Analysising the influence factors of single task pricing based on public packet system: An Empirical Study in China

Jiang-Jie Sun¹,³, Ping-Wang², Ya-Nan Du¹, Jun-Liu³*

¹Health Management College, Anhui Medical University, Hefei, Anhui, 230032, China
²Clinical Medical College, Anhui Medical University, Hefei, Anhui, 230601, China
³School of Humanistic Medicine, Anhui Medical University, Hefei, Anhui, 230032, China
*Corresponding author e-mail: ahmulj@163.com

Abstract: This study takes single task pricing as an example, analyzes the key factors of task pricing and constructs an index model of single task pricing through the data of 835 task cases. Through the empirical analysis of the new pricing model, we draw the conclusion: the new pricing model reduces the cost of the task and increases the completion rate. It has a certain guiding significance for the development and management of the public crowdsourcing platforms.

1. INTRODUCTION
The essence of a crowdsourcing system, a new virtual platform, lies in the aggregation of supply and demand information. The platform users can get a certain reward by completing corresponding tasks, so the platform can help a large number of freelancers (the membership) to meet their own needs as well as help businesses or organizers (the recruiters) solve problems. Therefore, crowdsourcing systems came into being in the “Internet+” era and they have become common and popular[1]. Crowdsourcing is a sourcing model in which individuals or groups obtain task orders and service. These task orders include single task orders and multi-task ones, from a large, relatively open and often rapidly-increasing group of internet users. How to ensure the contribution rate of those orders[2], how to control the recruiter’s costs, and how to maximize the interests of the members have become key problems urgently in need of a solution. Building a single task pricing model can not only reduce the cost and improve the completion rate of the task, but also has a certain guiding significance for the development and management of public crowdsourcing platforms and the exploration of this new economic development model of the Internet. An empirical analysis of the model is carried out using data in four cities of Guangdong, China.

In the market economy system, due to the constraints of the platform and the reality of the public crowdsourcing market, whether a member is willing to do the task or not is often related to a lot of factors such as economic cost, time cost, and difficulty level. Recruiters use a variety of ways to promote their members’ enthusiasm: raising the task price, reputation priority[3]. A series of scholarly studies have discussed the pricing mechanism of the market, but the literature lacks quantitative research on the pricing of public crowdsourcing tasks. This research is based on the mining of a large amount of information about the platform and trying to fix the price for a single task of the public crowdsourcing platform to improve the efficiency of the operation of the platform, which is of great significance.
The rest of the article is organized as follows: Section 1 introduces the theory of model. Section 2 study design. Section 3 reports the main results and is followed by a detailed discussion of the results in Section 4.

The exponential function model is one of the most widely used statistical models in modern science and technology, economy, risk management, and other fields[4]. In 2016, Bu applied the exponential function model to the evaluation of competitiveness[5]. In 2017, Boratyńska introduced the exponential function model to the prediction of Reserves[6]. Shen applied the exponential function model to health quality management and got an effective evaluation[7]. In this study, we attempt to establish a single task pricing model of public crowdsourcing platform with an exponential function.

Based on the existing literature, we propose four hypotheses which are related to single task pricing. The following hypotheses are postulated.

Hypothesis 1: Task distance is positively correlated with task pricing.
Hypothesis 2: A significant difference exists in task pricing when there are different task densities.
Hypothesis 3: Membership density is significantly related to task pricing.
Hypothesis 4: Membership credit is significantly related to task pricing.
Hypothesis 5: The task price is significantly influenced by four factors: task distance, task density, membership density, and membership reputation. (other factors are not considered in this study.)

2. Measures used

2.1 Sample
The data of this study come from the “photograph task” of CUMCM (China Undergraduate Mathematical Contest in Modeling) in 2017. The study considers data of four cities in Guangdong Province, China: longitude, latitude, membership information, task information, and other factors.

