



AFFINITY PROPAGATION CLUSTERING IN DEFINING LEVEL OF SERVICE CRITERIA OF URBAN STREETS

Prasanta Kumar Bhuyan, Smruti Sourava Mohapatra

Dept of Civil Engineering, National Institute of Technology Rourkela, India

Submitted 16 March 2012; resubmitted 26 July 2012, 9 December 2012; accepted 8 March 2013

Abstract. Defining Level Of Service (LOS) criteria is very important as this is the first step of LOS analysis but this is not well defined in Indian context. The analysis followed in India is basically adaptation of Highway Capacity Manual (HCM 2000) methodology which is more suitable for developed countries having homogenous traffic pattern. An attempt has been made in this study to define LOS criteria of urban streets for developing countries having heterogeneous traffic flow condition. Defining LOS is basically a classification problem and to solve it Affinity Propagation (AP), a very recently developed cluster algorithm is used. Inventory details and the required speed data are collected from five major street corridors of Greater Mumbai Region in India through the application of *Trimble GeoXT* Global Positioning System (GPS) receiver. Six validation parameters are used on Free Flow Speed (FFS) data to find the optimal number of clusters, which is required for the classification of street segments into number of classes. After that speed data collected during both peak and off-peak hours are averaged over street segments and clustered into six groups to get the speed ranges of different LOS categories. Using validation parameters, considering the physical and surrounding environmental characteristics it is found that street segments can be classified into four classes in Indian context as mentioned in Highway Capacity Manual 2000. However, the FFS range for urban street class IV (urban design category) is significantly lower because of varying road geometric characteristics. The speed ranges of LOS categories under urban street classes are proportionately lower to that values mentioned in HCM 2000 because of highly heterogeneous traffic flow on urban Indian roads. The travel speed data collection procedure using GPS is simple and accurate. In addition, AP clustering is highly efficient in terms of time saving and provides a very accurate solution to classification problems. Hence, both GPS and AP techniques can be applied in other countries to define the speed ranges of LOS categories considering the local conditions.

Keywords: urban street; level of service; travel speed; GPS, affinity propagation clustering; cluster validation.

Introduction

Level Of Service (LOS) concept germinated from the concept of 'practical capacity' presented in the 1950 Highway Capacity Manual (HCM 1950). In the HCM (1950) LOS was defined in terms of speed and travel time, traffic interruptions, freedom to manoeuvre, safety, driving comfort and convenience, and operating cost. In 1985 Highway Capacity Manual (HCM 1985) the concept of LOS was made more realistic by classifying LOS into six classes from 'A' to 'F', based on the combination of travel time and the ratio of traffic flow rate to the capacity, because travel time was recognized as a dominant factor of the service quality. In course of time the definition of LOS went into an evolutionary process and that is followed now is the definition stated in 2000 Highway Capacity Manual (HCM 2000). LOS in the HCM (2000) is defined as 'a quality measure describing operational

conditions within a traffic stream, generally in terms of service measures such as speed and travel time, freedom to manoeuvre, traffic interruptions, comfort and convenience'. The HCM (2000) also designates six levels of service for each type of facility, from 'A' to 'F', with LOS 'A' representing the best operating conditions and LOS 'F' the worst.

Traffic on urban Indian roads is highly heterogeneous in nature having different operational characteristics. In current scenario no suitable methodology is available for the determination of LOS of heterogeneous traffic conditions. Determining LOS for urban street is very much important as this is the first step of LOS analysis procedure. This affects the planning, design, and operational aspects of transportation projects as well as the allocation of limited financial resources among competing transportation projects. Therefore, there is a bare



need of determining the LOS criteria in Indian context.

For heterogeneous traffic condition in India, Marwah and Singh (2000) have classified LOS into four groups (I–IV). Similarly, Maitra *et al.* (1999) taking congestion as measure of effectiveness for prevailing heterogeneous traffic condition in India divided LOS into nine groups 'A' to 'I'. Baumgaertner (1996) pointed out that today motorists became more adapted to urban congestion so the traffic condition, which was viewed as intolerable 1960, now considered normal. Kita and Fujiwara (1995) stated LOS not to be a traffic operating condition but tried to find the relationship of LOS with driver's perception. Spring (1999) found service quality to be a continuous and subjective matter so the author opined it to be inappropriate to have distinct boundary or threshold value for a particular LOS. Flannery *et al.* (2005) suggested LOS does not completely represent a drivers' assessment on performance of urban streets. The author opined for inclusion of qualitative measures for defining LOS. Clark (2008) from a study upon the prevailing traffic condition of New Zealand suggested for a new LOS category to be termed as 'F+' or 'G'.

Shao and Sun (2010) categorized LOS into two parts: (1) level of facility supply and (2) level of traffic operation. Travel speed to free flow speed (FFS) ratio was considered as an evaluation index of traffic operation. Traffic operation categorized into different groups using Fuzzy set. Determination of LOS of urban street from user perception was carried-out by Flannery *et al.* (2008). Fang *et al.* (2011) determined speed-flow curves of different segments of an interchange by developing a simulation model using VISSIM software. LOS ranges from the speed-flow curves were determined by taking density as classification index. Arasan and Vedagiri (2010) through computer simulation studied the effect of a dedicated bus lane on the LOS of heterogeneous traffic condition prevailing in India. A state-of-the-art hybrid algorithm was developed by Čavar *et al.* (2011) to classify urban roads based on vehicle track and infrastructural data collected through Global Positioning System (GPS). From the study the authors found the limitations of traditional clustering algorithm in classifying the large amount of speed data. Not going with traditional research in which traffic flow is considered as the only parameter to access the LOS of traffic facilities, Tan *et al.* (2007) analysed the pedestrian LOS with physical facilities and traffic flow operation along with user perception. Limitations of LOS criteria of walkways proposed by HCM (2000) for China were found out by Cao *et al.* (2009). User perception taken into consideration for classification of LOS at urban rail transit passages and found the limit for LOS standards suitable for China is lower than that suggested by HCM (2000). Body size, culture, gender and age of user were found to be influencing factors for the LOS classification.

Defining LOS criteria is basically a classification problem and cluster analysis is a suitable technique that can be applied for the solution of it. For this cluster analysis large amount of FFS and Average travel speed data are required because according to HCM (2000)

LOS of urban street is a function of travel speed along street segments. Turner *et al.* (1998) found from their study floating car method to be very much susceptible to human error. With improvement computers, Distance Measuring Instrument (DMI) came as the solution of Floating car method. Benz and Ogden (1996) found limitation in this method relating to installation of the DMI unit and data storage problems. Recent research has demonstrated the feasibility of using GPS and Geographical Information System (GIS) technologies for automated inventory and travel time data collection. The new automated procedures provide consistency, finer level of resolution and better accuracy in measuring travel time and travel speed than traditional techniques. In addition, this automated data collection technique helps in collecting, updating and managing large inventory data. Gallagher (1996) described the opportunity of cost-effective travel time data collection by the use of GPS equipment.

