Frontiers in Business, Economics and Management
ISSN: 2766-824X | Vol. 18, No. 2, 2025

Research on Automobile Aftermarket Parts Pricing Model Based on Support Vector Machine

Tengjian Yang[1, 2], Chao Zhang[1, 2], Peipei Zhu[1, 2], Fan Zhang[1, 2, *]

[1]China Automotive Technology and Research Center Co., Ltd. Tianjin, China
[2]China Auto Information Technology (Tianjin) Co., Ltd. Tianjin, China
[*]Corresponding author: m13502037247_3@163.com

Abstract: In this study, a pricing model based on support vector machine is proposed to solve the problem of auto aftermarket parts pricing. By collecting and analyzing multi-source heterogeneous data, a "1+4+8" pricing index system was constructed, which includes parts cost, market demand and supply, brand and quality, material and performance. Support vector machine (SVM) is used to establish the pricing model of automobile aftermarket parts. The experimental results show that the pricing model based on support vector machine can effectively improve the pricing accuracy and provide data support for enterprises to formulate reasonable pricing strategies. This study provides a new idea and method for the pricing of automotive aftermarket parts, which is of great significance to enhance the competitiveness of enterprises.

Keywords: Automobile after-sale parts; Pricing model; Support vector machine.

1. Background

With the rapid development of the automobile industry and the increasingly fierce market competition, the pricing of automobile after-sale parts has become a key factor for enterprises to enhance their competitiveness. Traditional pricing methods often rely on experience and simple cost plus, which is difficult to adapt to the complex and changing market environment. In recent years, the rise of machine learning technology has provided new opportunities for the pricing of automotive aftermarket parts. By collecting and analyzing massive data, enterprises can more accurately grasp market demand, competitive situation and customer value, so as to formulate a more scientific and reasonable pricing strategy.

In today's auto aftermarket, the pricing strategy traditionally relies on several models such as cost plus, competition-oriented pricing and value pricing. The cost plus method is easy to operate, but it fails to fully consider the market demand and competitive dynamics; Although the competition-oriented pricing takes into account the market environment, it is easy to lead to the vicious circle of price war; Although value pricing focuses on the perceived value of customers, it faces quantitative problems in practice. These traditional methods show obvious shortcomings when dealing with the complexity and variability of the market environment. With the rapid development of big data and machine learning technologies, data-driven pricing strategies are gradually becoming the new trend in the industry. Companies are able to gather and analyze huge amounts of market, customer and operational data to gain insight into market demand, competitive landscape and customer value, providing strong support for the development of more accurate and reasonable pricing strategies. The integration of machine learning not only greatly improves the accuracy of pricing, but also gives enterprises the ability to quickly adapt to market fluctuations and realize dynamic adjustment of pricing.

In the field of parts prediction, machine learning and artificial intelligence methods are gradually showing their great potential. For example, Wu Xin et al. [1] innovatively proposed a multi-value chain model based on third-party platform data, aiming at the problems of relying on manual experience, single data and ignoring the upstream and downstream links of the supply chain in traditional parts procurement decision-making, which resulted in insufficient prediction accuracy and poor robustness. This model combines graph convolutional neural network and long and short term memory neural network. By integrating the historical data of the core manufacturer and the dealer, the prediction ability is improved, and the RMSE evaluation index has achieved a significant improvement of 3%. Similarly, Li Yuqing et al. [2] also pointed out that the traditional manual experience decision-making method of automobile manufacturing enterprises lacks scientific basis and is easy to lead to the imbalance between supply and demand of spare parts. They actively introduced intelligent technology for optimizing decision-making and developed an intelligent service system for spare parts inventory decision-making by making full use of real business data. This system provides scientific and accurate inventory optimization decision support for manufacturing enterprises, effectively helps enterprises optimize inventory management process, reduces inventory costs, improves profits, and significantly improves the intelligent level of inventory management, thus enhancing the market competitiveness of enterprises. Yan Weijun [3] et al. proposed a forecasting model based on long and short term memory (LSTM) in order to solve the two prominent and difficult problems plaguing auto parts enterprises, namely, the low timely supply rate and the serious inventory overhang. For auto parts with long replacement cycle and small data scale, a multi-model fusion prediction model based on machine learning was proposed, and the algorithm was verified by experiments, and the experimental results showed that it was relatively reliable. Therefore, the application of machine learning and artificial intelligence methods in the field of parts prediction not only improves the prediction accuracy and robustness, but also provides a scientific decision-making basis for automobile manufacturing enterprises, and promotes the development of inventory management to the direction of intelligence and high efficiency.

