Analysis of the Task Pricing based on Multiple Linear Regression Model for Photo-taking

Bang Zhang[1], Zhiyu Yang[1], Xinfang Zhao[2], Zhihong Ma[3], Cuiping Zhao[3]*
[1]College of Engineering and Technology, Tianjin Agricultural University, Tianjin, 300384, China
[2]College of Food Science and Biotechnology, Tianjin Agricultural University, Tianjin, 300384, China
[3]College of Basic Science, Tianjin Agricultural University, Tianjin, 300384, China
*Correspondent author: Cuiping Zhao

Abstract: In recent years, taking photos and making money has become popular in China. Pricing of tasks in this area has also become a hot topic of research. Therefore, this paper discusses the task pricing by using the multiple linear regression model which is packaged as the principle. The task of the multivariate linear model established by the residual test method is found: this model can reduce the corresponding task cost while improving the completion rate of the members.

Keywords: Task pricing; Nearby packing; Multiple liner regression model; Residual test method

1. Introduction

Taking pictures to make money is an important way to complete some research projects, which can save a lot of manpower, time and money for some research projects. Therefore, the pricing of taking pictures for money is also crucial. In recent years, many scholars have established a variety of pricing models to study the related pricing problems.

In order to solve the existing problems of pricing model, Peng Xiao[1] et al. proposed a virtual resource pricing model based on mixed game; Kun Zhou[2] et al used branch pricing model to study the pricing rules to reduce the operation cost of the aircraft; Targeting to the problem that pricing of public rental housing rent exists unreasonable problem in the formulation of the rent pricing standard , Yi-fei Cai[3] adopted a single multiple linear regression model to study it; Ya-ni Luo[4] et al proposed pricing model of multivariate regression method to study the stock; Ji-yan Zou[5] used option pricing method to study the pricing model of cost plus pricing; Yao-lin Zhang [6] et al used multiple linear regression model of pricing. In this paper, a multi linear pricing model based on the principle of near packaging is proposed to study the problem of task pricing.

2. The Establishment of the Model

2.1. Multivariable linear regression model

2.1.1. Basic idea

Multivariable linear regression mode, in real economic problems, a variable is often influenced by many variables. For example, family consumption expenditure is not only influenced by family disposable income, but also affected by many factors, such as family wealth, price level, interest rate of financial institutions and deposits.

2.1.2. Expression

The general form of the multivariate linear regression model is:

$$Y_i = b_0 + b_1 X_{1i} + b_2 X_{2i} + \ldots + b_k X_{ki} + m_i \ (i = 1, 2, \ldots, n)$$

“k” is the number to explain variables, $b_j (j = 1, 2, \ldots, k)$ is called the regression coefficient. The formula mentioned above is also called the random expression of the general regression function.

2.2. Establishment of task pricing model

2.2.1. Task data

According to the data provided by the 2017 Undergraduate Mathematical Contest in modeling, "photography for money making", we have completed 835 task data. In fact, multiple tasks may be chosen by users because of their relatively centralized location. One consideration is to combine these tasks together to package and release. Packaging ideas: take each completed task as the center to draw a circle of radius, and count the number of statistics of the regional task and memberships, and then get the average task limit of these memberships. Comparing the average limit with the task number ,if the number of tasks is less than the average limit, it indicates regional memberships can still take this packaged task; but if the number of tasks is more than the average limit, it reveals

ISSN: 2307-0692, Volume 6, Issue 6, December, 2017

that after these tasks being packaged, task completion rate is affected by the few number of the memberships accepting task in this region. So this paper shortens its radius to a certain distance, then repeats the above process until the number of tasks is less than the average limit, and then packages it into a task package. According to this theory, all unpackaged tasks are processed until all the tasks are packed.

The following steps are as follows:

Membership ship coordinates (x_{1j}, y_{1j}), Task coordinates (x_{1i}, y_{1i}).

The distance from each membership to each task:

$$r=\sqrt{\left(\left(x_{1j}-x_i\right)\times 82\right)^2+\left(\left(y_{1j}-y_i\right)\times 111\right)^2} \quad (1)$$

Memberships within the scope of each task（$r<l$）:

$$k=k+1;\ b_k=z_{1j} \quad (2)$$

Seek the average task limit of memberships within the specified scope:

$$a_1=\frac{1}{k}(b_1+b_2+\ldots+b_k) \quad (3)$$

use the formula (1) of distance, the distance from each task to each task 错误！未找到引用源。.

The number of tasks within the scope of each task（$r_1<l$）:

$$k_1=k_1+1 \quad (4)$$

Comparison of task number and average task limit:

If $k_1<a_1$, then the k_1 task can be packaged, and conversely, repeat the above steps, know $k_1<a_1$, and then pack it.

