

Research on Task Pricing of Crowdsourcing Platform

 Senrong Ma¹, Lei Wang¹, Jiaen Guo¹ and Gang Zhou^{2,*}
¹Students' Brigade, Naval Aviation University, Yantai Shandong, 264001, China

²School of Basic Sciences for Aviation, Naval Aviation University, Yantai Shandong, 264001, China

*Corresponding author

Abstract—With the development of the science and technology, a new business model for publishing tasks over the internet has emerged. This paper mainly focuses on the task completion of crowdsourcing platform and pricing scheme. By packaging tasks centrally, we establish a task pricing model based on partial least square regression and partial binary tree support vector machine. Traditional pricing schemes only consider the benefits of the platform but neglect the benefits of the users. Thus, it has the problem of low completion rate and dispersion of tasks. Comparing with the traditional pricing scheme, this model can not only improve the task completion rate but also maximize the benefits of both the task platform and the user. Besides, the model can avoid the problem of user congestion and make the model more secure and reliable.

Keywords—partial binary tree support vector machine; genetic algorithm; crowdsourcing platform

I. INTRODUCTION

With the rapid development of mobile internet technology, more and more enterprises choose to collect a wide range of data offline confirmation and collection work through the internet self-help platform. Through the platform, users receive and release the task to earn a reward. Different from the common pricing project, the task pricing of the internet crowdsourcing platform not only involves a large amount of data, covers a wide range of areas, and also has complex influence factors. Besides, it also needs to consider the problem of maximizing the interests of both task arranger and user. However, most of the traditional pricing model use profit optimization model and establish the correlation function to solve the problem. These models only take the maximizing profits of the merchant into account, but they neglect the users' idea. Therefore, the traditional model can't solve the bidirectional benefit maximization pricing demand based on big data platform. In order to adapt to the change of the task pricing, combining with the specific task needs, we present a pricing scheme based on partial least square regression and partial binary tree support vector machine model.

II. THE RULE OF TASK PRICING

A. Setting up Partial Least Square Regression Model

We consider the modeling of dependent variable is y_1 and x_1, x_2, \dots, x_m are individual independent variables, and finally establish the partial least square regression model $y_1 = b_1x_1 + b_2x_2 + \dots + b_mx_m$.

According to the reality of the subject, the setting of the independent variable is as follows:

TABLE I. INDEPENDENT VARIABLE DESCRIPTION

x_1	Mission location latitude
x_2	Mission location longitude
x_3	Member location latitude
x_4	Longitude of member location
x_5	Member credit value

Note that F_0 is a standardized variable for variable task pricing y and E_0 is the normalized matrix of the set of independent variables X .

$$F_{0i} = \frac{y_i - \bar{y}}{s_y}, i = 1, 2, \dots, L, \quad n \quad (1)$$

$$E_{ij} = \frac{X_{ij} - \bar{X}_j}{s_j}, j = 1, 2, \dots, L, \quad p \quad (2)$$

First, a component u_1 extracted from $u_1 = F_0 c_1, \|c_1\| = 1$, is the first spindle of F_0 , extract t_1 from F_0 , w_1 is the first spindle of E_0 , with the constraints of $\|w_1\| = 1, \|c_1\| = 1$ to find the maximum value of $w_1^T E_0^T F_0 c_1$, using the Lagrange algorithm to remember:

$$\text{Cov}(t_1, u_1) = \sqrt{\text{Var}(t_1) \text{Var}(u_1)} r(t_1, u_1) \rightarrow \max, \quad (3)$$

$$s = w_1^T E_0^T F_0 c_1 - \lambda_1 (w_1^T w_1 - 1) - \lambda_2 (c_1^T c_1 - 1) \rightarrow \max. \quad (4)$$

s is the partial respectively derivative of $w_1, c_1, \lambda_1, \lambda_2$, and then we make it zero, through derivation, it is possible to:

$$\left. \begin{aligned} E_0^T F_0 F_0^T E_0 w_1 &= \theta_1^2 w_1 \\ F_0^T E_0 E_0^T F_0 c_1 &= \theta_1^2 c_1 \end{aligned} \right\} \theta_1 = 2\lambda_1 = 2\lambda_2 = w_1^T E_0^T F_0 c_1, \quad (5)$$

w_1 is a unit eigenvector corresponding to the eigenvalue matrix $E_0^T F_0 F_0^T E_0$. On the other hand, c_1 is the unit eigenvector of θ_1^2 . θ_1^2 is the maximum value of the matrix of $E_0^T F_0 F_0^T E_0$. Then we get spindle w_1 and c_1 ,

Then we obtain the two regression equations of E_0 and F_0 for t_1 and then the regression coefficient vector is:

$$F_0 = t_1 r_1^T + F_1, \quad E_0 = t_1 p_1^T + E_1, \quad (6)$$

$$p_1 = \frac{E_0^T t_1}{\|t_1\|^2}, \quad r_1 = \frac{F_0^T t_1}{\|t_1\|^2}. \quad (7)$$

