An Improved Pricing of Taking Photos and Making Money Model Based on SVM and Clustering

Yujie Ren

Department of Automation, North China Electric Power University, Baoding 071000, China.

Abstract. The aim of this study is to improve the imbalance pricing problems of Self-service package platform based on mobile Internet. The primary emphasis in this paper is on improving the completion of the to-do task. Based on the pricing established by factor analysis and regression equation, the author adopts the idea of packing and clustering. An improved support vector machine and model identification are applied to optimize the fitting function of pricing. Experiments show that the new pricing model greatly improves the mission completion degree and reduces the operation cost of the enterprise. Such an improved pricing model has a good practical guiding significance.

Keywords: support vector; machine model; identification packing clustering.

INTRODUCTION

"Take photos and making money" is a self-service mode under the mobile Internet. The user downloads APP, registers as a member of APP, and then receives the task that needs to be photographed from APP (for example, users go to the supermarket to check the situation of a commodity's shelves) [1]. Meanwhile, the user earns APP's demarcated reward for the task. When multiple tasks are concentrated, it is more convenient for the user to perform several tasks at the same time. If the price of the task is constant, some more dispersed tasks will be more likely not to be completed. The author has compared the pricing of 835 completed and incomplete tasks. The author gets the factors of the correct pricing of tasks and the correlation degree. The improved model we applied is referred to as the vector machine learn the features of the completed task, pack and publish a centralized task. Later, clustering analysis of task points with Q type cluster analysis is one of the key steps. Our idea is to pack the centralized task at the same time, the total price is multiplied by a discount factor α and generate a new table of task variables. Another key step is support vector machine (SVM) to find the best discount factor by model identification based on the completed samples. This model can improve the quota of tasks and the degree of completion. Finally, the author improves the rationality of pricing model. The improved model has widely ductility and applicability.

FACTOR ANALYSIS

In order to measure the influence factors of pricing model:
Set:
x1: the distance between the members closest to a task.
x 2: the number of members close to the task (A distance parameter is determined. The number of members in each area near each task is determined by the circle delimitation range of each task location and the radius of **r.**)
x3: the reputation of the task (The average credibility of each member in this range.)
x4: the quota of the task (The average quota of each member in this range.)
After normalization of the 4 factors, we get standardized index vectors. The author calculates the elementary load matrix and the contribution rate of factors. We found that x1, x2 can be the most important factor for pricing rationality.

Contribution rate
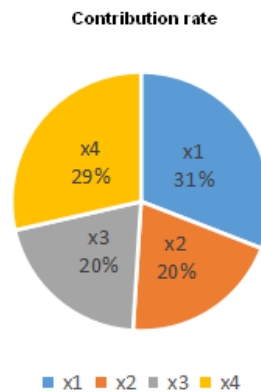
FIG 1. The Ratio of Each Independent Variable

Partial Least Squares Regression

Based on the regression equation of dependent variable and independent variable, a linear relationship is established. We use the regression function of completed tasks as the pricing function of the whole tasks. Then we get a new pricing model. [2]

$$y = 138.9672\ x_1 - 0.0427\ x_2 - 0.0020\ x_3 + 0.0146\ x_4 + 69.3633 \tag{1}$$

Among them, x_1, x_2, x_3, x_4 is the four factor we set before, y is the pricing of task.

It can be seen that all kinds of factors affect the weight of pricing factors. The distance between the task and the members is positively correlated with the price y. The number of people around the task is negatively related to the pricing y. The average reputation of the members around the task and the average quota of each member around the task are negatively correlated with the price y.

Conclusion:

The farther the member from the task is, the higher the price is.

The more the average number around the task, the greater the probability that the task will be completed, and the lower the price.

The smaller impact of a member's human factors on the pricing itself. But the better the member is, (The higher the reputation and reputation of the task) lower the price can be, so as to reduce the cost.

