東京大学大学院新領域創成科学研究科
社会文化環境学専攻

2020 年度

修　士　論　文

Improvement on online ride-hailing based on empirical GPS data

経験的 GPS データに基づくオンライン配車の改善

2020 年 7 月 10 日提出

指導教員　柴崎　亮介　教授
副 指導教員　瀬崎　薫　　教授

陳　瑾瑜

Jinyu Chen

Contents

List of figures

List of tables

Abstract

Ride-hailing, as a popular shared-transportation method, has been operated in many areas all over the world. Researchers conducted various researches based on global cases. They argued on whether car-hailing is an effective travel mode for emission reduction and drew different conclusions. The detailed emission performance of ride-hailing system depends on the cases. Therefore, there is an urgent demand to reduce the overall picking up distance during the dispatch. Moreover, most of the cases only analyze the emission pattern of mature ride-hailing systems. None of them provide a change pattern of a developing one. Discovering the emission pattern during the development can help understand how number of users in the system affect the emission performance and furtherly provide guideline of controlling the number of users to keep the system at a high-performance level. In this study, we answer these two demands by proposing two frameworks. 1. A cross simulation model combined with Gibbs sampling for a comprehensive computation. Based on the simulation results, we found a strong impact of user scale on the emission performance. The mean of void distance proportion varies from 3.69% to 31.75% under all situation simulation. Finally, based on this relationship, we provided a guidance for the computation of approximate user scale if the emission and efficiency performance of car-hailing is expected to be better than a threshold. 2. A optimization method combined with prediction model to minimize the global pick-up distance when solving the dispatch problem. We use Didi ride-hailing data on one day for simulation and found that our method can reduce the picking-up distance by 8.60% compared with baseline greedy algorithm. The proposed algorithm additionally makes the average waiting time of passenger more than 10 minutes shorter. The statistical results also show that the performance of our method is stable. Almost the metric in all cases can be kept in a low interval. We believe our findings can improve deeper insight into the mechanism of ride-hailing system and contribute to further studies.

Key words: Ride-hailing System, Simulation, Gibbs Sampling, Greedy Algorithm, Deep Learning, Optimization

Acknowledgements

First and foremost, I would like to express my great gratitude to my advisor Prof. Ryousuke Shibasaki. Shibasaki sensei gave me many advices on the design on the simulation of ride-hailing system. He also provided precious Didi ride-hailing record data for the simulation in this study. I am also grateful to my co-advisor, Prof. Kaoru Sezaki. Sezaki sensei also gave me many practical suggestions as research directions.

Thanks to Associate Professor Xuan Song, who is the leader of IPUC group. Song sensei often shares interesting topic with us, which helps us learn lots of methodologies of analyzing human mobility. Thanks to Associate Professor Haoran Zhang, Zipei Fan and Renhe Jiang, who taught me lots of useful programming skills, data processing knowledge and ideas to do research. Among them, Associate Professor Haoran Zhang helped me most. He gave me the experience of paper writing and conducting novel researches. Under his instruction, I successfully published two journal papers as the first author separately in Applied Energy and Journal of Cleaner Production. Here, I thanked him a lot for his patient instructions in these two years.

In addition, I thank all my friends in Shibasaki Lab: Thanks to Yuhao Yao, Xinlei

Lian, Bojia Chen, Zekun Cai, Wenjing Li and Tengyao Yu, with whom I enjoy the happy study life in the University of Tokyo. Thanks to senior students Shuzhe Huang, Dou Huang, Xiaoya Song, Tianqi Xiaa and Zhiling Guo, who helps me with my daily life in Japan.

Thanks to my parents support my life of studying abroad in Japan.

Due to the coronavirus outbreak all over the world. Japan is also influenced by the virus. From April to July, I stayed at my rent house and completed my master thesis. During this time period, Haoran Zhang usually phoned with me to talk about the details of my researches and gave me remote assistances. Also, my friends sometimes talked with me to give much mental support. Here, I am very grateful to all of you.

Improvement on online ride-hailing based on empirical GPS data

1. Introduction

As a great source of emission and energy consumption, transportation attracts the attention of many studies on emission reduction and energy-saving[1]. In recent years, sustainable transports in the city arise all over the world [2]. Among all kinds of sustainable transports, ride-hailing is a popular trend in big cities [3, 4]. Ride-hailing refers to the services that dynamically match drivers' supply and customers' demand and allow customers hire drivers to send them to destinations through online ride-hailing platforms, such as Uber (an American multinational transportation network company) and Didi Chuxing (the biggest ride-hailing service provider in China, in the following part of this paper, we will use Didi as the reference for it). Ride-hailing system does bring many benefits to citizens and the urban development. The average waiting time by adopting ride-hailing is dramatically shorter than the original taxi dispatch [5, 6]. Also, people live in the remote and rural areas where public transport can hardly reach can go to urban areas easily by ride-hailing system, which indirectly promotes urbanization go faster. Meanwhile, ride-hailing can provide more jobs for drivers [7] and improve driving efficiency [8]. As we all know, the emission by transportation can affect citizen's health and amenity of the city [9].

Spreading ride-hailing is a popular trend[10]. Currently, there are many places, where there is no mature ride-hailing service. There is still much vacancies in local ride-hailing market. When one ride-hailing service company tries to develop local market, there must be a phase of increasing of user scale, which refers to the number of users. The emission behavior of ride-hailing system during this phase is remained to be a puzzle. With the rise of user scale, the quantity of emission of system can be constantly increasing. The thing that is essential is how to observe and control the change that user scale brings to the efficiency of emission. Currently, most of studies focused their attentions on the emission behavior of developed ride-hailing systems. Few evidences discussed on the emission behavior of developing one. As a crowdsourcing service, the participation scale is one key factor to affect the emission performance of whole system. Thus, to promote a cleaner shared transportation, the clear pattern of the relationship between emission performance and user scale should be studied further.

In addition, there are some scholars exploring the efficiency of novel technologies in mature ride-hailing system. Korolko el al.[11] indicated that bipartite matching with time window batching and dynamic pricing can lower waiting time for both riders and drivers as well as capacity utilization, trip throughput and total welfare. However, they only consider the dispatching in real-time time window and didn't consider the travel demand in future. What's more, Afeche et al.[12] pointed out that the interference from service platform to avoid dispatching driver to area with low travel demand can be optimal. These two conclusions inspire us with the idea that if we can predict the distribution of travel demands in the future, whether it can help us optimize the dispatching of ride-hailing system, especially improving the utility of energy. Among current studies, there is no existing literatures that quantifying and discovering this improvement. This gap also needs to be fulfilled and furtherly instructs the development of ride-hailing system.

Operating such kind of research isn't an easy task. Solid and real travel demand data are required as the base of simulation and assessment. Next, reliable prediction model is in a dominant position in the whole simulation as imprecise prediction can bring misjudge to dispatching decision. In addition, the suitable dispatching should be designed carefully and the performance and applicability should be ensured.

The development of method for urban data mining [13] as well as more and more occurrence of works on GPS data mining enable us to analyze the emission performance from urban transportation GPS information [1]. In this paper, we adopted massive Didi GPS records and designed a simulation method to mainly achieve two tasks: (1) Mining the relationship between the emission performance of ride-hailing system and user scale and provide corresponding advices for promoting ride-hailing services. (2) Proposing a prediction-based dispatching method for ride-hailing system to improve the efficiency of system. We chose Chengdu City, a typical big city in China as the study case.

2. Related works

In recent years, ride-hailing is becoming a popular topic among researchers. They concerned the problem from various aspects. The mainstream of research fields can be presented as the impact the ride-hailing has brought to urban areas; regulation and policy on ride-hailing; future development expectation and efficiency and benefit of ride-hailing.

For impact of ride-hailing, scholars separated and discussed their options from various perspectives. Rayle et al.[5] indicated that in San Francisco, at least half of ride-hailing trips replaced traditional urban transportation mode like taxi, public transportation based on the comparison of survey data and trip data from ride-hailing and taxi. Henao et al.[14] discussed that 13% of survey respondents owned fewer cars because of ride-hailing after analyzing on survey and socio-demographic data from 311 passengers. Different from the conclusion by Li, Erhardt et al. [15] found that in San Francisco, 2016, 22% of traffic delay may be reduced without the operation of ride-hailing based on a simulation employing travel demand model.

For regulation and policy, Flores et al.[16] observed the process of appearance of ride-hailing on San Francisco street, the conflict between the ride-hailing companies and regulatory agencies, the resolution to the conflict through new and better registration framework. Onto higher level, Beer et al.[17] revealed that the regulations and their strictness vary in 15 different USA cities. What's more, the ride-hailing service providers trend to operate in areas with light and non-fingerprint information registered regulation. Both studies revealed the current status of regulation and provide the guideline for future decision making.

In future development expectation, autonomous vehicle is always an eternal topic. Bosch et al.[18] showed that autonomous service can reduce the cost of ride-hailing by 85% in Zurich. Wadud et al.[19] argued that the there is a potential in the benefit brought by vehicle automation, while this potential depends many factors like vehicle automation level and vehicle connection. Some scholars also focused on the development brought by electric vehicle technology. Tu et al.[20] battery electric vehicles with 200km could satisfy the travel demand of ride-hailing drivers up to 47% or 78% and 20% or 55% of total ride-hailing travel distance can be traveled by driven by electric vehicle. Level-2 charging available at home, work, and public parking can boost the acceptance up to 91% of drivers and 80% of distance.

About efficiency and benefit of ride-hailing, most scholars discussed on the environmental and energy saving behavior of ride-hailing. Reducing the vehicle distance traveled became an important ruler to evaluate the energy behavior of ride-hailing. Some conclusions told that, however, the introduction of ride-hailing brought the increase of vehicle distance traveled. Schaller[21] found that in 2017, the ride-hailing brought additional vehicle distance traveled summed up to 9.1 billion kilometers in nine US cities. Furtherly, some scholars put forward the point that the actual impact of ride-hailing depends. Tirachini et al. [22] did a Monto Carlo Simulation on the ride-hailing scenarios in Chile. They declared the if the mean occupancy rate is 2.9 pax/veh or higher, there is higher possibility that the ride-hailing can cause less travelling distance. Rodier[23] summarized that the factors that can decide whether the travelling distance can be reduced are auto ownership, trip generation, physical and legal limits to driving, mode choices and void vehicle relocation travel. The report emphasized that the void cruising distance can account for up to 20% of total travelling distance in high density urban areas and 60% in lower density suburban areas because of lower travel demand density. Therefore, an urgent effort is to decrease void cruising distance.

Dispatch and matching are the key issue to overcome this issue. The strategy of dispatch directly impacts the global void cruising distance. Recently, there exists a lot of literatures considering to optimize the dispatch strategy. Xu et al.[24] proposed an optimization dispatch approach considering maximizing the future gain of drivers, which refers to the matchable pairs of driver-order, based on mining the empirical order demand pattern. Their simulation and real-world application result show that the approach can maximize the Gross Merchandise Volume of total drivers. Feng et al.[25] improved the matching issue of ride-hailing from the perspective of waiting time of passengers. They developed a heuristic method and proved that the algorithm can give the near optimal solution. However, they didn't directly consider the optimal void cruising distance. Also, their strategy only considers the matching based on the current existing demand. With the emerging works concerning the problem of predicting the ride-hailing demand[26], we have some solid methods to forecast the travel demand in future scenarios. By merging the future knowledge of distribution of travel demand, we can improve the matching issue to minimize the void cruising distance, thus, minimizing the invalid energy consumption. However, there is scarce literature quantifying this improvement and give an analysis on the performance.

Especially, these studies mainly focused on analyses from historical evidences in places with nearly saturated ride-hailing system. They only operated simulations on complete dataset in places, where the ride-hailing services have been operated in a long time. The number of users that participate in ride-hailing every day is relatively stable. Thus, most of them consider improving the emission performance from the perspective of technologies. As a crowdsourcing system, the emission performance of ride-hailing is also deeply affected by the number of users. They failed to discover the relationship between the number of users and emission performance. The meaning of such study is to solve the problem that during the growth of riders in the system, how many registered drivers should be kept to reach a balance in both efficiency and emission performance Such gap also needs to be further studied.

