Estimation of Traffic Congestion Level via FN-DBSCAN Algorithm by Using GPS Data

Ahmet Can Diker[1], Elvin Nasibov[2]
[1]Dokuz Eylül University, Izmir, Turkey
[2]Ege University, Izmir, Turkey
[1]*dikerahmetcan@gmail.com*, [2]*elvin.nasibov@yahoo.com*

Abstract— **Determination of traffic congestion level is one of the fundamental problems in Intelligent Transportation Systems (ITS). In this paper, fuzzy based data mining technique, namely, Fuzzy Neighborhood Density-Based Spatial Clustering of Applications with Noise (FN-DBSCAN) was performed to cluster road segments with traffic congestion level. Data were collected from portable navigation device in probe car on selected roads in Izmir. Six clusters were obtained as a result of experimental study and these clusters were named traffic congestion levels. It is considered that this paper will provide a contribution to related work.**

Keywords— ***data mining; clustering; FN-DBSCAN; intelligent transportation systems; traffic congestion***

I. INTRODUCTION

Traffic congestion is one of the most fundamental issues in developed and developing countries. As a result of population growth, number of vehicle and driver in the traffic increases with each passing day and causes the traffic congestion. Such as Advanced Traffic Management Systems (ATMS), Emergency Management Systems (EMS) and etc. in ITS are important role to information about traffic congestion level [1]. The basis of ITS is the acquisition of traffic related data, which allows to judge how traffic is moving, with installation of sensor technologies for the collection of various types of data [2]. Road traffic congestion, are obtained by static sensors and dynamic sensors. In this study velocity, time and road information are collected by portable navigation device in probe vehicle as a dynamic sensor due to provided high quality in traffic monitoring.

There is no standard measurement to determine road traffic congestion level. Measurements using different metrics such as travel time, travel delay, travel time index, travel ratio index, etc. were developed for determination of the traffic congestion degree [2, 3]. These metrics were commonly used to estimate the congestion in each segment which has length with an equal interval. However, our study differs from previous studies in that we perform FN-DBSCAN method which is one of the fuzzy based data mining techniques using variable length of segments with traffic congestion level [3].

Nasibov and Ulutagay showed that FN-DBSCAN method had given more realistic results to recognize the stable duration intervals and the BIS stages in the measurement series [3]. Bispectral index scale (BIS) is a continuous processed electroencephalogram (EEG) parameter that correlates to the patient's level of brain activity and to investigate brain disorders. Traffic data resemble electroencephalographic records. Similarly, we will use also FN-DBSCAN in order to determine traffic congestion level in this paper.

Levels of traffic congestion were determined using different methodologies with several data source in many studies [4-6]. Pattara, Pongpaibool and Thajchayapong quantified three levels of traffic congestion according to the weighted exponential moving averages from GPS data [4]. Pongpaibool, Tangamchit and Noodwong evaluated traffic state from image processing data using fuzzy logic and adaptive neuro-fuzzy techniques [5]. Thianniwet, Phosaard and Pattara proposed a technique that included decision tree and sliding window in order to classify traffic state. Human perceptions were used to determine congestion level and data collected from mobile sensors and web cam [6]. Combinations of traffic flow data such as velocity, occupancy of road segment were processed by the dynamic fuzzy clustering method due to estimate the congestion in another study based on fuzzy approach [7]. But there are almost no fuzzy clustering techniques estimating traffic congestion level except that study in literature. With this aspect, our study will be a new approach for estimating the traffic congestion level utilizing only velocity and traveled road length via FN-DBSCAN method.

This paper is organized as follows: In Section II, detailed information about traffic data is given. In Section III, some background information about FN-DBSCAN is explained. In Section IV, traffic data is analyzed using FN-DBSCAN and its results are explained. In the last section, conclusion of this study is stated.