Frey and Dueck (2007) used Affinity Propagation (AP) algorithm to cluster images of faces and genes in microarray data. The authors found AP to cluster data with much lower error than other methods, and it did the clustering in less than one-hundredth the amount of time. Conroy *et al.* (2010) developed a semi-supervised AP algorithm for face-image clustering and Functional Magnetic Resonance Imaging (fMRI) volumetric pixel clustering. Xia *et al.* (2008) presents two variants of AP for grouping large-scale data with a dense similarity matrix. The local approach was Partition Affinity Propagation (PAP) and the global method was Landmark Affinity Propagation (LAP). Refianti *et al.* (2012) compared the accuracy and effectiveness of AP and *k*-means algorithm. The authors have found that AP to be more effective than *k*-means by implementing these algorithms on the relationship between the two variables i.e. Grade Point Average (GPA) and duration of Bachelor Thesis completion of Gunadarma University. Zhang and Zhuang (2008) presented a modified AP algorithm called voting-PAP, which is a method of clustering using evidence accumulation. Yang *et al.* (2010) used AP for classifying the large amount of remote sensing image data quite accurately. The authors found the algorithm to be very much efficient for clustering of data for which training data is not available. Yang *et al.* (2010) used this newly developed AP clustering algorithm in traffic engineering. The authors have proposed a model-based temporal association scheme and novel pre-processing and post-processing operations which together with AP make a quite successful method for vehicle detection and on road traffic surveillance. Zhang *et al.* (2012) proposed an instant traffic-clustering algorithm using AP to find points on road having similar traffic pattern. The authors found the algorithm to be suitable in predicting the traffic pattern and for finding the influence of traffic pattern at one point to that at another point. These are some of researches carried-out at different locations under different traffic conditions, which give a strong background for further research carried-out in this study in defining LOS criteria in Indian context.

The data set that was used in this study was obtained from 10–12 travel runs taken on five major urban corridors in the Mumbai City. The total length of these corridors is about 140 km. These corridors, on the whole, were divided into 100 street segments. Two urban street corridors of Kolkata City having varying geometric and surrounding environmental characteristics were also taken into considerations to find the applicability of this study for other cities of India. Out of two corridors considered, one corridor was Airport to Joka and the other corridor was Airport to Ulberia. These two corridors are approximately 80 km length; comprised of 50 street segments. The interesting fact on selecting these two cities for this study is that traffic composition and road geometric characteristics along with functionality brings the true variation that was required for this purpose.

All the speed data were collected by using *Trimble Geo XT* GPS receiver, which was fitted to a mid-sized car. After acquiring these speed data a very newly developed AP algorithm developed by Frey and Dueck (2007) was used twice; first to cluster FFS to get the speed ranges of different urban street classes and second to cluster average travel speed data during peak and off peak hour in both directions to get the speed ranges of different LOS categories. AP is known as a message-passing algorithm, where each data point is grouped by sending messages to all other data points in order to inform the relative attractiveness between each other. This newly developed algorithm has been used by various researchers for efficient clustering of data with the minimal error in very short time.

1. Affinity Propagation and Cluster Validation Measures

In this section a brief overview about the AP algorithm is discussed. To know the optimal number of clusters to classify the FFS six validation parameters, which were used in this research, are elaborated.

1.1. Affinity Propagation

AP is a theoretic clustering method recently developed by Frey and Dueck (2007). This algorithm simultaneously considers all of data points as possible exemplar (center point) where each message is sent to reflect the latest interest, which is owned by each data point to be able to select other data points as their exemplar. Through this message-passing process, an algorithm tests the possibility of all data points to become the center of cluster (exemplar). So that, every data point has the same opportunity to become an exemplar.

Fig. 1. shows the steps that are performed during AP clustering. Procedurally, the algorithm operates on three matrices: a similarity (s) matrix, a responsibility (r) matrix, and availability (a) matrix. Results are contained in a criterion (c) matrix. These matrices are iteratively updated by four equations, where i and k refer, respectively, to the rows and columns of the associated matrix.

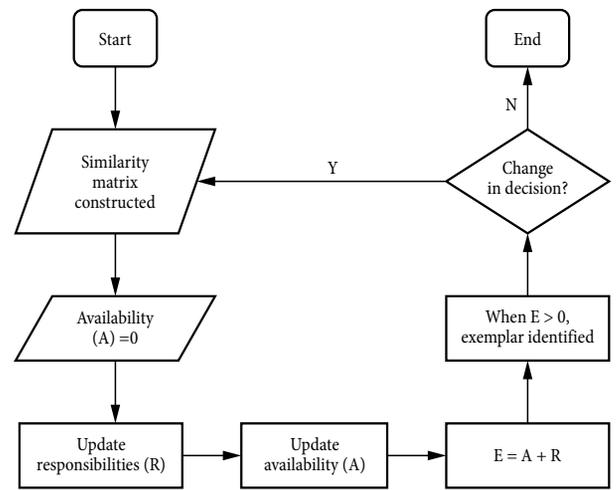


Fig. 1. Flowchart of AP Clustering

The algorithm is processed in the following seven steps to find out the exemplar, which is the cluster centre.

Steps:

1. Inputting similarity matrix $s(i, k)$: the similarity of point i to point k .
2. Initializing the availabilities $a(i, k)$ to zero: $a(i, k) = 0$.
3. Updating all responsibilities $r(i, k)$:

$$r(i, k) \leftarrow s(i, k) - \max_{k' \neq k} \{a(i, k') + s(i, k')\}. \quad (1)$$

4. Updating all availabilities $a(i, k)$:

$$a(i, k) \leftarrow \min \left\{ 0, r(k, k) + \sum_{i': i' \notin \{i, k\}} \max \{ 0, r(i', k) \} \right\}, \quad \text{for } k \neq i. \quad (2)$$

5. Adding the availabilities and responsibilities matrix to monitor the exemplar decisions. For a particular data point i ; $a(i, k) + r(i, k) > 0$ for identification exemplars.
6. If decisions made in step 3 did not change for a certain time of iteration or a fixed number of iteration reaches, going to step 5. If otherwise, going to step 1.
7. Assigning other data points to the exemplars using the nearest assign rule that is to assign each data point to an exemplar, which it is most similar to.