Simulation is a common method to study ride-hailing system. Maciejewski et al.[27] use the simulation setup of MATSim to simulate the proposed dispatching algorithm to prove the efficiency of their algorithm. Grau et al.[28] simulated on an agent-based ride-hailing dispatch model for better improvement on driver earnings, user cost and vacant versus occupied time. Mourad et al.[29] summaries various ride-hailing models for the simulation. In this study, we propose a Gibbs sampling-based simulation framework to further discover the emission behavior of ride-hailing system during its development. We believe our work can fill this gap and provide advice for ride-hailing service providers.

Therefore, there are some gaps in current studies:

(1) There is lack in solid and visible evidence on the change pattern of emission behavior when the number of user in the ride-hailing system changes.

(2) Considering the real-time matching of driver and passenger with the combination of optimization method and prediction model to minimize the void cruising distance as well as maximizing the energy utility

(3) The simulation of method on real-world data to improve the performance and applicability as well as an analysis on the spatial-temporal pattern of emission behavior of two different dispatch strategy.

In this paper, we will try to answer these questions by operating two simulations separately:

(1) Simulation on how user scale impact the emission performance and efficiency of ride-hailing system.

(2) Simulation on dispatch algorithm based on optimization method and prediction model and operating novel analysis.

3. Problem Description

3.1 Emission performance and user scale

3.1.1 Concept of emission performance

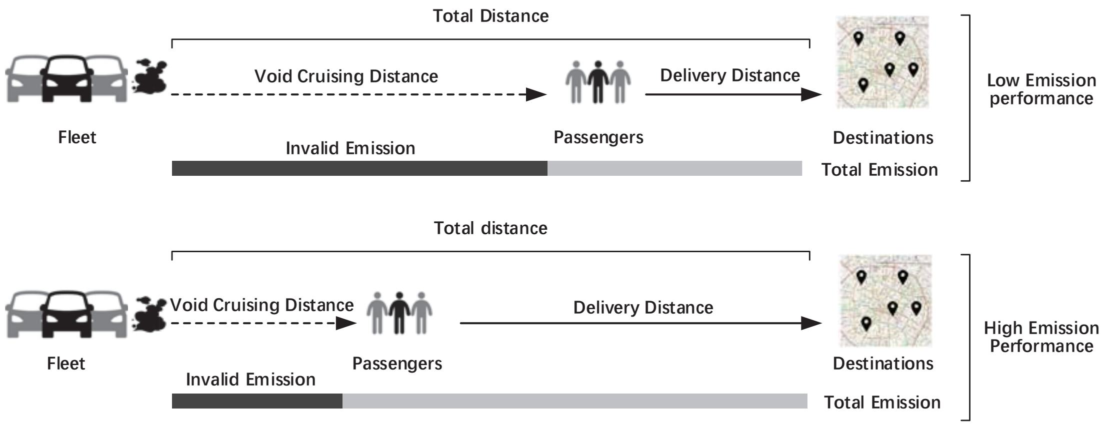


Figure 1. Illustration of emission performance

Before we give more details of the problem description, we need to give the concept of emission performance. It can be referred directly by invalid emission proportion or indirectly void cruising distance proportion. As shown in Figure 1, for each ride-hailing order request, the expected driver to finish the request must travel two trips, which refers to pick up trip and delivery trip respectively[30]. The distance of pick-up trip, which is referred by the dotted line in the figure is defined as the void cruising distance as it won't serve the transportation and create income to the driver; the distance of delivery trip, which is referred by the solid line is defined as the delivery distance. So, the exhaust emitted during the pick-up trip by the car is defined as invalid emission and the one during the delivery distance is defined as valid emission. thus, the proportion of void cruising distance to the total distance is the void cruising distance proportion and similarly, the proportion of invalid emission can be thought as the proportion of invalid emission to the total emission. Apparently, there is a relationship between the void cruising distance proportion and the invalid emission proportion. If we can get the void cruising distance proportion, the invalid emission proportion can also be computed, symbolizing the emission performance of car-hailing system.

The main problem that this work tries to figure out is the difference of emission performance, which is directly influenced by drivers' trajectories under different scales of user available in the system. To extent the local market of ride-hailing system, ride-hailing service provider would try many methods to attract users to system[31]. With the growing of regular users, the emission performance is also deeply affected. In this study, we will adopt a simulation framework to evaluate this impact and analyze the invalid emission proportion as the symbol for the emission performance. From result of analysis, we can have a clear observation on how user scale effect the emission performance of ride-hailing, thus, giving further advice on controlling the number of users in the system and keep the emission performance of system at a high level.

3.1.2 Emission performance and user scale

Here, we will illustrate how the change of user scale in the ride-hailing system affects the emission performance that inspires our method. We will choose three cases as the illustration sample as shown in figure 2. In figure, $\boldsymbol{d_j}$ is the driver $\boldsymbol{j}$, $\boldsymbol{p_i}$ is the passenger $\boldsymbol{i}$. Suppose a scenario during the development of a ride-hailing system (a) The situation that may happen under the supplement of passenger scale (b) The situation that may happen under the supplement of driver scale (c) The situation that may happen under the supplement of driver and passenger scales simultaneously. The solid line refers to the original routine before the reduction; the dotted line refers to the routine drivers probably take after the supplement. Yellow triangle refers to the origin of the order; blue circle refers to the destination; red rectangle refers to the origin of driver.

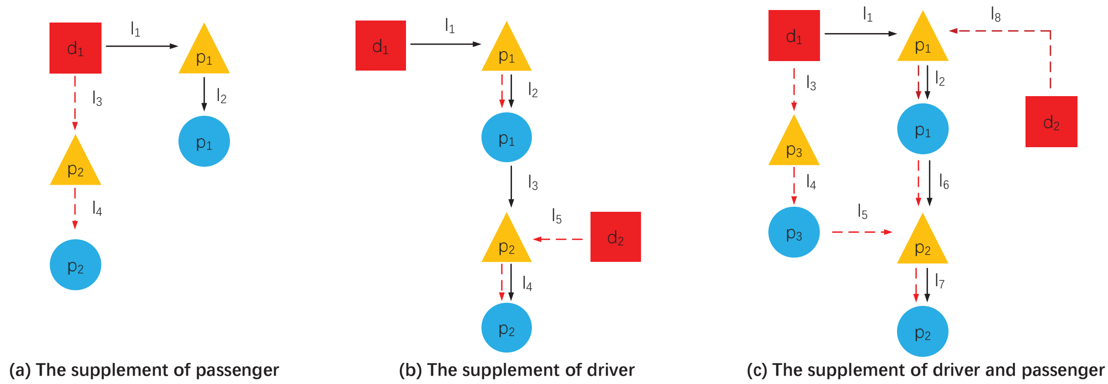


Figure 2. Illustration of relationship between emission performance and user scale

Figure 2(a) shows the situation of the supplement of passenger scale**.** Formerly, driver ***d₁*** is assigned to pick up the passenger ***p₁***. If there exists a new passenger ***p₂*** simultaneously and ***p₂*** is assigned to ***d₁*** to pick up, the driver ***d₁*** will turn to pick up ***p₂*** and give up ***p₁***. This kind of change causes driver ***d₁***'s travel distance changes from $l_1 + l_2$ to $l_3 + l_4$ and the void cruising distance proportion in this situation changes from $\frac{l_1}{l_1+l_2}$ to $\frac{l_3}{l_3+l_4}$. Thus, the emission performance of ***d₁*** changes. But it's hard to tell whether it increases or decreases.

Figure 2(b) shows the situation of the supplement of driver scale. Formerly, the order of passenger ***p₁***, ***p₂*** will be finished by driver ***d₁***. With the appearance of ***d₂***, ***p₂*** may be picked up by another driver ***d₂***. So the average void cruising distance proportion in this case changes from $\frac{1}{3}\left(\frac{l_1}{l_1+l_2} + \frac{l_3}{l_3+l_4}\right)$ to $\frac{1}{3}\left(\frac{l_1}{l_1+l_2} + \frac{l_5}{l_5+l_4}\right)$, which leads to the change of invalid emission proportion.

Figure 2(c) shows the situation of the supplement of both driver and passenger scales. When driver ***d₂*** and passenger ***p₃*** are added to the scenario, the passenger ***p₁*** may instead call for another driver ***d₂*** to pick him or her up. By doing this, driver ***d₂*** probably firstly turns to pick up passenger ***p₃*** and then picks up his original passenger ***p₂***. The average void distance proportion in this case changes from $\frac{1}{2}\left(\frac{l_1}{l_1+l_2} + \frac{l_6}{l_6+l_7}\right)$ to $\frac{1}{3}\left(\frac{l_8}{l_8+l_2} +\right.$

$\frac{l_3}{l_3+l_4} + \frac{l_5}{l_5+l_7}$). And of course, like former two situations, the emission performance shifts.

Besides the elaborations above, it should be pointed out that the scale of available drivers will also affect the emission performance from other perspectives. It can be illustrated in Figure 3. (a) the original situation (b) The scale of drivers decreases. (c) The scale of drivers decreases more. With the reduction of scale of drivers, the waiting time for a driver to take the order varies (maybe increases, maybe not). If one passenger waits for too long, he or she will be considered to give up the ride-hailing and take other transportation mode. Thus, affecting the emission performance of the ride-hailing system. With the same reason, the driver may also quit serving if he or she waits too long for taking orders.

Figure 3(a) gives the description of original situation. The three passengers $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ will be assigned to the three separate drivers $\boldsymbol{d_1}$, $\boldsymbol{d_2}$, $\boldsymbol{d_3}$. However, in Figure 3(b), when $\boldsymbol{d_3}$ is removed from the scene. Suppose $\boldsymbol{p_1}$, $\boldsymbol{p_3}$ will be assigned to the same driver $\boldsymbol{d_1}$ and $\boldsymbol{p_2}$ to $\boldsymbol{d_2}$. In this situation, $\boldsymbol{p_3}$ has to keep waiting until $\boldsymbol{d_1}$ drops $\boldsymbol{p_1}$ and comes to pick him or her. If the waiting time is beyond the tolerance of $\boldsymbol{p_3}$, $\boldsymbol{p_3}$ will cancel the order. Figure 3(c) is a deeper case: $\boldsymbol{d_2}$ is also removed from the scenario. The three passengers $\boldsymbol{p_1}$, $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ must be assigned to the only $\boldsymbol{d_1}$. At this point, the $\boldsymbol{p_2}$, $\boldsymbol{p_3}$ may both cancel their orders if the waiting time is beyond their tolerance respectively.

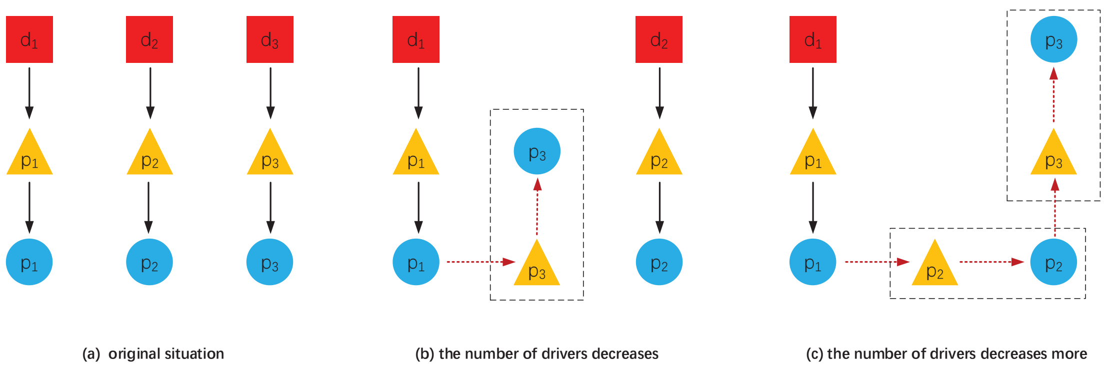


Figure 3. Illustration of affection of driver scale on the emission performance

In summary, whatever kind of change of user scale will both directly or indirectly influence the emission performance of ride-hailing system. It is important to study different scales of drivers and passengers and find out how the scale of drivers and the scale of passengers impact the emission performance to promote the future work. Due to the complex changes and multiple possibilities, a comprehensive simulation method is necessarily adopted to evaluate this relationship.

