

Research on taxi software policy based on big data

Daoming Feng

College of Mathematics and Computer Science, Xinyu University, Xinyu 338000, China

Abstract. Through big data analysis, statistical analysis of a large number of factors affect the establishment of the rally car index set, By establishing a mathematical model to analyze the different space-time taxi resource "to match supply and demand" degree, combined with intelligent deployment to solve the "taxi difficult" this hot social issues. This article takes Shanghai as an example, the central park, Lu Xun park, century park three areas as the object of study. From the "sky drops fast travel intelligence platform" big data, Extracted passenger demand and the number of taxi Kongshi data. Then demand and supply of taxis to establish indicators matrix, get the degree of matching supply needs of the region. Then through the big data relevant policies of each taxi company. Using the method of cluster analysis, to find the decisive role of the three aspects of the factors, using principal component analysis, compare the advantages and disadvantages of the existing company's programs. Finally, according to the above research to develop a reasonable taxi software related policies.

1 PROBLEM DESCRIPTION

Taxi is one of the important means of transport for the public, with the "Internet +" era, the popularity of mobile internet, there are a number of companies rely on the mobile Internet to build a taxi software service platform, to achieve the passenger and taxi drivers between the information exchange, making people in the daily life of the taxi to become more efficient and convenient, changed the way the traditional travel taxi, and thus to the general public's life caused a certain impact. However, on the other hand the taxi driver complained that labor intensity, income is relatively low, even the occurrence of a taxi driver strike, this reflects the taxi market management there are some problems. The existing taxi pricing is not reasonable, the speed is too high, leading to the entire taxi industry downturn. The long run will affect social stability, worthy of attention. China's cities in the next period of time, The scale will continue to expand, the population will continue to grow, people's living standards will continue to improve. The demand for a taxi will also change. How according to the city's population and travel intensity and the share of the taxi, To develop effective and practical programs in line with the strategic objectives of urban development, Plan a reasonable number of car rental, to maximize the people to meet the travel needs, at the same time, according to reduce environmental pollution and resource consumption. Coordinate the interests of all walks of life, is worthy of further study.

Taxi is one of the important vehicles of public travel, "Take a taxi difficult" is a hot topic of social concern. With the era of "Internet +", there are a number of companies rely on mobile Internet to establish a taxi service platform software, and information exchange

between passengers and taxi drivers, together with a variety of taxi subsidy scheme.

Through data analysis, research mathematical model the following questions:

- (1) try to establish a reasonable indicators, and analyze the different space-time taxi resource "to match supply and demand" degree.
- (2) analysis of each company whether the taxi subsidy program "to alleviate taxi difficult" helpful?
- (3) to a taxi service software platform for the newly created design a reasonable subsidy program.

2 Problem analysis

2.1 Analysis of the first question

For the first question, through the establishment of a city to study the G region of the supply and demand matching, the number is (D_1, D_2, \dots, D_G) , the public daily in 24 hours a day unit of the demand for indicators of the taxi $(Q_1, Q_2, \dots, Q_{24})$ and taxi supply index $(V_1, V_2, \dots, V_{24})$. Use the formula

$$D_i = \frac{V_{i,j} - Q_{i,j}}{Q_{i,j}} \quad i = 1, 2, 3 \dots G; j = 1, 2, 3 \dots 24 \quad (1)$$

the supply and demand matching model of a certain area in different time periods is established by analyzing the degree of "supply and demand matching" of taxi resources in different time and space. Through the analysis of supply demand data about time series, The image more intuitively illustrates the change in the supply demand relationship in each region, for further

results analysis. But also through the past few years the demand for taxis.

