IJTE, Vol.9, No. 1, Summer 2021, pp. 503-520

Research Paper

Optimizing Algorithm for Allocating Passengers in Shared Taxis

Shahriar Afandizadeh Zargari[1], Samim Shakoori[2], Hamid Mirzahossein[3,*], Mehrdad Karimi[4]

Received: 2021.04.05 *Accepted: 2021.08.20*

Abstract

The issue of sharing vehicles has been riding since the '70s, but the advent of smartphones has made it a competitive choice to other transportation modes in recent years. The lack of restrictions on the movement of Internet-based passenger sharing systems leads to patrolling numerous personal vehicles in the network; this exacerbates congestion in high-traffic areas. On the other hand, the significant presence of circulating taxis and their non-optimal performance have disrupted the normal flow of traffic during peak hours and have led to an increase in travel time. This paper outlines a novel optimization algorithm for sharing repetitive and pre-planned trips. This algorithm is implemented on the midtown area network of Manhattan, New York, USA. Three scenarios were defined to simulate common services' status with the base scenario (do-nothing), which makes comparing possible with indicators such as distance travelled, and taxi occupancy ratio determined by passenger coefficient. Results of the first scenario - sending the nearest car - shows a decrease of 10.51%, the second scenario - allocating passengers to the nearest taxi - shows an increase of 10.16%, and finally the third scenario - the proposed algorithm - shows an increase of 25.56% in total mileage compared to the base scenario. Moreover, by defining Sharing Importance Factor (SIF) and using the proposed algorithm, it is possible to organize round-trip taxis, service repetitive and pre-planned trips, and significantly reduce the distance travelled throughout the network, and finally increase the passenger coefficient.

Keywords: Taxi Sharing, Allocation of Passengers, Optimization Algorithm, Manhattan

---

[*] Corresponding author. E-mail: mirzahossein@eng.ikiu.ac.ir

[1] Professor, PhD, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

[2] MSc, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

[3] Assistant Professor, PhD, Department of Civil Engineering – Transportation Planning, Imam Khomeini International University, Qazvin, Iran.

[4] MSc, School of Civil Engineering, Iran University of Science and Technology, Tehran, Iran.

1. Introduction

Nowadays, due to the increasing development of cities, the physical development of urban infrastructure is no longer considered a cost-effective and desirable option, and as a rule, this issue leads to restrictions on supply. One of the available options is the implementation of Transportation System Management (TSM) strategies to improve mobility in the Urban Transportation Network.

Introducing a Ride-Sharing and Carpooling system based on touring taxis which have a significant contribution in the city network, as a subset of Travel Demand Management (TDM), which instead of emphasizing on increasing supply (street capacity), focuses on the use of existing infrastructure capacity, is the main foundation of the forthcoming article.

Since a significant percentage of cars in daily traffic are taxis [Josep Maria Salanova, Miquel Estrada, Georgia Aifadopoulou, & Evangelos Mitsakis, 2011] (for example, 23% in Tehran metropolis - Statistics of the Deputy of Transportation and Traffic of Tehran Municipality 2017), most of which empty taxis (more than 38% in Beijing [Hao Wang, Kai Zhang, Junhua Chen, Zhifeng Wang, Guijun Li, & Yuqi Yang, 2018], this situation poses two main problems for two groups of users; taxi drivers (more empty kilometers mean less profit); and citizens (traffic congestion and more pollution). The reason for the accentuation on touring taxis is due to the fact that they do not use their full capacity; the distribution of empty taxis in the city (especially during rush hour) is often not optimal, which leads to increased waiting time or inaccessibility to the requested service for the passenger which causes consequently inefficient taxi performance [Josep Maria Salanova, Miquel Estrada Romeu, & Carles Amat, 2014]. On the other hand, the high capability of taxis in changing the function of single-passenger to multi-passenger has made them a reliable option as a complement to public transit mode [Jaeyoung Jung, R Jayakrishnan, & Ji Young Park, 2016]. In recent years, thanks to the advancement of smartphone technology, people can request a taxi anytime, anywhere. Social media has a major effect on this phenomenon [C. Zhang, M. Dong, K. Ota, & M. Guo, 2016].

The majority of New Yorkers rely on public transportation and taxi services. The taxi and livery system in New York City (NYC) is the fourth largest transportation provider in the United States of America (USA) [Regulatory Reform Team, 2014]. According to NYC statistics, there are about 60,000 vehicles on Internet services such as Uber and Lift, which is a significant number compared to about 13,500 taxis operating on the city network (according to N's Taxi and Limousine Commission (TLC) data). Otherwise stated, for every taxi in NYC, there are 4 private cars to share the ride! Thus, ride sharing using a localized distributed coordination between the riders and the driver [A. Manjunath, V. Raychoudhury, S. Saha, S. Kar, & A. Kamath, 2021]. The benefits of ride-sharing are high loading rate, high operating efficiency, and less traffic resources, and to ease the trouble of getting a taxi in urban [Yi Cao, Shan Wang, & Jinyang Li, 2021]. There are many variations of demand and supply for such services, which requires efficient ride-matching strategies to guarantee optimal allocation of trips to drivers and users [Jayita Chakraborty, Debapratim Pandit, Felix Chan, & Jianhong Xia, 2020]. Wang in his study critiques the issue of travel sharing from a new perspective. He says the growth in the use of travel sharing apps such as Uber and DiDi has worsened the situation in recent years. In this regard, ride sharing and the term ridesourcing are commonly used in transportation researches [Scarlett T. Jin, Hui Kong, & Daniel Z. Sui, 2019].

The pickup decision of a taxi driver is linked to the taxi supply and demand in the NYC. Both taxis supply excess [Walter Skok & Juan Antonio Martinez, 2010; Wei Zhai, Xueyin Bai, Zhong-ren Peng, & Chaolin Gu, 2019] and deficiency [David Carvalho Teixeira Da Costa & Richard De Neufville, 2012; Wei Zhai, Xueyin Bai, Zhong-ren Peng, & Chaolin Gu, 2019] could happen. To put it in another way, patrolling a large number of personal vehicles on the network has increased traffic load, wasted energy, and ultimately increased greenhouse gas emissions. This situation means that in most cases, the performance of Internet-based ride-sharing services should be no different from that of a single-passenger car [Yazhe Wang, Baihua Zheng, & Ee-Peng Lim, 2018]. The most ideal situation in the discussion of travel sharing is to complete the number of passengers in taxies in order to use their full capacity to reduce the volume of traffic at the network level. Given the above, presenting an algorithm for allocating users to circular taxis, collectively, considering the growing need of urban communities for pre-planned commuting trips (which is the main contribution in daily trips on city roads) is the main purpose of this study; in a way that, while benefiting from their maximum capacity, a significant improvement in performance indicators could be achieved. Minimization of distance traveled at the network level will be used as the main indicator to evaluate different scenarios. By providing a central management system based on the proposed algorithm: the service of repetitive and pre-planned trips is organized and control over the operation of circulating taxis in the city could be attained. Put differently, by allocating a group of people to a taxi, one can take advantage of travel sharing and prevent unnecessary traffic of single-passenger private cars in the city; on the contrary, the movement of empty taxis on the network is controlled too.

