Identifying Traffic Congestion Pattern using K-means Clustering Technique

Md. Ashifuddin Mondal
Dept. of Computer Science and Engineering
Narula Institute of Technology
Agarpara, Kolkata-700109, India
ashifuddin.mondal@nit.ac.in

Zeenat Rehena
Dept. of Computer Science and Engineering
Aliah University
II A/27, Newtown, Kolkata-700160, India
zeenatrehena@yahoo.co.in

***Abstract*—With increase in urbanization and socio-economical growth, the number of vehicles in major metropolitan cities is increasing day by day. Therefore, traffic congestion is becoming a major concern of metropolitan cities all over the world. This results in tremendous air pollution, loss of valuable time and money of citizens. Hence, traffic congestion monitoring of different road segments is very essential for analyzing the problem associated with smooth mobility. Identifying the problematic road segments within the city is one of the important job for the transport authority to assess the road condition. That will assist the government agencies or policy makers to optimize traffic rules and regulations. This work identifies traffic congestion pattern which can classify the different road segments based on traffic density and average speed of vehicles. The traffic parameters are captured by in-road stationary sensors deployed in road segments. The proposed system uses k-means clustering algorithm to categorize the different road segments.**

***Index Terms*—Clustering, in-road sensors, traffic congestion pattern, k-means**

I. Introduction

The socio-economical growth and rural to urban migration of people are putting an enormous pressure on existing road infrastructure [1]. According to [2], urban population of India is increasing at an average rate of 3% per year and by the end of year 2021 it will reach 500 million from existing 300 million. The rapid growth in population results in huge number of vehicles in streets of urban areas. The number of registered vehicles in India has reached 142 million in 2011 from 0.3 million in 1951 [3]. Due to vehicular growth, traffic congestion has become a major problem in the roads of urban areas. Hence, traffic management authority facing huge problem to manage the traffic congestion. The road infrastructures of cities need to be managed efficiently to enhance the standard of living of citizens. The concept of smart city is evolving as an effective means to manage different challenges of cities. Smart mobility of vehicles is one of the major challenges that need to be addressed to deal with the problem of traffic congestion [4] on road. Traffic congestion causes several problems like increase in air pollution, travel time, fuel consumption and operational cost of vehicle. Road accident is also a consequence of traffic congestion. According to World Bank study [3], India experiences monetary loss of $6 billion a year due to traffic congestion. Hence, it is obligatory to design a traffic management system to deal with congestion problem.

The aim of this paper is to propose a traffic congestion pattern recognition system. It acquires real time road traffic information and categorizes the different road segments based on traffic density and average speed of vehicles. The proposed system uses K-means clustering technique [5] to segregate the road segments within the city. It collects the traffic information from in-road stationary sensors and feeds to the data analysis module to identify traffic congestion pattern without human intervention. The system will assist the transport authorities or policy makers to plan and design traffic management rules and regulations. It also helps in decision making regarding whether further infrastructure is required or not to cope with the congestion.

The rest of the paper is organized as follows: Section II highlights related work on traffic monitoring and analysis. Section III describes the proposed system design for traffic pattern identification. The simulation environment and result analysis of traffic pattern identification is presented in Section IV. Finally, Section V concludes the paper.

II. Related Work

A lot of research work has been carried out in the field of different aspects of intelligent transportation system (ITS) for proper road traffic management. Segregation of different road segments based on traffic congestion is one of the important task of ITS.

In [6], authors studied traffic congestion pattern of Beijing city. They have used traffic congestion index to show congestion intensity and used clustering method to identify congestion pattern. In [7], author proposed traffic pattern analysis technique based on GPS data. The author used statistical approach to infer traffic patterns and trends from GPS data. In another study [8], authors proposed k-means clustering method for segmenting road accident data. In [9], authors used support vector machine to recognize transport condition. The paper classifies transportation condition patterns by considering three attributes: traffic volume, average speed and occupation ratio. In another work [10], authors proposed data mining approach for identification of accident prone area in different road segments. They have used k-means clustering approach to form

978-1-7281-1253-4/19/$31.00 © 2019 IEEE

the different groups comprising with accident prone location based on frequency of road accidents. Authors also analyzed these data to identify the main cause of road accidents of those locations. A methodological technique [11] have been used in different road link which uses link-based capacity disruption values to identify the critical road segments. Also they have ranked the road segments based on criticality and asses the robustness of the road network.

