

# A Pricing Scheme for the Task “Take Photos to Make Money”

Qiao-Xu Qin<sup>1</sup>, Yuan-Biao Zhang<sup>1,2</sup>

<sup>1</sup>Innovation Practice Base of Mathematical Modeling, Electrical and Information College of Jinan University, Zhuhai, China

<sup>2</sup>Key Laboratory of Product Packaging and Logistics of Guangdong Higher Education Institutes, Jinan University, Zhuhai, China

Correspondence: Yuan-Biao Zhang, Innovation Practice Base of Mathematical Modeling, Electrical and Information College of Jinan University, Zhuhai, China.

Received: October 8, 2018

Accepted: October 24, 2018

Online Published: November 8, 2018

doi:10.5539/ibr.v11n12p26

URL: <https://doi.org/10.5539/ibr.v11n12p26>

## Abstract

In this paper, we first use the multiple stepwise regression to fit the price function and analyze the reason of the unfinished task from many angles. Then, by using the idea of "optimal neighbor", the matching model of task and member is established to optimize the original pricing model. Then, starting from the "small world" model, the actual packaging situation was simulated, and the price correction function was defined to reduce the impact of task distribution density on the completion situation, and the pricing scheme was further optimized. Finally, the new project is priced with an optimal package pricing formula, and the implementation effect is judged from the two perspectives of task completion and simulation results.

**Keywords:** simulation of multiple regression, matching, close to optimal, simulation

## 1. Introductions

### 1.1 Background Description

Crowdsourcing, a business model that has emerged on the Internet in recent years, refers to companies outsourcing work traditionally done by internal employees or external contractors to a wide range of Internet users. Professional services the crowdsourcing model applying information technology to service all of the link, to marketing, organization and management of outsourced work via the Internet users, rely on the expertise of Internet users to provide customers with professional services, the model introducing e-commerce, crowdsourcing professional services, for the traditional service industry has brought impact and change (Liu su, 2011).

"Take photos to make money" is a self-service model of mobile Internet. The user downloads the APP, registers as a member of the APP, and then gets the task that needs to be photographed from the APP (such as going to the supermarket to check the shelf situation of a certain product), and earns the remuneration as set by the APP for the task. This self-service crowdsourcing platform based on mobile Internet provides enterprises with various business inspection and information collection. Compared with the traditional market survey method, it can greatly save the investigation cost, effectively guarantee the authenticity of the survey data and shorten the investigation cycle. Therefore, APP becomes the core of the platform, and task pricing in APP is the core element. If the price is not reasonable, some tasks will be neglected, resulting in the failure of commodity inspection.

### 1.2 Research Design

Firstly, data of finished tasks and member information collected from an app are respectively annex 1 and annex 2. The task pricing rule of the project is studied, and the reason why the task is not completed is analyzed. This paper collected the data of the location, price and execution of the completed task on an app. In order to explore the pricing law, it is necessary to find the factors influencing the pricing, analyze the pricing law based on the influencing factors and the actual situation, and define the expression of pricing law. Then, based on the pricing rules and the actual situation, the factors that may affect the task completion are identified, and the specific analysis of these factors is conducted to summarize the reasons for the unfinished task.

Then, a new task pricing scheme is designed for the completed projects on the app and compared with the original scheme. From the first part of the analysis, whether in the pricing of tasks, the difficulty of completing

tasks, or in the reasons for unfinished tasks, the location of tasks always has a direct or indirect relationship with the pricing. In order to design a better pricing scheme, the new pricing scheme is designed based on the location relationship between tasks and members.

Finally, in the case of a concentrated task distribution, the pricing model when the task is packaged and released is considered and evaluated. It can be seen from the second part that the user pursues unit distance gain when completing the task. When the task distribution is more concentrated, even if the benefit of a single task is lower, there will be higher benefit per unit distance due to the task concentration. In the same way, when the task is distributed, the benefit per unit distance is lower even if the single task has a high profit. To further improve the model. The pricing model should be improved for task distribution.

## 2. The Establishment and Solution of the Model

### 2.1 The Task and Member Distribution are Clustered

There are many data processing methods, such as interval convergence processing, standardization processing, clustering processing, etc., and appropriate data processing methods can be selected according to specific data. This paper adopts *Z-score* standardization method to process data. The steps of *Z-score* standardization processing are as follows:

For the index value  $x_{ij} (i=1, 2, \dots, n; j=1, 2, \dots, m)$  of  $m$  items of  $n$  evaluated objects, the new standardized index value is

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{s_j}$$

Where,  $\bar{x}_j = \frac{1}{n} \sum_{i=1}^n x_{ij}$  is the average value of the  $J$ th index of  $n$  evaluation objects, and  $s_j = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2}$  is the mean variance of the  $J$ th index of  $n$  evaluated objects, that is,  $x_{ij}^* (i=1, 2, \dots, n; j=1, 2, \dots, m)$  is the unquantified index value.

