Menemui Matematik (Discovering Mathematics)
Vol. 45, No. 1: 38 - 55 (2023)

Multi Origin Single Destination Split Delivery Selective Open Vehicle Routing Problem for First-Mile Ridesharing Service to Increase Public Transportation Take-Up

Kien Hua TING[1], Lai Soon LEE[1,2,*], and Hsin-Vonn SEOW[3]

[1] *Laboratory of Computational Statistics and Operations Research, Institute for Mathematical Research, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia.*

[2] *Department of Mathematics and Statistics, Faculty of Science, Universiti Putra Malaysia, Serdang 43400, Selangor, Malaysia.*

[3] *Nottingham University Business School, University of Nottingham Malaysia Campus, Semenyih 43500, Selangor, Malaysia.*

[*] Correspondence: lls@upm.edu.my

ABSTRACT

This paper is motivated by the unbalance utilization rate of public transit which affects the take-up rate of public transportation. The authors advocate solving this via the first-mile ridesharing problem. The selective open vehicle routing problem is used to model the first-mile ridesharing problem. Constrained K-Mean and K-Mean clusterings are used to cluster the dataset to represent the number of available drivers to service the transit to the stations. In terms of the meeting point selection, it can either be at a mutual meeting point (centroid) or at one of the cluster's commuter residences (non-centroid). For this, two types of models, Multi Origin Single Destination Split Delivery Open Vehicle Routing Problem (MOSD-SDOVRP) (Centroid) and Multi Origin Single Destination Split Delivery Selective Open Vehicle Routing Problem (MOSD-SDSOVRP) (Non-Centroid), are discussed. The proposed models are evaluated and compared using CPLEX with the well-known Solomon benchmark dataset. The results will allow a smooth transit for commuters from their respective residences to the station to encourage a high take-up of public transportation.

Keywords: public transit; ridesharing; first-mile problem; split delivery.

INTRODUCTION

Urban transportation can be categorized into public transport and private transport. Public transport is a transportation system that is managed on a schedule and operated on established routes to serve the travel demand of the public. Private transport allows an individual to freely use a vehicle without sharing it and allows users to have more flexibility and freedom to travel. However, this kind of transportation is not environmentally friendly.

In response to the increasing environmental awareness, governments of various countries have implemented policies to encourage more people to use public transport. However, Malaysians tend to not consider public transit because the current system is inefficient, unreliable, not punctual, and crowded. It is crucial that an effective urban transportation system facilitates the commuting of people from their homes to their working places and vice versa. Due to the limited coverage of transit hubs, some commuters may find it difficult to get to or from the transit hub. This is the first mile or last mile problem respectively. "First-mile" and "Last-mile" can be described as the beginning and end of a commuter's transit trip.

In the wake of Covid-19, users will appreciate more privacy and comfort from not needing to share transport with others. Hence ride-hailing services such as Grab, MyCar, EzCab, Dacsee, and MULA have become Malaysia's preferable transport. Ride-hailing (also known as ride-sourcing) is a prearranged and on-demand transportation service in hiring personal drivers to meet commuters' ride requests by sending them to the exact place they need to go without sharing the rides with other ride requests. The public has begun to switch their mobility choice from public transit, towards ride-hailing to not worry about the first and last-mile connectivity to public transit.

This leads to the low utilization rate of public transit. At the same time, the increase in ride-hailing demand causes an increase in the number of vehicles on the road (Anderson et al., 2014; Hampshire et al., 2017).

To solve the unbalance utilization rate of public transit and ride-hailing, ride-hailing can be the first-mile/ last-mile service provider but with the condition that commuters must share the rides. First-mile service is much more crucial than last-mile service because it can be one of the main factors that can influence people in choosing either public transport or private transport to go to work. With the convenience of first-mile services, people tend to choose public transport which indirectly influences people to choose last-mile services (Tay et al., 2012). Hence, this paper mainly focuses on the first-mile problem. To further support a sustainable environment, a first-mile ridesharing problem that possesses a ride-share concept is proposed.

The remainder of the paper is organized as follows. The next section is a presentation of relevant literature that has made possible the models used in this paper. This is followed by the methodology of the study and the mathematical formulations of the problem. The computational experiments of the proposed models are performed and discussed. Finally, the conclusion is given.

INNOVATIVE RIDESHARING SOLUTIONS

Vehicle Routing Problem with Ridesharing

The authors suggest this innovation based on the combination of the work done previously by Li et at. (2018), Bian and Liu (2019 & 2020), Chen et al, (2020), and Ning et al. (2021). Li et al. (2018) designed an improved ridesharing system by including meet places and the users' preferred time windows. With the advent of meet sites, rideshare operators balanced the advantages of reducing the number of delays that occur along the route against the costs of more walking that certain commuters must endure being collectively picked up or dropped off. Bian and Lie (2019) suggested a first-mile ridesharing service that allows commuters to personalise their requirements on various aspects of the inconveniences they face, such as the number of people riding with them, the amount of additional time spent travelling and waiting time at the transit hub. This was expanded upon by Bian et al. (2020) to accommodate instant booking. An autonomous vehicle dispatch and ridesharing scheduling with advanced requests was developed by Chen et al. (2020) as a potential solution to the issue of the first-mile with a clustering technique to partition the pickup points to cut down computational time. Ning et al. (2021) found the best travel routes for a large number of commuters while taking into account the uncertainty of commuter demand, which ensures good Quality of Service (QoS) for a Passenger-Centric Vehicle Routing for First-Mile Transportation (PCVR-FMT).

Clustering Method

In terms of the clustering method in vehicle routing, the K-Mean clustering method in classifying drivers' origins and destinations according to their respective locations (Najmi et al., 2017; Miranda-Bront et al., 2017; Shu et al., 2021). Hence for this paper, the authors utilize this method. Czioska et al. (2019) studied shared demand-responsive transportation (SDRT) systems using meeting points which can bring a significant impact towards customer satisfaction. To prevent a combinatorial explosion during the assignment of commuters to the meeting site, customer requests are clustered under a specified cluster size. After customer clustering, meeting point selection and route optimization are determined.

An adaptive clustering approach is employed to determine the position of the bus stop based on historical consumer demand data (Shu et al., 2021). The K-means clustering technique is used to produce the first set of bus stop locations and then adjusted by dividing and merging clusters to further reduce the walking distance and maintain a balance between bus stops.

Split Delivery

Allowing split deliveries may result in substantial savings in terms of both the total distance travelled and the number of necessary vehicles. The adoption of a split-delivery restriction in the VRP problem was first suggested by Dror and Trudeau (1989). This was further researched by Archetti et al. (2008), Gulczynski et al. (2010), Gutiérrez-Jarpa et al. (2010), Tang et al. (2013), Archetti et al. (2015), Yan et al. (2015), Moshref-Javadi and Lee (2016), Chen et al. (2016), Wang et al. (2016), Bianchessi et al. (2019), Bortfeldt and Yi (2020), Ji et al. (2021), and Ferreira et al. (2021). Ferreira et al. (2021) presented a study about the Capacitated Vehicle Routing Problem with Two-Dimensional Loading Constraints (2L-CVRP) by allowing split delivery and green requirements.

