

Improved Logistic Regression Precision Marketing and Decision System Model Based on Rough Set

Yufeng Yang ^{1,*}, Xue Xia ² and Fuheng Wang ³

¹ School of Mathematics and Information Science, Hebei University, 071002, China;

² School of Management, University of Shanghai for Science and Technology, 200442, China;

³ School of Energy and Power Engineering, Lanzhou University of Technology, 730050, China;

* Corresponding Author

Abstract

Combined with the basic behavior information of users, we select the network activity index, online shopping index, online video index, travel index and financial management index as the basic characteristics of user behavior. The index values of the users who have purchased the mobile phone are screened out. Since there is basically no collinearity among the indexes, these indexes of the users who have purchased the mobile phone may have convergence. The greater the convergence, the more significant the influence of this index will be. We use the square difference analysis method to select the indicators. Taking whether the user buys the mobile phone as the dependent variable and the screened index as the independent variable, a binary logistic regression model is established, and the functional relationship between whether the user buys the mobile phone and the user's basic behavior characteristics is obtained. In order to explore the specific impact of these indicators, one of them changed slightly each time while the other indicators remained unchanged. The regression values before and after the change were compared to get the factor influence rate of each indicator. The larger the factor influence rate is, the greater the influence of this index on whether users buy this mobile phone. Finally, we get that the indexes with higher factor influence rate are online shopping index and travel index.

Keywords

Precision marketing, Information management, Logistic regression, Analysis of variance, Engineering mathematics.

1. Introduction

In today's digital information age, traditional marketing means can no longer meet the needs of the rapid development of enterprises. Faced with the emerging opportunities and increasingly fierce competition, new marketing means emerge as The Times require. Among them, precision marketing based on big data has been paid more and more attention.

Precision marketing is based on precise positioning, relying on modern information technology means to establish a personalized customer communication service system, to achieve measurable low-cost expansion of the road, is one of the core ideas of the attitude of the network marketing concept.

At this stage, mobile Internet users are showing a steady growth trend overall. Such a huge mobile Internet user group and steady growth momentum provide a guarantee for the number of users of mobile advertising, especially the development of mobile advertising in the background, and also indicates the huge potential of China's mobile advertising market in the future. With the rapid development of e-commerce and mobile payment, mobile phones have

become an indispensable tool in people's lives and work. Therefore, what kind of mobile phones to choose has become an issue that consumers pay attention to.

The sales headquarters of a certain brand of mobile phone wants to know consumers' purchase intention of this mobile phone so as to be able to carry out precise marketing. To this end, the Marketing Department conducted a relevant investigation and obtained the attached data. In order to carry out accurate marketing of the company's mobile phones, we need to establish a mathematical model to solve the following problems:

Preprocess the collected data and conduct descriptive statistical analysis.

Among the target users, some users purchased the mobile phone during the research period, but more users did not purchase it. As a sales department, you really want to know if the basic behavioral characteristics of the user have any impact? And analyze exactly how to affect.

Different network concerns will reflect different personal preferences of mobile phone consumers, resulting in different main motives for everyone to buy mobile phones. Different mobile phones also have different performance. The sales department would like to know whether personal preferences have an impact on mobile phone purchases. And analyze how it is affected.

In order to solve the above problems, first of all, we selected the users who purchased the mobile phone from the target users. Meanwhile, we selected the network activity index, online shopping index, online video index, maternal and child index, travel index and financial management index as the basic characteristics of user behavior. It is preliminarily considered that the collinearity among these six indicators may be relatively small, so we conducted variance analysis on the selected six indicators to screen out the first four indicators that have significant influence on whether to buy this mobile phone. Then, a binary logistic regression model is established with the user's mobile phone as the dependent variable and the selected index as the independent variable, and the relationship between whether the user buys the mobile phone and the user's basic behavior characteristics is obtained.

