Analysis of the Pricing Problem of Photographic Tasks Based on Multiple Linear Regression Models

Chunli Wang*, Lei Zhu, Licong Wang, Yutong Hu
School of Tourism Data, Guilin Tourism University, Guilin, Guangxi, China
**Corresponding Author.*

Abstract: This paper is for the economic pricing development of popular online labor crowdsourcing platform in China. Based on the real-time statistical data of photo-earning APP, a reasonable pricing scheme of comprehensive evaluation index system is established from the perspective of geographical location, scheduled task quota, task distribution and member distribution. The greedy algorithm, reformulation of equation, price weight model and power map of Excel are used for calculation and analysis.

Keywords: Crowdsourcing; Goal Programming; Task Pricing; Regression Model; Photo-earning;

1. Introducion

With the development of society, online crowdsourcing part-time jobs have become a new trend. Traditional market research costs a lot of money and has a long periodicity. To solve these problems, self-service labor crowdsourcing platform "photo-earning" emerges. On the one hand, the users can get the tasks that need to be photographed from the software. For example, they can go to the supermarket to check the shelves of certain products and earn a fee for the task. On the other hand, the enterprise also saves the investigation cost, and effectively guarantees the authenticity of the investigation data. In addition, the investigation period is shortened. But the pricing of the task becomes the core element of whether the task is completed or not. This requires the analysis of pricing rules and the establishment of mathematical models to get a reasonable pricing and the best results of task completion. And then it helps the product inspection to be successful.

Therefore, this paper will build a comprehensive system for the development level of "photo-earning" labor crowdsourcing benefit[1], and make corresponding research and analysis on the price setting level of part-time crowdsourcing by using pricing analysis and weighted pricing analysis.

2. Research Ideas

(1)The assumption is that we don't care how easy the task is.
(2)The assumption is that the task price is either very high or very low.
(3)The assumption is that there are no instances of members failing to complete tasks after receiving them.
(4)The assumption is that the usual route of action of members is not taken into account.
(5)The assumption is that members stay in the same region throughout the statistical period.

3. Multiple Regression Model

3.1 Research Idea

Start studying its pricing law, measure the quality of a model depending on its application effect. It can be represented graphically.

3.2 Greedy Algorithm to Solve the Local Optimal

To study its pricing law, using *MATLAB* to convert the data into longitude and latitude graphs, maps and scatter plots. According to the greedy algorithm[3-5], to find the local optimal solution and to make the scatter plot as shown in Figure 1.

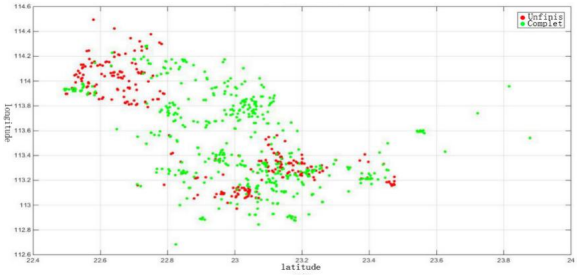

Figure 1. Task Profile

Planning the task profile into 9 areas[6], per 0.38510 degree of longitude is an interval and per 0.53802 degree of latitude is an interval as shown in Figure 2.

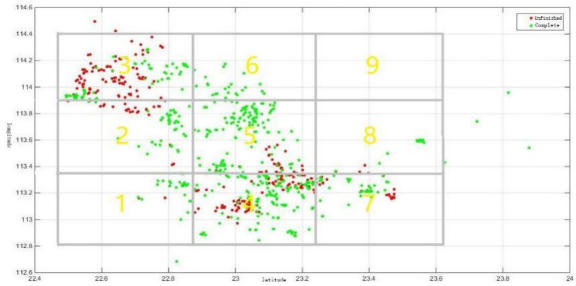

Figure 2. The Task Profile is Equally Divided

3.3 Processing of Multiple Regression Models

Process and apply from the known functions, respectively extracting 3 main data to establish a table as shown in table 1.

Assuming that the number of members in the independent variable area is W , the total number of tasks in the area is J , dependent variable is the average pricing in the area M ,so the multiple regression equation[7-9] is

$$M = \theta_1 W + \theta_2 J + \partial \quad \text{.....(1)}$$

The θ_1, θ_2 are dependent variables and ∂ is random error term. Putting the data in $MATLAB$ and t θ_1 hen obtained by multiple regression fitting: =-0.0048 、 θ_2 =-0.0049 、 ∂ =71.7574 and degree of fitting[10] R^2 =0.752>0.5, which obtain

$$M = -0.0048W - 0.0049j + 71.7574 \quad \text{.....(2)}$$

Table 1. Area Membership, Task Totals and Average Pricing Table

Partition	The total number of tasks	Membership	The average price
1	29	31	71
2	77	184	69
3	164	564	68
4	283	525	69
5	174	339	68
6	26	94	74
7	55	77	71
8	24	46	71
9	1	0	75

We can see that the equation has a high correlation with the data. Task pricing law is that as the number of members and the number of tasks increases in a area, the average pricing decreases accordingly.