To sum up, the core goal of this paper is to develop a machine learning technology-based automotive aftermarket parts pricing model, aiming to overcome the limitations of traditional pricing means. This model will integrate diversified data sources and information of different structures, and adopt advanced machine learning algorithms to create a pricing solution with self-learning and continuous optimization ability. The solution aims to provide companies with accurate and flexible pricing strategies to respond to market changes. This research is not only expected to significantly improve the efficiency and accuracy of pricing decisions of enterprises, but also aims to contribute theoretical insights and practical wisdom to the steady development of the automotive aftermarket.

2. Research Methods

The research method proposed in this paper includes four main parts: pricing index system construction, data collection, data preprocessing and feature engineering, and pricing model design based on machine learning model. The construction of the pricing index system is mainly based on the big data recommendation technology to determine the index system affecting the after-sales parts, and the data collection is based on the index system architecture to find the data; Data sources include but are not limited to relevant data websites and authoritative reports, etc. Data fields include cost data, sales data, etc.; In terms of data preprocessing and feature engineering, firstly, the collected data are processed with outliers, missing values and data normalization. In terms of feature engineering, based on the pricing index system, the feature engineering mathematical operation is carried out according to the principle of the simplest constraint to obtain the optimal feature data and pricing index system; The pricing model based on machine learning model is mainly based on the optimized index system and data to select the most suitable machine learning method. The research roadmap of this paper is shown in Figure 1 below.

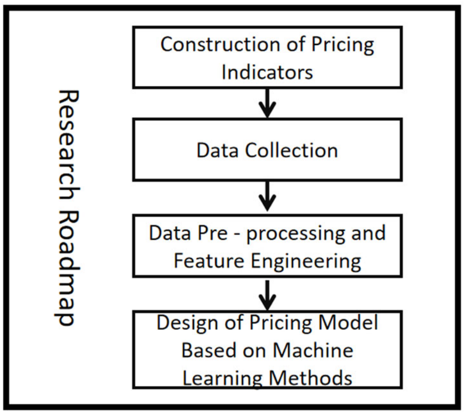

Figure 1. Research roadmap of this paper

2.1. Selection of parts pricing index

After the screening and recommendation of big data, the after-sales parts pricing index system selected in this paper is "1+4+8", that is, 1 first-level indicator, 4 second-level indicators and 8 third-level indicators. The after-sales parts pricing index system is shown in Figure 2 below.

