FAILURE ANALYSIS OF “PHOTOGRAPHING AND MAKING MONEY” TASK BASED ON B-P NEURAL NETWORK

1*Zeng Yunhui, 2Guo Hongfei, 3Zhang Yuanbiao, 4Huang Guoquan, 1Huang Shusheng, 1Chen Yilin, 1Shen Shiyue, 1Chen Zhengguo

1Bachelor Research Scholar, College of Intelligent Science and Engineering, Jinan University, No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province.

2Associate Professor, College of Internet of Things and Logistics Engineering, Jinan University, No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province.

3Associate Professor, College of Intelligent Science and Engineering, Jinan University, No. 206, Qianshan Road, Xiangzhou District, Zhuhai City, Guangdong Province.

4Professor, Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong, China.

*Corresponding Author

ABSTRACT

This article takes a new mobile self-service method under the mobile Internet to "photograph and make money" as the research object. First, using the Surfer software, the contour map of the task distribution and price is drawn. Guess that the original pricing scheme is not only related to the latitude and longitude of the task distribution. It is also related to factors such as the degree of matching between members and tasks, the economic level of the city where the task is located, and the distance of tasks from the central city. The above conjecture was verified by B-P neural network. Finally, the Pearson coefficient is used to analyze the relevance of the task pricing, and the reasons for the failure to complete the main value theory and price theory are analyzed. It includes not considering the impact of the member's reputation value on its completion task, the remote task location the low pricing and the impact of non-monetary value of self-service labor crowdsourcing on members' completion of tasks, etc. This paper provides a theoretical basis and guiding significance for the pricing of self-service labor crowdsourcing under the mobile Internet in the future.

Keywords: B-P Neural Network, Crowdsourcing, Matching model, Surfer, Task pricing
1. INTRODUCTION

"Photographing to make money" is a self-service mode under the mobile Internet. The user downloads the APP, registers as a member of the APP, and then receives a task from the APP that requires photographing (such as going to the supermarket to check the availability of a certain item). And earns that the APP has calibrated for the task. This mobile Internet-based self-service labor crowdsourcing platform provides enterprises with various business inspections and information collection. Compared with traditional market research methods, it can greatly reduce the cost of investigation, and effectively ensure the authenticity of survey data and shorten the investigation cycle. Therefore, APP becomes the core of the platform operation, and the task pricing in APP is its core element. If the pricing is unreasonable, some tasks will be left unattended and lead to the failure of the check of goods. Since the emergence of self-service labor crowdsourcing under the mobile Internet, the academic community has conducted extensive research on it. But overall, the academic research on task pricing mainly focuses on the use of cluster analysis to study the impact of price and distance on task pricing. For example, Yang Lianwu used Cluster analysis and Excel to analyze the impact of task location and price on task pricing\cite{1}. Chao Shigang used q-Cluster Analysis and MATLAB to analyze the impact of task location, quantity and price on task pricing \cite{2}. Zhu Ke used MATLAB and Cluster Analysis to analyze the impact of task location and member location on task pricing\cite{3}. Or on the basis of the number of members and the intensity of the impact on the task pricing, such as Zhu Jiaming used K-Means clustering and Least squares to analyze the correlation between the number of members, task point distance, task pricing and task completion\cite{4}. Li Yuan used Cluster Analysis and generalized linear regression model analyzing from the density and distance of the task and number of members \cite{5}. Zhong Xiangjia used the Multiple Regression and Optimization Model to analyze the impact of the task location, member location, membership density, and number of members around the task on pricing\cite{6}. In fact, task pricing is influenced by many factors, not only related to task price, distance and number of members, but also affected by the reputation value of members and the non-monetary value of crowdsourcing such as member satisfaction and member perceived value\cite{7}. This paper analyzes the reasons for pricing failure by using the theory of value theory and price theory through the matching model of construction task and member distribution, validating it through B-P neural network to make up for the shortcomings of previous pricing theory, providing theoretical basis and guiding significance for the self-service labor crowdsourcing pricing in the future mobile Internet.

