# City Taxi "Supply to Match" Degree Analysis and Solution Algorithm

Fengyu Han[1, a *] and Qiang Fu[1, b]

[1]Changchun University of Science and Technology, China Jilin Changchun 130022

[a]tuhanfengyu@163.com, [b]fuqiangwork@163.com

**Keywords:** Python; Principal component analysis; MATLAB; No-load ratio.

**Abstract.** In this paper, we focus on the need to solve the problem through reasonable index, based on a taxi drops quickly intelligent software platform, with Beijing as an example, Analysis the degree of “The matching of supply and demand” under different time and space. This paper has established three kinds of methods, they are analysis method, no-load ratio method and get the data through Python program [1]. And we also used MATLAB R2014b, Python, SPSS software, etc. Based on taxi no-load ratio as the major factors in control measure "the matching of supply and demand". By using principal component analysis to select reasonable factors that influences the matching of supply and demand. Get the internal relationships between the variables by SPSS software, and finally we selected taxi's distribution of quantity and demand as reasonable index. Drawing three-dimensional image with time and space as independent variables, impact index as the dependent variable through MATLAB, we use it to express no-load ratio and then reflect the degree of “The matching of supply and demand ” under different time and space.

## To Get the Data

First, model data is based on drops quick smart travel platform, platform data update in real time, so the data cached in the local, then we using python language to get the intelligent platform of data for parsing and processing. The data for each index in the platform is presented in an hour. In this paper, the city of Beijing is divided according to the rectangular ring in ideal, with latitude and longitude as the limiting condition, get every hour and every lot of taxi demand, taxi distribution and taxi costs that three indicators. Because data is obtained from the intelligent platform in real time, so the value displayed by the program will be changed according to the update of the website data.


Figure 1.

The following Fig. 2 is the operation of the Python programming file. Data in the picture is the total number of per hour’s taxi distribution requirement and fare in the second ring of Beijing city. In this paper, we study the time period of two hours, and variables units in the back are all in two hours.

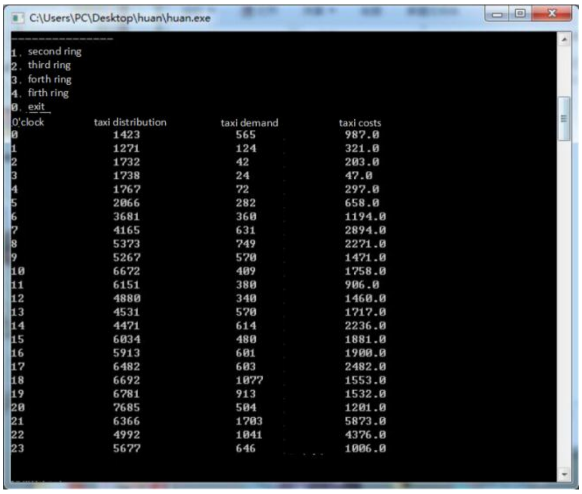


Figure 2.

## Principal Component Analysis of Influencing Factors

**Principal Component Analysis [2, 3].** Principal component analysis is a kind of data reduction technique, through research the internal dependence among numerous variables, then delete duplicate variables (variables that are closely related), integration into one or more of the relatively independent integrated factors, it is also a method used to reduce the dimension. The size of the information is often measured in terms of the square sum or variance, and the variance is measured in this paper.

Principal component analysis method mainly uses SPSS software, using Python program to get statistical data and then put it into SPSS statistic software. With taxi distribution, demand and total fare as the index, output multiple factors for analysis of data, then we can obtain the numerical value of the correlation matrix and the total variance explained.

Table 1 correlation matrix

| | | taxi distribution | taxi demand | taxi costs |
|---|---|---|---|---|
| related | taxi distribution | 1.000 | .781 | .647 |
| | taxi demand | .781 | 1.000 | .844 |
| | taxi costs | .647 | .844 | 1.000 |

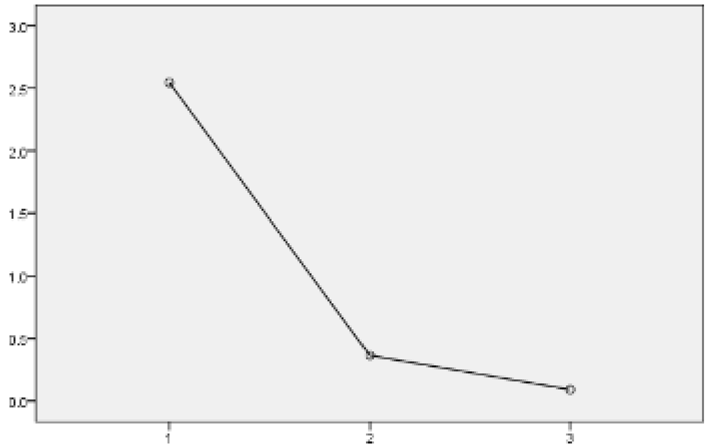
The correlation coefficient matrix table shows the correlation coefficient between each variable. In the correlation coefficient matrix table "relevant" column, the greater the value of the data, the more significant correlation. Among them, the demand for a taxi and the total fare are the strongest correlation.