According to these steps, 835 tasks are packaged into 485 tasks with MATLAB programming.

2.2.2. Factors affecting the pricing of tasks

First, the intensity of the task. The more intensive the tasks are, the more attractive to the memberships in the specified area. Second, the average value of the membership's credibility. This article takes into account the interests of the high-quality memberships, so the higher the average value of the memberships' credibility is, the more the quality memberships in the region are. Third, the convenience of traffic. It can be imagined that the more convenient the traffic of location of tasks is, the more benefits to completion of the memberships, thus improving the completion rate. Fourth, the distance of each task from its nearest membership. The closer the distance between the membership and the task, the higher the task completion rate is. The SPSS is used to analyze the correlation between these factors and the pricing of completed tasks. The intensity of the task: bee; the average of the membership's credibility: average; the convenience of the traffic: traffic; the distance from the task to its nearest membership: dis. (Table 1 Correlation Analysis)

Table 1. Correlation Analysis

	Pricing						
	Pearson correlation	Significance (double tail)	Number	Self-service sampling "c"			
				Slg.	Std.error	95% confidence interval	
						Lower Bound	Upper Bound
Average	0.107*	0.019	485	0.001	0.049	0.012	0.201
Bee	0.194**	0	485	0.001	0.048	0.109	0.295
Traffic	-0.094*	0.038	485	-0.007	0.048	-0.19	-0.005
Dis	0.107*	0.019	485	0.001	0.049	0.012	0.201

2.2.3. Find a solution about the factors of task pricing

Member coordinates (x_{1j}, y_{1j}), task coordinates (x_{1i}, y_{1i})

The intensity of the task

Use the formula (1) of distance to solve the distance from one task to another task r_1.

The number of tasks within each mandate ($r<l$):

$$k=k+1 \quad (5)$$

Area of the specified scope:

$$s=p_i\times l^2 \quad (6)$$

The intensity:

$$b_i=\frac{k}{s} \quad (7)$$

The average value of membership in the circle

Use the formula (1) of distance to solve the distance r_1 from one task to anther task.

Use the formula (2) of distance to solve the number K of membership within the limited range of task.

Use the formula (3) of distance to solve the average value a_1 of membership credibility within the prescribed scope.

(3)The convenience of transportation

Table 2. GPS of Downtown of five Cities

Symbol	GPS	Shenzhen	Dongguan	Conghua	Foshan	Guangzhou
x_{1j}	North latitude	114.06	23.02	23.33	23.02	23.08
y_{1j}	East longitude	22.61	113.45	113.33	113.06	113.14

Substitute the coordinates of the five cities into the formula (1) of distance to find the distance from one task to another task.

Find the shortest distance: $d = min(r)$

The distance from each task to its nearest member

Use the formula (1) of distance to find the distance from one task to another task.

Find the shortest distance: $a = min(r)$

3. Solution of the Model

The pricing Y and the intensity of tasks X_1, the average value X_2 of the credibility of the round members of the circle with the mission, the convenience X_3 of transportation, and the distance X_4 from each task to its nearest member have a great relation. The function in use fit the average value of the price Y and the task, the average value of the member's credit value in the fixed area, the convenience of traffic and the distance of the nearest member of each task.

$$Y = 93.5006 + 199.2712X_1 + 2.2252X_2 - 0.3289X_3 - 6.9243X_4 \quad (8)$$

4. Test of the Model

In order to verify the match between the data fit and real value, this paper USES an objective test method-residual test to test the model. The specific steps of the inspection are as follows:

The mean of the original sequence is calculated as follows: $y^{-(0)} = \frac{1}{n}\sum_{i=1}^{n} y^{(0)}(i) = 68.89$

Calculate the mean variance of the original data $y^{(0)}(i)$:

$$S_0 = \sqrt{\frac{S_0^2}{n-1}} = 2.6615$$

Among them: $S_0^2 = \sum_{i=1}^{n}[y^{(0)}(i) - y^{-(0)}]^2 = 701.29$

Calculate the mean of the residual $e^{(0)}$: $e^{-(0)} = 0.2129$

The mean variance of residual calculation is: $S_1 = \sqrt{\frac{S_1^2}{n-1}} = 0.4637$

Among them: $S_1^2 = \sum_{i=1}^{n}[e^{(0)}(i) - e^{-(0)}]^2 = 21.2873$

To calculate the accuracy of the model: $c = \frac{s_1}{s_0} = 0.1742$

Inspection: the inspection provisions of residual fitting precision grade table (see Table 2).