The main hypothesis of this study is the positive effect of sharing travel requests (duplicate and pre-planned) on improving network performance indicators (here total mileage) and the role of the proposed coefficient of importance of subscription in the process of allocating passengers to achieve it will be important. This hypothesis has been evaluated by implementing the algorithm using Python programming and NYC taxi data.

2. Literature Review

The outbreak of the 1973 oil crisis in the United States paved the way for alternative transportation options; the concepts of travel sharing and companionship were seriously raised during this period [Jørgen Aarhaug & Kåre Skollerud, 2014; Nelson D Chan & Susan A Shaheen, 2012]. On those days, Travel Demand Modeling (TDM) policies focused on pull and push strategies. Finding the best policy for the best location [Shahriar Afandizadeh Zargari, Hamid Mirzahossein, & Yi-Chang Chiu, 2016] and find the proper solutions like congestion pricing [Hamid Mirzahossein & Shahriar Afandizadeh Zargari, 2018] or ride sharing [Christopher K Brownell, 2013] are the achievements of this method. In following, Rayle et al. Examined the emerging phenomenon of travel sharing services and compared its performance with taxis and public transportation [Lisa Rayle, Danielle Dai, Nelson Chan, Robert Cervero, & Susan Shaheen, 2016]. According to their study, the main difference between taxis and travel sharing services is the waiting time. According to their findings, travel sharing services increase transportation options in dense cities (which are meeting the challenge of parking and public transportation restrictions); but the negative aspects of their activities in the city still need further investigation. From their point of view, the fact that travel sharing services make up for a percentage of public transport travel (especially for those who do not own a car) should be considered by city policymakers; therefore, one should be skeptical about the declining effects of these services on car use and car ownership.

Arhag and Sculrod categorized taxi services; Cruising taxis, Stand taxis, Dispatching markets, Contract taxis [Jørgen Aarhaug & Kåre Skollerud, 2014]. In most cases, taxis offer point-to-point service as part of public transportation, known as semi-private service.Shared taxis are common in many developing countries and several developed countries. This mode serves in different forms with different levels of laws and regulations in that area. By this definition, Salanova et al. State that models are an essential tool for decision makers in determining the main criteria for taxi services, such as fleet size, costs, etc. [Josep Maria Salanova, Miquel Estrada Romeu, & Carles Amat, 2014]; therefore, while reviewing the formulated models for modeling taxi services in urban areas, it has identified and analyzed the variables involved in the problem for the three main categories of taxi mode (static, touring and car delivery companies). According to their study, due to the complexity of formulating performance analysis variables for touring taxis, examining such taxis' performance is a priority.

In the concept of Car2work, Rego et al. Examined repetitive business trips and presented their proposed system and algorithms based on the concept of shared mobility [Robert Regue, Neda Masoud, & Will Recker, 2016]. They have proposed a new transportation system, which is to connect passengers to the workplace and guarantee a return trip, which is coordinated with the existing public transport network. In their system, travelers announce their trips in advance.

Lee et al. Introduced a two-stage passenger-sharing distribution system called taxi-pooling as a solution to increase the coverage of mass transit [Ker-Tsung Lee, Da-Jie Lin, & Pei-Ju Wu, 2005]. The study examines passenger sharing for a feeder system responsible for transporting passengers from multiple sources to one destination (mass transit stations as regional hubs). Their model sought to minimize the cost of operating a taxi and the travel time of a passenger.

Herbawi and Weber proposed a heuristic insertion algorithm for the Ride-Matching Problem with Time Windows (RMPTW) [Wesam Herbawi & Michael Weber, 2012]. Their proposed concept is like dynamic shared taxi subjects. The study proposed an objective function to minimize the total distance traveled by vehicles and maximize the number of adapted ride subscriptions within the maximum distance and time allowed. Genetic algorithms were used to solve the base, and innovative responses to new requests or suggestions received. In their system, riding offers were offered by personal drivers.

Al-Hassani et al. In a study presented their proposed service for ride sharing [Hadi El Hosni, Nourhan Farhat, Rakan Nimer, Nour Alawieh, Chadi El Masri, Mark Saroufim, Hassan Artail, & Joe Naoum-Sawaya, 2012]. The algorithm provided by them is responsible for identifying the most optimal taxi for the passenger. The algorithm allocates the best available taxi to the ride request using the concept of incremental cost, in this algorithm, after registering a new request in the system, the taxi with the least additional cost will be assigned to the new request.

Kramer and Krueger tested the efficiency of passenger sharing services versus taxis in their study [Judd Cramer & Alan B Krueger, 2016]. The basis for their comparison was mileage (while the passenger was in the car), defined as the capacity utilization rate. Their surveys in several cities show that the index defined for passenger sharing services (in this study, Uber) is significantly higher than the taxi service.

Factors that contribute to this occurrence: 1- Optimal passenger-driver adaptation algorithm, 2- Larger scale of passenger carrier companies (here Uber), 3- Inefficient taxi regulations, 4- Flexible workforce supply and pricing models which adapt to the supply-demand pattern of these companies throughout the day.

Komel et al. Solved the problem of taxi distribution to passengers using the Stable Marriage Problem (also Stable Matching Problem) [Michal Kümmel, Fritz Busch, & David ZW Wang, 2016]. This solution strategy contrasts with the conventional solution method, which responds to travel requests based on system registration priority. Komel further states that the main advantage of the stable marriage algorithm is the simultaneous (parallel) allocation of taxi travel requests, which improves taxi performance in all observed indicators (number of passengers transferred, distance traveled, passenger waiting time, fare and at the same time driver profit) follows.

Wang et al. Studied the effects of ride sharing in Singapore [Yazhe Wang, Baihua Zheng, & Ee-Peng Lim, 2018]. In their paper, they analyzed the effects of passenger sharing among those who had the same origin and destination when applying for a taxi in Singapore simultaneously by providing a simple but practical framework. Their proposed solution helped reduce users' costs, waiting time, and travel time during peak demand periods, and increased taxi drivers' revenue by serving multiple requests simultaneously. Their results show that passenger sharing can increase the service capacity of Singapore taxis by 20-25%; and reduce waiting times for passengers during peak hours. Additionally, they reduce the mileage by 2-3 km on an average taxi ride, which is about 20-30% of the average riding distance. As mentioned in the literature review, with the increasing growth of Internet taxi services, the provision of optimal fleet allocation algorithms in a manner that provides a central management system based on the proposed algorithms, trying to avoid unnecessary traffic in the city.

Based on the reviewed literature and as the optimizing algorithm for taxi distribution could be beneficial for autonomous taxi services [Tatiana Babicheva & Wilco Burghout, 2019], the following is the method proposed in this paper to improve the existing algorithms.

3. Methodology

The proposed algorithm for optimal allocation of passengers at specified time intervals (predefined time windows) is introduced in this section. Regarding time windows, it should be noted that, in small time windows, due to the small number of registered applications, the synergy of trips is reduced; on the other hand, by increasing the time window too much, the probability of canceling the synergy will increase. Jung et al. in their study considered a time window of 15 minutes [Jaeyoung Jung, R Jayakrishnan, & Ji Young Park, 2016]. Baren et al. have proposed a 10-minute time window, arguing that travelers have to wait in other situations, such as at a taxi stand or on the side of the road nevertheless [Benjamin Barann, Daniel Beverungen, & Oliver Müller, 2017].

