

Model construction of “earning money by taking photos”

Jingmei Yang

School of mathematics and statistics, Qujing Normal University, Qujing, Yunnan

*Corresponding author e-mail: 421327416@qq.com

Abstract. In the era of information, with the increasingly developed network, "to earn money by taking photos" is a self-service model under the mobile Internet. The user downloads the APP, registers as a member of the APP, and then takes a task that needs to take photographs from the APP and earns the reward of the task on the APP. The article uses the task data and membership information data of an already completed project, including the member's location and reputation value. On the basis of reasonable assumption, the data was processed with the MATLAB, SPSS and Excel software. This article mainly studied problems of the function relationship between the task performance, task position (GPS latitude and GPS longitude) and task price of users, analyzed the project's task pricing rules and the reasons why the task is not completed, and applied multivariate regression function and GeoQ software to analyze the data, studied the task pricing rules, applied the chart method to solve the complex data, clear and easy to understand, and also reality simulation is applied to analyze why the tasks are not completed. Also, compared with the previous program, a new task pricing program is designed for the project to obtain the confidence level by means of the SPSS software, to estimate the reasonable range of the task pricing, predict and design a new pricing program on the reasonable price range.

1. Problem analysis

In the era of information, with the increasingly developed network, "to earn money by taking photos" is a self-service model under the mobile Internet. The user downloads the APP, registers as a member of the APP, and then takes a task (such as to go to a supermarket to check the listing of a certain product) that needs to take photographs from the APP and earns the reward of the task on the APP. This kind of self-service crowdsourcing platform based on mobile Internet provides enterprises with a variety of business inspection and information collection. Compared with the traditional market research methods, it can greatly save the investigation cost, effectively ensure the authenticity of survey data and shorten the investigation cycle. Therefore, the APP has become the core of the platform operation, and the task pricing of the APP is its core element. If the price is not reasonable, some tasks will be not noticed by people, which led to the failure of commodity inspection.

Aimed at the task pricing rules and finding out the reasons why the task is not completed, task completion and the member location are clearly shown on the GeoQ software. Because pricing is a systematic exercise, it is able to initially guess that the price is related to the task location, the member location, the task completion, app downloading users and many other factors, which is a problem of the relationship between multiple variables, so the pricing rule is found by building linear or nonlinear mathematical models for multiple variables.



Aimed at the problem of designing a new task pricing program for the project, the statistical analysis of the completion and the unfinished quantity corresponding to the task price is carried out with help of EXCEL, and then the scatterplot is drawn according to the relationship between the GPS longitude and latitude, the task prices, from which the relationship between task execution, GPS latitude and longitude, task pricing can be found. The SPSS software is applied to estimate range of the reasonable task prices in the new program. Finally, data fitting was carried out with EXCEL to find out the correlation between task prices, GPS latitude and longitude, also the data fitting chart is used to design a new pricing program.

According to distribution of the known data and the location of members, through the GeoQ map and MATLAB, task completion and the location of members can be seen clearly on the map. Pricing is a systematic job, through which it is able to initially guess the relationship between pricing and task location, member location, task completion, app downloading and many other factors, and is a problem between multiple variable factors. It is also known that multivariate linear regression analysis mainly studies the correlation between a dependent variable and multiple independent variables, also known as multiple linear regression. It is a statistical method that reflects the rule that the number of phenomena or things changes according to a variety of linear or quantitative variations and establishes the quantitative relationship of linear or nonlinear mathematical models among multiple variables. Therefore, the principle of multiple regression analysis is applied for the discussion. Through the theoretical analysis, the rule of task pricing is studied rationally, and the reason of the task unfinished is analyzed with reality simulation.

2. Establishment and solution of the model

2.1. Model establishment for task pricing rules

The problem is to consider the influences of longitude and latitude changes of the task on the price, so the price is regarded as an independent variable, latitude and longitude are regarded as dependent variables, in line with regression of multiple dependent variables on the independent variables, so the multiple linear regression analysis is applied.

y is the price, $x_1, x_2, x_3, x_4, x_5, \dots, x_k$ are the dependent variables associated with the price, so the multivariate linear regression model is: $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k + e$

Where, b_0 is a constant term, b_1, b_2, \dots, b_k for unknowns parameters, namely the regression coefficient, e for random error terms. Parameter estimation of the multivariate linear regression model is the same as that of the unary linear regression equation, in which the least square method is applied to obtain the parameters on the prerequisite that the sum of squares of errors ($\sum e^2$) is the minimum.