3.1.3 Research framework

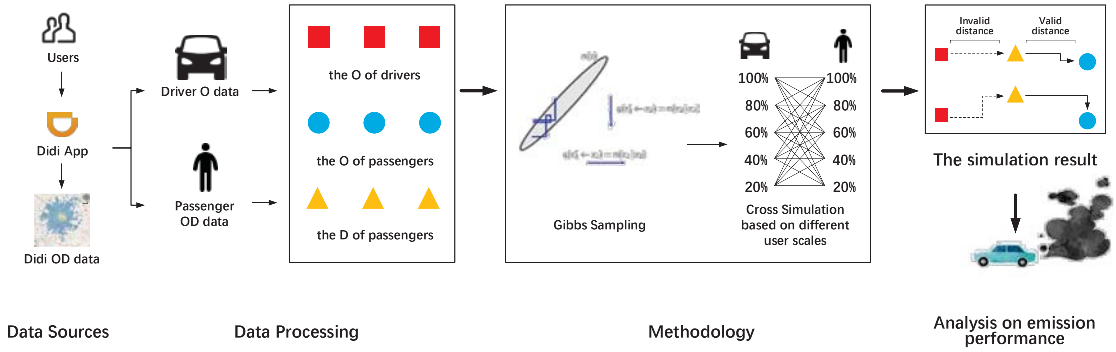


Figure 4. Flowchart of data processing, method used in this study and result to be achieved

Figure 4 illustrates the process of the whole simulation method. The solid line refers to the process direction. The Didi Apps in users' (both passenger and driver) phones collect users' GPS data. The data source is from Chengdu, China, where the ride-hailing service has been operated for over 6 years. According to the report made by Didi Media Research Institute[32], the times of ride-hailing services is beyond the local taxi services and ride-hailing system serves over 1.4 million times one day. Thus, the dataset is suitable for study. The full dataset is suitable to be considered as saturated. After receiving these data, we distract the origin of driver (referring to Driver O in the figure) as well as the origin and destination of passenger (referring to Passenger OD in the figure). These data also include the time stamps. In this paper, our method wants to find the relationship between the scale of users and the emission performance, our simulation method tries to study the situation under different user scale. We choose to sample from the dataset by certain percentage to stimulate the scenarios of different scale of regular users in ride-hailing system. Therefore, the result varies based on the partial dataset that is sampled in the simulation. So, instead of showing a single result, the mean of some times of simulation and their 95% confidence intervals will be illustrated. We use Gibbs sampling to sample these data by different percentages. The detail will be elaborated in the methodology part. For a full simulation, the percentage is determined by 100%, 80%, 60%, 40%, 20%. A cross simulation model is adopted to obtain the result we want under different combinations of scale of drivers and passengers to make the model more universal. It also needs to be noticed that since we focus on analyzing the minimal potential invalid emission proportion, so the situation of 100% scale of drivers and 100% scale of passengers will also be put in simulation instead of purely calculating from the dataset. The detailed methodology will be elaborated in the rest of method section.

The assumptions made in this part is shown as follows,

(1) In the sampling, the probability of each user to adopt the ride-hailing and sampling of data is in independent distribution. This can be referred by the principle of Gibbs sampling [33].
(2) Abnormal orders are not considered. Those trips with the same origin and destination or very short distance will be removed from the simulation set [34].
(3) The cost of each passenger and driver to adopt the ride-hailing will be considered. If a passenger waits for a driver to take the order longer than 15 minutes [34], the order will be cancelled.
(4) The rejection of order out of driver’s personal issue is not allowed [35].
(5) The participation of driver is in long term, which means that no driver will quit the system until the simulation is over. During the idle time, the driver will park their cars nearby the drop off location[36].

3.2 Ride-hailing dispatch based on prediction and optimization

3.2.1 Metric of evaluation

In this section, we will introduce the metric to evaluate the performance of proposed algorithm. As elaborated in the previous section, void cruising distance proportion is a good metric as it can clearly show and quantify the percentage efficiency of time, distance as well as energy. The equation of void cruising distance proportion can be implemented as:

$$P_{VD} = \frac{D_V}{D_V + D_D} \tag{1}$$

where, P_{VD} is the void cruising distance proportion; D_V is void cruising distance; D_D is delivery distance.

This metric is also adopted in the work by [14], where it is defined as efficiency. However, only comparison on void cruising distance proportion seems not enough. In work by Korolko[11], they also took the waiting time of passenger into consideration. This metric means to make sure the service quality because usually passenger won't wait for someone to pick him up for too long in real application. In this study, we also consider the cancellation of order when the someone wait for too long. The low cancellation rate Thus, in this study, we mainly consider three metrics listed as void cruising distance proportion, waiting time of a passenger to wait for a driver to pick him up and cancel rate of order.

3.2.2 Research Framework

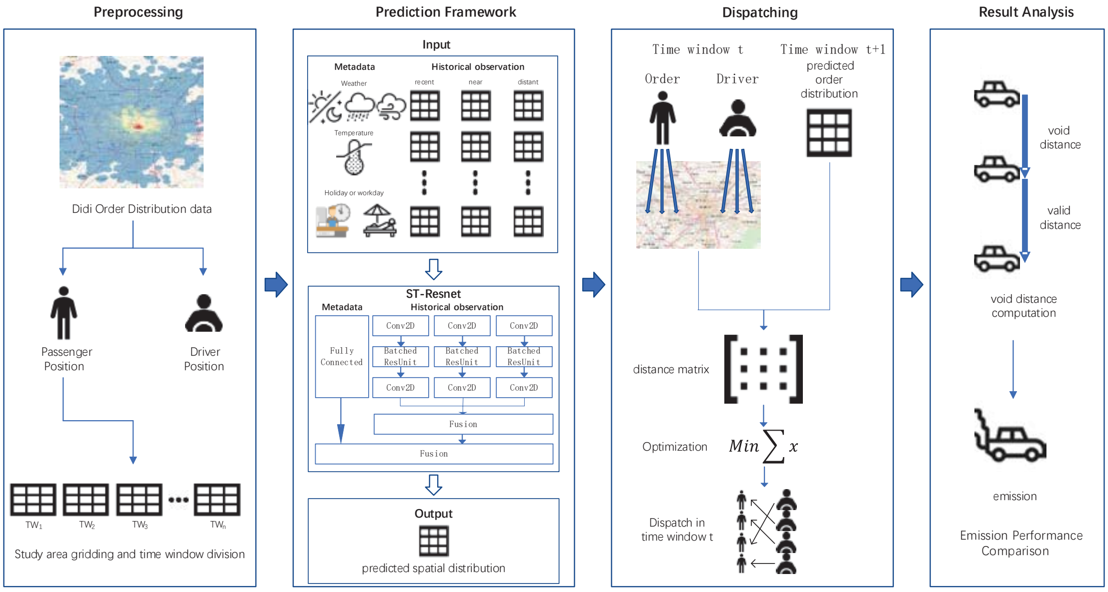
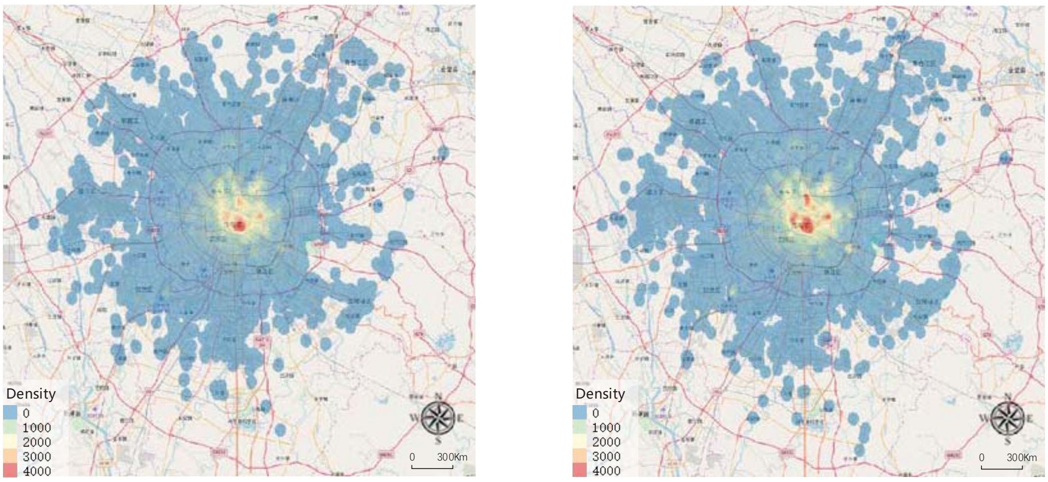


Figure 5. The framework used in this study

The research framework used in this part in shown as figure 5. The data preprocessing part is similar to the one in the part of evaluating the relationship between user scale and emission performance. But we added one more step, which is to preprocess the OD data of orders into the desired format of prediction model. The detailed steps will be elaborated in the Methodology part. The proposed methodology can be divided into two parts: the prediction part and dispatching part. Prediction part is mainly responsible for predicting the distribution of travel demand in the future based on a deep learning method. The input of this deep learning model requires historical observation corresponding metadata and the output is the predicted spatial distribution of riding demand. The dispatching part focuses on optimizing the dispatch under the consideration of minimizing void cruising distance proportion based on the predicted distribution result in the future. To better combine two parts. We adopt the time window method. The order and driver will be divided by serial time window as input[37]. For each time window, the prediction and optimization method will be separately operated once to decide the assignment of driver to order. To better show the utility of proposed algorithm. We will use greedy algorithm operated in time window as baseline and compare the performance of two algorithms in result analysis. The assumptions made in this part similar to the one in the part of Emission performance and user scale.

4. Study case



(a) origins of all orders (b) destinations of all orders

Figure 6. Thermal map of order OD data on one day in Chengdu

The GPS Dataset used in this study is provided by Didi. It includes the detailed information (Order ID, the time of the start and end of order, the longitude and latitude of origin and determination) except privacy information within one month in Chengdu City ranging from November 2nd, 2016 to November 30th, 2016 except for some vacancies on 10th and 8th. There are more than 6.6 million order records in the dataset. We exclude the orders with travel distance that are shorter than 500m, with same origins and destinations as well as same start time and end time and very few ones that travel among cities[38]. After the preprocessing, 75.4% of orders remains. They will be used for the simulation and analysis. The visualization of dataset in one day can be shown as Figure 6. Most of the origins and destinations are located at the city center. The rest are distributed in the surrounding area. The distribution of ODs covered almost all the urban area of Chengdu. The dataset can describe the case study well.

5. Methodology

5.1 Emission performance and user scale

5.1.1 Reassignment system

The definition of availability of a driver. If driver ***i*** is taking another passenger to the destination or on the way to pick up in the time interval that passenger ***j*** can wait for a driver to pick him up, the driver ***i*** is considered unavailable to passenger ***j***, else he is available.

When we sample a part of dataset by certain percentage to simulate the situation of different number of users in the system. It is necessary to reassign the drivers to pick up orders and simulate a process of operation. Then, the emission performance can be computed. Under different scales of drivers and passengers in the simulation, an order from a passenger may be reassigned to another available driver based on their original GPS information to maximize the efficiency of whole operation system. The reassignment algorithm is based on the shortest void distance principle which means when an order is given in the system, it will be assigned to the driver that is available and nearest to the position of the passenger. This reassignment method will ensure the least distance for a driver to pick passenger up. If all drivers are not available in the time period passenger wants to be picked up, the order will be cancelled.

The code form of the reassignment algorithm is shown as algorithm 1. Notice that the whole simulation code is similar with the form of Python code. The meaning of mathematical symbols can be found in Nomenclature.

Algorithm 1 —— Reassignment

Input: single piece of order *i* & D_s

Function: Find the available and most suitable driver for order *i*

Output: the record of reassignment

for each driver $\boldsymbol{j} \in \boldsymbol{D_s}$:

    if $\boldsymbol{t_a} \leq \boldsymbol{t_o} + \boldsymbol{\theta_{max}}$ (driver $\boldsymbol{j}$ is available to order $\boldsymbol{i}$):

        driver $\boldsymbol{i} \longrightarrow \boldsymbol{A_i}$

    else:

        go to the next driver $\boldsymbol{j}$ +1

    if $\boldsymbol{A_i} = \emptyset$:

        go to the next order $\boldsymbol{i}$ +1

    else:

        find the $\boldsymbol{md_i} \in \boldsymbol{A_i}$

        calculate $\boldsymbol{t_p}$

        assign the order $\boldsymbol{i}$ to $\boldsymbol{md_i}$

        update the original position and available time of $\boldsymbol{md_i}$ in $\boldsymbol{D_s}$

return the record of $\boldsymbol{md_i}$ & order $\boldsymbol{i}$

After an order is reassigned to the driver, the driver will be considered not available to other passengers until he or she finishes the job. Then the GPS position and time stamp of driver's OD will be updated according to the order. The reassignment will come to an end until all the orders are assigned to the drivers or cancelled.

