

MBE, 19(1): 1041–1057. DOI: 10.3934/mbe.2022048 Received: 24 August 2021 Accepted: 09 November 2021 Published: 29 November 2021

http://www.aimspress.com/journal/MBE

Research article

Dynamic balance between demand-and-supply of urban taxis over trajectories

Mingyang Liu^{1,*}, Junhao Han¹, Yushan Mei² and Yuguang Li¹

¹ College of Instrumentation and Electrical Engineering, Jilin University, Changchun 130012, China

² Changchun China Optical Science & Technology Museum, Changchun 130118, China

* Correspondence: Email: liumingyang@jlu.edu.cn.

Abstract: Urban taxi serves as an irreplaceable tool in public transportation systems. The balancing of demand-and-supply can be of significant social benefit, for which the equilibrium method for urban taxis, especially with dynamic trip demands, is not well studied yet. In this paper, we formally define the equilibrium problem and propose a coarse-grained dynamic balancing algorithm. It efficiently evaluates the trip demand distribution pattern and schedules supplies to more unbalanced regions. We first propose a density-based blocking algorithm to detect regions that are with more travel demands. A trip demand merging strategy is then proposed, which checks the correlation of trip demands to merge the trips into ones. To reduce the computation load, a lazy trip correlation strategy is devised to speed up the merging process. By calculating the defined balance factor, a scheduling algorithm is proposed to realize the trip merge and supply translocation based balancing approach. We evaluated our approach using a month of global positioning system (GPS) trajectories generated by 13,000 taxis of Shanghai. By learning the spatiotemporal distribution of historical taxi demand-and-supplies, we simulated an inflated trip demand platform. Tested on this platform with extensive experiments, the proposed approach demonstrates its effectiveness and scalability.

Keywords: demand-supply balancing; taxi trajectory; dynamic balance; urban traffic

1. Introduction

Taxi serves as a public transportation mode that cannot be replaced by neither the bus systems nor the subway transportation in urban environments. However, due to the urban commuting demand, most of the metropolis cities are facing with the problem of imbalance between taxi request and taxi supply. Especially in peak hours, taxi requests are typically much more than available taxi supply. Such imbalance lead to the resulting common phenomena that people who are hailing cabs always waste a lot of time on road sides for waiting an vacant taxi. It also means that potential riders compete for limited taxi resource nearby [1]. Meanwhile, instead of waiting at fixed locations, taxis drivers usually search around for riders, resulting in extensive fuel burning and unnecessary road space occupancy [2]. As the global information of taxi riders is not available to taxi drivers, the searching strategy could not be the optimal one [2]. There have some works that aims to extract the spatio-temporal travel patterns from the perspectives of both taxi supplies and the urban commuting demands. However, there has not been sufficient studies on the interplay between the taxi supplies and the urban commuting demands.

To address these issues of taxi demand-supply balance, we proposed a dynamic demand-andsupply balancing system, in which taxi riders could sense other riders' position nearby and originate a ridesharing, furthermore, regions that are dense of sufficient vacant taxis could be recognized and drivers could get the information of regions that are denser of taxi riders to decide their orientation. The main contributions of this paper include:

1) We first propose a novel density based layered blocking algorithm to learn the trip demand distribution pattern and detect regions that are dense of trip demands and small enough within walking distance. Furthermore, we define the notation of Balance Factor to describe the imbalance of each block that are dense of trip origins and destinations.

2) We present a trip merging and supply transition-based demand-and-supply balancing approach. A trip demand merging strategy is proposed in advance, which checks the correlation of trip demands originating from the same grid and in the same time slot to merge the trips into ones with acceptable additional incurred travel distance for the longer trip. To tackle the heavy computation load, a lazy trip correlation strategy is devised to speed up the merging algorithm. By calculating the defined balance factor, a supply transition scheduling algorithm is proposed to realize the balancing approach.

3) We evaluated the proposed approach in a practical setting by exploiting a real taxi trajectory dataset generated by 13,000 taxis in Shanghai during a period of one month. By learning the spatiotemporal distribution of historical demand and supply, we simulated an inflated trip demands platform. Extensive experiments on this platform have testified the effectiveness and scalability of our proposed method.

2. Related works

Related works can be categorized into two lines: the taxi demand-supply modeling and taxi service optimizing. The first one focuses on analyzing the travel pattern using the taxi trajectory data analysis, which also helps us understand the urban structure; another category aims to detect two types of regions, i.e., the over-served region and the under-served regions, which could provide insights for providing better taxi services.