In this article, after the passengers have registered their request in the specified time window, they are assigned to the nearest active taxi in the system, which is within its search radius; at this stage, the cost function, and a series of constraints step into the problem to ensure the optimal allocation of the final allocation. Once the taxi and the passengers assigned to it are identified, taxi routing is on the agenda, so that in total, the shortest distance is covered by the taxi. Simply put, the effect of passengers' permutation on the origin and destination has been seen. This ensures less mileage at the network level.

The Proposed algorithm can be classified as Greedy Algorithm. In the greedy method, achieving each step's goal is independent of the previous and next step. This means, at each stage to reach the goal, regardless of what choices were made in the previous stages, and what choices the current choice may lead to, the choice that seems to be the best choice may be made. The greedy method adds elements to the answer set every step, checks the constraints condition, and adds the next element if there is no problem. The algorithm is completed, and it is presented as an optimal answer set after meeting a specific final condition or if it is not possible to add another element to the set of answers. At each step, the greedy method selects one element from the set of elements available. The final element set includes this element as part of the solution to the problem. Steps for solving the problem [Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, & Clifford Stein, 2009].

- Select Procedure: An element is selected to be added to the answer set during this step. The criterion or procedure for selecting an element to add is its value. The most valuable value element is selected based on the type of problem.
- Feasibility Check: After selecting an element greedily, it should be checked whether it can be added to the previous set of answers. In some cases, the addition of an element violates one of the fundamental conditions of the problem. If adding this element does not violate any conditions, it will be added; otherwise, it will be left out, and another element will be selected based on the first step. The algorithm will end if there is no other element to select.
- Solution Check: During each step after completing step 2 and adding a new element to the answer set, it should be checked whether the desired answer was achieved? Once the final answer has not been obtained, the cycle continues in the next steps.

In this algorithm, allocation operations are performed statically; put the matter another way, the location of the requests is already recorded in the time window, and the allocation and routing is done based on it. The process which is introduced in Figure 1 is as follows:

The set P includes all requests (r) and active taxis (t); considering the effect of passenger permutation at the origin, the combination (c) that has the lowest cost and meets all the constraints (taxi search area and detour factor) is selected $\left((r_c, t), \forall c \in (3,4)\right)$. After that, the combination of requests r_c and taxi t are removed from the P set, and this process is repeated until no possible allocation remains.

Once the taxi and its passengers are identified, taxi routing will be on the agenda; the boarding and disembarking of passengers at the origin and destination, respectively, are done in such a way that the taxi covers total distance. It is important to note that the proposed algorithm emphasizes on the completion of the taxi capacity at the beginning of the trip. This is done by entering a parameter called the Sharing Importance Factor (SIF).Besides, according to the one-to-one, one-to-many, or multi-to-multiple classifications that researchers pursue in relation to passenger and taxi adaptation by solving the passenger sharing problem [Phathinan Thaithatkul, Toru Seo, Takahiko Kusakabe, & Yasuo Asakura, 2015] , it should be mentioned that the approach of this article is multi-to-multiple. Thus, the origin and destination of travelers are not necessarily the same. In short, it can be placed. (The scenario based on the proposed algorithm, as the third scenario, is one of the contributions of this paper).

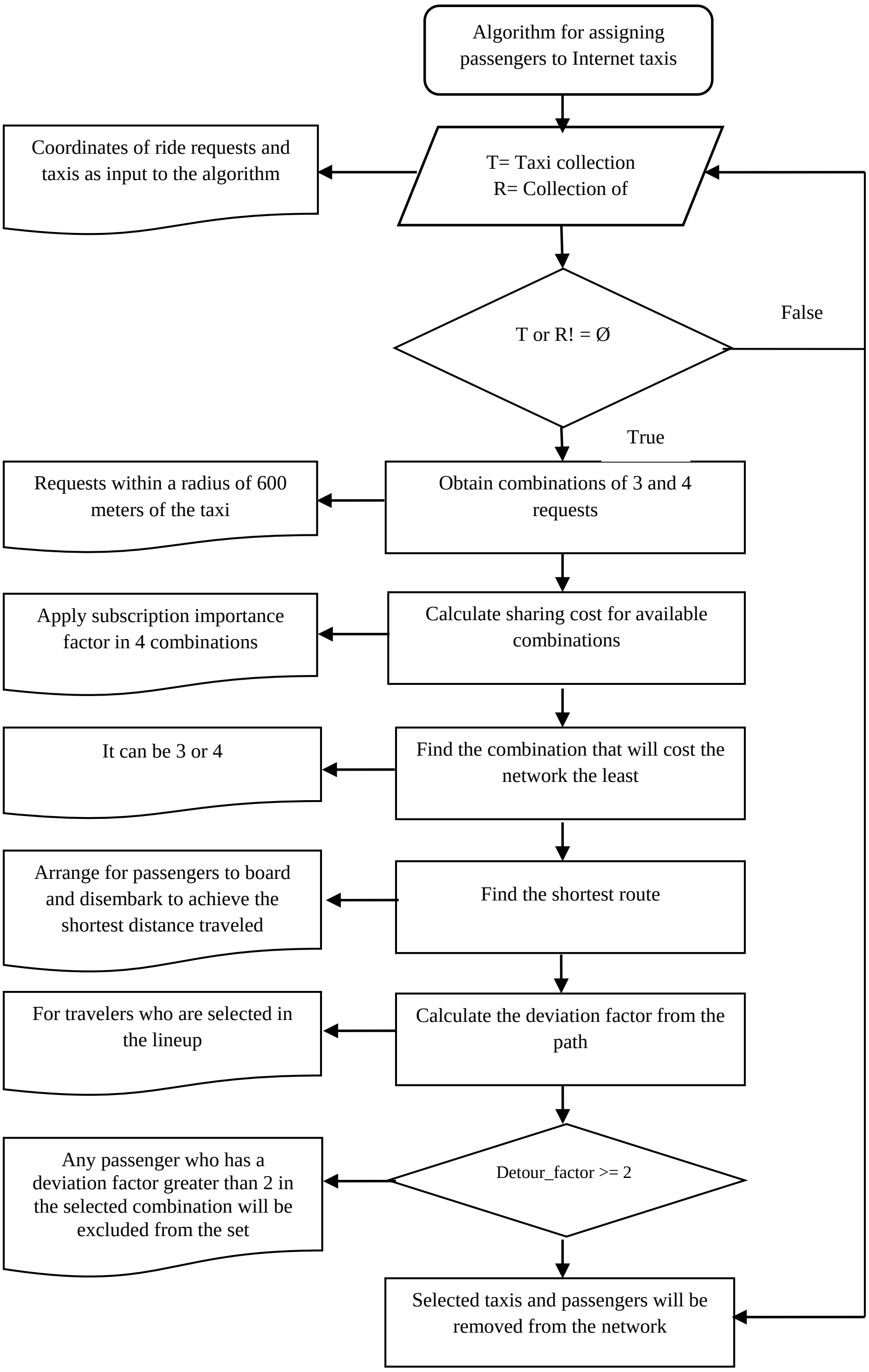

Figure 1. Flowchart of the proposed algorithm

3.1. Cost Function

The cost function is calculated in terms of the Euclidean Distance dimension (the shortest connecting distance between two points, using a straight line). In the proposed algorithm, the goal is to find the minimum amount among the passenger subscription candidates for each taxi. In the selection process, a combination of passengers (three or four) with the least cost to the network (less mileage) will be identified, while not violating the constraints of the issue.

