UDC: 656.022.9

DOI: http://dx.doi.org/10.7708/ijtte2022.12(3).06

DETERMINATION OF POTENTIAL FEEDER BUS STOPS FROM SMART CARD DATA AND CAPACITATED CLUSTERING ALGORITHM

H S Abdulrahman[1]

[1] *Civil Engineering Department, Federal University of Technology, Minna, PMB 65 Niger State, Nigeria*

Received 19 July 2022; accepted 22 August 2022

Abstract: Most of the literature that deals with feeder bus route design simply assumes the feeder bus stops are already established. The Feeder bus stops are established to connect some known origins to a known destination unlike the conventional bus stop location. In this study smart card data together with geo location of existing bus stops are used to obtain the origin (feeder bus stops) and destination (train station). The first step is to identify the conventional bus stops by isolating smart card users who use these stops to a particular destination. These conventional stops are numerous and hence a demand-constrained clustering method was developed to convert these stops into feeder bus stops. This method was then applied to data extracted from Izmir's database management system. This methodology was compared to well-known clustering algorithms and it performed reasonably well. This study can serve as a foundation in the application of smart data in the planning of feeder bus routes.

Keywords: feeder bus stops; smart card data, K means, K medoids, capacitated clustering algorithms.

1. Literature Review

Generally, bus stops are planned based on traditional surveys which are used to decipher the traveling patterns. Although they are quite useful, some of their lacking stems from; time-consuming and also capital intensive, they may also be biased because usually, passengers don't like giving information, especially regarding their privacy and social status. Due to these reasons and more, transit agencies carry out these surveys incessantly which may lead to poor planning of the entire system because these systems depend on quality data. The advancement in information communication technologies at the turn of the century has made it possible for the application of big data sources in the management, planning, and design of modern public transportation systems. Examples of big data sources especially relevant to the transportation sector include smart cards, database management systems, GPS and GIS systems, and have been applied in one form or the other to support the planning as well as operation of mass transportation systems Zhang *et al.* (2018).

The bulk of the work on this topic deals with conventional bus stop locations and not feeder bus stop location. Bachok *et al.* (2013) used GPS and GIS to identify and determine the location of passengers boarding and

[1] Corresponding author: abdul.hassan@futminna.edu.ng

alighting positions as well as conventional bus stops. Jahani *et al.* (2013). Developed a novel multi-objective bus stop location model for the Central Business District, Tehran, Iran. The candidate stops were clustered based on the walking distance by using K mean algorithms and their results laid emphasis on the importance of the weighted combination of candidate stops to find suitable stops. Chen *et al.* (2018), based their study on access coverage and mobility effects of bus stop location, therefore they used two objectives for the optimization bus stop location in a network. The optimal bus stop locations are selected from given bus stops along given bus routes. This gives an insight into planning bus stops from existing bus stop positions as they reduced the number of existing bus stops after optimization but this may not be very useful when planning a new bus stop from the beginning. Li and Chen (2016) semi-supervised clustering algorithm (Fuzzy C-mean) was used for bus stop location optimization, similarly Leksakul *et al.* (2017) determine stop locations by testing four different clustering algorithms on real data from large scale industrial factory in Thailand. The K-means with Maximin provided the best solution, as it minimized the number of bus stop locations. Perugia *et al.* (2011) propose a cluster routing approach to model both bus stop locations It uses a cluster routing representation to simultaneously determine bus stop locations and bus routing in urban street networks. Taplin and Sun (2020), in their study, a genetic algorithm was used to evaluate potential stops based on demand maximizing locations, the novelty of their work focuses on starting bus modeling from the stop unlike other studies because they are of the opinion that bus stops should be determined in terms of demand Shatnawi *et al*. (2020). This study is for conventional bus stops. It applied GIS and some metaheuristics optimization techniques to model the location of bus stops by removing redundant bus stops which are located at irregular distances thereby finding optimal travel time and serviceability of the stops. They also used existing bus stops therefore the bus stops were not designed from scratch. These studies focused on how to determine the location of conventional bus stops by using mostly optimization methods, clustering algorithms such as K-means, Fuzzy C-means, etc. without necessarily giving an idea of how the data was sourced but only mentioned the data used such as geolocation of existing stops, demand, and others like walking distance and so on.