1.2. Validation Measure

Cluster validity is concerned with checking the quality of clustering results. It has been mainly used to evaluate and compare whole partitions, resulting from different algorithms or resulting from the same algorithms under different parameters. A common application of the cluster validation measure is to determine the correct number of cluster for a set of data (Bensaid *et al.* 1996). Different validity measures have been proposed in the literature, none of them is perfect by oneself, and therefore several indices are used in this study, such as: C-Index (CI), Weighted inter-intra Index (WI), Hartigan Index (HI), R-Squared Index (RSI), Krzanowski-Lai Index (KLI) and Davies-Bouldin Index (DBI).

1.2.1. C-Index

CI is an index, which can be used to get optimal number of clusters from a particular set of data. CI is as follows (Hubert, Schultz 1976):

$$CI = \frac{D - D_{min}}{D_{max} - D_{min}}, \tag{3}$$

where: D, D_{min}, D_{max} are calculated as follows. Let p be the number of pairs that can be formed taking data points belonging to a single cluster. D is the sum of the distances between all these pairs of data points in a particular cluster. P is the number of possible pairs can be formed from the data set. These P numbers of pairs that can be ordered according to their distances and p pairs with largest and smallest distances between samples can be selected. D_{min} is the sum of the smallest distance of p and D_{max} is the sum of the largest distance of p . When pairs of samples with small distances are in same cluster, the index will be small so it corresponds to optimal number of clusters.

1.2.2. Weighted Inter-Intra Index

WI tries to find the optimal number of clusters with high overall quality of cluster $\phi^{(Q)}$ for a small number of clusters c . This index can be utilized to get the right number of clusters for a particular data set. The objective here is to maximize intra-cluster similarity and minimize inter-cluster similarity. The inter-cluster and intra-cluster similarities are given bellow (Strehl, Ghosh 2003):

$$inter(X, \lambda, i, j) = \frac{1}{n_i \cdot n_j} \cdot \sum_{\lambda_a=i, \lambda_b=j} S(X_a, X_b) \tag{4}$$

and

$$intra(X, \lambda, j) = \frac{2}{(n_i - 1) \cdot n_i} \cdot \sum_{\lambda_a=\lambda_b=i, b>a} S(X_a, X_b). \tag{5}$$

S is the similarity matrix. The inter cluster similarity is undefined for singleton cluster. The quality measure $\phi^{(Q)}$ is found out by the ratio of weighted average inter-cluster to weighted average intra-cluster similarity:

$$\phi^{(Q)}(X, \lambda) = 1 - \frac{\sum_{i=1}^c \frac{n_i}{n - n_i} \sum_{j \in \{1, \dots, i-1, i+1, \dots, c\}} n_j \cdot inter(X, \lambda, i, j)}{\sum_{i=1}^n n_i \cdot intra(X, \lambda, i)}, \tag{6}$$

where: i and j are cluster indices; n is the number of data points in the input matrix; X is the input matrix; X_a and X_b are two vertices; λ is a vector containing the output.

$\phi^{(Q)}$ can have highest value 1 and lowest value 0. The target function to achieve high quality $\phi^{(Q)}$ with lowest c is $\phi^{(T)} \in [0, 1]$ which is:

$$\phi^{(T)}(c) = 1 - \frac{2c}{n} \phi^{(Q)}(c). \tag{7}$$

Strehl and Ghosh (2003) stated that the highest value of Weighted Inter-Intra Index (WI) gives the optimal number of clusters for a given data set.

1.2.3. Hartigan Index

HI is applied to get the optimal number of clusters for a set of data. The index is calculated as follows (Hartigan 1975):

$$HI = (n - c - 1) \frac{err(c) - err(c + 1)}{err(c + 1)}, \tag{8}$$

where $err(c)$ describes the fitness of clustering result:

$$err(c) = \sum_{i=1}^c \sum_{a=1, a \in i}^n d^2(X_a, C_i).$$

In this formulation c is number of clusters and $c \geq 1$; n is the number of data points; d is the distance between data sample X_a and the centre C_i ; i, j are indices. The highest lowest value indicates the optimal number of clusters for the input data. Bolshakova and Azuaje (2006), Dudoit and Fridlyand (2002) found that the lowest value of CI and HI signifies the optimal number of clusters for a particular set of data.

1.2.4. R-Squared Index

RSI is defined as follows (Sharma 1995):

$$RSI = \frac{SS_t - SS_w}{SS_t}, \tag{9}$$

where:

$$SS_t = \sum_{j=1}^m \sum_{k=1}^{n_j} (X_A - \bar{x}_j)^2;$$

$$SS_w = \sum_{i=1 \dots c} \sum_{j=1 \dots m}^{n_{ij}} (X_A - \bar{x}_j),$$

where: c is number of clusters; m is the number of dimensions of data; n_j is the number of data values of j dimension; n_{ij} corresponds to number of data values of j dimension that belongs to cluster i ; \bar{x}_j is mean data value of j dimension; X_A chosen any arbitrary data point.

Halkidi *et al.* (2002a, 2002b) affirmed that from RI the optimal number of clusters can be found from the Index vs. Number of cluster graph. The author proposed that the optimal number of clusters is that point for which the parametric value on the graph goes downward.

1.2.5. Krzanowski-Lai Index

KLI is defined as (Krzanowski, Lai 1988):

$$KLI(c) = \left| \frac{DIFF(c)}{DIFF(c + 1)} \right|, \tag{10}$$

where:

$$DIFF(c) = (c - 1)^{\frac{2}{p}} W(c - 1) - c^{\frac{2}{p}} W(c)$$

and p denotes the number of features in the data set; c denotes the number of clusters and $c \geq 2$; $W(c)$ denotes the within cluster sum of square of the partition. Dudoit and Fridlyand (2002) suggested that the highest value of KLI gives the optimal number of clusters for a given data set.

1.2.6. Davies-Bouldin Index

DBI is a function of the ratio of the sum of within-cluster scatter to between-cluster separation. This DBI index is defined by (Davies, Bouldin 1979):

$$DBI = \frac{1}{c} \sum_{i=1}^c \max_{i \neq j} \left\{ \frac{D_c(i) + D_c(j)}{d_{ce}(i,j)} \right\}, \quad (11)$$

where:

$$D_c = \frac{\sum_i \|X_a - C_i\|}{N_i};$$

$$d_{ce} = \|C_i - C_j\|,$$

where: D_c , d_{ce} are intra-cluster and inter-cluster distances respectively. The intra cluster distances are calculated by average of pairwise distances from points in the cluster to the cluster centroid. The inter-cluster distance between two clusters is computed as the distance between their centroids. X_a any chosen arbitrary data that belongs to cluster i . N_i is number of data that belong to cluster i . C_i and C_j cluster centre of i and j cluster. So, when the cluster is compact and far from each other the ratio is small. Consequently, DBI is small for good cluster. Dimitriadou *et al.* (2002) found that optimal number of clusters for a data set is the number of clusters for which the DBI value is the lowest.