Due to the large amount of data, the SPSS software is applied to separately sample longitudes, latitudes and prices, and 300 sets of data are randomly selected for analysis, with x_1 for the longitude, x_2 for the latitude, y for the price, and the extracted data is correspondingly assumed and taken into the following equation, to solve the standard equation set of the regression parameters:

$$\begin{cases} \sum y = nb_0 + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum x_1 y = b_0 \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_1 x_2 \\ \sum x_2 y = b_0 \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{cases}$$

Value of b_0, b_1, b_2 can be obtained when the equation is solved.

$b = \frac{(x'y)}{(x'x)}$ can also be obtained with the following matrix method.

$$\begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_1 & \sum y \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2 y \end{bmatrix}$$

Estimated values of b_0, b_1, b_2 can be obtained, to establish the multiple regression linear model, in which residual sum of squares can be obtained. If the residual is too large, indicating that there is no dependency between them, on the contrary, if the residual is too small, indicating the existence of dependencies between the price and location. Through the comprehensive analysis, the residual sum of squares is too large, and the fitting degree is not high, so it is inferred that there is no interdependence between the price and location.

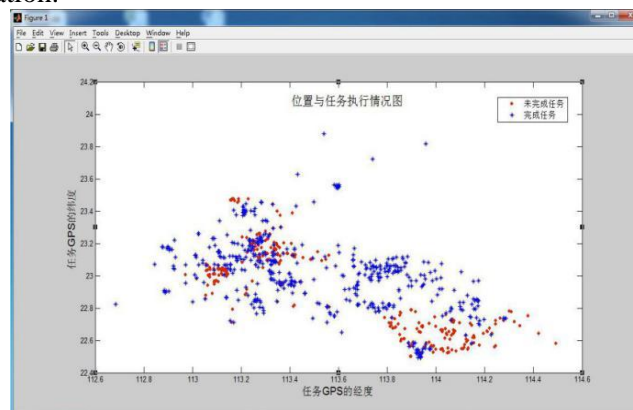


Fig.1

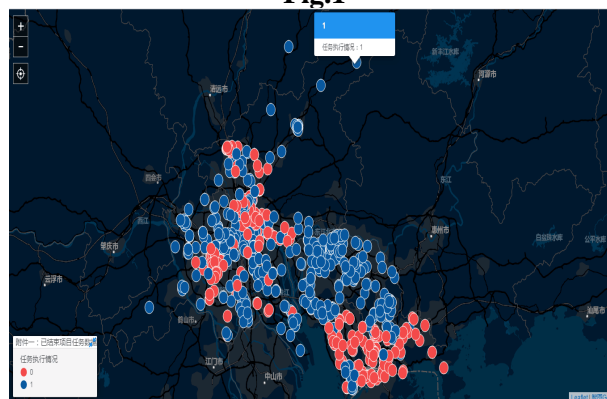


Fig.2

According to the GPS longitude, latitude, execution of the task, the distribution of the task execution is drawn on the map with GeoQ and MATLAB. On the figure, blue * points on the MATLAB map shows the location of the completed task, and the red dot indicates the uncompleted task (Figure 1).

On the GeoQ map, the blue dot indicates location of the task with the execution status of "1", and the red dot indicates the task with the execution status of "0" (Fig. 2). It can be seen from both figures that the blue dots are relatively scattered, that is, the locations where the completed tasks are scattered while the red dots are more concentrated than the blue ones, that is, the locations of the uncompleted tasks are more concentrated. Factors to be considered here are: member location, task price and other factors.

The same as above, distribution of the member location is shown with the MATLAB and GeoQ software, on which the member locations are shown to be centralized, it can be found through the combined analysis with the distribution of the member location, and task execution: part of the tasks close to the member location failed to be performed, and some task executions with a sparse location

distribution are successful, that is to say, task completion is not good in the area where the membership is centralized.