Few studies have concentrated on the stop location of a community shuttle or feeder bus service. Most of the literature that deals with feeder bus routes simply assumes the feeder bus stops are already established and they mainly focused on the optimization of the feeder bus routes as a consequence Shatnawi *et al.* (2020). Liu *et al.* (2014) used large-scale real-world taxi-based GPS data to plan a feeder bus stop for the airport. In their study, the mined origin and destination dataset consequently use a clustering technique to identify the feeder bus stops. Their main goal was to improve the accessibility of the feeder bus stops at the airport. Chen *et al.* (2018) research work is mainly concerned with the bus routing problem of a suburban bus route for airport access, thus, their method establishes stop location and route generation simultaneously. Because its focus is not on stop location per se, the number of required stops is also given, therefore not all stops are covered in the bus routes. Xiong *et al.* (2013) used a community shuttle route as an input for the heuristic algorithm proposed for positioning stops. The concept of minimizing passengers' walking distance was used such that each stop is matched with a corresponding zone only. Similarly, Cui *et al.* (2015) in trying to improve the efficiency of bus operations used a multi-objective optimization to determine the location of bus stops. Locations based on walking distance, the delay time of non-transit vehicles, transit traveling time adjacent to the intersections. This study was mainly concerned with the positioning of the bus stops as it connects to the subway station. Ren *et al* (2018) also focus on minimizing transfer time between two public transport modes so as obtain optimal bus stop location. Prah *et al.* (2018) and Dragan *et al.* (2011), unlike the conventional bus stop, this study solves the optimal bus allocation for school routing problems. The main algorithm is based on data clustering by using three-dimensional GIS data which accounts for the terrain. They also used Monte Carlo simulation together with the aforementioned data clustering to optimize bus stop allocation making sure the total walking distance does not exceed limits. Wei *et al.* (2021), optimizes feeder bus stops locations and assigned passengers to them by minimizing the total walking cost and the feeder bus stop locations were used to design feeder bus routes. In practice, feeder bus routes are constructed by adjusting existing conventional bus routes even though passenger demand for conventional bus stops and feeder bus stops differ. According to Ge *et al.* (2011), a survey carried out in Beijing, China shows that some buses that go through a railway station are always fully loaded when they arrive and therefore can accommodate passengers transferred by the railway station. Therefore, the feeder bus route should be planned based on feeder bus passenger demand at the stops. It is clear from these studies that a huge amount of data is available and this type of data can be found in big data sources such as smart card data, GPS data, and GIS data which can be used in the planning of feeder bus stop locations, etc. Also, most studies used clustering algorithms based on the geolocation of the stops without considering the demand and capacity of the clustered stops. Hence, the goal of the study is to use smart card data and capacitated clustering algorithms in the determination of feeder bus stop location.

2. Materials and Methods

This study focuses on the application of data sources (the geolocation of bus stops and train stations, passenger demand at bus stops) to estimate potential feeder bus stops. The remainder of this write-up will discuss, the case study, demand estimation at the conventional bus stops, estimation of potential feeder bus stops using clustering algorithms, and capacitated clustering. The number of bus stops is steadily increasing which are serviced by 337 routes with 1170 buses.

2.1. Case Study of Izmir Data Base Management System

Izmir a city in Turkey has an urban population of about 4,061,074; (Stat, 2014). It has one of the most developed and well-integrated public transit systems in the country because it can boast of quite a number of different and well-integrated modes. In Izmir, the urban mass transportation network is comprised of the public bus system, three rail systems (rapid transit, commuter rail Service, light rail service), and ferry service which operate in the inner bay. The electronic fare collection system in Izmir – is now known as Kent Kart (City Card in Turkish) The system employs contactless smart cards that are presented to a validator positioned alongside the driver in the bus entry passage. This device can read information from the smart card, such as identify card type, fare type, bus route and direction, boarding time, date, and in the case of suburban train records fare reduction because the suburban train fare system is based on distance, not time, and store the transaction data for transfer to the system server when needed. Primarily, this system is a fare collection system. Fig. 1 below shows the methodology adopted for feeder bus stop location.