Making the distribution of members and latitude and longitude into a member distribution map, and comparing it with task distribution map as shown in the figure 3.

Observing the image and find that Where image members are concentrated, task execution rates are lower. Applying the Partition concept in Figure 2 and put the regional task execution rate and Regional Members number in the same table as shown in table 2.

Table 2. Membership Number and Success Rate Data Table

Partition	The total number of tasks	Membership	The success rate
1	29	31	0.828
2	77	184	0.636
3	164	564	0.371
4	283	525	0.597
5	174	339	0.799
6	26	94	1
7	55	77	0.582
8	24	46	0.833

Assuming the dependent variable is task execution rate is O , and then the multiple regression equation is

$$O = \alpha_1 W + \alpha_2 J + \beta \quad \text{...(3)}$$

α_1, α_2 are unknown parameters,and β is random error term.

Putting the data in Table 2 into MATLAB for fitting obtain α_1=-0.001、 α_2=0.001, β=0.8341, so the multiple regression equation is

$$O = -0.001W + 0.001J + 0.8341 \quad \text{...(4)}$$

With the increase of the total number of tasks and the decrease of the number of members, the success rate will be higher. But ideally, it is clear that Figure (4) does not accord with this result. Entering the incomplete and completed the latitude and longitude of the incomplete and completed distribution into $power\ \ map$ of $Excel$ and Compare on a map, Comparing on the map.

The statistical data are located in the coastal area of Guangdong province, and the result is obtained by re-fitting:

$$O = 0.0020W - 0.0046J = 0.8097 \quad \text{....(5)}$$

Design projects in real data into new task pricing scheme, and compared with the original scheme to obtain Table 3.

Assuming that the average pricing is the dependent variable X , the independent variable is the total number of tasks S , membership

number is P , total membership Limit is Q , which obtain the Multiple regression equation

$$X = a_1 S + a_2 P + a_3 Q + e \quad .....(6)$$

a_1, a_2, a_3 are unknown parameters of independent variable, e is the error variable, putting the data in Table 1 into MATBAL to fitting to get formula

$$X = -0.0312S + 0.0075P + 0.0002Q + 71.6544 \quad (7)$$

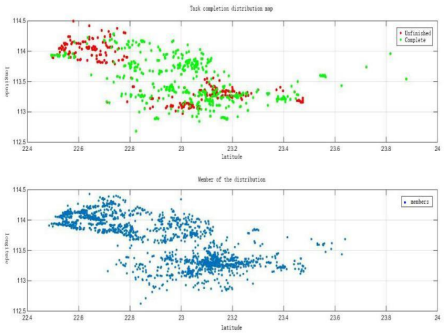

Figure 3. Membership Distribution and Task Comparison Chart

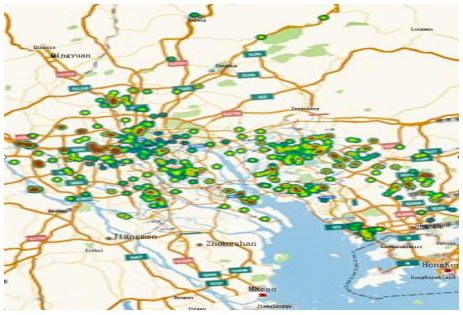

Figure 4. Color Difference Map of Unfinished Tasks and Prices

Figure 5. Completed Task and Price Chromatic Aberration Chart

Table 3. Partitioning Factor Pairs for Comparison

Partition	The total number of tasks	Membership	Members to limit	The average price
1	29	31	370	71
2	77	184	596	69
4	283	525	3742	69
5	174	339	3060	68
6	26	94	554	74
7	55	77	737	71
8	24	46	427	71

4. Greedy Algorithm Model

4.1 Research Idea

The release of the task package will be more demanding on the member's reputation value and prior limitation. Packaging 10 tasks, members below 10 will not be able to accept the package task, and new variables will appear that affect the execution rate.

