ISSN: 2688-9323

# Photo Money App Pricing Plan Revision Model

Lidong He[a], Zonghai Li[b], Wenao Ma[c], Mingyu Su[d], Zaiyin Ma[e]

School of Metallurgy, Northeastern University, Liaoning, China

[a]20192463@stu.neu.edu.cn, [b]20192507@stu.neu.edu.cn, [c]20192473@stu.neu.edu.cn,

[d]20192604@stu.neu.edu.cn, [e]20192661@stu.neu.edu.cn

## Abstract

**In this paper, by discovering the existing pricing law of photo money app, analyzing the reasons why photo tasks are not completed, and improving the reasons, we come up with a new pricing scheme with better comprehensive performance. After considering the packaging of tasks, the tasks are completed better. Finally, the new pricing scheme is applied to the uncompleted data and a higher completion rate is obtained. the model developed in this paper is mainly based on the pricing rationalization problem of the app "Photo Money". In fact, the model is applicable to the pricing scheme design in crowdsourcing economy, and can also be applied to many other similar scenarios such as Meituan take-out and DDT taxi, which have a wide application prospect.**

## Keywords

**K-means Multiple Linear Regression; Simulation; Hierarchical Clustering; Crowdsourcing Pricing Scheme.**

## 1. Background

In the Internet era, the crowd-sourcing economy has emerged, and the self-service mode represented by "Make Money by Taking Pictures" has flourished. Users become members of the "Photo Money" APP, receive and complete photo tasks, you can earn task fees. Companies use such platforms to complete business checks and information collection, not only to reduce the cost of investigation, but also has the advantages of real data, short survey cycle.

## 2. Problem Analysis

The rapid development of the Internet industry has given birth to many new things, including "photo money" this self-service model has become a fresh choice for many people to earn income in their spare time. Users register as members on the corresponding APP, receive and complete tasks, and provide companies with a lot of real and reliable survey data. The study of reasonable pricing of tasks in such APPs is a very meaningful issue, too high pricing does not improve the quality of results will only increase costs, while too low pricing is difficult to attract members and thus affect the completion rate of tasks. At the same time, the existing pricing mechanism is mostly set by enterprises and crowdsourcing platforms individually or collaboratively, with little consideration of members' income. Although such a pricing model can maximize the interests of enterprises and platforms in the short term, it will cause negative impacts such as decreasing members' experience and losing a large number of users in the long run. Therefore, the main goal of this paper is to fully analyze the data related to completed tasks (task coordinates, task pricing, and completion status) to extract useful variables and restore the pricing pattern of the platform. Then, other key factors affecting pricing are introduced into the model to design a new pricing scheme.