Table 3. Fitting Precision Grade Table

c value	The fitting precision
< 0.35	Good
< 0.50	Qualified
< 0.65	Barely qualified
> 0.65	Unqualified

Conclusion: based on the above prediction precision scale, $c = 0.1742 < 0.35$, the accuracy of fitting is shown in this paper, which indicates that the accuracy of fitting is good.

5. The Application of the Model

5.1. Factors that affect pricing

According to the data provided by the 2017 college students' mathematical modeling B contest "taking photos to earn money", the new project task data, a total of 2,066 tasks, too many tasks, and some distance too close, the mapping of the geographical location map is too large, only to intercept the distribution of the tasks of a city in Dongguan. (See Figure 1)

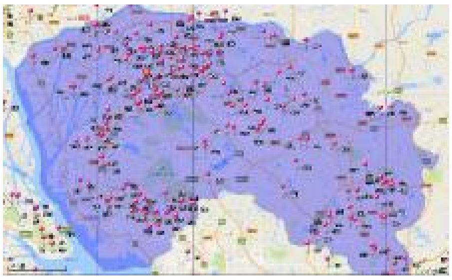

Figure 2. The Actual Location of the Task

Because there are so many tasks, pack it first and pack the 2,066 tasks into 667 tasks with the packaging model above.

5.2. Factors for solving task pricing

The 667 task data were solved through the above factors, and the data of the four factors were obtained, because the data was too large and only one part was displayed. (See Table 4).

Table 3. Data of Task Pricing Factors

Bee	0.840	0.649	0.688	0.802	1.006	0.013	0.980	0.955	0.649	0.955
Average	0.495	0.460	0.495	0.601	0.566	0.141	0.672	0.707	0.460	0.601
Traffic	21.456	27.396	19.787	26.198	22.770	36.759	24.452	24.428	20.432	23.764
Dis	0.495	0.460	0.495	0.601	0.566	0.141	0.672	0.707	0.460	0.601

5.3. Solution of pricing

Substitute the intensity of the task X_1 and the average value of the membership's credibility X_2, the convenience of traffic X_3, the distance of each task from its nearest membership X_4 of the data into the distance formula (8) and find out the new task pricing. Because 667 tasks are priced too much, only one part of the task is priced. (See Table 5)

Table 5. Mission Pricing

Mission number	Mission latitude of GPS	Mission longitude of GPS	Mission pricing	Mission number	Mission latitude of GPS	Mission longitude of GPS	Mission pricing
C0001	22.73	114.24	235.86	C0011	22.65	114.34	64.09
C0002	22.72	114.29	208.52	C0012	22.71	114.24	230.08
C0003	22.70	114.23	222.72	C0013	22.72	114.29	207.36
C0004	22.73	114.28	227.21	C0014	22.74	114.28	203.19
C0005	22.71	114.25	272.57	C0015	22.74	114.27	240.48
C0006	22.75	114.38	74.50	C0016	22.72	114.26	252.67
C0007	22.72	114.27	269.33	C0017	22.75	114.28	184.79
C0008	22.71	114.27	261.64	C0018	22.75	114.27	227.58
C0009	22.73	114.23	201.44	C0019	22.73	114.29	196.93
C0010	22.71	114.26	262.90	C0020	22.73	114.28	208.63

6 . Conclusions

Based on the traditional multiple linear regression models, on the basis of analysis and research, this paper proposes a clustering in nearby packaged as a principle of multivariate linear regression model. Residual test by this model, it is concluded that this model can not only improve the completion of tasks, but also can save costs. Therefore, the multivariate linear regression model established in this article can be widely applied to the task pricing problem.

References

[1] Peng Xiao , Zhi-gang Hu. Virtual resource pricing model based on mixed game in cloud environment [J]. Computer integrated manufacturing system, 2014,20(01):198-206.

[2] Kun Zhou, Hong-shan Xia. Aircraft scheduling model based on cooperative multiple task allocation and algorithm [J]. Journal of aviation, 2011, 32 (12) : 2293-2302.

[3] Yi-fei Cai. Taiyuan city public rental housing rent pricing research [D]. Shanxi University of finance and economics, 2013.

[4] Ya-ni Luo, Yong-jie Guo. The application of multiple linear regression models in stock pricing [J]. Accounting communication (finance), 2008,(09):26-28.

[5] Ji-yan Zou. China's housing accumulation fund loan guarantees loan pricing research [D]. Hunan University, 2012.

[6] Yao-lin Zhang, Li Shuping. Mergers and acquisitions of listed companies financial risk analysis and control [J], modern business, 2011 (32) : 228-229.