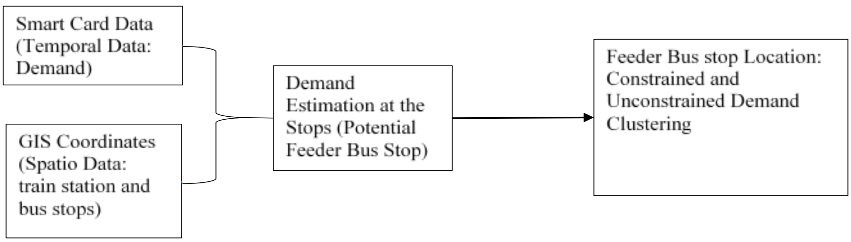

Fig. 1.
Methodology

2.2. Demand Estimation

This exercise is rigorous and requires a lot of computer software to prepare the smart data into useable data. It includes cleaning and coding the smart card data since not all collected data are relevant and there are also missing data. In this study, the 1,967,955 boarding data of Tuesday, the 8th of November 2018 in the smart card system of the Directorate General for Electricity, Water, Coal Gas, Buses and Trolleybuses (ESHOT) of İzmir Metropolitan Municipality was used. This is a combined usage of transit smart cards across all modes of transportation, taking into account both normal and transfer recordings. This information together with GPS data from the Vehicle Location system can be used in transportation planning (Deri and Kalpakci, 2014). Therefore, coordinates of the bus stop and train station were extracted and transfer time was calculated for each smart card data. In Izmir, the fares are charged based on time and it is less than or equal to 90 minutes. Therefore, if the transfer time is less than 90 minutes, then we isolate all bus stops that use that IZBAN station and no any other modes. The Flow chart for the algorithm is shown in Fig. 2.

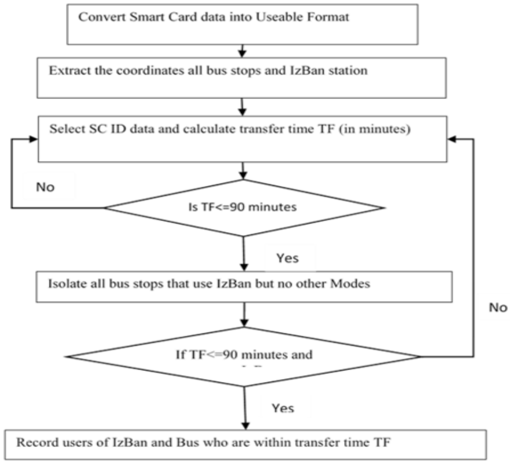

Fig. 2.
Algorithm for Demand at Origin Estimation for a known Destination

2.3. Feeder Bus Stop Location: Clustering Algorithms

Two popular data partition algorithms (K means and K medoids) are used. Both algorithms cluster data into groups by minimizing the distance between coordinate points and the centroid of the clusters. While K medoids select the centroid from one of the coordinate points as a centroid, K means selects the centroid from the average between the points in the cluster. Also, K medoid also uses dissimilarity measures which reduces the sum pairwise dissimilarity instead of euclidean distances used by K means. Importantly, it is assumed that the “k” number of clusters is known or evaluated by other methods or certain criteria. Therefore, for the purpose of this study, the procedure for determining the feeder bus stop location through clustering restricted by the capacity of the ensuing cluster is proposed. The procedure is stated below;

- Calculate the number of clusters; It is calculated based on the demand (di) at the stops and expected capacity of the cluster(C), k = sum of all demands at the stops / Expected Capacity of the Cluster;
- Select initial centroids: the initial k centroids are selected by arranging the stops based on their demand in their non-increasing order d1 > d2 > d3 > dn. Then the first k stops becoming k centroids;
- Select first k stops as initial centroids;
- Assign the stops to cluster:
  i. the Euclidean distances between each requester to all the k centroids are calculated. Group all the stops to the closest centroid.
- Centroid Calculation:
  i. If the capacity of the centroid is not exceeded, the initial k centroid is adopted;
  ii. If the capacity of a centroid is exceeded, that centroid is split into the required number to satisfy the cluster constraint.