2.2 Notation

| Symbol | Name              | Explanation                                                                 |
|--------|-------------------|----------------------------------------------------------------------------|
| $P$    | task pricing      | Price of a single task                                                      |
| $D$    | task distance     | Minimum distance between tasks in kilometers (km)                           |
| $\rho_T$ | task density     | Number of tasks within 10 km                                                |
| $\rho_M$ | membership density| Number of members within 10 km                                               |
| $C_M$  | member reputation | Percentage of tasks successfully completed by members in relation to the total number of tasks taken by members |
| $R$    | task completion rate | Proportion of tasks completed in a certain area                             |

3. Results

3.1 Correlation between dependent variables and independent variables for successful tasks
Using the statistical method discussed above, we obtain the correlation coefficient between the four independent variables and dependent variables and the correlation among independent variables for the successful tasks. These are shown in Table 2.

| Table 2. Relationships between dependent and independent variables for successful tasks |
|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|-----------------------------------------------|
| 1. Task pricing | Pearson Correlation | 1 | 2 | 3 | 4 | 5 |
| sig.(2-tailed) | N | 522 |
| 2. Task distance | Pearson Correlation | .239** | 1 |
3. Task density
Pearson Correlation 
-0.449** -0.328** 1
sig.(2-tailed) .000 .000
N 522 522

4. Membership density
Pearson Correlation 
-0.414** -0.302** .880** 1
sig.(2-tailed) .000 .000 .000 .000
N 522 522 522 522

5. Membership reputation
Pearson Correlation 
-0.250** -0.826** .483** .442** 1
sig.(2-tailed) .000 .000 .000 .000 .000
N 522 522 522 522 522

** Correlation is significant at the 0.01 level (2-tailed); *. Correlation is significant at the 0.05 level (2-tailed).

The analysis results show that the pricing for the successful task is positively related to the task distance, which verifies Hypothesis 1. The pricing is negatively correlated with task density, membership density, and member reputation, which validates Hypotheses 2-4. The correlation coefficient between the independent variables was significant. And in particular, the correlation coefficient between task density and membership density is 0.88, which is more than the minimum standard of 0.6. Our further analysis indicates that the lowest tolerance value is 0.094 and the highest VIF is 10.638. Accordingly, there may be a multicollinearity problem in our dataset[9].

3.2 Regression analysis between dependent variables and independent variables for successful tasks
To test the influence of a collinearity problem in the research, referring to the above analysis results, we further use stepwise regression on the influence factors of the successful task pricing and get the results shown in Table 3.

| Predictor | Step 1 | Step 2 |
|-----------|--------|--------|
|           | $\beta$/Coef | SE | $\beta$/Coef | SE |
| Task pricing |        |       |        |       |
| $D$ | 0.023*** | 0.003 | 0.026*** | 0.003 |
| $\rho_T$ | -0.013*** | 0.003 | -0.012*** | 0.003 |
| $\rho_M$ | -0.013*** | 0.003 | -0.013*** | 0.003 |
| $C_M$ | -0.031* | 0.014 |        |       |

Adjust $R^2$ 0.297 0.334

$F$ 106.932*** 140.237***

Notes: *, **, and *** indicate $P<0.05$, $P<0.01$, and $P<0.001$, respectively.

The above results verify Hypothesis 5.

3.3 Construction of single task pricing model
From the above correlation analysis results, we can set task distance, task density, membership density, and average reputation as independent variables, then set single task pricing as the dependent variable and based on the results of the above regression analysis, we can get a single task pricing Model 1 and Model 2:

$$\ln P = 4.209 + 0.023\ln D - 0.013\ln \rho_T - 0.013\ln \rho_M - 0.031\ln C_M --Model 1$$

and

$$\ln P = 4.151 + 0.026\ln D - 0.012\ln \rho_T - 0.013\ln \rho_M --Model 2$$
\( P \) denotes price and \( D \) denotes the shortest distance in the task. \( \rho_r \) denotes the task density, \( \rho_M \) denotes the membership density, and \( C_M \) denotes the member reputation.