II. DATA COLLECTION AND PREPARATION

Static and dynamic sensors are major sources of traffic. Static sensors collect the data from many devices (loop detectors, camera, infrared detectors, ultrasonic, radar detectors, etc.) which are located at certain points, especially equal interval, whereas dynamic sensors collect the location, velocity and direction information from equipments that are included GPS receiver. Although static sensors are used for data acquisition in many studies, in consequence of its installation and maintenance is very costly, researchers are interested dynamic sensors as alternative source [8]. Obtained traffic information (location, velocity, direction, weather, etc.) from GPS equipped probe vehicle is sent traffic center via cellular network such as GPRS and 3G, namely, Floating Car

Data (FCD). FCD can be used on-line and off-line regimes. The on-line data can be used to predict instantaneous traffic state and to inform traffic users. The off-line data, usually, are used for modeling traffic congestion, for road categorizing or segmentation. In this study, portable navigation device as dynamic sensor and GPS information from its log as off-line data are utilized to calculate congestion level in road segments.

Data such as date, time, latitude, longitude and velocity obtained by parsed NMEA sentences from portable navigation device is transferred in computer environment via MATLAB. Haversine formula which is compute length between two GPS coordinate is used to calculate traveled road length by probe car [9]:

$$D = 2R\sin^{-1}\left(\left[\sin^2\left(\frac{\phi_1-\phi_2}{2}\right) + \cos\phi_1\cos\phi_2\sin^2\left(\frac{\lambda_1-\lambda_2}{2}\right)\right]^{0.5}\right), \tag{1}$$

where D is the distance (in km) between two points on the earth identified by latitude φ and longitude λ (in radians) and R is the radius of the earth (in km); here, the geometric mean was used, that is, 6367.45 km.

III. FN-DBSCAN Algorithm

There are various clustering methods in the literature. Density-based approaches rely on the density of data points for clustering and have the advantage of generating clusters with arbitrary shapes and good scalability. Among density-based clustering methods, Density Based Spatial Clustering of Applications with Noise (DBSCAN), a well known clustering algorithm is generally used for spatial data analysis [10]. Fuzzy Joint Points (FJP) [11-12], Noise-Robust Fuzzy Joint Points (NRFJP) is variously an alternative to DBSCAN. And, Fuzzy Neighborhood Density Based Spatial Clustering (FN-DBSCAN) [3] algorithm which is the based on fuzzy, is come into use spatial data analysis by researcher [13].

In methods like DBSCAN, in order to determine the core points of clusters or noise points with notion of classical neighborhood analysis, if the number of points in a certain radius is larger than specified threshold. Instead of classical neighborhood thought, FJP like-methods (NRFJP etc.) use fuzzy neighborhood cardinality in order to determine core points [3, 11-12]. FN-DBSCAN integrates the advantages of speed of DBSCAN with a low complexity, and robustness of NRFJP algorithm due to used fuzzy relation in neighborhood analysis [11, 12].

Distance (ε), known as neighborhood radius, and minimum number of points (MinPts) which has to be found in ε -neighborhood are two parameters used for cluster assignment in classical DBSCAN. ε is specified to satisfy the condition $0 \le \varepsilon \le d_{\max}$, where

$$d_{\max} = \max_{x_i, x_j \in X} d(x_i, x_j). \tag{2}$$

where $d(x_i, x_j)$ is the distance between the points x_i and x_j . ε values vary for data set depending on the scale, and for this reason, some problem occurs in adjusting values of ε . Data are normalized to resolve this problem by using the following transformation:

$$x_i = \frac{x_i - x_i^{\max}}{x_i^{\max} - x_i^{\min}}, i = 1,\ldots,n, \tag{3}$$

where $x_i^{\min} = \min_{i=1,\ldots,n} x_i$ and $x_i^{\max} = \max_{i=1,\ldots,n} x_i$. After normalization, condition $0 \le \varepsilon \le 1$ will be satisfy in consequence of $d_{\max} \le 1$ condition satisfies.