The k-means method is used to cluster tasks and member distribution (Tingting Li, 2015). First, randomly choose  $K$  objects as the initial clustering center from the distribution coordinates and then loop. Then, the distance from each object to the central object is calculated according to the mean value of the central object, and the corresponding object is divided again according to the minimum distance. Finally, the mean value of each cluster is recalculated until the center of the cluster does not change, making the following formula minimal (Zhuo jinwu, 2014).

$$E = \sum_{j=1}^k \sum_{x_i \in w_j} \|x_i - w_j\|^2$$

Where  $m_i$  is the mean of the  $i$ th cluster and  $x_i$  is the variable in the  $i$ th cluster, so the distance between the variables in the cluster and the mean should be minimized. The clustering results are as follows:

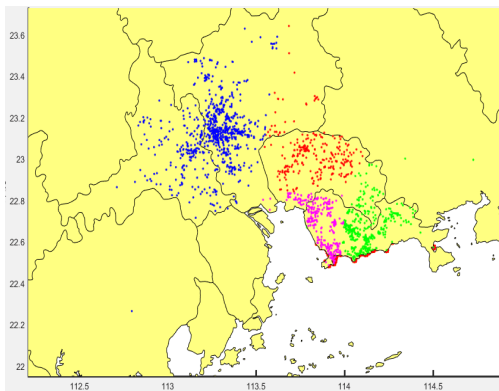


Figure 2.1. Membership profile

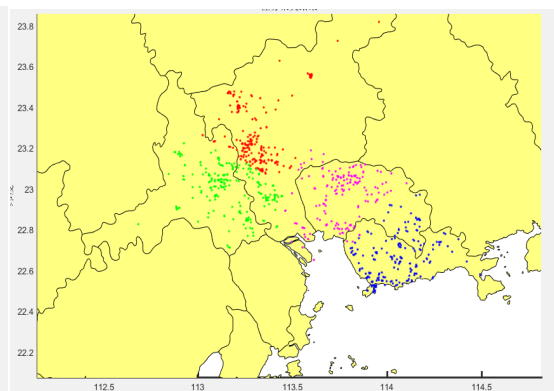


Figure 2.2. Task profile

### 2.2 Spatial Distribution of Task Pricing

With the help of surfer8.0 software, the spatial distribution of the task price in annex 1 is described, so as to obtain the spatial distribution diagram of the pricing of each task. The results are as follows:

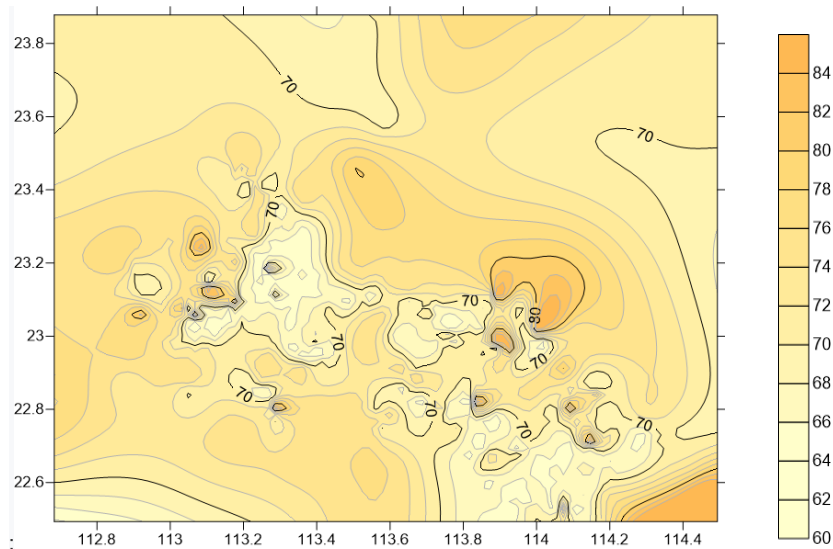


Figure 2.3. Task pricing diagram

In the figure above, most areas are relatively light in color, indicating that the area between 65 and 75 accounts for the majority of the task pricing, while the darker areas with higher pricing are mostly distributed in sheet form, and the task pricing near the darker areas is also relatively high, while the continuous areas are generally the same or similar in price.

From the actual analysis, the location of the task determines whether the platform members need to spend a certain amount of time and energy to arrive and complete, and the tasks that require members to spend more energy to complete are often located in remote locations, so the pricing of such tasks should be improved to motivate members to complete the task.

To sum up, there is a close relationship between task pricing and task location. Therefore, the longitude and latitude of the location coordinate level task point is used as the variable influencing the pricing.

### 2.3 Model of Task Pricing

According to the above analysis, the price of the task is closely related to the location of the task, so the longitude and latitude of the task point can be taken as the independent variable affecting the price.

In addition, the task price far away from the member location should be increased accordingly. When the price increase of a task does not match the distance increase, the task may not be completed. In order to motivate members to complete tasks, the price should be increased when the distance between tasks increases. Based on this, the p-d (price-distance) matching factor is defined as the third variable influencing the price. The matching factor is used to measure the matching degree of the change of task distance and the change of task list price. Only when the change degree of the two is consistent can the task be accepted and completed by members.