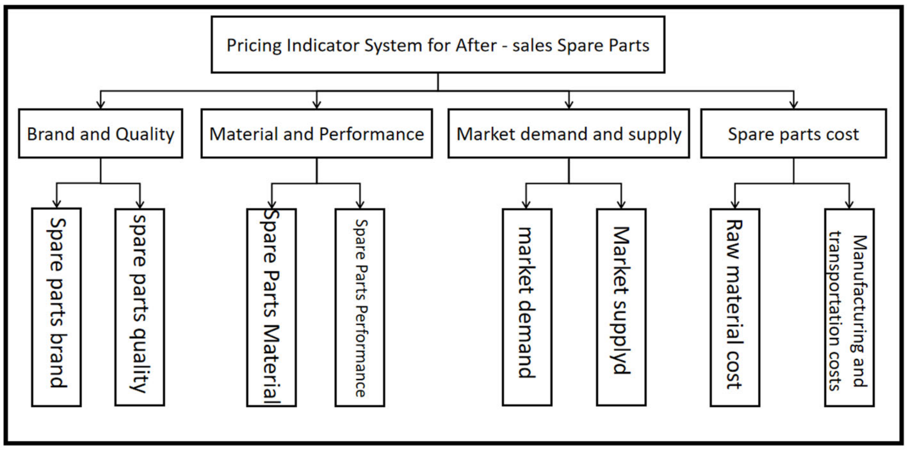

Figure 2. After-sale parts pricing index system

(1) Brand and quality

Accessory brand: Different brands of accessories have differences in price, quality, performance and so on. Well-known brands usually have higher quality assurance and after-sales service, but prices can also be relatively high.

Accessory quality: The quality of the accessory directly affects its service life and performance. Quality accessories are usually manufactured with better materials and processes, with higher durability and reliability.

(2) Material and performance

Accessory material: The material of the accessory determines its physical and chemical properties, which in turn affects its use effect. For example, brake pads made of aluminum alloy have better heat resistance and wear resistance than iron brake pads.

Accessory performance: The performance of the accessory includes function, efficiency, durability and so on. Accessories with high performance usually have better use results, but may also increase the cost of use.

(3) Market demand and supply

Market demand: The greater the market demand for a certain accessory, the higher the price of the accessory tends to be. On the contrary, if the market demand decreases, the price may decrease.

Availability: The availability of accessories can also affect their price. Prices usually go down when there is an oversupply; When supply is low, prices may rise.

(4) Cost of parts

Raw material cost: The cost of raw materials to make accessories.

Manufacturing and transportation costs: The price of the parts made based on the raw materials and the corresponding transportation costs.

2.2. Data collection

This paper uses computer-aided method for data collection, through computer technology to collect, organize and analyze data. The technology of data mining and web crawler is an important part of the computer aided method. Data mining can help researchers extract useful information from large amounts of data, while web crawlers can automatically grab data from web pages.

The crawler code is as follows:

```
import requests
from bs4 import BeautifulSoup
# Destination URL
url = 'http://example.com'
# send HTTP GET request
response = requests.get(url)
# Check if the request was successful
if response.status_code == 200:
    # Parse the HTML content
    soup = BeautifulSoup(response.content, 'html.parser')
    links = soup.find_all('a')
    for link in links:
        href = link.get('href')
        if href:
            print(href)
else:
    print(f" Request failed, status code: {response.status_code}")
```

(1) Destination URL: Replace the url variable with the URL of the web page you want to crawl.

(2) Request header: In a practical application, some websites may check the request header, such as User-Agent. If the request is blocked, the appropriate request header needs to be added.

(3) Procedural legality: Before crawling a website, you need to ensure compliance with the website's robots.txt file and the relevant terms of use. Do not overburden the website, and do not violate the law.

(4) Exception handling: In practical applications, more exception handling code needs to be added to deal with network problems, parsing errors, etc.

2.3. Data preprocessing and feature engineering

(1) Data preprocessing

Data preprocessing, also known as data cleaning, deals with missing values, outliers, duplicate values and other problems in the data to improve data quality. In this paper, missing values and outliers are mainly processed. Missing value processing: delete rows or columns with more missing values, or use interpolation methods (such as mean interpolation, median interpolation, mode interpolation, etc.) to fill in the missing values. In addition, model-based interpolation methods can also be used to predict missing values, such as regression models, K-nearest neighbor algorithms. Outlier processing: Identify outliers by visual methods (such as boxplot) or statistical methods (such as standard deviation based method), and choose the deletion, correction and other processing methods according to[4] the specific situation.

(2) Data transformation

Data transformation: through smooth aggregation, data generalization, normalization and other ways to convert the data into a form suitable for data mining. Common data transformation methods include standardization, normalization, logarithmic transformation, etc.

Normalization: Converting the eigenvalues of the data into a specific interval, such as converting the data into an interval with a mean of 0 and a standard deviation of 1. This can eliminate the impact of different features due to different dimensions, so that different features are comparable in subsequent analysis. Normalization: Mapping the data to the interval [0,1] so that all features fall within the same numerical range. Logarithmic transformation: When the data presents a skewed distribution, you can perform a logarithmic transformation on the data (such as taking the natural logarithm) to make the data distribution closer to the normal distribution, which is convenient for subsequent analysis and modeling[5].