SUPPORT VECTOR MACHINE, PACKING AND CLUSTERING

Cluster Analysis

Clustering of samples by Q type cluster analysis, we calculate the degree of similarity by Euclidean calculation of coordinate distance. [3]

$$d(x, y) = [\sum_{k=1}^{p} |x_k - y_k|^2]^{\frac{1}{2}} \tag{2}$$

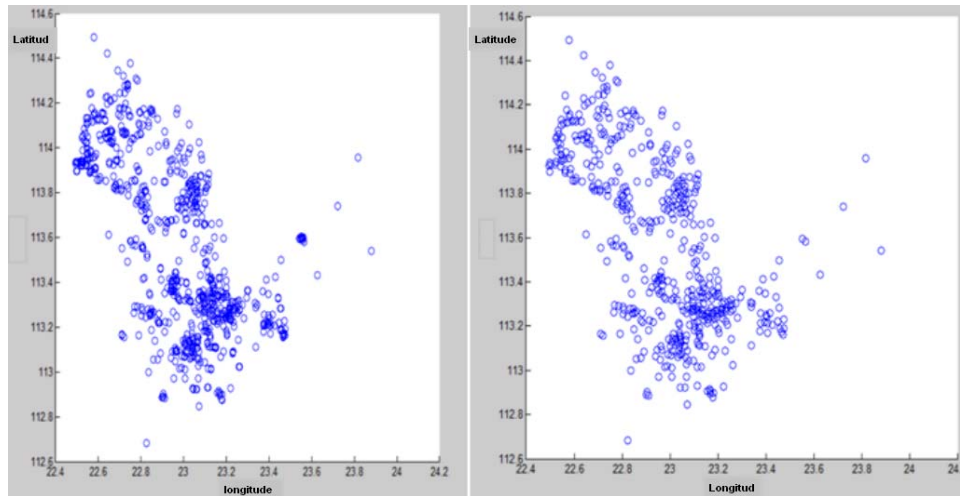The task point package is divided into 500 packages to reduce the number of tasks.

FIG 2. The Distribution of Samples (L is before Package, R is Post Package)

Generating the Task Variable Table with Post Package

After clustering analysis, the average longitude and latitude in the package is recalculated. The new latitude and longitude are used as the center latitude and longitude of the task package. The rest of the indicators, like classification function, are also recalculated after packaging [4] [5]. We calibrate the new influencing factors for the new packing task. That are x1, x2, x3, x4, x5, x6. A special description of x5, x6:

x5: fitting pricing of packing task, the calculation formula is as follows

$$p' = \alpha p \tag{3}$$

α is a discount factor, and p is the average pricing for the package.

$$p = \frac{\sum_{i=1}^{n} p_i}{n}, i = 1, 2, \dots, n \tag{4}$$

x6: task volume in a package
Summarize the above information to get a new data table:

TABLE 1. The New Variable Table after Packaging

Tab of package	x_1	x_2	x_3	x_4	x_5	x_6
1	0.0012	39	44.7676	3.1282	69.0664	1
2	0.0011	52	69.1296	3.5385	67.8647	1
3	0.0013	44	511.3565	19.6818	67.5054	1
4	0.0035	44	502.9087	19.3636	67.5534	2
5	0.0035	44	502.9087	19.3636	67.5534	2
…	…	…	…	…	…	…

Training Support Vector Machine

The classification function of the vector machine:

$$c(\tilde{x}) = \sum_{i \in I} \beta_i K(b_i, \tilde{x}) + b \qquad (5)$$

β is the weight coefficient of the classified function, $\beta_i = a_i y_i$

Determine the task to be completed: $c(\tilde{x}) \geq 0$, $\tilde{x} = 1$

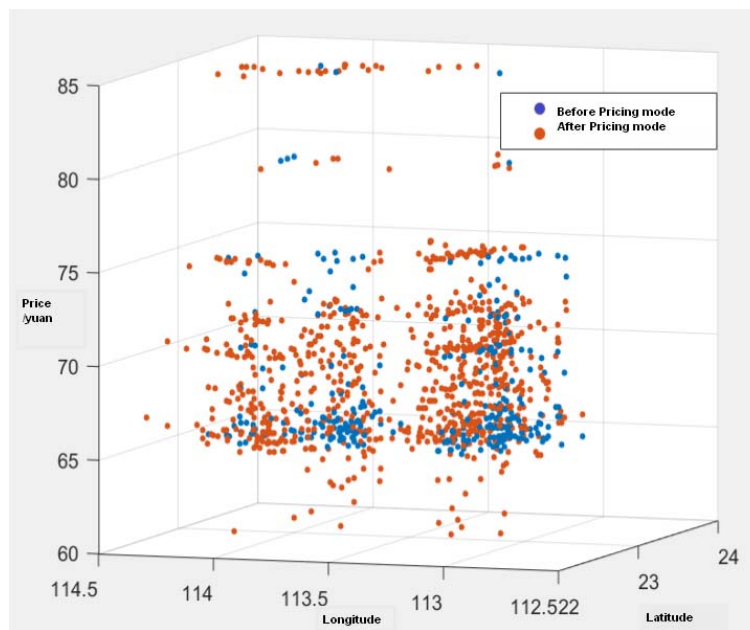Determine the task not to be completed: $c(\tilde{x}) < 0$, $\tilde{x} = 0$