## 3. K-means Method for Clustering Task Points

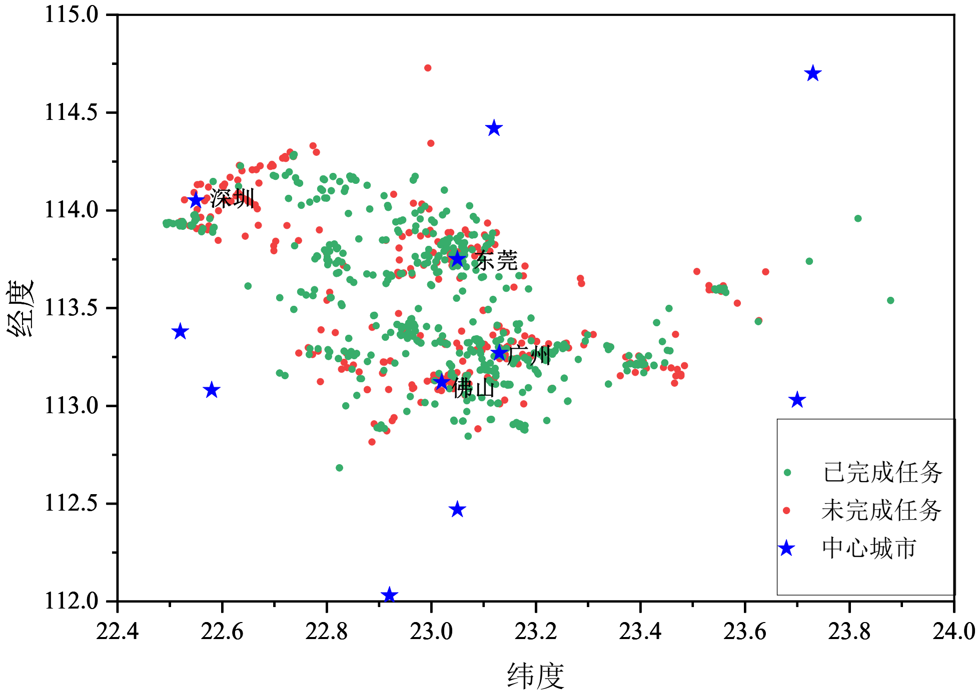


**Figure 1.** Task completion and distribution of prefecture-level city centers

As can be seen visually from Figure 1, the distribution of the tasks all have obvious aggregation characteristics. Combined with the analysis of the map of the region, it is not difficult to find that the tasks are mainly distributed within the four cities of Shenzhen, Dongguan, Guangzhou and Foshan. [1] According to experience, the closer the tasks are to the cities, the denser they are, and the further they are from the cities, the sparser they are. Therefore, we set the number of clusters$k$ set to 4, and using K-means method, the task points can be clustered into 4 classes based on their geographical locations, i.e., longitude and latitude, as shown in Figure 2.

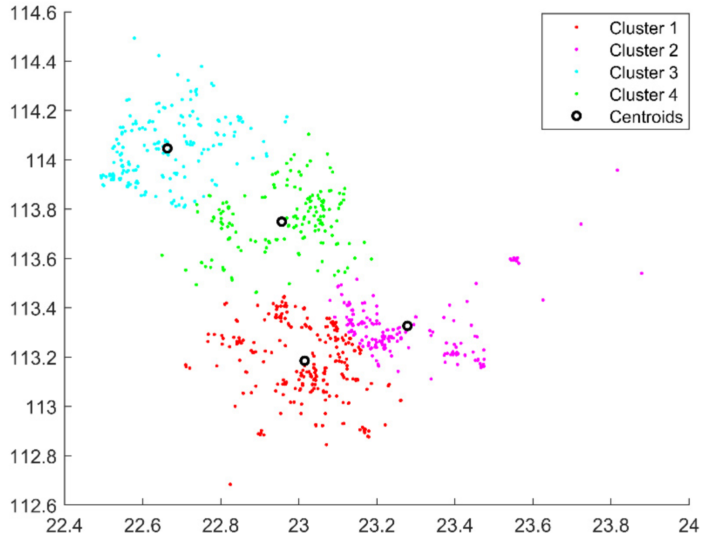


**Figure 2.** Distribution of tasks clustered by latitude and longitude

The latitude and longitude coordinates of the clustering centers are (23.015°N, 113.185°E), (23.278°N, 113.327°E), (22.663°N, 114.046°E) and (22.956°N, 113.749°E), corresponding to

Foshan (23.02°N, 113.12°E), Guangzhou (23.13°N, 113.27°E), Shenzhen (22.55°N, 114.05°E) and Dongguan (23.05°N, 113.75°E), respectively. 23.13°N, 113.27°E), Shenzhen (22.55°N, 114.05°E) and Dongguan (23.05°N, 113.75°E), respectively.

The similarity of the location coordinates of the clustering center and the city center verifies that the tasks are clustered and distributed in four cities, Shenzhen, Dongguan, Guangzhou and Foshan.

