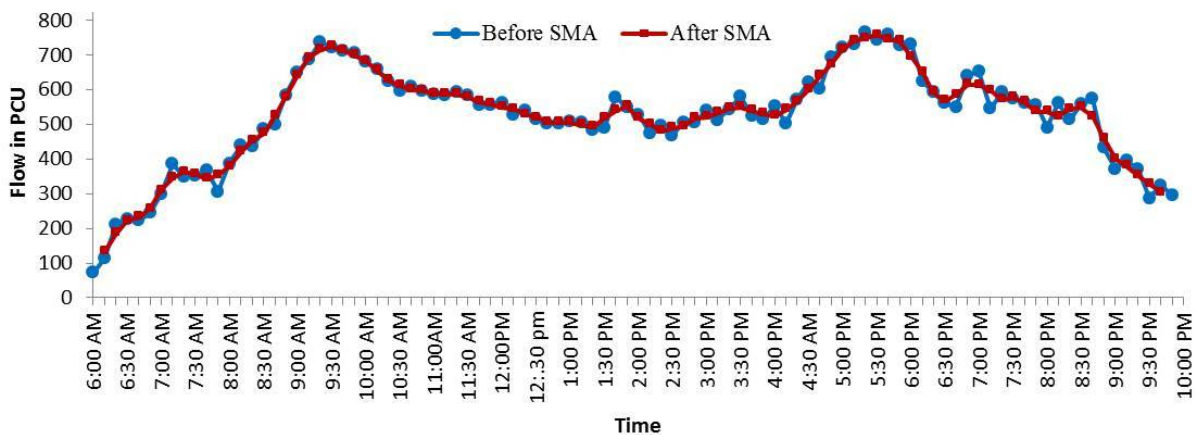


FIGURE5 Example of Simple moving average method applied on a traffic flow pattern of selected link (A6)

Once the smoothed traffic volume data was obtained, the volume data was clustered using k-means clustering algorithm to obtain peak and off-peak values. According to the scope of work, K value of three was more appropriate as it differentiates the peak volumes from average flows and free flows. The clustering was done using the statistical package of MATLAB. Based on the analysis, the peak periods are identified during 9:00 a.m. to 10:30 a.m. and 4:50 pm to 6:20 pm. However, it can be noted that few intervals 4:30 p.m. to 4:40 p.m. and 6:50 p.m. to 7:10 p.m. are also identified as peak. This is due to the short term variations. The clustering when performed after the application of SMA the peak periods are identified during 9:00 a.m. to 10:30 a.m. and 4:50 pm to 6:20 pm. Although there is no change in the peak durations identified it can be observed the short term variations are eliminated (**FIGURE 6**).