3. Results and Discussion

3.1. Demand Estimation

In total 321 stops used the train station for the period for this particular date as shown in Fig. 3. Not all stations have significant demand therefore the stops were chosen based on the card usage at the various stops. The selected stops were limited to those stops with at least 3 passengers per hour thereby reducing the stops to 87 stops with a total demand of 633 passengers per hour as shown in Fig. 3. Feeder bus services differ from conventional bus services in that they are mostly servicing suburban areas with high commuter volumes. This implies most passengers gather in one location to proceed to a somewhat common destination i.e. bus stops may be crammed. Therefore, the need to locate suitable positions for siting feeder bus stops and it is also important to take into consideration the capacity of the stops themselves in terms of passenger volumes.

Fig. 3.
(A) Stops that are Associated with a Train Station and (B) Significant Stops associated with the Train Station

3.2. Feeder Bus Stop Location: Clustering Algorithms

The clustering considered, has 87 existing bus stops whose demands are known and are distributed in (x, y) coordinates. The K medoid algorithm tries to establish the suitable cluster centroid by using existing nodes (bus stop) unlike the Kmeans algorithm which creates a new centroid. Using the K medoid allows to use of existing bus stop locations for the siting of the feeder bus stop location, however, because the demand at every bus stop exists, the new point might be cramped based on demand which might be un desirable for users. Table 1 shows the results of the clustering algorithm with the demand and number of clusters in the brackets. It can be seen that some stops have quite a number of stops close together and thus resulting in cramped clusters with high demands. These stops regardless of which algorithm is used have high passenger demand if we assume the capacity of each potential bus stop is 50 passengers per hour. Because both K mean and K medoid algorithms are based on strictly distance they isolated stops that are far from others see Figure 4.

In this case, since the initial centroids are determined based on their demands and making their point of interest, we cluster the remainder of the nodes around these points of interest by finding the distance for all remaining nodes to the selected initial centroids. After the shortest distance to all centroids is assigned, a group of clusters formed around the initial centroids, and Table 1 ensued. But we have a constraint that no cluster demand shall exceed 50 passengers per hour. Hence, any cell whose demand is exceeded must be split into more centroids to accommodate the excessive demand. Centroids 1,2 4 and 13 are split further to accommodate more demand and the final feeder bus nodes respective demands are shown in Table 2 below.

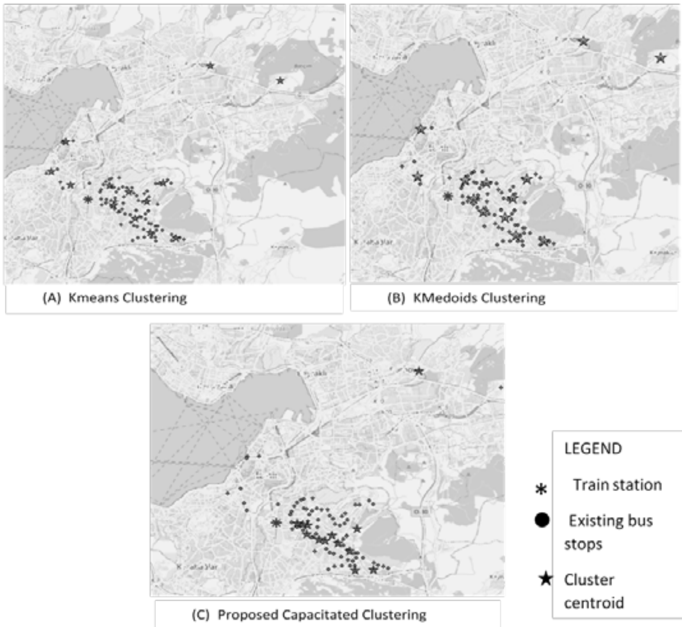

Fig. 4.
Clustering Algorithm Results (A) Kmeans, (B) Kmedoids, and (C) Proposed Capacitated Clustering