METHODOLOGY

In this section, the flow of the study, clustering method, and meeting point selection are discussed. The authors use the well-known Solomon benchmark datasets for Vehicle Routing with Time Windows problem (Solomon, 1987). These consist of randomly generated datasets (R sets), clustered datasets (C sets) and mixed random and clustered datasets (RC sets). Problem sets R1, C1 and RC1 have a short scheduling horizon and allow only a few commuters per route. These are suitable to represent the different conditions of commuters' residence locations.

Flowchart

First, a set of modified Solomon's datasets from Solomon (1987) are used in this study. The dataset will go through data clustering either by K-Mean clustering or constrained K-Mean clustering to obtain the clusters or groups of commuters. Once these groups are obtained, the selection of meeting points based on either centroid or non-centroid is carried out (see Figure 1). A centroid method (MOSD-SDOVRP) allows picking a meeting point which requires all commuters to walk a short distance to. A non-centroid method (MOSD-SDSOVRP) chooses the residence of one of the members of that cluster as a meeting point. Both methods' walking distance is shorter than walking to the nearest public transit spot. Both methods will be solved using the CPLEX with the branch and cut method.

One of the novelties of this study is the inclusion of the maximum walking distance to identify the most favourable number of clusters or groups of commuters for pick-up for each dataset. The maximum walking distance is 10 units (equivalent to 100 meters) which is the maximum distance between the residences of the commuters and the centralized pickup point within each cluster/meeting point (centroid case), or the maximum distance between the residences of the commuters within each cluster/meeting point (non-centroid case).

Clustering Method

The clustering analysis can be divided into two main groups: hierarchical and partitional (see, Figure 2). They generally depend on providing prior knowledge or information of the exact number of clusters for each dataset to be clustered and analysed. A hierarchical clustering algorithm is further subdivided into agglomerative and divisive methods, while the partitional has four subdivisions: hard or crisp clustering method, fuzzy method, mixture, and square error. The application of hierarchical clustering in large-scale datasets is limited because it has high computational complexity. A drawback of hierarchical methods is that those that are practical in terms of time efficiency require memory usage proportional to the square of the number of groups in the initial partition (Fraley and Raftery, 1998). There are some limitations as well in partitional methods such as the user having to specify in advance the number of clusters, data-dependent, and often converging to suboptimal solutions.

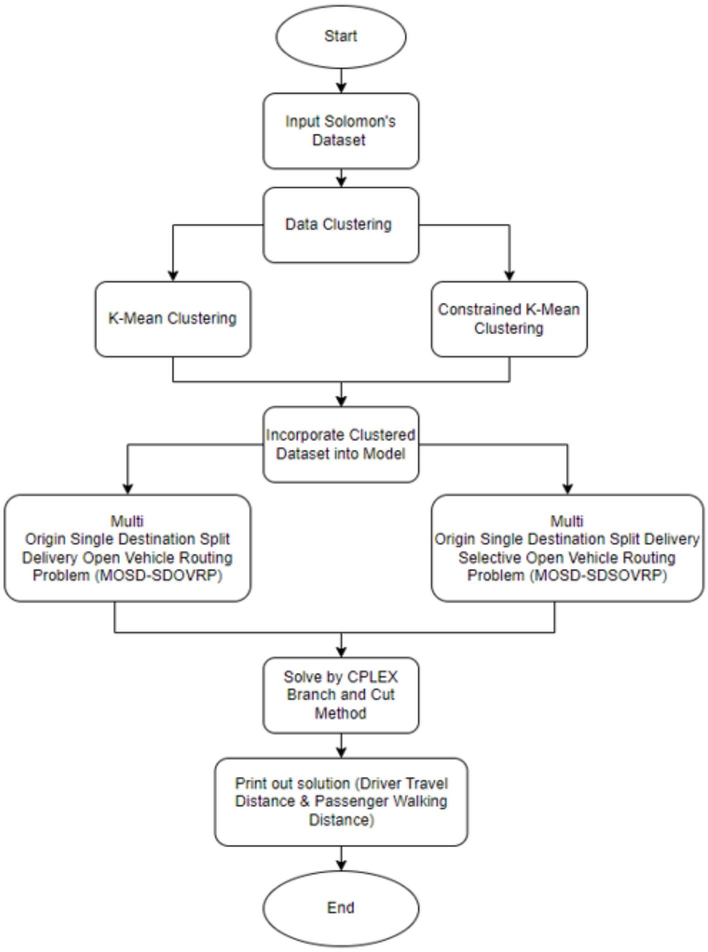

Figure 1: Flowchart of this Study

In this paper, K-Mean clustering and constrained K-Mean clustering, which are under partitional methods in clustering, are selected. The main advantage of these methods is they are easy to implement and have low time complexity for large datasets (Das et al., 2007). Also, constrained K-Mean clustering allows for the clustering of data sets with a fixed number of people in each cluster (the number of clusters can also be fixed). For instance, if the minimum and the maximum limit within each cluster are set to be 2 and 3 respectively. Then, a dataset with 25 commuters will have a maximum number of clusters of 13 and the minimum number of clusters is 9. K-Mean clustering, on the other hand, is a simple iterative hill-climbing algorithm which only requires a number of clusters (Garai and Chaudhuri, 2004). Since there is no cluster size limit, therefore a 25-commuter dataset can have a cluster size of 1 to 25.

Meeting Points Selection (Centroid and Non-Centroid)

In terms of solving this first-mile ridesharing problem, a new approach called the Selective Open Vehicle Routing Problem (SOVRP) is used. The centroid-based clustered dataset was solved by Multi Origin Single Destination Split Delivery Open Vehicle Routing Problem (MOSD-SDOVRP) while the non-centroid-based clustered dataset was solved by Multi Origin Single Destination Split Delivery Selective Open Vehicle Routing Problem (MOSD-SDSOVRP) (see, Figure 3). In the case of centroid, the coordinate for the center of each cluster is defined as the pickup point of the cluster. All the commuters within the cluster must walk to the center of the cluster or fixed pickup point. Therefore, MOSD-SDOVRP is the suitable model for demonstrating the centroid case scenario problem. While in the case of non-centroid, either one of the nodes within each cluster will be the possible pickup point. Hence, the model of MOSD-SDSOVRP is used in demonstrating non-centroid cases in which either one of the nodes within the cluster will be picked under the model.