(2) Feature engineering

Screening features: Evaluate the necessity of the existence of indicators through special importance. If the feature value of a feature is unique or nearly unique, then the feature is invalid and can be deleted;

$$n = (1 - \frac{1}{set(feature)})\% \tag{1}$$

Formula (1) is the mathematical model for calculating the contribution degree of an influential factor, n is the contribution degree of an influential factor, set is the de-duplication function of the eigenvalue of an influential factor, and $feature$ is the set of eigenvalue.

Combined features: By combining multiple features, new features are generated to provide richer information. Combining features can reveal the interactions between features, thereby improving the performance of the model.

Feature selection: Select the features that are most representative or best able to distinguish different categories from the original feature set. Feature selection can reduce redundant features, reduce model complexity, and improve model performance.

2.4. Design of pricing model

Based on the above index system, this paper uses the learning algorithm of support vector machine to build the pricing model. Support vector machine conducts regression[6] by finding the optimal hyperplane, which is suitable for high-dimensional space and nonlinear problems. The steps of support vector machine are[7] as follows:

The first step is to construct the objective function of SVM. The objective function is formula (2).

$$\begin{aligned}&\min_{w,b,\zeta}\frac{1}{2}\|w\|^2+\mathrm{C}\sum_{i=1}^{n}\zeta_i\\&s.t.\ \ y_i\bullet(w^T\bullet\phi(x_i)+b)\geq 1\text{-}\zeta_i,i\text{=}1,2,\cdots,n\\&\quad\zeta_i\geq 0,i\text{=}1,2,\cdots,n\end{aligned}\tag{2}$$

In the second step, the Lagrange function with relaxation factor is constructed. The Lagrange function is shown in formula (3).

$$L(\mathrm{w,b,a},\zeta,\mu)\text{=}\frac{1}{2}\|w\|^2+\mathrm{C}\sum_{i=1}^{n}\zeta_i\text{-}\sum_{i=1}^{n}a_i(y_i\bullet(w^T\bullet\phi(x_i)+b)\text{-}1+\zeta_i)\text{-}\sum_{i=1}^{n}\mu_i\zeta_i\tag{3}$$

Step 3: Take the partial derivative of w,b,,. ζ

Step 4: Formula (3) into the objective function formula (2), construct formula (4)

$$\min_{w,b,\zeta}L(\mathrm{w,b},\zeta,a,\mu)\text{=}\sum_{i=1}^{n}a_i\text{-}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jy_iy_j\bullet(\phi(x_i)\phi(x_j))\tag{4}$$

Step 5: Find the maximum value of formula (4) with respect to a to get formula (5)

$$\begin{aligned}&\max_{a}\sum_{i=1}^{n}a_i\text{-}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jy_iy_j\bullet(\phi(x_i)\phi(x_j))\\&s.\mathrm{t.}\ \sum_{i=1}^{n}a_iy_i\text{=}0\\&\quad a_i\geq 0,\\&\quad C-a_i-\mu_i\text{=}0\\&\quad \mu_i\geq 0,i\text{=}1,2,\cdots,n\end{aligned}\tag{5}$$

Step 6: Construct the dual problem of the maximum value problem, that is, the minimum value, the minimum value objective function as follows: (6)

$$\begin{aligned}&\min_{a}\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}a_ia_jy_iy_j\bullet(\phi(x_i)\phi(x_j))\text{-}\sum_{i=1}^{n}a_i\\&s.\mathrm{t.}\ \sum_{i=1}^{n}a_iy_i\text{=}0\\&\quad 0\leq a_i\leq Ci\text{=}1,2,\cdots,n\end{aligned}\tag{6}$$

Step 7: Solve the optimal solution a^*, the optimal solution formula is as formula (7).

$$\begin{aligned}&w^*\text{=}\sum_{i=1}^{n}a_i^*y_i\phi(\mathrm{x}_i)\\&b^*\text{=}\frac{\max\limits_{i:\ y_i=-1}w^*\bullet x_i+\min\limits_{i:\ y_i=1}w^*\bullet x_i}{2}\end{aligned}\tag{7}$$

Note: When calculating b*, you need to use a vector that satisfies the conditions $0<a_j<C$ .In practice, all values of the support vector are averaged and denoted as b*.