$$Min\, sc_{taxi} + sc_{origin} + sc_{destination} \quad t_i, (i = 1, \ldots, m) \quad r_{jo}, (j = 1, \ldots, n) \quad r_{jd}, (j = 1, \ldots, n) \tag{1}$$

$$cor_o = \frac{\sum_{j=1}^{n} (r_{jo})}{k}, \forall k \in (3,4) \tag{2}$$

$$cor_d = \frac{\sum_{j=1}^{n} (r_{jd})}{k}, \forall k \in (3,4) \tag{3}$$

$$[\![sc]\!]_{taxi} = t_i - [\![cor]\!]_o \vee \tag{4}$$

$$sc_{origin} = \sum_{j=1}^{k} \left| r_{jo} - cor_o \right|, \forall k \in (3,4) \tag{5}$$

$$sc_{destination} = \sum_{j=1}^{k} \left| r_{jd} - cor_d \right|, \forall k \in (3,4) \tag{6}$$

The notation used to formulate the cost function is as follows:

t_i: Taxi coordinates i

r_{jo}: Coordinates of origin of the passenger j

r_{jd}: Coordinates of passenger destination j

cor_o: Coordinates of the passenger origin center

cor_d: Coordinates of the center of passenger destinations

sc_{taxi}: Taxi-related sharing cost in terms of Euclidean distance $t_i$from cor_o

sc_{origin}: The sharing cost related to travelers at the origin in terms of Euclidean distance $r_{jo}$from cor_o

$sc_{destination}$: The sharing cost related to travelers at the destination in terms of Euclidean distance $r_{jd}$from cor_d

3.2. Constraints

In the category of shared riding, people (both drivers and users) face a trade-off scenario. In the case of shared taxis, which is the subject of this study, one of the situations in which users examine this balance is the issue of detour. In other words, every user has a predefined threshold in mind. In a study [Jaeyoung Jung, R Jayakrishnan, & Ji Young Park, 2016], this value was considered equal to 2; this means that if the detour factor for a trip is more than 2, users are not willing to do that trip. In this article, this means that due to the position of taxis in the network, the passenger will not be included in the shared transportation system. Similarly, this issue can be examined from the network perspective, with an increase in the number of trips which exceed this threshold, the defined index (distance traveled on a network scale) strays from the optimal state. That is to say, the proposed algorithm does not emphasize on the synergy of all users. Therefore, users who do not have a taxi at the end of the allocation issue; they could pay additional fees and travel solitarily. Alternatively stated, this limitation guarantees a high distance for the passenger (between the origin and the destination) and prevents excessive diversion, which is achieved due to the simultaneous allocation of more passengers to one vehicle. In addition, to reduce the search space of taxis and prevent situations in which diversion from the path of users is out of the normal state, a kind of initial classification of users is needed. In this way, by imposing a restriction on the search radius of taxis (600 meters), it is possible to examine the composition of users whose case has the potential to exceed the defined limit for deviating from the route.

detour factor = (mileage traveled by the passenger in shared mode) / (mileage traveled by passenger in basic mode)

Basic mode: The distance that the passenger travels individually from the origin to the destination (in simpler terms, the shortest route between the origin and the destination of the passenger).

3.3. Sharing Importance Factor (SIF)

One of the most innovative approaches in this paper has been the definition of SIF. Since the cost function (on average) is obtained for 3-person combinations less than 4-person combinations, the preference in the selection stage of the selected combination is with 3-person combinations, and it is not possible to compete for 4-person combinations. This means that the potential and capacity of taxis are not fully utilized. In order to increase the chances of 4-person allocation to reduce taxi traffic as much as possible, a coefficient called the subscription importance coefficient (due to the higher importance and value that this coefficient gives to the selection of 4-person allocations in the taxi and passenger search process) enters the problem-solving process. SIF depends on three parameters: network structure, taxi position, and passenger position. Improvement of the proposed indicators (less mileage and higher passenger coefficient) than other scenarios can be expected from considering SIF. In this paper, the passenger coefficient is considered as the car occupancy ratio without considering the driver. To find the optimal value of this coefficient, it is necessary to consider an optimization problem in the problem-solving process.

The following is the selection process of the selected combination, with and without considering the coefficient of the importance of subscription.

$withoutSIF$:

$$sc4 = \frac{sc_{taxi} + sc_{origin} + sc_{destination}}{4} \tag{7}$$

$$sc3 = \frac{sc_{taxi} + sc_{origin} + sc_{destination}}{3} \tag{8}$$

$$sc4 > sc3 \rightarrow sc3willbeselected \tag{9}$$

$withSIF$:

$$sc4 = \frac{sc_{taxi} + sc_{origin} + sc_{destination}}{4} \tag{10}$$

$$sc3 = \frac{SIF * sc_{taxi} + sc_{origin} + sc_{destination}}{3} \tag{11}$$

$$sc4 < sc3 \rightarrow sc4willbeselected \tag{12}$$

3.4. Routing

After selecting and assigning the passengers to the desired taxi, to find the best possible route for boarding and disembarking the passengers, the routing is done in such a way that the minimum distance is traveled in the total taxi trip. In other words, it is necessary to put the effect of passenger permits at the origin and destination into account, regardless of the order of the registration of requests in the routing issue. In this part, a simpler case of the Traveling Salesman Problem (TSP) should be solved [Thomas H Cormen, Charles E Leiserson, Ronald L Rivest, & Clifford Stein, 2009] (Routing problem with emphasis on the order of boarding at the origin and disembarking at the destination). Undeniably, one of the differences between the routing problem in this article and the traveling salesman problem is that it does not return to the starting point. Additionally, because the search space in this issue is small (maximum 5 nodes at the origin and 4 nodes at the destination; as a result, the permutations at the origin and destination are calculated separately, which in turn contributes to the smaller search space.). The approach taken in this case is to implement a precise algorithm to test all possible permutations to find the shortest path. If there is no taxi to respond to the registered request, the request is rejected and the passenger tries to register the request again. If there is no request within a certain time frame, no allocation will be made.