2. Study Corridor and Data Collection Techniques

This section divided into two parts. The first part briefly describes the study corridor from where the speed data as well as road inventory data were collected. The second part elaborates upon the details of data collection technique adopted for this study.

2.1. Study Corridor

For this study, the commercial capital of India Mumbai City was considered. The city has a linear pattern of transport network having predominant North–South commuter movements. South Mumbai houses various work places so during morning time people move towards South for work and during evening hours, they return towards North to their homes in the Suburbs of Mumbai. So In this study five major corridors were chosen out of which four are North–South corridors and one is east-west corridor. The North–South corridors are Eastern express highway extends up to South (Corridor-1), LBS Road extending up to South via Ambedkar road (Corridor-2), Western express highway extending up to marine drive (Corridor-3), SV road extending up to South via Veer Savarkar road (Corridor-4) and the only East–West corridor is Versova–Andheri–Ghatkopar–Vashi (VAGV) (Corridor-5). These five corridors are overlapped on the GIS base map of Greater Mumbai are shown in Fig. 2. Two urban street corridor of Kolkata City, which is in the eastern part of India, was studied to find the applicability of this research in other Indian cities. These two corridors studied for this research are Airpot to Joka and Airpot to Ulberia.

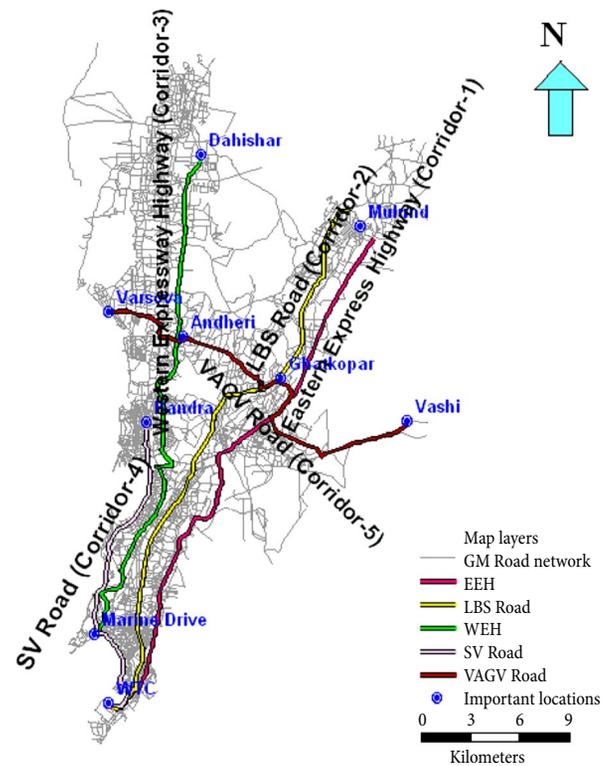


Fig. 2. Map showing selected corridors of greater Mumbai

2.2. Data Collection

Trimble Geo-XT GPS receiver, which logs data continuously in a time interval of one second, was fitted to a mid-sized vehicle typically a car. The data acquired by GPS provides both spatial and temporal references from which various traffic parameters were derived, including travel time and travel speeds. In order to get unbiased data sets three mid-sized vehicles were used and help of different drivers on different days of the survey work was taken. This exercise was executed because operational efficiency of one mid-sized vehicle varies from other; similarly driving skills of a driver differs from other.

Three types of data sets were collected. The first type is roadway inventory details. In this survey road inventory details on segments like segment number, number of lanes, median types, pedestrian activity, roadside development, access density, construction activity, speed limit, separate right turn lane, number of flyovers, date and day of data collection and segment length were collected. During the collection of inventory details proper segmentation technique was applied, which is the directional stretch of road section immediately after signalized intersections to the location point immediately after the next signal. These road inventory data are used to find the coherence between the clustering results in classifying urban street segment into four classes and the speed ranges of different LOS categories. This analysis is elaborated in section 3.

The second type of survey was to find the FFS. Before finding the FFS data a detailed 24 hours traffic count survey conducted at 45 stations on 7 screen lines to know the duration in a day when traffic volume is less than or equal to 200 veh/ln/hr. It was found that

free flow traffic condition (less than 200 veh/ln/hr) is approaching at 12 mid-night and all road sections are having free flow traffic conditions from 1 AM to 5 AM. Hence FFS for all these corridors were collected during these hours. On these FFS data AP clustering was applied to get the FFS ranges of urban street classes. Values of Cluster validation parameters were calculated using these FFS data set and also analysed to get the optimal number of clusters for the input data set. The third type of data that was collected during the survey was congested travel speed. Number of trips covered for each direction of travel and for the study hours (peak, off-peak and free-flow) on each corridor is at least 3 and sometimes it is up to six trips. These average travel speed on road segments data were clustered by the same AP clustering algorithm to get the speed ranges of different LOS categories of four different urban street classes. Both these two types of speed data that were collected on study Corridor-3 are shown in Table 1. This particular Urban street corridor consists of 15 street segments. FFS of each segment, which was collected during 1 AM to 5 AM, is given in column 2 of the Table 1. Travel speed data that were collected during peak and off-peak hours and averaged over the street segments under Corridor-3 is shown in columns 3–6.

Relevance of the current study was further strengthened by considering the traffic of Kolkata City, which is in the eastern part of India. Two corridors having varying geometric and surrounding environmental characteristics were taken into considerations i.e. one corridor was Airport to Joka and the other corridor was Airport to Ulberia. These two corridors are approximately 80 km length; comprised of 50 street segments. The true variation required for the purpose was brought by studying the traffic composition and road geometric characteristics of the two cities.

3. Results and Analysis

In this research six validation parameters as discussed earlier were used. For determination of the parametric value of validation measures, FFSs data acquired through GPS receivers, which are averaged over street segment, and their cluster centres found from AP cluster analysis is used as input to the validation parameters. Values of validation parameters computed from 2 to 7 numbers of clusters. The resulted values are plotted in Fig. 3. These six validation parameters are used to know the optimum number of cluster for FFS data set. Optimal number of clusters is basically concern with the quality of cluster obtained by applying a particular clustering algorithm to a particular data set. Every algorithm has its natural way of classification of the data set into number of groups. When a data set is clustered into its optimal number of clusters the quality of the cluster is best as the variation between the data points belonging to a particular cluster is minimal.