## 4. Multiple Linear Regression Analysis of Pricing Laws

According to the formation mechanism of equilibrium price of goods in microeconomics, when pricing tasks, demand and supply should be considered first. The platform releases tasks as demand, while members accept and complete tasks as supply. When the supply of members around the task is too much, it will cause the market supply to exceed the demand and lead to the decrease of the task price. [2]

Further analysis shows that supply capacity is not only related to the number of members, but also to the quality of members. In the context of crowdsourcing economy, membership quality can be expressed by the reputation value provided by the platform.

Analyzed from the user's point of view, if the task location is isolated and far from the central location, then the cost of the user to complete the task is higher and the user's willingness to want to go to complete the task is lower, i.e., the supply will be reduced and the price should be raised accordingly.

Summing up the above three analyses, we can get that there are three indicators that may affect the task pricing, which are: the distance from the task point to the cluster center, the number of members around the task point and the average reputation of members around the task point. [3]

Establish the multiple linear regression equation.

$$S = \varepsilon + \omega_1 d + \omega_2 n + \omega_3 c \tag{1}$$

where S is the pricing of the task and$d$ task point is the Euclidean distance to the corresponding clustering center, and$n$ is the number of members within 10 km of the task point as the center, and$c$ is the average of the member reputation within 10 km of the task point.

The regression solution and significance test were performed with Stata software, and the results obtained are shown in Table 1.

**Table 1.** Multiple linear regression solution results

| Assignment price tag | Coef. | Robust Std. Err. | t | P>\|t\| | [95% Conf.] | |
|---|---|---|---|---|---|---|
| Distance to cluster center $d$ | 0.079518 | 0.017452 | 4.56 | 0 | 0.045262 | 0.113774 |
| Number of surrounding members $n$ | -0.021236 | 0.001487 | -14.27 | 0 | -0.024156 | -0.018315 |
| Average reputation of surrounding members$c$ | 0.00144 | 0.000422 | -3.42 | 0.001 | -0.002276 | -0.000617 |
| Constants $\varepsilon$ | 70.12636 | 0.459399 | 152.65 | 0 | 69.22464 | 71.02809 |

From Table 1, it can be seen that$d$ , the $n$ and $c$ are all significant less than 0. 01, which verifies the rationality of the multiple linear regression model. As a result, the multiple regression analytic equation was obtained as follows.

$$S_0 = 70.126 + 0.080d - 0.021n - 0.001c \tag{2}$$

ISSN: 2688-9323

Where $S_0$ denotes the existing program pricing. With equation (2), the pricing law can be derived as follows: the pricing is based on 70.126, and three factors are considered based on this, which are the distance from the task point to the center of the cluster, the number of members around the task point and the average reputation of members around the task point, and the distance from the task point to the center of the cluster is an indicator to increase the pricing, and the number of members around the task point and the average reputation of members around the task point are indicators to decrease the pricing.

## 5. Analysis of the Reasons Why the Task was not Completed

1) Level of economic development.

Using the GDP per capita to measure the economic level of the four cities, we can get that Dongguan's GDP per capita is only 48.29% of Shenzhen's, but the average pricing is 4.67% higher than Shenzhen's. This leads to the conclusion that for cities with higher economic levels, the reason for their tasks not being completed is that the pricing is too low resulting in low attractiveness to members who are not inclined to take tasks issued by the platform.

2) Remote location.

