

Optimal Design of Crowdsourcing Project Pricing

Lizhi Yu¹, Dongkai Chen¹, Huangding Zhu², Zhe Xuan² and Xusheng Kang^{1,*}

School of computer and computer science, Zhejiang University City College, Hangzhou 310015, PRC

School of engineering, Zhejiang University City College, Hangzhou 310015, PRC

*Corresponding author

Abstract—Aiming at the problem of the pricing and completion of crowdsourcing project, we set up a project task pricing design model and solve this model by using genetic algorithm and logistic regression. Meanwhile, we use the differential equation to improve this model.

The project task pricing design model is a multi-objective nonlinear programming model. We introduce the concept of company profit to limit the price and regard the profit as our target function. In addition, we consider the members' credit value, the distance between the members to the mission point, the number of members within the mission area and the possibility of members performing tasks to limit target function. In consideration of the possibility of the member performing the task, we used the logistic method to fit it. When considering the integrated reputation of members, we use the entropy weight method to determine the coefficients of each index. Finally, we get the target function values in different cases by using Genetic Algorithm. Meanwhile, we use the differential equation method to set up the price floating model to optimize our existing model.

Through the model we established, we can ensure that the tasks issued by the company can be carried out with a higher completion rate in practice, and obtain better profit.

Keywords-component; crowdsourcing; multi-objective nonlinear programming; Entropy method; Genetic Algorithm ; Logistic regression

I. INTRODUCTION

With the popularity of mobile internet and O2O(Online to Offline) business model, crowdsourcing has become a new e-commerce model[1]. Various types of crowdsourcing platform are closely related to people's lives. Crowdsourcing is an open innovation, and its success depends on platform pricing. It is inevitable that some tasks will be undone if the price of these tasks is too low. Thus, it's essential to attach importance to making a proper plan on pricing.

Nowadays the platform pricing often relates to some factors such as the number of people around the task, people's credit value and the price that people are able to accept. Crowdsourcing companies take various factors into consideration. There are still some tasks which are not accepted. We use a nonlinear programming model which is based on Genetic Algorithm hope to mine the data and formulate a better platform pricing solution.

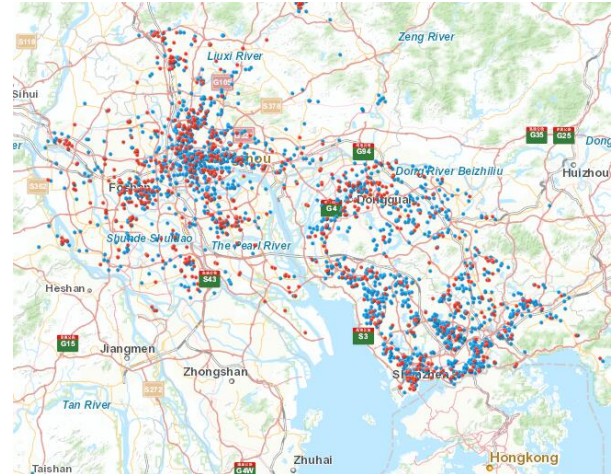


FIGURE I. THE GEOGRAPHICAL MAP ON THE LOCATION OF MEMBERS AND TASKS

As shown in the figure above, we map the data to the actual map one by one, using two colors to represent the task publishing location and the member location information respectively. The red dots in the figure represent the geographic location where the task is posted and the gray dots refer to the geographical location of the corresponding member.

II. PRICE PLANNING MODEL

A. Main Idea about the Model

We have the data including the location of each mission (Latitude and longitude coordinates), the location of each member (Latitude and longitude coordinates), the price that each member could accept, the reputation value of each member and the scheduled task start time. We want to maximize the benefits of crowdsourcing companies. Thus, we establish a revenue objective function. Then we take the above factors into consideration to limit the result. Finally, we use Genetic Algorithm to solve this problem[2,3].