2. METHODS

2.1 Data acquisition and assumptions
The data in this paper is derived from the 2017 China Undergraduate Mathematical Modeling Contest B [8], which is the task data of a completed project, including the location, pricing and completion of each task, as well as member information data, including the location of the member, reputation value, reference to the task given by its reputation start booking time and booking limit, in principle, the higher the member's reputation, the higher the priority to start the selection task, the greater the quota (The task assignment is actually based on the proportion of the reservation limit). For ease of understanding and analysis, we assume that:

- Suppose the earth is an ideal sphere;
- Regardless of the impact of the altitude factor of the task location, the weather factor on the day of the task release, social factors such as political events on the day of the release on member orders.
- Assume that the user must complete the task after accepting the task, that is, there is no case of discarding the task after accepting the task;
- Assume that the data given by the title is true and reliable;
- Do not consider the impact of member subjective factors on the order;
- Most members do not regard the completion of tasks as the main source of life.

2.2 Drawing of price contour maps

In order to explore the relationship between the current task pricing and the geographical location, the Suffer software is used to draw the contour map of the position, including latitude and longitude.
Observing the above figure, we can find that there are multiple centers in the figure, guessing it is the location of the city center; each center is divergent, and the guessing task pricing is related to the location of the city center, but the radiation range of different centers is not concentric. It is suspected that it is interfered by the matching of the distribution of members and tasks. In summary, guessing the task pricing rules is related to the following factors: the location of the city, the distance from the city center point, and the matching of membership and task distribution. The appeal hypothesis will be verified one by one below.

2.3 Verification of the relationship between pricing and the location of the city

Observed from Figure 1, there are multiple centers, so this paper guesses that the task distribution exists in multiple cities. Use Google Maps to get the latitude and longitude of the task location in Appendix 1, as shown in Figure 2.
Among them, the yellow dot in the figure represents the latitude and longitude of the task position in Appendix 1. As can be seen from the figure, the tasks are basically concentrated in the four regions of Guangzhou, Shenzhen, Dongguan and Foshan in Guangdong Province, China.

K-Means clustering method is introduced here to cluster the task pricing: The basic idea of K-means clustering analysis method is to divide a large number of high-dimensional data points into l clusters according to their data characteristics, and extract clusters. The center point serves as a prototype for the data. Each prototype point is the most representative point in each cluster, and then l prototype point can represent all state information of the reconstructed signal. The basic calculation process \[^9\] is as follows:

1) In the reconstructed phase space, the vibration signal is reconstructed into k high-dimensional phase points as shown in equation (1). Initially determine the number of cluster centers l and the initial distribution of cluster centers;
2) According to the nearest distance principle, the phase points of the reconstructed signal are allocated to the corresponding cluster centers to form a new cluster;

3) Calculate the center of the newly generated clusters;

4) Repeat steps 2) and 3) until the position of the cluster center is stable and classify the phase points;

5) Calculate the distance $J(C_k)$ between each cluster center and the phase points belonging to the cluster, and accumulate the distance between each cluster center to get the overall distance $J(C)$:

$$J(C_k) = \sum_{x_i \in C_k} \| x_i - \mu_k \|^2$$

$$J(C) = \sum_{k=1}^{K} J(C_k)$$

6) According to the above steps calculated $k=1 \sim 8$ separately in the above steps, the overall distance $J(C)$ curve when $k$ increasing. As the $k$ increases, the phase points can always be grouped closer to the center of the cluster, and the overall distance will decrease. In general, the $k$ value at which the overall distance drop is mitigated can be selected as the number of cluster centers.

K-Means cluster analysis will divide it into four categories, and the final cluster centers of various types are shown in Table 1. The number of cases in each cluster is shown in Table 1.

### Table 1: Final cluster center

<table>
<thead>
<tr>
<th>cluster center</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task GPS latitude</td>
<td>22.665</td>
<td>23.423</td>
<td>22.957</td>
<td>23.061</td>
</tr>
<tr>
<td>Task GPS longitude</td>
<td>114.047</td>
<td>113.335</td>
<td>113.746</td>
<td>113.224</td>
</tr>
</tbody>
</table>

Table 1 shows the latitude and longitude coordinates of each center point after dividing into four categories. After consulting Google Maps, it is found to coincide with Guangzhou, Shenzhen, Dongguan, and Foshan in the geographical area.
Using Table 1, in combination with SPSS, the categories of the four categories are classified, and the scatter plot is plotted in Figure 3.