According to the above discussion, it is found that some of the authors used k-means clustering technique to form groups of different accident prone areas, while this paper aims to segregate the road segments depending on density and average speed and thus helps the traffic authority to identify the problematic road segments in terms of traffic congestion.

III. Proposed System

In this section traffic congestion pattern recognition system has been proposed. The main aim of the work is to cluster the road segments based on different traffic parameters namely traffic density and average speed of the vehicles. Each cluster will contain road segments of same type in terms of density and average speed. This traffic pattern may be considered as a very useful tool to policy makers in decision making, defining traffic rules and regulations.

A. *Overview of the System:*

The fundamental units of proposed system are: data acquisition unit, data transmission unit and data processing unit. Figure 1 shows the proposed system to identify traffic congestion pattern.

a) Data Acquisition Unit: For proper strategies and planning, precise and real time data acquisition is important. Different ways are there to collect traffic information like: CCTV, video camera, infrared and floating car data. Here, the system considers in-road stationary sensors to collect data from roads.

b) Data Transmission Unit: This unit is responsible of transmitting the acquired data through wireless or wired communication from on-site to the remote server for further processing.

c) Data processing Unit: The traffic data that has been collected by sensors are processed by processing unit to identify problematic road segments in a road network. The system uses K-means clustering technique to form a number of clusters of different road segments. This knowledge can be used by policy makers effectively for proper planning to minimize traffic congestion.

In the next section K-means clustering technique is explained.

B. *Overview of k-means Clustering Technique:*

K-means clustering [6] is one of the simplest unsupervised machine learning techniques used to partition the given data set in to k number of clusters in which each objects of similar

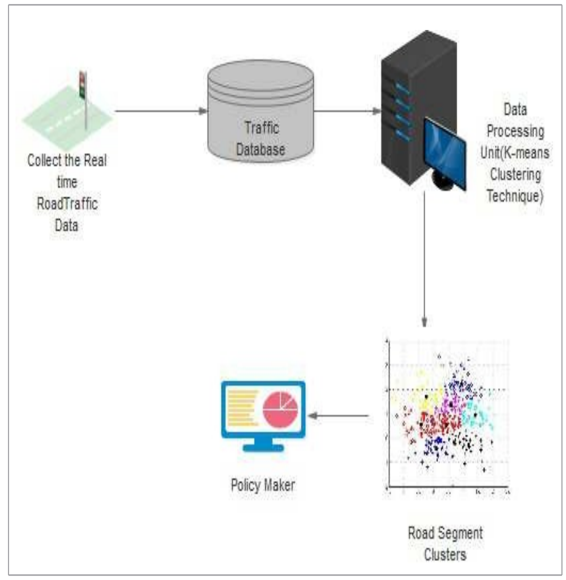

Fig. 1. System for identifying traffic congestion pattern

Algorithm 1: K-means Clustering

Input: Dataset $D = d_1, d_2, d_3, ..d_n$ containing n objects, number of cluster k

Output: Set of clusters

1. Randomly chose k no. of data from D as a initial centroids.
2. Calculate the Euicldean distance between each object $d_i \in D$ and every centroid and assign the d_i to the nearest cluster.
3. Recalculate the centroid by mean of all objects in a newly formed cluster.
4. Repeat *steps 2 to 3*, until the centroid stop moving.
5. Stop

The proposed system forms four clusters based on density and average speed: high density-low speed (cluster 1), medium density-low speed (cluster 2), medium density-moderate speed (cluster 3) and low density-high speed (cluster 4).

IV. Simulation Environment

To simulate the proposed technique, python has been used. The proposed system considers four clusters, hence the value of k in k-means clustering algorithm is 4. Table I shows the sample traffic data which are fed into data processing module and it forms four different groups of road segments using k-means clustering technique.

Figure 2 shows the initial graphical presentation of traffic

data. Figure 3 shows the clusters of different road segments formed using k-means clustering
type are placed in single cluster. Algorithm 1 shows the K-means algorithm used in this work.

