Development analysis of domestic new energy vehicles based on machine learning method —— Take BYD for Example

Yan Sun, Kaiyang Zhao, Junyan Liao, Mengxin Wang
(School of Applied Science, Beijing Information Science & Technology University, Beijing 100192,China)

Abstract: In view of the rapid development of new energy vehicles, the price of fuel vehicle raw materials greatly affected by the world situation, and the domestic vigorous support for the development of new energy vehicles, taking BYD, a leading domestic new energy vehicle enterprise, as an example, the principal component analysis method and cluster analysis method were used to select the most representative 5 types of models from BYD's 97 models. The results of this paper will help consumers choose new energy vehicle models, help car companies improve the market competitiveness of brands, and provide opinions for the government on the formulation of new energy policies.

Keywords: New energy; Principal component analysis; Cluster analysis

Introduction

New energy vehicles began at the beginning of the 21st century, due to the breakthrough of battery technology, countries began to apply electric vehicles on a large scale. In recent years, new energy vehicles have gradually become an important choice for people to buy cars. Researchers at home and abroad have studied new energy vehicles from the aspects of economics, sociology, psychology and so on. Most researchers study new energy vehicles from the macro level, think and put forward valuable opinions from the perspective of corporate profitability or government policy promulgation, etc., and study new energy vehicles from the micro individual perspective of people who are ready to buy new energy vehicles.

For consumers, their awareness of environmental protection is a significant factor affecting whether they buy new energy vehicles. In addition, the use of relatives and friends will directly affect people's psychological perception of value. Due to the limited knowledge, the group with low education pays more attention to the policy of car purchase, while the group with high income and high education pays more attention to the later use benefits of re-energy vehicles. Since the government's increase in purchase subsidies is a short-term effect, subsidies should be increased in the supporting facilities of new energy vehicles, so as to change the consumption path of new energy vehicles with minimum consumer expenditure, and then increase its demand. At the same time, the society should improve consumers' sense of identity for new energy vehicles and enhance consumers' satisfaction. According to Bayson, the government should issue policies to encourage consumers to buy new energy vehicles and accelerate the construction of charging infrastructure. At the same time, the government's reasonable mix of fiscal and tax policies, advances in industrial production technology and shifts in consumer preferences can greatly affect the speed of development of new energy vehicles.

Regarding the development status of new energy vehicles, Gu Wenjie and others believe that new energy vehicles are one of the important ways to cover fuel vehicles and promote the development of energy replacement. Zhang Xianjing mainly studies the development prospects of new energy automobile enterprises from the perspective of their profitability, debt paying ability and operating ability. However, this research method is only applicable to short-term operation. An enterprise may have a high market share and a good brand reputation, but the degree of scientific and technological innovation is not enough, and its high cost under market competition leads to its lack of profitability.

In this paper, China's new energy leading car enterprise BYD as an example, through the principal component analysis method on the basis of the above research, to find the main factors affecting the sales of new energy vehicles. And through the sales data of the enterprise from July 2018 to December 2022, the machine learning method is used to predict its sales volume in early 2023, which provides useful help for consumers to buy cars, automobile enterprises to improve the market competitiveness of the brand, and the government to formulate relevant new energy policies.

I. Data Description

This paper selects the research objects through the sales data of new energy vehicles of major automobile brands from January 2018 to December 2022 on websites such as "Auto Home", "Owner's Home" and "Orient Fortune Net".

In order to analyze what aspects consumers pay more attention to when purchasing cars, the next part is word cloud image analysis. As shown in Figure 1, the size of the word indicates how many times the word appears in the user's review. In the purchase process of cars, "cost performance" is the most priority for consumers, while "model", "mileage", "power" and so on are more concerned aspects. Therefore, in the next factor analysis, this paper selected price standardization, maximum horsepower standardization, total power standardization, battery capacity standardization, pure electric range standardization and other five indicators for principal component analysis.

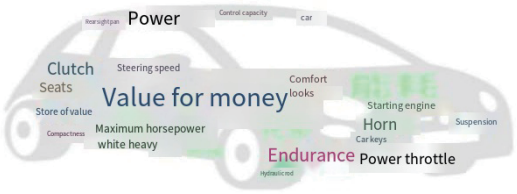

FIG. 1 Cloud map of automobile analysis words

II.Principal component analysis and cluster analysis

1. Principal component analysis

Principal component analysis is an effective way to reduce the dimension of data. It is a statistical analysis method that combines several related indicators into a comprehensive index. The original multiple indicators are recombined into several new uncorrelated comprehensive variables to replace the original variables, which can retain the original information as much as possible and simplify the operation.

In order to eliminate the difference caused by the different values between the parameters, the five parameters of price, maximum horsepower, total power, capacity and mileage are normalized first, and the normalization adopts min-max method, as shown in formula (1)

$$\frac{x-\min(x)}{\max(x)-\min(x)} \quad (1)$$

Where, x is the parameter value, $\min(x)$ is the minimum value of this parameter, $\max(x)$ is the maximum value of this parameter.