For analysis of each index is how to influence the user whether to purchase the mobile phone, every time we give one of the indicators of a small changes, the corresponding values of logistic regression, the time to get the index to the user whether to buy the phone, the sensitivity of the influence of size, and each index is how to affect the user whether to buy the phone.

2. The establishment and solution of the two-classified logistic model

2.1. The theoretical basis of logistic regression model

Since the probability of Logistic model can only be 0 and 1, we make the following transformation to it:

$$\text{logit}(p) = \ln \frac{p}{1-p}$$

and so,

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \varepsilon$$

The solution is as follows:

$$p = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k x_k)}$$

The value P represents the probability that the result is 1 (i.e., the user buys the phone); β_j regression coefficient of each variable x_j for the first j explained variable;

2.2. Parameter estimation of logistic regression

We use the maximum likelihood estimation method. The maximum likelihood estimation method is to maximize the parameter value θ of the function $L(\theta)$ as the estimated value of the parameter θ .

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \max L(x_1, x_2, \dots, x_n; \theta)$$

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta)$$

It is more difficult to directly solve this formula, so we use the Newton-Raphson iterative algorithm (Newton-Raphson iterative algorithm): it is obtained by analyzing its geometric meaning, that is, the slope of the tangent line.

2.3. Determine the main influencing factors and establish the model

We establish the 0-1 variable of task completion as the dependent variable, 0 means the user has not purchased the mobile phone, 1 means the user has purchased the mobile phone; taking the factors we defined earlier as the independent variables, they are the network activity index and the online shopping index *lp gjtugd* model of online video index and financial index.

2.4. Model testing

The network activity index, online shopping index, online video index, travel index and financial management index obtained by us were put into the logistic model to get the following results:

Table 1 Model fitting test table

Model fitting statistics		
Guidelines	Intercept only	Covariant
AIC	39. 550	12. 842
SC	43. 910	38. 999
-2 Log L	37. 550	0. 842

It can be seen from the above table that the intercept and covariate values corresponding to AIC, SC, and -2LOGL are relatively small, which shows that the model fits well, that is, the selection of indicators and the functional relationship are relatively reasonable.

Table 2 Model fitting test table

Test the global zero hypothesis: beta=0			
inspection	Bangla	freedom	Pr> Bangla
Likelihood ratio	36. 7085	5	<. 0001
score	27. 3509	5	<. 0001
Wald	2. 3179	5	0. 0836

From the above table, we can see that all the numbers in the column value of pr>chi-square are less than 0.05, which indicates that the model fits well.

Table 3 Maximum likelihood estimation table of model parameters

Maximum Likelihood Estimation Analysis					
parameter	freedom	estimate	error	Wald angl	Pr> Bangla
Intercept	1	-24.891	18.2766	1.8548	0.0532
x1	1	0.0019	0.00150	1.5997	0.0059
x2	1	0.0031	0.00244	1.6604	0.0275
x3	1	0.0039	0.00288	1.8342	0.0456
x4	1	0.025	0.0180	2.0197	0.0553
x5	1	0.0030	0.00605	0.2583	0.0013

From the above table, we can see that all the numbers in the column pr>chi-square are less than 0.05, so the fit of each indicator is better, indicating that the model is reasonable. The obtained logistic expression is as follows:

$$\ln \frac{p}{1-p} = 0.19x_1 + 0.031x_2 + 0.39x_3 + 0.025x_4 + 0.30x_5 - 24.891$$

P for the user to purchase the probability of the mobile phone, y 1 for network active index, y 2 as the index of the network shopping, y 3 for online video index, y 4 for travel index, y 5 for financial index.

