Research on Optimization Configuration of Taxi Resources Based on Matching Model

Nanxuan Wang¹, Wentao Zhang², Shuqing Weng³

¹School of Insurance and Economics, University of International Business and Economics, Beijing, 100029.
²College of Electrical Engineering and Control Science, Nanjing Tech University, Nanjing, Jiangsu, 211800
³The second clinical medical college, Nanchang University, Nanchang, Jiangxi, 330036

Abstract. “Difficulty in taxiing” has always been a concern of the society in recent years. This paper discusses the current situation of resource allocation of taxis and the impact of taxi company subsidy programs on the difficulty of taxiing, and then introduces a better subsidy program. We first collected data based on the Drip, fast intelligent travel platform, plotted the image, calculated the average of the taxi distribution and passenger demand at each time point, and carried out tracing and analysis. On this basis, the matching model is established: the idea of nearest neighbor” is introduced, and the time points that are difficult to get a taxi are concentrated at 8-10 am, 16 pm, and 19 pm. Then analyze the subsidy program data, taxi drivers and passengers after using the taxi software, taxi vacancy rate, passenger waiting time and changes in the economic benefits of taxi drivers and passengers, etc., after analysis, we found the existence of subsidy programs can fully mobilize the enthusiasm of drivers and passengers to use the software, so that the number of people using the software continues to increase, which not only shortens the waiting time for passengers to take a taxi, but also reduces the driver's vacancy rate, so it can effectively alleviate the problem of difficult driving.

1. Introduction
Taxi is one of the important means of transportation for people in the city. However, with the increase of urban population density and vehicle density, it is very inconvenient for traffic operation, so even if the number of taxis in a city is greater than the demand volume, there will also be problems with taxis [1]. In order to alleviate this problem, coupled with the arrival of the "Internet +" era, a number of companies such as Didi, KuaiDi, etc. rely on the mobile Internet to establish a taxi service platform, passengers can make a reservation in advance, the driver can be based on the appointment location and the time location, information exchange, in order to win the use of users, companies have introduced different subsidy programs [2]. Therefore, it is very important to study the difficulty of taxis in different time periods and different places to optimize the allocation of taxi resources [3].
2. Establishment and Solution of Matching Models

2.1. Data Collection and Processing

We collected the taxi distribution data and the passengers' demand data for every hour from September 1st, 2015, from 1st to 24th, and after finishing, we drew two-dimensional scatter plots and scatter plots. The dot matrix size reveals the number of taxis or the amount of passenger demand. As can be seen from the figure, there is a certain taxi distribution and passenger demand at any time, and even there is a large passenger demand at midnight. Throughout the whole day, it can be found that the peak period of passenger demand is 8-9 in the morning and 19 in the evening. At this time, the phenomenon of "difficulty in taxiing" is more likely to occur.

In order to better check the matching of taxis and passengers in each time period, we calculated the average value of taxi distribution and passenger demand at each time point. The average value is shown in Figure 1. As can be seen from Figure 1, in general, the average number of taxis at various points in Chengdu is greater than the average demand of passengers, which is about 2 times different. Therefore, on average, there is no problem of "difficult to taxi a car". As can be seen from Figure 1, taxis are the least distributed at 5 am, with an average of about 20 vehicles, which is the lowest point in a day. The highest point of taxi distribution appeared around 23 o'clock in the middle of the night, and there were about 50 taxis, indicating that the nightlife of Chengdu citizens is quite rich. From 7:00 am to 18:00 pm, the taxis are basically stable at around 40 vehicles. Compared with taxis, the demand for passengers presents a different pattern. 8-9 in the morning is the time when passengers have the most demand. On average, there are nearly 15 demand; in addition, there is a small peak of passenger demand at 18-19 in the afternoon, with an average of about 10; during the rest of the day, during the daytime from 11 o'clock to 17 o'clock in the afternoon, and 20-22 o'clock in the evening, it is basically stable around 8 demand; from 23 o'clock in the middle of the night to 6 o'clock in the next day, the passenger demand is the lowest, about 5 or so.

