Research on Sales Strategy of Electric Vehicles based on Logical Regression and TOPSIS Model

Rupei Li, Keke Ji*, Shengcheng Deng

College of automation, College of industrial Internet, Chongqing University of Posts and Telecommunications, Chongqing, 400065

*Corresponding author. Email: jikeke220411@163.com

Abstract. The development of new energy vehicles represented by electric vehicles has a broad market prospect and is an effective way to solve energy and environmental problems. In this paper, outliers are obtained according to the range of scores, variance, population, and economic situation, and they are regarded as missing data or deleted directly, and then the missing outliers are obtained by using the average method. Then, it is used SPSS to count each index of each kind of data, describe each kind of data, get the sample statistics of each index, and finally calculate the normalized scores of the three brands and the satisfaction ranking of the target customers through the TOPSIS evaluation method. Then the data a1~a8 and B1~B17 are discussed, respectively. Firstly, the oversampling method is used to equalize the samples, then the continuous features are discretized, and then the logical regression model is used to eliminate the features to get the final results.

Keywords: Electric vehicle, TOPSIS, logical regression.

1. Introduction

The new energy vehicle industry is a strategic emerging industry. Represented by electric vehicles, vigorously developing new energy vehicles is of great significance in reducing China's dependence on foreign oil and alleviating environmental pollution pressure [1]. It also drives scientific and technological innovation in relevant fields and boosts the manufacturing industry's transformation and upgrading. It is an effective way to solve energy and environmental problems and has broad market prospects.

However, the electric vehicle is a new thing. Consumers still have some doubts in some fields, such as battery problems, compared with traditional vehicles. These doubts will affect consumers' attitudes towards electric vehicles, and consumers' attitudes are the basis and premise of consumers' purchase behavior [2]. Therefore, establishing a model to study the sales strategy of target customers of electric vehicles is particularly important.

2. Descriptive analysis of data

This paper collects the relevant data of three new energy electric vehicles for mathematical modeling and analysis.

3. Analysis and processing of missing data

The first kind of missing data is the excluded outlier. The mean value can be obtained from the remaining data in the column. The outlier is located as the alternative value of the outlier. It was found that the experience data of the target customers numbered 0001, 0100, 0480, and 1964 had abnormal values. Then, the A1 of the target customer numbered 0001 was 753 04 points [3], which was more than 100 points. Data error and available through calculation were 77 90274 points. The target customer with No. 0100 rated A7 as 7 88 points, which is too far from the average value, and 77 points can be used through the calculation 61091 points; The A5 of the target customer with No. 0480 is 605 63 points, more than 100 points, data error, available through calculation 76 88063 points; The

A3 of the target customer with number 1964 is 703 00 points, more than 100 points, data error, available through calculation 76 05052 points instead.

The second missing data is the missing value of B7 in Appendix 1, that is, the number of children of target customers, which can be calculated from the data of B5, B6, and B7. In B6, case 5 means "married, with children, not living with parents," so the number of children in the family is reduced by 2, case 6 means "married, with children, living with parents," so the number of children in the family is reduced by 4, and case 7 means "divorced/widowed," so the number of children in the family is 0. Case 8 means "other." The number of children cannot be calculated through this case, and there is only one case, 8 in the missing data, so this group of experienced data of this customer is deleted.

4. Descriptive analysis of data

The satisfaction data of target customers with eight types of data are exported through SPSS. After preliminary statistics and SPSS analysis of the collected data, On the whole, the average value of each index is as follows: a4 > a2 > a1 > a7 > a6 > a8 > a5 > a3. According to the variance and standard deviation, the fluctuation range of data a3 is the largest, while data A1 is the smallest, ranging from ±8.492016589540524 to ±9.922597136139986.

Table 1. Discrete Data Statistics In B

		B1	B3	B5	B6	B7	B9	B11	B12	Brand type	Willingness to buy
Number of cases	Effective	1626	1626	1626	1626	1626	1626	1626	1626	1626	1626
	defect	0	0	0	0	0	0	0	0	0	0
average value		1.69	1.61	3.44	4.56	.85	5.46	4.39	4.86	1.79	.05
Median		2.00	1.00	3.00	5.00	1.00	6.00	4.00	4.00	2.00	.00
Mode		2	1	3	5	1	6	4	2	2	0
standard deviation		.464	.741	1.089	1.406	.606	.841	1.517	2.499	.560	.215
Min		1	1	1	1	0	3	1	1	1	0
Max		3	6	6	7	3	8	9	11	3	1

5. TOPSIS evaluation method compares customer satisfaction with different brands of cars

We calculate the average value of the same index of each brand and take the calculated average value as the score of this index. The results are as follows:

Table 2. The score of each indicator of the three brands

Brand	a1	a2	a3	a4	a5	a6	a7	a8
Joint venture	77.6095	78.1261	75.8682	78.8431	77.4220	77.7070	77.9813	77.8597
Autonomy	77.8971	77.8527	76.0472	78.4946	76.8806	77.6743	77.6046	77.1413
New forces	77.0225	77.2540	74.6508	77.7504	75.9256	77.2625	77.3936	77.3402