2.2 Analysis of the second question

For the second question, through the Baidu search can understand the taxi company to ease the taxi and their own profits promulgated by some of the policies (here we listed nine more representative of the policy), we can through clustering analysis These nine policies are divided into three categories, namely:

- (1)taxi promotion policy for software;
- (2)taxi company incentives for non-peak car customers;
- (3)the taxi driver's reward;

And then through the questionnaire survey and expert interviews, respectively, the three main factors of fuzzy comprehensive evaluation, the establishment of models for analysis.Through these three programs as a program layer to explore the distribution of indicators of weight, and then get the level of comprehensive sort, and then compare Get the merits of the existing company's programs.

2.3 Analysis of question three

On the basis of the third question and the first and the second question, the author puts forward a relatively reasonable scheme in combination with the relevant actual situation and related regulations in combination with the weight of the three main decision-making schemes.

3 General assumptions and definitions

3.1 Assumptions

- Assuming that the demand of passengers every day,like the law does not consider the impact of holidays,
- Assuming that the taxi company has a reasonableallocation of taxis,
- Assuming that the latitude and longitude within a certain range can well represent the vicinity of a region,
- Assumptions consistent commuting time on the ground of general corporate or enterprise,
- Assuming that the results which using the big data analysis can effectively reflect the supply demand situation in the region.

3.2 Terms and definitions

Table 1. Terms and Definitions

$D_1, D_2, D_3 \dots D_G$	1 to G regions of the matching degree
$Q_1, Q_2, Q_3 \dots Q_{24}$	The demand matrix for different time periods in a given day
$V_1, V_2, V_3 \dots V_{24}$	Taxis supply index matrix for different periods of time in a day
X_{ij}	Sample matrix

R_j	Policy factors
W_j	Weights

4 Establishment and solution of the model

4.1 Problem one model is established and solved

4.1.1 The establishment of evaluation indicators

Taxi traffic has different eigenvalues as time changes and location displacements, and is analyzed on a day-by-day basis. According to the number of months, the day is divided into $T_1, T_2, T_3 \dots T_{24}$, a total of 24 time units, assuming that the day's passenger demand for Q, then 24 hours a day in the unit of passenger demand corresponding to $Q_1, Q_2, Q_3 \dots Q_{24}$, to meet:

$Q_1 + Q_2 + Q_3 + \dots + Q_{24} = Q$; The number of empty taxi is V, the number of empty taxis in a certain period of time is marked, assuming that it is $V_1, V_2, V_3 \dots V_{24}$, then:

$V_1 + V_2 + V_3 + \dots + V_{24} = V$; Assuming that the supply and demand of G regions are matched, then the number is $D_1, D_2, D_3 \dots D_G$, for G regions, The demand Q for a day in the area is a matrix and the number of empty taxi numbers V (ie supply number) is a matrix, respectively:

$$\begin{bmatrix} V_{1,1} & V_{1,2} & \dots & V_{1,24} \\ V_{2,1} & V_{2,2} & \dots & V_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ V_{G,1} & V_{G,2} & \dots & V_{G,24} \end{bmatrix} \begin{bmatrix} Q_{1,1} & Q_{1,2} & \dots & Q_{1,24} \\ Q_{2,1} & Q_{2,2} & \dots & Q_{2,24} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{G,1} & Q_{G,2} & \dots & Q_{G,24} \end{bmatrix} \quad (2)$$

And then use the formula:

$$D_i = \frac{V_{i,j} - Q_{i,j}}{Q_{i,j}} \quad i = 1, 2, 3 \dots G; j = 1, 2, 3 \dots 24 \quad (3)$$

you can find the matching of different regions j supply and demand. If the absolute value of D_i is closer to zero, it means that the matching degree of supply and demand is better. If the absolute value of D_i is larger, the greater the matching degree of supply and demand is. If D_i is greater than zero, the supply is oversupply, D_i is less than zero, the amount.