5.1.2 Gibbs Sampling for the generation of simulation samples

Sampling method plays an important role in the whole simulation process. In this study, the Purely randomly selecting one part of data and putting them into simulation is not suitable because the simulation result deeply depends on the selection of data. The multiple possibilities should be considered. Additionally, the spatial-temporal pattern of demand and server varies due to regions, so blind sampling can't assure the full characterization of simulation results. As a result, a Gibbs Sampling is adopted in this study. Gibbs sampling can generate different simulation samples and consider the issue of full sampling. Ride-hailing is a high dimensional system. The behavior of each user, either passenger or driver, can cause a chain effect on the operation of whole system. Thus, we can imagine the relationship between the emission behavior and users as a function: emission behavior = f ($user_1$, $user_2$, … …, $user_n$). Each parameter of function is a single user. Thus, it is obvious that the function is high dimensional. The sampling on the emission behavior is difficult. Gibbs sampling is used to solve such sampling problem with high dimension. The principle of Gibbs sampling is to simulate the function step by step and slowly cover the main value of function. The sampling method randomly select a set of parameters for the first time of simulation (Suppose the number of parameters is n). Then, it will randomly change one parameter and remain the rest n-1 ones, then operate another time of simulation. This process will be iterated repeatedly. In each iteration, the simulation result will be recorded. If the variance of the simulation result in the last several times of simulation is lesser than a threshold, the sampling process can be thought to be completed and come to an end. In this study, the dimension of problem is very high, which can be up to 250, 000 and down to 50,000. Therefore, if replacing one user at one time of sampling is unreasonable. It can hardly affect the final result. Instead, we choose to replace thousands of users after each time of simulation. The detailed quantity depends on the percentage of users used in the simulation. In this study, firstly, we randomly pick a certain percent of data in the whole dataset as the simulation samples. The rest of dataset will be used as backup set. After one time of simulation, a part of simulation samples will be randomly replaced by the data in the backup set for another time of simulation. To ensure the full consideration, the whole simulation process will be finished until it is considered to satisfy the convergence condition. In this study, $\boldsymbol{e}$ refers to the variance of the mean of all the void cruising distance in last five times of simulation. When $\boldsymbol{e}$ is smaller than the convergence condition, the whole sampling process can be considered to be completed. Algorithm 2 gives the code form of Gibbs sampling.

Algorithm 2 —— Gibbs sampling

Input: $\boldsymbol{O}$, $\boldsymbol{D}$, $\boldsymbol{P_d}$ & $\boldsymbol{P_o}$

Function: find the distribution of result of simulation

Output: the results of all times of iterations

$\boldsymbol{O_s} \longleftarrow$ randomly selected $\boldsymbol{P_o}$ of data $\in \boldsymbol{O}$

$\boldsymbol{D_s} \longleftarrow$ randomly selected $\boldsymbol{P_d}$ of data $\in \boldsymbol{D}$

$\boldsymbol{O_b} \longleftarrow$ the rest (1 - $\boldsymbol{P_o}$) of data in $\boldsymbol{O}$

$\boldsymbol{D_b} \longleftarrow$ the rest (1 - $\boldsymbol{P_d}$) of data in $\boldsymbol{D}$

while($\boldsymbol{e} > \boldsymbol{e_{max}}$):

  for each order $\boldsymbol{i} \in \boldsymbol{O_s}$:

    record of reassignment $\longleftarrow$ **Reassignment**(order $\boldsymbol{i}$, $\boldsymbol{D_s}$)

  calculate the results obtained

  calculate the $\boldsymbol{e}$

  randomly selected a part of data $\in \boldsymbol{O_s} \longleftrightarrow$ randomly selected a part of data $\in \boldsymbol{O_b}$

  randomly selected a part of data $\in \boldsymbol{D_s} \longleftrightarrow$ randomly selected a part of data $\in \boldsymbol{D_b}$

return the results in all times of iterations

5.1.3 Cross simulation module generation

Since the aim of this method is to find the difference of emission performance under different scale of users, a full simulation is needed to get the complete result. Thus, a cross simulation of different scales of samples is adapted. The original Data sets of drivers and passengers are randomly sampled by the percentage of 20%, 40%, 60%, 80% and 100% separately. Then, the cross simulation will be applied. It means all 5 kinds of percentage of drivers and passengers will group with each other as the input of simulation model. Then, the cross simulation will give the output under different scenarios we need for analysis.

The flowchart form of the whole simulation process is shown as Figure 7. Certainly, the whole simulation process will also include the step to deal with data processing. $\boldsymbol{k}$

in the figure refers to the percentage of scale of drivers; ***l*** in the figure refers to the percentage of scale of orders.

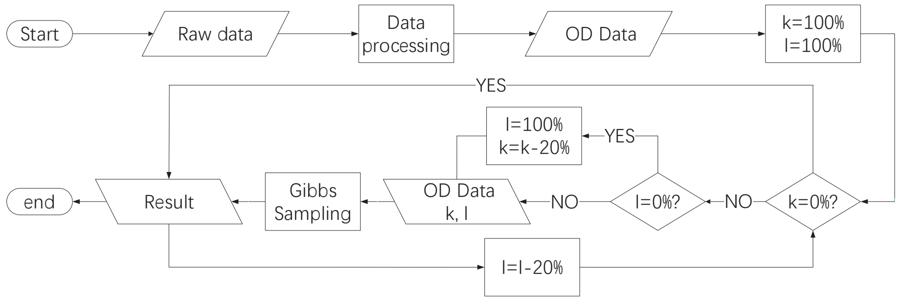


Figure 7. Flowchart of whole simulation process

5.1.4 Equation for result computation

In the result of every time of simulation, except the OD of each assigned or cancelled order and driver, the results obtained from the cross simulation includes the distance, time interval that driver travels to pick up the passenger and complete order; the time length that drivers wait for an order to come; the time that passengers give the orders. The orders that originally existed in real world but cancelled in the simulation. Thus, we can compute the void distance proportion, total distance that might be travelled or not be travelled.

After obtaining these results, we can compute the invalid emission proportion, the quantity of invalid emission, the quantity of total emission and other results. The equations we use to link these two parts are from the COPERT (COmputer Program to calculate Emission from Road Transportation), a manual written by European Environmental Agency. Since the measured standard which COPERT complies with is similar to the one adopted in China, the model is suitable for evaluating vehicle emission in Chinese cities [39, 40]. This manual provides the methodology and equations to calculate the emission by the cars and they are suitable to all countries. This manual contains methods to compute various kinds of emission. The common equation of emission computation can be written as:

$$Q_{i,k} = \sum_j q_{i,j,k} l_j \quad (2)$$

where, the $\boldsymbol{Q_{i,k}}$ refers to the quantity of emission $\boldsymbol{k}$ of vehicle $\boldsymbol{i}$; $\boldsymbol{q_{i,j,k}}$ refers to the hot emission factor of emission $\boldsymbol{k}$ vehicle $\boldsymbol{i}$ during trip $\boldsymbol{j}$; $\boldsymbol{l_j}$ refers to the length of trip $\boldsymbol{j}$. $\boldsymbol{q_{i,j,k}}$ can be calculated by:

$$q_{i,j,k} = \frac{\alpha_k v_{i,j}^2 + \beta_k v_{i,j} + \gamma_k}{\varepsilon_k v_{i,j}^2 + \zeta_k v_{i,j} + \eta_k} \tag{3}$$

The detailed information to calculate the emission factor $\boldsymbol{q_{i,j,k}}$ can be found in work by [38]. Since the quantity of emission mainly depend on the driving distance, the pattern among each kind of emission is same. Thus, in this study, we will broadly elaborate on the emission pattern. Then, the total invalid emission quantity, invalid emission proportion can be calculated by the equation set (4)

$$\begin{cases} P_{void} = \frac{d_{void}}{d_{total}} \\ P_{e_{invalid}} = f(P_{void}) \\ E_{invalid} = EP_{e_{invalid}} \end{cases} \tag{4}$$

where, P_{void} is the proportion of void distance; d_{void} means the void distance that one driver travel to pick up one passenger and d_{total} is the total distance a driver travels to finish an order from a passenger. It includes $d_{invalid}$ and the distance that driver travels from the origin of passenger to destination. $P_{e_{invalid}}$ is the proportion of vehicle invalid emissions; $E_{invalid}$ is the total quantity of vehicle invalid emission. $f(x)$ is the function turns proportion of void distance to proportion of invalid emission. E is total vehicle emission.

5.2 Ride-hailing dispatch based on prediction and optimization

5.2.1 Time window Division:

Time window division is a classic method used in many real-time dispatch problems[41]. One characteristic of this method is easy to understand and implement. The illustration of time window division is shown as figure 8.

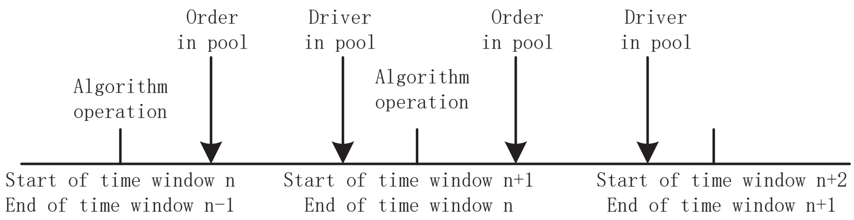


Figure 8. Time window division

The process of time window division can be treated as group of time line. The end of each time window is also the beginning of last one. Suppose the length of time window is l, number of time window is N, and the start time of simulation is 0. During the period of time window n, $n \in [1, 2, \dots \dots, N]$, the orders and available drivers given between time $(n - 1)l$ and nl will be collected in the matching pool. Then, at the end of time window nl, the simulation algorithm will be operated to give the matching result. In this study, we choose the length of time window to be 5 minutes, which makes 8352 time windows.

5.2.2 Baseline: Greedy algorithm:

The baseline algorithm we choose is the greedy algorithm. Greedy algorithm is a classical algorithm used in many real-time dispatch studies of pick-up and delivery problems[35, 42]. In principle, the algorithm will iterate over every travel demand and find it the closest driver who can pick up the order or follow the rule of first-come-first-serve[43]. When an order is given into the system, the algorithm will search the current candidate drivers and assign the nearest one to order. Generally speaking, the algorithm only considers the optimized solution for each single object. Although this method is easy to implement and manage, it is naturally uncoordinated and tends to prioritize immediate passenger satisfaction over the global supply utilization. In the Result and analysis part, we will illustrate the performance of greedy algorithm.

5.2.3 Prediction Model

Travel demand prediction is currently a hot research topic in the field of Computer

Science[26, 44]. The main stream of current methods is deep learning. Deep learning model tries to stimulate the thinking process of human. People can feed the training materials to the deep learning model and make the model learn how to generate the desired output. The general structure of deep learning is shown in figure 9.

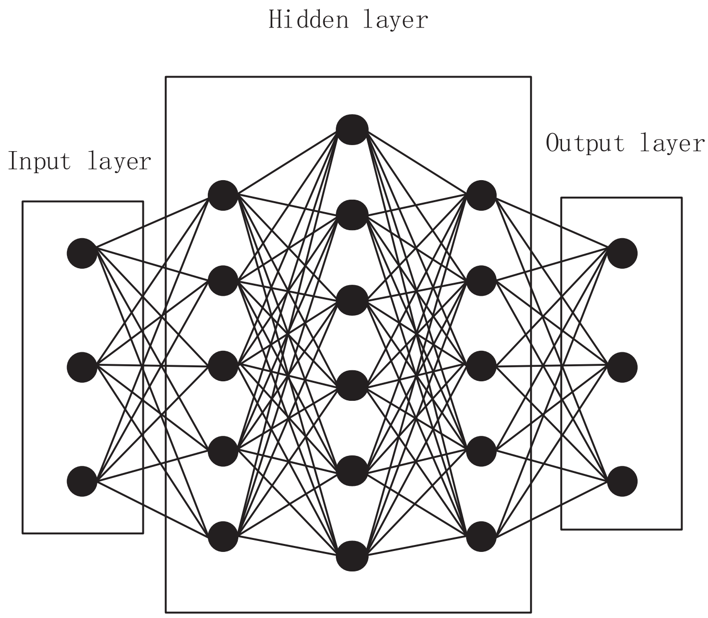


Figure 9. General structure of deep learning

The input layer is used to input the features. The hidden layers contain the parameters for computation. The output layer will give the desired output. During the training process, the model will try to find the optimal parameters in the hidden layers and minimize the difference between the ground truth and computed output.

