International Journal of Global Economics and Management
ISSN: 3005-9690 (Print), ISSN: 3005-8090 (Online) | Volume 7, Number 1, Year 2025
DOI: https://doi.org/10.62051/ijgem.v7n1.15
Journal homepage: https://ijgem.org

An Intelligent Automotive Marketing Monitoring Platform Based on Vehicle Network Data

Yafei Wang, Dehuan Kong, Xue Zheng *, Fan Zhang

China Auto Information Technology (Tianjin) Co., Ltd., Tianjin, China
*Corresponding Author: zhengxue@catarc.ac.cn

ABSTRACT

With the rapid development of Internet of Vehicles technology, the application of Internet of Vehicles data in the automotive sales field has gradually become an important means to improve sales efficiency and customer experience. This paper aims to study and construct a monitoring system for intelligent automotive sales systems based on connected vehicle data. First, handle the missing values and outliers according to the different characteristics of the Internet of Vehicles data; Then, feature engineering is carried out based on the distribution trend of the data and the relationship between feature dimensions; Finally, build an intelligent sales system that includes the data collection layer, data transmission layer, data storage and processing layer, business logic layer and application layer. The research results show that the system can effectively improve the accuracy and efficiency of car sales, provide strong support for the decision-making of car enterprises, and has significant practical significance and application value.

KEYWORDS

Internet of Vehicles data; Feature dimension; Feature engineering; Intelligent sales systems for automobiles

1. INTRODUCTION

With the rapid development of information technology, significant progress has been made in Internet of Vehicles technology. By building a three-dimensional interactive network of vehicle-to-vehicle, vehicle-to-road, and vehicle-to-cloud, vehicles can generate up to terabytes of multi-dimensional data assets every day, covering core dimensions such as vehicle operation status, driving behavior, location information and vehicle performance. In the face of increasingly prominent personalized demands in the consumer market and fierce industry competition, the traditional car dealership system is facing three core challenges: first, the "uniform" product matching dilemma caused by ambiguous customer profiles; Second, the conversion rate bottleneck caused by the decline in the effectiveness of offline channels; Third, brand value dilution due to service homogeneity. Precision marketing solutions based on big data mining from the Internet of Vehicles are becoming a strategic breakthrough to restructure the relationship between people and vehicles and break through the growth ceiling. Some automakers are beginning to use Internet of Vehicles data for customer profiling analysis. By collecting data such as customers' driving habits, frequency of use, and vehicle preferences, they build customer profiles to achieve precision marketing.

The emergence of Internet of Vehicles data brings new opportunities to the automotive sales sector. Through in-depth mining and analysis of these data, intelligent automotive sales can be achieved, and the monitoring capabilities of the sales system can be enhanced to better meet customer needs and improve the market competitiveness of enterprises. For example [1], in order to explore the application of big data technology in the development of vehicle reliability test specifications, the Internet of Vehicles big data was used to conduct characteristic analysis in the field of vehicle reliability test specification development, including the analysis of user behavior characteristics, the statistics of the number of vehicle component operations, the prediction of the proportion of user road types, the identification of user driving styles, and the formulation of user working conditions. It provides new ideas and reference value for the rational development of its test specifications. In [2] order to study the driving behavior of electric vehicle users, a driving behavior evaluation model based on Internet of Vehicles big data is proposed. It analyzes and studies the driving behaviors of electric vehicle users from the dimensions of driving habits, safety coefficients, and energy consumption respectively. The results showed that the adopted method could effectively evaluate the actual driving behavior of users.

This study aims to build a monitoring system for intelligent car sales based on Internet of Vehicles data, deeply explore the value of Internet of Vehicles data in car sales, and use advanced technical means to achieve real-time monitoring and analysis of car brand sales. By analyzing customer behavior, we aim to accurately grasp customer needs and achieve personalized marketing; By monitoring market trends, it provides a basis for the company's strategic decisions. The study is of great practical significance and can help automotive companies improve sales performance, increase customer satisfaction and gain an edge in the fierce market competition. At the same time, it provides theoretical support and practical reference for the further application of Internet of Vehicles technology in the automotive sales field and promotes the deep integration of the automotive industry with information technology.