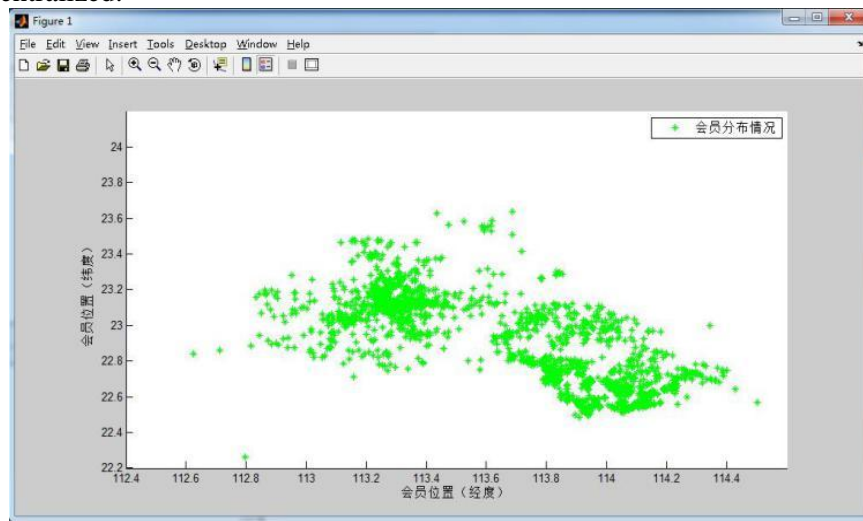


Fig.3

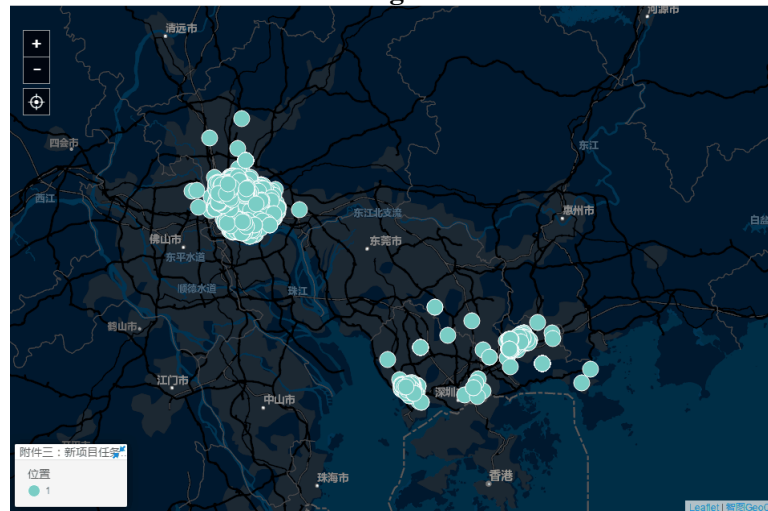


Fig.4

The relationship chart between the task price and the task execution is drawn with EXCEL, shown on the figure, and the orange points on the abscissa represents that the task is completed, while the dot with no color indicates that the task is not completed. (Fig.5)

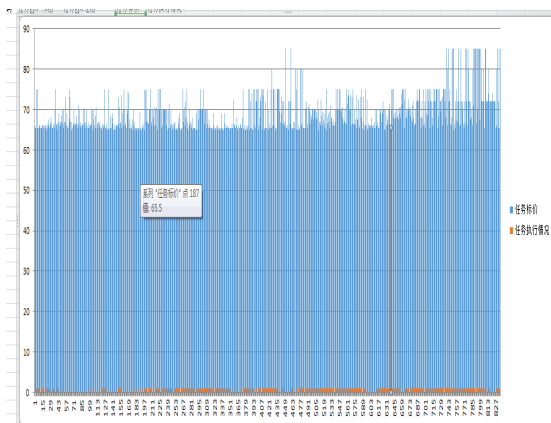


Fig.5

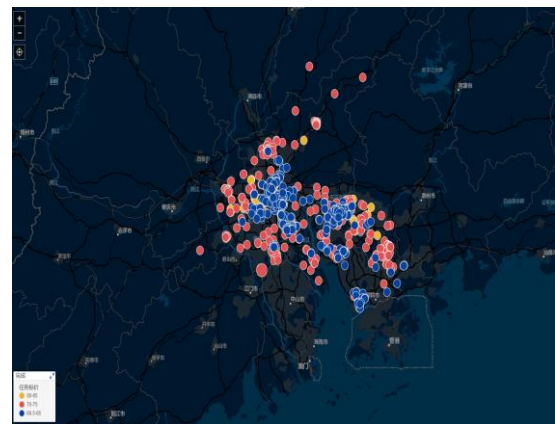


Fig.6

The ordinate represents the marked price, and it can be seen the average value of the marked prices is around 65 yuan. (Fig.5)