3.5. Scenario Planning

To compare the results of passenger allocation based on the proposed algorithm, several scenarios are defined in order to simulate the status of common services among the possible scenarios; 1. Nearest Vehicle Dispatch (NVD) based algorithm scenario which has the most common use in real applications; 2. Scenario based on allocating passengers to the nearest taxi cooperatively; 3. Scenario based on the proposed algorithm in this article and finally 4. Do-nothing scenario, which is traveling with a personal car, is considered a basis for comparing scenarios. To illuminate the first and second scenarios, it should be noted that currently the first scenario (the nearest car delivery algorithm) has the largest share in online taxi delivery systems [Jaeyoung Jung, R Jayakrishnan, & Ji Young Park, 2016]. Therefore, in the first scenario, when the passenger registers his ride request in the system, the algorithm starts searching to find the closest car to the passenger's origin coordinates. In this case, the goal is to provide the answer as soon as possible. After selecting the nearest taxi, the optimal route for boarding and delivering the passenger is calculated. Since this algorithm considers only the shortest possible path, it does not require a complex sending algorithm. The purpose of defining this scenario (after this referred to as Scenario 1) is to consider the current situation in allocating passengers to Internet taxis. In this case, the vehicle is assigned to only one request (the relevant request can include one or more people) and does not consider other requests' status at the passenger allocation stage. This means ignoring and not using the sharing potential, which can be a good criterion in the non-sharing mode in the results comparison phase. In the second scenario, after the passengers register their request in the designated time window, they are assigned to the nearest active taxi in the system, which is within its search radius. Appears obviously, the cost function and restrictions do not apply to the selection of passengers. Alternatively stated, the taxi is trying to provide service with maximum occupants and capacity. The order of boarding and disembarking passengers is in such a way that it brings the shortest distance to the taxi in total. The motive of defining this scenario (hereinafter referred to as Scenario 2) is to describe a situation in which there is no cost and constraint function. Otherwise stated, the question is, what is the result of the allocation is made without considering a specific criterion in the passenger selection stage. Consequently, the algorithm of this type of allocation is simple and simultaneously easy to implement

4. Results

This section examines the potential of the proposed algorithm for shared allocation of users, using New York taxi data and the Manhattan Midtown area network as a case study.

4.1. Case study

To validate and evaluate the potential of the proposed algorithm, the data set used in this article is data collected by the NYC City Taxi & Limousine Commission (TLC) and made available to the public. This dataset contains information on the 2009 trips made by green and yellow taxis, which include several hundred million trips. The data set includes information on taxi license, driver's ID, type of taxi meter, start and end time of each trip, number of passengers per trip, travel time, distance traveled in miles, GPS coordinates of boarding and disembarking, fare, additional costs, taxes, tips, duties, and total expenses paid. 6/15/2016 was chosen as a normal day for review, so there was no specific event in the desired period. Meanwhile, trips to the southwest to the coordinates [40.739814, -74.000233] and the northeast to the coordinates [40.762440, -73.960054] and were limited to the period 7:30 to 19:30 (MIDTOWN MANHATTAN area - consisting of 33124 thousand lines) .73.27% of the trips in the desired period were made as a single passenger; This shows the high potential of taxi sharing to improve network flow and side effects (lower fuel consumption, reduced pollution, etc.). The average distance traveled is 1599.19 meters, while about 74.66% of the trips show less than 2 km.9.93% use of taxis for short trips (less than 800 meters) is also significant. Figure 2 shows the status of the source-destination pair.

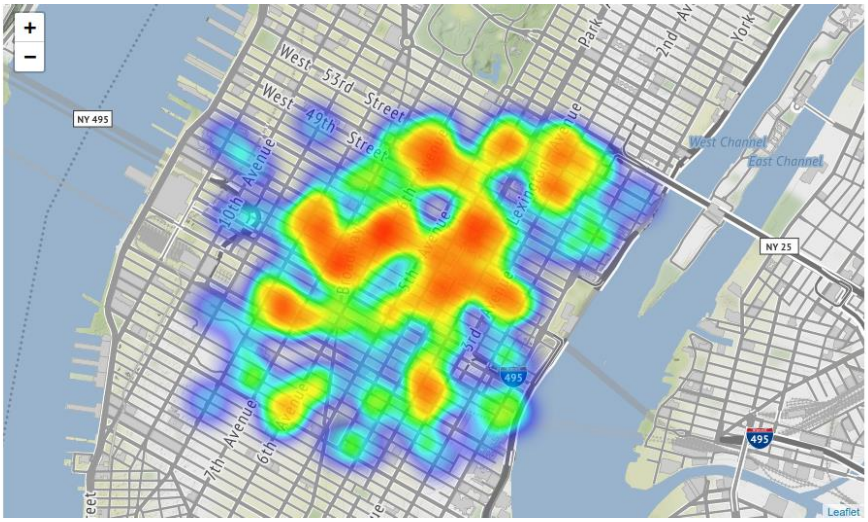

Figure 2. Heat Map (Heat Map) Location of the source-destination pair in the region

4.2. Data Preprocessing

In this paper, in order to increase the quality of input data and consequently improve the output results, data cleaning operations (Data Cleaning) were applied as one of the main techniques to remove invalid inputs (eg data with similar origin and destination, without passenger travels.). In this regard, shorter than 400 meters trips and unreal length trips (for example, longer than two hours) were removed from the data set. Raw data preprocessor eliminated 5% of all trips. The data were then clustered and repeated paths were found using the Apriori algorithm [Mohammed Al-Maolegi & Bassam Arkok, 2014].

4.3. Clustering

Inertia or within-cluster sum-of-squares was used to determine the optimal number of clusters in segregated clustering. Based on this index, the data were divided into 10 clusters. As well, the average of the total profile index (Silhouette) as a criterion for evaluating the correctness of clustering has been calculated to be equal to 0.25. After determining the number of clusters, in two-hour intervals, the clustering algorithm was applied to users’ travel data and the users' travel distribution and arrangement was obtained in the form of category centers.

4.4. Implementing Apriori Algorithm

The Apriori algorithm is used to explore the relationships between transactions in a data set.In this regard, after constructing a data set containing transit nodes (routing between source-destination pairs), while determining the minimum backup, the combination of links which have more than 2% backup was calculated by the algorithm and by performing a recursive process of frequent sub-routes which are at least 1000 meters long, during 7:30 to 19:30 with intervals of two hours, are marked by the network of passages. Figure 3 shows these routes for 6 different intervals.

7:30 to 9:30 -1559 meters 9:30 to 11:30 -1285 meters 11:30 to 13:30 -1222 meters

13:30 to 15:30 -1311 meters 15:30 to 17:30 -1450 meters 17:30 to 19:30 -1121 meters

Figure 3. Recurring routes at six different times of the day

4.5. Implementing Scenarios

The main philosophy of the proposed service is based on benefiting the optimum capacity of taxis. This approach reduces the number of taxis and their traffic at the road network level. As much as the restrictions allow, each taxi must start with a full passenger (the priority is four-passenger taxi), which is possible by reducing the number of single, double, and triple taxis. As shown in Figure 4, changes in the number of active taxis for different SIFs are noticeable, and as the SIF increases, the chances of four-person allocations increases (which is evident in the decrease in the number of three-passenger taxis), while responding to most travel requests. (71%), the number of required taxis is reduced in overall. This condition converges at SIF = 1.4. This result can also be seen in changes in the passenger coefficient for different SIFs (Figure 5); meaning, as the number of taxis required decreases, the passenger coefficient increases. As a result, the efficiency of the shared taxi system will increase.