Table 1

Passenger Demand/Hour at the Centroids (Number of Stops)

Cluster	Kmeans	Kmedoids	Capacitated
C1	152(12)	10(1)	67(6)
C2	10(1)	38(5)	73(8)
C3	19(3)	16(1)	42(3)
C4	38(5)	18(4)	63(11)
C5	48(11)	73(7)	47(9)
C6	127(15)	121(15)	48(7)
C7	8(2)	19(3)	26(2)
C8	121(15)	140(12)	21(3)
C9	42(9)	75(8)	44(9)
C10	17(4)	37(8)	34(4)
C11	23(6)	31(5)	38(9)
C12	16(1)	17(4)	32(4)
C13	12(3)	39(8)	97(18)

3.3. Discussion of Results

To analyze the results, we evaluate the clustering results by summarizing the clustering by a quality score and also based on the purpose of the clustering carried out. The quality scores used are calculated by understanding the cohesion of how near the data points in a cluster are to the cluster centroid (Intra cluster distance) and also the distance between the centroids of different clusters (inter-cluster distance). It is well desired a good cluster should maximize inter-cluster distance and minimize intra-distance cluster. These distances are then used to calculate the Dunn index which is basically a ratio of inter-distance (separation) and intra-distance (compactness), and the Davies-Bouldin index which is the ratio intra distance (compactness) to inter-distance (separation).

Table 2

Average Distance from Stops to Centroids (intra-cluster Analysis)

Cluster	Kmeans	Kmedoids	Capacitated
C1	2931	0	649
C2	0	264	708
C3	3926	0	576
C4	3851	550	782
C5	1811	329	400
C6	1686	519	408
C7	4500	222	4077
C8	2384	447	841
C9	1877	328	1143
C10	3186	432	463
C11	1017	439	460
C12	0	350	283
C13	3146	404	1989

The clustering algorithm that produces a collection of clusters with the smallest Davies–Bouldin index is considered the best algorithm based on this criterion, while the algorithms that produce clusters with a high Dunn index are more desirable. Fig. 5 shows the cluster analysis and the proposed algorithm has the lowest inter-cluster distance while the K means has the largest distance while for the Intra distance Kmedoids algorithm has the lowest value. This may be largely due to the fact that the proposed algorithm pays attention to locating the centroids based on demand and it is expected that passengers that use the train station are closer to the station. To go further, see Fig.6 below. Based on both Indexes, the K medoids Algorithms performed better than our proposed method, even though our proposed method performed better than the K means algorithm. These analyses do not really show how useful these algorithms especially in this specific problem of estimating the location of feeder bus stops. While The K means algorithm locates the feeder bus stop generating a mean centroid for all the clusters, the K medoids use the existing centroid and try to find which centroid is more at the center of the cluster. Interesting enough our main focus is to develop feeder bus stops by locating important centers of demand and grouping them together. Therefore, the purpose of achieving this with 13 clusters based on the expected capacity of each centroid is shown in the table below. For all the algorithms some stops exceeded the estimated capacity. This is expected as K means and K medoids are mainly dependent on distance and even in our proposed method, the number of stops has to be expanded by dividing any stops that exceeded the capacity into the desired number of stops to accommodate the capacity. In all our method can be very useful in planning bus stop locations especially by considering how close they are but also how important demand points are, how close they are and the capacity of the final clusters see Fig. 7 for final bus stop location with their passenger demand. It is important to note that when evaluating clustering algorithms to emphasize the purpose to which the clustering is done (which is highly subjective) even though the statistics used to compare our algorithms to k means and k medoids can relate relevant information with regards to whether the clustering is bad or not.