2. RESEARCH ROUTE

The research route of this paper mainly consists of four parts. The first part is the characteristics of Internet of Vehicles data, mainly introducing the characteristic dimensions of Internet of Vehicles data; The second part is data preprocessing, dealing with abnormal and missing data, and constructing a standard database with unified rules and complete data; The third part is the architecture of the intelligent sales system, which is mainly divided into modules such as data collection and preprocessing, data storage and management; The fourth part is intelligent monitoring of the system, which analyzes and monitors in real time key indicators such as vehicle model sales trends and price sales trends. The roadmap for this study is shown in Figure 1.

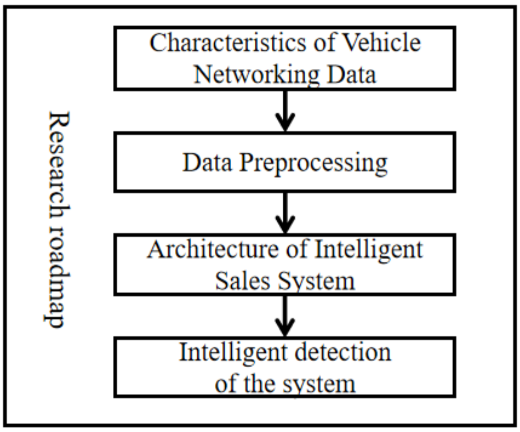

Figure 1. The research roadmap of this paper

2.1. Characteristics of Internet of Vehicles Data

Internet of Vehicles data covers all kinds of information generated and circulated in the vehicle network environment, and its data sources are diversified, mainly including real-time operating parameters collected by on-board sensor devices, interaction records generated by on-board entertainment information systems, as well as industry analysis data and user feedback information obtained through market research channels. The characteristics of Internet of Vehicles data are shown in Table 1 below.

Table 1. Characteristics of Internet of Vehicles Data

Serial numbers	Characteristics	Description
1	Multi-source heterogeneity	The Internet of Vehicles data system is multimodal, including structured data with standardized fields (such as vehicle travel latitude and longitude data), semi-structured data with partial organizational formats (such as log sequences generated by in-vehicle systems), and unstructured data containing original content such as spatio-temporal trajectories and consumer research.
2	The ocean of data	With the widespread application of intelligent connected technologies, the number of connected vehicles has continued to rise, and each intelligent connected vehicle is constantly generating data while in motion, leading to an explosive growth in data volume
3	Timeliness characteristics	The Internet of Vehicles data has the feature of instant synchronization, which ensures that critical information such as vehicle status and environmental perception data is transmitted and processed quickly after generation, forming a dynamically updated data stream.
4	Correlation	There are intricate correlations among the various data elements in connected vehicle data, such as a direct causal relationship between vehicle speed and fuel consumption, and a close correlation between driving behavior and vehicle failure rate.

2.2. Data Preprocessing

Data preprocessing is a key prerequisite for Internet of Vehicles data analysis, aiming to transform raw Internet of Vehicles data into analyzable and modelable high-quality data assets. In view of the multi-source heterogeneity, real-time dynamics and other characteristics of Internet of Vehicles data, the data preprocessing flowchart in this paper is shown in Figure 2.

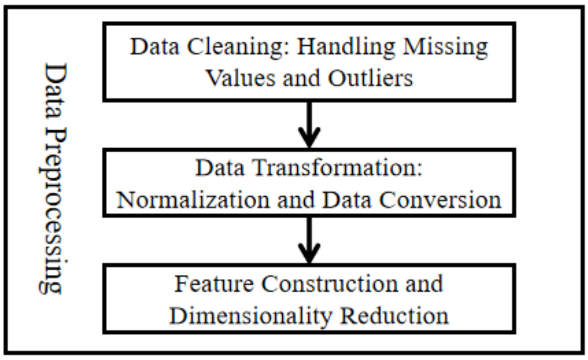

Figure 2. Flowchart of data preprocessing in this paper

2.2.1. Data cleaning

(1) Outlier detection and processing

Outliers are observations in a data set that deviate significantly from the majority of data points, and their sources are complex and diverse, possibly due to factors such as the characteristics of the data generation mechanism itself, errors in data collection or recording, mistakes in data processing, or actual extreme phenomena.