2.1. The taxi demand-supply modeling

Urban trip demand modeling is an essential thesis for transportation system improving, city structure understanding, and regular traffic flow prediction. Yuan J. et al. propose a system prototype that aims to obtain the optimal routes in terms of driving time with two types of information, i.e., the derived traffic patterns and the knowledge of the driver behavior. Specifically, the proposed prototype system utilizes the extracted traffic patterns from historical traffic information and further incorporates

dynamic traffic context to predict future traffic conditions [3]. By using the taxi trajectory dataset, Liu et al. propose to utilize the typical network community detection method to explore the spatial structures of both commuting trips, and they further investigate the city structure based on the commuting trips [4]. Furthermore, the collection of urban taxicabs trajectories provides materials for analyzing and modeling the balance between taxi demand and supply. The studies in [5,6] adopt typical regression models, i.e., the Poisson model, the quasi-Poisson model, and the negative binomial model, to elucidate the taxi pick-ups based on a collection of taxi trip records during a time period of 10 months from New York city. The negative binomial model has been verified to be the most appropriate one for the over-dispersed count data after sufficient analysis. Yang C. et al. adopt another variable, i.e., the number of taxi drop-offs, to further represent the taxi supply dynamics, which is due to the consideration that each drop-off also denotes an available supply. Kang C. et al. have systematically studied the spatial patterns of an urban of taxis supply in Wuhan, China, where the taxi trajectories are used, and the modeling is based on the non-negative matrix decomposition [7]. They find that the taxi drivers follow a remarkably self-organized operation pattern in the supply regions, which perform as the reactions to the sectored distribution of daily commuting trips. It has been observed that taxi drivers tend to operate within a single specific service area and a small proportion of taxis drives throughout different urban areas, which serves as a significant shifting tool that connect the remote urban areas and satisfy the demand of long-distance trips. Generally, this study provides a significant insight for helping recognize the drivers' collective intelligence when they do not know the global information of taxi supply and demand. But it only considers the demand-supply status in the granularity of a district and disregard the unsupplied taxi demand.

Studies on taxi demand-supply level are necessary for improving taxi service, meanwhile, there still exist several challenges about how to detect the service disequilibrium. Firstly, there exists inherent uncertainties in urban taxi service systems. Such uncertainties mainly relate several aspects. One is that the actual time of a taxicab request is uncertain. Another aspect is that we do not know the requests that have not been satisfied. But such unsatisfied requests are quite important when we model the taxi service disequilibrium. Besides, the taxi location between two reported GPS coordination is lost, which makes the counts of available taxis are uncertain for that time interval. Another challenge is how to design a model that could detect the over-served regions, as well as the under-served regions. Moreover, the model should fully relate the taxi demand with the taxi supply. Huang et al, partitioned the researched region of Shanghai into equal-sized cells and they assumed that the pickup time of requests follows a normal distribution, and a normal distribution is also used for estimating the unknown positions between two trajectory point [8]. Then they proposed a probabilistic counter method to count the observed requests and taxis in a time interval <t1, t2>. Afterward, they proposed a reasonable estimate of the unserved request by simulating the random injection of requests and attempting to match them with a cruising taxi within an arbitrary 100 meters, according to which to estimate the disequilibrium level and detect the over-served and under-served regions. Based on taxicab trajectories, Zhou Q. et al. proposed to model the taxicab demand in urban regions. Specifically, the propose use the taxicab service rate as an indicator to define the taxicab demands [9]. The analysis results over a real trajectory dataset have showed the weekday exhibits a higher the normality number than that on the weekends. Such an observation results result from the actual fact that weekdays own a relatively regular commuting pattern. Shao D. et al. also proposed to describe the demand-supply level with an indicator that denotes how fast free taxis are taken [10]. The effectiveness of the proposed indicator is identified by obtaining the survey dataset of average waiting time of taxi passengers.

2.2. Taxi service optimizing

Taxi service optimization has attracted attention in both academic researches and real industrial applications, such as the taxi fleet management in urban areas. In fact, the supply-demand equilibrium of the taxi system was studied as early as 1998. The studies have focused on modeling the behaviors of passenger searching [11–13]. Studies in these works find that the average taxi utilization decreases sharply with the number of taxis operating, and that the higher the taxi utilization, the larger the average customer waiting time. While, with the growth of cities and technologies, the ITS (intelligent transportation system) is employed widely. For example, with the pervasive usage of smart phones, the taxi service has come to the era of e-hailing, which is popular globally. The utilization of e-hailing applications has exhibited excellent performance in reducing the average taxi waiting time. It also significantly increases the taxi utilization [14]. A distributed taxi advisory dispatch system (TADS) was proposed in [15], which estimates regions that are under-served. These under-served regions are further advised to taxicabs for obtaining a better balance at a global view. While, in this system, some clients use WiFi devices to communicate with taxis to signal the taxi demand, and taxi drivers also could communicate with each other to exchange the region's service information. As well, with considering the constrains of supply time, capacity, and monetary, some taxi-sharing systems, most of which are smartphone applications at the client side, have been developed to tackle the real-time ride requests from taxi passengers, and some of these systems allow the ridesharing [2,16].