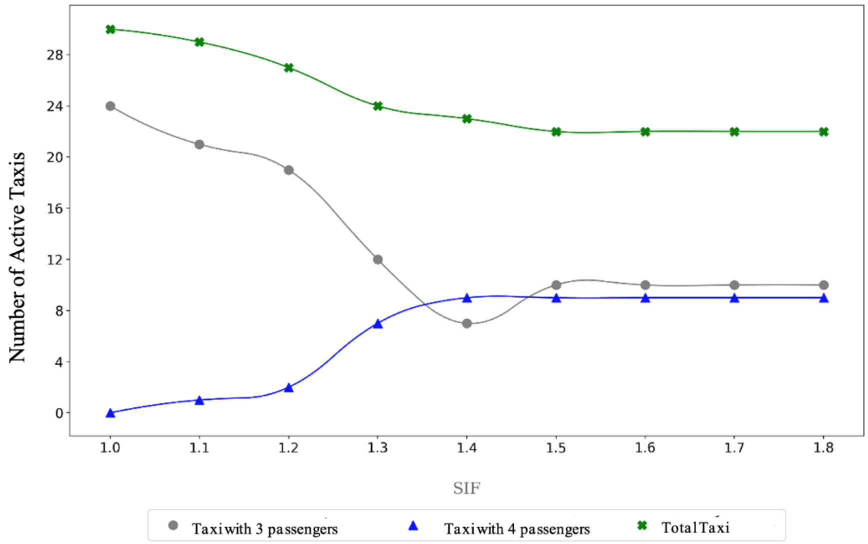

Figure 4. Changes in the number of active taxis with increasing SIF

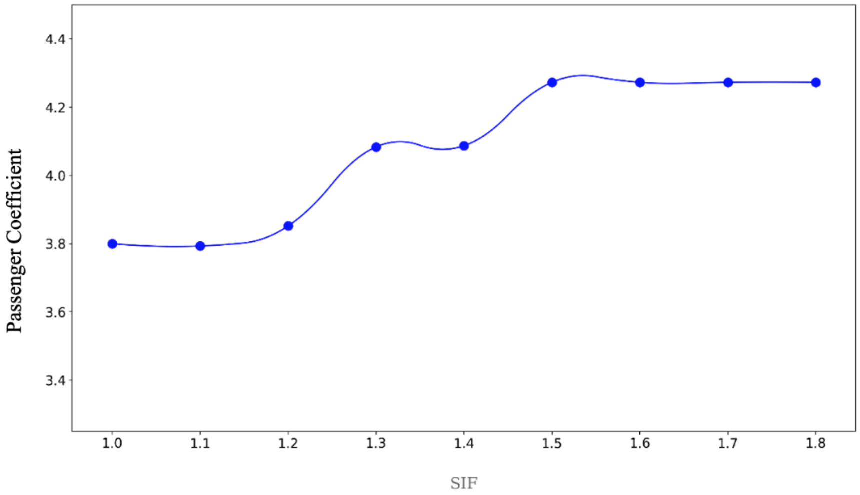

Figure 5. Changes in passenger coefficient with increasing SIF

One of the main goals was to improve the mileage index on a network scale. As shown in Figure 6, as expected, the value of this index decreases with increasing SIF, and at SIF = 1.4 it assumes the minimum value. The situation can be interpreted as declared that, with SIFs less than 1.4, due to less stringency on completing the taxi capacity at the beginning of the mission, there is no improvement in this index. However, with SIFs greater than 1.4, overemphasizing, selected four-person combinations, and increasing its weight in the selection phase between candidate combinations will increase the mileage index value. This increase is due to the excessive removal of users from the shared system. Not only can that the status of this index, in comparison with other scenarios, be seen in Figure 7. The distance traveled in SIF = 1.4 for the scenarios of doing nothing, 1, 2, and 3 are 164234, 181547, 147537, and 122241 meters, respectively. The three scenarios show the perfection of -10.51, 10.16, and 25.56% in total mileage, respectively, compared to the Do-Nothing scenario; according to the results, scenario 3 (based on the proposed algorithm) can be determined superior.

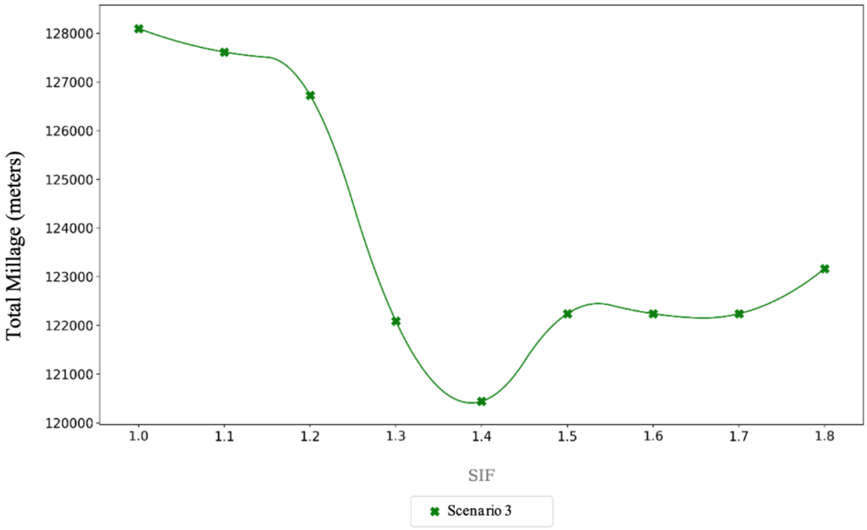

Figure 6. Changes in the total mileage in Scenario 3 with increasing SIF

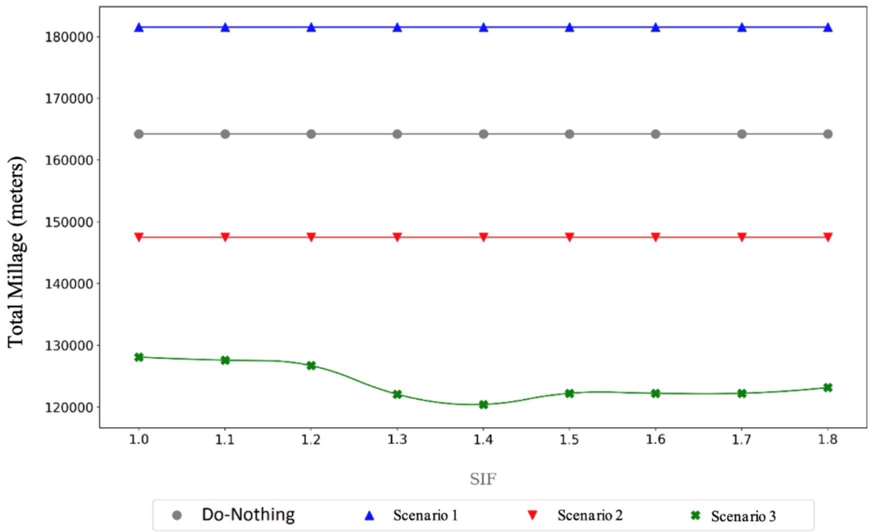

Figure 7. Changes in the total mileage in different scenarios with increasing SIF

Following the mileage index, its average value per user can be obtained (Figure 8). A decrease in this index was expected with an increase in SIF. However, a noteworthy point occurs at SIF = 1.4; Where SIF is not yet involved, and in its absence, the average value of the mileage index is even worse than Scenario 2.