4.1.2 Data collection and processing

This problem is mainly research and analysis In the "Internet +" era under the influence of some taxi software with the Internet to build a taxi software service platform to achieve the passenger and taxi drivers between the information exchange, influence and change the traditional way of taxi Circumstances, the taxi industry in different time and space regions of the impact of supply and demand matching. But also in the existing "didi taxi", "fast taxi", "haha carpool" and other software, the biggest difference in the market share of

the taxi, so the data from the title of "sky dripping fast intelligent travel platform," the large data, Out of Shanghai within a day of passenger demand and the number of vacant taxi part of the data. See Appendix 1 for specific data. (In Appendix 1, each data is composed of a square bracket which consists of four parts separated by semicolons, such as: ["", 121.4861, 31.2269, 861], the first data is double quotation marks, Is estimated to be a code, the second third, respectively, said latitude and longitude, the fourth represents the latitude and longitude of the needs and supply).

In this paper, by selecting the Shanghai Central Park (longitude: 121.46906000000001 latitude: 31.23169), and then within a certain range of the central park (ie, longitude: 121.46 latitude: 31.23) within a day of travel and demand data, City Lu Xun Park (longitude: 31.27, latitude: 121.49) Century Park (longitude: 31.21, latitude: 121.55) The specific data are as follows:

Table 2.Center Park Taxi Supply Demand Table

Time	Demand	Supply	Time	Demand	Supply
0:00	54	30	12:00	299	199
1:00	60	35	13:00	274	224
2:00	66	43	14:00	271	211
3:00	77	50	15:00	204	156
4:00	104	66	16:00	195	105
5:00	137	90	17:00	392	235
6:00	171	150	18:00	221	202
7:00	259	203	19:00	206	152
8:00	322	300	20:00	249	186
9:00	291	250	21:00	216	143
10:00	266	203	22:00	204	125
11:00	235	171	23:00	67	21

Table 3. Lu Xun Park Park Taxi Supply Demand Table

Time	Demand	Supply	Time	Demand	Supply
0:00	151	45	12:00	352	266
1:00	158	63	13:00	341	231
2:00	163	71	14:00	285	227
3:00	181	76	15:00	257	189
4:00	209	93	16:00	171	146
5:00	215	114	17:00	198	175
6:00	221	166	18:00	353	211
7:00	289	215	19:00	335	231
8:00	360	265	20:00	351	254
9:00	563	261	21:00	309	218
10:00	426	235	22:00	260	154
11:00	311	213	23:00	169	75

Table 4.Century Park Taxi Supply Demand Table

Time	Demand	Supply	Time	Demand	Supply
0:00	59	23	12:00	141	83
1:00	72	21	13:00	137	62
2:00	68	26	14:00	131	50

3:00	71	30	15:00	123	47
4:00	85	34	16:00	120	42
5:00	93	39	17:00	123	63
6:00	91	42	18:00	134	78
7:00	101	57	19:00	145	92
8:00	124	83	20:00	153	101
9:00	130	89	21:00	113	92
10:00	266	203	22:00	204	125
11:00	235	171	23:00	67	21

4.1.3 Explanation and explanation of the phenomenon

Through the above data table can be obtained $V_{i,j}, Q_{i,j}$ supply and demand matrix, through

$$D_i = \frac{V_{i,j} - Q_{i,j}}{Q_{i,j}} (i = 1, 2, 3 \dots G; j = 1, 2, 3 \dots 24) \quad (4)$$

Calculated The supply matching degree of the central park $D_1 = 0.3132$;

Lu Xun Park supply demand matching degree $D_2 = 0.5226$;

Century Park supply demand matching degree $D_3 = 0.7373$;

Further, as long as we have obtained the supply demand data for each region, we can use a similar approach to find the supply demand matching relationship in the corresponding region.

Through the calculation results can be drawn:

$$D_1 < D_2 < D_3,$$

That is: the center of the park around the scope of the supply needs of the best matching, the scope of the century around the park supply demand matching the worst degree; and D_1, D_2, D_3 are greater than zero, so you can know the three areas of the supply demand of the taxi are Is the supply is greater than the demand, indicating that the taxi industry there is a greater competitive edge, on the other hand can also show that customers can hit the car in the more time, to meet the requirements of customer taxi!