Table 2 Total variance explained

| Element | Initial eigenvalue | | | Extract the square and load | | |
|---|---|---|---|---|---|---|
| | Sum | Square % | Accumulative % | Sum | Square % | Accumulative % |
| 1 | 2.545 | 84.834 | 84.834 | 2.545 | 84.834 | 84.834 |
| 2 | .365 | 12.156 | 96.990 | .365 | 12.156 | 96.990 |
| 3 | .090 | 3.010 | 100.000 | .090 | 3.010 | 100.000 |

The table of the total variance explained contain characteristic values variance percentage (contribution rate) and cumulative variance (cumulative contribution rate) which are obtained through principal component analysis. Because the sum of all the eigenvalues is equal to the number of variables in the input sample. So the greater the eigenvalues, the greater the contribution rate and the amount of information contained in the initial variable. We can see in the column of the sum of squares, more than 96.99% of the original variable information is concentrated in the first two principal components. Therefore, we can use the first two variables to instead of all to conduct a principal component analysis.

Graph 3. Scree plot




The number of principal components can be determined by a table combined with scree plot. The first two steep, next gentle, it is shown that we can use the first two variables to instead of all to a conduct principal component analysis.

**Establish a "Supply to Match" Model [4, 5]**

**Draw Graph of Relationships.** With the above data, using MATLAB to construct a graph that the independent variables were the time (every two hours) and Beijing loop area, the dependent variable respectively taxi requirement and taxi distribution. Put two graphs in the same axis and observe their relationships.

Fig. 4 is a figure that the independent variable is taxi requirement and taxi distribution in different space stability different time.

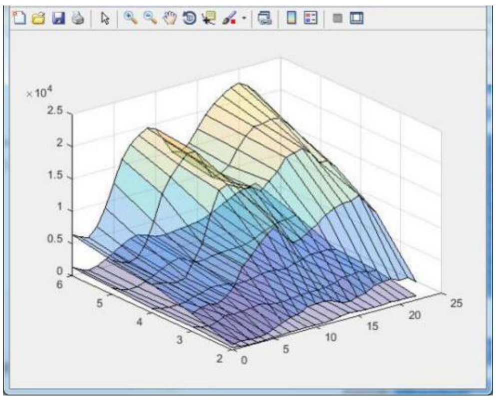


Figure 4.

According to observation that the taxi requirement is directly proportional to the taxi distribution. In order to determine which is the common influence factors by taxi requirement and taxi distribution, So we find the factor by derivation of the equation.

**Drawing the No-load Ratio Graph [6, 7].** Taxi bear permanent population's transportation turnover (W):

$$W = R_{sum} \cdot A_{sum} \cdot P \cdot D \quad (1)$$

In formula: R - total urban population. A - Residents travel times every two hours. P- The proportion of travel by taxi on the structure of inhabitant mode share. D - The average travel distance of the residents by taxi.

And P also is called taxi sharing ratio:

$$P = \frac{Q_{\text{out}}}{Q_{\text{sum}}} \quad (2)$$

In formula: $Q_{\text{sum}}$ - Total number of urban residents to travel. $Q_{\text{out}}$ - Number of residents traveling by taxi. $Q_{\text{sum}}$ is determined by the total urban population and Residents travel times every two hours, so the formula:

$$Q_{\text{sum}} = R_{sum} \cdot A_{sum} \quad (3)$$

According to Eq.1, Eq.2 and Eq.3, a new formula is obtained.

$$W = R_{sum} \cdot A_{sum} \cdot P \cdot D = Q_{sum} \cdot P \cdot D = Q_{out} \cdot D \quad (4)$$

Taxi's effective mileage can be calculated by the following formula:

$$L_{\text{effective}} = \frac{W}{S} = \frac{Q_{\text{out}} \cdot D}{S} \quad (5)$$

'S' is an average number that passenger takes the taxi at one times. And the following formula can also express $Q_{\text{out}}$ [8,3].

$$Q_{\text{out}} = \frac{\text{m} \cdot S \cdot N \cdot W}{10000} \quad (6)$$

In formula:m - Take a taxi times per two hours. S - The number of passengers per vehicle is within two hours. N- The total number of the taxi. W - Taxi operating ratio. According to formulas of (5) and (6), we can obtain a Eq.7.

$$L_{effective} = \frac{Q_{out} \cdot D}{S} = \frac{\frac{\text{m} \cdot S \cdot N \cdot W}{10000} D}{S} = \frac{\text{m} \cdot N \cdot W \cdot D}{10000} \quad (7)$$