Many researchers developed various deep learning neural networks that concern this problem. In recent years, there exists a lot of achievements, like convolutional LSTM neural network[45]. The prediction model we use in this study is the ST-Resnet[46]. This is a deep learning neural network based on residual unit. The structure of ST-Resnet is shown as figure 10.

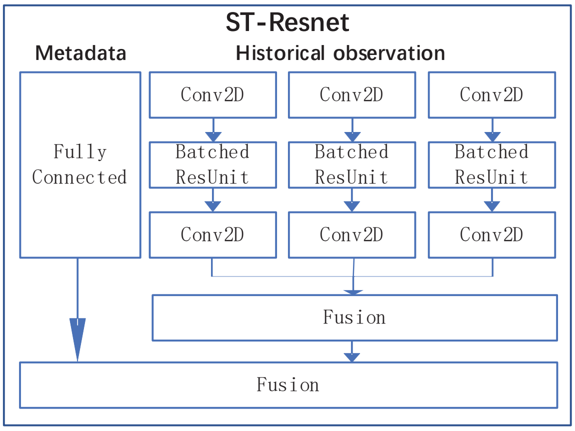


Figure 10. The graphical structure of ST-Resnet

The input of this prediction model is divided into two parts, which are separately historical observation and metadata. The output is the spatial distribution in the future. The desired input format of historical observation is matrix in essence. The preprocessing is needed to convert spatial data to matrix. Firstly, we extract the order data in each time window. Next, we apply the regional grid method[47]. The concept is to convert the spatial distribution data to image-like data, which is in the form of matrix. We divided the study area with square grid with the size of 3690 m * 3690 m. This makes the whole study area gridded with size of 40 * 40. Then, we count the quantity of orders in the area of each cell in each time window. The quantity of orders is the element of each matrix. Finally, we can get a sequence of matrixes abstracted from spatio-temporal data. Since the length of time window is 5 minutes, from preprocessing of the data in one month, we can get totally 8352 matrixes. In the prediction model, the input of observation is divided into three sequences, separately recent, near and distant. Recent: a sequence of continuous matrixes of historical observation closely before the time window we want to predict; near: a sequence of continuous matrixes of historical observation that is one day before near; distant: a sequence of continuous matrixes of historical observation that is one week before recent. We choose the length of sequence to be 6. Thus, if we use $x_n, n \in [2022,8351]$ to represent the output, the input can be represented as

$$X\begin{cases}\{x_i | i \in [n-6, n-5, n-4, n-3, n-2, n-1]\} \\ \left\{x_j \middle| j \in \begin{bmatrix} n-288-6, n-288-5, n-288-4, n-288-3, n-288-2, \\ n-288-1 \end{bmatrix}\right\} \\ \left\{x_z \middle| z \in \begin{bmatrix} n-2016-6, n-2016-5, n-2016-4, n-2016-3, \\ n-2016-2, n-2016-1 \end{bmatrix}\right\}\end{cases} \tag{5}$$

The input dimension of historical observation is 40 x 40 x 6 x batch size.

Another part of the input is the metadata. Generally speaking, metadata includes all the information that can have impact on the spatial-distribution of order. In the original paper, the author used the weather data and date information. Thus, in this study, we marked the hour that time window is located in in one day, the day in one week, the week in one year of each time window, the mark of whether the day is holiday or not(holiday is marked as 1; workday is marked as 0) as the date information. Meanwhile, we also use the weather and temperature data as the weather data. The table of detailed weather data is shown as follows:

Table 1. Weather data used in the prediction model

Weather	Highest temperature of day	Lowest temperature of day
[partly cloudy, cloudy, sunny, little rainy]	[9, 22]	[4, 13]

The temperature data will be rescaled to [-1, 1] by equation:

$$\frac{tem - tem_{min}}{tem_{max} - tem_{min}} \times (t_{max} - t_{min}) + t_{min} \tag{5}$$

where, *tem* is the temperature we want to rescale; tem_{min} is the minimal observed temperature; tem_{max} is the maximal observed temperature; t_{max} is the upper bound of rescaled interval; t_{max} is the lower bound of rescaled interval.

We have introduced all the input data needed in this prediction model. In the following part, we will elaborate on how the model deal with input and get trained for prediction.

For metadata part:

The date information and weather information will be separately turned to numerical data by one-hot encoding method[48]. Then, two one-hot encoded data will be concatenated together as one matrix. The feature of metadate information will be extracted by a fully-connected layer implemented as:

$$Y = Relu(aX + b) \tag{6}$$

Y is the output feature; *X* refers to the on-hot encoded input numerical data; *a* is the transform matrix; *b* is the bias matrix; *Relu* is activation function which can be implemented as:

$$Relu(x) = \begin{cases} 0, & x < 0 \\ x, & x \geq 0 \end{cases} \tag{7}$$

The output feature will be then reshaped to the size of 40 x 40 x 6 as the same size of historical observation.

For historical observation part:

As shown in figure 8, there are three individual input channels for recent, near and distant. We will introduce the structure of one channel as all three channels are the same. The first layer is a 2D convolutional layer with the kernel size of 3 x 3 and 64 filters that extract the feature of input sequence to the matrix of size 40 x 40 x 64. Then, the following part is a sequence of residual units. The job of each residual unit is to deeper analyze the features. Suppose we adopt *l* residual units in the sequence. The universal equation can be implemented as:

$$X_{l+1} = X_l + F(X_l; \theta_l) \tag{8}$$

where, X_{l+1} is the output of residual unit; X_l is the input; *F*(*x*) is the function of residual function; θ_l refers to all the trainable parameters in the residual function.

The residual function contains two groups of combination of *Relu* and convolution. Thus, this makes the equation of residual function:

$$Y = F(X_l; \theta_l) = Relu(W * X + b) \tag{9}$$

where, Y is the output of residual function; W is the transform matrix; * is the convolution operation; b is the bias matrix.

Notice that between two groups, there is a batch normalization layer[49].

After an iteration in the sequence, the result will go through the final 2D convolutional layer and fused together. The method of fusion is to add the matrixes from three channels together to one. Then, this one matrix will be added with the reshaped output feature of metadata. Finally, the summed-up feature will be handled by a Tanh function and turned to be the output of prediction result.

During the training process, we use the former 20% of dataset as the test set and latter 80% as the training set. The optimizer for the gradient descent is Adam[50], which has shown a better performance among all the optimizers. Adam tries to solve the problem of gradient descent with the idea of moment. The first and second moment can be computed as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{1}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) {g_t}^2 \tag{2}$$

where m_t and v_t are moving averages, g_t is gradient on current mini-batch, and β is a new introduced hyper-parameter of the algorithm. They have really good default values of 0.9 and 0.999 respectively. The vectors of moving averages are initialized with zeros at the first iteration. Thus, $m_0 = 0, v_0 = 0$.

In principle, the first and second order will be corrected by:

$$\widehat{m_t} = \frac{m_t}{1-{\beta_1}^t} \tag{3}$$

$$\widehat{v_t} = \frac{v_t}{1-{\beta_2}^t} \tag{4}$$

Finally, the parameter of model can be updated by:

$$w_t = w_{t-1} - \alpha \frac{\widehat{m_t}}{\sqrt{\widehat{v_t}}+\epsilon} \tag{5}$$

w is the parameter of model; α is the learning rate and ϵ is the hyperparameter that is 0.0001 in default.

By far, we have completed the introduction of prediction model. In the result analysis, we will illustrate the accuracy of prediction model.

5.2.4 Optimization

Optimization is a classical mathematical method used in many research fields including the ride-hailing[51]. Generally speaking, the concept of optimization is to optimize the objective function and find the globally optimal solution. Then performance of optimization algorithm deeply depends on the knowledge of conditions, which, in this case, is the information of drivers and orders. More information can help us get more optimal solution.

The core of the dispatch strategy used in this study is that when we consider the dispatch problem in time window n, besides the existing knowledge of drivers and orders in time window n, the predicted distribution of orders in time window n+1 will also be taken into account. Thus, the objective function is to minimize the global pick up distance in both time windows:

$$\min(f = \sum_i \sum_j S_{i,j} D_{i,j}) \tag{10}$$

subject to:

$$\sum_j S_{i,j} \leq 1 \tag{10.1}$$

$$\sum_i S_{i,j} \leq 1 \tag{10.2}$$

where, $S_{i,j}$ is the decision variable that decide whether driver i to pick up passenger j or not; $D_{i,j}$ is the distance of driver i to pick up passenger j.

Constraint 10.1 aims to ensure that one driver can be maximally assigned with one order; constraint 10.2 aims to assure that one order can be maximally assigned with one driver. Here, we can choose to impose one more constraint:

$$\sum_i \sum_j S_{i,j} = \max(\mathrm{I}, \mathrm{J}) \tag{10.3}$$

This constraint can serve the purpose of trying best to satisfy orders with current available drivers. The difference is that without the constraint, a part of drivers will not be dispatched to the order in the optimized solution because there may be an order that is much closer to him, while with the constraint, if there is no other candidate driver for the order, the driver will be dispatched in the current time window. In this study, we will also compare the performance of algorithm with and without the constraint.

From the target function, the problem is an ILP (integer linear programming) problem. The basic method to solve ILP is Simplex algorithm[52]. Its basic concept is to firstly construct an initial solution, which is a feasible and finite solution. If the initial solution isn't the globally optimal one, then, the algorithm will introduce non-base variables to replace a base variable for a better solution. The iteration is repeated until the globally optimal one is found. The process of optimization algorithm shown as flow chart 11.

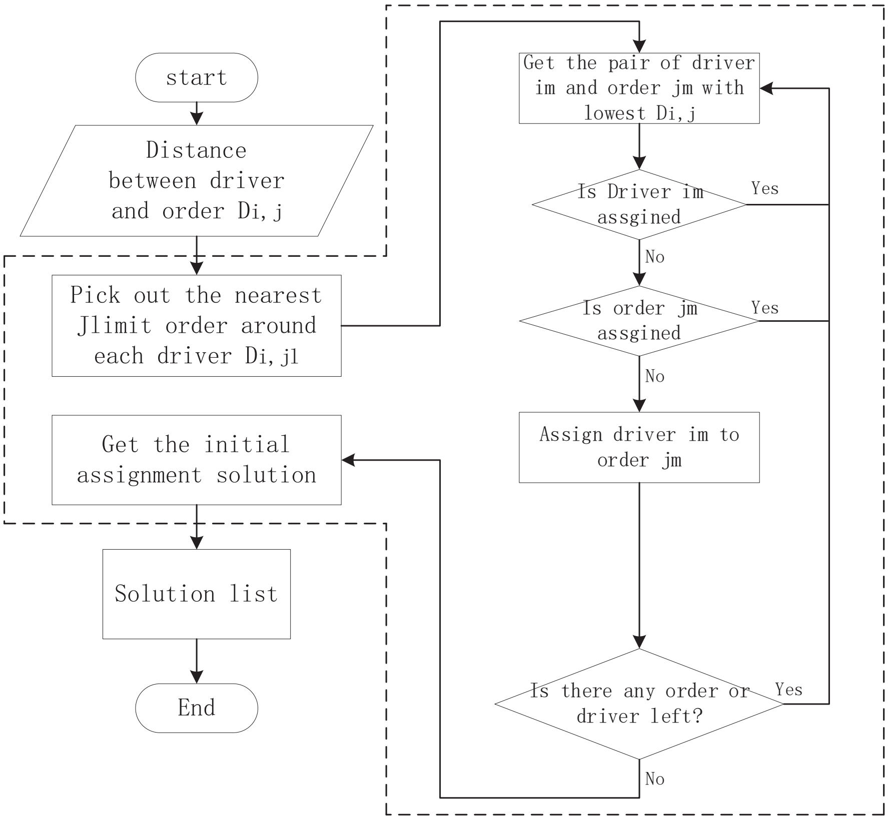


Figure 11. Flow chart of optimization algorithm

In the first step, we construct the distance matrix of each pair of driver and order. The driver list only contains the available drivers in current time window *n*; the order list contains orders both in current time window *n* and future time window *n*+1. The column of matrix refers to the list of drivers and the row refers to the list of orders. The element $D_{i,j}$ means the distance between the position of driver *i* and origin of order *j*. Because we can only predict the number of orders in each cell, we lack the information of exact spatial distribution of orders in each cell. Therefore, it's hard to compute the exact distance between predicted orders and existing drivers. To solve this problem, we furtherly divide each cell of 40 x 40 grid into a 10 x 10 grid. (See figure 12).