Based on k-means clustering technique mentioned in algorithm 1, the system categorized these data into several clusters.

algorithm. Figure 3 depicts

four clusters of road segments with different colours. The colour black indicates high density-low speed (cluster 1), red indicates medium density-low speed (cluster 2), and magenta represents medium density-moderate speed (cluster 3) and

TABLE I
TRAFFIC PARAMETERS

Traffic Parameters	
Density (No. of cars/km)	Avg. Speed (km/hr.)
90	15
60	30
60	30
55	33
75	20
80	18
..	..
..	..
39	73
40	71
80	18
38	74

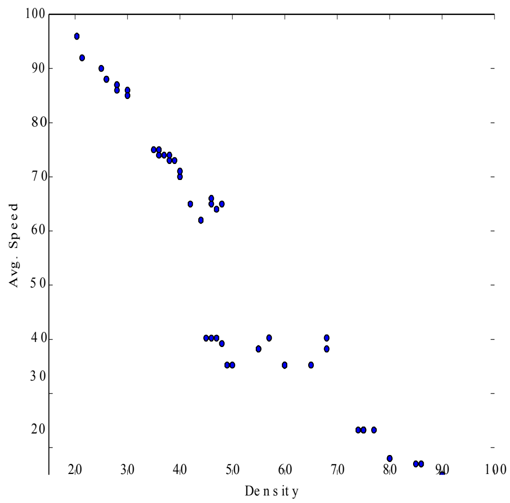

Fig. 2. Sample Road Segment Traffic Data

blue represents low density-high speed (cluster 4).
The *elbowpoint* concept is being used to determine the number of clusters. The main objective is to minimize the sum of squared distance (SSE). SSE is defined as the sum of square distance between each data points of the cluster and its centroid. Sum of squared distance is plotted against number of clusters k in Figure 4. SSE will be decreased when number of cluster increases. The *elbowpoint* is the point where the rate of decrease of mean distance i.e. SSE will not change significantly with increase in number of clusters. Table II shows the corresponding value of SSE with respect to number of clusters. Figure 4 indicates the *elbowpoint* based on sample traffic data shown in Table I.

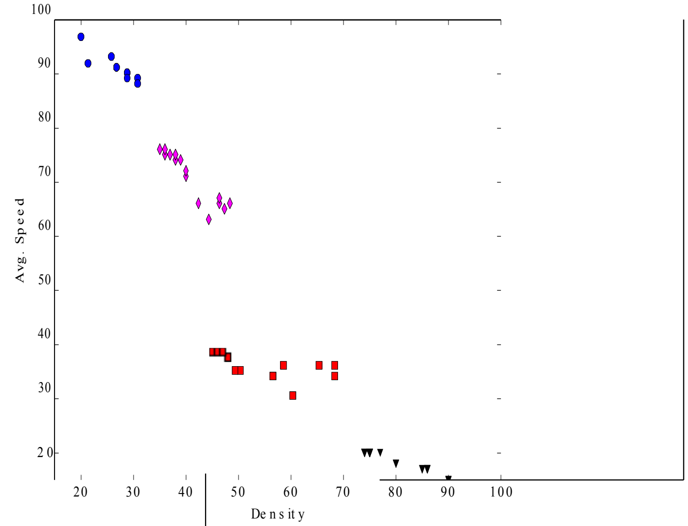

Fig. 3. Clustering of Road Segments

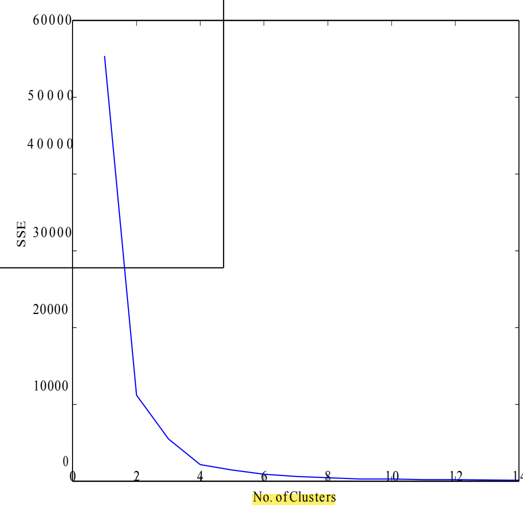

Fig. 4. Elbow Point of the System

TABLE II
VARIATION OF SSE

No. of Cluster (K)	SSE
1	55376.431372549036
2	11205.166153846149
3	5483.7009523809511
4	2136.4142857142856