All these researches are conducted on the road network of Beijing using taxi GPS data. In some ways, these researches are a kind of demand-supply balancing approach. Similarly, in [17], it presents exploratory evidence of how "ride-sourcing" services (app-based, on-demand ride services like Uber and Lyft) are used in San Francisco. The findings indicate that, despite many similarities, taxis and ride-sourcing differ in user characteristics, wait times, and trips served. While ride-sourcing replaces taxi trips, at least half of ride-sourcing trips replaced modes other than taxi, including public transit and taxi driving. These findings fill an important gap in our understanding of this emerging travel mode on which publicly available data remains scarce. In China, recent researches on taxi service are mainly conducted by Microsoft Asia. Yuan J. et al. provide recommendations to both taxi drivers and passengers, which mobilizes them and reduces the disequilibrium of the demand and supply on the level of road-segment [18]. On the one hand, a parking place detecting approach is proposed, the parking places stand for the location where taxi drivers usually wait for passengers with their taxi parked. Given the geo-position and time of a taxicab looking for passengers, they suggest the taxi driver with a location, towards which drivers are most likely to pick up a passenger as soon as possible and maximize the profit of the next trip. On the other hand, they provide people expecting to take a taxi with the locations (within a walking distance) where they are most likely to find a vacant taxicab. Thus the proposed recommender system help taxis find passengers more quickly and people take a taxi more easily, therefore balances the taxi demand-supply level to some extent [18–20]. Furthermore, they propose a time-dependent landmark graph, where a node is a road segment frequently traversed by taxis, to model the intelligence of taxi drivers and the properties of dynamic road networks. Based on this graph, they design a routing algorithm to compute the practically fastest route [21].

3. Methods

In this section, we firstly present the definitions of necessary concepts used throughout our paper,

and formally state the focal problem to be solved. In our paper, we adopt a practical trajectory database to evaluate the supply-demand equilibrium strategy.

3.1. Preliminary

Definition 1 (Taxi trajectory). The taxi trajectory is a sequence of GPS points pertaining to the taxi's sampling location over time. Each point $Tr_i \in Tr$ consists of a tuple $\langle (x, y, t), f \rangle$ with location (x, y), location reporting time t, and the taxi's occupancy status f, where (x, y) is a pair of spatial coordinates representing latitude and longitude, f = 1 means the taxi within passengers, otherwise f = 0.

Note that the flag f bound to each trajectory position is essential for judging the taxi occupation state, which is adopted to extract the origin and destination point and corresponding time slot of the passenger.



Figure 1. Taxi trajectory.

Definition 2 (Trip demand). As shown in Figure 1, a trip demand is a 2-tuple $Td = \langle (x_o, y_o, t_o), (x_d, y_d, t_d) \rangle$, where both (x_o, y_o, t_o) and (x_d, y_d, t_d) are geo-spatial points respectively represent the pick-up and drop-off position. All other points between these two points own the same occupation state of f = 1. In detail, *O-D* pair represents a trip demand that starts at (x_o, y_o, t_o) and ends at location (x_d, y_d, t_d) . Therefore, all trip demands of urban taxi passengers are extracted from the taxi trajectory database.

Definition 3 (Trip orientation correlation). Given two trip demands as $Td = \langle (x_o, y_o, t_o), (x_d, y_d, t_d) \rangle$, $Td' = \langle (x'_o, y'_o, t'_o), (x'_d, y'_d, t'_d) \rangle$. When the distance between two origin points dist($(x_o, y_o, t_o), (x'_o, y'_o, t'_o)$) $\langle \lambda_o$, and time span between two origins $|t_o - t'_o| \langle \lambda_t$, where, λ_o, λ_t are respectively the distance threshold and the time lag limitation threshold. Afterwards, the correlation between two trip demands is just cosine distance on trajectories cosine $\langle Td, Td' \rangle$, calculated as below [22]:

$$Corr_{Td,Td'} = \frac{(x_d - x_{ae})^2 + (y_d - y_{ae})^2 + (x'_d - x_{ae})^2 + (y'_d - y_{ae})^2 - (x_d - x'_d)^2 - (y_d - y'_d)^2}{2 \times \sqrt{(x_d - x_{ae})^2 + (y_d - y_{ae})^2} \sqrt{(x'_d - x_{ae})^2 + (y'_d - y_{ae})^2}}$$
(1)

where $(x_{ae}, y_{ae}) = min_{(x,y)}(Td.t, Td'.t)$. It means the origin point that arrive earlier.

Definition 4 (Taxi supply). A taxi status is changed to unoccupied when the flag f change from 0 to 1. A Taxi Supply is a tuple $Ts = \langle (x_s, y_s, t_s), duration \rangle$, as shown in Figure 1, (x_s, y_s, t_s) means the start point of the taxi status change to be vacant, and the duration represents the time span that the taxi remain vacant, which is equal to f = 0.

3.2. Framework

Figure 2 shows the framework of our system that consists of two primary components:

Offline mining: As illustrated in the left column, this stem starts from the trajectory dataset. The trip demand detection module extracts the origin and destination points of the historical trajectories to build the trip demand dataset. It also learns the statistic features of the historical trajectories and trip demands, which is used in the layered blocking module. Specific to the positions that are in the same time slot, we partition them using the proposed blocking method to learn the distribution of trip demands, which represent the knowledge of demands' origin and destination distribution. The blocking results are stored as block distributed pattern of historical trip demands in order to be adopted in online dynamic balance between taxi demand and supply.

