Research on Task Pricing Strategy of "Taking Photos to Make Money" APP

Haichao Hu

Abstract. "Taking photos to make money" APP is a kind of self-service mode under the mobile Internet mode, so it is of practical significance to study the task pricing for the later improvement of service. Based on this, this paper firstly determines the main indexes and calculation methods that may affect task pricing, then uses the one-way analysis of variance to determine the significant differences caused by each index variable on task pricing, and then uses multiple regression analysis to consider the influence of multiple factors on task pricing. As a result, the index variable should be removed. Then the relationship between the other three indicator variables and task pricing is analyzed. And then you get the equation. Finally, the unfinished data in the original data were selected, and the influence of index variables on task pricing was observed by using multiple regression analysis. The reasons for unfinished tasks are analyzed mainly because the shortest distance between members and tasks is not considered.

1 Introduction

With the continuous development of Internet technology, "taking photos to make money" has become a self-service model under the mobile Internet. This self-service crowdsourcing platform based on mobile Internet greatly saves survey costs, effectively guarantees the authenticity of survey data and shortens the survey cycle\(^{(1)}\). In this service mode, task pricing is the core of the platform. If the pricing is not reasonable, some tasks will be neglected, resulting in the failure of commodity inspection.

2 Determination of main indicator variables and its calculation method

In order to study the task pricing rule of the platform; Firstly, this paper determined the dependent variable Y as the task pricing, and selected some dependent variables with statistical relationship with Y according to the original data, namely: task density \(X_1\), member density \(X_2\), member credibility \(X_3\), and the shortest distance between member and task \(X_4\).

In essence, one-way analysis of variance (ANOVA) uses the processing method selected as the center to build a circle with radius \(r\). MATLAB is used to calculate the number of tasks \(m_i\) in the region and the sum of the member's credibility in the region. Take the number of tasks as its task density \(MD_i\) and member credibility \(hi\), that is: \(MD_i=\sum m_i.\) This paper takes a task point as the center point, construct a circle with radius \(r\), calculate the number of members in the area by MATLAB, and divide by the corresponding area, namely: 

\[
p_i = \frac{n_i}{s}
\]

Where, \(n_i\) is the task number and the total number of peripheral members of the \(A_i\), s is the area of the surrounding area that is determined with the membership as the center. In addition, this paper can directly use \(n_i\) represent the membership density of task number \(A_i\). Subsequently, this paper found that the radius \(r=5\mathrm{km}\) was the most reasonable.

The shortest distance between the member and the task is calculated according to the Euclidean distance algorithm. The specific distance conversion formula is as follows:

\[
d_{\text{min}} = \sqrt{(111(x_{1m} - x_{2n}))^2 + (102.2(y_{1m} - y_{2n}))^2}
\]  

\(x_{1m}\) is the latitude of task \(A_m\), \(y_{1m}\) is the longitude of task \(A_m\), \(x_{2n}\) is the latitude of member \(B_n\), \(y_{2n}\) is the longitude of member \(B_n\).

Therefore, the distance of task \(A_m\) is: \(\text{d}_{\text{min}}=\min\{d_{\text{min}}\}\).

3 The principle and analysis of one-factor analysis of variance

In essence, one-way analysis uses the processing method

| Index variable | \(X_1\) | \(X_2\) | \(X_3\) | \(X_4\) |
|---------------|-------|-------|-------|-------|
| Index variable meaning | Member density | Task density | Member credibility | Minimum distance between member and task |

Table 1. Index variable scale

1 School of economics Shanghai university Shanghai, China

13476554205@163.com

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).
of statistical inference to calculate the F statistic and conduct the F test\[^3\]. Where, the total sum of variance squares is denoted as SST and decomposed into two parts: One part is the deviation caused by four index variables, denoted as SSA.

The other part is SSE caused by the inside of a single index variable. There is:

\[
SST = SSA + SSE
\]

Including:

\[
SSA = \sum_{i=1}^{k} n_i (\bar{x}_i - \bar{x})^2
\]

In formula (3), \(k\) is the number of four index variables; \(n_i\) is the sample size of the \(i\)th index variable, all of which are 835 data.

Thus it can be seen that the sum of sample dispersion squares between groups is the sum of squares between the mean of the number of index variables and the mean of the population, which reflects the influence of different index variables on price\[^4\][\(^5\)]. By:

\[
SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2
\]

It can be seen that the sum of squared deviations in the group under a single indicator variable is the sum of squared deviations between each data and the average value of the horizontal group, reflecting the influence of a single indicator variable.