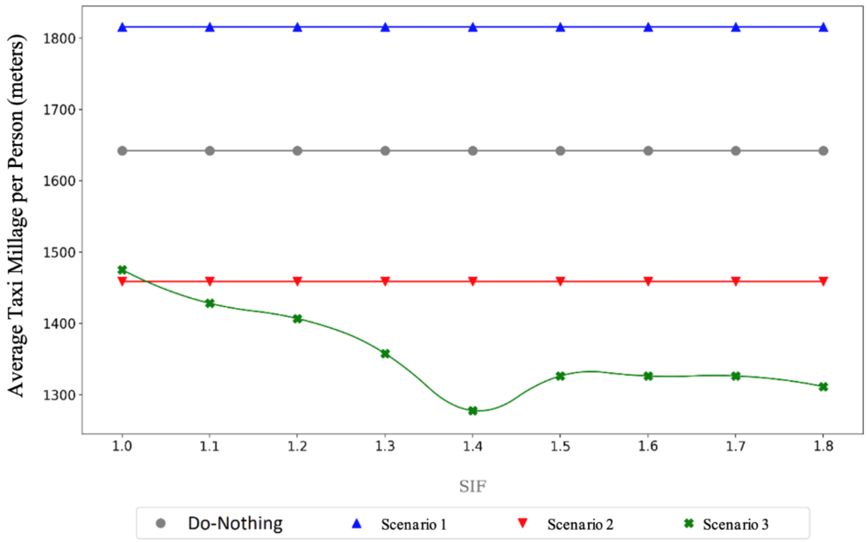

Figure 8. Average changes in taxi mileage per person in different scenarios with increasing SIF

Since all the indicators defined in this study are defined from a network perspective, it is possible that some of the requests will not be shared; because in a certain period of time with the known source-destinations location, their joint movement reduces the optimal network performance (mileage). In other terms, traveling in a private car or as a single passenger in that particular situation is in the network's interest. This implies that SIF has the task of identifying requests that the network does not need to assign shared services. Thus, by eliminating 13% of users in the current allocation issue, the system will operate optimally. It should also be noted that 16% of users are not in the area of taxi response (Figure 9).

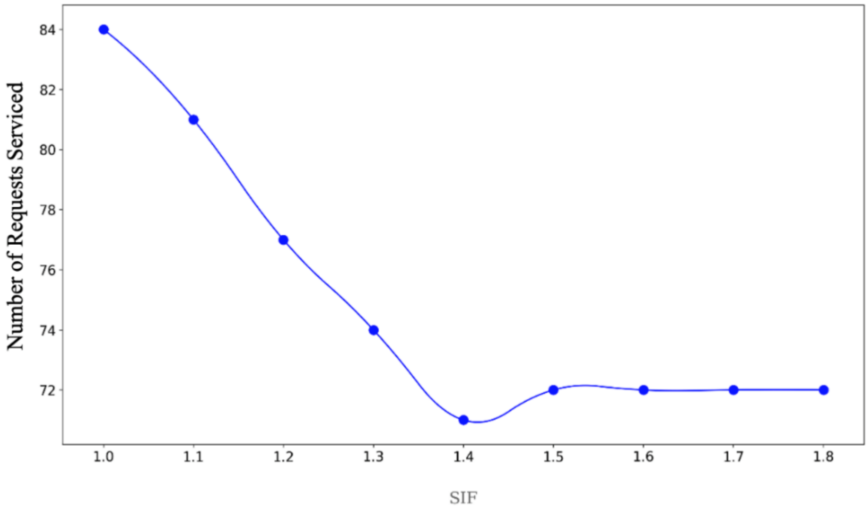

Figure 9. Number of requests serviced with increasing SIF

The detour factor changes in Figure 10 show the interference between the User Equilibrium perspective and the System Optimal. As in the case of traffic allocation (traffic flow distribution), users must work together to minimize network costs in order to balance the system and achieve the optimal system; here, too, several people have to make sacrifices and travel longer distances while sharing the trip to help improve the network status (reduce mileage). The increase in detour factor in SIF = 1.4 can be explained by the increase in the number of users who travel with more deviations. In the optimal SIF, the amount of self-sacrifice of users is in such a way that less than that amount will lead to a decrease in the desirability of the index, and more than that amount will mean unnecessary self-sacrifice from the system's point of view.

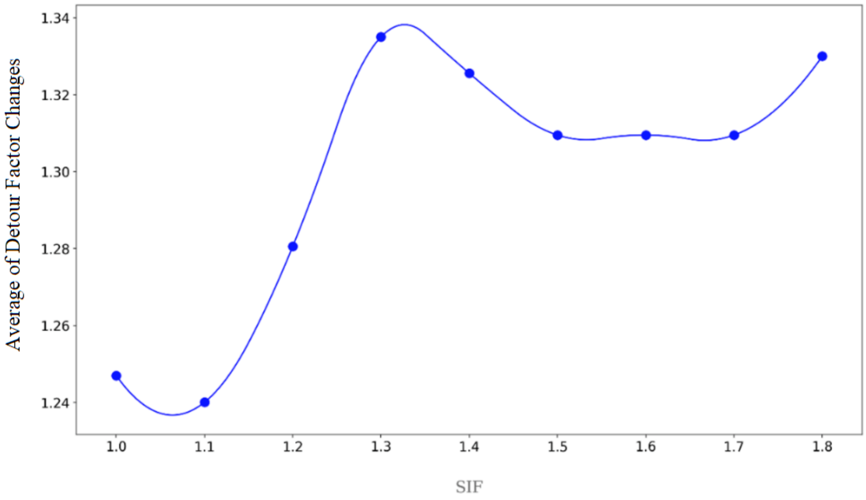

Figure 10. Detour factor changes with increasing SIF

In the discussion of the reasons and the importance of adding the SIF coefficient to the process of solving the proposed algorithm, the following can be mentioned: If SIF does not enter the problem (SIF = 1.4), no 4-person assignment will be made. In other terms, the capacity of taxis is not fully utilized. In addition, in the absence of SIF, since the maximum capacity of taxi sharing is not used, it can leave the mileage index unchanged compared to other scenarios. To maximize system performance, SIF also helps to identify requests that do not require shared services and instead offers individual travel at an additional cost. Finally, as the interpretation of the results shows, the scenario based on the proposed algorithm (Scenario 3) was selected as the best scenario. In this scenario, while entering the SIF coefficient and examining a series of possible combinations in the basic search space, the best allocation is made to maximize the system optimization (minimization of mileage). Similarly, eliminating the roaming taxis and circulating cars, which can be achieved with the help of the central management system; results in a reduction in taxi traffic at the network level.

5. Conclusions

This paper outlines today's urban communities' growing need to perform repetitive trips by presenting a novel optimization algorithm. It was shown whether such travel requests could be serviced jointly while improving network performance indicators. As illustrated, based on the proposed algorithm the third scenario (Scenario 3) was selected as the best one. This scenario improved the allocation of passengers to taxis and, consequently, the state of the network compared to conventional services, which were seen in the form of the introduced scenarios, by using taxis more efficiently. This algorithm can also be used for developing metropolises such as Tehran because, like most developing cities, it suffers from imbalances in travel demand and transportation supply. Therefore, such solutions can, in addition to organizing the taxi service, also take on the role of managing the travel demand. The results and achievements of this research are summarized as follows:

- The proposed algorithm's main application could be observed in a central passenger allocation system based on the algorithm; to serve repetitive and pre-planned trips jointly (This issue can also be raised in the form of organizing circling taxis).
- Considering the less used algorithms in the field of transportation (Apriori clustering), it was tried to present a platform for providing solutions for a more appropriate distribution of taxis by identifying the distribution and arrangement of trips, which is considered a kind of data preprocessing.
- Introducing Sharing Importance Factor (SIF): This coefficient, as the innovation of this article, has the task of identifying a set of requests that are not required to be shared by the network. By giving importance to the higher occupant ratio, it prefers to reduce the total distance of the network due to the diversion of individuals. Therefore, its presence in the solution process improves the network performance (here mileage).
- Reduced mileage, and possibly the consequent network-time travel time compared to conventional ride-sharing services achieved using the proposed algorithm.
- Diminution of traffic volume at the network level due to the increase of the taxi's passenger coefficient will increase the efficiency of the shared taxi system and significant economic and environmental benefits.

