# Construction of performance evaluation model of production vehicle based on big data

Yuan Liu[a,b,*], Rui Shi[a,b,1]
[a]China Automotive Technology And Research Center Co., Ltd.; [b]China Auto Information Technology (Tianjin) Co., Ltd. China, Tianjin, 300300

## ABSTRACT

With the continuous development and progress of automobile design technology, most of the main engine manufacturers no longer develop new models through large-scale reverse mature models. However, the independent research and development and production capacity of automobile manufacturers are still obviously insufficient. Automobiles designed and produced independently are far behind foreign automobiles in terms of first failure mileage, average failure interval mileage, average life, reliable life and rated life. Data mining draws lessons from the knowledge of statistics, machine learning, database and other fields in its methods. Data mining was first carried out by experts in the field of database, and getting up early in the database is a very difficult professional technology to master. Based on the big data theory, this paper establishes a performance target evaluation model, and uses this model to set the performance target of mass production vehicles to improve their reliability. In this way, in the early development stage of the mass production vehicle, the performance target of the mass production vehicle can be formulated under the consideration of the market, which can not only reflect the market competitiveness of the quality, but also improve the reliability of the performance target. The results show that RF is more suitable for this experiment than DT and KNN models, and has better classification performance on data sets.

**Keywords:** Big data; production vehicle; performance evaluation

## 1. INTRODUCTION

After the 21st century, people's quality of life has been continuously improved, the demand for auto-abundant has been increasing, the sales hall of auto-abundant has soared, and the auto-abundant industry has developed rapidly. However, automobile quality problems appear constantly, and complaints about automobile quality increase. Reliability is an important index to measure automobile quality. Only by comprehensively and systematically analyzing the application status of reliability technology in China's automobile industry can we find out the problems, prescribe the right medicine and improve the quality of domestic automobiles, which is of great practical significance to China's automobile industry. Therefore, it is particularly important to set a reliable and competitive performance target [1-2].

With the continuous development and progress of automobile design technology, at present, most main engine manufacturers no longer develop new models through large-scale reverse mature models [3]. Professional automobile design companies are generally divided into six sections: general/general layout design, body design, chassis design, interior and exterior decoration design, and electrical/air conditioning design. With the increasing number of automobile design practitioners and the maturity and popularization of design technology, major engine manufacturers have gradually established their own design and research centers [4-5]. At present, China's automobile production and sales scale has ranked first in the world, but the independent research and development and production capacity of automobile manufacturers are still obviously insufficient. The automobiles designed and produced independently are far behind foreign automobiles in terms of first failure mileage, mean time between failures, average life, reliable life and rated life. How to improve the reliability of automobiles has become the key for domestic automobile enterprises to become bigger and stronger [6].

*liuyuan@catarc.ac.cn; [1]shirui@catarc.ac.cn

Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023), edited by Srikanta Patnaik, Tao Shen, Proc. of SPIE Vol. 12779, 1277914 · © 2023 SPIE · 0277-786X · Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2688726

Working condition research itself is based on statistical samples, and the ultimate goal is to make the test results of energy consumption test regulations conform to the average value used by users. The main engine factory needs to carry out development work for median sample working conditions. Data mining is a very common method, which uses computers to analyze and process data, find out the unique laws in a large number of data, and then get feedback to help people conduct rapid and accurate analysis [7-8]. By studying the application of big data and its related technologies in the performance evaluation of production vehicles, this paper improves the accuracy and efficiency of performance evaluation of production vehicles, and provides new analytical means and methods for performance evaluation of production vehicles.

## 2. RESEARCH METHOD

### 2.1 Parameter requirements of performance analysis model of mass production vehicle

The application of reliability technology in domestic automobile product development is still lacking. In the process of automobile design, it still relies on experience design, the reliability design system is not perfect, the relevant reliability management tools are not used, and the popularity of reliability knowledge is not enough; The distribution law of automobile faults can be approximately described by normal distribution, lognormal distribution, exponential distribution and Weibull distribution. It is generally believed that the first failure mileage of complex repairable systems follows Weibull distribution [9]. In the performance evaluation, it also needs to be carried out according to the requirements of test procedures and test specifications. After the test vehicle reliability driving project is completed, it is necessary to disassemble and inspect the internal structure of each assembly for other abnormal phenomena. In the performance evaluation of production vehicles, it is necessary to record and count the test conditions, faults and other test data for the performance evaluation and evaluation of production vehicles.

The reliability and durability of automobiles depend greatly on the reliability and durability of engines. The overhaul of engine is the main sign that determines the overhaul of automobile. There are many factors that affect the durability of the engine, and there are also many failure modes. The mileage of the car driven by the engine during the stable wear period. Generally, it is the starting mileage of deteriorating wear period minus the ending mileage of early wear period. The pattern of performance evaluation system of mass production vehicle in automobile design and development is shown in Figure 1.