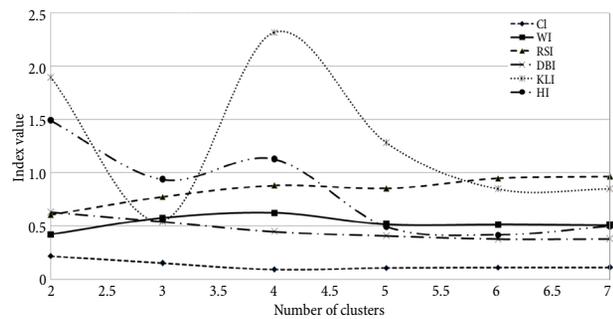


Fig. 3. Validation measures for optimal number of clusters using AP clustering

Table 1. FFS and average travel speed during peak and off peak hours of Corridor-3

Segment No.	Average FFS [km/hr]	Duration and direction of travel			
		M–N–S	M–S–N	E–N–S	E–S–N
Average travel speed [km/hr]					
1	71.14	28.64	28.63	30.5	37.33
2	61.56	19.24	30.08	24.51	27.28
3	65.91	40.41	45.19	44.94	40.43
4	43.19	38.17	47.06	41.68	37.29
5	77.36	37.04	56.11	50.39	44.61
6	57.92	16.88	27.14	17.35	38.27
7	58.83	28.74	32.13	20.68	24.65
8	44.27	20.17	27.01	25.47	47.95
9	54.27	27.39	33.21	27.93	20.15
10	54.39	15.50	30.10	12.87	15.86
11	56.68	21.80	19.22	44.39	16.63
12	43.00	22.50	25.99	26.39	22.21
13	24.94	25.91	22.84	11.31	8.85
14	31.38	13.21	31.49	31.50	23.66
15	29.62	17.67	15.27	24.40	8.25

Notes: M–N–S – Morning–North–South; M–S–N – Morning–South–North; E–N–S – Evening–North–South; E–S–N – Evening–South–North.

By knowing the optimum number of cluster the urban street segments can be classified into that number of Urban street classes. It is always considered that lesser number of clusters is preferred as optimal cluster if variation in parametric values between two consecutive numbers of clusters is minimal. Fig. 3 shows Index value vs Number of clusters for CI, WI, RSI, DBI, KLI and HI. It can be seen that CI is lowest for 4 numbers of clusters and beyond 4 number of cluster the variation in parametric values is very minimal so 4 is taken as the optimal number of clusters for CI. Further, it can be clarified that when the FFS data are classified into four groups, the pairs of FFS data with small distances (difference in speed values) falls under each group and hence sum of these differences in values for the four groups are minimal. From weighted inter-intra index vs. Number of cluster graph, it is clearly visible that, the index is having highest value for 4 number of cluster. So 4 is chosen as the optimal number of clusters going according to the literature. For four number of cluster the similarity in FFS values within a group is maximized and the similarity among four groups is minimized. RSI graph goes downward beyond four clusters giving the optimal number of clusters as 4. In this case when FFS data clustered into 4 numbers of clusters sum of the square of FFS values of a particular cluster is maximum. From graph, it can be observed that KLI is highest for 4 numbers of clusters so 4 is chosen as the optimal number of clusters. When FFS values are clustered into 4 numbers of groups the difference between FFS values of all clusters to their respective cluster centre is minimal. For DBI optimal number of clusters is found to be 6 because according to literature number of cluster for which index value is lowest is chosen as optimal number of clusters. So if the FFS data set will be clustered into 6 numbers of groups the cluster centroid of each cluster will be far from each other and all the FFS values coming under a particular cluster will have the least distance from the cluster centroid. Parametric value of HI is scaled down by a factor 100 and resulted value vs Number of clusters graph is illustrated in Fig.3. HI value is lowest for 5 numbers of clusters. 5 can be chosen as optimal number of clusters as literature states that number of cluster giving the lowest HI value is taken as optimal number of clusters. This index calculates the total square error of all the clusters and determines the index value for 2 to 7 number of

clusters. When the FFS data are clustered into 5 numbers of clusters total square errors is found to be minimal.

In simple words, these parameters determine the quality of cluster obtained by classifying the FFS data into number of groups. It is found that when the input data set is clustered into optimal number of clusters the quality of cluster is best because street segments clustered into a particular group has the least difference in FFS and also street segments clustered into different cluster possess the maximum difference in FFS. Out of six validation parameters considered in this study, four parameters give the optimal cluster value as 4 which is also same as suggested by HCM (2000). That is the reason for which in this research the urban street segments were classified into four Classes by using the AP algorithm.

Hundred urban street segments of five urban street corridors were analysed in this research. So to get the FFS ranges of different urban street class FFS of these 100 urban street segments were clustered using AP. AP which is known for a very flexible and accurate clustering technique is utilized for this purpose. Speed data are given to the algorithm in the form of a similarity matrix. As from validation parameter analysis it was found the optimal number of clusters to be 4, number of exemplar chosen is four for clustering o FFS. The distribution of data after AP clustering is shown in Fig. 4. Different symbols represent the class of the urban street. Four different colours used for four different urban street classes. From Fig. 4. It can be seen most of the street segments belong to urban street class-I and urban street class-II. Very few street segments belong to urban street class-I.

AP clustering is used for the second time after classification of urban street segments into four different urban street classes. For four classes of urban street classes LOS speed ranges were defined using AP clustering. In this clustering direction, wise average travel speed data of both peak and off peak hours were given as input to the clustering algorithm.

In Fig. 5 the clustering result are compiled. The legends in Fig. 5a–d gives the speed ranges for the six LOS categories obtained by using AP clustering. The speed ranges for each individual LOS categories elaborated in Table 2. The speed ranges of LOS categories (A–F) expressed in percentage of free-flow speeds were found to be approximately 80, 62, 48, 35, 23 and 18–23 respec-