Considering that there is a linear relationship between the factors, the variables are transformed into a set of linearly uncorrelated variables by orthogonal transformation. The original number of variables was 5. After linear combination of matrix (orthogonal transformation), 5 new components were formed. The first 2 components contributed 93.614% of the data variation in total, and the feature roots were all greater than 1, so the first 2 components were extracted as the main components. Look at the component matrix of the two factors and rotate it.

The rotated component matrix is the coefficient matrix obtained after the original data is mapped to the factor space after the factor analysis is performed, and it reflects the linear relationship between these several factors and the original variable. As can be seen, all variables have load values greater than 0.4 in their own rows, indicating that all five variables are useful. Price, maximum horsepower total power in component 1 quotient load is the largest; Battery capacity and pure electric mileage are the largest in component 2. Note that component 1 is highly correlated with the three indicators of price, maximum horsepower and total power, and these three are about the cost performance problem, which can be named “cost performance factor”. Component 2 is extremely related to “capacity and mileage” and is named “convenience factor”.

2. Cluster analysis

Common clustering methods include K-means clustering and systematic clustering. Because BYD new energy vehicle models are variegated and complex, this paper chooses systematic clustering method as the clustering method. According to the two main factor scores of 97 BYD models, the indicators are standardized for systematic clustering analysis. Ward system clustering method is similar to the idea of analysis of variance, through the characteristics of the research object to carry out comprehensive classification, the advantage of this method is to make each category after the combination of the group deviation square and the added value of the minimum, effectively highlighting the similarity within the category and the difference between the categories, reflecting the difference between different objects. The clustering method can be used to effectively classify the vehicles under study, so as to excavate the characteristics of each type of vehicle. After cutting with distance 4, more models are clustered in the first, fourth and fifth categories. The third and fourth categories are broad categories, which contain more models, while the second and third categories are small categories, which contain less models.

The representative models of each category are listed in Table 1 for consumers’ reference.

Table 1 List of representative models of BYD by type

Category	Representative model
category 1	Dolphin 2021 405km Liberty
	dolphin 2021 405km fashion edition
	Yuan Pro 2021 401KM Love for Life Edition
category 2	Song PLUS 2021 DM-i 110KM Signature PLUS
	Qin PLUS 2021 DM-i 120KM Premium
	Song PLUS 2021 DM-i 110KM Signature PLUS
	Tang New Energy 2021 DM-i 112KM Premium

category 3	SONG PLUS New Energy 2021 DM-i 100KM 4WD Flagship PLUS
	Tang New Energy 2021 DM-i 112KM Premium
	Han 2020 EV Extended Range Deluxe
category 4	Han 2021 EV Standard Range Deluxe
	Song PLUS New Energy 2021 EV 505KM Flagship
	Han 2020 EV 4WD High Performance Flagship
	Tang New Energy 2021 EV 4WD High Performance Edition Premium
category 5	Tang New Energy 2021 DM 2.0T 4WD High Performance Edition
	Genesis Flagship

III. Closing remarks

Through principal component analysis and cluster analysis, this paper makes clear the choice of BYD's various models for consumers, and also provides a reference for car companies to develop new models from the perspective of consumer demand. Byd brand, as a representative of domestic new energy vehicles, will have good sales performance in the future, reflecting the broad prospects for the development of new energy vehicles.

Reference literature:

[1] Chunli He. Research on Market Demand and Policy Orientation of New Energy Vehicles [D]. Southwest University of Finance and Economics,2020.

[2] Wenyuan Wang,Haofei Wang,Yizhuo Wu,Ying Hua. Research on Influencing factors of consumption behavior of new energy vehicles in Changzhou City [J]. China Business Theory,2019(20):206-208.

[3] Chuang Li,Lulu Ye,Liping Wang. The impact of new energy vehicle consumption promotion policy on the purchase intention of potential consumers [J]. China Management Science,2021,29(10):151-164.

[4] Yunxia Li. Research on the incentive effect of China's new energy vehicle subsidies on consumers [D]. Dongbei University of Finance and Economics,2018.

[5] Li Li,Chujing Yang. Optimization of policy effect of consumption of new energy vehicles in China based on international comparative perspective [J]. Guangxi Social Sciences,2017(02):77-82.

[6]Bryson J R . An international journal business service firms, service space and the management of change business service firms, service space and the management of change[J]. Entrepreneurship & Regional Development, 1997, 9(2):93-112.

[7] Zheyuan Liang, Wei Le. Impact of fiscal and taxation policies, technological progress and consumer preference on the development of new energy automobile industry [J]. Finance and Economics,2022(11):56-65.

[8] Wenjie Gu,Kuan Fu. Research on the development of new energy automobile industry based on Big Data analysis [J]. Internal Combustion Engine and Accessories,2022(05):163-165.

[9] Xianjing Zhang. Analysis of New energy Automobile industry -- taking BYD Company as an example [J]. China Storage and Transportati on,2023(01):151-152.

[10]Frederic, B. Bulletin of mathematical biophysics[J]. The Journal of Symbolic Logic,1944,9(2):49-50. (in Chinese)