F statistics is the ratio of the average sum of squares between groups to the average sum of squares within groups, and the calculation formula is:

\[
F = \frac{SSA / (k-1)}{SSE / (n-k)}
\]

From the formula of F value calculation, it can be seen that if different levels of control variables have a significant impact on task pricing, then the sum of squares of inter-group dispersion of indicator variables must be large, and the F value will be relatively large\[^6\]. On the contrary, if the different levels of control variables do not significantly affect the indicator variables, then the influence of the sum of squares of deviations within a single indicator variable group will be relatively large and the F value will be relatively small\[^2\].

The following is the result file output by SPSS:

Table2. One-way anova table of task density and task price

| Task Pricing | Sum of Squares | df | Mean Square | F     | Significance |
|--------------|----------------|----|-------------|-------|--------------|
| Groups       | 7611.766       | 106| 71.809      | 5.578 | .000         |
| Intraclass   | 9372.737       | 728| 12.875      |       |              |
| Variance     | 16984.503      | 834|             |       |              |

Table3. Single factor anova table of credit accumulation and task price

One-way Analysis of Variance

| Task Pricing | Sum of Squares | df | Mean Square | F     | Significance |
|--------------|----------------|----|-------------|-------|--------------|
| Groups       | 15625.512      | 608| 25.700      | 4.274 | .000         |
| Intraclass   | 1358.991       | 226| 6.013       |       |              |
| Variance     | 16984.503      | 834|             |       |              |

Table4. Single factor anova table of credit accumulation and task price

One-way Analysis of Variance

| Task Pricing | Sum of Squares | df | Mean Square | F     | Significance |
|--------------|----------------|----|-------------|-------|--------------|
| Groups       | 16699.065      | 791| 21.364      | 10.752| .000         |
| Intraclass   | 85.438         | 43 | 1.987       |       |              |
| Variance     | 16984.503      | 834|             |       |              |

Through the calculation by SPSS, it can be seen that the relationship between the four variables and the task price is significant, and the four variables are the influencing factors of the task price.

4 Construction and processing results of multiple regression analysis model

Since task price \(Y\) may be affected by task density \(X_1\), member density \(X_2\), member credibility \(X_3\), shortest distance between member and task \(X_4\), etc., this paper need to further discuss the parameter estimation and goodness of fit evaluation of the model.

Since the random error term should satisfy the Gauss-Markov hypothesis, the multivariate linear regression model of \(Y\) and \(X_i\) can be constructed.

The multiple linear regression model was built by RStudio, and the linear regression equation of the overall level \(Y\) for the four independent variables was obtained as:

\[
y = 73.87 - 0.1531x_1 - 0.04361x_2 - 3.078x_3 - 5x_4 - 0.001842x_4
\]

(R6)

RStudio\[^7\] is used for significance test analysis, and the running results are shown in figure 1:

![Fig1. Results of the significance test of global multiple linear regression](image)

The result can be obtained: the corresponding P value of the variable is 2.13*10^-15, 3.58*10^-10, 0.00191, 1.54*10^-5. It can be seen that the correlation between the member's credibility and the task price is weak, the shortest distance between member and the task has a
strong correlation with the task price, and the task density and the member density have a strong correlation with the task price. The value of F statistic is 99.22, and the value of P < 2.2*10^-16 < 0.05, so it can be judged that the established regression equation is significantly effective.

Since the significance of the member's credit degree X3 is relatively weak compared with other factors, a mathematical model is established when the member's credit degree is deleted:

\[ y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_4 + \beta_4 X_5 + \theta. \]

RStudio software is used for significance test analysis, and the running results are shown in figure 2:

Calculated by RStudio software, the linear regression equation of overall level Y to 3 independent variables can be obtained when member's credit degree is deleted as follows:

\[ y = 74.283745 \cdot 0.184611x_1 + 0.050587x_2 - 0.014579x_4 \]  
(9) Satisfies the significance test.

Similar to the above, the shortest distance between the member and the task with the least significant factor was deleted, and the linear regression equation of the overall level Y to the two independent variables was obtained as:

\[ y = 73.42451 - 0.17140x_1 - 0.04943x_2 \]  
(10) Satisfies the significance test.

The incomplete data were selected and the multiple linear regression model was established by R software. Thus, the linear regression equation of the overall level Y to the four independent variables was obtained as:

\[ y = 72.36 - 0.1089x_1 - 0.0367x_2 - 0.000244x_3 - 0.01638x_4 \]  
(11) The significance test analysis was carried out with R software, and the operation results were shown in FIG. 4:

The significance test analysis was carried out with R software, and the operation results were shown in figure 3:

Calculated by RStudio software, the linear regression equation of overall horizontal Y to 3 independent variables is as follows:

\[ y = 73.903953 - 0.159596x_1 - 0.049827x_2 - 0.018505x_4 \]  
(7) Satisfies the significance test.