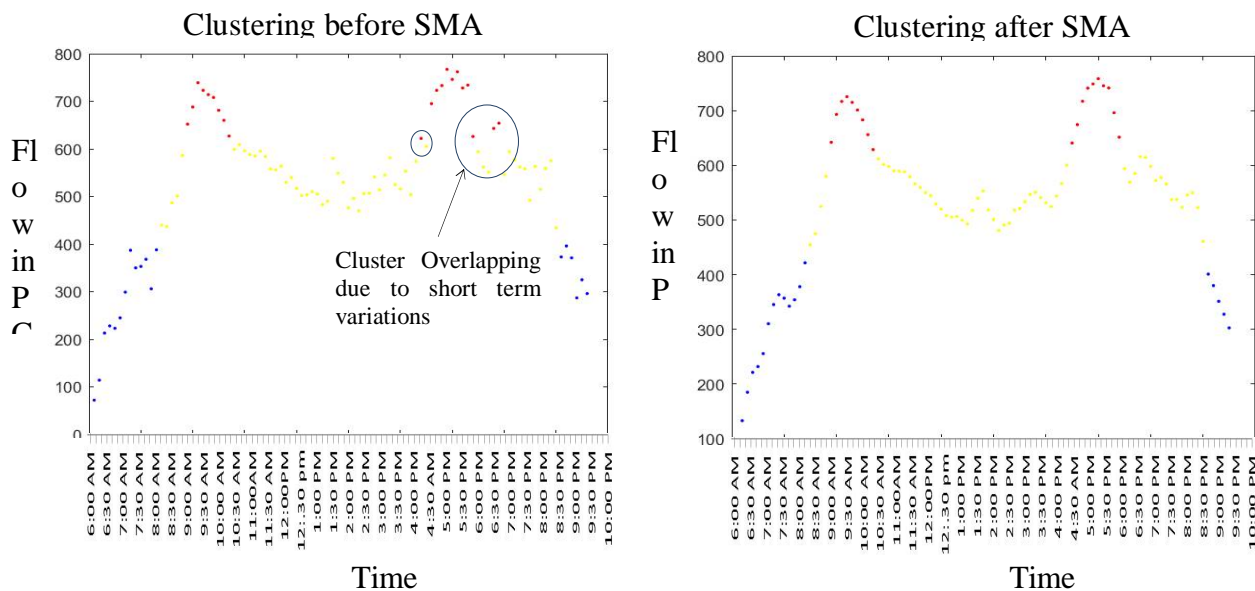


FIGURE 6 K-means clustering before and after SMA

FIGURE 7 shows the improvement in cluster structure of the link A6 before and after application of SMA respectively using silhouette plot (Silhouette value 0.689 to 0.702). Although the improvement is marginal SMA was able to reduce the short term variations in the traffic flow.

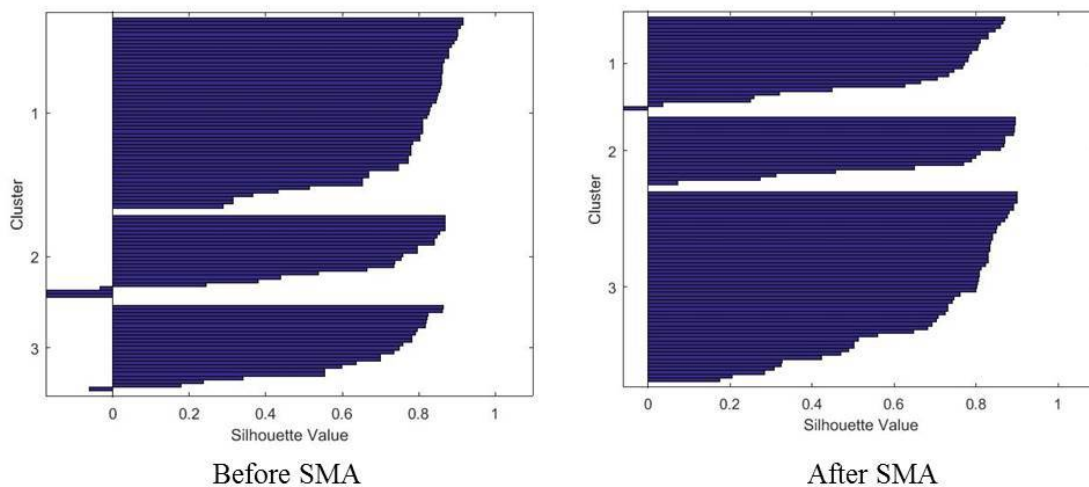


FIGURE7 Cluster structure before and after the application of SMA

Using the technique mentioned above it is possible to obtain the peak and off peak hours of a road segment. It was observed from the graphical plot that for different links the peak flow was recorded at different times. Hence, in order to identify the peak and off peak hours of a city/ CBD which has various road segments of varying capacity leading to and from the area, a slightly different methodology needs to be adopted. As the main objective of this study was to find the peak and off-peak durations of the city, only the arterial roads which lead to the city (29) were considered for further analysis. A total of 10 arterial road links (A1-A10) approaching the CBD area were selected as to represent the traffic around the area. As SMA was found to be instrumental in reducing the short term variations it was applied for all the selected links. From the selected links to formulate one representative traffic flow pattern of the city, all the ten arterial link volumes were first normalised on a 0-1 scale as shown in **FIGURE 8**.