Online balancing: As illustrated in the right column, travelers send their trip demands that consist of an origin point and a destination point with corresponding pick-up time slot. The origin and destination points are matched to the blocks of corresponding time slot using the learned block distribution pattern. A demand mutual sensing module is introduced here to make trippers sense each other in the same block of the same time slot, and it get the feedback of other trips that denotes if they agree merging with each other. Then it delivers the trips' information to the next module of trip demand merging, which determines whether two trips could be merged and feedback to trippers. Meanwhile, the specific block is estimated using the defined balancing factor. Then a balancing algorithm is adopted to realize taxi supply translocation from blocks that are sufficient of vacant taxis to blocks that are relatively shorter of taxi supply.



Figure 2. Balancing system framework.

In Section 3.3, the trip demand detection is described, the Subsection "spatial-temporal trip demand distribution" describes both modules of "Statistic Learning" and "Layered Blocking". The module "Block Distribution Pattern" is described in "Block-Based Balance Level Estimating". In Section 3.4, we first describe a demand expanding strategy, where the translocation is also described. Secondly, the trip demand merging is described. Finally in Section 3.4, the modules of "Balancing Degree Analyzing of Blocks" and "Taxi Supply Translocation" are described in the subsection "Demand-and-Supply Balancing Algorithm".

3.3. Spatial trip

Trip demand detection from trajectories: In this paper, we utilize a monthly collection of trajectory data from 13,000 taxis in the City of Shanghai, China. The study area covers all of the Shanghai districts except for Chongming Country, as in Figure 3. As illustrated by the trip demand definition and the system framework, in order to extract a trip demand with an origin point and a destination point, all taxi trajectory records are traversed to find a variation of occupation state (f value of a point). Ultimately, we obtain all taxi passenger trips and then assign each trip origin or destination into associated blocks got with the proposed blocking algorithm. Let us briefly illustrate the idea of our system with Figure 3. It depicts the distribution of origin points and destination points of particular time slot. It shows some areas have plenty of destinations that represent available taxis with only a few origins that denote occupied taxis, like C_1 , while the situation is opposite in some other areas, such as C_2 . The heat map of the ratio of destination points to origin points suggests a demand-supply imbalance in various regions around Shanghai. Meanwhile, the distances between O_1 and O_2 , O_3 and O_3 are respectively within walking distance. Therefore, the trip merging strategy comes to mind easily only if their travel orientation is similar.



Figure 3. The Shanghai study area.

When partition the city into square grids of 1 km. By analyzing the statistical distance between origins of different trips, we found that the distance follows a normal distribution and mean distance is 444 meters.

Spatial-temporal trip demand distribution: In order to learn the spatial-temporal pattern of taxi trips, we partition the origin points and destination points in the same time slot into blocks with small area and high density, which is essential for demand-supply balancing. Therefore, a density based blocking algorithm is proposed in this paper, which divides the points into blocks with different density level, as shown in Figure 4.

As the figure describes, the minimum vertex that covers of all origin and destination points in the same time slot is divided into 2×2 grids and the density is calculated and is justified whether it is in the density interval $[\rho_{low}, \rho_{up}]$. If the density is lower than ρ_{low} , then the grid is deleted, which means that it is no longer considered in later iteration. If it is larger than ρ_{up} , then it is reserved in this layer. Otherwise, the grids are taken into next iteration of partition and each grid is divided into 2×2 smaller grids. After grid dividing, the grid and its neighbors in the same layer, which are less than walking distance, are linked together by labeling an identical block number using the region growing algorithm that are widely applied in image segmentation problem. By intergrading the grids and its 3×3 neighbors, grids of the same layer are set as blocks. A higher layer represents smaller blocks and higher density. The ultimate blocking result of origin and destination points in the same time slot is achieved by overlaying all blocking results of each layer.



Figure 4. Density level-based blocking illustration.



Figure 5. Distribution of trip demand and balance factor over block.

Block-Based balance level estimating: This part is to estimate the imbalance between taxi demand and supply, and to identify the reasonable definition of the balance factor. As the drop-off status of a taxi means that it turns to be vacant, it's worth noting that the defined balance factor within a block of a time slot is estimated by the ratio between destination points and origin points instead of using the vacant taxis number as the numerator.

Definition 5 (Balancing factor within block). Given a set that contains points of trip demand origin and destination, we block all the points using the proposed blocking algorithm and get a set of blocks $B_k^t = \{O_1, O_2, ..., O_r, D_1, D_2, ..., D_s\}$. Here, the total number of origin and destination points in B_k^t is $N_{B_k^t} = r + s$. Assume that, in the dynamic balancing process, a number of $r_+ = \sum_{t' < t} \sum_{B_{k'} \in B^{t'}} |B_{k'}^{t'} \sim B_k^t|$ destination points are scheduled to block B_k^t , which means the vacant taxis are reallocated to B_k^t , where the symbol \sim denotes transition, while $r_- = \sum_{t''>t} \sum_{B_{k''} \in B^{t''}} |B_k^t \sim B_{k''}^{t''}|$ destination points are scheduled to other time slots or other blocks from B_k^t , which means the vacant taxis' location $B_{k''}^{t''}$ substitutes for the original block B_k^t . Then, the balance factor of a block B_k^t is defined as below, which is a system indicator that is the basis for our demand-supply balancing strategy.