Step 8: Calculate the separation hyperplane, the hyperplane formula is formula (8).

$$w^*\phi(\mathrm{x})+b^*\text{=}0\tag{8}$$

Step 9: Decision function

$$f(\mathrm{x})\text{=sign}(w^*\phi(\mathrm{x})+b^*)\tag{9}$$

Input the pre-processed data into the model and train multiple pricing models through feature selection and parameter optimization. In order to further improve the performance of the model, the method of model integration is also used to weighted average the prediction results of different algorithms to obtain more stable and accurate prediction results. The model can not only provide accurate pricing suggestions, but also quantify the impact of various influencing factors on prices, providing strong support for enterprises' pricing decisions.

3. Result Analysis

Through the feature importance analysis of the model, it is found that market demand and supply are the most influential factors, followed by brand and quality, and cost has a relatively small impact. This result is consistent with the current customer-centric trend of the automotive aftermarket, and also reflects the ability of the machine learning pricing model to capture factors that are difficult to quantify with traditional methods. In addition, the model also reveals the difference in price elasticity between different regions and different seasons, providing a basis for enterprises to formulate differentiated pricing strategies.

4. Conclusion

The support vector Machine-based automotive aftermarket parts pricing model constructed in this study can effectively improve the accuracy and scientificity of pricing by integrating multi-source heterogeneous data and corresponding data preprocessing and cleaning methods. The experimental results show that the model can capture the factors that are difficult to be quantified by traditional pricing methods, such as customer value and market trend, and provide strong support for enterprises to formulate more reasonable pricing strategies. However, there are still some limitations in this study. The model is highly dependent on data quality and does not consider the influence of external environment such as macroeconomic factors. Future studies can further expand the data sources, introduce more external variables, and explore more advanced algorithms such as deep learning to continuously optimize the performance of the pricing model.

References

[1] Wu Xin. Research on Multi-value chain Auto Parts demand Forecast based on Graph Neural Network [D]. Sichuan university, 2023. DOI: 10.27342 /, dc nki. Gscdu. 2023.000580.

[2] Li Yuqing. Research and System Implementation of After-sale Parts Inventory Decision based on Deep Reinforcement Learning [D]. Southwest jiaotong university, 2023. DOI: 10.27414 /, dc nki. Gxnju. 2023.000291.

[3] YAN Weijun. Research and Realization of Supporting Auto Parts sales forecasting method [D]. University of Electronic Science and Technology of China, 2019.

[4] Ling Ning, Guo Jin. Cleaning and visualization of meteorological big data over one million based on Python [J]. Modern Information Technology, 2025, 9 (01): 100-103+109. DOI:10.19850/j.cnki.2096-4706.2025.01.020.

[5] Fang Li-Zhen, Cao Xin-Yu. Intelligent data cleaning and pretreatment algorithm of the information system in colleges and universities study [J]. Computer programming skills and maintenance, 2025, (01) : 115-117. DOI:10.16184/j.cnki.com.prg. 2025.01.003.

[6] FENG Ziang, SHAO Jiayu, Zhang Ning. Research on Flame recognition Algorithm based on Support Vector Machine [J]. Industrial Control Computer, 2025, 38 (01): 79-81.

[7] WANG Rongchao, DUAN Hongguang, LIU Tingqin, et al. SSS sequence detection model based on Residual Convolutional neural Networks-support vector Machines [J/OL]. Radio communication technology, 1-9 [2025-02-11]. http://kns.cnki.net/kcms/detail/13.1099.TN.20250114.1846.006.html.