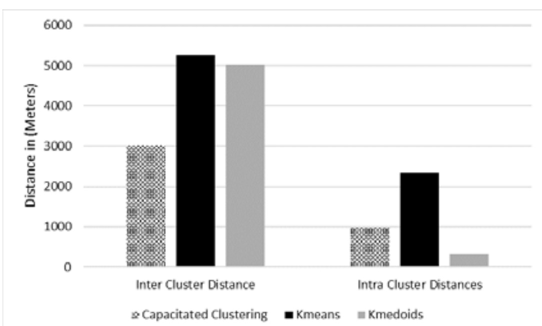

Fig. 5.
Inter/Intra Cluster Distance

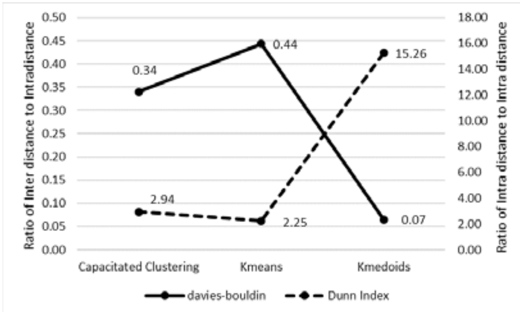

Fig. 6.
Clustering Analysis (Index)

Clusters	Demand per
1	37
1b	30
2a	37
2b	36
3	42
4a	21
4b	42
5	47
6	48
7	26
8	21
9	44
10	34
11	38
12	32
13a	48
13b	49

Fig. 7.
Final Feeder Bus Stop Location and their Demands

4. Conclusions

In conclusion, we proposed a methodology for planning the location of feeder bus stops from existing conventional bus stops taking into consideration the passenger demand going towards a specified destination (train station). The first part involves the determination of potential feeder bus stops by identifying the location of smart card users who are destined to a particular train station through an algorithm developed in this study. The second part equally follows a set of steps that allows for potential stops to be clustered based on the specified capacity of the potential feeder bus stops. The results of our capacitated clustering procedure were compared with famous clustering algorithms K Means and K Medoids algorithms. The procedure focuses mainly on high-demand stations, therefore, the final feeder bus stops were compact while on the K medoids clustering tends to be best in terms average distance of the clusters to the centroids and Kmeans have the worst average distance from clusters to centroids. Further analysis could be carried out to limit the capacity of the bus stops such that any stop whose demand exceeds the limits could split into n desirable stops. This will help in reducing the impact of the demand on the average waiting cost. This method shies away from conventional clustering algorithms which do not take into account the demand of the clusters and also from complex optimization techniques presented in the literature. To the best of our knowledge little literature, exist with regard to using smart card data in the determination of feeder bus stops. As for future endeavors, the feeder bus stops determined can be used to plan feeder bus route networks subsequently, since their designs are mainly based on passenger demand locations. Also, other forms of data may be useful such as mobile data records which may be able to pinpoint passenger demand locations themselves rather than bus stop locations. It is worthy of note that a form of o multi-objective optimization may be necessary in order to capture the varying stakeholders' perspectives in siting of feeder bus stops.

References

Bachok, S.; Ponrohono, Z.; Osman, M.M.; Bohari, Z.A. 2013. Gps/Gis Identification of Potential Bus Stop Locations and Passenger's Access and Egress Points. In *Third International Conference on Geotechnique, Construction Materials and Environment*, 13-15.

Chen, J.; Wang, S.; Liu, Z.; Chen, X. 2018. Network-level optimization of bus stop placement in urban areas, *KSCE Journal of Civil Engineering* 22(4): 1446-1453. doi.org/10.1007/s12205-017-0075-2.

Cui, Y.; Chen, S.K.; Liu, J.F.; Jia, W.Z. 2015. Optimal locations of bus stop connecting subways near urban intersections, *Mathematical Problems in Engineering* 2015: 537049. doi.org/10.1155/2015/537049.

Deri, A.; Kalpakci, A. 2014. Efficient Usage of Transfer-based System in Intracity Bus Transit Operation: Sample of Izmir, *Procedia-Social and Behavioral Sciences* 111: 311-319. doi.org/10.1016/j.sbspro.2014.01.064.

Dragan, D.; Kramberger, T.; Lipičnik, M. 2011. Monte Carlo simulation-based approach to optimal bus stops allocation in the municipality of Laško, *PROMET-Traffic&Transportation* 23(4): 265-278. doi.org/10.7307/ptt.v23i4.129.