![Figure 1](image)

Figure 1. All-day taxi distribution and passenger demand mean map of Chengdu, September 9, 2015

2.2. Model Establishment

From the previous analysis, we can see that, on average, there is no problem of "difficult to taxi a car". However, due to the uneven distribution of demand for taxis and passengers, some areas are very concentrated, while the distribution points of individual passengers are very isolated. Obviously, this will also lead to the phenomenon that the car cannot be hit. To this end, we mainly focus on passengers, establish a matching model, and measure the degree of "difficulty in taxiing" to analyze
the taxi problem in Chengdu. It is not difficult to see that at each time point, the specific location of the data we collected is different. This aspect reflects the dynamic and random nature of the data. On the other hand, it also makes it difficult for us to match taxis and passengers. How to establish a reasonable model and indicators to measure the degree of "difficulty in taxiing"? Here, we introduce the idea of "nearest neighbors" to match. The so-called "nearest neighbor" idea is based on each time point, each passenger demand point, looking for the nearest taxi to match. However, from the perspective of market economy, we believe that the more reasonable matching model should be based on the map distance (of course, it is better to be based on the reachable map distance, but since the data is difficult to obtain, the map spherical distance is temporarily considered here). Match the taxi circle at a certain distance from the passenger's point, such as 500 meters, 1 kilometer, or 2 kilometers.

Record the passenger data structure as \{\{P_i(t), P_s(t), P_t(t)\}, i=1, 2, ..., n.\} here, \(P_s)\) indicates the longitude of the i-th passenger data point, \(P_r\) indicates the latitude of the i-th passenger data point, \(P_i\) represents the demand for the i-th passenger data point, i represents the i-th passenger data point, and n represents the number of passenger data points collected. Here we consider different points of time, because at different time points, we will get different sets of passenger data points, and the number of different passenger data points n. For the convenience of expression, we can add time factor in the symbol to get passenger data point collection for.

\{\{P_i(t), P_s(t), P_t(t)\}, i=1, 2, ..., n(t).\}, where t=1, 2, ..., 24.

among them, \(P_s(t)\) indicates the longitude of the i-th passenger data point at the t-th time point, \(P_r(t)\) represents the latitude of the i-th passenger data point at the t-th time point, and \(P_i(t)\) represents the demand amount of the i-th passenger data point at the t-th time point.

Correspondingly, we can get the data point set of the taxi:

\{\{T_j(t), T_r(t), T_t(t)\}, j=1, 2, ..., m(t).\}

Where \(j\) is the jth taxi data point and \(m\) is the number of taxi data points. \(T_r(t)\) indicates the longitude of the j-th taxi data point at the t-th time, \(T_r(t)\) indicates the latitude of the jth taxi data point at the t-th time point, and \(T_t(t)\) indicates the number of taxi distributions of the jth taxi data point at the t-th time point.

Let \(d(A, B)\) denote the spherical distance between two points A and B, C be the specified neighborhood radius, and define the i-th passenger data point at the tth point \(A_i\). The match is

\[
N_{it} = \begin{cases} 
1, & \sum_{j \in d(A_i(t), B_j(t))} T_j(t) < P_i(t) \\
0, & \sum_{j \in d(A_i(t), B_j(t))} T_j(t) = P_i(t) \\
-1, & \sum_{j \in d(A_i(t), B_j(t))} T_j(t) > P_i(t) \end{cases}
\]

Where \(t=1, 2, ..., 24, B_j(t)\) indicates the jth taxi data point at the tth time. It is not difficult to see that if the sum of taxis that do not exceed C from the passenger point is greater than the passenger demand, there will be no difficulty in taxiing, and the definition matching degree is equal to -1. If the sum of the taxis that do not exceed C from the passenger's point is less than the passenger's demand, it indicates that there is a problem of difficulty in taxiing, and the matching degree at this time is defined as 1. If they are exactly equal, the matching degree is defined to be equal to zero.

Let the first point A pass, the latitude is (LonA, LatA), and the second point B has the latitude and longitude (LonB, LatB). According to the 0 degree warp line, the east longitude is positive (Longitude), the west longitude take the negative longitude value (-Longitude), the north latitude takes 90-latitude value (90-Latitude), the south latitude takes 90+ latitude value (90+Latitude), then the two
points after the above processing are counted as (MLonA, MLatA) and (MLonB, MLatB). Then
according to the triangle derivation, the following formula for calculating the distance between two
points can be obtained:
\[ C_0 = \sin(MLatA) \times \sin(MLatB) \times \cos(MLonA-MLonB) + \cos(MLatA) \times \cos(MLatB) \]
\[ \text{Distance} = R_1 \times \arccos(C_0) \times \pi/180 \]
Here, \( R_1 \) same as the Distance unit, if the radius is 6371.004 kilometers, then the unit of Distance is
kilometers.