![Regional distribution Map](image)

**Fig. 3: The region obtained by K-Means cluster**

Figure 3 shows a visual representation of the regional representation of K-Means clustering. Among them, black represents the Shenzhen area, blue represents the Guangzhou area, red represents the Dongguan area, and green represents the Foshan area. Observing the above chart, we can see that the clustering results obtained by pricing have a one-to-one correspondence with the geographic administrative division. It can be judged that the location of the city does have an impact on the pricing of the task. Further analysis of the economic situation and task pricing of the four cities located, we can find that the task pricing is roughly positively related to the economic situation of the city.

2.4 Verification of the relationship between pricing and the distance of task distribution from the urban center and the matching degree between members and tasks.

i. Calculation of the distance of the task distribution from the urban center.

Access to information\[10\], the spherical distance between two points is

\[
l = r \times \arccos C_0, \quad \text{among them,}
\]

\[
C_0 = \sin(90 - LatA) \times \sin(90 - LatB) \times \cos(LonA - LonB) + \cos(90 - LatA) \times \cos(90 - LatB),
\]

\(l\) indicates the spherical distance between the two places, the longitude and latitude of point A is \((LonA, LatA)\), the longitude and latitude of point B is \((LonB, LatB)\), \(r\) is the radius of the
earth, the radius of the Earth is 6371.004 km. The partial spherical distances between the task and the urban center in Dongguan are shown in Table 2.

Table 2: The partial spherical distance between the task and the center in Dongguan

<table>
<thead>
<tr>
<th>Task number</th>
<th>Task GPS latitude</th>
<th>Task GPS longitude</th>
<th>Central longitude</th>
<th>Central latitude</th>
<th>L (km)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0070</td>
<td>22.7386</td>
<td>113.8184</td>
<td>113.7457</td>
<td>22.9568</td>
<td>25.3800</td>
</tr>
<tr>
<td>A0078</td>
<td>22.7386</td>
<td>113.8184</td>
<td>113.7457</td>
<td>22.9568</td>
<td>25.3800</td>
</tr>
<tr>
<td>A0125</td>
<td>22.7704</td>
<td>113.8551</td>
<td>113.7457</td>
<td>22.9568</td>
<td>23.5614</td>
</tr>
<tr>
<td>A0126</td>
<td>23.1677</td>
<td>113.6652</td>
<td>113.7457</td>
<td>22.9568</td>
<td>24.8581</td>
</tr>
<tr>
<td>A0133</td>
<td>23.1861</td>
<td>113.5975</td>
<td>113.7457</td>
<td>22.9568</td>
<td>29.6690</td>
</tr>
</tbody>
</table>

**ii** Calculation of the matching degree between members and tasks.

Looking at Figure 1, it is found that there is no divergence distribution of concentric circles around each center point, and the distribution of members and tasks are roughly coincident. It can be explained that the pricing law also refers to the matching degree between members and tasks[^11]. Here, the definition of the matching degree used in the original pricing in this paper is $\varphi_1$.

$$\varphi_1 = \frac{n_s}{d_{\text{min}}}$$

Among them, $\varphi_1$ indicates the matching degree used in the original pricing of the task, $d_{\text{min}}$ indicates the shortest distance of each task distribution corresponding to all member locations, $n_s$ indicates the number of members in the area of the circle centered on the task point and the radius of a certain distance. Referring to the statistics of China's population activity released by the National Bureau of Statistics[^12], set the distance to 2.4km.