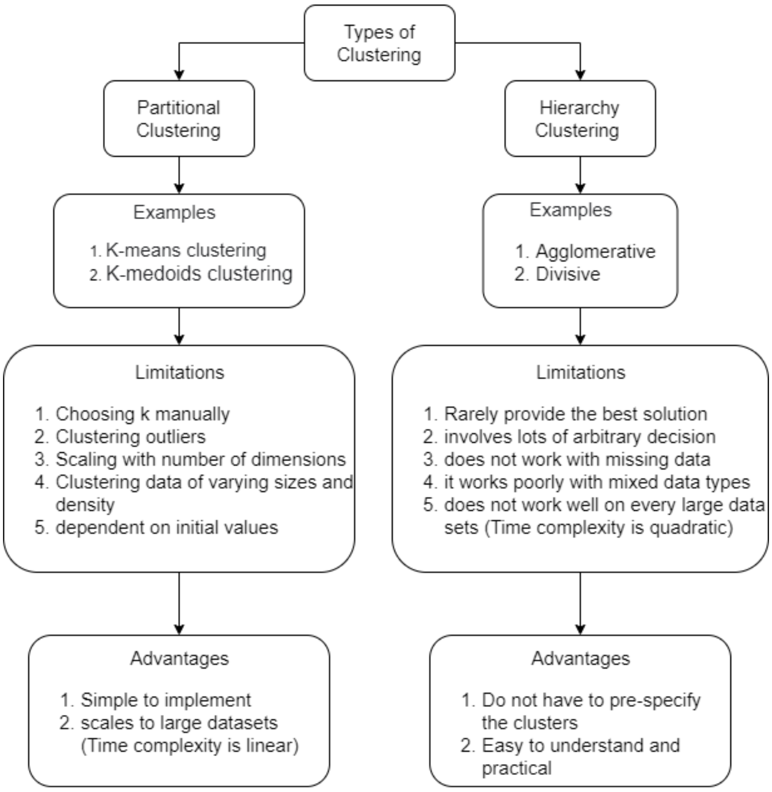

Figure 2: Types of Clustering Methods

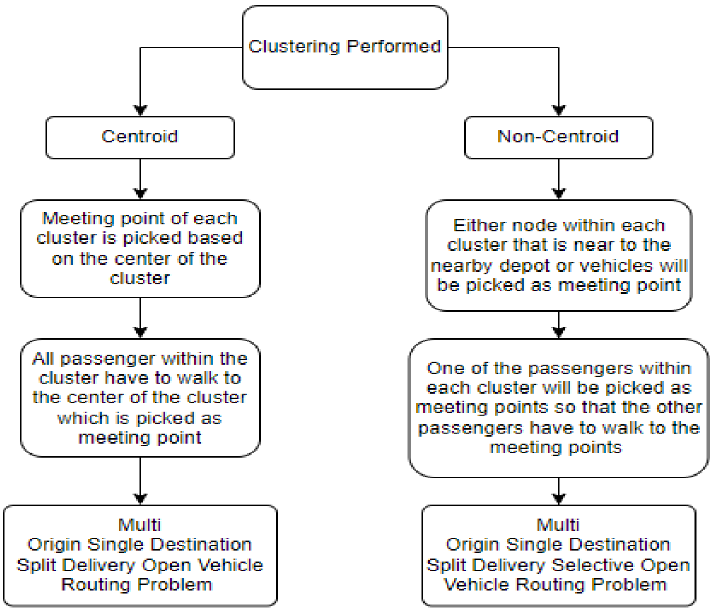

Figure 3: Meeting Point Selection

MATHEMATICAL FORMULATION

The mathematical formulation of MOSD-SDOVRP is a common vehicle routing problem with split delivery which refers to work by Vornhusen & Kopfer (2015) work. On the other hand, the mathematical formulation of MOSD-SDSOVRP is the combination between split delivery and selective concept which refer to studies by Sabo et al. (2020) and Vornhusen & Kopfer (2015). The formulation for MOSD-SDSOVRP is considered one of the novelties of this paper since there are no existing studies that have split delivery together with a selective concept.

Multi Origin Single Destination Split Delivery Open Vehicle Routing Problem (MOSD-SDOVRP)

The MOSD-SDOVRP is defined over an undirected graph $G = (V, E)$ with vertex set $V = \{0,1, \dots, n\}$.

K = a fleet K of homogeneous vehicles k ,
D = set of depot,
De = set of destination nodes,
Q = capacity of the vehicle,
qi = demand of the node i ,
$x_{ij}^k = \begin{cases} 1, \text{if the vehicle } k \text{ traverse the edge } (i,j) \\ 0, \text{if the edge } (i,j) \text{ is not part of any route} \end{cases}$,
$y_i^k = \begin{cases} 1, \text{if demand point } i \text{ is visited by vehicle } k \ (y_0^k = 1) \\ 0, \text{otherwise} \end{cases}$,
$w_i^k =$ delivery amount at demand point i by vehicle k $(w_0^k = 0)$,
$u_i^k =$ dummy continuous variables for subtour elimination constraints .
The MOSD-SOVRP can be formulated as follows:

$$min \sum_{i \in V} \sum_{\substack{j \in V \\ i \neq j}} \sum_{k \in K} x_{ij}^k \cdot Dist_{ij} \tag{1}$$

subject to

$$\sum_{\substack{i \in V \\ i \neq j}} \sum_{k \in K} x_{ij}^k \geq 1, \forall j \in V \setminus \{D + De\} \tag{2}$$

$$\sum_{\substack{j \in V \\ i \neq j}} x_{ij}^k = \sum_{\substack{j \in V \\ i \neq j}} x_{ji}^k, \forall k \in K, \forall i \in V \setminus \{D + De\} \tag{3}$$

$$u_i^k - u_j^k + n \cdot x_{ij}^k \leq \text{n} - 1, \forall i \in V \setminus \{De\}, \forall j \in V \setminus \{D + De\}, \forall k \in K, i \neq j \tag{4}$$

$$\sum_{j \in V \setminus \{D\}} x_{ij}^k = y_i^k, \forall k \in K, \forall i \in V \tag{5}$$

$$\sum_{i \in V \setminus \{D + De\}} w_i^k \leq Q \ \ \forall k \in K \tag{6}$$

$$q_i \cdot y_i^k \geq w_i^k, \forall i \in V \backslash \{D + De\}, \forall k \in K \tag{7}$$

$$\sum_{k \in K} w_i^k = q_i \ , \forall i \in V \tag{8}$$

$$\sum_{\substack{i \in V \backslash \{D+De\} \\ i \neq j}} \sum_{j \in De} x_{ij}^k = 1, \forall\ k \in K \tag{9}$$

$$\sum_{\substack{i \in V \backslash \{D+De\} \\ i \neq j}} \sum_{j \in D} x_{ij}^k = 0, \forall\ k \in K \tag{10}$$

$$w_i^k \geq 0, \forall k \in K, \forall i \in V \tag{11}$$

Constraint (2) imposes that any node can be visited at least once, while constraint (3) is the flow conservation constraints while (4) is the subtour elimination constraints. Constraint (5) ensures variables x_{ij}^k and y_i^k are linked. Constraint (6) is capacity constraint to ensure the quantity delivered by each vehicle does not exceed the vehicle capacity, while constraint (7) ensure that vehicles must not fetch any customers which are destined for customers they do not visit. Constraint (8) ensures that the entire demand of each customer is satisfied. Constraint (9) makes sure that all nodes must end at destination and (10) ensures that all nodes must end at depot.

Multi Origin Single Destination Split Delivery Selective Open Vehicle Routing Problem (MOSD-SDSOVRP)

The MOSD-SDSOVRP is also defined over an undirected graph $G = (V, E)$ similar to MOSD-SDOVRP with the following additional variables.

$C =$ set of clusters,
$d_c =$ total demand of cluster,
$TD_j =$ updated demand of the nodes within the same cluster j ,
$z_c^k = \begin{cases} 1, \text{if total demand of the cluster } c \text{ is visited by vehicle } k \\ 0, \text{otherwise} \end{cases}$,
$v_j = \begin{cases} 1, \text{if one of the nodes within the cluster is visited} \\ 0, \text{otherwise} \end{cases}$,
$w_c^k =$ delivery amount at total demand cluster c by vehicle k $\left(w_0^k = 0\right)$,
$\lambda_{ic} =$ if the node $i \in V$ belongs to cluster $c \in C$ $(\lambda_{ic} = 1)$ or not $(\lambda_{ic} = 0)$.