4.2 The Solution of Greedy Algorithm

Greedy algorithm divided the total region into the same nine regions[14], and figured out the ratio of the number of members in each region with a predetermined task quota of more than 10 and more than quota 2 to the total number of members in the region, and got Table 4:

Table 4.The Ratio of the Number of Members to the Total Number of Members in the Region

Partition	The total number of members	Above quota 10	Quota 2 and above	Quota above 10 (%)	Quota 2 above (%)
1	31	12	28	39	90
2	184	12	119	7	65
3	564	60	388	11	69
4	525	92	378	18	72
5	340	63	253	19	74
6	94	10	80	11	85
7	77	21	66	27	86
8	46	16	36	35	78

Counting the number of members and the total number of tasks in each region, and got Table 5:

Table 5. List of Membership and Task Totals

Partition	Membership	The total number of tasks
1	31	29
2	184	77
3	564	164
4	525	283
5	339	174
6	94	26
7	77	55
8	46	24

The linear regression equations are listed according to Table 5:

$$J = \omega W + \varepsilon \quad \text{.....(8)}$$

The ω and ε are unknown parameters, the data processed by MATLAB can be obtained as follows:

$$J = 0.3984W + 12.1976 \quad \text{....(9)}$$

Calculating the goodness - of - fit[10]: R^2 =0.909. The ratio of the number of members with a limit of more than 2 to the total number of members in the region is used as the weight[10-12] of the number of packaged tasks in each region, and combined the number of members in each region to made into table 6:

The linear regression fitting of the data in Table 6 is as follows:

$$W = b_1 T + c \quad \text{....(10)}$$

The T is the weight value of the number of tasks packed in each region, and the b c is the unknown parameter value. The data in the table are fitted with MATLAB to get:

$$W = -1649.45T + 1508.886 \quad \text{.....(11)}$$

5. Weight Model

5.1 Research Idea

If the multiple regression pricing model is added to the Weight Model of Greedy Algorithm, the correct pricing solution will emerge.

5.2 The Solution of Weight Model

List the membership whose quota is greater than 2 in each region. Compared with the total number of membership in the region, the membership with the quota of 1 cannot book the packing task, so the data weight is obtained. Figure out the number of packaged published tasks from the number of tasks in each region, and then appropriately reduce the price of packaged tasks , and appropriately increase the price of unpackaged tasks.The reduced and increased prices can be obtained by taking the weighting of each region into the simultaneous equations 1, in which the weighting of each region can be inserted. Based on the analysis of multi-objective programming, the simultaneous equations (12) is obtained

$$\begin{cases} J = 0.3948\ W + 12.1976 \\ W = -1649.45\ T + 1508.886 \end{cases} \quad \text{.....(12)}$$

By analyzing the membership data table of network data, it is found that the membership with quota of 1 account for 35% of the total number of members. The package publishing task will reduce the interest of the members with low quota on the platform and lead to the loss of the platform membership.

Table 6. Membership Number and Weight Relationship Table

Partition	Membership	The weight
1	31	0.9
2	184	0.65
3	564	0.69
4	525	0.72
5	340	0.74
6	94	0.85
7	77	0.86
8	46	0.78

6. Synthesis of Weight Model and Greedy Algorithm

6.1 Research Idea

Work out task pricing solution for new projects. We imported the online data into the map through Excel Power Map for observation, and found that except for a few tasks scattered, the rest tasks were concentrated in Haizhu District and Tianhe District of Guangzhou, Nanshan District and Longgang District of Shenzhen.

6.2 Weight Model and Greedy Algorithm

According to the regression equation analysis, this text uses the Weight Model analysis and Greedy Algorithm to solve the local optimal solution. It is assumed that there are three factors: regional economy Z, regional membership number P and regional task number S, which will affect the quotation task price. As shown in:

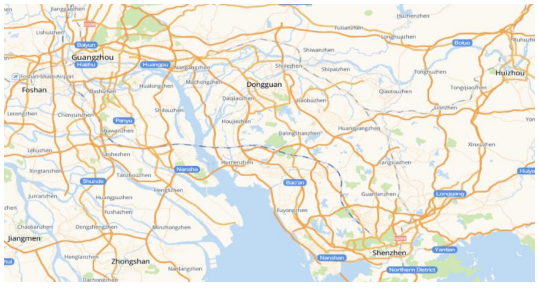

Figure 6. The Task Distribution Map in Annex Iii

According to the National Bureau of Statistics of the People's Republic of China, the recent economic conditions[15]and traffic conditions of each region are shown in Table 7. Through the analysis of data 2 and 3, the number of membership and tasks in these five regions (Baiyun District, Haizhu District, Longgang District, Nanshan District, Tianhe District) can be counted.