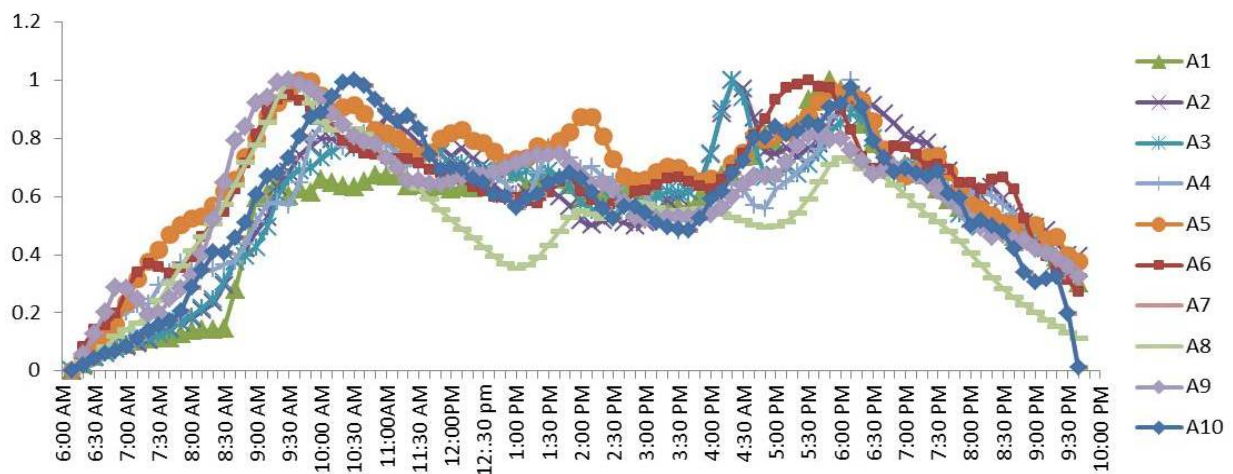


FIGURE8 Normalised flow pattern of selected road links

Then the mean of the normalised data was used as the representative of the traffic flow pattern of the city (**FIGURE 9**). Normalising the flow pattern of individual links prior to estimating the mean flow

brings all the links to a same scale keeping the individual traffic pattern intact and gives equal weight to all the selected links. The obtained representative flow pattern is used for further analysis.