B. Develop a Target Return Function

$$\begin{cases} \min \sum p_i x_i \\ \max \sum p_i \end{cases} \quad (1)$$

We first consider that the pricing model in this problem ought to belong to the multi-objective programming. When we analyze the formula (1), p_i is the probability of each task execution, x is the price under the condition of each probability, it is clear we need to increase the probability that each task will be executed as much as possible, while minimizing the total cost of the company as much as possible, allowing the company to get better earning. Then we transform multi-objective programming model into single-objective programming model in order to solve this problem better.

$$\min a \sum p_i x_i - b \sum p_i \tag{2}$$

$$a + b = 1 \tag{3}$$

In this single-goal planning objective function, and x have the same meaning as above. While a and b , both of them are coefficients of this equation which are set manually in order to find the optimal result. According to the above equation, we find that if we want to get a high profit, we ought to increase the probability as much as possible, in the meanwhile maintain the price on a proper level. If the price of each task is too high, it may decrease the profit of the company. However, if it is too low, it may cause no one to accept and complete this mission

C. Restrictions

$$p_i = \frac{1}{(1 + e^{-t_i})} \tag{4}$$

When we think about the success possibility of each task, we use Logistic regression method to help us build an authentic model in huge amounts of data to fit an appropriate value[4].

$$t_i = \alpha x_i' + \beta H_i' + \frac{\sigma}{S_i'} + \rho n_i' \tag{5}$$

According to the principle of the Logistic model and combined with the actual situation, we conclude the variable t is related to four key factors, which are the price of each task, the value of each member's reputation, the distance between each member and task, and the density around a task. We can find it is obvious that except distance which is negatively related to the objective value, the remaining factors are positively correlated with the objective value.

$$\theta_{ij} = \sin(y_i) \sin(y_j^*) \cos(x_i - x_j^*) + \cos(y_i) \cos(y_j) \tag{6}$$

$$\bar{S}_i = \frac{1}{n} \sum R \arccos(\theta_{ij}) \tag{7}$$

Cause (x_j^*, y_j^*) indicates the latitude and longitude coordinates of its corresponding task, while (x_i, y_i) indicates each member's latitude and longitude. Through the latitude and longitude and the distance between two points conversion, we can easily get the value of S_{ij} .

Then we regard the location of each task as a center point in a circle, drawing a length of 2km as its radius.

$$\bar{H}_i = \frac{1}{n} \sum (C_1 \cdot m_i + C_2 \cdot c_i + \frac{C_3}{t_i}) \tag{8}$$

In order to consider variable H comprehensively, we take three factors into consideration as H 's judging criteria which are each member's task limitation, reputation and the time of begin a task. After determining this relationship, we apply entropy method to calculate each variable's coefficient. Then sum them up and limit their total value in zero to one by standardization. Using the entropy method, we get the value of each constant [3]. $C_1=0.2069, C_2=0.6060, C_3=0.1872$.

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \tag{9}$$

$$H_i' = \frac{H_i - H_{\min}}{H_{\max} - H_{\min}} \tag{10}$$

$$\bar{S}_i' = \frac{\bar{S}_i - \bar{S}_{\min}}{\bar{S}_{\max} - \bar{S}_{\min}} \tag{11}$$

$$n_i' = \frac{n_i - n_{\min}}{n_{\max} - n_{\min}} \tag{12}$$

Through above operation, we have got four relative values, index of relative price, index of relative reputation value, index of relative distance between each member and task, index of relative member density around a task. And all of these values are processed in normalization in order to minimize the impact of different dimensions. These relative indicators play a significant role in this model

$$t_i = 2.27x_i' + 2.07H_i' + \frac{1.29}{S_i'} - 2.98n_i' \quad (13)$$

After determining the equation for t , we pass the given dataset and according to the principle of the logistics model and the corresponding solution process, we need to solve for the maximum likelihood θ of the log-likelihood function of the logistic regression, so we use the method of gradient ascending to the data set, the local optimal solution of the corresponding variables is obtained by using Matlab. The coefficients of the corresponding variables are as follows (13).

According to the above description and calculation, we integrate the equations in the above steps and various constraints, after a multi-party verification of the feasibility of the model[5], ultimately establish the final pricing model, as detailed in (14).