Hence, the MOSD-SDSOVRP can be formulated as follows:

$$min \sum_{i \in V} \sum_{\substack{j \in V \\ i \neq j}} \sum_{k \in K} x_{ij}^k \cdot Dist_{ij} \tag{12}$$

subject to

$$\sum_{\substack{j \in V \\ i \neq j}} x_{ij}^k = \sum_{\substack{j \in V \\ i \neq j}} x_{ji}^k, \forall k \in K, \forall\, i \in V \setminus \{D + De\} \tag{13}$$

$$u_i^k - u_j^k + n \cdot x_{ij}^k \leq \mathrm{n} - 1, \forall i \in V \setminus \{De\}, \forall j \in V \setminus \{D + De\}, \forall k \in K, i \neq j \tag{14}$$

$$\sum_{\substack{i \in V \setminus \{D+De\} \\ i \neq j}} \sum_{j \in V \setminus \{D+De\}} \lambda_{jc} \cdot x_{ji}^k = z_c^k, \forall k \in K, \forall c \in C \setminus \{D + De\} \tag{15}$$

$$\sum_{k \in K} w_c^k = d_c \ , \forall c \in C \setminus \{D + De\} \tag{16}$$

$$d_c \cdot z_c^k = w_c^k, \forall c \in C \setminus \{D + De\}, \forall k \in K \tag{17}$$

$$\sum_{c \in C \setminus \{D+De\}} w_c^k \leq Q \ , \forall k \in K \tag{18}$$

$$\sum_{\substack{i \in V \setminus \{D+De\} \\ i \neq j}} \sum_{j \in De} x_{ij}^k \leq 1, \forall\, k \in K \tag{19}$$

$$\sum_{\substack{i \in V \setminus \{D+De\} \\ i \neq j}} \sum_{j \in D} x_{ij}^k = 0, \forall\, k \in K \tag{20}$$

$$\sum_{\substack{i \in V \setminus \{D\} \\ i \neq j}} \sum_{k \in K} x_{ji}^k \geq \left\lceil \frac{TD_j}{Q} + 1 \right\rceil \cdot v_j, \forall\, j \in V \setminus \{D + De\} \tag{21}$$

$$\sum_{j \in V \setminus \{D\}} \lambda_{jc} \cdot v_j = 1, \forall c \in C \setminus \{D + De\} \tag{22}$$

$$\sum_{j \in V \setminus \{D+De\}} \sum_{i \in D} x_{ij}^k = 1 \ , \forall k \in K \tag{23}$$

$$w_c^k \geq 0, \forall k \in K, \forall c \in C \setminus \{D + De\} \tag{24}$$

Constraint (13) is the flow conservation constraints and (14) is the subtour elimination constraint. Constraint (15) ensures variables x_{ij}^k and y_i^k are linked, while constraint (16) is to ensure that the entire demand of each cluster is satisfied. Constraint (17) impose that vehicles must not fetch any customers which are destined for customers they do not visit. Constraint (18) is capacity constraint to ensure the quantity delivered by each vehicle does not exceed the vehicle capacity. Constraint (19) makes sure that all nodes must end at destination and (20) ensures that all nodes must end at depot. Constraint (21) and (22) ensure that only one of the nodes within each cluster is selected which is one of the novelties of this study. Constraint (23) ensures that all vehicles are fully utilized.

COMPUTATIONAL EXPERIMENTS

To solve the proposed mathematical formulations of MOSD-SDOVRP and MOSD-SDSOVRP, the CPLEX with the branch and cut method is used. The models are coded using Phyton language and the computational experiments are performed on a laptop computer running on Intel® Core™ i5-8250U CPU @ 1.60GHz-1.80 GHz, with 8GB RAM of memory. Both models have investigated 12 modified problem instances from the well-known Solomon benchmarks dataset. There are three sets of the dataset which are the C, R, and RC datasets. Each set of the dataset has 4 different numbers of meeting points. The capacity of each vehicle is 3 commuters. Table 1 showed the computational results for centroid case (MOSD-SDOVRP). In column 1, “Dataset” represents the modified Solomon dataset which are C, R and RC. For example, R101-25 represents R type of dataset with 25 nodes. In column 2, “Clustering Method” represents the types of the clustering method. Column 3 “Cluster Size” represents the cluster size, especially for constrained K-Mean. Column 4, “Clusters” represents the number of clusters produced under the stated clustering method. Column 5, “Total Demand” represents the total number of demands for each node. The total demand is always equal to the number of nodes due to the setting of only one customer demand on each node. Column 6, “Vehicles” represents the number of vehicles that will be utilized for each dataset. Column 7, “Status” represents the obtained solution whether it is feasible, optimal or unknown. Column 8, “CPU(s)” represents the computational time (seconds) of CPLEX. Column 9, “Total WD for Total Demand (unit)” represents the total walking distance (x 10m) for the commuters. Column 10, “Radius Cluster (unit)” represents the maximum travel distance (x 10m) for each commuter.

Figure 4 shows the obtained number of clusters/meeting points for the centroid case of both constrained K-Mean and K-Mean clustering. As mentioned early, constrained K-Mean clustering limits the number of commuters within each cluster to between 2 to 3. It is obvious that the number of clusters will be higher as compared to the number of clusters produced by K-Mean clustering as the latter does not have a cluster size limitation. Looking at Figure 5, constrained K-Mean clustering produced a much higher number of clusters due to the limit in cluster size. Hence, the total distance for all commuters to walk to the pickup point will be much lower as compared to K-Mean clustering. So, as shown in Figure 5, constrained K-Mean can produce a lower total commuter walking distance for all types of datasets. Hence, it is much more favourable to use constrained K-Mean in clustering the dataset from the commuter point-of-view. Overall, constrained K-Mean is the best clustering method for a shorter total walking distance for commuters, especially in the centroid case. K-Mean clustering will challenge the welfare of the commuters as the total walking distance for them will be much higher.

Table 2 showed computational results for non-centroid case (MOSD-SDSOVRP). The descriptions for each column are mentioned previously in Table 1. In Table 2, due to the complexity of the model, the solution for the dataset of 80 and 100 nodes are unknown even after 2 hours (7200 CPU seconds) of computation by the CPLEX. In the case of the non-centroid method, the solution for the C dataset is still comparable when discussing which clustering method is more suitable to be used. However, it is hard to tell which clustering method is superior for the R and RC dataset due to the input parameter of the radius cluster, or which is the maximum distance for commuters to walk. The maximum walking distance is 100m but in the case of R101-25, R101-50 and RC101-50, the clusters cannot be obtained due to that constraint. Hence, to have a clustered dataset, the maximum walking distance parameter had to be increased. For example, in R101-25 using constrained K-Mean in clustering the dataset for the centroid case, it is calculated that the maximum number of clusters/group of commuters is 13 and the minimum is 9. In the non-centroid case of R101-25, cluster 12 with the lowest maximum walking distance (155.5m) is selected. This is one of the limitations of constrained K-Mean clustering which affects the maximum walking distance for commuters.