Based on k-means clustering analysis, the tasks in annex 1 are classified into four categories according to the coordinates (figure 2-1), and the average price of each type of task is calculated.

The ratio of the price and the clustering average price of each specific task to the price and the average price of all tasks is defined as the degree of price change, that is

$$P_C = \frac{P_i - \bar{P}}{\sum_{i=1}^{835} (P_i - \bar{P})}$$

$P_C$  is the degree of price change,  $P_i$  is the Price for task i,  $\bar{P}$  Is the average price of the category to which task item belongs.

Again using K - means method according to its location in the annex ii member into four classes (figure 2-2), clustering center of purple category is (22.91, 22.91), clustering center of the red category for (22.99, 22.99), blue category of clustering center for (23.16, 23.16), clustering center of the green category for (22.64, 114.07). The task location to the member center distance clustering and the ratio of the sum of all the tasks to the member center distance is defined as the distance change, due to the topic given data point location change is very small, You can almost view it as a plane, where the Euclidean distance is calculated to measure the distance between two points. That is

$$D_c = \frac{\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sum_{i=1}^{835} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}$$

The ratio between the degree of price change and the degree of distance change is defined as the p-d interaction factor, which is expressed by the formula

$$PD = \frac{P_c}{D_c} = \frac{\frac{P_i - \bar{P}}{\sum_{i=1}^{835} (P_i - \bar{P})}}{\frac{\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}{\sum_{i=1}^{835} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}} \quad (1)$$

To the degree of the degree of price change and the distance change using z - score method standardization, to eliminate the influence of the dimension, after processing of numerical plug (1) calculate the each task points matching values, based on the above analysis, define the task for the multivariate linear regression equation model is established, with task price as the dependent variable, the longitude, latitude, and the matching factor for the multiple linear regression equation of independent variables can be obtained as follows:

$$P = \beta_0 + \beta_1 * La + \beta_2 * Lo + \beta_3 * PD$$

Where  $La$  is the latitude of a task point,  $Lo$  is the longitude of the task point, and  $PD$  is the p-d interaction factor.

There are some mutant values in the p-d interaction factors. Although it may be reasonable to consider the occurrence of these values from the actual situation, it will affect the fitting effect of the model. Therefore, these singular data should be removed when fitting the equation (annex 2). In the previous 90% or 742 data, multiple regression was solved by matlab and the optimized regression equation can be obtained by using the regression toolbox in matlab

$$P = -199.5307 + 3.41324La + 1.6699Lo - 0.08189PD$$

#### 2.4 Model of the Factors Influencing the Complexity of the Task

Besides the distance of the task and the matching degree of the price, the difficulty of completing the task also affects the setting of the price. The harder the task is, the higher the price should be. According to the conclusion that the task is more difficult to complete when it is farther away from the member location, this paper measures the difficulty of completing the task by the distance between the task and the member center, and takes the difficulty as an influence factor to indicate the influence of the difficulty of the task on the pricing.

$$\text{The difficulty level of task } i = \frac{\text{Distance from cluster center for task } i}{\text{Average distance from all points to cluster center}}$$

In terms of the formula:

$$d_i = \frac{\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} * 835}{\sum_{i=1}^{835} \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}}$$

$d_i$  is the difficulty level of task I,  $\sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2}$  Is the distance from the cluster center for task i.

Under the condition of considering the impact of task difficulty on pricing, the model of pricing can be obtained as

$$P = (-199.5307 + 3.41324La + 1.6699Lo - 0.08189PD)d_i \quad (2)$$

The effect of equation (2) is tested with the remaining 10% data as above, and the average value of prediction error is 6.10% (annex 2). Compared with the model without the influence factor, the error value of the model is reduced to a certain degree by 6.94%. Therefore, it can be judged that the effect of the pricing rule is better after adding the influence factor of the difficulty of the task.

### 2.5 The Cause of the Unfinished Task

According to annex I, the project has a total of 834 tasks, among which 521 tasks have been completed, with a completion rate of 62.47%. There are a certain number of unfinished tasks. According to the above analysis, the location of a task has the greatest impact on the completion status. Secondly, the price level also determines whether a task is attractive. The matching degree of price-distance also affects the members' willingness to accept a task to some extent. Finally, the limit of the member's acceptance of the task and the credibility of the member will also affect the completion of the task. Therefore, the reasons for unfinished tasks are analyzed from the following four perspectives:

#### (1) Geographical location

The clustering results are drawn according to the distribution of completed and unfinished tasks. From the completion degree of different regions, it can be concluded that there is a great relationship between task completion and geographical location. At the same time, the coefficients of the two variables of longitude and latitude can be obtained from the regression equation are relatively large, so the pricing has a great relationship with longitude and latitude, namely the location, while the pricing can directly determine whether the members are willing to complete, so the geographic location has a great relationship with whether the task can be completed.