In classical DBSCAN algorithm, neighborhood ($N_{x_i}(x_k)$) of the point x_i within given ε radius, is obtained following formula:

$$N_{x_i}(x_k) = \begin{cases} 1, & \text{if } d(x_i, x_k) \le \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

$N_{x_i}(x_k)$ is described with sum of the membership degrees of points to neighborhood set within ε radius in fuzzy situation. The neighborhood membership degrees of the points vary distances from core point, although, there is no difference between points within same neighborhood radius of core point in classical DBSCAN. This case provides the fundamental advantages for fuzzy approach. For example, in Figure 1, x_1 and x_2 which have the same number of neighbors within a given ε radius, accordingly, two points have equally dense with respect to the crisp neighborhood. But, density of cluster in which involved x_2 points is more than other, since x_2 has higher membership degree to be core point than x_1 .

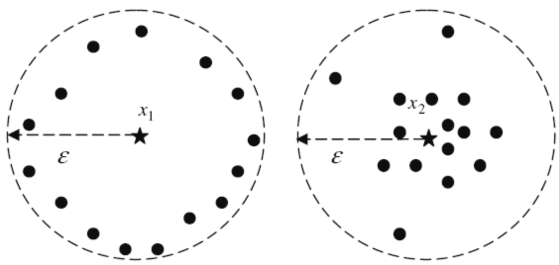

Figure 1. Points are similar according to crisp neighborhood cardinality, but dissimilar according to fuzzy aproach [3].

The main advantage of using FN-DBSCAN algorithm instead of DBSCAN is that various neighborhood membership functions which regularize different neighborhood sensitivities can be utilized [14]. So the FN-DBSCAN method could be more robust to the scale and density variations of the datasets.

Three parameters are required to perform FN-DBSCAN: fuzzy neighborhood membership function, distance parameter (ε) and minimum set cardinality (ε_2). The linear neighborhood membership function used in FN-DBSCAN algorithm in our study is described as:

$$N_x(y) = \begin{cases} 1 - \dfrac{d(x,y)}{d_{\max}}, & \text{if } d(x,y) \le \varepsilon, \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The minimal threshold of neighborhood degrees is determined as ε_1 parameter. If ε parameter fixes the maximal neighborhood radius, ε_1 parameter fixes the minimal neighborhood membership degree has the following relationship with ε [3]:

$$\varepsilon = d_{\max}(1 - \varepsilon_1). \tag{6}$$

The fuzzy neighborhood set of point $x \in X$ with ε_1 parameter is fuzzy set determined as:

$$FN(x;\varepsilon_1) = \{(y, N_x(y)) | N_x(y) \ge \varepsilon_1\}, \tag{7}$$

where $N_x : X \to [0,1]$ is any membership function that determines neighborhood relation between points. Instead of sum of distances between points and ε parameter are used to calculate cardinality as in DBSCAN, sum of fuzzy neighborhood degrees of points to the neighborhood set within ε radius are used to calculate fuzzy set cardinality of FN-DBSCAN algorithm. A point x is called, fuzzy core point with ε_1 and ε_2 parameters using fuzzy set cardinality if

$$cardFN(x;\varepsilon_1) \equiv \sum_{y \in N(x;\varepsilon_1)} N_x(y) \ge \varepsilon_2 , \tag{8}$$

where $N(x;\varepsilon_1)$ is a neighborhood set.

The pseudocode of the FN-DBSCAN algorithm is given below.

FN-DBSCAN algorithm.

Step 1. Set the cluster assignment for all points as unclassified.

Step 2. Specify the memberships threshold parameters ε_1 and ε_2, set t=1.

Step 3. Find an unclassified fuzzy core point p with parameters ε_1 and ε_2.

Step 4. Mark p to be classified. Start a new cluster C_t and assign p to this cluster.

Step 5. Find all the unclassified points in the neighborhood set $N(p;\varepsilon_1)$ and add to seed set created S.

Step 6. Get a point q in S, mark q to be classified, assign q to the cluster C_t, and remove q from S.

Step 7. Check if q is a fuzzy core point with parameters ε_1 and ε_2, add all the unclassified points in the neighborhood set $N(p;\varepsilon_1)$ to the set S.