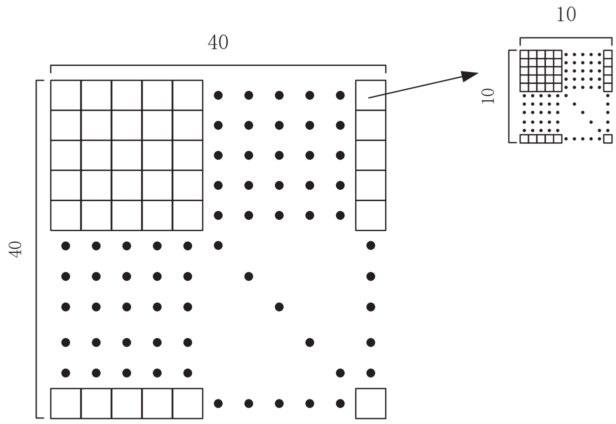


Figure 12. Cell division

Each cell of larger grid can be furtherly divided by a 10 x 10 grid. Then, we do statistics on the historical spatial distribution of orders in each 10 x 10 grid. Each cell of smaller grid will be marked by a probability that an order may exist in this cell. Based on the statistic result, we can estimate where the orders may be located in if there are orders predicted in the larger cell. The predicted orders will be allocated in the cell with higher probability. We assume the location of predicted orders at the spatial center of cell of 10 x 10 grid. After that, we can compute the distance between driver and predicted orders. Then, the following part is more like a greedy algorithm, where we pick out the pair of order and driver in the ascending order of distance to construct an initial solution of matching. In the next step, we try to improve this solution by iteration until we get the optimal solution. Since the result is the decision variable $S_{i,j}$, all variables are non-negative, which satisfies the standard form of simplex algorithm. The flow chart of simplex algorithm is shown as Figure 13.

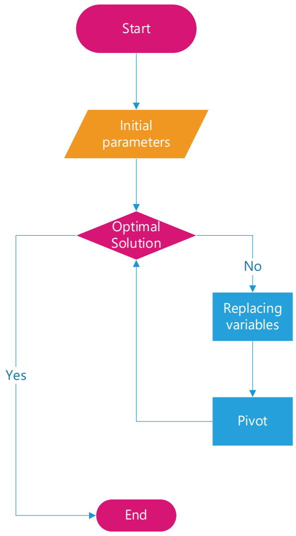


Figure 13. The flowchart of simplex algorithm

The algorithm will firstly check whether the initial solution is the optimal one. If not, the algorithm will introduce basic variables into the constraints and turn the constraints and target function to set of equations. Then, the algorithm will randomly choose a variable with positive coefficient for example x_i to increase. Next, the algorithm will choose the strictest equation with x_i and rewrite the equation to one with only x_i on the left-hand of equation. This is used to replace all x_i. This is called pivot. Finally, if all the coefficients are non-positive, the solution reaches the optimal one, else the above process will be repeated.

6. Result analysis

6.1 Emission performance and user scale

6.1.1 Void Distance Proportion Analysis

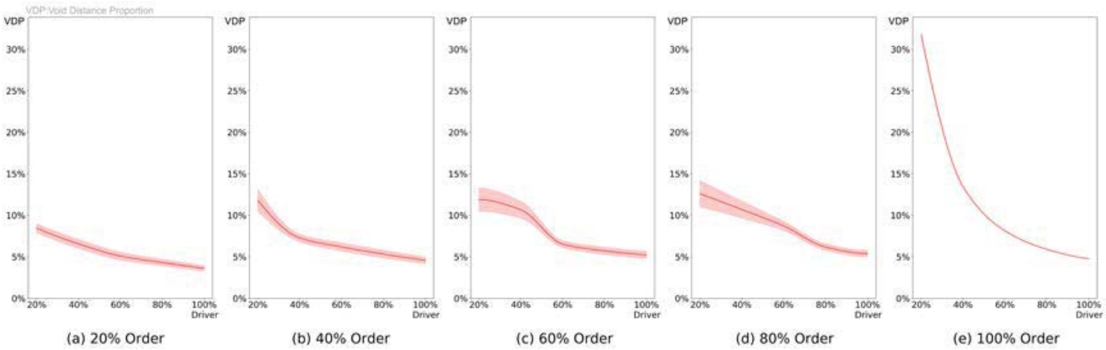


Figure 14. Plot of void cruising distance proportion

The plots of simulation results are shown in Figure 14. The red solid curve refers to the mean of void cruising distance proportion. The area with the same color is the 95% confidence interval. The 5 subfigures show the difference of void cruising distance proportion when the scale of orders is set to a certain percentage. (a) Orders = 20% (b) Orders = 40% (c) Orders = 60% (d) Orders = 80% (e) Orders = 100%. If we remain the scale of orders, with the increase of scale of drivers, the average void cruising distance proportion is decreasing in the five subfigures. It infers that when the mobility demand in the area is determined, the supplement of drivers to the shared transportation system will help decrease the mean void cruising distance proportion. By the comparison of these five figures, the linear relationship weakens by the increase of percentage of orders. When the scale of travel demand is at the 100% level, the void distance proportion increases rapidly with the decrease of scale of drivers; while it increases much more stably at the 20% level. This tells us that when the travel demand in the system is large in an area, the addition of demand satisfiers will effectively reduce the void cruising distance, thus, reducing the invalid emission. Here is the possible explanation. Consider the situation where the travel demand is huge while the servers are far not enough, the cars are busy trying to satisfy every travel demand in the system as soon as possible, so when an order is given into the system, there are few candidate drivers in the matching sequence. As common knowledge, it is hard to find a driver who can pick up the order with a short void cruising distance. From the analysis above, the effectiveness of supplement to improve the emission performance depends on the total travel demand, the greater the demand is, the more effective the supplement is. Besides, if we compare situations when the scale of drivers is set to 100%, even though the scale of orders varies, the void cruising distance proportion doesn't change much and the confidence intervals are still obvious in the figures until the scale of orders mounts to 100%. This indicates that the scale of orders is the leading factor of determining the void distance proportion. It is mainly because that order is the served object in the system. The ODs of orders served in the system won't change under any circumstances—The passenger will only quit the system when his or her travel demand is not satisfied. While the ODs of drivers are deeply influenced by the different orders in the system. So, when the travel demands in the system are relatively stable in the system, the void distance proportion is relatively determined.

These findings tell us that in real cases, when ride-hailing system is initially introduced in one area, an evaluation of ride-hailing market not only serves the purpose of assessing the potential travel demand and commercial value[53], the policy making of balancing the number of registered drivers in the system is also necessary. The operators can observe the void cruising distance proportion of historical service record to judge whether the current quantity of driver is enough to serve the travel demand. Overall, the relationship between the void cruising distance proportion and number of drivers is not a linear one. We can observe that the change of void cruising distance proportion is insensitive to the growth of driver numbers in the system when the proportion of driver is equal or larger than the one of orders. This indicates that the number of drivers has reached an equal or higher level with the number of passengers. The effort of making policies to attract more drivers is unnecessary. In this case, the critical ratio of number of drivers to orders is 1:9. What's more, we also found a rapid interval before the critical ratio, where the increase of driver number can more effectively decrease the void cruising distance proportion. These two key features can be a clear sign of the suitable number of drivers in the system. In the future, deeper parameters that can affect the ratio and overall model can be explored to transfer the conclusion to other areas.

6.1.2 Emission performance Analysis

After obtaining the simulation results, we can compute the invalid emission proportion of NO_x. The plots of calculation results are shown in Figure 15.

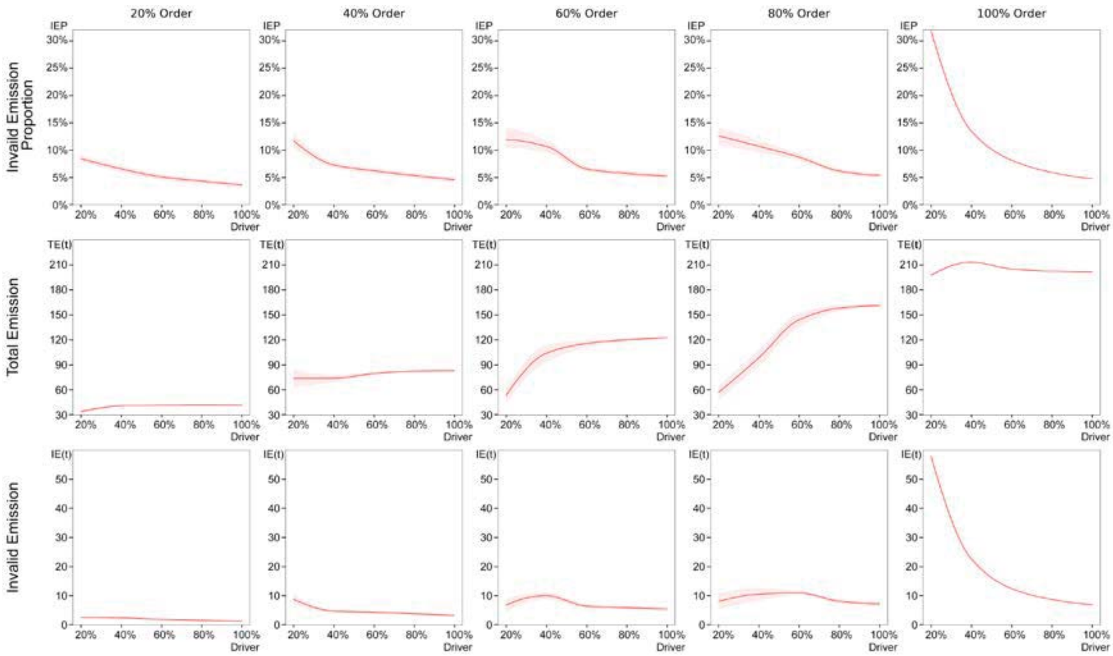


Figure 15. the emission of NOx

The trend and shape of invalid emission proportion curve is very similar to the one of void distance proportion, despite the variety of type of emission and quantity. So, we can say that the invalid emission proportion is closely related to the void distance proportion, a positive correlation. Thus, we can conclude that the emission performance is closely related to the user scale and the larger the demand scale is, the more sensitive the emission performance is to the driver scale. By far, we can bridge the impact of user scale on the emission performance. This impact is important for the future decision. With this impact, if some metrics that ensure the high efficiency of ride-hailing system want to be limited in a threshold, the scale of users that may satisfy this request can be computed. For example, if car-hailing service provider wants to restrict the average void distance into a certain value, we can find this high efficiency area of scale of users for decision makers as a guidance.

6.1.3 High efficiency area computation under metric constraints

This subsection aims at providing a sample to obtain the probable area of the user scale if we want some metrics to be considered to be limited into a range and does a little discussion. We choose three metrics for the discussion: maximal average invalid emission proportion, maximal average order cancel proportion and maximal average waiting time for a driver to take an order. The first metric is meant to ensure the emission performance wouldn't be worse than a certain value; the second metric is meant to ensure the interest of the ride-hailing service platform and the travel satisfaction proportion of passengers who adopt ride-hailing; the third metric is meant to ensure the interest of driver. Now we set two metric sets to simulate two scenarios at the purpose of showing the difference of area. Reference metric set: Average invalid emission proportion ≤ 9.2% Average order cancel proportion ≤ 35.8% Average waiting time for a driver to take an order ≤ 1.75h; higher performance metric set: Average invalid emission proportion ≤ 7.6%; Average order cancel proportion ≤ 13.2%; Average waiting time for a driver to take an order ≤1.50h.