Here is a list of the relationship between the number of surrounding members, the sum of quotas and the matching degree corresponding to the original pricing[^13], as shown in Table 3.
Table 3: The number of surrounding members, the sum of quotas, and the matching degree corresponding to the original pricing

<table>
<thead>
<tr>
<th>Task number</th>
<th>Nearest membership distance</th>
<th>Matching degree</th>
<th>Quotas</th>
<th>Number of surrounding members</th>
</tr>
</thead>
<tbody>
<tr>
<td>A0001</td>
<td>1.50400655</td>
<td>21.27650309</td>
<td>208</td>
<td>32</td>
</tr>
<tr>
<td>A0002</td>
<td>0.306524512</td>
<td>117.445746</td>
<td>68</td>
<td>36</td>
</tr>
<tr>
<td>A0003</td>
<td>0.497424585</td>
<td>86.44526493</td>
<td>231</td>
<td>43</td>
</tr>
<tr>
<td>A0004</td>
<td>1.084076606</td>
<td>1.844888073</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>A0005</td>
<td>0.298842935</td>
<td>204.120603</td>
<td>145</td>
<td>61</td>
</tr>
<tr>
<td>A0006</td>
<td>0.401977178</td>
<td>2.487703416</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 3 is the relationship between the number of surrounding members, the sum of quotas and the matching degree corresponding to the original pricing.

iii. Using B-P neural networks to verify the relationship between the pricing and the distance of the task distribution from the urban center and the matching degree between members and tasks.

The B-P Network, also known as the Back-Propagation Neural Network, continuously corrects the network weights and thresholds by training the sample data to make the error function fall in the negative gradient direction and approach the desired output. In this paper, it is conjectured that the impact factors related to pricing may be introduced into the model by using B-P Neural Network. A part of data is used to train the network to obtain prediction data about the pricing, and by comparing with the actual data, it is judged whether the pricing-related impact factors of the conjecture are reasonable.

The basic principles of prediction by using B-P Neural Networks include three parts: prediction model, rolling optimization and feedback correction \[14\]. As far as the basic principle of predictive control is concerned, the predicted model that satisfies the requirements can be extracted from the controlled object, it can be applied to any linear or nonlinear system. According to this point, this paper proposes a nonlinear predictive control algorithm, the basic framework of which is shown in Figure 4.
Fig. 4: Predictive control structure diagram based on B-P neural network

Prediction model as the B-P Network can reflect the nonlinear time-discrete system\textsuperscript{[14]}, the input and output signals known to the controlled object are used as the input signals of the network, and the output of the B-P Network is used as the future output of the predicted controlled object. A general nonlinear process can be described by the following time-discrete equation:

\begin{equation}
\begin{aligned}
y(k) &= f(y(k-1), \ldots, y(k-n), y(k-n), u(k-1), \ldots, u(k-m))\\
y(k+1) &= f(y(k), \ldots, y(k-n+1), u(k), \ldots, u(k-m+1))\\
&\vdots\\
y(k+p) &= f(y(k+p-1), \ldots, y(k+p-n), u(k+p-1), \ldots, u(k+p-m))
\end{aligned}
\end{equation}

The approximation model of equation (8) is established by using the B-P Neural Network, as shown in Fig. 5.
The p-step prediction model can be obtained by p-iteration of the one-step prediction model:

$$\begin{align*}
Y(k + p) & = F[y(k), \ldots, y(k - n + 1), u(k + p - 1), \ldots, u(k), \ldots, u(k + m - 1)]
\end{align*}$$

According to the B-P neural network model above, the location of the task, the distance of the task distribution from the urban center, the matching degree of the member and the task are introduced into the model. The four cities are respectively tested by the neural network, and only the Dongguan area is drawn here as shown in Figure 6.
Fig. 6: Training error map of Neural Network in Dongguan area

It can be seen from Fig. 6 that after 20 times of training, the best verification performance is 0.18682, which is within an acceptable range. It also shows that there is some invisible relationship between task pricing and project latitude and longitude, and the relative position of members and projects. Then, divide the data in the Dongguan area into two parts, 2/3 of the data is used as the training data, and the rest is used as the predicted data. The comparison between the predicted value and actual values is shown in Figure 7.

Fig. 7: Neural Network forecast and actual comparison of task price in Dongguan
Among them, the blue line indicates the predicted value and the red line indicates the actual value.