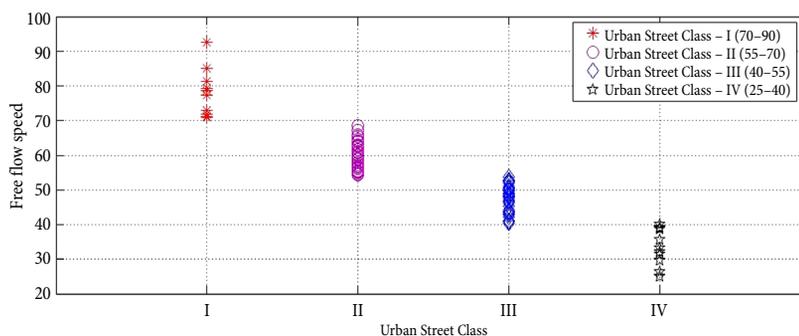


Fig. 4. AP clustering of FFS for urban street classification

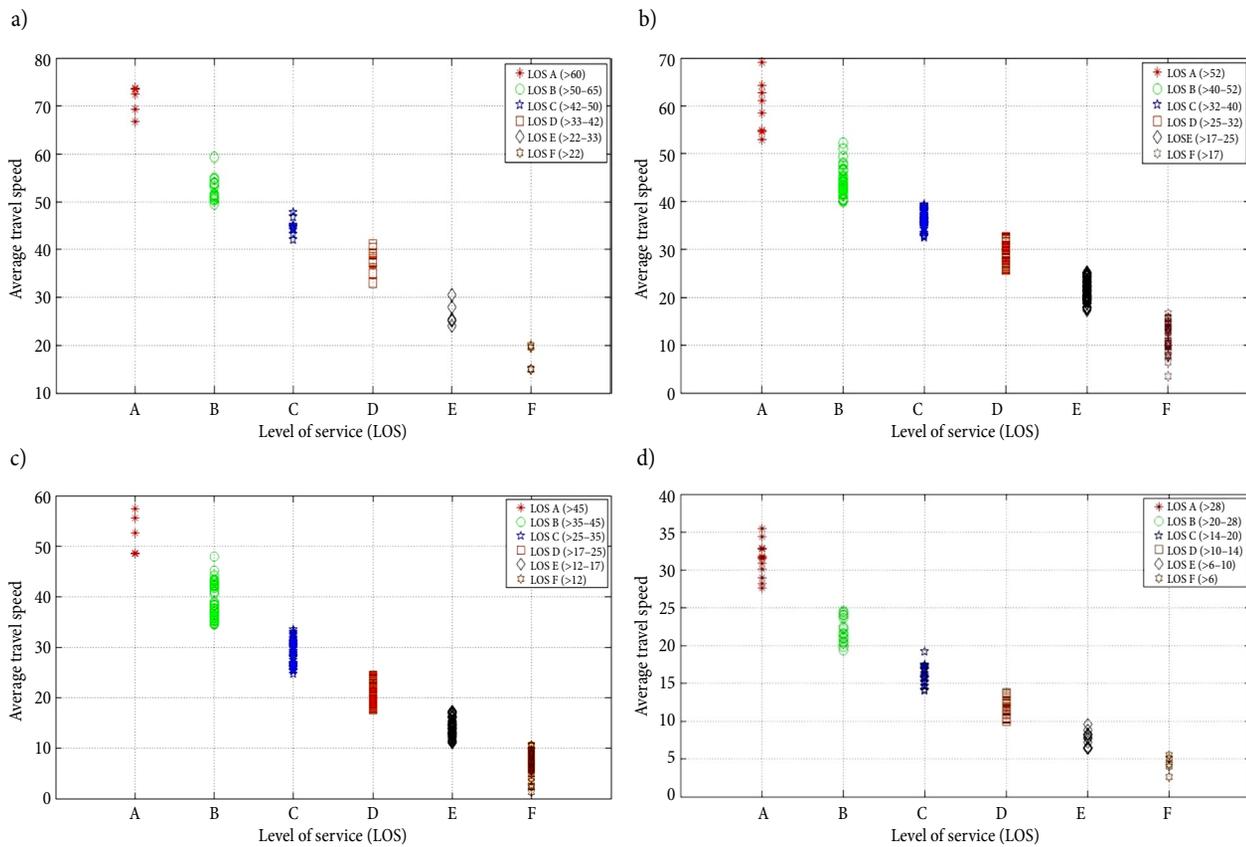


Fig. 5. LOS of urban street classes (I-IV) using AP clustering on average travel speed: a – LOS of urban street class-I; b – LOS of urban street class-II; c – LOS of urban street class-III; d – LOS of urban street class-IV

tively in the present study. Whereas, in HCM (2000) it has been mentioned these values are 90, 70, 50, 40, 33 and 25–33 percent respectively.

The coherence of the clustering result in the classification of urban streets into four classes and speed values into six LOS categories is verified with the road inventory data that was collected during the survey. The road inventory data considered here are the number of lanes, median type, access density, roadside development, on street parking and pedestrian activity. These geometric characteristics and surrounding environmental characteristics provide sufficient information on physical char-

acteristics of each road segment. Out of the five urban street corridors, physical characteristics of street segments under urban street Corridor-3 are shown in the Table 3 for information.

Table 3 contains various geometric and surrounding environmental characteristics of 15 street segments of Urban street Corridor-3 that were acquired during inventory survey and also the different LOS experienced on these street segments. Geometric and surrounding environmental characteristics were checked to know whether the experienced LOS on a particular road segment during a travel run actually satisfies to the field condition or not. From this analysis it is inferred that on a segment having good traffic flow, geometric and surrounding environmental conditions good LOS ‘A’, ‘B’, ‘C’, can be experienced by the commuter. The commuter have to experience poor LOS like ‘D’, ‘E’ and ‘F’ when traffic flow condition is poor and surrounding environmental conditions are not so good. So travel speed the parameter used to determine LOS to be quite capable of defining LOS categories for a particular urban street segment.

In order to check the application of this LOS criteria; data collected from Kolkata City were tested. FFS and average travel speed during both peak and off-peak hours on each of segments on both corridors were calculated. The street segments were classified into four classes based on free-flow speed, geometric and surrounding environmental characteristics. In addition, levels of ser-

Table 2. Urban street speed ranges for different LOS proposed in Indian conditions by AP clustering method

Urban street class:	I	II	III	IV
Range of free-flow speed (FFS) [km/hr]:	from 90 to 70	from 70 to 55	from 55 to 40	from 40 to 25
Typical FFS [km/hr]:	75	60	47	35
LOS	Average travel speed [km/hr]			
A	>60	>52	>45	>28
B	>50–60	>40–52	>35–45	>20–28
C	>42–50	>32–40	>25–35	>14–20
D	>33–42	>25–32	>17–25	>10–14
E	>22–33	>17–25	>12–17	>6–10
F	>22	>17	>12	>6

Table 3. LOS of urban street classes in different physical and traffic characteristics on Corridor-3