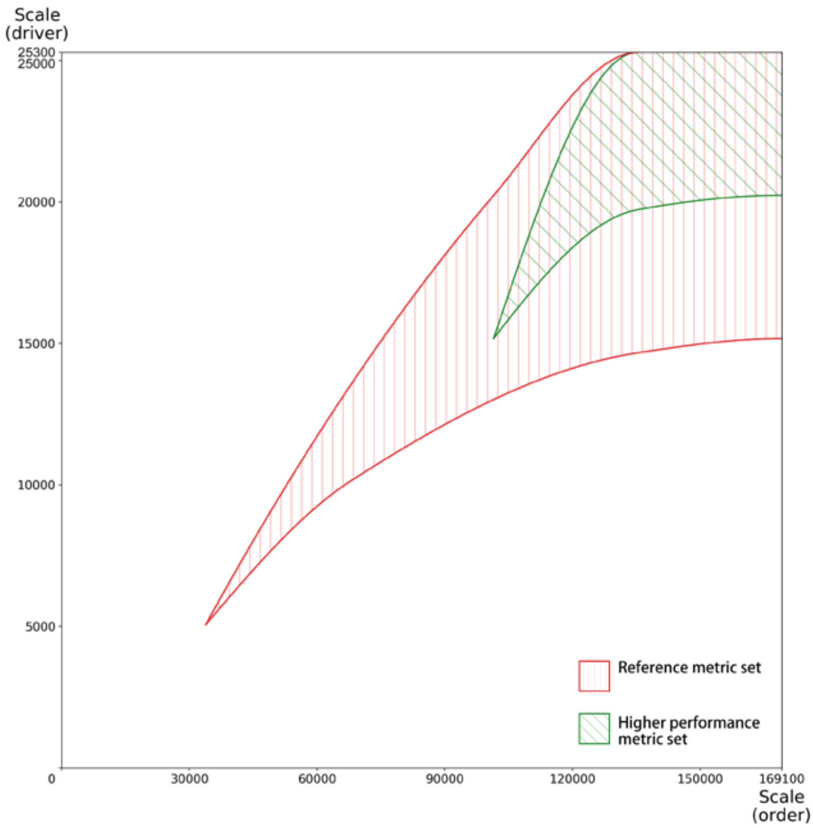


Figure 16. The area of scale of driver and order under two metric set

The area of the corresponding scale of drivers and orders can be shown in figure 16. The red one refers to the reference metric set and the green one refers to the higher performance metric set. In reference metric set, when the scale of orders and drivers are both greater than about 135,300 (near 80%) and 15,200 (near 60%) respectively, a square area, under any combination of quantity of driver and order, these three metrics can be satisfied anyway. When both are under this quantity, the area is restricted in a valley down straight to the situation where there are about 5100 drivers and 33,800 orders. It is near around the 20% of drivers to 20% of orders case. In higher performance metric set, we restrict the three metrics tighter, then we get the area of corresponding scale shown in Figure 10. It can be seen that the square area shrinks to around 20,300 (80%) for the axis of scale of driver. The valley area extension is limited to the point where X axis is approximately 101,400 (60%) and Y axis is roughly 15,200 (60%). These two cases indicate that if some metrics want to be realized, we can mine the history data and find the approximate user scale in the ride-hailing system.

6.2 Ride-hailing dispatch based on prediction and optimization

6.2.1 Result of prediction

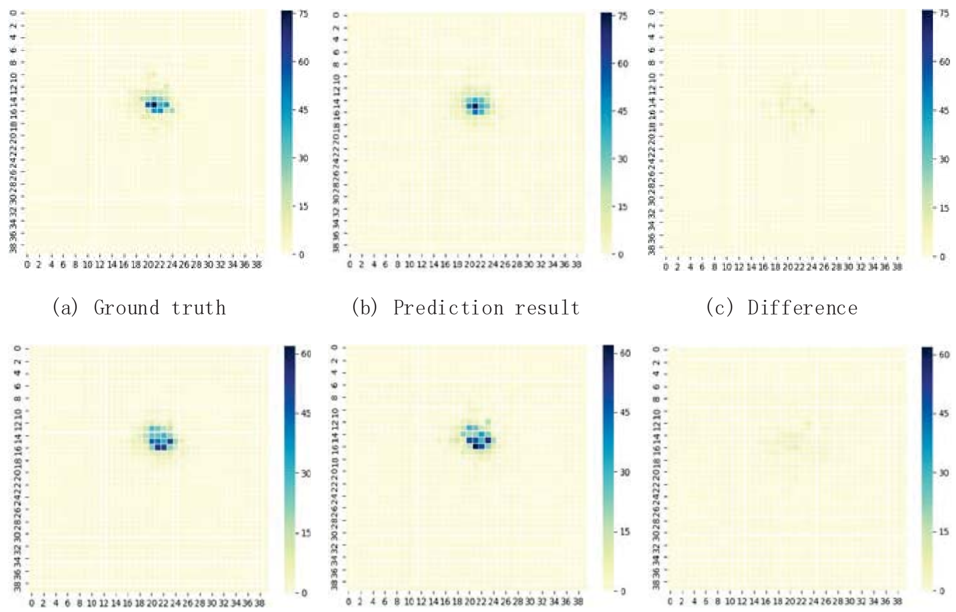
An important parameter of ST-Resnet is the number of residual units used in the model. In the original paper, author indicated that the more residual units in the neural network, the deeper the neural network is, the accuracy the prediction is. However, more residual units mean more memory usage during the training and slower training. To find a balance between the accuracy and computation resources, we choose the number of residual units to be 21. There are totally over 4.4 million trainable parameters in the model, which is quit a large quantity.

In the field of computer science, multiple forms of losses are used to evaluate the performance of prediction model like mse (mean squared error) and mae (mean absolute error). However, these losses are usually used to compare the performance among different prediction models and hardly give a direct impression on the accuracy. In this study, we adopt the mape (mean absolute percentage error) as the metric of accuracy, which can directly show the differences, is computed by:

$$A_{diff} = |A_{GT} - A_{PR}| \tag{11}$$

where, A_{diff} is the difference matrix; A_{GT} is the ground truth; A_{PR} is the prediction result.

After completing the training, we compute the mape of test set to be 0.01169, which means that for each cell of grid, the difference between prediction result and ground truth is about 1.169%. In addition, we also visualize the prediction result and ground truth shown as figure 17.



(a) Ground truth (b) Prediction result (c) Difference

Figure 17. Visualization of prediction result (a) Ground truth matrix (b) Predicted matrix (c) the difference between ground truth and prediction

The figures on left side are the ground truth matrixes of the spatial distribution of order distribution; the figure in the middle is the corresponding prediction result; the figure on the right is the heatmap of difference between ground truth and prediction result. We can see from the figure that there is little difference between the ground truth and prediction result. The prediction is relatively accurate and enough to put into simulation.

6.2.2 Result of dispatch

Here, we will separately introduce the performance of greedy algorithm, proposed optimization algorithm and proposed optimization algorithm with additional constraint. The performance of three algorithms are shown in table 2.

Table 2. Comparison of different algorithms by metric

Metric \ Algorithm	Baseline: Greedy algorithm	Proposed method without optional constraint	Proposed method with optional constraint
Average waiting time of passenger	654.36 s	219.23 s	217.02 s
Average void cruising distance proportion	12.31%	3.83%	3.70%
Proportion of cancelled orders	48.35%	0.00149%	0.00%

From table, we can see that the greedy algorithm shows a poorer performance. The average waiting time of passenger is over 10 minutes. At the same time, almost half of the orders will be cancelled in the simulation. This is mainly because in the baseline algorithm, we don’t provide priority to the orders that have been waiting long enough. In real world application, for commercial purpose, ride-hailing dispatch platform may provide priority to the orders that have been waiting for a long time. Besides, when operating the dispatch, the Didi dispatch platform will provide each order to several candidate drivers to raise the chances that the order will be taken. Thus, the cancel rate will be much lower in real application. In simulation, the proportion of cancelled orders of optimization method is much lower than the one of optimization method. Besides, the average of waiting time of passenger in optimization method is only 1/3 times of greedy algorithm. This shows that the optimization algorithm can more easily and quickly answer the travel demand from customers. What’s more, from the perspective of void cruising distance proportion, the optimization method shows a better performance, which is lesser than 1/3 times of greedy algorithm. The overall result shows that the proposed algorithm surpasses the traditional greedy algorithm. The statistic of distribution of metrics is shown as figure 18.

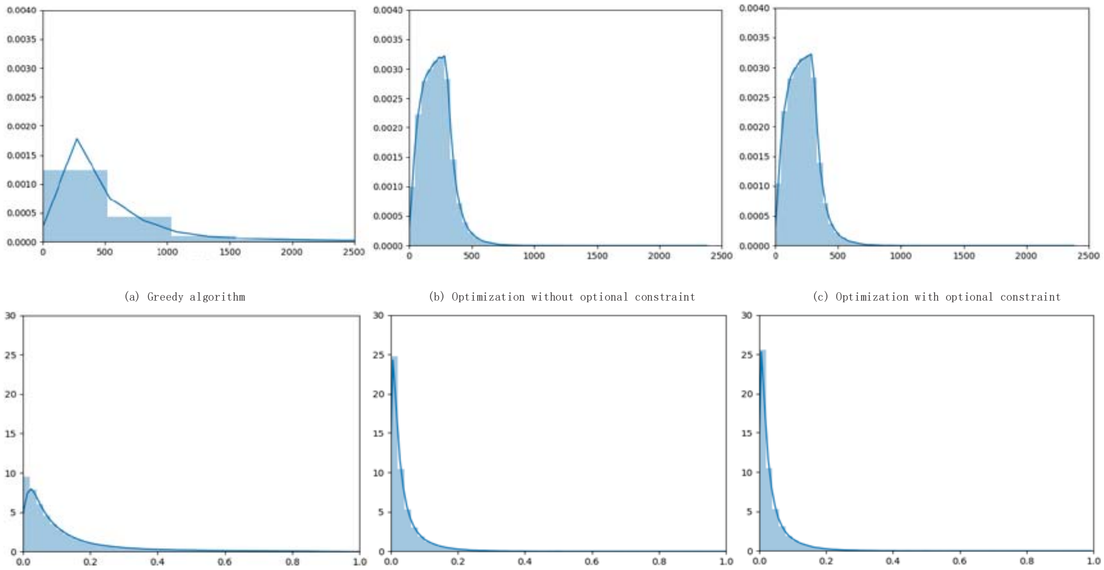


Figure 18. The statistic result of metrics

From left to right, the figures show the result of separately greedy algorithm, optimization method without optional constraint and with optional constraint. The upper figures show the waiting time of passengers and lower ones show the void cruising distance proportion. The x axis is the value of metric and y axis can be treated as the "probability". The larger the y value is, the higher probability is. The integrate of result of greedy algorithm is small is mainly because the high cancellation proportion. We can have a clear vision that the proposed algorithm can effectively suppress the waiting time and void distance proportion of most cases into a low interval. The waiting time in all of the cases is under 1000 seconds, which is about 16 minutes; the highest void cruising distance proportion is about 25%. while on the other hand, there are some extreme cases in the greedy algorithm. Some have been waiting for over 2500 seconds. In a small part of cases, the picking up distance is near half of delivery distance. This furtherly proves the stability of performance of proposed algorithm.

If we do a transverse comparison between the optimization method with and without the optional constraint, we can find that the optimization algorithm with optional constraint performs better than without constraint. Smaller proportion of cancelled orders and shorter waiting time is natural. We also notice that it accidentally brings lesser void cruising distance proportion. To better understand the mechanism behind this, we conduct an experiment under the perfect prediction, which means that there is no error in the prediction result. The result of metrics is separately, average waiting time of passenger: 219.31 s; average void cruising distance proportion: 3.83%; proportion of cancelled orders: 0.00%. We can see from the comparison that the only difference is that perfect prediction help eliminate the cancelled orders in the simulation. This can furtherly prove the high accuracy of prediction. What may cause the difference between the result of optimization algorithm with and without optional constraint is probably the uncertainty of distance in the prediction. In the process of solving the optimization problem, we compute the distance between each order and driver in both the current and predicted time window and construct a distance matrix. Because we lack the information of exact spatial distribution of orders in each cell. This causes many uncertainties in the simulation and affects the performance of optimization. Since currently, most of prediction of travel demand is mainly based on the area gridding method, we recommend more to pay main effort on optimizing the current dispatch problem. The prediction result aims to provide guidance on which group of drivers are better choices of dispatching at present.

7. Conclusion and future work

Under the background of energy-saving and emission reduction, many studies focus on the traffic emissions[54]. It is pointed out that void cruising for the next passenger caused a lot of unnecessary exhaust emission[55].

The main idea of this study is divided into two parts. First is to propose a simulation method to evaluate the impact the scale of users in the ride-hailing system on the emission performance and provide guideline of controlling number of users for ride-hailing service providers. To better adopt the method, massive Didi data from Chengdu in one day were used as the dataset. The method included the data process—extraction of OD of orders and drivers, Gibbs sampling, cross simulation and result computation.