E_1 and F_1 are the residual matrices of the two regression equations, using the residuals matrix E_1 and F_1 instead of E_0 or F_0 . The same method is used to calculate the second axes w_2 , c_2 and the second principal components. If the m operation is performed, there will be:

$$E_0 = t_1 p_1^T + t_2 p_2^T + \dots + t_m p_m^T + E_m, \quad (8)$$

$$F_0 = t_1 r_1^T + t_2 r_2^T + \dots + t_m r_m^T + F_m. \quad (9)$$

III. APPLICATION OF MODEL

According to the thought steps of partial least square regression algorithm, we obtain the regression equation of task pricing:

$$y_1 = -0.0034x_1 - 0.0606x_2 - 0.0169x_3 + 0.0245x_4 - 0.0760x_5. \quad (10)$$

TABLE II. VARIABLE CORRELATION COEFFICIENT MATRIX

	x_1	x_2	x_3	x_4	x_5	y_1
x_1	1.0000	-0.9795	-0.0862	0.0764	-0.2262	0.0694
x_2	-0.9795	1.0000	0.0862	-0.0815	0.2528	-0.0975
x_3	-0.0862	0.0862	1.0000	-0.7378	0.0903	-0.0474
x_4	0.0764	-0.0815	-0.7378	1.0000	-0.1909	0.0539
x_5	-0.2262	0.2528	0.0903	-0.1909	1.0000	-0.0838
y_1	0.0694	-0.0975	-0.0474	0.0539	-0.0838	1.0000

According to the algorithm step, we obtain the final result of the task pricing through the classification of classifier. Set up scatter diagram for the post pricing task, and the distribution of the pricing results as shown in figure 1 and figure 2.

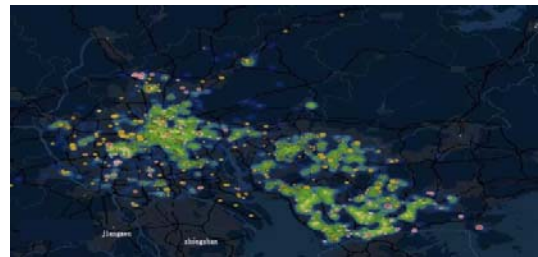


FIGURE I. COMPLETION OF TASKS



FIGURE II. TASK COMPLETION QUANTIZING

From the above picture, we can also get the analysis of the reasons for the unfinished task: from the distribution of scatter points, the correlation between the completion of the task and the bid price is relatively large, and the correlation with the member distribution is relatively weak.

IV. ESTABLISHMENT OF PARTIAL BINARY TREE SUPPORT VECTOR MACHINE

The idea of this method is to use partial two tree structure, a class of samples is extracted as the first classification at every time, thus, the classification of k problem is transformed into the classification of $k - 1$ problem, after that, we use the double

support vector machine to implement classification for each node. The algorithm is divided into four steps:

1. Initialize the data.
2. Generate binary tree structure.
3. TVSM training.
4. Test pricing results.

V. APPLICATION OF MODEL

For the specific pricing items, we use the partial binary tree support vector machine to make scheme. The specific member information and the task position are shown in the attachment.

Set the distance standard value between the task point and the member position as r_0 , the distance between the task point and the member position is r , set a target function $F(n)$ which is the ratio of the distance between the task point and the member position and the standard distance:

$$F(n) = \frac{r}{r_0} \quad (11)$$

Use $F(n)$ to represent the fitness value between each task and the members who can accept the task. $T(n)$ is the maximum value used to store the tasks accepted by each member.

The specific steps of implementing the algorithm are as follows:

1. Generate the first member who accept the task, and calculate the fitness of each task to the member and the maximum limit value of the task.

2. Set the probability of each task selected by a member to be P_i ,

$$P_i = \frac{F(i)}{\sum_{i=1}^n F(i)} \quad (12)$$

3. Use the roulette method to make members choose tasks randomly.

4. Randomly select a pair of tasks and members to cross the operation to calculate the value of the fitness value and the number of tasks for the members.

5. Update the task information. Mark the completed tasks as 0, meanwhile the uncompleted tasks are set to 1, after that carryout the cross operation between the tasks and the members.

6. Select the largest fitness value and compare it with the fitness of the current optimal solution. If the fitness value of the optimal solution is greater than the optimal solution, replace the

current optimal solution, and update the membership information and the number of accumulated.

7. Loop iteration, when the number of times reaches the setting maximum, jump out of the loop, outputs the optimal solution.