Final Price Planning Model

$$\begin{aligned} & \min a \sum p_i x_i - b \sum p_i \\ & \left\{ \begin{aligned} x_i' &= \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \\ H_i' &= \frac{H_i - H_{\min}}{H_{\max} - H_{\min}} \\ \bar{S}_i' &= \frac{\bar{S}_i - \bar{S}_{\min}}{\bar{S}_{\max} - \bar{S}_{\min}} \\ n_i' &= \frac{n_i - n_{\min}}{n_{\max} - n_{\min}} \\ t_i &= 2.27x_i' + 2.07H_i' + \frac{1.29}{S_i'} - 2.98n_i' \\ s.t. \left\{ \begin{aligned} p_i &= \frac{1}{(1 + e^{-t_i})} \\ H_i &= \sum (0.2069 \cdot m_i + 0.6060 \cdot c_i + \frac{0.1872}{t_i}) \\ \bar{S}_i &= \frac{1}{n} \sum R \arccos(\theta_{ij}) \\ \theta_{ij} &= \sin(y_i) \sin(y_j^*) \cos(x_i - x_j^*) + \cos(y_i) \cos(y_j) \\ a + b &= 1 \\ 0 \leq x_i &\leq 100 \end{aligned} \right. \end{aligned} \right. \quad (14) \end{aligned}$$

D. Using Genetic Algorithm to Solve the Model

In order to solve this model, we use genetic algorithm to help us get the global optimal solution of the model as much as possible. According to the genetic algorithm, we maintain a potential group to perform a multi-directional search in

order to maximize profit, increase the probability of task execution, and establish a reasonable price.

To test the correctness and validity of the model, we selected the 835 tasks as samples of our tests.

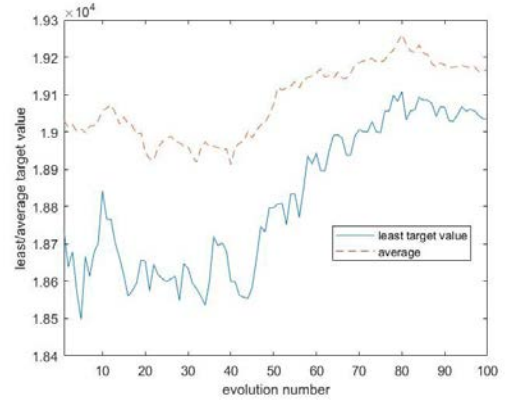


FIGURE II. THE RESULTS OF GENETIC ALGORITHM(A=0.6,B=0.4)

As we can see from the figure II(a=0.6,b=0.4), we set the population as 50 at the beginning of the genetic algorithm, select the task price in Annex 1 as unknown quantity on each chromosome, and terminate the calculation after iterating 100 times. The result is a graph of the yield, where the solid line represents Is the largest profit in the 50 chromosomes, the dotted line represents the average profit of the population, through the above calculation we can get the company's profit after the adoption of the new program.

TABLE I. THE VALUE OF EACH DATA IN DIFFERENT SITUATIONS

Different Situation	The data				
	a	b	least target value	average completion rate	average price
Situation1	0.5	0.5	23400	0.7482	74
Situation2	0.6	0.4	18700	0.7487	74.16
Situation3	0.7	0.3	13800	0.7463	74.05
Situation4	0.8	0.2	9000	0.7462	74.03
Situation5	0.9	0.1	4200	0.7469	74.17

This table I describes the results obtained when we took different values of a and b . It can be observed that different a and b values have a more or less significant effect on the result, therefore choose a good value of a and b becomes significant.

Although we can't determine the specific value of a and b in real situation, but we can look up the table or manually select the value of a and b to get better results.

III. PRICE FLUCTUANT MODEL

For the price scheme formulated in Part 2, based on the actual situation, we found that even if the probability of a task being completed by the model is large, there will still be members who will not accept the task or other objective factors that cause the task not to be accepted Cause the task abortion. In response to this situation, we have envisaged a

new model improvement program, we set up a new price-floating model to deal with the task is not accepted.

Firstly, we consider that tasks are not accepted in relation to time. When a task is issued, the longer it takes, the lower its probability of being accepted if the price remains in its original position. Therefore, our consideration of time is a factor that affects the price that we can reformulate when the mission is not accepted.