The no-load ratio means that in the proportion of deadhead kilometers accounted for total vehicle mileage. And the no-load ratio formula as follows.

$$k = \frac{L_{\text{sum}} - L_{effective}}{L_{sum}} = 1 - \frac{L_{effective}}{L_{sum}} \quad (8)$$

In the above formula, the total mileage of taxis ($L_{\text{sum}}$).

$$L_{\text{sum}} = T \cdot V \cdot N \quad (9)$$

In Eq.9: T - Taxi driving time per two hours. V - Average travel speed of taxis. N - The total number of the taxi. Lastly, put the total mileage of taxis ($L_{\text{sum}}$) in the no-load( k ), and we can obtain a formula (Eq.10).

$$\text{k} = 1 - \frac{\frac{N \cdot W \cdot D \cdot \text{m}}{10000}}{T \cdot V \cdot N} = 1 - \frac{m \cdot W \cdot D}{10000 \cdot \text{T} \cdot \text{V}} \quad (10)$$

Assuming that W, D, T and V are quantitative, m is variable. What's more the times of taking a taxi(m) to be the taxi demand(G) and the trip distribution of occupied taxis(C) combined effect. As mentioned, proportional relationship between G and C. And We know that the success ratio of taxi is 80% around[9,10]. So we can get a formula.

$$\mathrm{m} = k'\frac{C \times 80\%}{G} \tag{11}$$

In this formula, $k'$ is linearly dependent coefficient. According to the formula (Eq.11) that the times of taking a taxi per two hours(m) can obtain values. When $k'$ is one, we can draw a three-dimensional graphics which independent variable are a period of time and the different of ring road and the dependent variable are no-load (shown in Fig. 5).

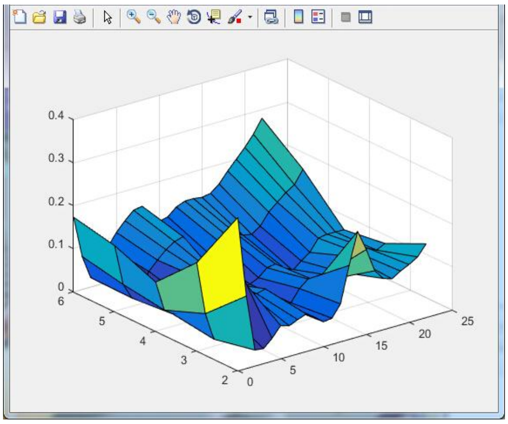



Figure 5.

**Analysis of No-load Ratio Method.** No-load ratio measured the degree of "supply to match". According to the equation, the larger the value of taking a taxi times per two hours, the lower the no-load ratio, and vice versa. It can be got from the three-dimensional image that the no-load ratio is very low after ten pm in the 5th ring road, and the actual load achieves the highest at zero o'clock in 2th, 3th and 4th ring road. According to the observation can come to the conclusion that the actual load ratio is gradually raising trend from 6am to 8am, 10am to 12am and 2pm to 0am. Which proves the number of passengers in the morning and evening peak is very large.

The lower the no-load ratio is, the higher the enterprises and operators of taxi business economic benefits are, but for people it will be more difficult to take a taxi, and also caused the situation that the market demand exceeds supply. On the other hands, the higher the no-load ratio is, the easier to take a taxi, but the revenue of the taxi operators will not be optimistic. These two cases respectively illustrate that the taxi supply and demand are imbalance. Only in the case of a reasonable load ratio, keeping the balance of the number of the taxi and the local residents travel demand. In other words, to achieve reasonable and effective allocation of resources, and convenient for residents to travel, at the same time to ensure the operation of enterprise and managers have a reasonable economic benefit.

## References


[1] Information on https://www.python.org.

[2] D.G Luo. S Feng: Journal of Railway Science and Engineering,Vol.6(2009) No.1, p87.(In Chinese)

[3] W Ying. S Jie. H Jing: School of Civil Engineering, Vol.33(2014) No.2, p93.(In Chinese)

[4] L.N Liu, Y.Y Chen, W.G Zhang: Transport Engineering & Safety, (2010) No.224,p89. (In Chinese)

[5] X.M Feng: Study o*n the Development Scale of Taxi Under the Balance of Supply and Demand* (Master, Southwest Jiaotong University, China 2010). (In Chinese)

[6] X.S Li. Z.Z Chen: Statistical Research, Vol.27 (2008) No.8.(In Chinese)

[7] S Landau. B Everitt: *A Handbook of Statistical Analysis Using SPSS* (Florida,Chapman & Hall, 2004).

[8] S.T Mu: *The Research on the Prediction Model and Solution Algorithm of Urban Taxi Number* (Master, Jilin University, China 2014). p.25. (In Chinese)

[9] Information on http://view.163.com/special/reviews/taxiapp0424.html.

[10] Information on http://auto.hexun.com/2013-05-07/153858870.html.