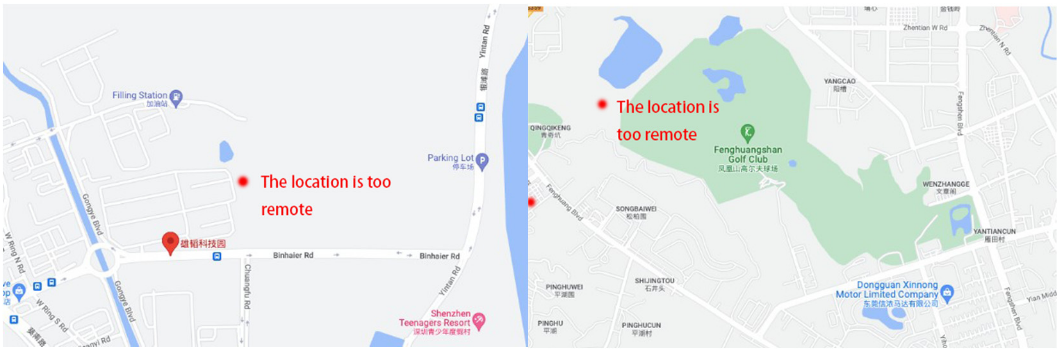


**Figure 3.** Some of the geographically remote mission sites

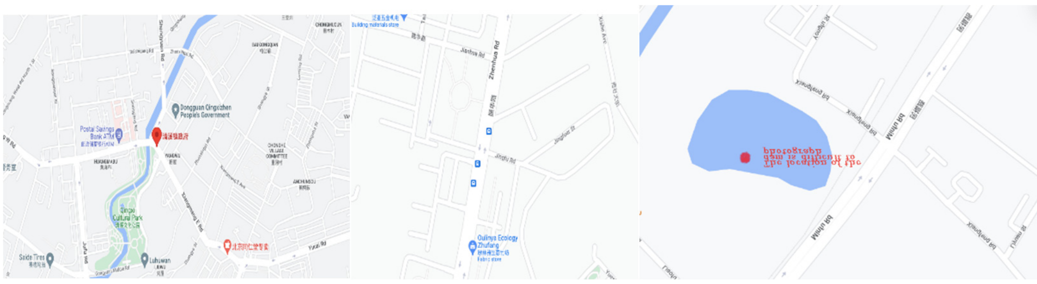

**Figure 4.** Some of the hard-to-film mission sites

Statistically, the task completion rate outside the urban areas of the three cities has decreased to a certain extent compared with the obtained task completion degree of the three cities citywide. Geographic remoteness has a significant impact on task completion. This leads to the conclusion that geographical location remoteness is one of the important factors for task non-completion. In addition, we found 10 schools, reservoirs, and fire departments through the satellite map survey of 0 tasks, all of which are difficult to access. There are also 10 training grounds, charging plants, industrial parks, etc. For the simple completion of the photo task, these places are difficult to complete the photo task for a variety of reasons, or the location itself is difficult to access. Analysis there are 16 parks, scenic spots class tasks, 17 from the satellite map to observe in the city within the green task is not completed may be the reason for such

ISSN: 2688-9323

places photo task is more difficult, involving certain professional related knowledge, so members also do not tend to accept such tasks.

3) Credibility value and quota issues.

It is calculated that the number of unfinished tasks in Guangzhou and Foshan City around members with too low quotas both account for about 50% of the total number of unfinished tasks. This effect leads to the conclusion that the improper distribution of creditworthiness value and quota is one of the important factors for the uncompleted tasks.

## 6. Adjusting the Weight of Price Influencing Factors

From the multiple linear regression model of 5.1.2, the distance from the task point to the center of the cluster, the number of members around the task point and the average reputation of members around the task point are considered comprehensively. Therefore, in order to ensure the improvement of the task completion rate while minimizing the platform's expenses, the weights of the distance from the task point to the cluster center and the average reputation of the members around the task point are raised and the weights of the number of members around the task point are lowered, and the pricing equation after adjusting the weights is as follows.

$$S = 70.126 + 0.090d - 0.010n - 0.010c \tag{3}$$

Based on the pricing criteria after adjusting the weights, we consider the economic level factors of each city and integrate the total number of members and the cost of completing tasks for members in each city to reflect the differences of each city. We use the price transfer method to correct the base pricing by considering the above 3 factors, averaging the standardized values of the 3 factors, and multiplying them by the adjustment factor as the city$i$ the price transfer function of.

$$\sigma_i = \frac{\mu}{3}(\alpha_i + \beta_i + \gamma_i) \tag{4}$$

Among them$\sigma_i$ is the city$i$ the price transfer function of the city, and$\mu$ is the adjustment factor, and$\alpha_i$ ,$\beta_i$ and$\gamma_i$ are the city's$i$ s gross product, i.e., GDP, membership, and per capita wages normalized.