Location of street segments		FFS [km/hr]	Geometric characteristics			Surrounding environmental characteristics			Average travel speeds and LOS								
Segment No.	Urban street class		Segment length [km]	Lanes	Median type	Access density	Road side development	Parking	Pedestrian activity	M–N–S LOS	M–S–N LOS	E–N–S LOS	E–S–N LOS	LOS			
1	I	71.1	2.2	6-LD	barrier	low	medium density	significant	some	28.6	E	28.6	E	30.5	E	37.3	D
2	II	61.5	1.8	6-LD	barrier	low	medium density	some	some	19.2	E	30.1	D	24.5	E	27.3	D
3	II	65.9	6.4	6-LD	barrier	low	low density	no	little	40.4	B	45.2	B	44.9	C	40.4	B
4	III	43.2	4.3	6-LD	fence	medium	medium density	some	some	38.2	B	47.1	A	41.7	B	37.3	B
5	I	77.3	4.6	6-LD	barrier	low	low density	no	some	37.0	D	56.1	B	50.4	B	44.6	C
6	II	57.9	0.8	6-LD	plantation	medium	low density	no	some	16.9	F	27.1	D	17.4	E	38.3	C
7	II	58.8	0.99	6-LD	plantation	medium	low density	no	some	28.7	D	32.1	B	20.7	E	24.7	D
8	III	44.2	0.38	4-LD	mountable	medium	high density	significant	usually	20.2	D	27.0	C	25.5	C	48.0	A
9	II	54.2	8.15	4-LD	mountable	medium	low density	some	little	27.4	D	33.2	C	27.9	D	20.2	E
10	II	54.4	0.9	4-LD	plantation	medium	low density	some	little	15.5	F	30.1	D	12.9	F	15.9	F
11	II	56.7	1.4	3-LU	no	medium	low density	some	little	21.8	E	19.2	E	44.4	B	16.6	F
12	III	43.0	0.79	2-LU	no	high	high density	significant	usually	22.5	D	26.0	C	26.4	C	22.2	D
13	IV	24.9	0.37	2-LU	no	high	high density	significant	usually	25.9	B	22.8	B	11.3	D	8.9	D
14	IV	31.4	0.90	2-LU	no	high	high density	significant	usually	13.2	D	31.5	A	31.5	A	23.7	B
15	IV	29.6	0.95	2-LU	no	high	high density	significant	usually	17.7	C	15.3	C	24.4	A	8.3	E

Notes: LD – lane divided; LU – lane undivided; M–N–S – Morning–North–South; M–S–N – Morning–South–North; E–N–S – Evening–North–South; E–S–N – Evening–South–North.

vice provided by the street segments during peak and off peak hours were estimated using Table 2. The percentage of travel runs under different levels of service categories found in urban street classes in Kolkata City during the survey period are shown in Table 4.

Table 4. Percentage of travel runs under levels of service categories of urban street classes in Kolkata City

LOS	Urban street class			
	I	II	III	IV
A	52.78	0.00	5.88	40.00
B	22.22	22.22	9.80	24.44
C	13.89	16.67	19.61	15.56
D	2.78	27.78	31.37	2.22
E	2.78	27.78	15.69	8.89
F	5.56	5.56	17.65	8.89

Conclusions

In order to define the speed limits of LOS categories of urban streets of developing countries especially for India having heterogeneous traffic flow a very recently developed and effective technique such as AP clustering algorithm is used in this study.

An in-depth research is carried-out by applying six cluster validation parameters on FFSs and considering the geometric and surrounding environmental characteristics of street segments it is found that urban street segments can be classified into four classes (I–IV); the same numbers of street classes are also mentioned in HCM (2000).

Applying AP clustering on FFSs data in Indian context it is found that FFS ranges of urban street classes are proportionately lower than that mentioned in HCM (2000). Heterogeneous traffic flow and roads having varying geometric and surrounding environmental characteristics are the major reasons for these lower values in FFSs. The speed range of each particular LOS category is found to be significantly lower than that mentioned in HCM (2000). Speed ranges of different LOS categories in terms of percentage FFS are also found to be lower than that mentioned in HCM (2000).

For LOS 'A' speed range was found to be 80 percentage of FFS which is 90 percentage according to HCM (2000), because unwanted movement of pedestrians along and across the road sections produces undesirable side friction that constrain travelers to reduce vehicular speed. Roadside vendors and on-street parking occupy a substantial portion of road sections which compels the commuter to move through a narrower space is also can be inferred to be one of the reason.

From the AP clustering of FFS data it can be seen a less number of roads in Mumbai are of high-speed design (street class-I) or highly congested (street class-IV). More number of road segments is of suburban (street class-II) or intermediate (street class-III) type. It can be suggested that Greater Mumbai region needs substan-

tial geometric improvements to mitigate the burden on urban road infrastructure because of ever growing vehicular traffic volume.

From this study, the applicability of GPS in collection of speed data with high precision in short time is established. So this tool can be exploited by other developing and developed countries to collect speed data and cluster analysis can be applied to define the speed ranges of LOS categories of their own rather than following some values, which are not completely appropriate for the local condition.

Hence, this study result has strong application in transportation planning, operation and design. Rather using mid-sized vehicles for data collection purpose this study can be extended to collect data using different types of vehicles and further study can be carried-out.

References

- Arasan, V. T.; Vedagiri, P. 2010. Study of the impact of exclusive bus lane under highly heterogeneous traffic condition, *Public Transport* 2(1–2): 135–155. <http://dx.doi.org/10.1007/s12469-010-0021-x>
- Baumgaertner, W. E. 1996. Level of service: Getting ready for the 21st century, *ITE Journal* 66(1): 36–39.
- Bensaid, A. M.; Hall, L. O.; Bezdek, J. C.; Clarke, L. P.; Silbiger, M. L.; Arrington, J. A.; Murtagh, R. F. 1996. Validity-guided (re)clustering with applications to image segmentation, *IEEE Transactions on Fuzzy Systems* 4(2): 112–123. <http://dx.doi.org/10.1109/91.493905>
- Benz, R. J.; Ogden, M. A. 1996. Development and benefits of computer-aided travel time data collection, *Transportation Research Record* 1551: 1–7. <http://dx.doi.org/10.3141/1551-01>
- Bolshakova, N.; Azuaje, F. 2006. Estimating the number of clusters in DNA microarray data, *Methods of Information in Medicine* 45(2): 153–157.
- Cao, S.; Yuan, Z.; Zhang, C.; Zhao, L. 2009. LOS classification for urban rail transit passages based on passenger perceptions, *Journal of Transportation Systems Engineering and Information Technology* 9(2): 99–104. [http://dx.doi.org/10.1016/S1570-6672\(08\)60059-8](http://dx.doi.org/10.1016/S1570-6672(08)60059-8)
- Čavar, I.; Kavran, Z.; Petrović, M. 2011. Hybrid approach for urban roads classification based on GPS tracks and road subsegments data, *Promet – Traffic & Transportation* 23(4): 289–296. <http://dx.doi.org/10.7307/ptt.v23i4.131>
- Clark, I. 2008. Level of service F: is it really bad as it gets?, in *Proceedings of the IPENZ Transportation Group Conference*, 2–5 November 2008, Puke Ariki, New Plymouth, New Zealand, 198–203.
- Conroy, B.; Xi, Y. T.; Ramadge, P. 2010. A supervisory approach to semi-supervised clustering, in *2010 IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, 14–19 March 2010, Dallas, TX, 1858–1861. <http://dx.doi.org/10.1109/ICASSP.2010.5495368>
- Davies, D. L.; Bouldin, D. W. 1979. A cluster separation measure, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1(2): 224–227. <http://dx.doi.org/10.1109/TPAMI.1979.4766909>
- Dimitriadou, E.; Dolničar, S.; Weingessel, A. 2002. An examination of indexes for determining the number of clusters in binary data sets, *Psychometrika* 67(1): 137–159. <http://dx.doi.org/10.1007/BF02294713>