Set i (i = 1,2,3, i = 1,2,3,4,5,6,7,8) as the satisfaction score of the first brand and the second index:

$$X = \begin{bmatrix} 77.6095 & 78.1261 & 75.8682 & 78.8431 & 77.4220 & 77.7070 & 77.9813 & 77.8597 \\ 77.8971 & 77.8527 & 76.0472 & 78.4946 & 76.8806 & 77.6743 & 77.6046 & 77.1413 \\ 77.0225 & 77.2540 & 74.6508 & 77.7504 & 75.9256 & 77.2625 & 77.3936 & 77.3402 \end{bmatrix} \quad (1)$$

After standardization, the score of each brand becomes

Table 3. The score of each index of the three brands after standardization

Brand	0.5781	0.5802	0.5800	0.5809	0.5824	0.5785	0.5797	0.5804
Joint venture	0.5802	0.5781	0.5813	0.5783	0.5784	0.5783	0.5769	0.5751
Autonomy	0.5737	0.5737	0.5707	0.5728	0.5712	0.5752	0.5754	0.5765
New forces	0.5781	0.5802	0.5800	0.5809	0.5824	0.5785	0.5797	0.5804

Let Z^+ be the column vector composed of the maximum value of each column element, and Z^- the column vector composed of the minimum value of each column element.

This paper assumes the distance between the first evaluation object (i=1,2,3) and the maximum value.

$$D_i^+ = \sqrt{\sum_{j=1}^{8}\left(Z_j^+ - Z_{ij}\right)^2} \tag{2}$$

The distance from the minimum value is used for the I evaluation object (i=1, 2, 3).

$$D_i^- = \sqrt{\sum_{j=1}^{8}\left(Z_j^- - Z_{ij}\right)^2} \tag{3}$$

The score for the i object that is not normalized is:

$$S_i = \frac{D_i^-}{D_i^+ + D_i^-} \tag{4}$$

Table 4. Score and ranking of three brands

Brand	D^+	D^-	Non-normalized score	Normalized score	Ranking
Joint venture	0.0025	0.0199	0.8870	0.5455	1
autonomy	0.0080	0.0164	0.6726	0.4137	2
New forces	0.0209	0.0015	0.0664	0.0408	3

6. Sales Analysis of different factors on all kinds of Brand Electric vehicles

6.1. Establishment of model

Logistic regression [4] is a classification model in machine learning. Due to the simplicity and efficiency of the algorithm, it is widely used in practice. It belongs to the "generalized linear model":

$$y = g^{-1}\left(w^T x + b\right) \tag{5}$$

Considering the binary classification task, its output flag is $y \in \{0,1\}$, and the predicted value z=wx+B generated by the linear regression model is a real value. Therefore, it is necessary to convert the real value Z to a 0/1 value. The ideal is ". unit step function":

$$y = \begin{cases} 0, z < 0 \\ 0.5, z = 0 \\ 1, z > 0 \end{cases} \tag{6}$$

If the predicted value Z is greater than 0, it will be quantified as a positive example[5]. If it is less than zero, it will be judged as a negative example, and if the predicted value is the critical value, it

can be judged arbitrarily. However, the step function is not continuous and cannot directly act on g-(). Therefore, another function is considered to replace the step function, namely the sigmoid function:

$$y = \frac{1}{1+e^{-z}} \tag{7}$$

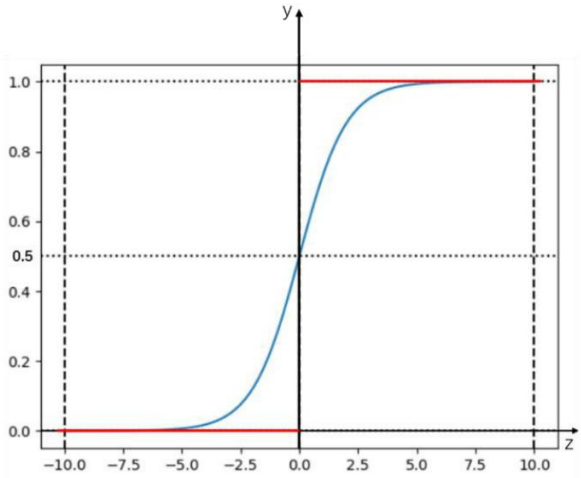

Figure. 1 Image corresponding to the sigmoid function

The most basic way to solve the parameters is the gradient descent method, as long as the cost function and parameters are written θ Gradient formula. The gradient descent formula is:

$$\mathrm{J}(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m} y^{(i)} log h_\theta\left(x^{(i)}\right) + \left(1-y^{(i)}\right)\log\left(1-h_\theta\left(x^{(i)}\right)\right)\right] \tag{8}$$

$$\theta_j = \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta) \tag{9}$$

$$\frac{\partial}{\partial \theta_j} J(\theta) = \frac{1}{m}\sum_{i=1}^{m}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right) x_j^{(i)} \tag{10}$$