In summary, the formula for calculating the pricing of the new pricing scheme is as follows.

$$S_1 = 70.126 + 0.090d - 0.010n - 0.010c + {\sigma_i}' \tag{5}$$

Where $S_1$ indicates new program pricing.${\sigma_i}'$ is the city$i$ the normalized price transfer function, i. e. $\sigma_i$ mapped to the interval [-1,1] to obtain${\sigma_i}'$ .

In order to compare the differences in the effects of the old and new pricing schemes, we conducted simulations of member selection tasks and compared the platform costs and task completion under different pricing schemes.[4]

It is assumed that members with high reputation pick more tasks on average per day. It may be assumed that: users with a reputation value less than or equal to 2 are inactive and do not participate in task selection; users with a reputation value$x$ is greater than 100 and the scheduled task limit$l$ Members with a reputation value greater than or equal to 5 select 5 tasks per day; the rest of the members select 1 task per day.

Number of member selection tasks$t$ The calculation formula is as follows.

ISSN: 2688-9323

$$t(x,l)=\begin{cases}5, & x>100\ and\ l\geq 5\\ 0, & x\leq 2\\ 1, & other\end{cases} \tag{6}$$

The process for members to select tasks is as follows.

1. each member selects the task with priority given to the 10 tasks closest to him/her, in descending order of distance, and if all 10 tasks are selected, other tasks with a distance of less than 30km will be considered.

2. on the premise of 1, when the nearest 10 tasks have all been selected, the selection is made in descending order of price.

3. After choosing a task, members will compare the task price with their psychological expectation. If it is less than the psychological expectation value, you will not take the task; if it is greater than or equal to the psychological expectation value, you will take the task.