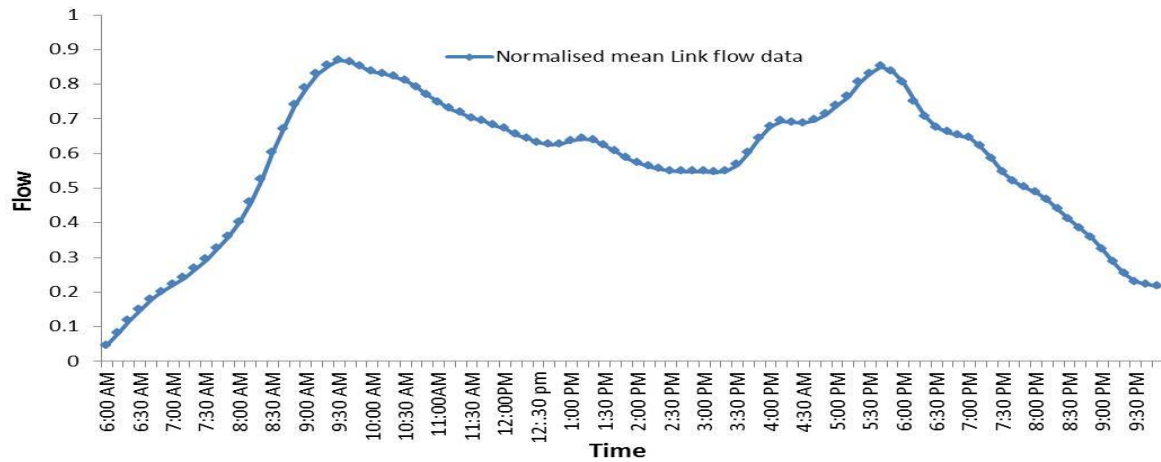


FIGURE 9 Mean of normalised flow of selected road links

For clustering the time series data, k-means clustering was used similar to that of a single road link. Based on the analysis, the two peak periods are observed during 8:50 am – 12:00 pm and 4:00 pm – 6:20 pm as reported in **FIGURE 10**. It is interesting to note that the peak periods were observed for more than 3 hours in the morning and more than 2 hours in the evening. Although when traffic flow pattern for individual links were considered, the peak period was generally found to be less than 2 hours. This can be attributed to the fact the peak periods of different links selected do not occur at the same time, but have a slight shift in their peak durations depending upon their respective locations. However, the observed results indicate that around the CBD area extended peak periods can be noticed.

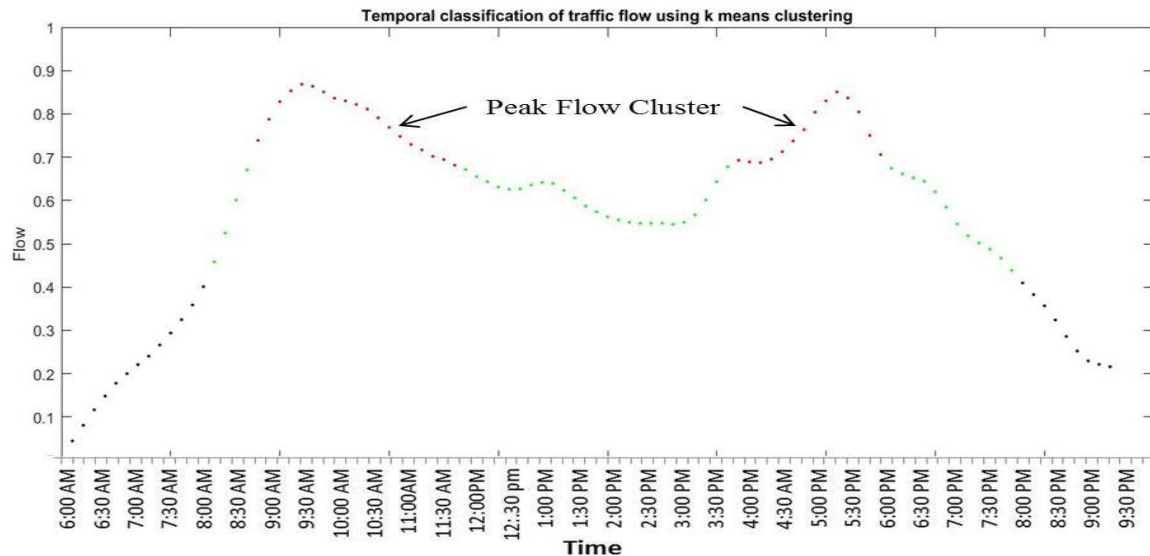


FIGURE 10 Clustering of traffic flow by k-means clustering

Additionally, to check the consistency of the clusters obtained, silhouette value was estimated and was found to be 0.7247 and plotted in **FIGURE 11**, which indicates a very strong cluster structure.