It can be seen from Figure 7 that the network predicted value of the task pricing is not much different from the actual value in the Dongguan area, and the difference between the overall predicted value and the actual value is about 5%. If a certain factor is considered in the neural network or the influence of a certain factor is considered less, the predicted value of the finally obtained neural network will be greatly different from the actual value [15]. The overall error of the current experimental results is within 5%, further illustrating that the factors considered in the search for the pricing law of Annex I are more reasonable, which is the location of the task, the distance of the task distribution from the urban center, the matching degree of the member and the task which are introduced into the model, will definitely affect the pricing of the task.

3. RESULTS AND DISCUSSION

The Pearson correlation coefficient is also called the Pearson product-moment correlation coefficient, which is a linear correlation coefficient. The Pearson correlation coefficient is a statistic used to reflect the linear correlation between two variables. The correlation coefficient is represented by r, where n is the sample size, which is the observed and mean of the two variables. r describes the degree of linear correlation between two variables. This paper uses the Pearson correlation coefficient to analyze the relationship of the location of the task, the distance between the task distribution from the city center point, the matching degree between the member and the task and the task pricing, from a qualitative point of view.

\[
    r = \frac{1}{n-1} \sum_{i=1}^{n} \left( \frac{X_i - \bar{X}}{s_X} \right) \left( \frac{Y_i - \hat{Y}}{s_Y} \right)
\]

(10)

3.1 Analysis of the correlation between task price and membership and task matching

The classification analysis of the four cities, the calculation of the correlation between the pricing and the degree of matching between the members and the project, limited by the length, only the correlation analysis of the listed cities in Dongguan is shown in Table 4.
Table 4: Correlation table of task pricing and member-task matching degree in Dongguan City

<table>
<thead>
<tr>
<th>Task pricing</th>
<th>Matching degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson pertinence</td>
<td>1</td>
</tr>
<tr>
<td>significance</td>
<td>.003</td>
</tr>
<tr>
<td>N</td>
<td>173</td>
</tr>
</tbody>
</table>

Matching degree

| Pearson pertinence | -0.207** |
| significance       | 1 |
| N                 | 173 | 173 |

**. When the confidence (double test) is 0.01, the correlation is significant.

Table 4 shows the correlation table between the task price and matching degree of Dongguan City. It can be seen from the table that the significance is 0.003 and the Pearson coefficient is negative, which indicates that the task pricing and the matching degree between members and tasks have significant negative correlation.

3.2 Correlation Analysis between Task Price and Task Distribution Distance from City Center Point

The classification of four cities, the calculation of the correlation between the task price and the distance from the center point of the task, limited by space, only the correlation analysis of Dongguan city is shown in Table 5.

Table 5: Correlation table of task pricing and the distance between task location and central point in Dongguan City

<table>
<thead>
<tr>
<th>Task pricing</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson pertinence</td>
<td>1</td>
</tr>
<tr>
<td>significance</td>
<td>.000</td>
</tr>
<tr>
<td>N</td>
<td>173</td>
</tr>
</tbody>
</table>

**. When the confidence (double test) is 0.01, the correlation is significant.
Table 5 shows the correlation table between the task pricing and the distance from the center point of Dongguan. From the table, it can be seen clearly that the significance is 0.000 and the Pearson coefficient is positive, which indicates that the task pricing and the task distribution which are far from the urban center point have significant positive correlation.

### 3.3 Pricing differences between cities

After classifying four cities, calculating the average pricing between different cities as shown in Table 6.

<table>
<thead>
<tr>
<th></th>
<th>Dongguan</th>
<th>Guangzhou</th>
<th>Shenzhen</th>
<th>Foshan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average pricing (RMB)</td>
<td>70.12</td>
<td>70.68</td>
<td>68.61</td>
<td>68.92</td>
</tr>
</tbody>
</table>

As can be seen from Table 6, there is still some degree of difference in the average pricing between different cities. From this, it is concluded that the pricing law of the region can be qualitatively described as the following aspects: the pricing of different cities will be affected by the living level of the region; the farther the task is from the center of the city, the higher the pricing; the membership and tasks the higher the match, the lower the price.

