Analysis of Task Pricing Under Crowdsourcing Mode

Jinhua Ding¹, Ying Xiong², Fan Hu³, Liya Cao⁴

¹School of Statistics and Applied Mathematics, Anhui University of Finance and Economics, Anhui, 233000, China

Abstract. In light of the task pricing issues in terms of the new crowdsourcing model, we constructed the Global Optimization Model, Marquardt Algorithm, Back Propagation Neural Network Model and Fuzzy Neural Network Model at the basis of the cluster analysis, K-means Clustering and other theories and methods. This paper also designed the Fuzzy Neural Network Model by means of MATLAB programming, 1st opt and SPSS software to settle the best pricing scheme, promoting several conclusions such as the completion rate of the original pricing is related to the local population structure and other factors. The task packing mode is recommended to start when the distance between task points is less than 4000 meters and tasks packaging is effective to improve the completion rates etc.

1. Introduction

With the continuous development of mobile Internet technology, a new type of commercial crowdsourcing model has emerged. "Taking photographs to make money" is a typical product of the Internet self-service crowdsourcing model. After downloading the APP and registering an account, customers can select the task and upload the photos as required so as to earn remuneration. This brand new e-commerce mode has more and more market vitality and competitiveness. As soon as the company enters the platform, it can release tasks right away. The cost of data collection will drop greatly, and it will bring substantial income to C-terminal customers. The task bid is the core of the model, which is related to the basic elements of the task such as the completion of the task, the quality of the task completion, and the cost of the product. It can thus be seen that the new labor crowdsourcing mode[1] has huge market potential, among which task pricing is a technical problem that the platform is currently eager to overcome.

2. Data Sources and Model Hypothesis

The research content and related data of this article come from the 2017 National Undergraduate Mathematical Modeling Contest. In order to facilitate researching and solving problems, the following hypotheses are proposed: (1) suppose that the number of works that the task publisher expects has a positive influence on the bid; (2) suppose that the average quoted price of the outsourcing industry is positively related to the bid of the task publisher; (3) suppose that the difficulty of the task has a positive influence on the bid of the task publisher; (4) suppose that the average bid of similar tasks in the platform has a positive influence on the bid of the task publisher; (5) suppose that the number of task contractors in the platform has a reverse effect on the bid of the task publisher.

3. The Law of Project Task Pricing

3.1 Research Approach

Using the idea of clustering analysis [2], the task points were divided into three cities according to...
latitude and longitude information. The global optimization and the Marquardt algorithm model under the classification data were constructed respectively. Using 1stOpt software to solve the problem, the task pricing model required by the problem was obtained. The task completion rate is counted out according to different pricing divided by the fixed value, and finally the reasons why the task is not completed are analyzed.

3.2 Modeling
(1) Data classification and preprocessing: Use Google Maps to get an approximate range of all data distributions. Use the Internet to query the latitude and longitude ranges of the four cities. They are Guangzhou, Shenzhen, Dongguan and Foshan. According to the geographical zoning data, the 836 data in Annex 1 were cluster analyzed. It is not accurate to simply take the latitude and longitude range as the standard for division. There are some users who belong to two municipal divisions. In this case, duplicate values need to be removed. The final number of valid users in Guangdong is 242, 367 in Dongguan, and 176 in Foshan.

(2) Create a coordinate system: Take latitude as the vertical axis $y$, longitude as the horizontal axis $x$, the line that is perpendicular to plane $x, y$ and intersect with x-axis and y-axis as the z-axis. Let the data of z axis represent the price. Use MATLAB to obtain the changes in task pricing when customers are at different latitudes and longitudes. As shown in the following figure, it can be found that with the increase of latitude and longitude, the task price will also increase. In low-latitude high-longitude areas, there are a few tasks with high prices.

![Figure 1. The three-dimensional relationship diagram of the latitude, longitude and the task price](image)

3.3 The Expression of Law of Pricing of Crowdsourcing Service in Three Cities
First of all, carry out qualitative analysis and use MATLAB to conclude preliminarily that: In the high-longitude regions of Guangzhou, the pricing of tasks is often higher, and the increase is more obvious. In the low-latitude and low-longitude area, the task pricing is the lowest. As shown below:

![Figure 2. Quantitative analysis of task pricing in Guangzhou](image)

Then perform quantitative analysis. Use the 1stOpt software, import the relevant data into the software, and use the Marquardt algorithm and global optimization algorithm to establish the optimal equation model as:
Calculate the algebraic relationship between the two independent variables of task pricing and latitude and longitude, that is, obtain all the coefficient values in the equation. The resulting law of pricing and fitting effect of actual value is as follows:

\[
z = p_1 + p_2x + p_3x^2 + p_4x^3 + p_5x^4 + p_6x^5 + p_7y + p_8y^2 + p_9y^3 + p_{10}y^4 + p_{11}y^5
\]

From the figure above, it can be seen that the equation fits the actual situation well. The quantitative relationship between latitude and longitude and task pricing in Guangzhou is:

\[
z = -6095.45 - 574.739x + 132.8605x^2 - 1803348x^3 + 0.22597x^4 - 0.00077x^5
\]

\[-1053.57y + 44660.64y^2 - 3957.24y^3 + 131.4493y^4 - 1.55129y^5\]