$$bf_{B_k^t} = \frac{s + r_+ - r_-}{\Delta \times r - M_{B_k^t}} \tag{2}$$

Here, Δ is the inflating parameter of historical trip demand that will be introduced in Section C. And $M_{B_k^t}$ is the number of merged trips that will be elaborated in Section C. As in Figure 5, it shows that the blocks of denser trip demands are of small balance factor value, which means more unbalanced. This fits ordinary commonsense observations. Figure 6 is an example of the spatial distribution of the trip demand and the corresponding balance factor.

3.4. Demand and supply balancing approach

Demand expanding for the balance platform: Note that the taxi GPS trajectory dataset only

reveals the number of trip demands that have got served. In reality there are also many taxi demands unsatisfied and disappeared due to the shortage of taxis. In order to take such taxi demands into consideration and validate our proposed balance model under practical settings, we introduce a system parameter Δ , which stands for the ratio of actual total number of taxi demands to the trip demand number extracted from the historical dataset.

Additionally, the historical trajectory dataset conceals rich information regarding 1) the spatialtemporal distribution of the trip demand origin points and destination points, and 2) the mobility patterns of taxi riders. In this paper, we evaluate a coarse-grained demand-and-supply balancing approach that means that we disregard the road network.

Previously, the origin and destination points of the same time slot in weekdays or weekends/holidays are separately mixed and then blocked with different density level. Obviously, the block distribution represents spatial-temporal taxi trip demands distribution, including trip origins and destinations, and the demand mobility patterns can also be described by transition probability between blocks. Thus, we mine the trajectory dataset to build an experimental platform, which generates inflated trip demands of different time slots for the balancing model and preserves the transition pattern between blocks. Figure 6 illustrates the demand expanding process.



Figure 6. Distribution and transition between blocks of the inflated demand.

Assume that we blocked trip demand origins and destinations and got eight blocks B_{1-8} in three continuous time slots.

Given the system inflation parameter Δ , both point number in each block B and each grid g is inflated, and the inflating process guarantees the transition probability of $B_i^{t_k}$ to $B_j^{t_{k'}}$ remains unchanged as shown in Eq (3).

$$\forall g \in B_{k}^{t}, g' \in B_{k'}^{t'}, t' \geq t, \begin{cases} pr\left(\sum \Delta \times B_{k}^{t} \rightsquigarrow \sum \Delta \times B_{k'}^{t'}\right) = pr\left(\sum B_{k}^{t} \rightsquigarrow \sum B_{k'}^{t'}\right) \\ pr(\sum \Delta \times g \rightsquigarrow \sum \Delta \times g') \neq pr(\sum g \rightsquigarrow \sum g') \\ pr(\Delta \times B_{k}^{t} \rightsquigarrow \Delta g') \neq pr(B_{k}^{t} \rightsquigarrow g') \end{cases}$$
(3)

where the transition from B_i^t to $B_{k'}^{t'}$, i.e., $\sum B_k^t \longrightarrow \sum B_{k'}^{t'}$ is defined according to the origindestination pair of the trip demand $Td = \langle (x_o, y_o, t_o), (x_d, y_d, t_d) \rangle = \langle 0, D \rangle$, and the transition probability is defined as Eqs (4) and (5).

For example, the transition probability from B_1 to B_6 is the $n_{1,13;14;16} + n_{2,15}$ divided by $n_{1,3} + n_{1,8;10} + n_{1,13;14;16} + n_{2,3;4;5} + n_{2,6;7} + n_{2,9} + n_{2,11;12} + n_{2,15} + n_{2,18} + n_{2,19;20}$.

According to Eq (3), this transition probability remains unchanged before and after demand inflating, which means that the demand distribution and mobility are preserved same with historical dataset on the level of our block. While, the transition probability from B_2 to g_9 is changed, as well as the transition probability from g_3 to g_9 , or g_{12} to g_9 .

$$Sgn(P, B_{k}^{t}) = \begin{cases} 1, P. (x, y) \in B_{k}^{t} \land P. t \in t \\ 0, P. (x, y) \notin B_{k}^{t} \end{cases}$$
(4)

$$pr(\sum B_k^t \rightsquigarrow \sum B_{k'}^{t'}) = \frac{\sum_{Td = \langle O, D \rangle} Sgn(O, B_k^t) Sgn(D, B_{k'}^{t'})}{\sum_{Td = \langle O', D' \rangle, B_{k'}^{t''}, t'' \geq t} Sgn(O', B_k^t) Sgn(D', B_{k''}^{t''})}$$
(5)