6.2. Hypothesis test of logistic regression model

The test methods include the likelihood ratio test, score test, and Wald test. The likelihood ratio test is introduced here. The original hypothesis is tested as $H_0 : \beta_1 = \beta_2 = \ldots = \beta_m = 0$. That is, the overall regression coefficients of all independent variables are 0. The likelihood ratio test statistic G is:

$$\mathrm{G} = -2\left(\ln L_1 - \ln L_0\right) \tag{11}$$

When the overall test of the logistic regression model is meaningful, it is also necessary to test the hypothesis of the regression coefficient of each independent variable in the model to test the contribution or function of each independent variable to the model. The commonly used hypothesis test method is Wald-χ 2 Inspection. The original hypothesis is tested as $H_0 : \beta_i = 0$. That is, the overall regression coefficient of each independent variable is 0. Test statistics

$$\chi^2 = \left(\beta_i / S_{\beta_i}\right)^2 \tag{12}$$

6.3. Solution of model

Logical regression belongs to the generalized linear model, which can increase the nonlinearity of the model and improve the expressive power of the model by discretizing the continuous variables. Meanwhile, it is easy for the last iteration of the model. Because a1~a8 is a value with a total score

of 100, all are continuous values, and data B has continuous and discrete values, the continuous values are discretized to obtain:

Table 5. A1 ~ A8 discretization

Data type	Initial value	Value after discretization
a1~a8	0-40	1
	40-60	2
	60-70	3
	70-80	4
	80-90	5
	90-100	6

Table 6. Discretization of B2

Data type	Initial value	Value after discretization
B2	0-4	1
	4-10	2
	>10	3

(As an example)

Step1: according to the purchase intention and the brand of electric vehicles, the target customers are divided into customers who have bought electric vehicles and those who have not bought electric vehicles.

Step2: perform a logical regression operation on the data of a1~B17 in Appendix 1.

Step3: eliminate the features whose significance is greater than 0.05in the operation result.

Step4: perform a logical regression operation on the remaining features.

Step5: the feature is eliminated according to the significance.

Step6: Repeat the above steps until all features pass the significance test.

6.4. Test and solution of the model

Table 7. Omnibus test of model coefficients

		χ^2	Freedom	Significance
Step 1	Step	1121.880	25	.000
	Part	1121.880	25	.000
	Model	1121.880	25	.000

The omnibus test of the model coefficient passes, the model can be established, and the model is generally meaningful.

Table 8. Variables in joint venture brand a1~a8 equation

		B	Standard error	Wald	Freedom	Significance	Exp(B)	EXP(B) 95% confidence interval
								lower limit
Step 1[a]	a1	.085	.023	13.405	1	.000	1.088	1.040
	a2	.072	.024	9.219	1	.002	1.075	1.026
	a3	.147	.018	64.440	1	.000	1.158	1.117
	a4	-.038	.019	3.954	1	.047	.963	.928
		-21.677	1.647	173.259	1	.000	.000	

It can be seen from the above table that among a1~a8 these eight experience data, A3 has the largest coefficient, so A3, that is, economy, has the greatest impact on the sales of electric vehicles.

For joint venture brands, the economy of electric vehicles and customers' marriage and family situations have the most significant impact on the sales of electric vehicles. The proportion of the annual expenditure of electric car brands and comfort to the total household income; For new power brands, the economy of electric vehicles and the nature of the customer's unit have the greatest impact on the sales of electric vehicles.

7. Conclusions

After data pre-processing, this paper uses SPSS to count each index of each kind of data, describes each kind of data, and obtains the sample statistics of each index. Finally, the topsis evaluation method calculates the normalized scores of the three brands and the satisfaction ranking of the target customers and concludes that the target customers are most satisfied with the joint venture brand, followed by independent brands, followed by the new power brand. Then, the data a1~a8 and B1~B17 are discussed, respectively. Firstly, the oversampling method is used to equalize the samples, then the continuous features are discretized, and then the logical regression model is used to eliminate the features. Using SPSS to solve: for the joint venture brand: the economy of electric vehicles and customers' marriage and family conditions have the greatest impact on the sales of electric vehicles. For independent brands: the comfort of electric vehicles and the proportion of customers' annual mortgage expenses to total household income have the greatest impact on electric vehicle sales; for new power brands: the economy of electric vehicles and the nature of customers' units have the greatest impact on electric vehicle sales.

References

[1] Si Shoukui, Mathematical Modeling algorithm and Application (second Edition), National Defense Industry Press, 2011.8.

[2] Yu Xiaofen and Fu Dai, Summary of Multi-index Comprehensive Evaluation methods, Statistics and decision-making, Zhejiang University of Technology, 2004.

[3] Yue Ruixiang Tang Yong, Analysis of influencing factors of Automobile sales and its Countermeasures, Journal of Beijing Institute of Economic Management, 2011.

[4] Mao Yi, Chen Wenlin and Guo Baolong, logical regression model based on density estimation.

[5] Sheng Zhaohan and Liu Bingxiang, decision Tree method for customer churn crisis Analysis, Journal of Management Science, 2005