Similarly, this paper gives the quantitative relationship between latitude and longitude and task pricing in Dongguan is:

\[
z = -8955430 + 52538.75x + 14649.47x^2 + 8.864163x^3 - 0.32255x^4
\]

\[+ 0.001397x^5 - 213611y + 9405.165y^2 - 138.022y^3\]

The quantitative relationship between latitude and longitude and task pricing in Foshan is:

\[
z = 46951.9 - 4077.28y + 88.653y^2 - 0.07427\left(\frac{1}{x} + 36.2784\right)
\]

\[1 - \frac{1}{3.03825}\]

The equations above reflect the law between crowdsourcing task pricing and latitude and longitude in Guangzhou, Dongguan and Foshan.

4. Task Pricing Model and Evaluation Under Centralized Location Situation

4.1 Research Approach

The pricing model required to be modified with combination of membership information due to releasing the task jointly. It should start with combined packaging approach to determine the packaging principle. The fact of essence in pricing model in foreword pricing is to modify and evaluate the results based on predictive modeling.

4.2 Modeling

(1) Determine packaging principle: The core of packaging is the distance between task points, thus the premier element considered is that whether it is appropriate to pack within the distance of task points when combining. Assumed the real distance between task points is S, the relative distance between two task points in the latitude and longitude framework is \(d(x, y)\), \(\lambda\) is Conversion coefficient between latitude and longitude and actual distance. (with Every 0.0001 degree, the distance difference is about 10 meters.), then the conversion formula for task points distance is:

\[S = \lambda \times d(x, y)\]

Set the threshold value for the transformation of task point distance, and make sure to start the packaging mode when the distance between task points is less than 4000 meters.
(2) Preprocessing of members’ data: The pre-processing of member data gives priority to the real-time distance between members and task points. Like taxi-hailing apps in daily life, it is always the nearest driver in near range to passenger picking up the order. By converting the relative distance of longitude and latitude between the member and the task point and taking personal data of the member into consideration to prejudge the data type, (Table 2) then coming into unified disposal.

| Distance between task point and member | Reserved task limitation | Reserved task beginning time | Reputation Value |
|---------------------------------------|--------------------------|----------------------------|------------------|
| Type Judges                           | Cost Type                | Efficient Type             | Cost Type        |

In all 1877 data points, according to the standard that the data that is too small and shown with the form of scientific notation, the data that conform to the standard after unification of credit value is eliminated, then selecting 1% sample randomly by using SPSS MODELER, seeds of random data is 3776858.

(3) Improvement of pricing model: The original pricing model is based on latitude and longitude, the direction of setting is to consider from the objective distance of APP platform. The first step of modification is to divide pricing model which takes distance into consideration into two levels, the first level is a single task point that is not packaged, and the second level is n task points which were packed already. The biggest difference between a packaged task and a single task is that packaged task has an additional value as \( \beta \) based on stacking single task reward. (set \( \beta \) to 0.2, stands for stacked price increased by 20% under packaging situation) Weighted four index of member data preprocess to get the appropriate subjective stacking model.

For setting the above price as base price in the modification of model, the member’s own personal profile is seen as gold coin stacking requirement. Setting the weight ratio in objective basic pricing and subjective stacking requirement to get a new pricing model.

Setting \( XE \) as data after pre-processing with limitation, \( SJ \) as data after time-preprocess, \( XY \) as data after pre-processing with limitation, \( JL_n \) as data after distance-preprocess. Here according to experience, assign \( a \) to 0.95, \( b_1 \) to 110, \( b_2 \) to 120.

Then getting pricing model after modification:

\[
\begin{cases}
  a \times \text{original objective pricing results} + b_1 \times (c_{ij} \times XE + c_{ij} \times SJ + c_{ij} \times XY + c_{ij} \times JL_n) & \text{Single taskpoint} \\
  a \times (1+\beta) \times \text{superposition of original objective pricing results} + b_2 \times (c_{ij} \times XE + c_{ij} \times SJ + c_{ij} \times XY + c_{ij} \times JL_n) & \text{Task package}
\end{cases}
\]

(4) Operating Example: Selecting 5% randomly in raw data by using SPSS MODELER software, seeds of random selecting data is 364517(28 groups of data in total). Find the distance between each task point and choose one group of points that has closest distance as an example according to original packaging principle. Table 19 is diagram of distance between each task point.
According to the table, the closest points are between group 12 and 16.