Data Availability Statement

Some or all data, models, or codes that support the findings of this study are available from the corresponding author upon reasonable request. Also, the relevant code and data can be found at www.github.com/samim-sh/shared-taxi

6. References

- Aarhaug, Jørgen, & Skollerud, Kåre. (2014). Taxi: different solutions in different segments. Transportation Research Procedia, 1(1), 276-283.

- Afandizadeh Zargari, Shahriar, Mirzahossein, Hamid, & Chiu, Yi-Chang. (2016). Quick Link Selection Method by Using Pricing Strategy

Based on User Equilibrium for Implementing an Effective Urban Travel Demand Management. Promet-Traffic&Transportation, 28(6), 605-614.

- Al-Maolegi, Mohammed, & Arkok, Bassam. (2014). An improved apriori algorithm for association rules. arXiv preprint arXiv:1403.3948.
- Babicheva, Tatiana, & Burghout, Wilco. (2019). Empty vehicle redistribution in autonomous taxi services. EURO Journal on Transportation and Logistics, 8(5), 745-767. doi:https://doi.org/10.1007/s13676-019-00146-5

- Barann, Benjamin, Beverungen, Daniel, & Müller, Oliver. (2017). An open-data approach for quantifying the potential of taxi ridesharing. Decision Support Systems, 99, 86-95.

- Brownell, Christopher K. (2013). Shared autonomous taxi networks: An analysis of transportation demand in nj and a 21st century solution for congestion. Princeton University, April.

- Cao, Yi, Wang, Shan, & Li, Jinyang. (2021). The Optimization Model of Ride-Sharing Route for Ride Hailing Considering Both System Optimization and User Fairness. Sustainability, 13(2), 902. Retrieved from https://www.mdpi.com/2071-1050/13/2/902

- Chakraborty, Jayita, Pandit, Debapratim, Chan, Felix, & Xia, Jianhong. (2020). A review of Ride-Matching strategies for Ridesourcing and other similar services. Transport Reviews, 1-22.

- Chan, Nelson D, & Shaheen, Susan A. (2012). Ridesharing in North America: Past, present, and future. Transport Reviews, 32(1), 93-112.

Cormen, Thomas H, Leiserson, Charles E, Rivest, Ronald L, & Stein, Clifford. (2009). Introduction to algorithms: MIT press.

- Cramer, Judd, & Krueger, Alan B. (2016). Disruptive change in the taxi business: The case of Uber. American Economic Review, 106(5), 177-182.

- El Hosni, Hadi, Farhat, Nourhan, Nimer, Rakan, Alawieh, Nour, El Masri, Chadi, Saroufim, Mark, Artail, Hassan, & Naoum-Sawaya, Joe. (2012). An optimization-based approach for passenger to shared taxi allocation. Paper presented at the SoftCOM 2012, 20th International Conference on Software, Telecommunications and Computer Networks.

- Herbawi, Wesam, & Weber, Michael. (2012). Modeling the multihop ridematching problem with time windows and solving it using genetic algorithms. Paper presented at the 2012 IEEE 24th International Conference on Tools with Artificial Intelligence.

- Jin, Scarlett T., Kong, Hui, & Sui, Daniel Z. (2019). Uber, Public Transit, and Urban Transportation Equity: A Case Study in New York City. The Professional Geographer, 71(2), 315-330. doi:10.1080/00330124.2018.1531038

- Jung, Jaeyoung, Jayakrishnan, R, & Park, Ji Young. (2016). Dynamic shared- taxi dispatch algorithm with hybrid- simulated annealing. Computer- Aided Civil and Infrastructure Engineering, 31(4), 275-291.

- Kümmel, Michal, Busch, Fritz, & Wang, David ZW. (2016). Taxi dispatching and stable marriage. Procedia Computer Science, 83, 163-170.

- Lee, Ker-Tsung, Lin, Da-Jie, & Wu, Pei-Ju. (2005). Planning and design of a taxipooling dispatching system. Transportation research record, 1903(1), 86-95.

- Manjunath, A., Raychoudhury, V., Saha, S., Kar, S., & Kamath, A. (2021). CARE-Share: A Cooperative and Adaptive Strategy for Distributed Taxi Ride Sharing. IEEE Transactions on Intelligent Transportation Systems, 1-17. doi:10.1109/TITS.2021.3066439

- Mirzahossein, Hamid, & Zargari, Shahriar Afandizadeh. (2018). A Combined Model of Congestion Toll Pricing Based on System Optimization with Minimum Toll. Tehnički vjesnik, 25(4), 1162-1168.

- Rayle, Lisa, Dai, Danielle, Chan, Nelson, Cervero, Robert, & Shaheen, Susan. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transport Policy, 45, 168-178.

- Regue, Robert, Masoud, Neda, & Recker, Will. (2016). Car2work: Shared mobility concept to connect commuters with workplaces. Transportation research record, 2542(1), 102-110.

- Salanova, Josep Maria, Estrada, Miquel, Aifadopoulou, Georgia, & Mitsakis, Evangelos. (2011). A review of the modeling of taxi services. Procedia-Social and Behavioral Sciences, 20, 150-161.

- Salanova, Josep Maria, Romeu, Miquel Estrada, & Amat, Carles. (2014). Aggregated modeling of urban taxi services. Procedia-Social and Behavioral Sciences, 160, 352-361.
Team, Regulatory Reform. (2014). Case Study: New York City Taxis Regulatory Reform for the 21st Century City.

- Thaithatkul, Phathinan, Seo, Toru, Kusakabe, Takahiko, & Asakura, Yasuo. (2015). A passengers matching problem in ridesharing systems by considering user preference. Journal of the Eastern Asia Society for Transportation Studies, 11, 1416-1432.

- Wang, Hao, Zhang, Kai, Chen, Junhua, Wang, Zhifeng, Li, Guijun, & Yang, Yuqi. (2018). System dynamics model of taxi management in metropolises: Economic and environmental implications for Beijing. Journal of environmental management, 213, 555-565.

- Wang, Yazhe, Zheng, Baihua, & Lim, Ee-Peng. (2018). Understanding the effects of taxi ride-sharing—A case study of Singapore. Computers, Environment and Urban Systems, 69, 124-132.

- Zhang, C., Dong, M., Ota, K., & Guo, M. (2016). A Social-Network-Optimized Taxi-Sharing Service. IT Professional, 18(4), 34-40. doi:10.1109/MITP.2016.71