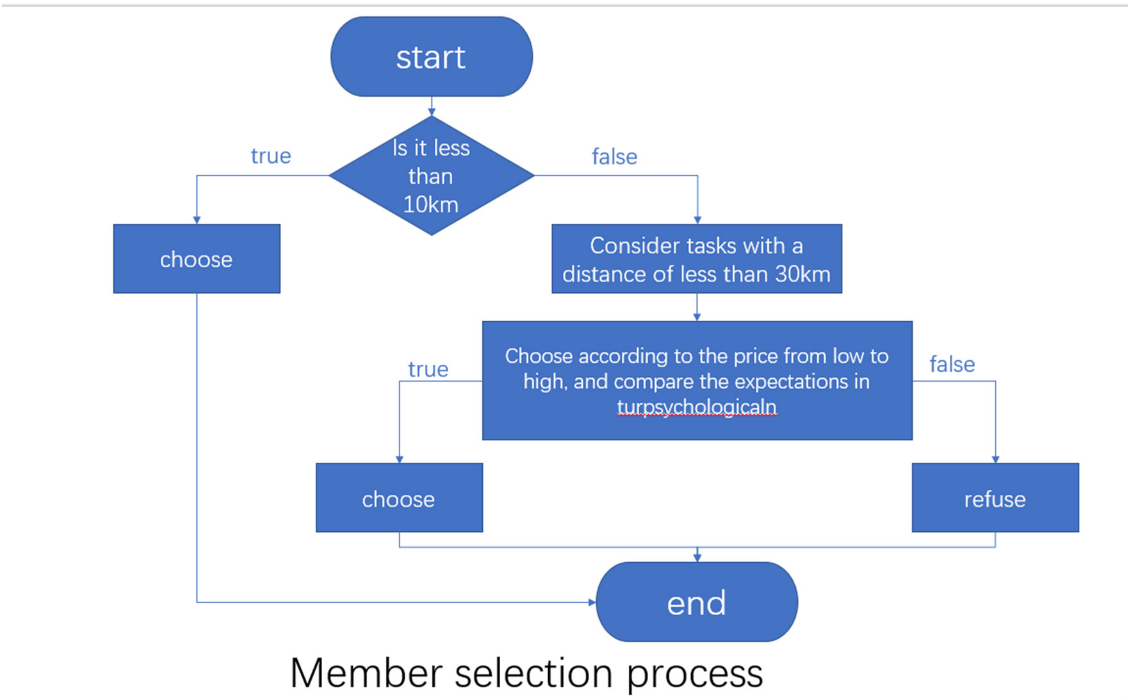


**Figure 5.** The process of member selection task

Assuming a 22-day work month and a 7-h work day, the number of hours worked in a month is 154 h. The monthly per capita wage in the city is$\gamma$ then the earnings per hour of work$c$ is calculated by the following formula

$$c=\frac{\gamma}{154} \tag{7}$$

Member psychological expectation value$\rho$ The calculation formula is as follows.

$$\rho=p_0+c*(t*n+d*\vartheta) \tag{8}$$

Of which$p_0$ is the base pricing, and$c$ is the revenue per hour worked, and$t$ is the average time required to complete a task, and$n$ is the number of tasks, and$d$ is the distance from the member to the task point, and$\vartheta$ is the factor that converts distance to time.

let $p_0 = 45$ ,$t = 0.5$ , and$n = 1$, and $\vartheta = 0.005$, simulations are performed and the simulation results are obtained as follows.

**Table 2.** Comparison of Simulation Results and Analysis of Old and New Pricing Schemes

| Region | Original actual completion rate | Original simulation completion rate | Simulation error | New Simulation Completion Rate | Relative change | Absolute change |
|---|---|---|---|---|---|---|
| Shenzhen | 36.22% | 31.63% | | 76.53% | | |
| Foshan | 67.16% | 91.04% | | 92.37% | | |
| Guangzhou | 53.54% | 32.32% | | 74.35% | | |
| Dongguan | 95.38% | 91.91% | | 86.71% | | |
| Overall | 62.51% | 63.35% | | 82.52% | | |

## 7. Hierarchical Clustering Approach for Task Packing

In order to pack the densely distributed and mutually neighboring task clusters, we select the hierarchical clustering algorithm, define the inter-class distance as the minimum distance between two classes of data, and set the termination condition of clustering as the minimum class spacing over a threshold value of 3 km. we convert the latitude and longitude data of task points into right-angle coordinates in kilometers, so the Euclidean distance is selected to measure the distance between different classes (similarity). The packing process of neighboring task groups is completed by continuously clustering the task points that satisfy the conditions.

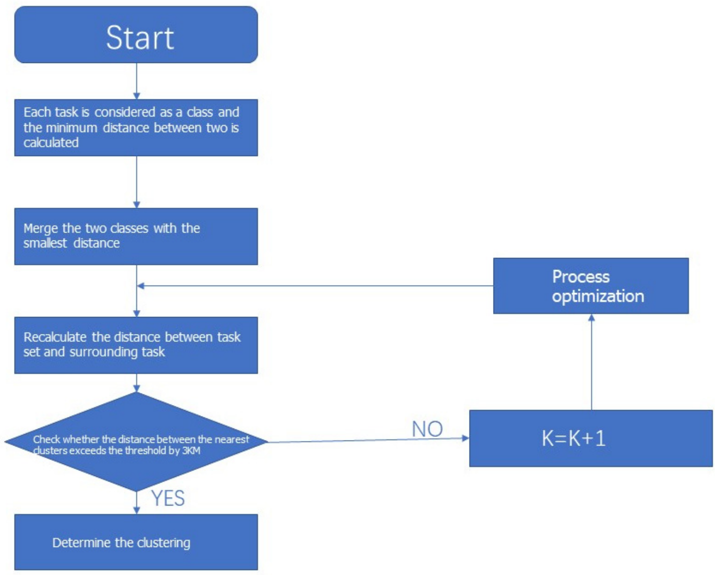


**Figure 6.** Flow chart of hierarchical clustering algorithm

After the hierarchical clustering of all completed projects, the total number of classifications was obtained as 429, including 175 task packages and 254 unpackaged tasks, and the maximum number of tasks within the task packages was 15.