- Dudoit, S.; Fridlyand, J. 2002. A prediction-based resampling method for estimating the number of clusters in a dataset, *Genome Biology* 3(7): 1–21. <http://dx.doi.org/10.1186/gb-2002-3-7-research0036>
- Fang, J.; Han, Y.; Zhou, R.-G. 2011. The capacity of urban expressway interchange, in *ICCTP 2011: Towards Sustainable Transportation Systems*, 2208–2216. [http://dx.doi.org/10.1061/41186\(421\)220](http://dx.doi.org/10.1061/41186(421)220)
- Flannery, A.; Roupail, N.; Reinke, D. 2008. Analysis and modeling of automobile users' perceptions of quality of service on urban streets, *Transportation Research Record* 2071: 26–34. <http://dx.doi.org/10.3141/2071-04>
- Flannery, A.; Wochinger, K.; Martin, A. 2005. Driver assessment of service quality on urban streets, *Transportation Research Record* 1920: 25–31. <http://dx.doi.org/10.3141/1920-03>
- Frey, B. J.; Dueck, D. 2007. Clustering by passing messages between data points, *Science* 315(5814): 972–976. <http://dx.doi.org/10.1126/science.1136800>
- Gallagher, J. 1996. Travel time data collection using GPS, in *Proceedings of the National Traffic Data Acquisition Conference*, 5–9 May 1996, New Mexico, 147–162.
- Hartigan, J. A. 1975. *Clustering Algorithms*. John Wiley & Sons Inc. 366 p.
- Highway Capacity Manual*. 1950. Transportation Research Board.
- Highway Capacity Manual*. 1985. Transportation Research Board.
- Highway Capacity Manual*. 2000. Transportation Research Board. 1134 p.
- Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. 2002a. Cluster validity methods: part I, *SIGMOD Record* 31(2): 40–45. <http://dx.doi.org/10.1145/565117.565124>
- Halkidi, M.; Batistakis, Y.; Vazirgiannis, M. 2002b. Cluster validity methods: part II, *SIGMOD Record* 31(3): 19–27. <http://dx.doi.org/10.1145/601858.601862>
- Hubert, L.; Schultz, J. 1976. Quadratic assignment as a general data analysis strategy, *British Journal of Mathematical and Statistical Psychology* 29(2): 190–241. <http://dx.doi.org/10.1111/j.2044-8317.1976.tb00714.x>
- Kita, H.; Fujiwara, E. 1995. Reconsideration on the level of service and a proposed measure, in *Proceedings of the 15th Annual Meeting of Japan Society of Traffic Engineers (JSTE)*, Japan, 25–28.
- Krzanowski, W. J.; Lai, Y. T. 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering, *Biometrics* 44(1): 23–34. <http://dx.doi.org/10.2307/2531893>
- Maitra, B.; Sikdar, P.; Dhingra, S. 1999. Modeling congestion on urban roads and assessing level of service, *Journal of Transportation Engineering* 125(6): 508–514. [http://dx.doi.org/10.1061/\(ASCE\)0733-947X\(1999\)125:6\(508\)](http://dx.doi.org/10.1061/(ASCE)0733-947X(1999)125:6(508))
- Marwah, B. R.; Singh, B. 2000. Level of service classification for urban heterogeneous traffic: a case study of Kanapur metropolis, in *Proceedings of the Fourth International Symposium on Highway Capacity*, 27 June – 1 July 2000 Maui, Hawaii, 271–286. Available from Internet: <http://onlinepubs.trb.org/onlinepubs/circulars/ec018/ec018toc.pdf>
- Refianti, R.; Mutiara, A. B.; Juarna, A.; Ikhsan, S. N. 2012. Analysis and implementation of algorithm clustering affinity propagation and k-means at data student based on GPA and duration of bachelor-thesis completion, *Journal of Theoretical and Applied Information Technology* 35(1): 69–76.
- Shao, M.; Sun, L. 2010. United Evaluation model of traffic operation level for different types of urban road, *Journal of Tongji University* 38(11): 1593–1598.
- Sharma, S. 1995. *Applied Multivariate Techniques*. Wiley. 512 p.
- Spring, G. S. 1999. Integration of safety and the Highway Capacity Manual, in *Proceedings of the Fourth International Symposium on Highway Capacity*, 27 June – 1 July 2000 Maui, Hawaii, 63–72. Available from Internet: <http://onlinepubs.trb.org/onlinepubs/circulars/ec018/ec018toc.pdf>
- Strehl, A.; Ghosh, J. 2003. Relationship-based clustering and visualization for high-dimensional data mining, *INFORMS Journal on Computing* 15(2): 208–230. <http://dx.doi.org/10.1287/ijoc.15.2.208.14448>
- Tan, D.; Wang, W.; Lu, J.; Bian, Y. 2007. Research on methods of assessing pedestrian level of service for sidewalk, *Journal of Transportation Systems Engineering and Information Technology* 7(5): 74–79. [http://dx.doi.org/10.1016/S1570-6672\(07\)60041-5](http://dx.doi.org/10.1016/S1570-6672(07)60041-5)
- Turner, S. M.; Eisele, W. L.; Benz, R. J.; Holdener, D. J. 1998. *Travel Time Data Collection Handbook*. Texas Transportation Institute. 341 p. Available from Internet: <https://www.fhwa.dot.gov/ohim/start.pdf>
- Xia, D.-Y.; Wu, F.; Zhang, X.-Q.; Zhuang, Y.-T. 2008. Local and global approaches of affinity propagation clustering for large scale data, *Journal of Zhejiang University SCIENCE A* 9(10): 1373–1381. <http://dx.doi.org/10.1631/jzus.A0720058>
- Yang, C.; Bruzzone, L.; Sun, F.; Lu, L.; Guan, R.; Liang, Y. 2010. A fuzzy-statistics-based affinity propagation technique for clustering in multispectral images, *IEEE Transactions on Geoscience and Remote Sensing* 48(6): 2647–2659. <http://dx.doi.org/10.1109/TGRS.2010.2040035>
- Zhang, X.; Wu, F.; Zhuang, Y. 2008. Clustering by evidence accumulation on affinity propagation, in *19th International Conference on Pattern Recognition (ICPR 2008)*, 8–11 December 2008, Tampa, FL, 1–4. <http://dx.doi.org/10.1109/ICPR.2008.4761213>
- Zhang, B.; Xing, K.; Cheng, X.; Huang, L.; Bie, R. 2012. Traffic clustering and online traffic prediction in vehicle networks: a social influence perspective, in *2012 Proceedings IEEE IN-FOCOM*, 25–30 March 2012, Orlando, Florida, USA, 495–503. <http://dx.doi.org/10.1109/INFCOM.2012.6195790>