3.4 Simulation of pricing model

3.4.1 Pricing simulation. By using the MATLAB software to fit the model, we can get Figure 1.

As can be seen from the above graph, the pricing data of Model 1 and Model 2 are more stable than the original price data. Through further analysis of the statistics of 835 tasks' raw data and fitting results, we find the total price of the raw data is 57707.50 (RMB) with a variance of 4.5100; the total price of Model 1 is 57130.44 (RMB) with a variance of 2.5379; and the total price of Model 2 is 57260.60 (RMB) with a variance of 2.5197. The total cost of Model 1 is lower than that in Model 2. The total cost of Model 2 is lower than original cost. This implies that using Model 1 as a single task pricing can achieve the purpose of saving cost.

3.4.2 Success rate simulation. For the successful tasks, we use price as the independent variable and consider the rate of completion of a single task (with the task for the center, 10 km as the radius of the sphere, and a single task completion rate) as the dependent variable to do linear regression. We then substitute the original data information into Models 1 and 2 to calculate the corresponding new price and simulate the completion rate of Models 1 and 2. The simulation results are shown in Figure 2.
Figure 2 Comparison of single task completion rate

The above graph shows that the completion rate data of Models 1 and 2 are more stable than the original completion rate data. Through further statistics of the original completion rate data and the fitting results of 835 tasks, we find that the average completion rate of the original single task is 59.35%, of the single task of Model 1 is 71.68%, and of the single task of Model 2 is 71.65%. Note that Models 1 and 2 have little difference in their completion rates. Compared with the original pricing, the completion rate has increased by 18%, and the completion rate of Model 1 for a single task is slightly higher than that of Model 2.

4. Conclusion

Based on the analysis of the constructed single task pricing model and the statistical analysis results above, we can draw the following conclusions.

The factors affecting the single task pricing of the public crowdsourcing system are mainly as follows: mainly include four factors: task distance, task density, membership density and reputation. The larger the task distance, the higher the task pricing, which is consistent with the social practice mechanism. Model 1 is a single task pricing model that is very useful. It not only improves the single task completion rate but also reduces the task completion cost.

Acknowledgments

The work was supported in part by the Natural Science Foundation of Anhui Province (1908085MG233), Quality Engineering Research Projects of the Anhui Department of Education about Wisdom Classroom (2018zhkt180), the Natural Science Foundation for the Higher Education Institutions of Anhui Province of China (KJ2019A945, KJ2016A369, KJ2016A367, and KJ2016A372), and the Humanities and Social Science Research Projects of Key Bases in Sichuan Province (SCYG2018-10).

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] Marcus A, Wu E, Madden S, et al.(2011) Crowdsourced Databases: Query Processing with People, Fifth Biennial Conference on Innovative Data Systems Research, Asilomar, USA, pp: 9-12. http://www.zentralblatt-math.org/zbmath/en/search/?q=an%3A11053583.

[2] Cooper S, Khatib F, Makedon I, et al. (2011) Analysis of social gameplay macros in the Foldit cookbook. ACM. New York. pp. 9-14. https://www.researchgate.net/publication/241623760.
[3] Singer Y, Mittal M. (2013) Pricing mechanisms for crowdsourcing markets. ACM. New York. pp:1157-1166. https://www.researchgate.net/publication/262206057.

[4] Meeusen W, Broeck J V D. (1977) Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. International Economic Review, 18:435-444.

[5] Bu Hongyun, Tao Lingling. (2016) The Competitiveness Evaluation of High-tech Industry Cluster Based on the Exponential Function——A Case Study of Beijing-Tianjin-Hebei. Journal of Industrial Technological, & Economics, 19-27.

[6] Boratyńska A. (2017) Robust Bayesian estimation and prediction of reserves in exponential model with quadratic variance function. Insurance Mathematics, & Economics, 76-86.

[7] Shen J M, Dang Y G, Zhou W J, et al. (2016) Grey dynamic multiple attribute correlation decision-making model based on exponential function. Control, & Decision,31: 1441-1445.

[8] Mcerlane F, Foster H E, Armitt G, et al. (2017) Development of a national audit tool for juvenile idiopathic arthritis: a BSPAR project funded by the Health Care Quality Improvement Partnership, Rheumatology, 57:140-151.

[9] Cheng P, Wei J, Marinova D, et al. (2017) Adoption of Protective Behaviours: Residents Response to City Smog in Hefei, China. Journal of Contingencies, & Crisis Management, 1–12.