ISSN: 2688-9323

## 8. Pricing Correction Model

The farther the distance of the member to the task, the more the number of tasks, the more the cost of the member to complete the task. Therefore, task packing can save members' cost. The more the number of tasks contained in the packaged task set, the more the member gains in distance, and accordingly, the more the price of the tasks in the packaged task set can be reduced.

From the point of view of supply and demand, the more members around the packaged task set, the more abundant the supply, and therefore, the more the price of the packaged task set tasks can be reduced accordingly.

Based on the above analysis, the creation of packaged task sets$i$ the degree of price reduction$\delta_i$ , the number of members around the task set$n$ and the number of tasks contained in the task set$m$ The functional relationship between

$$\delta_i(n) = \begin{cases} 0, & n = 1,2 \\ 10\%, & n = 3,4,5 \\ (n+m)\%, & n > 5 \end{cases} \tag{9}$$

Set packaged task set$i$ the pricing of$S_i$ calculated as follows.

$$S_i = (1-\delta_i)\sum_{j=1}^{n} s_{ij} \tag{10}$$

Where $s_{ij}$ is the set of packaged tasks$i$ of tasks $j$ is the pricing of the tasks in the packaged task set.

From equation (9), the set of packed tasks$i$ the reduced pricing is $\delta_i \sum_{j=1}^{n} s_{ij}$ . These savings are now distributed to the 10 unpacked tasks that are closest to the packed task set. Because the further away from the center of the packed task set, the less the task is driven by the packed task set, it needs to be made more attractive by compensating more prices to improve the overall completion rate. the 10 tasks closest to the packed task set$i$ Of the nearest unpacked tasks, the task $j$ The distribution ratio of $q_{ij}$ are.

$$q_{ij} = \frac{u_j}{\sum_{j=1}^{n} u_j} \tag{11}$$

Where $u_j$ is the distance of the jth task of these 10 task points from the center of the task set

In summary, the pricing model after the packaging correction is

If a task is packaged for assignment, the task pricing for the task set it is in is

$$S_b = (1-\delta)\sum_{j=1}^{n} S_{1j} \tag{12}$$

where $S_b$ is the pricing of the packaged task set, and$\delta$ is the degree of price reduction for the task set, and $S_{1j}$ is the task set task$j$ of the new program pricing.

If the task is not packaged for assignment, it is priced at

$$S_c = S_1 + \sum_{i=1}^{N} (q_i \delta_i \sum_{j=1}^{n} s_{ij}) \tag{13}$$

Where $S_c$ is the pricing of unpacked tasks, the $S_1$ is the pricing of the new solution for the task, and $\delta_i \sum_{j=1}^{n} s_{ij}$ is the set of tasks$i$ is the pricing for the reduced set of tasks, and $q_i$ is the set of tasks$i$ the proportion of the assignment to the task, the$n$ is the number of tasks in the task set$i$ the number of tasks contained in the task set, and$N$ is the number of packed task sets that generate assignments for the task.

## References

[1] Google GPS satellite positioning [EB/OL]. [2018-04-01]. http: / /www. gpsspg. com/maps. html.

[2] Zheng Jun, Lin Manjia, Hu Rong. Research on task pricing model in spatial crowdsourcing environment [J]. Economic Mathematics,2020,37(02):104-110.

[3] Pudong P, Fan Chongjun, Yuan Guanghui, Yang Yunpeng. Elite swarm algorithm and crowdsourcing pricing model considering stakeholders [J]. Computer Application Research, 2019, 36(04): 1039-1044.

[4] Lu, Bing'er, Zhu, Chen-Chen, Tian, Yuan-Ming, Wang, Zhi-Yong. A revised model of pricing scheme for photo money app [J]. Experimental Science and Technology, 2019, 17(01): 1-5+11.