Table 1: Computational Results for Centroid (MOSD-SDOVRP)

Dataset	Clustering Method	Cluster Size	Clusters	Total Demand	Vehicles	Status	CPU (s)	Total WD for Total Demand (unit)	Radius Cluster (unit)
C101-25	Constrained K-Mean	Max:3, Min: 2	9	25	9	Optimal	22.41	44.372 (= 443.72m)	10 (= 100 m)
C101-25	K-Mean	-	3	25	9	Optimal	0.23	92.213	10
C101-50	Constrained K-Mean	Max:3, Min: 2	18	50	17	Feasible	7200.00	93.370	10
C101-50	K-Mean	-	5	50	17	Optimal	104.19	198.112	10
C101-80	Constrained K-Mean	Max:3, Min: 2	28	80	27	Feasible	7200.00	167.104	10
C101-80	K-Mean	-	8	80	27	Optimal	2073.67	341.249	10
C101-100	Constrained K-Mean	Max:3, Min: 2	36	100	34	Feasible	7200.00	211.370	10
C101-100	K-Mean	-	10	100	34	Feasible	7200.00	437.960	10
R101-25	Constrained K-Mean	Max:3, Min: 2	11	25	9	Optimal	54.91	151.918	10
R101-25	K-Mean	-	11	25	9	Optimal	100.66	146.559	10
R101-50	Constrained K-Mean	Max:3, Min: 2	17	50	17	Optimal	1869.13	287.332	10
R101-50	K-Mean	-	17	50	17	Feasible	7200.00	269.470	10
R101-80	Constrained K-Mean	Max:3, Min: 2	27	80	27	Feasible	7200.00	395.846	10
R101-80	K-Mean	-	19	80	27	Feasible	7200.00	454.900	10
R101-100	Constrained K-Mean	Max:3, Min: 2	34	100	34	Feasible	7200.00	408.060	10
R101-100	K-Mean	-	22	100	34	Feasible	7200.00	496.670	10

RC101-25	Constrained K-Mean	Max:3, Min: 2	9	25	9	Optimal	46.20	57.552	10
RC101-25	K-Mean	-	3	25	9	Optimal	0.28	127.546	10
RC101-50	Constrained K-Mean	Max:3, Min: 2	20	50	17	Feasible	7200.00	102.590	10
RC101-50	K-Mean	-	6	50	17	Optimal	10.13	234.630	10
RC101-80	Constrained K-Mean	Max:3, Min: 2	28	80	27	Feasible	7200.00	327.320	10
RC101-80	K-Mean	-	18	80	27	Feasible	7200.00	374.650	10
RC101-100	Constrained K-Mean	Max:3, Min: 2	36	100	34	Feasible	7200.00	367.150	10
RC101-100	K-Mean	-	19	100	34	Feasible	7200.00	515.560	10

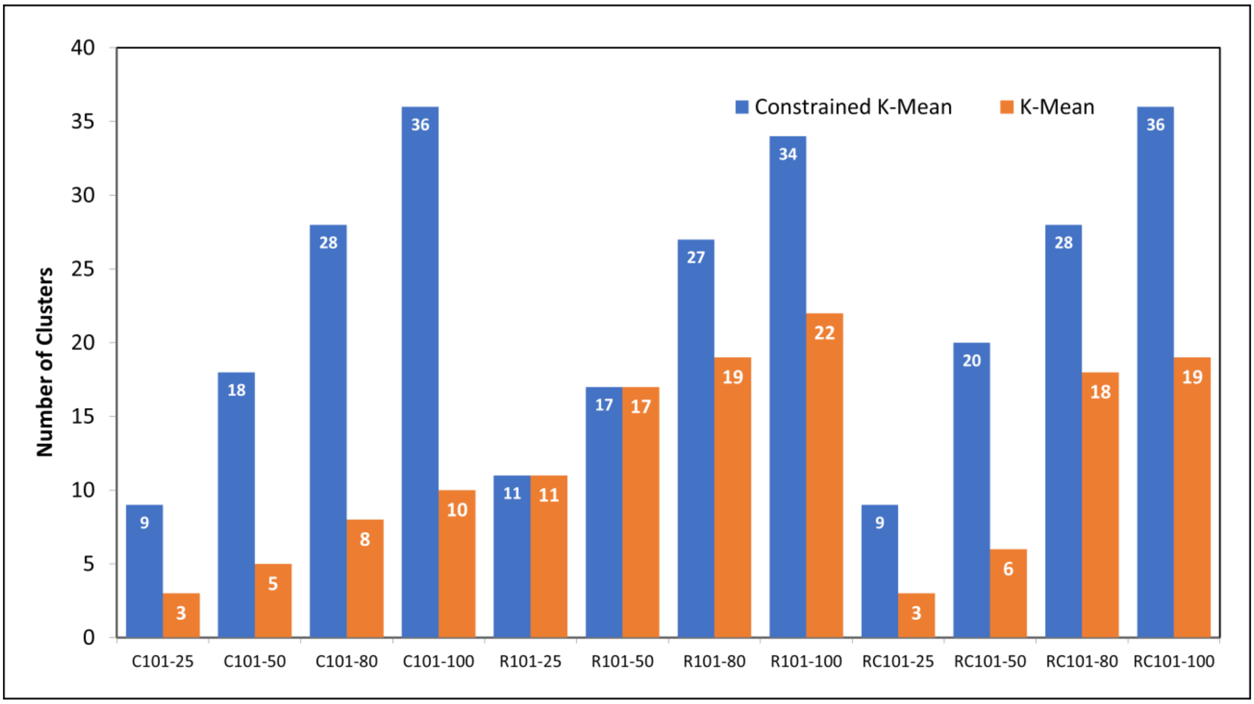

Figure 4: Centroid: Constrained K-Mean versus K-Mean (Number of Clusters)