2.3. Model Solving
The \( \text{d}(A, B) \) described above can be replaced by the distance formula Distance \( (A, B) \) here. Thus, the
neighborhood radius \( C \) can take a series of values. This article takes 59 scales of 100 meters, 200
meters, ..., 1000 meters, 2 kilometers, ..., 50 kilometers, and is labeled 1, 2, 3, ..., 59.

| Serial number | 1 | 2 | 3 | ... | 9 | 10 | 11 | ... | 58 | 59 |
|---------------|---|---|---|-----|---|----|----|-----|----|----|
| Radius c (unit: km) | 0.1 | 0.2 | 0.3 | ... | 0.9 | 1 | 2 | ... | 49 | 50 |

It can be expected that the smaller the radius of the domain, the greater the probability that the
matching degree will be taken as 1. As the radius of the neighborhood gradually increases, the
matching degree will change from 1 to -1, that is, the vehicle cannot be hit in a small range. However,
if a large range of taxis participate in the competition for passenger orders, the probability that
passengers will get to the car will increase significantly.

Thus, we define the following metrics for taxi difficulty:
\[ m_{N_i}(t) = \min_{k}^* N_{i}^{ik} (t) = -1, k = 1, 2, 3, ..., 59+ \]

among them, \( N_{i}^{ik} (t) \) indicates the radius of the neighborhood, \( C_i \) indicates the i-th passenger data
point at the tth time. The metric \( m_{N_i}(t) \) is the serial number corresponding to the minimum radius
required for the taxi to be difficult to reach. The larger the serial number indicates that a taxi with a
larger radius is required to participate in the transportation of passenger data points to satisfy the
passengers demand for taxis.

We also plot the difficulty scatter plot for passenger data points for all time periods, as shown in
Figure 2. As can be seen from Figure 2, taxis generally within 2 km can meet the passengers' demand
for taxis. Most of them can be satisfied within 1 km. Only a few data points are needed, and it takes
about 10 km or more. For example, a taxi of more than 30 kilometers can be satisfied. These data
points are the hardest data points for taxis.
Figure 2. Passengers’ difficulty in taxiing in all time periods

We plotted the bar chart of the difficulty of passengers taking taxis at different times. As can be seen from the figure, taxis can meet the needs of most passengers at any time within a distance of no more than 1 km. During the morning rush hour, it appears at 8:00-9:00 in the morning, and at 19:00 in the evening, there will be a night peak period; the daytime is basically flat, and the frequency of 1:5-5 in the middle of the night is the lowest.

Figure 3. Bar chart of passengers' difficulty in different time periods

3. Subsidy Policy and Difficulty in Taxiing
Through the collected information and data, this section mainly discusses the relationship between the subsidy policy of taxi software and the difficulty of taxiing, and finds out the new problems brought about by the changes in the existing taxi software subsidy policy. The detailed analysis is as follows:

3.1. Comparative Analysis of Subsidies For Passengers By Dripping Taxis and Fast Taxis
According to the collected data, the subsidies for drip and quick taxis to passengers are shown in Table 2 below.
Table 2. Comparison of discounts for passengers on ditute taxis and fast taxis in January-February 2014

| time          | Didi taxi subsidy program                  | Fast taxi subsidy program                  |
|---------------|--------------------------------------------|--------------------------------------------|
| January 10    | 10 yuan per order                          | Less 0 yuan                                |
| January 20    | Continue to reduce 10 yuan per order       | Continue to reduce 10 yuan per order       |
| February 11   | 5 yuan per order                           | 10 yuan per order                          |
| February 17   | 10 yuan per order, the new driver's first 50 yuan | Less than 11 yuan per order               |
| February 18   | 12-12 yuan per order                       | 13 yuan per order                          |
| March 3       | 12-12 yuan per order                       | 10 yuan per order                          |
| March 18      | 5 yuan per order                           | 5 yuan per order                           |
| March 22      | 5 yuan per order                           | 3-5 yuan per order                         |
| March 23      | 3-5 yuan per order                         | Continue to reduce 3-5 yuan per order      |