Trip merge strategy: As described by the definition of Trip orientation correlation, when the origin coordination is close and the trip starting time is in the same slot, the correlation value closing to 1 represents that two trips could be merged in our framework. The distance threshold and time lag threshold are respectively the grid size of the 12th layer and time slot granularity. Generally, the trip merging aims to decrease the average total travel distance, though it may increase the travel distance for single travel. Here, we propose a lazy trip correlation calculating method that considers the Euler distance between the origin and destination points. Assuming the travel distance $d_2 > d_1$, and $O_{ae} = min_{O_1,O_2}(O_1, t, O_2, t)$ is the origin point of the earlier trip demand. In order to assure that the increased distance of Td_2 is accepted by the trippers, we add a trip merging limitation of $\alpha < 90^\circ$, and the increasing distance ratio is set as the system parameter as shown in Eq (6).

$$\gamma_{d_1,d_2} = \frac{1}{d_2} \sqrt{d_1^2 + d_2^2 - 2d_1 d_2 \cos\alpha} \tag{6}$$

The arriving time slot of Td_2 is estimate as Eq (7).

$$dest_Time_{(Td_1,Td_2)} = \left[t_{D_1} + \frac{\|D_1D_2\|}{O_{ae}D_1} (t_{D_1} - t_{O_1}) \right]$$
(7)

Demand-and-supply balancing algorithm: According to the definition of Balancing factor within block, we sort the balancing factors (bf) of all blocks in time slot t, and the bf based demand-supply balancing algorithm is shown as Algorithm 1 below. As grids in 11th layer of the blocking algorithm sis 136.19 m × 50.92 m, which is regarded to be reached with walking distance, and this is the distance threshold λ_0 for trip merging.

$$max_{x} \frac{\sum Sgn(D_{i},B)-x}{\sum Sgn(O_{i},B)} \ge \frac{\sum Sgn(D_{i},B')+x}{Sgn(O_{i},B')}$$
(8)

Table 1. Algorithm 1.

Algorithm 1: <i>bf</i> _Balancing(<i>t</i>)								
Static Input: trip demand distribution pattern <i>distr_B^t</i>								
	Inflated trip demand $\{I_{Td} = < O_{B_{k'}^{t}} D_{B_{k'}^{t'}} > \},$							
distance ratio threshold γ , bf variation threshold σ .								
Output: updated distribution of blocks B^{t+1}								
1	1 foreach demand pair $I_{Td_1}, I_{Td_2}, O_1, O_2 \in g_k^t$ and $dist(O_1, O_2) \le \lambda_o$ // trip merging							
2	$d_1 = distance(O_1, D_1); d_2 = distance(O_1, D_1);$							
3	3 if $\gamma_{min(d_1,d_2),max(d_1,d_2)} < \gamma$							
4	$k^{\prime\prime} = longer_k(I_{Td_1}, I_{Td_2});$							
5	$t'' = dest_Time(I_{Td_1}, I_{Td_2});$							
6	end if							
7	add to $Block(B_{k''}^{t''}, location(max_D(d_1, d_2));$							
8	foreach block B_k							
9	$bf_{B_k} = \sum Sg$	$m(D_i, B_k) / \sum Sgn(O_i, B_k);$						
10	$0 \qquad \sigma^{t} = variance(bf_{B_{1}}, bf_{B_{2}}, \dots, bf_{B_{N}}), N = \parallel distr_{B_{k}}^{t} \parallel;$							
11	11 while $\sigma^t < \sigma$							
12	B = n	$\max_{B_k} bf_{B_k};$						
13	B' = r	$nin_{B_{k'},k'\neq k}dist(B,B_{k'});$						
14	x = tr	$cansitNum_{B,B'};$	//Eq (8)					
15	Trans	location(random(B, x ,{D}),B')	//	sufficient supply translocation				
16	end while							
17	go to 8.							
18	go to 1							
19	9 return $bf_Balancing(t+1)$;							

4. Experiment

4.1. Data and metrics

We perform the experiments using the taxi trajectory dataset contains the GPS trajectory recorded by over 13,000 taxis during a period of 30 days spanning from April 1st to April 30th in the year of 2015. After trip detection, there are more than 11.23 million occupied trips. The performance of the devised demand-supply balancing system is evaluated in the perspective of effectiveness, described by following effectiveness measurements.

Proportion of merged trips (PMT) is the merged trips percentage of total trip demands. Relative Decrement of mean BF value (RDMBF) and Relative decrement of BF variance (RDBFV) are applied

to evaluate the effectiveness of our defined balancing factor. *Satisfaction proportion (SP)* is the fraction of trip demands that get satisfied in the balancing process (the merged trips are given the priority to be satisfied and two merged trips are regarded as two satisfied demands).

We compare the performance of the original system that means non-merging and non-balancing approach with three flavors of our proposed balancing approach: Non-Merging With-Translocation approach, the With-Merging Non-Translocation, the With-Merging With-Translocation. We added another system parameter η to represent the percentage of travelers that receive the trip merging strategy with others.