Through the above data form, and further, we through the excel to draw the supply chain of each region needs to change the map, you can more intuitive image of the supply needs of the changes observed, the specific image below:

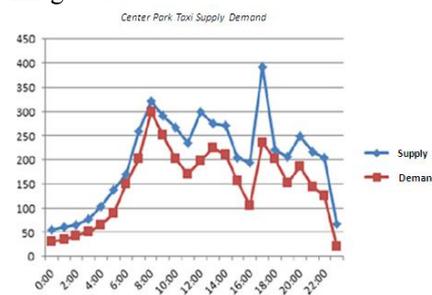


Fig. 1. Center Park Taxi Supply Demand

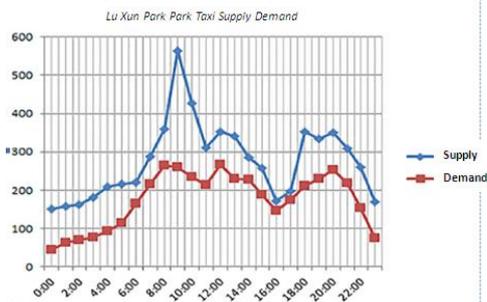


Fig. 2.Lu Xun Park Park Taxi Supply Demand

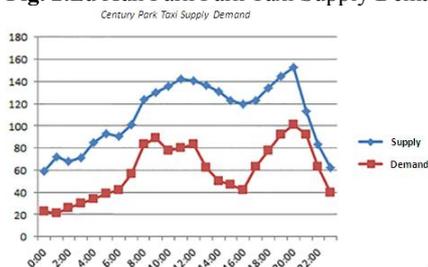


Fig. 3. Century Park Taxi Supply Demand

Through the image, you can intuitively know: the supply is greater than the demand, and generally in the morning 7: 00-9: 00 and 18: 00-20: 00, these times the supply and demand of the smallest difference between that In the case of a taxi company taxi supply certain circumstances, 7: 00-9: 00:00 and 18: 00-20: 00 These two time periods are commuting to the peak of commuting, so the passengers are relatively more; In other times, people flow is relatively small, relatively less car customers, and some customers will choose the subway or bus, which caused a certain impact on the taxi industry.

4.2 Problem 2 model building and solving

4.2.1 Taxi company related subsidy policy

February 17, 2014 Di tick hit the car, and its Wechat to pay the third round of marketing activities officially opened, to restore subsidies and has been strengthened. Beijing, Shanghai, Shenzhen, Hangzhou drivers with Wechat to pay the collection fee, 10 yuan per single, 10 per day, the other city drivers every day The first 5 single prize 5 yuan each, after 5 single 10 yuan per single prize. It is noteworthy that the Di Di car extra reward new users: passenger first single by 15 yuan, the first driver of the first prize 50 yuan.

4.2.2 Utilization of Big Data on subsidy programs analysed

This paper is to study the various regions of Shanghai taxi supply matches extent,Through the network platform data on the Shanghai area of the public taxi situation and the impact of the subsidy policy,Converted to the main factor of the final total score U_i , and then through the formula.

$$W_i = \frac{U_i}{\sum_{i=1}^3 U_i} \quad (5)$$

to determine the weight of each factor.Finally, by summing up the statistics, the final score of each factor is obtained .

Table 5. Policy weight score

R_i	Index value	Total score U_i	Specific gravity W_i
Taxi software promotion R_1		654	31. 53%
Each time period Subsidy for car customers R_2		723	34. 86%
The reward for the driver R_3		697	33. 61%

From the table, we can see that the proportion of subsidies to non-peak passenger customers is the largest, which reflects the common problems of taxi companies and the general public. Secondly, the proportion of incentives to drivers is the largest, indicating that the taxi drivers Subsidies to make the driver can more actively meet the needs of customers; the last is the promotion of taxi software policy favored by customers, but the result is relatively low. The reason is based on the analysis because some older and some younger people are less accustomed to using taxi software to spend.