From the result obtained by the simulation, we found that under the circumstance of a certain scale of travel demands in the scenario, the void distance proportion was decreasing with the increasing of scale of drivers and when the scale of travel demands increased. But, the trend of rising of the void cruising distance proportion varies by different scales of travel demands. The greater the travel demand is, the more rapidly the void distance proportion increases. It indicated that the greater the travel demand in an area is, the more efficient the supplement of driver is to reduce the void distance proportion. Moreover, we find two key features of judging the suitable number of drivers in the system, where before is a rapid interval and after is a gentle one. Since there is a strong relationship between the void distance proportion and invalid emission proportion, we also plotted the relationship between the invalid emission proportion as well as quantity and the scale of drivers under the certain quantity of travel demand. Then, we found an extremely similar pattern. At this point, we can have a clearer vision on the impact of the emission performance and user scale. When Ride-hailing service providers try to introduce ride-hailing system into a new area, they can use this relationship to estimate the suitable number of registered drivers in the system as guideline to attract or control the number of drivers or passengers. Finally, we also proposed a sample for the high efficiency area computation of scale of users in the ride-hailing system under certain metric constraints. Two metric sets are determined to find the different area of scale of users under different metrics.

The second part is to propose a dispatch method based on both prediction and optimization method to improve the efficiency of ride-hailing system. We use the same dataset for simulation. We firstly preprocess the data into desired input format of historical observation needed by the prediction model and collect the metadata for additional input. Next, we adopt the ST-Resnet as the deep learning neural network for prediction and successfully train the prediction model. The rescaled MAPE is rather little and enough for simulation of dispatch algorithm. Then, we introduce the dispatch algorithm based on optimization method. We state the target function to optimize and the constraint including the optional one which can impose a full satisfaction of orders.

We use the greedy algorithm which is used widely in current real-time dispatch system as the baseline and compare the performances. We find that the proposed method outperformances the baseline and shows a good stability of performance on the evaluation metrics, which proves a great potential in real-time application. In addition, we also find that the algorithm shows better performance with the optional constraint. This is mainly because the uncertainties in the location of predicted orders. Thus, we suggest imposing the constraint that maximize the number of served orders in solving the current dispatch problem. By far we have successfully answered and filled research gaps mentioned in literature review.

There are certainly some limitations in this work. For example, the simulation method in evaluating the impact of user scale is very dependent on regional historical data. Now if we want to transfer this model to other cities, we may need other local historical ride-hailing data. In the future work, we can improve this model by some critical city information, such as, the landmark position to make this model more common. In the part of dispatch, we mainly use the statistical result to estimate the location of predicted orders in the future. Adopting smaller cell can help get more accurate location. However, it will inevitably make the spatial distribution matrix sparser, thus, making the model hard to be trained. In the future, if better prediction method like GCN(graphical convolutional network) can be developed more well, we can improve this limitation by adopting them.

Reference

1. Zhang H, Song X, Xia T, et al. Battery electric vehicles in Japan: Human mobile behavior based adoption potential analysis and policy target response. *Applied Energy*. 2018;220:527-535.

2. Zhang H, Song X, Long Y, et al. Mobile phone GPS data in urban bicycle-

sharing: Layout optimization and emissions reduction analysis. *Applied Energy*. 2019;242:138-147.

3. Clewlow RR, Mishra GS. Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States. *University of California, Davis, Institute of Transportation Studies, Davis, CA, Research Report UCD-ITS-RR-17-07*. 2017;

4. Clewlow RR, Mishra GSJUoC, Davis, Institute of Transportation Studies, Davis, CA, Research Report UCD-ITS-RR-17-07. Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States. 2017;

5. Rayle L, Dai D, Chan N, Cervero R, Shaheen S. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*. 2016;45:168-178.

6. Li B, Zhang D, Sun L, et al. Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset. IEEE; 2011:63-68.

7. Guo Y, Xin F, Barnes SJ, Li X. Opportunities or threats: The rise of online collaborative consumption (OCC) and its impact on new car sales. *Electronic Commerce Research and Applications*. 2018;29:133-141.

8. Cramer J, Krueger AB. Disruptive change in the taxi business: The case of Uber. *American Economic Review*. 2016;106(5):177-82.

9. Wang Z, Chen F, Fujiyama T. Carbon emission from urban passenger transportation in Beijing. *Transportation Research Part D: Transport and Environment*. 2015;41:217-227.

10. Clewlow RR, Mishra GS. Disruptive transportation: The adoption, utilization, and impacts of ride-hailing in the United States. 2017;

11. Korolko N, Woodard D, Yan C, Zhu H. Dynamic pricing and matching in ride-hailing platforms. *Available at SSRN*. 2018;

12. Afeche P, Liu Z, Maglaras C. Ride-hailing networks with strategic drivers: The impact of platform control capabilities on performance. *Columbia Business School Research Paper*. 2018;(18-19):18-19.

13. Calegari GR, Celino I, Peroni D. City data dating: Emerging affinities between diverse urban datasets. *Information Systems*. 2016;57:223-240.

14. Henao A, Marshall WE. The impact of ride-hailing on vehicle miles traveled. *Transportation*. 2019;46(6):2173-2194.

15. Erhardt GD, Roy S, Cooper D, Sana B, Chen M, Castiglione J. Do transportation network companies decrease or increase congestion? *Science advances*. 2019;5(5):eaau2670.

16. Flores O, Rayle L. How cities use regulation for innovation: the case of Uber, Lyft and Sidecar in San Francisco. *Transportation research procedia*. 2017;25:3756-3768.

17. Beer R, Brakewood C, Rahman S, Viscardi J. Qualitative analysis of ride-hailing regulations in major American cities. *Transportation Research Record*. 2017;2650(1):84-91.

18. Bösch PM, Becker F, Becker H, Axhausen KW. Cost-based analysis of autonomous mobility services. *Transport Policy*. 2018;64:76-91.

19. Wadud Z, MacKenzie D, Leiby P. Help or hindrance? The travel, energy and carbon impacts of highly automated vehicles. *Transportation Research Part A: Policy and Practice*. 2016;86:1-18.

20. Tu W, Santi P, Zhao T, et al. Acceptability, energy consumption, and costs of electric vehicle for ride-hailing drivers in Beijing. *Applied Energy*. 2019;250:147-160.

21. Schaller B. The new automobility: Lyft, Uber and the future of American cities. 2018;

22. Tirachini A, Gomez-Lobo A. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. *International Journal of Sustainable Transportation*. 2020;14(3):187-204.

23. Rodier C. The effects of ride hailing services on travel and associated greenhouse gas emissions. 2018;

24. Xu Z, Li Z, Guan Q, et al. Large-scale order dispatch in on-demand ride-hailing platforms: A learning and planning approach. 2018:905-913.

25. Feng G, Kong G, Wang Z. We are on the way: Analysis of on-demand ride-hailing systems. *Available at SSRN 2960991*. 2017;

26. Yao H, Wu F, Ke J, et al. Deep multi-view spatial-temporal network for taxi demand prediction. 2018:

27. Maciejewski M, Bischoff J. Large-scale microscopic simulation of taxi services. *Procedia Computer Science*. 2015;52:358-364.

28. Grau JMS, Romeu MAE, Mitsakis E, Stamos I. Agent based modeling for

simulation of taxi services. *Journal of Traffic and Logistics Engineering*. 2013;1(2):159-163.

29. Mourad A, Puchinger J, Chu C. A survey of models and algorithms for optimizing shared mobility. *Transportation Research Part B: Methodological*. 2019;

30. O'Neil RJ, Hoffman K. Decision diagrams for solving traveling salesman problems with pickup and delivery in real time. *Operations Research Letters*. 2019;47(3):197-201.

31. Cheng X, Fu S, Yin G. Does subsidy work? An investigation of post-adoption switching on car-hailing apps. *Journal of Electronic Commerce Research, Forthcoming*. 2016;

32. Institute DMR. Report on big data of intellegent travel in Chengdu. 2016;

33. Marcotte D, Allard D. Gibbs sampling on large lattice with GMRF. *Computers & Geosciences*. 2018;111:190-199.

34. Vazifeh MM, Santi P, Resta G, Strogatz S, Ratti C. Addressing the minimum fleet problem in on-demand urban mobility. *Nature*. 2018;557(7706):534.

35. Arslan AM, Agatz N, Kroon L, Zuidwijk R. Crowdsourced delivery—A dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science*. 2018;53(1):222-235.

36. Transportation ICfEa. Online Ride-Hailing Network and Fuel Consumption: A Driver's Perspective. 2017;

37. Zheng L, Chen L, Ye J. Order dispatch in price-aware ridesharing. *Proceedings*

of the VLDB Endowment. 2018;11(8):853-865.

38. Sui Y, Zhang H, Song X, et al. GPS data in urban online ride-hailing: A comparative analysis on fuel consumption and emissions. *Journal of Cleaner Production*. 2019;227:495-505.

39. Cai H, Xie S. Estimation of vehicular emission inventories in China from 1980 to 2005. *Atmospheric Environment*. 2007;41(39):8963-8979.

40. Lang J, Cheng S, Zhou Y, Zhang Y, Wang G. Air pollutant emissions from on-road vehicles in China, 1999–2011. *Science of the Total Environment*. 2014;496:1-10.

41. Yan C, Zhu H, Korolko N, Woodard D. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)*. 2019;

42. Khazbak Y, Fan J, Zhu S, Cao G. Preserving location privacy in ride-hailing service. IEEE; 2018:1-9.

43. Zhang R, Pavone M. Control of robotic mobility-on-demand systems: a queueing-theoretical perspective. *The International Journal of Robotics Research*. 2016;35(1-3):186-203.

44. Wang D, Yang Y, Ning S. Deepstcl: A deep spatio-temporal convlstm for travel demand prediction. IEEE; 2018:1-8.

45. Chu KF, Lam AY, Li VO. Travel demand prediction using deep multi-scale convolutional LSTM network. IEEE; 2018:1402-1407.

46. Jia C, Wang S, Zhang X, Wang S, Ma S. Spatial-temporal residue network based in-loop filter for video coding. IEEE; 2017:1-4.

47. Ma X, Dai Z, He Z, Ma J, Wang Y, Wang Y. Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*. 2017;17(4):818.

48. Cassel M, Lima F. Evaluating one-hot encoding finite state machines for SEU reliability in SRAM-based FPGAs. IEEE; 2006:6 pp.

49. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:150203167*. 2015;

50. Kingma DP, Ba J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:14126980*. 2014;

51. Sühr T, Biega AJ, Zehlike M, Gummadi KP, Chakraborty A. Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. 2019:3082-3092.

52. Klee V, Minty GJ. How good is the simplex algorithm. *Inequalities*. 1972;3(3):159-175.

53. Frei C, Hyland M, Mahmassani HS. Flexing service schedules: Assessing the potential for demand-adaptive hybrid transit via a stated preference approach. *Transportation Research Part C: Emerging Technologies*. 2017;76:71-89.

54. Huang Z, Cao F, Jin C, Yu Z, Huang R. Carbon emission flow from self-driving tours and its spatial relationship with scenic spots–A traffic-related big data method. *Journal of cleaner production*. 2017;142:946-955.

55. Luo X, Dong L, Dou Y, et al. Analysis on spatial-temporal features of taxis' emissions from big data informed travel patterns: a case of Shanghai, China.

Journal of cleaner production. 2017;142:926-935.

Nomenclature

$\boldsymbol{O}$	Set of all order data
$\boldsymbol{D}$	Set of all driver data
$\boldsymbol{i} \in \boldsymbol{O_s}$	Set of order data for simulation
$\boldsymbol{j} \in \boldsymbol{Ds}$	Set of driver data for simulation
$\boldsymbol{P_o}$	Percentage of all order data for simulation
$\boldsymbol{P_d}$	Percentage of all driver data for simulation
$\boldsymbol{O_b}$	Set of backup order data
$\boldsymbol{D_b}$	Set of backup driver data
$\boldsymbol{A_i}$	Available driver set for order $\boldsymbol{i}$
$\boldsymbol{md_i}$	Driver with the shortest distance to order $\boldsymbol{i}$ in $\boldsymbol{A_i}$
$\boldsymbol{e}$	Mean variance of result
$\boldsymbol{e_{max}}$	Convergence condition
$\boldsymbol{\theta_{max}}$	Max waiting time of a passenger to wait for a driver to take the order
$\boldsymbol{t_o}$	Time when a passenger gives an order
$\boldsymbol{t_a}$	Time when a driver is available
$\boldsymbol{t_p}$	Pick up time