a. Detection of outliers in normally distributed data

Normal distribution data outlier detection uses the Z-Score method, which calculates the distance between data points and the mean and measures the degree of deviation with the standard deviation (σ). (The threshold is usually set at ±3σ), and the calculation formula is as shown in Formula (1).

$$Z = \frac{x_i - \mu}{\sigma} \tag{1}$$

x_i is the original data, μ is the mean, and σ is the standard deviation. $|Z| > 2$ is a mild outlier, $|Z| > 3$ is an extreme outlier (the industry generally considers $|Z| > 3$ as an outlier).

b. Detection of outliers in non-normally distributed data

For non-normally distributed data, the IQR (interquartile range) method is used to detect out [3] liers. The formula for locating outliers is shown in Formula (2) below. The meaning of the formula parameters of the interquartile range method is shown in Table 2.

$$\begin{aligned} Z_{up} &> Q_3 + 1.5 \times IQR \\ Z_{low} &< Q_1 - 1.5 \times IQR \end{aligned} \tag{2}$$

Table 2. Meaning of formula parameters of interquartile range method

Parameters	Meaning
Z_{up}	Outliers (upper outliers)
Q_3	Indicates the upper quartile, that is, all samples arranged from large to small, at the 25% position
IQR	The calculation formula is: $Q_3 - Q_1$
Q_1	Represents the lower quartile, that is, all samples arranged from large to small, at the 75% position
Z_{low}	Outliers (upper outliers)

c. Outlier data processing

The method for handling outliers in this article is: eliminate the outliers and treat them as missing values. The method for handling missing values is shown below.

(2) Detection and processing of missing values

a. Detection of outlier data

Quickly locate the distribution and proportion of missing values through descriptive statistics or data aggregation.

b. Outlier data processing

First: The deletion method, that is, directly deleting samples or features that contain missing values (row/column deletion). This method is based on the assumption of data integrity and is suitable for features or samples with an extremely high proportion of missing values (e.g. More than 80%).

Second: Interpolation, that is, filling the missing values with the mean or median of the features. This method is used for numerical data, data sets where the feature distribution is not significantly skewed. The calculation formula is as shown in Formula (3).

$$x_f = \frac{\sum_{i=1}^{n} x_i}{n} \tag{3}$$

2.2.2. Data transformation

(1) Data normalization

Data normalization is one of the key steps in data preprocessing, which aims to transform feature data of different ranges and dimensions into numerical values at a uniform scale, eliminating dimensional differences and numerical levels between variables, and making it easier for the model to learn the patterns in the data.

The principle of data normalization: Scale the data to a specified range to eliminate the influence of dimensions. It can be used for datasets with a relatively uniform distribution of values and no obvious outliers. Maximum-minimum normalization [4] calculation formula such as Formula (4)

$$x^{'} = \frac{x - \min(x)}{\max(x) - \min(x)} \times (n\_\max - n\_\min) + n\_\min \tag{4}$$

Among them, $x^{'}$ is the normalized data, $\max(x)$ is the maximum value of the original data, $\min(x)$ is the minimum value of the original data, $n\_\max$ is the maximum value of the normalized data, $n\_\min$ is the minimum value of the normalized data.

(2) Data transformation

Power transformation is an effective tool for adjusting distribution patterns in data preprocessing, suitable for skewed data and models that need to satisfy the normality assumption. Most of the user feedback and market research data in this paper conform to the skewed normal distribution, so power transformation is required. The formula for calculating the power transformation is as in Formula (5).

$$x(\lambda) = \begin{cases} \frac{x^{\lambda} - 1}{\lambda}, \lambda \neq 0 \\ \log(x), \lambda = 0 \end{cases} \tag{5}$$

Among them, for the λ transformation parameters, the optimal value needs to be determined through cross-validation (usually within the search range of [-3, 3]).

2.2.3. Feature construction and dimensionality reduction

(1) Feature construction

By constructing new features [5] through two or more dimensions, such as the longitude and latitude of vehicle operation in Internet of Vehicles data, the activity radius and activity center can be constructed to facilitate understanding the characteristics of consumers' lives. The formula is shown in (6).

$$V_{new} = \langle V_1 \oplus V_i \oplus V_n \rangle \tag{6}$$

Formula (6) meaning: multiple dimensions combined into one feature dimension.