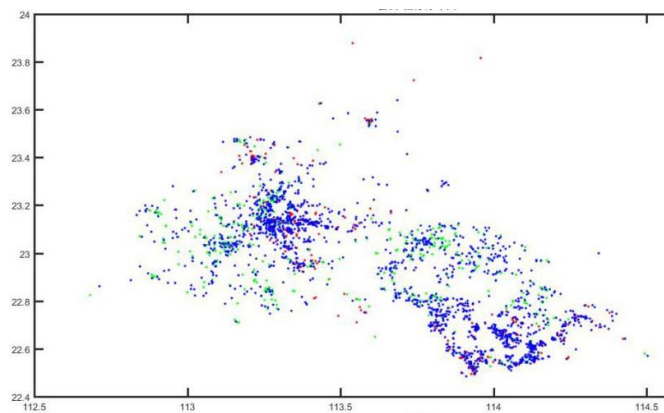


Figure 2.4. Task completion map

The red dots represent unfinished tasks, the green dots represent completed tasks, and the blue dots represent members' positions.

By common sense, the closer you get to a densely populated area, the easier it will be. Topic given the task of position in central guangdong province roughly, observe the task completion of the task distribution can be found that there are several regional degree is higher, respectively in longitude  $113.3^\circ \sim 113.6^\circ$  and  $113.6^\circ \sim 113.8^\circ$  and  $113.9^\circ$  and  $114.2^\circ$  range. Corresponding to the actual longitude, the region between these three longitude degrees is exactly three cities, namely guangzhou, dongguan and shenzhen. Therefore, the conclusion of higher completion degree of urban area can be obtained compared with other regions.

From a practical perspective, densely populated areas tend to be economically developed regions(Lina Wang & Aoran Xu, 2015), (Xiaorong Jiang & Shenglan Qang, 2017), which are more convenient to carry out various surveys and to reach the survey areas, so the number of people receiving tasks and the completion degree of

tasks are relatively high. Non-urban areas are relatively less complete due to the inconveniences in various aspects of the survey process. Therefore, the unfinished task may be caused by the unreasonable geographical location of the task.

#### (2) Reasons for price

Price is the most important factor to determine whether a task can be accepted and completed. The task is divided into two groups according to its completion, and the price tag of the two groups is analyzed for difference. Since the two groups of data are numerical data and there is no correlation, k-w test is adopted to analyze the difference between the two groups of data. The test results showed that P value was less than 0.005 and the original hypothesis was rejected, so there was a significant difference between the two groups of prices. Therefore, the unfinished task may be caused by unreasonable price setting.

#### (3) Reasons for price-distance match

According to the above analysis, when the task is located far away from the urban area, the difficulty will increase accordingly. At this point, the platform needs to increase the price of this task before anyone can accept it. When the distance of the task increases and the price does not increase accordingly, the task will be difficult to complete. Therefore, the price and location mismatch of the task will also make it difficult to complete. This factor is reflected in the P-D (price - distance) matching factor above.

#### (4) Reasons for membership

As can be seen from the attachment given in the title, members with high reputation in the platform are preferred to start the task selection, and the task quota is larger. Theoretically, people with high credibility should accept more tasks, but in fact, the platform stipulates the scheduled task limit for everyone. As a result, some tasks cannot reach people with high credibility, and finally, some tasks may not be completed.

According to the observation in annex 2, there are a large number of members with low reputation. Although members with high reputation can choose tasks first, due to the limitation of the scheduled task limit and the difference in number, there will still be many tasks that are selected by members with low reputation, resulting in the failure to complete the task (Huang limei, 2016).

The number and distribution of members will also lead to the low degree of task completion. In large cities like guangzhou and shenzhen, the number of members will far exceed the number of tasks, so there will be member competition for tasks and member resource waste with low reputation. In non-large cities such as dongguan, the number and distribution of members are well matched with the task, and there will be no waste or competition, which may be the reason why the task completion level in dongguan is much higher than that in guangdong and shenzhen.

### 3. Establishment of New Pricing Scheme

#### 3.1 Improvement Direction of the New Scheme

First, from the regression equation in question (1), it can be known that the pricing and location of the task have a direct relationship, and the p-d interaction factor also reflects the relationship between location and price. The influencing factors also measure the difficulty of the task from its position. Therefore, it can be judged that the main reason why the task cannot be completed is that the task is located at a far distance and is difficult to complete. Although the price has been improved to a certain extent, it is still not enough to motivate members to complete the task.

Therefore, from the perspective of task location, draw a circle with task coordinates as the center. According to the principle of nearby assignment, find the minimum radius to reasonably allocate each task. The longer the radius is, the more difficult it is to complete the task, so the price should be raised to make members accept the task. The task is priced according to the length of the radius.

#### 3.2 Establishment of Dynamic Partition Pricing Scheme

Build a matching model centered on the task location, and match each task with the member nearest to him according to the principle of "optimal proximity". The principle of "proximity optimization" is to assign tasks to members closest to the task point within a circle of radius<sup>1</sup>.