4.2.3 The solution to the problem three

From the second analysis of the results we can see: When the taxi company's decision-makers in the development of policies, we should give priority to the factors from different time to consider the improvement of customer satisfaction,and then combined with the driver's reward and punishment system.Finally, in combination with the promotion of taxi software policy,making the relevant policies can not only reflect the broad masses of customers the taxi consumption needs,but also to maximize the profitability of the taxi company to achieve a win-win situation.Thereby promoting attracting more consumer spending;Finally, through the careful consideration of the company's leadership, the following three policies have been drawn: (1) every Monday to Thursday 06: 00-10: 00, 20 yuan per subsidy; Every Monday to Thursday 16: 00-22: 00, 10 yuan per subsidy. Every Friday 06: 00-22: 00, 30 yuan per subsidy. Every Friday 16: 00-24: 00, 25 yuan per subsidy. Every Saturday, Sunday 0: 00-2: 00, each compensation 10 yuan. Every Saturday, Sunday 11: 00-24: 00, 10 yuan per subsidy. (2) the incentive for the driver according to the weekly travel volume and the customer under the number of equal weight and reward:On the completion of 40 weekly or more weekly travel more than 1,500 km of the driver, an additional reward of 200 yuan;On the completion of 50 per week or more than the weekly

travel volume of more than 1,800 km of the driver, an additional reward of 500 yuan; On the weekly completion of 80 single or more weeks more than 2200 km trip driver, an additional reward of 850 yuan; On the completion of 100 per week or more than the weekly travel volume of more than 2500 km of drivers, additional awards 1,200 yuan;
(3) the use of taxi software then subsidize customers ranging from 8-20 yuan. Daily subsidy three single, and for seven consecutive days using taxi software more than ten times, to carry out the activities of sending calls. In every day among all the use of taxi software, customers randomly selected part of the 5-10 yuan per person to pay as a subsidy.

7. Li Juan. Research on the optimization of traditional taxi return and low speed driving cost [D]. South China university of technology, 2016.

5 Result analysis

5.1 This article is worth learning from

In this paper, large data analysis, with persuasive; The model of this paper directly through the taxi company's supply and demand relationship to illustrate the degree of matching analysis, the theory is conducive to the realization; In this paper, clustering analysis, with a wide range of applicability; The establishment of the planning model with universal, combined with local data can be directly promoted; Due to the limitation of the data span, we can only analyze the data given in the topic, and the model is rather rough and omitted some cases.

5.2 Need to be improved

Data processing needs to be further precise; The collection of inquiries on the policy is not perfect.

References

1. Zhiqiang Qu, jing Qiao. Effects of Big Data Processing on E-commerce [J]. Chinese Business Theory, 2015, (16): 61-63..
2. Jie Jia, Zongzhe Sun, Ming-ming Mao, Yubin Bao, Research on MapReduce Large Data Processing Platform and Algorithm [J]. Journal of Software, 2017, 28 (03): 514-543.
3. Li xiaofei. Research on big data processing system based on cloud computing technology [J]. Journal of changchun engineering college (natural science edition), 2014, 15(01).
4. Juntong Wang, Shitong Wang, Zhaohong Deng. There are Some Problems in Cluster Analysis [J]. Control and Decision, 2012, 27 (03): 321-328.
5. Deng xue, Jiaming Li, Chen Junyang, jun-feng zhao. Ahp weight calculation method for the analysis and application study [J]. Journal of mathematics practice and understanding, and 2012 (7) : 93-93. [2017-09-28]..
6. Liu xiao. Research on development strategy of didi taxi company [D]. Guangdong university of finance and economics, 2016.