(2) Feature dimension reduction

Feature dimension reduction is the process of transforming a high-dimensional feature space into a low-dimensional one through some mathematical method. Its core objective is to reduce the number of features and the complexity of the data while retaining as much key information as possible, thereby improving model efficiency, avoiding overfitting, and enhancing the interpretability [6-7] of the data.

Key features such as speed, acceleration, average acceleration, fuel consumption, sudden acceleration, sudden deceleration and energy consumption can be extracted from vehicle driving data to reflect the user's driving habits, vehicle performance and driving smoothness index. For example, the driving stability index can be measured by the variance of acceleration, the smaller the variance, the smoother the driving. The formula for calculating the driving stability index is shown in Formula (7).

$$P = \frac{1}{N}\sum_{i=1}^{N}(a_i - \bar{a})^2 \tag{7}$$

Where, in formula (7) a_i is the acceleration at the i-th time point, $\bar{a}$ is the average acceleration, and N is the total number of time points.

2.3. Intelligent Sales System Architecture

The automotive intelligent sales system based on Internet of Vehicles data adopts a layered architecture design, mainly including the data acquisition layer, data transmission layer, data storage and processing layer, business logic layer and application layer. The intelligent sales system architecture is shown in Figure 3.

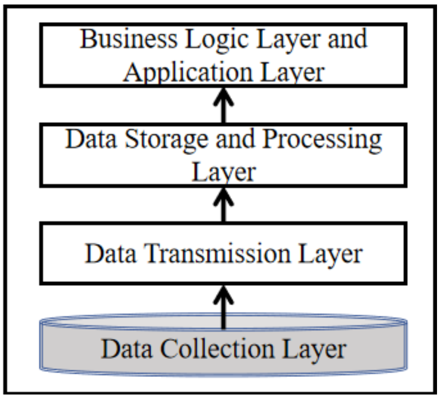

Figure 3. Architecture of the intelligent sales system

(1) Data collection layer: Responsible for collecting data from multiple data sources such as Internet of Vehicles devices, sales systems, and customer relationship management systems (CRM). Connected vehicle devices collect vehicle operation data, driving behavior data, etc. through various sensors.

(2) Data transmission layer: The collected data is transmitted securely and quickly to the data storage and processing layer using wireless communication technology. To ensure the reliability of data transmission, technical means such as data encryption and data verification are adopted.

(3) Data Storage and processing layer: Big data storage technology is adopted to store massive amounts of Internet of Vehicles data, and data mining, machine learning and other technologies are utilized to clean, preprocess, analyze and mine the data. The data processing layer refers to the data preprocessing module mentioned in Section 2.2 above, which processes missing values and outliers through mathematical methods, transforms data based on the distribution characteristics of Internet

of Vehicles data, and uses the relationship between feature dimensions, feature construction and feature dimension reduction.

(4) Business logic and application layer: Based on the data analysis results provided by the data storage and processing layer, develop personalized sales strategies and provide precise product recommendations for customers.

2.4. System Intelligent Monitoring

2.4.1. Intelligent recommendation module

The intelligent recommendation function is realized based on vehicle-network-related data information combined with machine learning algorithms. By analyzing the customer's historical behavior data, interests and preferences, and current market trends, recommend the vehicles and related services that best meet the customer's needs. For example, for a customer who frequently pays attention to new energy vehicles and focuses on vehicle range, recommend new energy vehicle models with longer range and provide relevant charging facility information and preferential policies. The intelligent recommendation module can also make dynamic adjustments based on the customer's real-time behavior, such as when the customer is browsing a certain model, the system automatically recommends related competing models and comparison analysis reports to help the customer make a smarter purchase decision.