---

<sup>1</sup>The allocation of taxi resources in the era of "Internet +", 2015 excellent paper of Chinese university students' mathematical modeling contest.

Since the data obtained from the appendix of the subject is the longitude and latitude coordinates of the task point, and the radius is the plane radius when the circle is drawn, it is necessary to convert the longitude and latitude to the plane distance. By referring to the data, the following method can be used to calculate the distance between any two points on the surface of the map<sup>2</sup>

The Longitude and Latitude coordinates of point A are (LonA, LatA), and the Longitude and Latitude of point B are (LonB, LatB). According to the datum of Longitude of 0 degree Longitude, the value of Longitude of east Longitude, minus Longitude of west Longitude, 90-latitude of north Latitude and 90+Latitude of south Latitude are (MLonA, MLatA) and (MLonB, MLatB).Based on the triangulation, the following formula for calculating the distance between two points can be obtained:

$$K_0 = \sin(\text{MLatA}) * \sin(\text{MLatB}) * \cos(\text{MLotA} - \text{MLotB}) + \cos(\text{LatA}) * \cos(\text{MLatB})$$

$$D_0 = R_1 * \text{arc cos}(K_0) * \text{Pi} / 180$$

$D_0$  Is the spherical distance between two points,,  $K_0$  Is the radius of neighborhood,  $R_1$  has the same units as  $D_0$ ,

If we use 6371.004 kilometers as the radius, then  $D_0$  is in kilometers.

The distance between the two points mentioned above can be correspondingly replaced by the distance formula  $D_0(A, B)$  here. Thus, the radius of the neighborhood is a set of values.

Since the assignment is based on the proportion of the quota, So the data structure of the member is represented as  $V(j)(Lo(j), La(j), PM(j), j = 1, 2, \dots, 1877)$ ,  $V(j)$  represents the data structure of the  $j$ th member,  $Lo(j)$  is the Longitude for the  $j$ th member,  $La(j)$  is the latitude for the  $j$ th member,  $PM(j)$  is the percentage of the  $j$ th member's scheduled task limit in the total task limit.

$$PM(j) = \frac{M(j)}{\sum_{j=1}^{1877} M(j)}, \quad j = 1, 2, 3, \dots, 1877$$

Corresponding to the member's data structure, the data structure of the task is represented as  $T(i)(Lo(i), La(i), PP(i), i = 1, 2, \dots, 835)$ ,  $T(i)$  represents the data structure of task  $i$ ,  $PP(i)$  is the proportion of the price of task  $i$  in the total task price.

$$PP(i) = \frac{M(i)}{\sum_{i=1}^{835} M(i)}, \quad i = 1, 2, 3, \dots, 835$$

It can be predicted that the larger the radius is, the more difficult it is to find matching members, and a higher price should be defined for this task. According to the above formula of calculating spherical distance by longitude and latitude, the shortest distance between any two task points in annex I is 0.17km and the longest distance is 13.50km. Therefore, the shortest distance and the furthest distance are used as extreme values of the radius taken, and 40 radii are taken at intervals of 0.33km to test the minimum radius that each task point can match to the member one by one.

Table 3.1. The value of the radius and corresponding serial number

Number (k)	1	2	3	.....	39	40
Radius (km)	0.17	0.50	0.83		13.17	13.50

For  $K_0(k, k = 1, 2, \dots, 40)$  and then define the radius of the tasks  $i$ th  $Mat(i)$  matching degree, because of the task assignment is, in fact, according to the proportion of reserve quota when to, when the price of the  $i$ th a task

<sup>2</sup>The allocation of taxi resources in the era of "Internet +", 2015 excellent paper of Chinese university students' mathematical modeling contest.

in the price of all tasks in  $PP(i)$  is higher than the proportion of greater than or equal to the first  $j$ th a member of the limitation of scheduled tasks in addition limitation when the proportion of  $PM(j)$ , can match, the matching degree is defined as 1, otherwise the matching degree is defined as 1. Use the formula expressed by:

$$Mat(i, K_0) = \begin{cases} 1 & PP(i) < PM(j) \\ -1 & PP(i) \geq PM(j) \end{cases}$$

$Mat(i, K_0)$  is the match degree of the  $i$ th task when the radius is  $K_0(k, k = 1, 2, \dots, 40)$

According to the principle of "optimal proximity", the goal is to determine the price of a task point according to the minimum radius of a task point. Therefore, the programming model can be established with the minimum matching as the objective function, so as to obtain the minimum radius of each task point. The planning model is as follows:

$$k_{\min}(i) = \min_k \{Mat(i, k) = -1, k = 1, 2, \dots, 40\}, i = 1, 2, \dots, 835$$

$$S.t. \quad Mat(i, K_0) = \begin{cases} 1 & PP(i) < PM(j) \\ -1 & PP(i) \geq PM(j) \end{cases}$$

$k_{\min}(i)$  is the sequence number of the minimum radius taken for the task point  $i$ ,  $Mat(i, k)$  is the match degree of the first task point at the  $i$ -th radius.