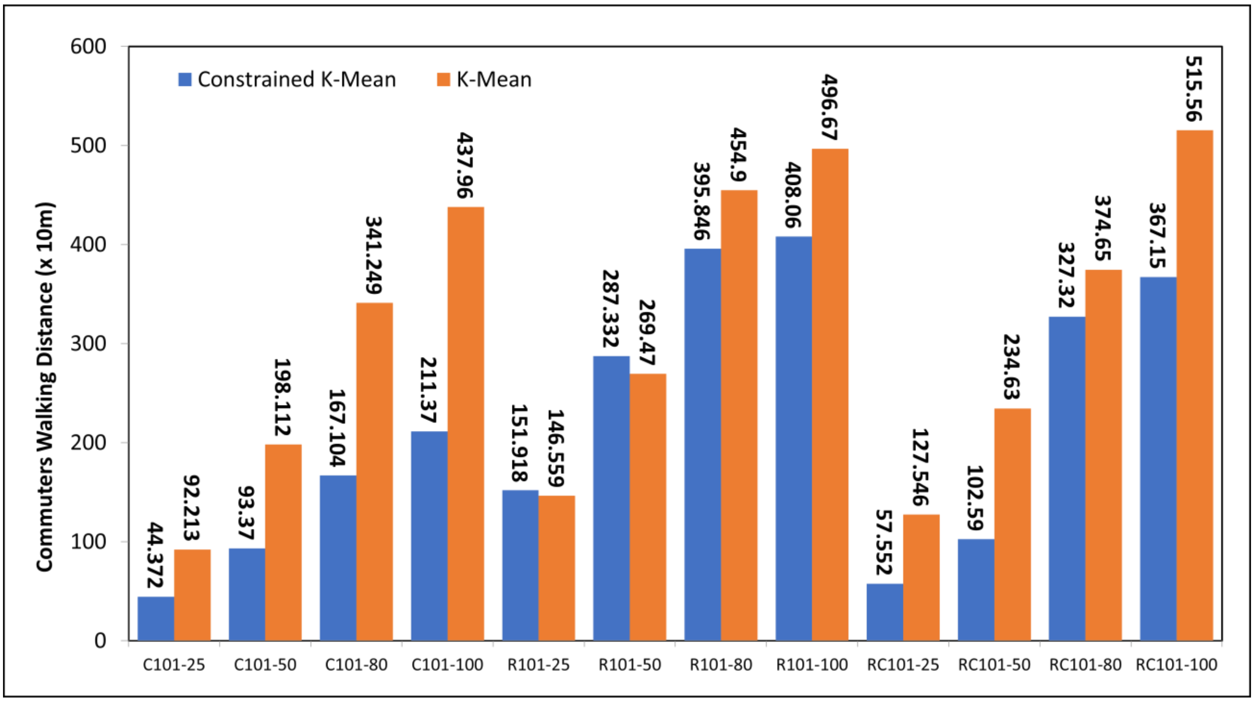

Figure 5: Centroid: Constrained K-Mean versus K-Mean (Total Commuters Walking Distance)

Table 2: Computational Results for Non-Centroid (MOSD-SDSOVRP)

Dataset	Clustering Method	Cluster Size	Clusters	Total Demand	Vehicles	Status	CPU(s)	Total WD for Total Demand (unit)	Radius Cluster (unit)
C101-25	Constrained K-Mean	Max:3, Min: 2	9	25	9	Optimal	263.88	51.53 (= 515.3m)	10 (= 100m)
C101-25	K-Mean	-	5	25	9	Optimal	26.13	82.993	10
C101-50	Constrained K-Mean	Max:3, Min: 2	18	50	17	Feasible	7200.00	113.250	10
C101-50	K-Mean	-	9	50	17	Feasible	7200.00	202.230	10
C101-80	Constrained K-Mean	Max:3, Min: 2	30	80	27	Unknown	-	-	10
C101-80	K-Mean	-	17	80	27	Unknown	-	-	10
C101-100	Constrained K-Mean	Max:3, Min: 2	37	100	34	Unknown	-	-	10.44
C101-100	K-Mean	-	25	100	34	Unknown	-	-	10
R101-25	Constrained K-Mean	Max:3, Min: 2	12	25	9	Optimal	156.98	144.626	15.55
R101-25	K-Mean	-	20	25	9	Optimal	70.28	44.140	10
R101-50	Constrained K-Mean	Max:3, Min: 2	24	50	17	Feasible	7200.00	217.250	11.18
R101-50	K-Mean	-	34	50	17	Feasible	7200.00	108.880	10
R101-80	Constrained K-Mean	Max:3, Min: 2	39	80	27	Unknown	-	-	12.72
R101-80	K-Mean	-	49	80	27	Unknown	-	-	10
R101-100	Constrained K-Mean	Max:3, Min: 2	41	100	34	Unknown	-	-	13
R101-100	K-Mean	-	50	100	34	Unknown	-	-	10
RC101-25	Constrained K-Mean	Max:3, Min: 2	9	25	9	Optimal	74.64	70.840	10
RC101-25	K-Mean	-	6	25	9	Optimal	23.06	88.350	10

RC101-50	Constrained K-Mean	Max:3, Min: 2	15	50	17	Feasible	7200.00	142.790	13
RC101-50	K-Mean	-	20	50	17	Feasible	7200.00	120.810	10
RC101-80	Constrained K-Mean	Max:3, Min: 2	37	80	27	Unknown	-	-	16.401
RC101-80	K-Mean	-	39	80	27	Unknown	-	-	10
RC101-100	Constrained K-Mean	Max:3, Min: 2	40	100	34	Unknown	-	-	13.15
RC101-100	K-Mean	-	36	100	34	Unknown	-	-	10

Figure 6 shows the proposed number of clusters/groups of commuters under the non-centroid method for both constrained K-Mean and K-Mean clustering. Both clustering methods performed as expected in the datasets C101-25, C101-50, and RC101-25 where constrained K-Mean produced more clusters than K-Mean clustering. However, in the R101-25, R101-50, and RC101-50 datasets, constrained K-Mean produced a lower number of clusters. This is totally different from other datasets. The reason is that they are clustered under a relaxed maximum walking distance for commuters. For example, in R101-25, the relaxed maximum walking distance is 155.5m.

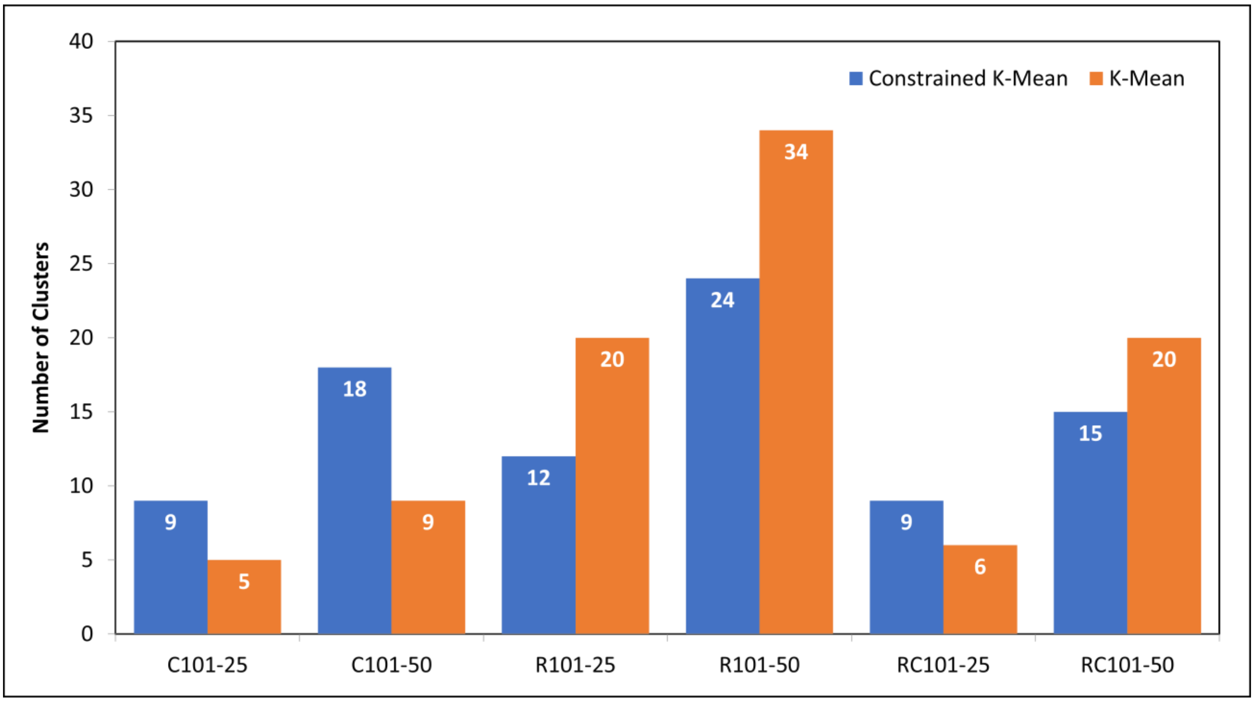

Figure 6: Non-Centroid: Constrained K-Mean Versus K-Mean Clustering (Number of Clusters)

The results in Figure 7 show the obtained total commuters' walking distance for the non-centroid. Based on the C101-25, C101-50, and RC101-25 sets in non-centroid cases, constrained K-Mean can ensure a lower total walking distance for commuters under 10 meters. Conversely, for R101-25, R101-50, and RC101-50, due to the relaxed maximum walking distance, the K-Mean clustering dataset tends to have a much higher total walking distance.