The block diagram of the specific procedure is as follows:

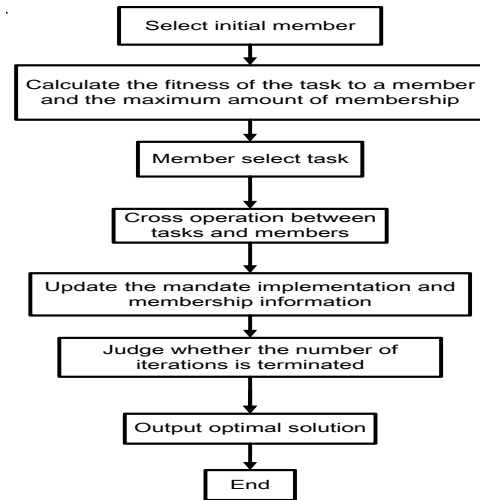


FIGURE III. BLOCK DIAGRAM

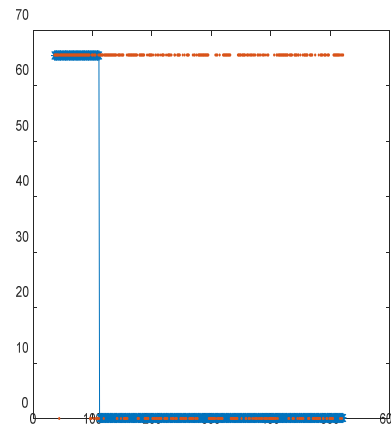


FIGURE IV. TEST RESULT

A. Test Classification Results

As 4 shown in the figure, the test graph intuitively gives the pricing results after the first SVM classifier. The upper blue dot set represents the priced task, the orange dot set represents the task pricing situation after being classified by the SVM classifier, and the horizontal coordinates represent the numerical code of the task label.

According to the algorithm step, we obtain the final result of the task pricing through the classification of classifier. Set up scatter diagram for the post pricing task, and the distribution of the pricing results as shown in figure 5. By calling the function of the SVM toolbox in MATLAB, the sample is classified according to appendix. After the classification, the tasks of

different pricing are distinguished by different colors, and the scatter diagram is drawn in figure 6.

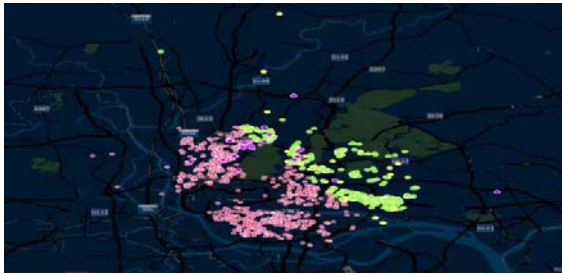


FIGURE V. TASK PRICING SCATTER PLOT

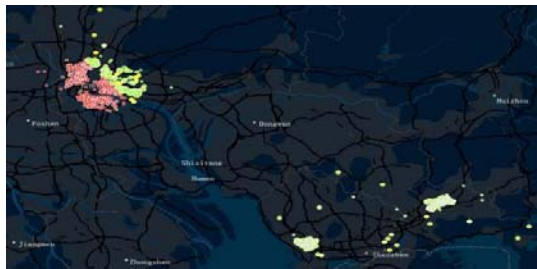


FIGURE VI. PRICING RESULTS SCATTER PLOT

B. SPSS Correlation Test

Correlate the obtained pricing scheme with the argument set consisting of all arguments to obtain the correlation test results as shown in Table 3.

TABLE III. CORRELATION TEST

		argument	
		price	
price	Pearson correlation	1	.823
	Significant (bilateral)		.796
	N	2067	2067
argument	Pearson correlation	.823	1
	Significant (bilateral)	.796	
	N	2067	2067

The test results show that the final pricing is significantly correlated with the set of independent variables at 0.05 level, and the Pearson correlation coefficient is 0.823, and result verifies the feasibility of the scheme.

VI. CONCLUSION

By using the partial least square regression and partial binary tree support vector machine model, this paper presents a new secure and reliable pricing scheme based on the service crowdsourcing business model on the internet, which can better adapt to the complexity of crowdsourcing business data and the complexity of the region. Meanwhile, this model can improve the task completion rate and make platform and user reach profit

maximization. Besides, the model has good practicability and generalization. Through the adjustment and expansion of project parameters, the model can also complete the pricing of other crowdsourcing projects. We hope that the model can be used more in the future research.

REFERENCES

[1] Shoukui Si. Mathematical modeling algorithm and application [M]. Beijing: national defense industry press, 2011:39-63.
 [2] Bei lei Qin, Application of partial least squares regression in hydrological analysis [J]. Journal of Sichuan University (Engineering Science Edition). 2003:116-117.
 [3] Juan ying Xie, Zhang Bingquan, Wang Wan-z, Multiclass classification algorithm based on double support vector machine. Journal of Nanjing University (Natural Science) 2011:355-360.