With matlab, the minimum radius corresponding to each task point in annex 1 can be obtained. After analysis, the minimum distance is 0.17km, which should correspond to the minimum price of 65 yuan, and the maximum distance of 13.50km, which should correspond to the maximum price of 85 yuan. Because the farther the distance, the higher the price, so the distance and the price one-to-one correspondence, can get the most suitable price that should be set for each task point in annex I. Limited by space, only part of the result is shown below.

Table 3.2. Comparison of data indexes between the original scheme and the new scheme

Task number	latitude	longitude	Price		completion	
			The original plan	new plan	original price	new price
A0100	22.61177508	114.0380587	66.5	66.0	0	0
A0101	22.73006657	114.0666477	69.0	66.5	0	1
A0102	22.65235376	113.9299125	67.5	70.0	0	0
A0103	22.71365257	114.0652375	68.5	66.0	0	0
A0104	22.56073431	113.8900779	67.5	70.0	1	1
A0105	22.67343353	113.8257052	67.0	65.0	0	0
A0106	22.59155498	114.1259102	66.5	67.5	0	1

### 3.3 Comparison of Pricing Schemes

The pricing of each task under the new pricing scheme is calculated above, and the task completion degree under the new scheme should be compared with the task completion degree under the original pricing rules to analyze whether the new scheme has effectively improved the situation that the task cannot be completed. Use surfer8.0 software to make the price distribution diagram under the new pricing scheme:



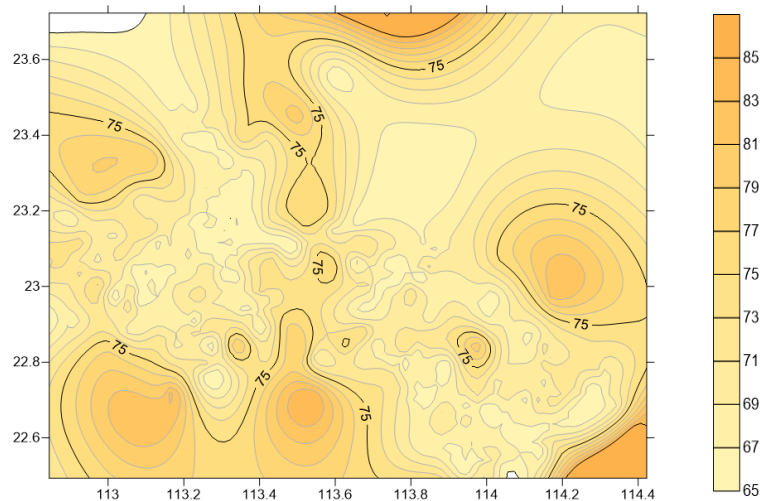


Figure 3.1. Pricing scenarios for different regional missions under the new programme

Compared with the original price distribution diagram (figure 2-3), it can be seen that the dark areas are obviously increased, so the task price under the new scheme is improved compared with the old one.

Although the price under the new plan has been increased, there will still be members who think the price is low and do not finish the task. Therefore, it is necessary to establish a price discrimination value. When the task price is higher than the discrimination value, the task can be completed; otherwise, the task cannot be completed. This paper measures task price with unit distance gain.

In fact, the differentiating value of the price can be interpreted as the minimum and minimum that members are willing to complete the task. For completed tasks, the value of this differentiation is the ratio of the gain to the distance of the task with the maximum distance; for unfinished tasks, this critical value is the ratio of the gain to the distance of the task with the minimum distance.

If this condition holds, the distinction between completed and unfinished tasks should be similar. The differentiated price is 6.3 yuan /km after calculation. The formula is as follows:

$$CP(i) = \begin{cases} 1 & P_r > P_{rm} \\ 0 & P_r < P_{rm} \end{cases}$$

$CP(i)$  is the completion status of task  $i$ ,  $P_r$  is the actual unit distance gain,  $P_{rm}$  is the differential value of the distance gain. When the actual task price is higher than the differentiated value, the task can be completed, which is denoted by "1"; otherwise, the task cannot be completed, which is denoted by "0".

This evaluation standard is used to compare the price of each task under the new pricing scheme with the price difference to obtain the degree of completion. A total of 835 missions were completed, with 565 completed tasks and a completion rate of 67.66%. Under the original pricing scheme, there were 835 tasks, of which 521 were completed, with a completion rate of 62.47%. Compared with the original pricing scheme, the completion degree of the new pricing scheme is increased by 8%, indicating that the new pricing scheme is effective.

#### 4. Task Packaging Model

It can be seen from the above analysis that whether a member can complete the task depends largely on the income per unit distance when he completes the task. The more concentrated task distribution enables members to complete more tasks in one area, thus the benefit per unit distance will be improved; The distributed task distribution requires members to go to different regions to complete the task, which will reduce the income per unit distance of members, thus causing users to scramble to choose the tasks distributed together. To further optimize the model, we need to adjust the distributed intensive and dispersed tasks on the price, and minimize the impact of the location distribution on the task completion.