In order to find out the reasons for the unfinished task, the completion of the task was divided into those who completed the task and those who did not.① When considering the completion of the task, the multiple linear regression model of Y and Xi is also constructed⑧.

The significance test analysis was carried out with R software, and the operation results were shown in figure 3:

Calculated by RStudio software, the linear regression equation of overall level Y to the two independent variables was obtained as:

\[ y = 72.36 - 0.1089x_1 - 0.0367x_2 - 0.000244x_3 - 0.01638x_4 \]  
(11) The significance test analysis was carried out with R software, and the operation results were shown in FIG. 4:

From the output results, it can be seen that:

1. the corresponding P value of the variable \( \beta_1 \), \( \beta_2 \), \( \beta_3 \), \( \beta_4 \) is 3.31*10^-5, 9.97*10^-6, 0.04202, 0.00354, the value of the F statistic is 31.36, and the P value < 2.2*10^-16 < 0.05, it can be considered that the established regression equation is significantly effective.

Then, in the case of deleting the least significant member's credibility and the shortest distance between member and task, the linear regression equation is:

\[ y = 72.329440 - 0.113786x_1 - 0.041635x_2 - 0.015379x_4 \]  
(12) All of them meet the significance test.

Based on the above analysis, it is concluded that under the overall situation, there is a significant relationship between task pricing and task density, member density, and the shortest distance between member and task, using the functional equation:

\[ y = 73.903953 - 0.159596x_1 - 0.049827x_2 - 0.018505x_4 \]  
(14) to describe the law of task pricing.

By observing the information of unfinished tasks, it can be seen that the unfinished tasks are concentrated in
the tasks with low price, which is closely related to the credibility of members and the shortest distance between members and tasks. After the fast division of tasks, it can be seen that there is a significant relationship between price and task density and member density, while the relationship between price and member credibility and the shortest distance between member and task is weaker. For this reason, the functional relationship satisfying this kind of price is found by using the multiple regression model respectively in the case of task completion and task not completed:

Finished task: \[ y = 73.42451 - 0.17140x_1 - 0.04943x_2 \]

Unfinished task: \[ y = 71.039164 - 0.069950x_1 - 0.04753x_2 \]

The comparison between the two functions shows that the two curves are very close, and the task price is related to the task density. The member density has a significant relationship, and considering the influence of four factors, the incomplete task is related to the shortest distance between the member and the task in the overall situation, and the impact is large. The member considers the location of the task, resulting in the unfinished task.

5 Conclusions

In order to study the pricing law of items in the original data, we firstly determine the main indexes that may affect the task pricing and find out the calculation method of the index variables. Then, MATLAB was used to calculate the results of each indicator variable, and task pricing constituted a one-to-one correspondence. The significant differences and changes caused by each index variable to the task price were determined by using one-way analysis of variance. Then the model was improved and the effect of multiple factors on task pricing was considered by multiple regression analysis. From this, it is concluded that there is little correlation between task pricing and the member's credit rating of the indicator variable. After eliminating the member's credit rating, the relationship between the other three indicator variables and task pricing is analyzed and the equation is obtained. Then select the unfinished data in the original data and observe the influence of index variables on task pricing by using multiple regression analysis. The reasons for unfinished tasks are analyzed mainly because the shortest distance between members and tasks is not considered.

References

1. Jiang Qi-yuan, Xie Jin-xing, Mathematical Model, Beijing: Higher Education Press, 2011.
2. Song Zhi-gang, He Xu-hong, Xie Lei-lei, Practical Course of SPSS Statistical Analysis (2nd Edition), Beijing: Posts And Telecommunications Press, 2013.
3. Jeffrey m. Wooldridge, Introduction to Econometrics, Beijing: China People's Publishing House, 2010.
4. Wang Zhi-xin, MATLAB Program Design And Its Mathematical Modeling Application, Beijing: Science Press, 2013.
5. Ronald d. ockey, SPSS is Actually Very Simple[M]. Beijing: China Renmin University Press, 2010.
6. Tang Nian-sheng, LI Hui-qiong, Applied Regression Analysis, Beijing: Science Press, 2014.
7. Brain Dennis, R Language Beginner's Guide. Beijing: Posts and Telecommunications Press, 2016.
8. Cai Suo-zhang, Principle and Method of Mathematical Modeling. Beijing: Ocean Press, 2000.