Step 8. Repeat Steps 6 and 7 until the set is empty.

Step 9. $t = t + 1$ and repeat Steps 4-7 until there is no unclassified fuzzy core point.

Step 10. Mark all unclassified points as noise.

End.

IV. EXPERIMENTAL STUDY AND ITS RESULTS

Third largest city of Turkey, Izmir, which has already one of the main troubles by reason of traffic congestion, was partly examined by selecting route from Bornova to Karşıyaka (Fig. 2). Data collected from probe car that equipped portable navigation device on selected roads which are Ankara, Anadolu and Altınyol Roads in peak hours.

In order to avoid noisy data, route was selected without traffic lights on roads. Moreover, selected route is about 11.4 km. and in contrast with short distance, it is one of the key roles due to connected south to north and northeast to north.

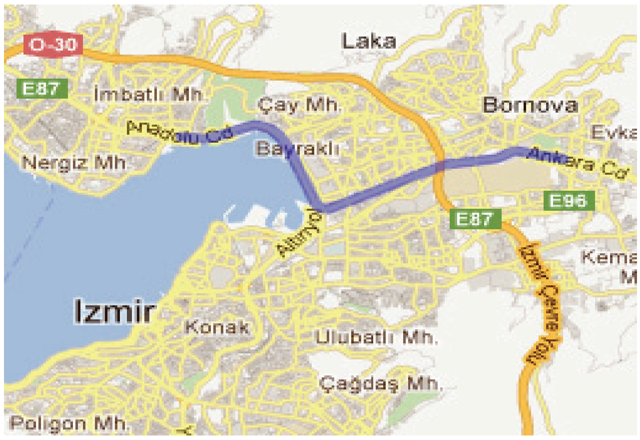

Figure 2. Selected Route for Experimental Study

Speed limit on route is 70 km/hr except crossroad, and in large part of road has three lanes. Moreover there exist some variations by particular road on selected route. Therefore, the probe car was driven under the speed limits of the road. Driving the vehicle according to speed limits caused some confusion. For example, if the driver slows down the car while part of the current road has lower speed limit than previous part of the road, circumstance of road traffic may be considered as congested, although light traffic condition. Therefore, in order to decrease effect of this case, velocity values were normalized to [0,1] interval, considering speed limits of the road using (3). Travelled road lengths are also included to analysis by normalizing.

Travel time is about nine minutes with driving without exceeding speed limit of selected roads if driver obeys the traffic rules. Probe car was driven same lane throughout the journey so as to not constitute variability arising from changing lanes and it arrived end point in sixteen minutes after departure from first point.

ε_1 and ε_2 values were specified as respectively, 0.95 and 0.15. The experimental results for GPS data are given Fig. 3. As it is seen in Fig. 3 traffic data are separated to six clusters. So road is divided segments which have different length for each segment. These segments can be considered as traffic congestion levels, and average of segments represented different line in Fig. 3 According to over ten years experienced driver's impressions, the results of using FN-DBCAN algorithm to collected data from portable navigation device were verified. Also correctness of the clustering results is validated by the expert visually.

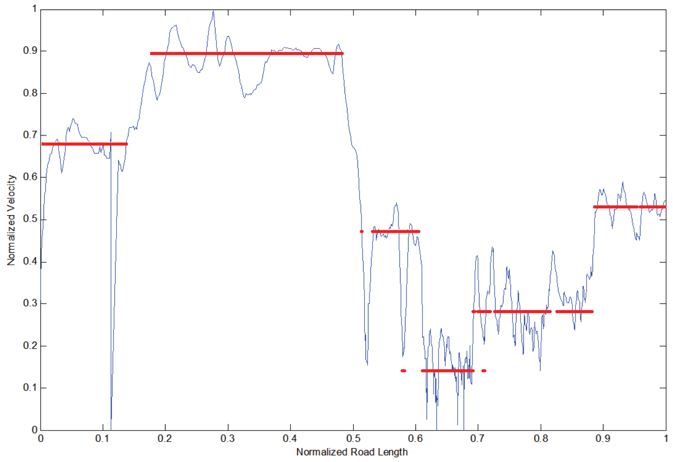

Figure 3. Clustered congestion level

V. Discussion

Traffic congestion increases time loss, energy consumption, air pollution etc., and for this reason, it is important topic in traffic management. Studies about traffic state will provide better manage traffic with decision making for more sustainable transportation system.