FIG 3. Comparison of Pricing Modle before and After

It can be seen that the improved pricing modle is more discretized and more consistent. The members and tasks of different geographical locations make the pricing modle more reasonable

Determine the Best Discount Factor

Support vector machine recognition for convenience, at the time of recognition, the total price of packaging task is divided into n before synthesis. The recognition function trained by vector machine determines whether the task after re pricing is completed. Draw different discount factor α.P-Q as an index, the higher the difference is, the more the number of tasks is completed, but it is saturated.

P: The number of tasks completed

Q: The number of tasks uncompleted

In the selection of the kernel function of the vector machine, the error rate of selecting the Gauss kernel first is higher, the standard difference function of the support vector is checked and the linearity is higher, so the linear support vector machine is used. [6] Error rate ee_rate reduced to 0.03, with obvious improve.
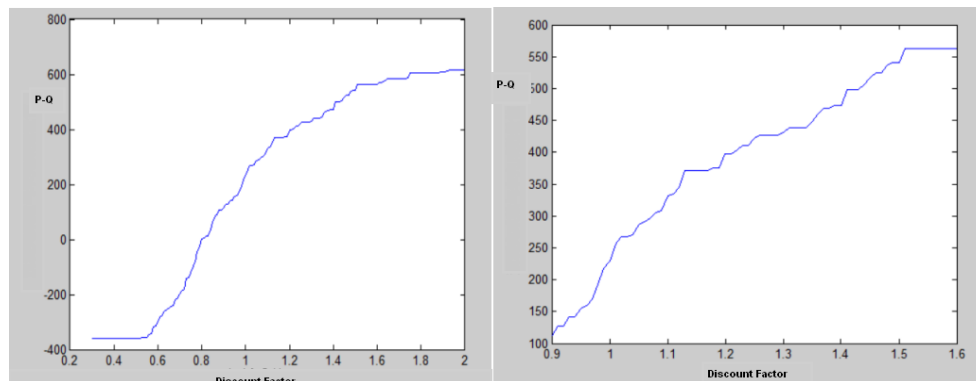
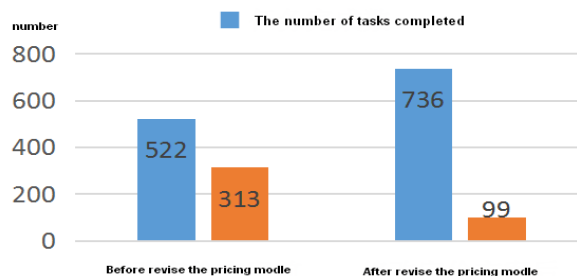**FIG 4.** The Trend Chart of P-Q Index with α and its Partial Enlargement Map

CONCLUSION

FIG 5. Comparison of Task Completion after Packaging

α is 1, the price is priced at the original packing price. It can be found that the difference is near 200, which indicates that the pricing effect of the original scheme is generally effective under the condition of packing. When α is in 1.2-1.4, the proportion of task completion is significantly improved, and the discount coefficient is significantly corrected. If we want to have more tasks to fulfill, we must raise the price appropriately if we have cluster packing. If we want to save the cost of pricing, the number of tasks will be reduced. The choice depends on the different development stages and goals of the company. The relationship between them is given here, which can provide a reference for the company's next goal. This model can add other factors with higher correlation, such as the wage level in the concentrated area, the convenient traffic situation, and so on, with higher ductility.

REFERENCES

1. Two-Sided Markets. Armstrong, M, Wright, J. The New Palgrave Dictionary of Economics. 2008.
2. Shoukui Si, Xijing Sun. Modeling of mathematics and application. Beijing: National Defense Industry Press, 2011.
3. Yu Li. Q Clustering algorithm based on Bayes Method. [J]. Computer and Digital Engineering, 2007 (07): 16-17+7.
4. Wu Qing. Research on support vector machine learning algorithm based on optimization theory [D]. Xi'an Electronic and Science University, 2009.
5. Chih-Wei Hsu, Chih-Jen Lin A Simple Decomposition Method for Support Vector Machines [J]. Machine Learning. 2002 (1-3).
6. A Tutorial on Support Vector Machines for Pattern Recognition [J]. Christopher J.C. Burges. Data Mining and Knowledge Discovery. 1998 (2).