2.5. Analysis of the results

We can see the linear relationship between whether the user buys the mobile phone and the online activity index, online shopping index, online video index, maternal and child index, and financial index:

$$\ln \frac{p}{1-p} = 0.19x_1 + 0.031x_2 + 0.39x_3 + 0.025x_4 + 0.30x_5 - 24.891$$

We calculate the odds ratio (odds ratio), the formula for calculating the odds ratio is:

$$OR = \frac{p_1/(1-p_1)}{p_0/(1-p_0)}$$

In order to judge the degree of consideration of each factor, we added 1 to the number of each factor. For example, we added 1 to x1 first. According to the formula of logistic regression model, we obtained the probability P1 of completing the task after factor x1 increased by 1, and obtained the following formula ①:

$$\ln \frac{p}{1-p} = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \tag{1}$$

The original formula is ②:

$$\ln \frac{p_0}{1-p_0} = \beta_0 + \beta_1x_1 + \dots + \beta_px_p \tag{2}$$

Divide formula ① by formula ② to get odds ratio (OR):

Because the coefficients from x 1 to x 5 are all greater than 0, the odds ratios of the indicators are all greater than 1, indicating that all four factors have an impact on whether the user purchases the mobile phone.

Therefore, we conclude that the user’s basic behavior characteristics have an impact on whether the user purchases the phone.

3. Formulation of sales model

Through variance analysis of user behavior basic indexes, through logic regression analysis to establish the user whether to buy the phone and the relationship between these indicators. In order to determine how these indicators affect whether the user buys the mobile phone, we change one indicator at a time, increase or decrease this indicator by 1 unit, and the other indicators remain the same. We compare the Y values before and after the change, and the ratio obtained is recorded as the factor influence rate.

$$r_i = \frac{y}{y_0}$$

The higher the factor influence rate is, the greater the influence of the index on whether the user buys the mobile phone.

Table 4 Factor influence rate of each index

index	Original model value	Model value after change	Factor influence rate
Network Activity Index	10.57	8.92	0.84
Online Shopping Index	5.85	10.59	1.81
Online Video Index	9.00	7.56	0.84
Travel index	5.22	9.49	1.82
Financial Index	10.16	10.34	1.02

It can be seen from the above table that the factor influence rate of online shopping index and travel index is relatively large. The factor influence rate of online shopping index is relatively large, indicating that the more times a user purchases things online, the more likely he is to buy our mobile phone. The travel index is probably because the travel index refers to the number of times a user goes out to buy something. However, the factor influence rate of network activity index and online video index is small, because users with high network activity index surf the Internet more but buy less things, and watching videos on the Internet has little influence on users' purchase of mobile phones.

4. Conclusion

Therefore, from the perspective of precision marketing, when a user's online shopping index and travel index are high, we can promote our mobile phone to them. In this case, it is more possible for potential customers to become real customers.

References

- [1] Shi Qihong, Research on Telecom Operators Using Big Data to Carry out 4G Mobile Phone Precision Marketing Jiangxi University of Finance and Economics 2017
- [2] Xie Zhipeng, Zhang Qing, Liu Zongtian. Decision Tree Generation Based on Rough Set Theory [J]. Computer Engineering and Applications, 2000, 11:26-28.
- [3] Miao Duoqian, Wang Huan. Construction method of multivariate decision tree based on rough set [J]. Journal of Software, 1997, 06: 26-32.
- [4] Xie Zhipeng, Zhang Qing, Liu Zongtian. Decision Tree Generation Based on Rough Set Theory [J]. Computer Engineering and Computer Engineering, 2000, 11:26-28
- [5] WANG Xiaoping. Research on Decision Tree Optimization Algorithm Based on Rough Set [D]. Sichuan Normal University, 2013

- [6] Yu Chun, Design and Implementation of Financial Data Analysis System Based on Data Mining Technology, University of Electronic Science and Technology of China, 2014
- [7] Ren Jinluan, Precision Marketing of TV Programs Based on Big Data, Modern Communication Magazine, 2015
- [8] Yang Yudan, Lu Long. User-oriented Brand Communication Application Strategy--Research on Precision Marketing Strategy of Mobile Advertising [J].Brand Research,2017(06):43-51.