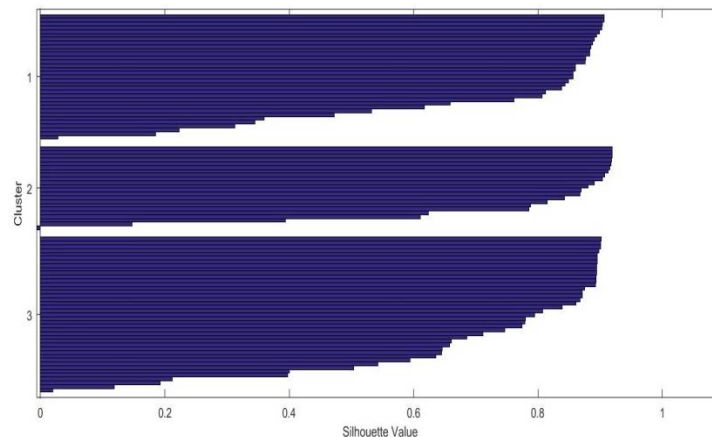


FIGURE 11 Silhouette plot of clusters obtained from K-Means clustering

5.0 CONCLUSION

The study demonstrates a cluster based approach for temporal classification of urban traffic flow pattern in order to identify the peak and off-peak durations with reference to an urban area in India. The study classifies the within-day traffic flow pattern by considering inherent temporally continuous nature of traffic flow data, and uses a non-hierarchical clustering algorithm called K-means clustering. The work has been demonstrated by considering the multi-modal urban transportation network of Bhubaneswar city, India. In order to carry out the study, traffic volume data was obtained on 10 minute intervals for a period of 16 hours, from 06:00 am to 10:00 pm on weekdays. The traffic flow data was smoothed in order to reduce the short-term variations using Simple Moving Average (SMA) of the power of three. The peak and off-peak durations were identified using an unsupervised k-means clustering technique. Silhouette analysis was undertaken to validate the cluster's structure, and the result exhibits a strong cluster structure. The peak durations identified for the city are 8:50 am to 12 noon (morning peak) and 4pm to 6:20 pm (evening peak). Although, this result is case specific, the approach demonstrated may be used for rational identification of time periods of traffic flow in other urban areas. Further, this approach may be applied to identify peak and off-peak hours of a smaller location such as a road segment. The other clustering approaches may be explored to compare and validate the current cluster structure for identification of peak and off-peak durations of traffic flow in a multi-modal urban transport environment.

ACKNOWLEDGMENT

Authors would like to acknowledge The Ministry of Human Resource Development, Government of India for funding the study as a part of 'Future of Cities' project to improve the quality of urban mobility.

REFERENCES

1. Boillot, F., and M. Papageorgiou. A real time coordinated optimal control approach for urban traffic networks. In: Second International Capri Seminar on Urban Traffic Networks, 1992, pp. 753–766.
2. Gartner, N.H., C. Stamatiadis, and P.J. Tarnoff. Development of advanced traffic signal control strategies for intelligent transportation systems: multilevel design Transportation Research Record, Vol. 1494, 1995, pp. 98-105
3. Jiang, X. and H. Adeli. Wavelet Packet-Autocorrelation Function Method for Traffic Flow Pattern Analysis. *Computer-Aided Civil and Infrastructure Engineering*, Vol. 19(5), 2004, pp.324-337.
4. Weijermars, W. A. M. *Analysis of urban traffic patterns using clustering*. University of Twente, 2007.
5. Turner, S. M., W. L. Eisele, R. J. Benz, , and D. J. Holdener. *Travel time data collection handbook* (No. FHWA-PL-98-035), 1998.