4.2. Results

The overall evaluation result of the metrics described above is shown as Table 1. In the evaluation, our proposed approach serves 20.86% additional taxi users while the defined balance factor average value decreasing by 1.58% and the variance decreasing by 3.2% compared with non-balancing (when the inflating ratio of taxi trip demand number is 3 and the ratio of trip merging acceptance is 0.6). Evidently, the satisfaction proportion of trip demands increases obviously when it applies our system, and compared to the Non-Balancing method, the With-Balancing method achieve much more increase of satisfaction proportion, which means when we adopt the proposed balancing factor based supply translocation algorithm, more trip demands can be satisfied. Furthermore, it also achieves much better value of RDMBF and RDBFV. The PMT value shows that 41.7% trips are merged when the system parameters are set as $\Delta = 3$, $\eta = 0.6$, $\gamma = 0.3$.

	Non-Merging,	Non-Merging,	With-Merging,	With-Merging
	Non-Translocation	With-Translocation	Non-Translocation	With-Translocation
PMT	_	—	0.417165	0.417165
RDMBF	_	0.013181		0.021149
RDBFV	_	0.039198		0.086995
SP	0.333333	0.65367	0.476190	0.881090

Table 2. Evaluation results of the study case ($\Delta = 3$, $\eta = 0.6$, $\gamma = 0.3$).



Figure 7. Proportion of merged trips with different parameters.

The metrics of different time slot with different system parameters are shown in Figures 7–10, and grid-based balancing is added to compare which block-based balancing, which partition the whole study area into all grids of identical size of 1 km. As shown in Figure 7, the proportion of merged trips achieves higher value when the system parameter $\Delta = 4$, compared to $\Delta = 3$ or $\Delta = 5$. This is because that when the trip demands are 4-inflated, the total number reach the amount of taxi supplies. When Δ and γ are fixed, the proportion achieves higher value when η is larger. It shows no significant difference of the proportion value when γ varying. It's worth noting that the other three metrics are also not quite sensitive to parameter γ . This reveals that γ is weakly correlated to the system performance though we instinctively regard that the distance increment ratio is an essential factor for trip merging.

Figures 8 and 9 show that the RDMBF and RDBFV remain more than zero that means the proposed balancing approach indeed improves the balance status of the taxi demand-supply system, as well, it certifies the reasonable definition of balance factor as a measurement of system balance status. The balance status achieves more improvement for ordinary time period than peak periods. And the RDBFV shows that the balance variance improves during almost time frame.



Figure 8. Relative decrement of mean BF value with different parameters.



Figure 9. Relative decrement of BF variance with different parameters.

As in Figure 10, the SP value also displays more sensitive to Δ while shows a little difference when η or γ changes separately. It can be seen that during the morning rush hour, the demand satisfaction ratio is not sensitive to the influence of the expansion parameters. While, after ten o'clock in the morning and during the evening peak hours, SP value is larger with small Δ . SP is not sensitive to η and γ . Therefore, after adopting the proposed balance model, under the condition of fixed taxi number, the taxi hailing success rate achieves 1.2 times of none balance strategy.



Figure 10. Satisfaction proportion of trip demands with different parameters.

5. Discussion and conclusions

This paper mainly devised a coarse-grained approach for urban taxi demand-supply balancing, after proposing a density-based blocking algorithm that partition trip demands into blocks of different density level. The supply translocation part is based on the defined balancing factor that is adopted to estimate the demand-supply balance level of a block in a given time slot. We evaluate our approach utilizing a real trajectory dataset generated by 13,000 taxis of Shanghai in one month. The experimental results demonstrated the effectiveness of our approach in estimating the demand-supply balance level and improving the satisfaction proportion of trip demands.

We regard the system as a coarse-grained taxi demand-supply balancing prototype because either the scheduling or trip merging process is implemented on the granularity of blocks. There are several potential future directions for this framework. Firstly, we will try to implement the framework on the level of the road network, so the travel distance and the merged trip destination time could be calculated with real path distance and more reasonable time consumption calculation. Secondly, with the development of mobile applications, such as the Didi Taxi and the Uber, the real taxi request that gets served and unserved can be collected, which can substitute for the trip demand inflating part.

Acknowledgments

The work was supported by the Jilin Provincial Science & Technology Department (Grant No. 20170204057GX), Jilin Provincial Education Department Science and Technology Project (Grant No. JJKH20190924KJ), Study on the influences of specific Environment on the Biological characteristics of Winter athletes (Grant No. 2018YFF0300405), and supported by Graduate Innovation Fund of Jilin University (Gran No. 101832020CX184).

Conflict of interest

The authors declare there is no conflict of interest.