We carried out different approach using FN-DBSCAN in order to cluster road segments according to road traffic congestion. Six clusters were observed in Fig. 3, and expert was visually verified correctness of segmentation in view of the fact that velocity of probe car. Although data were used as offline, this paper will provide contribution concept of historical average velocity which is required in many studies to determine traffic congestion. However, FN-DBSCAN algorithm may be modified by regarding historical average velocity to analyze for online traffic data.

On the other hand, there was no traffic light on selected road due to eliminate effect of traffic lights in this paper. And probe car was driven about sixteen minutes. We consider that our future work about traffic congestion will be concentrated on noisy and large data set from the roads which have traffic lights.

References

[1] L. Wei-Hsun, T. Shian-Shyong and S. Wern-Yarng, "Collobrative real-time traffic information generation and sharing framework for the intelligent transportation systems", Information Sciences, vol. 180, 2010, pp 62-70.

[2] W. Andreas, W. Shangbo, G. H. Bruck and J. Peter, "Traffic congestion estimation service exploiting mobile assisted positioning schemes in GSM networks", Procedia Earth and Planetary Science, vol. 1 (1), 2009, pp. 1385–1392.

[3] E. N. Nasibov and G. Ulutagay, "Comparative clustering analysis of bispectral index series of brain activity", Expert Systems with Applications, vol 37 (3), 2010, pp. 2495-2504.

[4] W. Pattara-atikom, P. Pongpaibool and S. Thajchayapong, "Estimating road traffic congestion using vehicle velocity", Proc. 6th International Conference on ITS Telecommunications, 2006, pp. 1001-1004.

[5] P. Pongpaibool, P. Tangamchit, K. Noodwong, "Evaluation of road traffic congestion using fuzzy techniques," Proc. of IEEE TENCON, 2007, Taiwan, pp. 1-4.

[6] T. Thianniwet, S. Phosaard and W. Pattara-Atikom, "Classification of road traffic congestion levels from GPS data using a decision tree algortihm and sliding window", Proc. of the World Congress on Engineering, vol. 1, London, 2009, pp. 105-109.

[7] Z. Yin and W. Junli, "A Study on urban traffic congestion dynamic predict method based on advanced fuzzy clustering model", International Conference on Computational Intelligence and Security, 2008, pp. 96-100.

[8] D. Fagan and R. Meier, "Highly-dynamic, pervasive monitoring of traffic congestion levels", Proceedings of ITRN2010, University College Dublin, Ireland, 2010.

[9] R. W. Sinnott, "Virtues of the Haversine", Sky Telescope, vol. 68 (2), 1984, pp.159.

[10] M. Ester, H. P. Kriegel, J. Sander and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." In Proceedings of the second international conference on knowledge discovery and data mining, 1996, pp. 226–231.

[11] E. N. Nasibov, "An alternative fuzzy-hierarchical approach to cluster analysis", Proc. Seventh Internat. Conf. on Application of Fuzzy Systems and Soft Computing, Siegen, Germany, 2006, pp. 113–123.

[12] E. N. Nasibov, G. Ulutagay, "On the fuzzy joint points method for fuzzy clustering problem", Automatic Control and Computer Sciences, vol. 40 (5), 2006, pp. 33–44.

[13] J. K. Parker and J. A. Downs, "Footprint generation using fuzzy-neighborhood clustering", GeoInformatica, vol. 16, 2012, 1-15.

[14] E. N. Nasibov and G. Ulutagay, "Robustness of density-based clustering methods with various neighborhood relation", Fuzzy Sets and System, vol. 160, 2009, pp. 3601-3615.