Figure 8 shows the obtained total commuters walking distance for centroid versus non-centroid. In the C dataset, the non-centroid method performed better in both clustering methods but with a little sacrifice of the total walking distance for commuters. However, in the case of the centroid for both clustering methods, it does ensure a lower total walking for commuters.

In the R dataset, the centroid scenario tends to sacrifice a lot of walking distance for commuters for both clustering methods. Constrained K-Mean tends to outperform others, especially in the non-centroid scenario, but this is due to the relaxed maximum walking distance. Overall, constrained K-Mean is still performing better in a centroid for the R dataset and is more suitable to be applied.

In RC dataset, the non-centroid condition seems to have better results compared to the centroid condition, regardless of which clustering method is applied. The K-Mean clustering performed better in saving total walking distance for commuters. Conversely, the constrained K-Mean had a higher total walking distance for commuters.

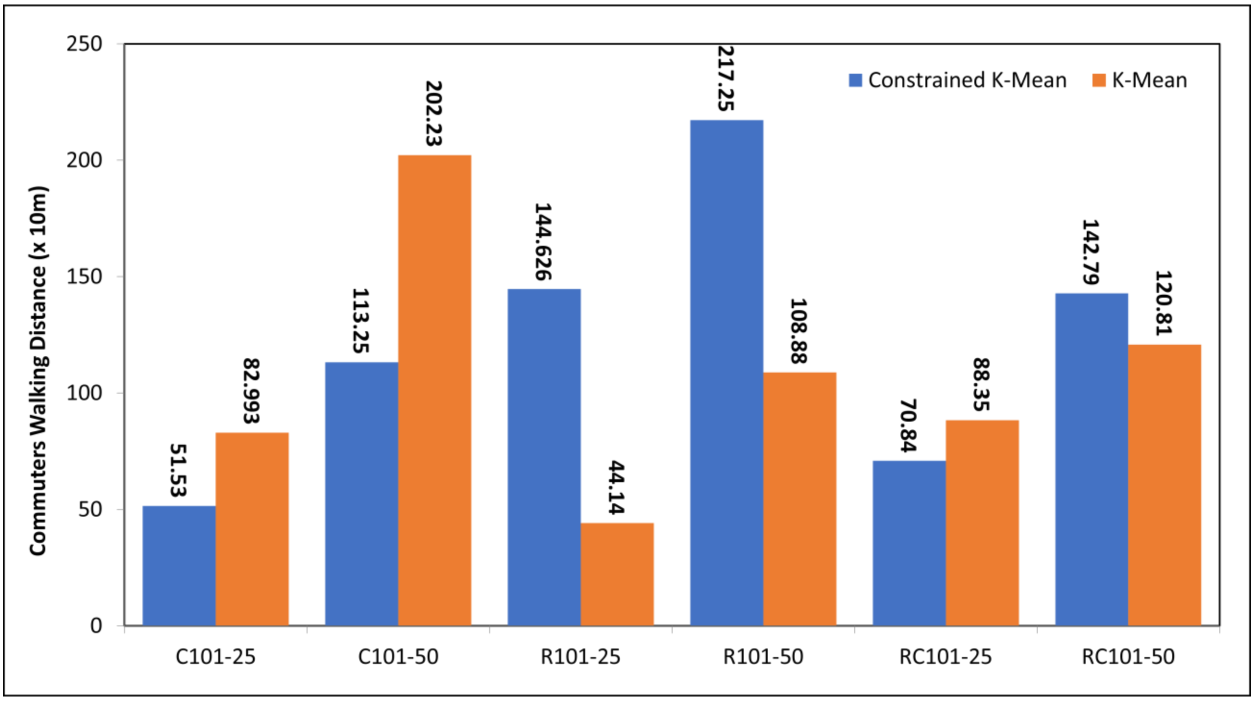

Figure 7: Non-Centroid: Constrained K-Mean versus K-Mean Clustering (Total Commuters Walking Distance)

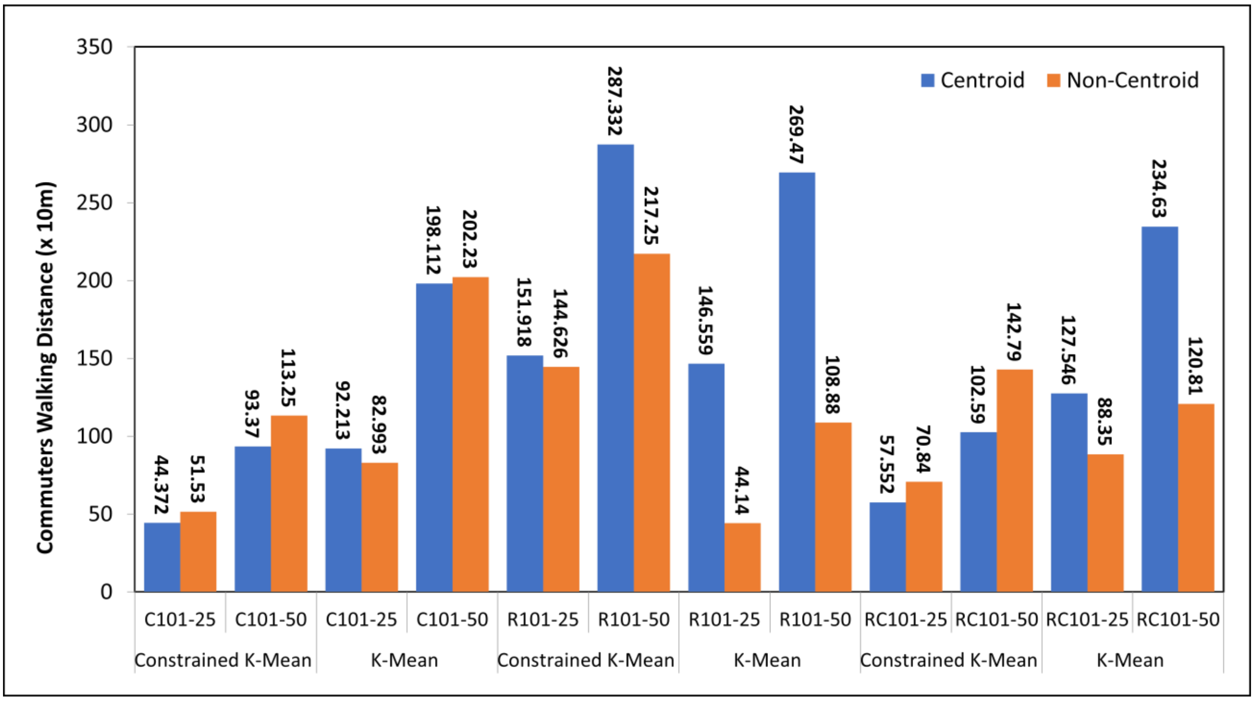

Figure 8: Centroid Versus Non-Centroid (Total Commuters Walking Distance)

CONCLUSION

The model allows for a choice to cater for a ridesharing driver perspective or a commuter: a driver's perspective and dealing with the C type dataset; then non-centroid condition must be implemented. In terms of the clustering method, constrained K-Mean should be utilized. If it is the commuter's perspective and dealing with the C dataset, centroid condition with constrained K-Mean should be picked. For the R type dataset, the centroid condition with constrained K-Mean clustering method must be implemented to ensure the system is driver-oriented. If one considers the commuter's perspective in dealing with the R type dataset, a non-centroid condition with constrained K-Mean should be picked. If one is dealing with the driver's perspective in dealing with the RC type dataset, then a non-centroid condition must be implemented. In terms of the clustering method, constrained K-Mean should be utilized. When one is looking from the commuter's perspective in dealing with the RC type dataset, the non-centroid condition with K-Mean should be picked.