Secondly, we consider that to re-establish a new price for a task that has not yet been completed, the price of the completed task around the task should have a corresponding reference value. That is, we can set the price around the task has been completed as we re-develop the task of the reference conditions, which can be to some extent, to re-increase the task completion rate, but also protect the company will not be over-repricing to reduce the revenue.

$$\begin{cases} \frac{dp}{dt} = r(1 - \frac{p}{k}) \\ r = \left| \frac{\bar{p} - p_0}{p_0} \right| \end{cases} \quad (15)$$

Where t represents the time that has elapsed since the task was released and p represents the price at which the task was reprogrammed. r is a related constant obtained by customizing the price of the tasks that have been completed around the periphery, and k is a set coefficient. The higher the price, the smaller the rate of growth and will eventually converge to the termination price set by the company pend.

After calculation, find p on the expression of t :

$$p = (p_0 - k)e^{-\frac{rt}{k}} + k \quad (16)$$

Find the limit of p when t takes positive infinity:

$$\lim_{t \rightarrow \infty} p = \lim_{t \rightarrow \infty} (p_0 - k)e^{-\frac{rt}{k}} + k = k \quad (17)$$

Based on the above analysis, the limit should be the termination price pend.

$$k = P_{end} \quad (18)$$

So the final price change model is:

$$\begin{cases} p = (p_0 - P_{end})e^{-\frac{rt}{P_{end}}} + P_{end} \\ r = \left| \frac{\bar{p} - p_0}{p_0} \right| \end{cases} \quad (19)$$

According to the above model, the company can adjust the task price in real time according to the termination price of different tasks, attract the member orders to complete the task, and maximize the completion rate of the task while ensuring the company's return.

We set the starting price as 60, the ending price as 100, r as 1, and get the price fluctuant curve as follows:

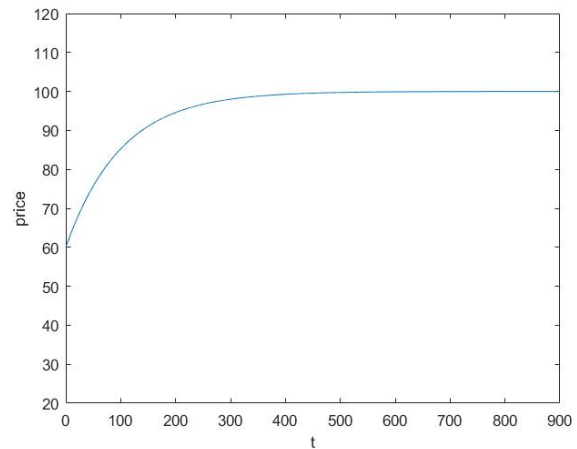


FIGURE III. PRICE FLUCTUANT CURVE

IV. CONCLUSION

In this paper, the pricing of crowdsourcing projects is achieved through the establishment of a multi-objective programming model, which results in high mission completion and low corporate spending. Also established a price fluctuant model of differential equations to further promote the completion of the task, and ultimately brought benefits to the company.

ACKNOWLEDGEMENTS

Thanks to professor Xusheng Kang for his guidance on this article. His rigorous academic style deeply moved us. He gave us a lot of valuable advice and once again heartfelt thanks to professor Kang.

REFERENCES

- [1] Marina Sergeeva, Daniel Delahaye, Catherine Mancel, Andrija Vidosavljevic. Dynamic airspace configuration by genetic algorithm[J]. Journal of Traffic and Transportation Engineering, 2017, 4(3): 300-314.
- [2] Hongbin Wang, Mian Liu. Design of robotic visual servo control based on neural network and genetic algorithm[J]. International Journal of Automation & Computing, 2012, 9(1): 24-29.

- [3] Yang, L. and Y. Qian, A sparse logistic regression framework by difference of convex functions programming. *Applied Intelligence*, 2016. 45(2): 1-14.
- [4] Szlapczynski, R., A New Method of Planning Collision Avoidance Manoeuvres for Multi-Target Encounter Situations. *Journal of Navigation*, 2008. 61(2): 307-321.
- [5] Wenqi Zhang. Evaluation method of urban sustainable development based on entropy method[P]. 2017 5th International Conference on Frontiers of Manufacturing Science and Measuring Technology (FMSMT 2017),Taiyuan,2017.