References

- 1. D. Liu, S. F. Cheng, Y. Yang, Density peaks clustering approach for discovering demand hot spots in city-scale taxi fleet dataset, in 2015 IEEE 18th IEEE International Conference on Intelligent Transportation Systems, IEEE, (2015), 1831–1836. doi: 10.1109/ITSC.2015.297.
- 2. S. Ma, Y. Zheng, O. Wolfson, Real-time city-scale taxi ridesharing, *IEEE Trans. Knowl. Data Eng.*, **27** (2015), 1782–1795. doi: 10.1109/TKDE.2014.2334313.
- 3. J. Yuan, Y. Zheng, X. Xie, G. Sun, Driving with knowledge from the physical world, in *Proceedings of the 17th SIGKDD Conference on Knowledge Discovery and Data Mining*, KDD, (2011), 316–324. doi: 10.1145/2020408.2020462.
- 4. X. Liu, L. Gong, Y. Gong, Y. Liu, Revealing travel patterns and city structure with taxi trip data, *J. Transp. Geogr.*, **43** (2013), 78–90. doi: 10.1016/j.jtrangeo.2015.01.016.
- 5. C. Yang, E. J. Gonzales, Modeling taxi demand and supply in New York city using large-scale taxi GPS data, in *Seeing Cities Through Big Data*, Springer, (2017), 405–425. doi: 10.1007/978-3-319-40902-3_22.
- 6. C. Yang, E. J. Gonzales, Modeling taxi trip demand by time of day in New York city, *Transp. Res. Rec. J. Transp. Res. Board*, **2429** (2014),110–120. doi: 10.3141/2429-12.
- 7. C. Kang, K. Qin, Understanding operation behaviors of taxicabs in cities by matrix factorization, *Comput. Environ. Urban Syst.*, **60** (2016), 79–88. doi: 10.1016/j.compenvurbsys.2016.08.002
- 8. Y. Huang, J. W. Powell, Detecting regions of disequilibrium in taxi services under uncertainty, in 20th International Conference on Advances in Geographic Information Systems, (2012), 139–148. doi: 10.1145/2424321.2424340.
- 9. Q. Zhou, J. Zhang, J. Li, S. Wang, Discovering regional taxicab demand based on distribution modeling from trajectory data, in 2014 IEEE 17th International Conference on Computational Science and Engineering, IEEE, (2014), 1605–1610. doi: 10.1109/CSE.2014.296.
- D. Shao, W. Wu, S. Xiang, Y. Lu, Estimating taxi demand-supply level using taxi trajectory data stream, in 2015 IEEE International Conference on Data Mining Workshop, IEEE, (2015), 407– 413. doi: 10.1109/ICDMW.2015.250.
- 11. H. Yang, S. C. Wong, A network model of urban taxi services, *Transp. Res. Part B Methodol.*, **32** (1998), 235–246. doi: 10.1016/S0191-2615(97)00042-8.
- 12. H. Yang, S. C. Wong, K. I. Wong, Demand-supply equilibrium of taxi services in a network under competition and regulation, *Transp. Res. Part B Methodol.*, **36** (2002), 799–819. doi: 10.1016/S0191-2615(01)00031-5.
- 13. H. Yang, T. Yang, Equilibrium properties of taxi markets with search frictions, *Transp. Res. Part B Methodol.*, **45** (2011),696–713. doi: 10.1016/j.trb.2011.01.002.
- 14. F. He, Z. J. Shen, Modeling taxi services with smartphone-based e-hailing applications, *Transp. Res. Part C Emerging Technol.*, **58** (2015), 93–106. doi: 10.1016/j.trc.2015.06.023.
- F. C. Choo, M. C. Chan, A. L. Ananda, L. S. Peh, A distributed taxi advisory system, in 2012 12th International Conference on ITS Telecommunications, IEEE, (2012), 199–204. doi: 10.1109/ITST.2012.6425165.
- S. Ma, Y. Zheng, O. Wolfson, T-share: a large-scale dynamic taxi ridesharing service, in 2013 IEEE 29th IEEE International Conference on Data Engineering, IEEE, (2013), 410–421. doi: 10.1109/ICDE.2013.6544843.

- L. Rayle, D. Dai, N. Chan, R. Cervero, S. Shaheen, Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco, *Transp. Policy*, 45 (2016), 168–178. doi: 10.1016/j.tranpol.2015.10.004.
- 18. J. Yuan, Y. Zheng, X. Xie, G. Sun, T-Drive: enhancing driving directions with taxi drivers' intelligence, *IEEE Trans. Knowl. Data Eng.*, **25** (2013), 220–232. doi: 10.1109/TKDE.2011.200.
- 19. J. Yuan, Y. Zheng, L. Zhang, X. Xie, G. Sun, Where to find my next passenger, in ACM Conference on Ubiquitous Computing, (2011), 109–118. doi: 10.1145/2030112.2030128.
- N. J. Yuan, Y. Zheng, L. Zhang, X. Xie, T-finder: a recommender system for finding passengers and vacant taxis, *IEEE Trans. Knowl. Data Eng.*, 25 (2013), 2390–2403. doi: 10.1109/TKDE.2012.153.
- J. Yuan, Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, et al., T-drive: driving directions based on taxi trajectories, in ACM Sigspatial International Symposium on Advances in Geographic Information Systems, San Jose, (2010), 99–108. doi: 10.1145/1869790.1869807.
- 22. P. Xia, L. Zhang, F. Li, Learning similarity with cosine similarity ensemble, *Inf. Sci.*, **307** (2013), 39–52. doi: 10.1016/j.ins.2015.02.024.



©2022 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)