2.4.2. Sales data analysis and forecasting module

Conduct in-depth analysis of the sales data of a certain model in the automotive market, and through techniques such as statistical analysis and data mining, uncover the underlying patterns and trends in the sales data to provide support for sales decisions. For example, analyze the sales situation in different regions and different sales channels, identify the regions and channels with better sales performance, and provide a basis for market expansion and channel optimization; By analyzing customers' purchasing behavior, predict customers' purchase intentions and purchase times, and develop marketing strategies in advance. At the same time, use methods such as time series analysis and regression analysis to predict future sales trends to help enterprises make reasonable production plans and inventory management. Sales data analysis and forecasting module

2.4.3. Customer behavior monitoring model

By monitoring the customer's browsing behavior on the car APP website or mobile application, such as the time spent on browsing pages, browsing sequence, browsing frequency, etc., the customer's interests and priorities can be understood. For example, if a customer spends a long time on the page of a certain model and visits the detailed parameters and configuration information of that model multiple times, it indicates that the customer has a high interest in that model, and the salesperson can follow up in a timely manner to provide more detailed product information and sales services. By analyzing the customer's browsing behavior path, the page layout and product display of the website or application can be optimized to enhance the customer's browsing experience and facilitate the customer's purchase decision.

2.4.4. Market trend monitoring indicators

Monitor the changing trends of sales volume and sales revenue of different vehicle models within a certain period, and analyze the influencing factors of vehicle model sales trends, such as changes in market demand, product launches by competitors, and adjustments in policies and regulations. By monitoring the sales trends of vehicle models, enterprises can adjust their product strategies in a timely manner, optimize the product structure, increase the production and promotion of best-selling models, and improve or eliminate models that do not sell well. For example, if the sales volume of a certain new energy vehicle model shows a rapid growth trend, the enterprise can increase the production capacity of that model, increase the investment in marketing promotion, and launch more derivative models; If the sales of a traditional fuel model continue to decline, the company may consider upgrading it or gradually reducing the production scale of that model.

3. CONCLUSIONS AND ANALYSIS

Connected vehicle data has great application value in automotive marketing. The development of Internet of Vehicles technology enables cars to generate large amounts of data in real time, covering multiple dimensions such as vehicle operation status, driving behavior, location information, and user preferences. Through in-depth mining and analysis of Internet of Vehicles data, automotive companies can achieve precise customer insights and positioning. Customer profiles built based on Internet of Vehicles data are more comprehensive and accurate, clearly presenting the basic attributes, consumption behavior, driving habits, location information and user preferences of customers, providing strong support for enterprises to formulate personalized marketing strategies. By analyzing Internet of Vehicles data, enterprises can accurately predict customer needs, achieve personalized marketing, and improve marketing effectiveness and customer satisfaction.

In the future, it will be necessary to constantly improve and innovate data analysis algorithms and models to enhance the efficiency and accuracy of data mining. Explore more advanced machine learning algorithms and deep learning models to achieve more precise personalized marketing and improve marketing effectiveness and customer satisfaction.

REFERENCES

[1] Lu Haibo, Chen Chengfeng, Chen Chunyan. Research on the application of big data from Internet of Vehicles in automotive reliability testing [J]. Automotive Practical Technology, 2024, 49 (18): 87-91. DOI:10.16638/j.cnki.1671-7988.2024.018.016.

[2] Li Zonghua, Zhai Jun, Wang Xianjun, et al. Electric vehicle driving evaluation application based on Internet of Vehicles data [J]. Automotive Practical Technology, 2022, 47 (13): 15-20. DOI:10.16638/j.cnki.1671-7988.2022.013.005.

[3] Wu, Z., Yang, J., Zhang, F. et al. Exploring automobile brand power: A multi - dimensional approach. Environ Dev Sustain (2024). https://doi.org/10.1007/s10668-024-05089-0.

[4] Zhang Fan, Guo Yaxin, Yang Jing, et al. Research on prediction of electric quantity based on GBDT+ feature engineering method [J]. Electron Quality, 2020, (01): 1-4.

[5] Y. Jing, H. Huiying, Z. Fan and Z. Yishu, "Research on Automobile User Association Recommendation Algorithm Based on Feature Attribute and Matrix Decomposition," 2024 International Conference on Computers, Information Processing and Advanced Education (CIPAE), Ottawa, ON, Canada, 2024, pp. 594-600, doi: 10.1109 / CIPAE64326.2024.00114.

[6] Liu Yulin, Bai Yang, Cui Bin, et al. A review of automated feature engineering for machine learning [J]. Computer Applications and Software, 2025, 42 (01): 1-10+40.

[7] Sheng Zhongsong. User Purchase prediction based on feature learning and ensemble algorithms [D]. Sichuan university, 2021. DOI: 10.27342 /, dc nki. Gscdu. 2021.000464.