Finally, the location “1” of task execution is counted, and the task price is divided into sections, with price of the blue area between [65,69.5], price of the red area between [70,75], price of the yellow area between [80,85]. Compared with the previous member distribution map, it can be seen that at locations where the members are densely populated, the prices are lower. (Fig.6)

Through comprehensive analysis of the above chart, the task pricing rules can be roughly obtained:

The pricing program is related to the distribution of member locations. Where members are densely populated, the prices are lower. This may be because the number of members is too large and there are also many people who are willing to work. Therefore, the price is lowered. For areas with a more scattered membership, the price is on the average or higher.

2.2. Analysis of reasons for the unfinished task

It can be seen from the figure, there is a close relationship between the completion and price of a task:

(1) The task completion rate is lower in the locations where members are centralized, while higher in the locations where members are scattered, possibly due to the lower price, which may affect the enthusiasm of members; or the members may give up because of failure to winning the order.

(2) Should the pricing be reasonable for the member's location, but the completion rate of is not high in the area where the members are centralized, indicating that difficulty coefficient of the task assigned by the merchant is large, or the member does not receive the accurate task information.

2.3. Model establishment and solution of the new task pricing program

2.3.1. Statistics for completion under the task price

Firstly, statistics of completed and uncompleted quantity of the task execution under each task price is carried out, with red marking the task prices for the 3 tasks with the largest completed quantity, green marking the task prices for the 3 tasks with the smallest uncompleted quantity. From the table, it can be found that the completed quantity of the tasks with a higher price is not necessarily higher, and the possible reasons for why the task cannot be completed are the difficulty of the task, traffic convenience, flow of people and so on.

Table 1

Task price	Completed quantity	Uncompleted quantity
85	24	3
80	9	4
75	59	19
74.5	1	1
74	1	4
73.5	3	2
73	7	3
72.5	7	2
72	42	18
71.5	4	1
71	3	1
70.5	9	2
70	11	19
69.5	6	2
69	12	7
68.5	6	5
68	25	5
67.5	17	6
67	18	20
66.5	35	28
66	46	57
65.5	76	74
65	35	30

2.3.2. Relationship between task pricing and task location

Given that under different longitudes and latitudes, the difficulty of the task, the traffic convenience, the flow of people and so on are the same, the scatter plot is drawn according the location and price of the task. It can be seen that the range of the task prices is concentrated between 65-75, shown as below:

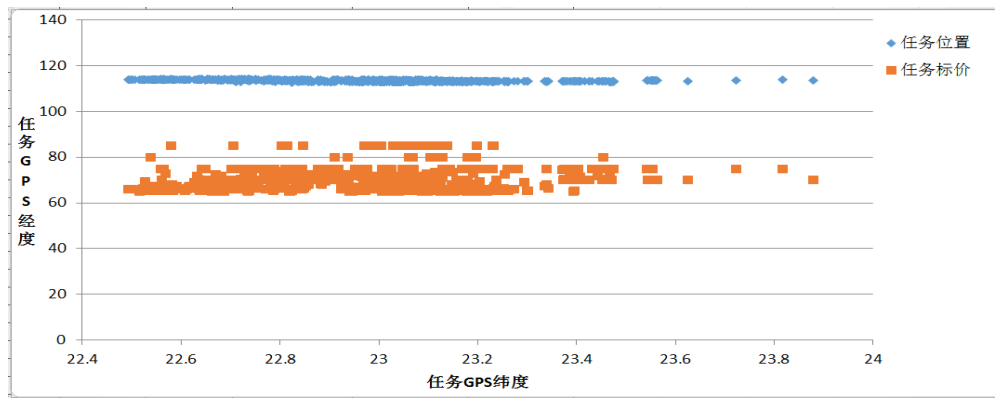


Fig.7

2.3.3. Estimation for the reasonable range of the task price

Firstly, task prices are classified and collected, in order to estimate the reasonable task pricing at different latitudes and longitudes, the following random samples are taken for the overall $N(\mu, \sigma^2)$, and the confidence interval of μ is calculated to be 0.95. Also, for the test task price, whether the overall average task price is significantly different from the price of each task at different latitudes and longitudes, and analysis can be carried out according to the data in the table.