Ge, Y.; Zhao, J.J.; Bian, Y.; Rong, J. 2011. Feeder bus stop selection within an integrated feeder bus planning framework. In *ICCTP 2011: Towards Sustainable Transportation Systems*, 2854-2865. doi.org/10.1061/41186(421)284.

Jahani, M.; Mehdi, H.S.; Ghatee, M.; Jahanshahi, M. 2013. A novel model for bus stop location appropriate for Public Transit Network Design: The case of Central Business Districts (CBD) of Tehran, *International journal of smart electrical engineering* 2(3): 133-141. Available from Internet: <https://www.sid.ir/en/journal/ViewPaper.aspx?id=380586>.

Leksakul, K.; Smutkupt, U.; Jintawiwat, R.; Phongmoo, S. 2017. A heuristic approach for solving employee bus routes in a large-scale industrial factory, *Advanced Engineering Informatics* 32: 176-187. doi.org/10.1016/j.aei.2017.02.006.

Li, H.G.; Chen, Y.S. 2016. Study on enterprise shuttle bus location and route optimization: An integrated approach, *J. Univ. Electron. Sci. Technol. China* 18: 68–73. doi:10.14071/j.1008-8105(2016)04-0068-06.

Liu, Y.; Jia, G.; Tao, X.; Xu, X. 2014. A stop planning method over big traffic data for airport bus shuttle, In *IEEE Fourth International Conference on Big Data and Cloud Computing*, 63-70.

Perugia, A.; Moccia, L.; Cordeau, J.F.; Laporte, G. 2011. Designing a home-to-work bus service in a metropolitan area, *Transportation Research Part B: Methodological* 45(10): 1710-1726. doi: 10.1016/j.trb.2011.05.025.

Prah, K.; Keshavarzsaleh, A.; Kramberger, T.; Jereb, B.; Dragan, D. 2018. Optimal bus stops' allocation: a school bus routing problem with respect to terrain elevation, *Logistics and sustainable transport* 9(2): 1-15. doi: 10.2478/jlst-2018-0006.

Ren, G.; Yu, Z.G.; Yuan, C.Q.; Xue, H.; Jiang, Q.Y. 2018. Optimal Bus Stop Location to Coordinate Transfer between Urban Rail Transit and Feeder Bus near Urban Road Intersection. In *CICTP 2017: Transportation Reform and Change—Equity, Inclusiveness, Sharing, and Innovation*, 2847-2855. Reston, VA: American Society of Civil Engineers, doi.org/10.1061/9780784480915.299.

Shatnawi, N.; Al-Omari, A.A.; Al-Qudah, H. 2020. Optimization of Bus Stops Locations Using GIS Techniques and Artificial Intelligence, *Procedia Manufacturing* 44: 52-59. doi.org/10.1016/j.promfg.2020.02.204.

Stat, T. 2014. *Information Society Statistics.* Turkish statistical institute. Available from Internet: <http://www. Tuik. Gov. Tr/PreIstatistikTablo. Do>.

Taplin, J.H.; Sun, Y. 2020. Optimizing bus stop locations for walking access: Stops-first design of a feeder route to enhance a residential plan, *Environment and Planning B: Urban Analytics and City Science* 47(7): 1237-1259. doi.org/10.1177/2399808318824108.

Wei, M.; Liu, T.; Sun, B. 2021. Optimal Routing Design of Feeder Transit with Stop Selection Using Aggregated Cell Phone Data and Open Source GIS Tool, *IEEE Transactions on Intelligent Transportation Systems* 22(4): 2452-2463. doi: 10.1109/TITS.2020.3042014.

Xiong, J.; Guan, W.; Song, L.; Huang, A.; Shao, C. 2013. Optimal routing design of a community shuttle for metro stations, *Journal of Transportation Engineering* 139(12): 1211-1223. doi.org/10.1061/(ASCE)TE.1943-5436.0000608.

Zhang, X.; Zhang, Q.; Sun, T.; Zou, Y.; Chen, H. 2018. Evaluation of urban public transport priority performance based on the improved TOPSIS method: A case study of Wuhan, *Sustainable Cities and Society* 43: 357–365. doi.org/10.1016/j.scs.2018.08.013.