REFERENCES

Anderson, D. N. (2014). "Not just a taxi"? For-profit ridesharing, driver strategies, and VMT. *Transportation*, *41*(5), 1099-1117.

Archetti, C., Savelsbergh, M. W., & Speranza, M. G. (2008). To split or not to split: That is the question. *Transportation Research Part E: Logistics and Transportation Review*, *44*(1), 114-123.

Archetti, C., Bianchessi, N., & Speranza, M. G. (2015). A branch-price-and-cut algorithm for the commodity constrained split delivery vehicle routing problem. *Computers & Operations Research*, *64*, 1-10.

Bian, Z., & Liu, X. (2019). Mechanism design for first-mile ridesharing based on personalized requirements part I: Theoretical analysis in generalized scenarios. *Transportation Research Part B: Methodological*, *120*, 147-171.

Bian, Z., Liu, X., & Bai, Y. (2020). Mechanism design for on-demand first-mile ridesharing. *Transportation Research Part B: Methodological*, *138*, 77-117.

Bianchessi, N., Drexl, M., & Irnich, S. (2019). The split delivery vehicle routing problem with time windows and customer inconvenience constraints. *Transportation Science*, *53*(4), 1067-1084.

Bortfeldt, A., & Yi, J. (2020). The split delivery vehicle routing problem with three-dimensional loading constraints. *European Journal of Operational Research*, *282*(2), 545-558.

Chen, P., Dong, X., & Xiao, J. (2016, August). Study on split delivery open vehicle routing problem within two splits. In *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp. 724-728). IEEE.

Chen, S., Wang, H., & Meng, Q. (2020). Solving the first-mile ridesharing problem using autonomous vehicles. *Computer-Aided Civil and Infrastructure Engineering*, *35*(1), 45-60.

Czioska, P., Kutadinata, R., Trifunović, A., Winter, S., Sester, M., & Friedrich, B. (2019). Real-world meeting points for shared demand-responsive transportation systems. *Public Transport*, *11*(2), 341-377.

Das, S., Abraham, A., & Konar, A. (2007). Automatic clustering using an improved differential evolution algorithm. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *38*(1), 218-237.

Dror, M., & Trudeau, P. (1989). Savings by split delivery routing. *Transportation Science*, *23*(2), 141-145.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, *41*(8), 578-588.

Ferreira, K. M., de Queiroz, T. A., & Toledo, F. M. B. (2021). An exact approach for the green vehicle routing problem with two-dimensional loading constraints and split delivery. *Computers & Operations Research*, *136*, 105452.

Garai, G., & Chaudhuri, B. B. (2004). A novel genetic algorithm for automatic clustering. *Pattern Recognition Letters*, *25*(2), 173-187.

Ghiani, G., & Improta, G. (2000). An efficient transformation of the generalized vehicle routing problem. *European Journal of Operational Research*, *122*(1), 11-17.

Gulczynski, D., Golden, B., & Wasil, E. (2010). The split delivery vehicle routing problem with minimum delivery amounts. *Transportation Research Part E: Logistics and Transportation Review*, *46*(5), 612-626.

Gutiérrez-Jarpa, G., Desaulniers, G., Laporte, G., & Marianov, V. (2010). A branch-and-price algorithm for the vehicle routing problem with deliveries, selective pickups and time windows. *European Journal of Operational Research*, *206*(2), 341-349.

Hampshire, R. C., Simek, C., Fabusuyi, T., Di, X., & Chen, X. (2017). Measuring the impact of an unanticipated suspension of ride-sourcing in Austin, Texas. *SSRN Electron. J*.

Ji, B., Zhou, S., Samson, S. Y., & Wu, G. (2021). An enhanced neighborhood search algorithm for solving the split delivery vehicle routing problem with two-dimensional loading constraints. *Computers & Industrial Engineering*, *162*, 107720.

Li, X., Hu, S., Fan, W., & Deng, K. (2018). Modeling an enhanced ridesharing system with meet points and time windows. *PloS One*, *13*(5), e0195927.

Miranda-Bront, J. J., Curcio, B., Méndez-Díaz, I., Montero, A., Pousa, F., & Zabala, P. (2017). A cluster-first route-second approach for the swap body vehicle routing problem. *Annals of Operations Research*, *253*(2), 935-956.

Moshref-Javadi, M., & Lee, S. (2016). The customer-centric, multi-commodity vehicle routing problem with split delivery. *Expert Systems With Applications*, *56*, 335-348.

Najmi, A., Rey, D., & Rashidi, T. H. (2017). Novel dynamic formulations for real-time ride-sharing systems. *Transportation Research Part E: Logistics and Transportation Review*, *108*, 122-140.

Sabo, C., Pop, P. C., & Horvat-Marc, A. (2020). On the selective vehicle routing problem. *Mathematics*, *8*(5), 771.

Shu, P., Sun, Y., Xie, B., Xu, S. X., & Xu, G. (2021). Data-driven shuttle service design for sustainable last mile transportation. *Advanced Engineering Informatics*, *49*, 101344.

Solomon (1987). Solomon's VRPTW benchmark problems. https://people.idsia.ch/~luca/macs-vrptw/problems/welcome.htm.

Tang, J., Ma, Y., Guan, J., & Yan, C. (2013). A max–min ant system for the split delivery weighted vehicle routing problem. *Expert Systems With Applications*, *40*(18), 7468-7477.

Tay, H. (2012). *Cycling infrastructure as a first mile solution for mass transit access in Singapore: A study of MRT ridership in Singapore towns* (Doctoral dissertation, Massachusetts Institute of Technology).

Vornhusen, B., & Kopfer, H. (2015, September). Emission vehicle routing problem with split delivery and a heterogeneous vehicle fleet. In *International Conference on Computational Logistics* (pp. 76-90). Springer, Cham.

Wang, X., Golden, B., Wasil, E., & Zhang, R. (2016). The min–max split delivery multi-depot vehicle routing problem with minimum service time requirement. *Computers & Operations Research*, *71*, 110-126.

Yan, S., Chu, J. C., Hsiao, F. Y., & Huang, H. J. (2015). A planning model and solution algorithm for multi-trip split-delivery vehicle routing and scheduling problems with time windows. *Computers & Industrial Engineering*, *87*, 383-393.