Looking at the data of the subsidy program, we can conclude that the subsidy for drivers of Didi taxis is 10 yuan on January 10 and 0 yuan on fast. Then, the drip-driving car showed a trend of decreasing first and then increasing. After March 3, the subsidy began to drop to 0 yuan. This shows that the company began to increase the subsidy in order to win the user's use, and when the user group reached a certain share, the passenger's rewards are gradually canceled. At this time, most passengers are accustomed to using taxi software to take taxis. Therefore, the subsidy will be reduced, and the passengers' use of the taxi software will be reduced, but it will not drop significantly. Especially during the peak period, passengers are more willing to reduce waiting time. Use taxi software to reduce the difficulty of taxis; the subsidy for fast taxis has been increasing since the beginning, in order to compete with Didi taxis, to win more users, and to start to reduce subsidies after the user's changes and drops a taxi is similar.

From the data of the subsidy program, the subsidy for drivers of Didi taxis began to stabilize at 10 yuan/single, and then began to gradually decrease. The subsidy for fast taxis is generally more than that of drip taxis. The purpose is to seize market share. At the time of scale, the subsidy is reduced to zero. These include:

When the subsidy is high, the driver is more willing to accept the order. This can reduce the waiting time of the passengers. However, for passengers who do not use the software to place an order, it is difficult for them to get to the car, which makes it more difficult for the passengers to take a taxi. For the group of elderly and children, there will be a situation where the car cannot be found at all.

After the subsidy fell to zero, according to the survey, the enthusiasm of the driver to grab the order fell, especially during the peak of the taxi, the guests may be pulled at any time and place, so in general, the difficulty of taking a taxi is similar to the situation without the taxi software.

3.2. Discussion on The Influence of Passenger Waiting Time and Taxi Vacancy Rate on “Difficulty of Taxiing”

For DDT taxis, since the subsidy began on January 10, 2014, the changes in the number of registered users and the average daily order volume from January to March 2014 are shown in Table 3.

Table 3. Drip taxi users, daily orders account for user data

| date      | User number | Average daily order | Average daily order/number of users |
|-----------|-------------|--------------------|------------------------------------|
| January 10 | 22 million  | 350,000            | 1.6%                               |
| February 9 | 40 million  | 1.83 million       | 4.6%                               |
| February 24 | 82.6 million | 3.16 million      | 3.8%                               |
| March 27   | 100000000  | 521.83 million     | 5.2%                               |

From Table 3, it can be concluded that from January to March 2014, as the subsidy progressed, the
number of orders continued to increase, and the percentage of "orders/users" stabilized at 4%-5%. Through the estimation of market share, as of December 2014, the cumulative account size of China's taxi app reached 172 million, of which the fast taxi market share was 56.5% and the Drip taxi was 43.3%, indicating that the subsidy policy was extremely high. The big one attracts the registration and ordering of taxi users; the change of fast taxis is similar to that of dripping taxis.

According to the survey, after using the taxi software, 94.0% of the passengers waited for 10 minutes, and the proportion of passengers who waited for more than 10 minutes decreased by 29.9%. After using the taxi software, 90.3% of the drivers thought that the idling rate was reduced. 55.0% of the drivers thought that the monthly vacancy rate fell below 10%, and 41.2% of the drivers thought that the monthly vacancy rate dropped by 10-30%, 3.9% of the drivers. It is considered that the monthly vacancy rate has dropped by more than 30%. In summary, the driver's vacancy rate dropped, indicating that the passenger load rate increased and the passenger's waiting time decreased, indicating that taxiing became easier. Therefore, through the previous analysis, there is a subsidy policy, and the subsidy is higher, drivers and passengers are more willing to go. Using the taxi software, the driver's idling rate is reduced and the passenger waiting time is also reduced. Both drops make it easier to take a taxi.

3.3. Discussing The Impact of The Subsidy Program on The Degree of “Difficulty In Taxiing” From The Economic Benefits of Passengers and Drivers.

According to the survey, since January 2014, fast taxis and drip taxis have been competing with each other in subsidies, and the total amount of subsidies between the two parties is more than 2.5 billion. Except for subsidized drivers, most of the subsidies are obtained by passengers. According to the survey, 81.0% of passengers believe that they save on taxi fares.