##### 4.1 Task Packaging Model Based on "Small World"

It can be seen from the observation data that the amount of data given in the appendix of the question is large, and it is difficult to determine the packaging model directly from the density of original data analysis. After observation analysis, as to make the task of distribution of middle latitude for  $23^\circ \sim 23.2^\circ$ , longitude  $113.2^\circ \sim$

113.4 ° area more reasonable distribution of tasks and membership on this article selects the area as a "small world", it is a small world and realistic factors, just model is relatively simple, easy to analysis. You can simulate realistic task packaging by analyzing small world models. In the small world, the grid is used to divide the small world into 5\*5 regions. The results are as follows:

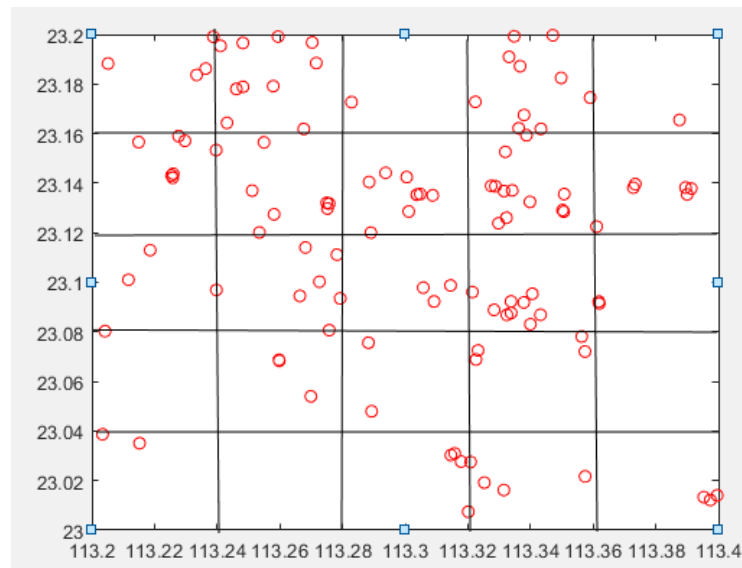


Figure 4.1. The regional division diagram of the "small world" model

Tasks in a grid are relatively concentrated and can be packaged, and tasks in a grid can be considered as a package. When not packaged, the task is independent, task of data structure for price} {the longitude, latitude, task, when in the grid data is packaged as a "package", after the package has become a task points set, the package of data structure is defined as {latitude and longitude, the package price, package contains singular}, parcel is belong to latitude and longitude of the center of the grid longitude and latitude.(see annex iv for the specific values of each package data structure in the small world). In this way, each package is equivalent to a task point, so that we can consider the pricing model of combining tasks and distributing them together.

#### 4.2 The Creation of a Price Correction Function

According to the above analysis, the packages with a relatively concentrated task distribution will lead to the selection of users, while the willingness of parcel users with a relatively dispersed task is not very strong. Considering this reason, the task is more dispersed task completion after packaging will become low, in order to reduce the effects of task distribution density of the results of packaging, need to adjust the price, different package for intensive task distribution to adjust prices, making as much as possible and complete the package can be member to choose.

As the title suggests, in principle, the more creditworthy a member is, the higher the priority of starting a task, and the higher the priority. High-priority members will give priority to tasks with high income per unit distance. As a result of delayed selection, members with low priority are often unable to select some tasks with high income. Over time, this phenomenon will lead to higher priority members with higher priority and less priority members will not be able to receive tasks.

Based on the above considerations, this paper defines a price correction factor to adjust the price.

Each grid represents a unit area. The more tasks within the unit area, the denser the task distribution within the package, and the less tasks within the unit area, the more dispersed the task distribution within the package. Under normal circumstances, the price of the package should be the sum of the price of each task within the package. In order to reduce the effect of task distribution density, the price of package should be reduced to reduce the choice willingness of members. For packages with distributed tasks, the price should be increased to encourage more members to choose.

The above idea of price correction is expressed in the mathematical equation, that is, when the number of tasks contained in the package is small, the price of the package should be greater than the value added directly to the price; when the number of tasks contained in the package is large, the price of the package will be less than the value added directly to the price.