### 3.4 Analysis of unfinished causes

The distribution of the member's geographic location and unfinished tasks is plotted as a scatter plot\[16\], and a comprehensive consideration of the member's reputation value is shown in Figure 8.

![Distribution of members and unfinished tasks](image)

**Fig. 8: Distribution of members and unfinished tasks**
Among them, the red dot represents the unfinished task, and the black dot represents the member's geographic location. The larger the radius of the black dot, the higher the reputation value of the user.

Observing the image, it can be seen that in the region with a large black dot, that is, within a range where the user reputation value is high, there are fewer unfinished tasks in the region. The area with dense red dots has a small distribution of black spots. Considering comprehensively, the reason for inferring that the task is not completed can be divided into two aspects: price theory and value theory:

(i) Price theory:

In this question, the commodity in the crowdsourcing refers to the result of the member completing the crowdsourcing task; the currency performance of the labor condensation of the member completing the crowdsourcing task refers to the pricing that the member deserves. Only when the supply and demand balance between the two can the task be successfully completed, and the reasons for its incompleteness are as follows:

- The task distribution itself is relatively remote and the price is low. The time cost and labor cost required to complete the task cannot be obtained with the members, so the task is not received;

- When the pricing rules are formulated, the impact of the user's reputation value is not considered: because the higher the member's reputation, the higher the priority, the greater the quota. In other words, users with high credibility first choose the phenomenon that is closer to him and higher in price, so that users with relatively lower reputation can only choose among the tasks with far distance and low price, so that they are not willing to take orders, causing the task to fail.

(ii) Value theory:

The general value theory discusses the meaning of various things in the objective world for the survival and development of human beings, that is, whether something is worthwhile or not. Under this premise, crowdsourcing not only obtains the currency that reflects the value, but also includes the non-monetary value of the individual's psychological satisfaction, such as satisfying the psychology of hunting, learning new knowledge and new skills, and obtaining higher credit scores.

Therefore, some situations in which the task is accepted may be that the task distribution point
happens to be where the member is going, that is to say, the unfinished task may also be caused by the place where the member is going and the task point is not smooth.

4. CONCLUSIONS

This paper first uses the Surfer software to draw the contour map of the task pricing, and guesses the influencing factors of the task pricing according to the results of the contour map. The analysis of the task pricing may be related to the location of the task, the distance of the task distribution from the city center point, and the matching degree of the member and the task. Then, by studying the existing task pricing, constructing the matching model of the task and the member distribution, and verifying the conjecture through the B-P neural network. Finally, the Pearson coefficient is used to analyze the task pricing from a qualitative perspective. It is judged that there are two main reasons for the failure of the task pricing. First, from an economic point of view, that is, certain tasks are difficult to reach in the areas where members are difficult to reach, but the prices are low, which makes the task difficult to complete. Second, from a non-economic perspective, that is, from the commercial nature of crowdsourcing: most members complete tasks mostly with the attitude of entertainment and the like is completed by the way. If the task is distributed in an area with a small range of member activities, even if the price is high, the task may not be completed. It makes up for the shortcomings of the previous pricing theory, providing theoretical basis and guiding significance for the self-service labor crowdsourcing pricing under the mobile Internet in the future.

Due to space limitations, there are some shortcomings in the research process, such as not considering the difficulty of the task, the matching degree of the task's predetermined limit and demand, the existence of competitors and other complex factors on the unfinished task make that the reasons for the incomplete tasks are not comprehensive. Then the research in this paper makes up for the shortcomings of previous research on task pricing, laying a foundation for the pricing of crowdsourcing tasks in the future.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (51475095), the Fundamental Research Funds for the Central Universities (21618412), Key Project of Guangdong Natural Science Foundation (2016A030311041), 2015 Guangdong Special Support Scheme (2014TQ01X706), High-level Talent Scheme of Guangdong Education Department (2014-2016), the Guangdong Natural Science Foundation (2017A030313401).
REFERENCES