Table 7. Statistical Table of Regional Variables

Partition	Number of jobs	Membership	District GDP (100 million yuan)	Economic level weight
Baiyun district	514	93	1640	0.12
Zhuhai district	368	92	1550	0.12
Longgang district	287	55	2636	0.20
Nanshan district	239	50	3714	0.28
The tianhe district	432	97	3801	0.28
The sum of the	1840	387	13341	1.00

Through 5.1, the incomplete multiple regression equation of task price (2) can be obtained, and the initial pricing table 8 can be obtained:

Table 8.The Initial Price

Partition	The initial price
Baiyun district	68.79
Zhuhai district	69.51
Longgang district	70.09
Nanshan district	70.35
The tianhe district	69.18

The economic level weight t in Table 7 is added to the initial pricing m to obtain the reasonable pricing M_1, as shown in Table 9.

$$m-10+10*t=M_1$$

Table 9. Reasonable Pricing

Partition	The initial price
Baiyun district	59.99
Zhuhai district	60.71
Longgang district	62.09
Nanshan district	63.15
The tianhe district	61.98

7. Conclusion

In this paper, the statistical regression model, objective programming model and weight model are used for calculation and analysis, and a large number of data are used for processing. In this way, the results obtained are relatively reliable.The conclusions are as follows : (1) Crowdsourcing is developing very fast in the coastal areas of Guangdong, but very slowly in the central and western regions. Therefore, we can vigorously impel it in the central and western regions to promote the development of part-time Internet crowdsourcing tasks, and improve people's livelihood. (2) Pricing rationality solves a lot of real problems, so as to help members better complete tasks, thereby promoting the rapid development of the Internet economy. (3) It is necessary to dig into the local development advantages of Guangdong coastal areas and further strengthen external economic communication and cooperation. Besides, we will increase fiscal input to strengthen the economy. At the same time, we will optimize the Internet infrastructure structure in the eastern and western regions and actively promote the development of the Internet system.

Acknowledgement

The research is supported by: Research Project of Institute of Information Technology of GUET（XJ202024）,Guangxi University's young and middle-aged teachers'Basic Research Ability Improvement Project (2019KY1046), Guangxi University's young and middle-aged teachers' Basic Research Ability Improvement Project(2019KY1037), Institute of Information Technology of GUET Research Project (B201911), 2020 Project of Three-wide education of Institute of Information Technology of GUET(2020SQ03).

References

[1] Peng Zhang, Ruoyu Lu. On Incentive Mechanism of Crowdsourcing Innovation——By Principal-agent Theory

[J]. Technical economy and management research, 2012, (06):45-48.
[2] Qiyuan Jiang ,Jinxing Xie. Mathematical Model[M]. Beijing:Higher Education Press,2003
[3] Jingjing Zheng, Zhinian Luo.Resource Management Based on Greedy Algorithm for Carrier Aggregation in LTE-A System. Computer engineering. (2017):56-60
[4]Shangshang Li."Dynamic vehicle routing problem analysis based on improved Greedy Algorithm." Modern economic information. (2017):345.
[5] Changhai Nie ,Jing Jiang."Optimization of Configurable Greedy Algorithm for Covering Arrays Generation."Journal of Software.(2013):55-69
[6] Jie Zhang, Abing Lv, Tao Li, Xueming Zhou, Huanchun Lou. Large scale goal programming model for development of Daqing oilfield[J]. Journal of Harbin Institute of Technology,1999,(04):11-15.
[7] Xianyou Fan, Chaoyang Li, Chuanhua Yang, Xiaoqin Cao."Short-term water level prediction method based on multiple linear regression analysis."Port & Waterway Engineering. (2014):177-180
[8] Zheng Liu , Guangzhe Deng ,Xiaojun Liu. Application of multivariate Regression Analysis in Coal Seam Gas Content Prediction[J]. Ming Safety & Environmental Protection , 2013(05):57-60.
[9] Haibo Jiang, Yanru Li.General Mathematical Model in Regression Analysis[J]. Mathematics in Practice and Theory, 2006, (01):174-180.
[10] Yuying Jiang ,Qiang Liu. Several ingenious algorithms for mathematical expectation of discrete random variables [J]. College Mathematics, 2008, (05):153-154.
[11] Chuanyin Xie.Landslides hazard susceptibility evaluation based on weighting model[J]. Journal of Central South University(Science and Technology), 2011, 42(06):1772-1779.
[12] Bing Gao,Lin Kong,Biao Xie,Wenji Wang,Siyang Qu,Meina Liu,Qiuju Zhang."Establishment and Application of Weighted Probabilistic Principal Component Analysis Model."Chinese Journal of Health Statistics. (2018):4-7.
[13] Lijun Yan, Junmou Qiao, Jian Xu, Hui Liang."A linear goal programming model for weight calculation in fuzzy AHP."Manufacturing Automation. (2018):113-115+123
[14] Zhidong Liu,Hua Chen,Yongming Zou,Hua Zhang. Particle swarm optimization (PSO) to solve multi-objective programming model of oil field measures[J]. Journal of Oil and Gas Technology, 2009, 31(01):261-263.
[15] Zhiping Fan, Guofen Hu. A Goal Programming Method for Multiple Attribute Decision Making with Intervals[J]. Journal of Industrial Engineering and Engineering Management, 2000, 014(004):50-52.