The proposed system helps to identify the critical or problematic road segments based on road traffic data (traffic density and average speed of vehicles) and it assists the transport authority and policy makers to assess the robustness of the road network. For example, traffic management authority should pay attention to the cluster 2 (in red color) which signifies that the average speed of vehicles in the road segments of that cluster is low instead of medium density, whereas average speed of vehicles in the road segments belonging to cluster 3 (in magenta color) is moderate in medium density. It may happened because the road segments of cluster 2 may not be wide enough to accommodate many vehicles or may be traffic

978-1-7281-1253-4/19/$31.00 © 2019 IEEE

rules are not obeyed properly by travelers. Hence, traffic management authority can try to find out the reason behind traffic congestion in the road segments belonging to cluster 2 and can take necessary steps to overcome the problem.

V. Conclusion

The paper proposes smart traffic congestion pattern recognition system based on k-means clustering technique to classify different road segments. It clusters into four categories based on traffic density and average speed of the vehicles. The system will assist the policy makers to identify the problematic road segments and will reform strategies regarding traffic rules and regulations to reduce traffic congestion. The decision can be taken whether new infrastructure (flyover, bypass road) is required or not to minimize traffic congestion as well as to minimize road accidents.

References

[1] Z. Rehena and M. Janssen, "Towards a framework for context-aware intelligent traffic management system in smart cities," in *Companion Proceedings of the The Web Conference 2018*, ser. WWW '18. Republic and Canton of Geneva, Switzerland: International World Wide Web Conferences Steering Committee, 2018, pp. 893–898. [Online]. Available: https://doi.org/10.1145/3184558.3191514

[2] S. Singh, "Urban transport in india: Issues, challenges, and the way forward," *European Transport - Trasporti Europei*, 12 2012.

[3] T. Rawal and V. Devadas, "Intelligent transportation system in india- a review," *Journal of Development Management and Communication*, vol. 2, no. 3, Apr. 2015.

[4] A. Monzon, "Smart cities concept and challenges: Bases for the assessment of smart city projects," in *Smart Cities, Green Technologies, and Intelligent Transport Systems*, M. Helfert, K.-H. Krempels, C. Klein, B. Donellan, and O. Guiskhin, Eds. Cham: Springer International Publishing, 2015, pp. 17–31.

[5] S. Na, L. Xumin, and G. Yong, "Research on k-means clustering algorithm: An improved k-means clustering algorithm," in *2010 Third International Symposium on Intelligent Information Technology and Security Informatics*, April 2010, pp. 63–67.

[6] H. Wen, J. Sun, and X. Zhang, "Study on traffic congestion patterns of large city in china taking beijing as an example," *Procedia - Social and Behavioral Sciences*, vol. 138, pp. 482 – 491, 2014, the 9th International Conference on Traffic and Transportation Studies (ICTTS 2014). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877042814041469

[7] E. Necula, "Analyzing traffic patterns on street segments based on gps data using r," *Transportation Research Procedia*, vol. 10, pp. 276 – 285, 2015, 18th Euro Working Group on Transportation, EWGT 2015, 14-16 July 2015, Delft, The Netherlands. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2352146515002641

[8] P. A. Nandurge and N. V. Dharwadkar, "Analyzing road accident data using machine learning paradigms," in *International conference on I-SMAC*, 2017.

[9] R. Yu, G. Wang, J. Zheng, and H. Wang, "Urban road traffic condition pattern recognition based on support vector machine," *Journal of Transportation Systems Engineering and Information Technology*, vol. 13, p. 130136, 02 2013.

[10] S. Kumar and D. Toshniwal, "A data mining approach to characterize road accident locations," *Journal of Modern Transportation*, vol. 24, no. 1, pp. 62–72, Mar 2016. [Online]. Available: https://doi.org/10.1007/s40534-016-0095-5

[11] J. Sullivan, D. Novak, L. Aultman-Hall, and D. Scott, "Identifying critical road segments and measuring system-wide robustness in transportation networks with isolating links: A link-based capacity-reduction approach," *Transportation Research Part A: Policy and Practice*, vol. 44, pp. 323–336, 06 2010.