Since the taxi software provides a subsidy for each driver, this provides the driver with a source of fare increase income. According to the survey, after using the taxi software, 92.0% of drivers believe that the average income per order has increased. Among them, 48.0% of drivers believe that the average income per order increases by 10-30%, 51.8% of drivers think that the average income per order increases by less than 10%, and 0.2% of drivers think that the average income per unit increases by more than 30%. Therefore, both in terms of passengers and drivers, economic gains have been made. Therefore, in this economic subsidy environment, both parties are more willing to use the taxi software. According to the above conclusion, the software usage rate is higher. Through the above analysis, we classify the urban population n: the elderly population N1 and the general population N2.

(1) For the elderly group N1 (about 20%), because this part of the population uses less smartphones, that is, even if the taxi software provides subsidies, it is still difficult to solve their "difficult to taxi" problem.

(2) For the general population N2, can be drawn with the presence or improvement of subsidies for passengers and taxi drivers, fully mobilize their enthusiasm for using software, so that the number of people using software will continue to increase, thus shortening the waiting time for passengers to take a taxi. It also reduces the driver's vacancy rate, so it can effectively alleviate the difficulty of taxiing. However, as the subsidy decreases, the passengers and drivers' enthusiasm for the use of taxi software begins to decrease, especially during peak hours, drivers do not need to use software to receive orders. The vacancy rate is also very low; while the passengers basically use the taxi software, it is still difficult to take a taxi at the peak, which indicates that the change and adjustment of the subsidy program needs further discussion.

Due to the use of taxi software to book a taxi, especially during peak hours, drivers can enjoy a taxi subsidy. Therefore, people who do not have an appointment to take a taxi can hardly get a car during peak hours, so many cities such as Shanghai and Jinan are strictly prohibited from using taxi software during peak hours. This also indicates that the subsidy policy for taxi software companies needs further discussion.
3.4. Discussion on Subsidy Program

3.4.1. Scheme 1 simulation evaluation
Option 1: Give a certain cash or point reward for traveling in the low peak period and successfully hitting the car.

We have already pointed out that the morning rush hour occurs at 8-9 pm and the late rush hour is 19 pm. To illustrate whether the subsidy program 1 is reasonable, we assume that passengers who are expected to travel at 8 o'clock are encouraged in subsidy program 1. Next, travel to 7 am in advance to see if it is possible to significantly alleviate the difficulty of taxiing. After calculation, we found that if we follow the s-curve model and finally stabilize the impact rate of around 5%, it will not affect the above difficulty level defined. That is, if only the passengers avoid the peak period, and the subsidy program has little impact, it is difficult to alleviate the difficulty of taxiing. The reason is to compare and analyze the taxi map. We found that no matter which time period, the distribution pattern of taxis is basically in a constant trend, which is roughly those areas. Therefore, in order to ease the difficulty of taxiing, we have to start with a taxi.

3.4.2. Scheme 2 Simulation Evaluation
Based on the evaluation of the scheme 1, we proposed the scheme 2 to let the taxi run. I hope that the scheme 2 can effectively alleviate the problem of taxiing.

Option 2: Encourage taxis to go to crowded places. The taxi software can tell the driver which area has a large demand for taxis through real-time analysis of big data. Through certain cash or point rewards, drivers are encouraged to grab orders from crowded places.

At 8 o'clock in the morning, the passenger demand in Chengdu are about 72.6% of the passenger data points of the passengers, and about 12.33% of the passenger data points. There are only a handful of data points that require more than 30 taxis.

It is not difficult to see that after the use of Option 2, the difficulty of taking the car at the 9th data point (that is, the data point with the largest demand for taxis in the front) is significantly reduced, from the original difficulty level of 11 to the current 1, and other data points. The change is almost zero. This shows that the subsidy scheme 2 for passengers' intensive points can better alleviate the difficulty of taxiing.

References
[1] Luo Congkun. Research on taxi taxi software problem [J]. Journal of Lanzhou Education College, 2015 (3): 58-62.
[2] Wang Guangqi. Suggestions on strengthening the comprehensive management of taxi software [J]. Urban Transport, 2014 (5): 98-102.
[3] Hu Jianjun. University clustering algorithm based on nearest neighbor priority [J]. Journal of Sichuan University (Engineering Science Edition), 2004 (6): 65-68.