6. Patel, C. R., and G. J. Joshi. Capacity and LOS for urban arterial road in Indian mixed traffic condition. *Procedia-Social and Behavioral Sciences*, Vol. 48, 2012, pp. 527-534.
7. Batterman, S., R. Cook, and T. Justin. Temporal variation of traffic on highways and the development of accurate temporal allocation factors for air pollution analyses. *Atmospheric Environment*, Vol. 107, 2015, pp. 351-363.
8. Hunt, P. B., D. I. Robertson, R. D. Bretherton, and R. I. Winton. SCOOT-a traffic responsive method of coordinating signals (No. LR 1014 Monograph), 1981.
9. Gartner, N. H., P. J. Tarnoff, and C. M. Andrews. Evaluation of optimized policies for adaptive control strategy. *Transportation Research Record*, Vol. 1324, 1991.
10. Williams, B., P. Durvasula, and D. Brown. Urban freeway traffic flow prediction: application of seasonal autoregressive integrated moving average and exponential smoothing models. *Transportation Research Record: Journal of the Transportation Research Board*, 1644, 1998, pp. 132-141.
11. Smith, B. L. Forecasting freeway traffic flow for intelligent transportation systems application. *Transportation Research Part A*, Vol. 1(31), 1997, pp. 61.
12. Williams, B. M., & L. A. Hoel. Modeling and forecasting vehicular traffic flow as a seasonal stochastic time series process (No. LTVA/29242/CE99/103), 1999.
13. Park, B., D. H. Lee, and I. Yun. Enhancement of time of day based traffic signal control. In *Systems, Man and Cybernetics, 2003 IEEE International Conference*, Washington, DC, Vol. 4, 2003, pp. 3619–3624.
14. Yang, L. C., L. Jia, , Q. J. Kong, and W. X. Zhu., Method of automatic programming traffic intervals based on artificial immune algorithm. *KongzhiLilunyuYingyong/Control Theory & Applications*, Vol. 23(2), 2006, pp. 193–198.
15. Guo, R., and Y. Zhang. Identifying time-of-day breakpoints based on nonintrusive data collection platforms. *Journal of Intelligent Transportation Systems*, Vol. 18(2), 2014, pp. 164-174.
16. Hauser, T. A., & W. T. Scherer. Data mining tools for realtime traffic signal decision support & maintenance. In *Systems, Man, and Cybernetics, 2001 IEEE International Conference*, Tucson, AZ, Vol. 3, 2001, pp. 1471–1477.
17. Smith, B. L., W. T. Scherer, T. A. Hauser, and B. B. Park. Data-driven methodology for signal timing plan development: A computational approach. *Computer–Aided Civil and Infrastructure Engineering*, Vol. 17(6), 2002, pp. 387–395.
18. Wang, X., W. Cottrell, and S. Mu. Using k-means clustering to identify time-of-day break points for traffic signal timing plans. In *Intelligent Transportation Systems, Proceedings, 2005 IEEE*, Vienna, Austria, pp. 586–591.
19. Ratrou, N. T. Subtractive clustering-based K-means technique for determining optimum time-of-day breakpoints. *Journal of Computing in Civil Engineering*, Vol. 25(5), 2010, pp. 380–387.
20. Ratrou, N. T. Developing optimal timing plans for cyclic traffic along arterials using pre-timed controllers. *Urban Transport XVII: Urban Transport and the Environment in the 21st Century*, Vol.116, 2011, pp. 367.
21. J. Lee, J. Kim, and B. B. Park. A genetic algorithm-based procedure for determining optimal time-of-day break points for coordinated actuated traffic signal systems. *KSCE Journal of Civil Engineering*, Vol. 15(1), 2011, pp. 197–203.
22. Indian Roads Congress (IRC). (1990). “Guidelines for capacity of urban roads in plain areas.” IRC: 106, New Delhi, India.
23. Xu, L., P. C. Ivanov, K. Hu, Z. Chen, A. Carbone, and H. E. Stanley. Quantifying signals with power-law correlations: A comparative study of detrended fluctuation analysis and detrended moving average techniques. *Physical Review E*, Vol. 71(5), 2005, 051101.

24. NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, updated in 2012, accessed on 09/05/17.
25. Kaul, R., & R. S. Chowdhury. Applied Statistics: Economic statistics. Department of Statistics, Lady Shri Ram College for Women, New Delhi, 2007.
26. Bratchell, N. Cluster analysis. *Chemometrics and intelligent laboratory systems*, Vol. 6(2), 1989, pp. 105-125.
27. Rousseeuw, P. J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, Vol. 20, 1987, pp. 53-65.
28. Department of Statistics, University of California. Performing and Interpreting Cluster Analysis, 2007. Available at <https://www.stat.berkeley.edu/~spector/s133/Clus.html> accessed on 09/05/17
29. Stathopoulos, A., and M. Karlaftis. Temporal and spatial variations of real-time traffic data in urban areas. *Transportation Research Record: Journal of the Transportation Research Board*, Vol. 1768, 2001, pp. 135-140.