① Propose the original hypothesis and alternative hypothesis: for the test task price, there is no significant difference between the overall average task price and price of each task at different latitudes and longitudes H_0 there is a significant difference between the overall average task price and price of each task at different latitudes and longitudes.

② Confirmation of the test statistic:

$$t(n-1) \sim \frac{\bar{x} - \mu}{s/\sqrt{n}},$$

$t(n-1) \sim \frac{\bar{x} - \mu}{s/\sqrt{n}}$, \bar{x} is the average value of the sample, n for the number of samples, s for sample standard deviation.

③ It can be obtained through calculation:

$$\bar{x} = 71.087, \quad s = 4.7210, \quad df = 23 - 1 = 22,$$

$$\alpha = 0.95, \quad \frac{\alpha}{2} = 0.025, \quad t_{0.025}(5) = 2.571$$

$$-t_{0.025}(5) \sim \frac{\bar{x} - \mu}{s/\sqrt{n}} < t_{0.025}(5)$$

$$-2.571 \sim \frac{\bar{x} - \mu}{s/\sqrt{n}} < 2.571$$

$$\bar{x} - s/\sqrt{n} t_{0.025}(5) < \mu < \bar{x} + s/\sqrt{n} t_{0.025}(5)$$

$$71.087 - \frac{4.7210}{\sqrt{23}} * 2.572 < \mu < 71.087 + \frac{4.7210}{\sqrt{23}} * 2.571$$

$$69.045 < \mu < 73.128$$

4 It can be concluded that the reasonable range of task price in different latitudes and longitudes is [69.045,73.128].

5 This question is a two-sided test, if the significance level of the test is $\alpha = 0.05$, and the critical value $t_{0.025}(5) = 2.571$, $|t| = 72.214 > 2.571$ can be obtained with reference to the distribution table, showing that the t value is within the accepted domain, namely the alternative hypothesis is accepted. Therefore, it can be considered that there is a significant difference between the task price and the overall average task price at different latitudes and longitudes.

Also, it can be obtained on the SPSS software:

Table 2

单个样本统计量						
	N	均值	标准差	均值的标准误		
任务标价	23	71.087	4.7210	.9844		

单个样本检验						
检验值 = 0						
	t	df	Sig. (双侧)	均值差值	差分的 95% 置信区间	
					下限	上限
任务标价	72.214	22	.000	71.0870	69.045	73.128

In the above test results, Sig. (two-sided) = 0.90400 > 0.05, so the alternative hypothesis is accepted, in which it is considered that there is a significant difference between the task price and the overall average task price at different latitudes and longitudes.

(Note: In the case of a significance level of $\alpha = 0.05$, the Sig value in the result of the hypothesis test is compared with 0.05. If the Sig value is less than 0.05, the test result would be that the original hypothesis is rejected. If the Sig value is greater than or equal to 0.05, the test result would be that the original hypothesis is accepted)

2.3.4. Data fitting for the price, the GPS longitude and latitude data of the task

It can be obtained from the above equation that task pricing, GPS latitude and longitude are correlated, and the relationship between the three is assumed as: $y = c_1x_1^2 + c_2x_2 + c_3$

The fitting results can be obtained through data fitting between the task price and GPS latitude directly with EXCEL:

$$c_1 = 0.1836, c_2 = -6.2265, c_3 = 115.23, R^2 = 0.0145$$

The obtained relation formula is as follows: $y = 0.1836x_1^2 - 6.2265x_2 + 115.23$

The data fitting chart is shown as below in Fig.8:

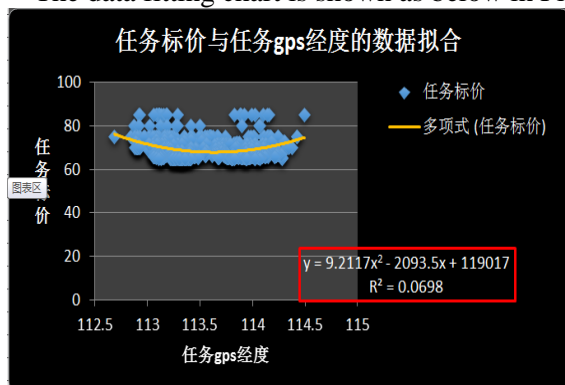


Fig.8

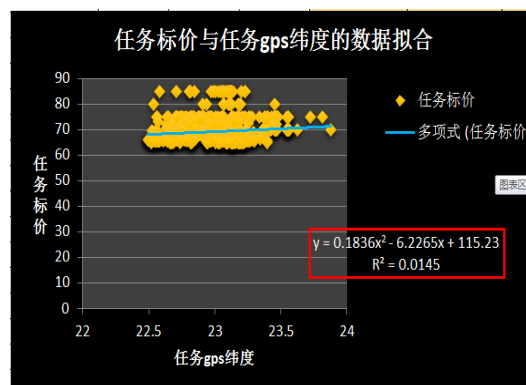


Fig.9

Similarly, the task price and GPS longitude, longitude are correlated, the relationship between the three is assumed as: $y = c_1x_1^2 + c_2x_2 + c_3$

The fitting results can be obtained through data fitting between the task price and GPS latitude directly with EXCEL

$$c_1 = 9.2117, c_2 = -2093.5, c_3 = 119017, R^2 = 0.0698$$

The obtained relation formula is as follows: $y = 9.2117x_1^2 - 2093.5x_2 + 119017$. Data fitting chart (Fig.9):

From the above calculation, it can be concluded that the task prices are concentrated in the range of 69-73, of which the GPS latitude is about 22.5 and the longitude is about 113.9, so the error fluctuation range is small. It can be seen from the solving process with SPSS and EXCEL that the reason for the failure of the priced task is lack of consideration on the impact of the GPS location on the priced task (GPS longitude and GPS latitude).

Based on the above judgment, compared with the original program, it will greatly improve the task execution rate, and factors that have influences on the GPS latitude and longitude are fully considered, which can well meet the demand of users.

3. Evaluation and improvement of the model

3.1. Evaluation of the model

3.1.1. Data statistics and evaluation

The problem of finding out the rules for the task pricing is to carry out statistics and analysis for the existing data, analyze the data with multiple regression functions and the GeoQ software. The method to solve the complex data rules with chart is easy to understand. The process is rigorous, the result is reliable and convincing.

3.1.2. Evaluation of data processing

The problem of designing a new task pricing program, in which the SPSS software is applied to obtain the confidence interval and estimate the reasonable range of the task price. To design a new price program focusing on estimation of the reasonable range of the price is helpful to explain the completion rate of the execution on the new program.

3.2. Improvement of the model

In view of many factors taken into account, there would be some incomplete consideration. For accuracy of the data, because there is no definite equation to show the relationship between the data, as a result, the accuracy is low and there are errors in the calculation conclusion.

The parameters of the regression model are not comprehensively selected, and some minor problems are not taken into account. For example, the economic status, the traffic convenience, the flow of people and other factors at the location where the task is assigned, and the factors including economic status and the traffic convenience should be considered. However, these statistics are complicated and can add a great deal of difficulty in modeling.

References

- [1] Han Zhonggen. A Practical Course of Mathematical Modeling. [M]. Beijing: Higher Education Press, 2013.
- [2] Edited by w. F. Lucas (U.S.A), translated by Cheng Lizhi et, al. Discrete and System Models. [M]. Beijing: National University of Defense Technology, (1996).
- [3] Zhang Wentong. A Basic course of SPSS Statistical Analysis (The second edition) [M], Beijing: Higher Education Press, 2011.
- [4] Jiang Qiyuan, Xie Jinxing, Ye Jun. Mathematical Model [M]. Beijing: Higher Education Press (The fourth Edition), 2003.
- [5] Zhang Shengqin. A Practical Course of MATLAB 7.0 [M]. Beijing: Mechanical Industry Press,

2006.

- [6] Website for the task execution status:

<https://media.geoq.cn/media/share/pc.html?appid=ded4f973-e4fa-4380-98f9-6f88797ff016>

- [7] Website for the task prices:

<https://media.geoq.cn/media/share/pc.html?appid=7ba8148e-42e1-43d5-9d85-5d68930ddbcd>