Table 2. Two task point coordinates and forecast price

| Position          | Latitude | Longitude | Bid price |
|-------------------|----------|-----------|-----------|
| Task point of 12  | 23.550796 | 113.5963  | 73.06701  |
| Task point of 16  | 23.553808 | 113.5953  | 73.09741  |

That is, if it is not packaged, the original objective pricing result of the group 12 is 73.06701; the original objective result of group 16 is 73.09741. Unify the distance between member which data had been excluded and group 12, group 16 respectively. Weight the member correlation data by coefficient of variation. Then select some random preprocess data. Weight selected data and illustrate results of subjective stacking table

Table 3. Subjective stacking results

| Member | NO.12 | NO.16 | NO.12-16 Package |
|--------|-------|-------|------------------|
| 1      | 0.1867| 0.1868| 0.1769           |
| 2      | 0.0536| 0.0536| 0.0565           |
| 3      | 0.019 | 0.019 | 0.024            |
| 4      | 0.0133| 0.0133| 0.0161           |

For member 1, the final price given by the pricing model when corresponding line is 12 (single task point) is: \( \sum_{j=1}^{12} \frac{C_{ij} \times XE + c_j \times SJ + c_{ij} \times XY + c_{ij}^{12} \times JL_{ij}}{12} \) getting result of 0.1769 when corresponding line is 12-16 (packaging task point) is \( \sum_{j=1}^{16} \frac{C_{ij} \times XE + c_j \times SJ + c_{ij} \times XY + c_{ij}^{12-16} \times JL_{12-16}}{16} \)

According to the modified pricing model, we concluded that for member 1, the final price given by the pricing model is 89.9507 when the 12th task point is selected independently, the final price given by the task pricing model is 89.9905 when the 16th task point is selected alone. The compensation is 179.9412 when the two tasks are not packaged. For the task 12-16, the final price given by the pricing model is 196.6253. From data quantification, we know that simply for member 1, the compensation of doing a packaged task increased by 9.2720 % over doing two tasks.

The original price of the 12th and the 16th points belongs to the section which task completion rate is 74.40 %. The modified task pricing model without packaging, with an increase of the completion rate approaching 19.4758%, has reached the completion rate of 88.89 % of the original rate. Besides, with the help of the packaging pricing model, the completion rate is expected to exceed 22 %. Therefore, the modified model can largely improve the task completion rate and it has a tremendous positive effect.

4.3 Model Optimization

According to the longitude and latitude of the task points given in appendix 3, we utilized k - means clustering to divide the task points into three clusters. Through the MATLAB software, the longitude and latitude coordinates of the center point are obtained. Besides, the distance from each task point to the central task point—generally, the closer to the center task point, the higher the task pricing—is
obtained from the longitude and latitude of the task points given in appendix 3. Then we operate the pricing model concluded from Question 2 and optimize the model after operating the GPS data from appendix 3.

Figure 5. Member points and task point distribution

The figure indicating the distance between longitude and latitude is obtained through the data in annex 3 by k - means clustering [4] with MATLAB software. As shown in the figure, the distribution of task points has a concentrated trend.

Figure 6. The distance between the longitude and latitude of appendix 2

Three different cluster center points coordinates could be directly obtained by the software program, and we use M、N、O to represent these points respectively. The final results are as shown in the table below:

Table 4. Coordinates of center points

| GPS  | M       | N       | O       |
|------|---------|---------|---------|
| Latitude | 23.1232 | 22.6495 | 23.2084 |
| Longitude | 113.3503 | 114.0810 | 113.2779 |

We could also figure out the Euclidean distance, provided as the basic data for subsequent analysis, between each task point and the cluster center point through the center coordinates.

When the GPS data are substituted into the task pricing model obtained by the problem two, we found that he data fluctuation are too extreme and such values might not compatible with the authentic data. Through expert pricing method, the model is further improved. When multiple the constant - 0.0085 on the basis of the original formula, the price fluctuation is small, and the value is more

\[
y = -0.0085(40934.3 + 98603.34 \ln(x) + \frac{1126391}{\ln(x)} + 15571.04(\ln(x))^2 + \frac{2027395}{\ln^2(x)} + 1544.573\ln^3(x) \\
+ \frac{2280955}{\ln^3(x)} + 87.048\ln^4(x) + \frac{1445805}{\ln^4(x)} + 2.120501\ln^5(x) + \frac{390859}{\ln^5(x).poster.png})
\]
practical. That is:

5. Conclusion
Crowdsourcing is an increasingly important mode to price tasks. In general this study finds that completion rate of the original pricing is related to the local population structure and some other factors. We also recommend to start task packing mode when the distance between task points is less than 4 kilometers to improve efficiency of the crowdsourcing. Specifically, this paper gives the quantitative relationship between latitude and longitude and task pricing in Guangzhou, Dongguan and Foshan to illustrate our model, which was proved to be effective.

Reference
[1] Cross-efficiency aggregation in DEA models using the evidential-reasoning approach[J] . Guo-liang Yang, Jian-bo Yang, Wen-bin Liu, Xiao-xuan Li. European Journal of Operational Research . 2013 (2)
[2] A collaborative and integrated platform to support distributed manufacturing system using a service-oriented approach based on cloud computing paradigm[J] . Omid Fatahi Valilai, Mahmoud Houshmand. Robotics and Computer Integrated Manufacturing . 2013 (1)
[3] A literature review of decision-making models and approaches for partner selection in agile supply chains[J] . Chong Wu, David Barnes. Journal of Purchasing and Supply Management . 2011
[4] Crowdsourcing systems on the World-Wide Web[J] . Anhai Doan, Raghu Ramakrishnan, Alon Y. Halevy. Communications of the ACM . 2011 (4)