According to the data structure wrapped in the small world, the function image is fitted as follows:

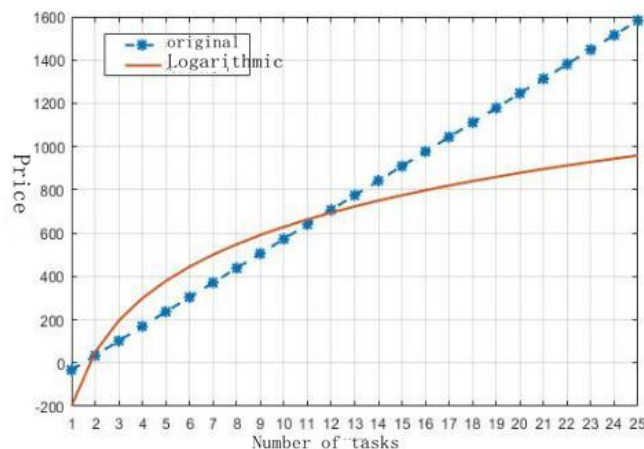


Figure 4.2. A fitting diagram of the package data

As shown in the figure, if the singular number contained in the package is taken as the independent variable and the package price obtained by simple summation is the dependent variable, the blue line in the figure above can be obtained. Based on the above considerations, it can be seen from the mathematical knowledge that the growth rate of the logarithm function is decreasing with the independent variable, so the logarithm function is adopted to correct the price. By fitting the logarithmic function with data, the logarithm formula is  $y = \ln x - 200$ . It can be seen from the curve that when the singular number of the package is below 12, the revised price of the package with less tasks increases to a certain extent, which will improve the hospital where members complete the package with less tasks. For packages with more than 12 tasks, the revised price will be reduced to a certain extent, which will reduce the number of members competing for such packages. Therefore, theoretically, the new plan effectively considers the impact of packaging on pricing and further optimizes the pricing model.

#### 4.3 Model Comparison

Theoretically, the packaged pricing model takes into account more comprehensive factors, making the model more optimized. It is also necessary to analyze the effects of the model based on actual data.

Using matlab to solve the pricing under the new scheme, the package is equivalent to a single task in the previous problem. Based on the second part, the completion of the task is measured based on the idea of distinguishing values, and the change of completion degree is analyzed. The calculated completion rate of the task under this scheme is 72%, which is 4.34% higher than the original completion rate, indicating that the pricing scheme considering the packaging situation is better.

### 5. Evaluation and Improvement of the Model

#### 5.1 Advantages of the Model

- (1) In the first part, the p-d interaction factor and the difficulty of the task were added to consider the impact on price, making the perspective more comprehensive.
- (2) In the second part, the measurement index to evaluate the matching degree of the task is given, and the task is priced according to the measurement index, so the pricing is more accurate.
- (3) In the third part, the analysis of the general situation of package pricing through the small-world model not only simplifies the analysis process, but also is representative.

#### 5.2 Disadvantages of the Model

- (1) Due to the small amount of data given in the annex, the impact of some factors cannot be specifically considered, leading to the inability to consider the factors affecting pricing more comprehensively.
- (2) In practice, there may be some influences of personal preference and some random factors when members complete tasks, so there may be some deviations in the results obtained.

(3) In order to simplify the problem, when describing the correlation between factors, more empirical formulas were used instead of more in-depth analysis of the correlation between factors.

### 5.3 Improvement Direction of the Model

Given fewer data angles, more specific data from different perspectives should be collected when possible, such as members' personal preference, members' free time and the correlation between various factors, etc., so as to get a more accurate pricing scheme by pricing tasks from multiple perspectives.

## 6. Conclusion

In this paper, the price function is fitted by means of multiple stepwise regression, and the reasons for the unfinished task are analyzed, including the difficulty level of task completion, distance, the relationship between task quantity and population density and so on. Then, the model of matching task and member is established by using the idea of "optimal neighbor". Then, starting from the "small world" model, the actual packaging situation was simulated, and the price correction function was defined to reduce the impact of task distribution density on the completion situation, and the pricing scheme was further optimized. The train of thought to provide the service provides the reference, the company pricing on the task, should be mainly based on to the difficulty of the task, and the distance of the city, and population density on pricing, will be closer at the same time, ease similar tasks package, which can be profitable, and can improve the completion of tasks.

## References

- Calculate the distance according to the longitude and latitude of two points.  
<http://www.cnblogs.com/yicsfwhh/archive/2010/12/20/1911232.html>.2012-12-20
- Huang, L. M. (2016). Research on user credit and message pushing algorithm based on crowdsourcing platform [D]. Guangdong normal university of technology, 2016.
- Lina, W., & Aoran, X. (2015). University students' employment to second-tier cities transfer phenomena research - in chongqing, for example [J]. *Journal of the Chinese Market*, 26, 113-121.
- Liu, S. (2011). Application research of professional services crowdsourcing model [D]. Beijing university of technology, 2011. <https://doi.org/10.1190/978-1-56080-284-6.15>
- The allocation of taxi resources in the era of "Internet +", 2015 excellent paper of Chinese university students' mathematical modeling contest.
- Tingting, L. (2015). Research on improving k-means clustering algorithm [D]. Anhui university, 2015.
- Xiaorong, J., Shenglan, Q., & Yongchun, Y. (2017). Research on urban population flow network in China -- based on baidu LBS big data analysis [J]. *Population and Development*, 23(1), 13-23
- Zhuo, J. W. (2014). Application of MATLAB in mathematical modeling. Beijing: Beijing university of aeronautics and astronautics press, 2014.

## Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).