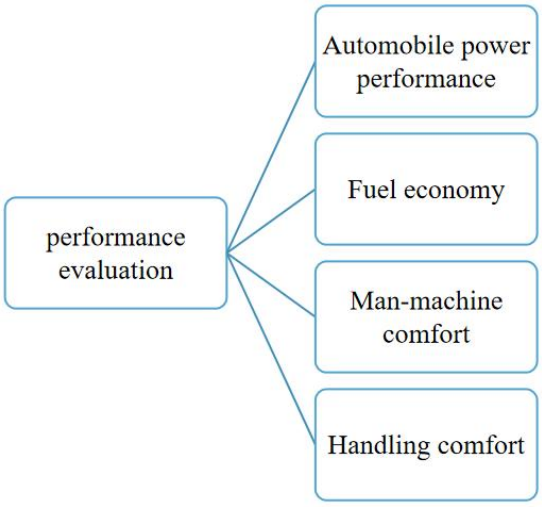


Figure 1. Performance evaluation system of mass production vehicle

As an efficient means of transportation, the transportation efficiency of automobile depends largely on the dynamic performance of automobile, so the dynamic performance index is the most basic and important index among various performance indexes of automobile, and the quality of dynamic performance directly affects the use of automobile on the

road. The efficiency of automobile transportation can be expressed by the driving force that can be distributed per unit mass of automobile to remove wind resistance.

$$D = \frac{F_t - F_w}{mg} \tag{1}$$

$F_t$ is the ramp resistance, $F_w$ is the clothing rolling resistance, and $m$ is the mass of the vehicle under the calculated load condition, in kg; $g$ is the acceleration of gravity in m/s $^2$.

Fuel economy is one of the basic properties of automobiles. Oil is the main energy source of modern industry, especially transportation. Good fuel economy can reduce the use cost of cars, save energy and protect the environment. In the actual calculation, the engine's external characteristic power is the actual calculated value obtained from the resistance power and the transmission efficiency, and then divided by 90% to obtain the engine's external characteristic power [10].

The ordinate of the engine universal characteristic curve is usually the engine torque, and the engine torque and engine power are converted by the following formula:

$$T = \frac{P_e \times 9550}{n} \tag{2}$$

Where: $T$ is the torque of the engine, and the unit is N·m. $P_e$ is the engine power in kW. $n$ is the engine speed, r/min.

The quality evaluation model is based on the equivalent volume and kerb quality of production vehicles, and with the help of the basic big data theory, the kerb quality is targeted in the early development of automobiles. Through a large number of data measurement of competing cars, it is found that there are great differences in $\lambda$ for different models of SUVs and cars. In general, when the $\lambda$ of SUVs is 0.662 and the $\lambda$ of cars is 0.5, it is more in line with the actual situation and the error of calculation results is the smallest [11]. Equivalent volume formula:

$$v = \left(\lambda(a-b)+d\right)bc/10^9 \tag{3}$$

Where: $a, b, c$ is vehicle length, vehicle width and vehicle height, mm; $d$ is the wheelbase, mm; $v$ is equivalent volume, m$^3$.

### 2.2 Construction of performance evaluation model for mass production vehicles

When developing a new component or vehicle model, designers should incorporate the concept of reliability design and embody it in the design process of the product. While pursuing high performance and structural advancement, they must base themselves on the reliability of the product and adopt new technologies, materials and processes to reassure users of the reliability, durability and safety of the product, thus improving the quality and market competitiveness of the product. Data mining technology can play an active role in evaluating the performance of mass production vehicles. In the development of enterprises, it is an important measure to choose a scientific and efficient way to improve efficiency. Through the application of data mining technology, it is convenient for enterprises to understand and analyze relevant data about automobile manufacturing. In data analysis, you can clearly see the whole database data.

Data mining draws lessons from the knowledge of statistics, machine learning, databases and other fields in methods. Data mining was first carried out by experts in the field of databases, and getting up early in the database is a very difficult professional technology [12]. With the development and commercial popularization of database technology, its operation is becoming simpler and simpler, and the reduction of learning cost makes more and more researchers in other fields begin to use and master database technology. Since then, data mining has gradually become a hot topic in other fields. Functions such as reliability prediction can be realized. In engineering practice, they are often combined and applied, which is convenient to give play to their respective advantages and complement each other, and can receive remarkable benefits with twice the result with half the effort.

Based on the big data theory, this paper establishes a performance target evaluation model, and uses this model to set the performance target of mass production vehicles to improve their reliability. In this way, in the early development stage of the mass production vehicle, the performance target of the mass production vehicle can be formulated under the

consideration of the market, which can not only reflect the market competitiveness of the quality, but also improve the reliability of the performance target.

Artificial neural network uses its nonlinear mapping idea and parallel processing method, and can express the related knowledge of input and output with its own structure. It is a multi-input and single-output nonlinear element, and its input-output relationship can be described as:

$$Y = f\left(\sum_{j=1}^{n} w_{ji} x_j - \theta_i\right) \tag{4}$$

Where $j = 1,2,\cdots,n$ is the input signal from other cells, indicating the connection weight from cell $j$ to cell $i$, and $n$ is the number of input signals. $Y$ is the neuron output. $f(\cdot)$ is called the transfer function.

Data mining has a wide range of applications, involving different fields, such as supermarket customer resource analysis, student data analysis and so on. In addition, there are different models and algorithms for data mining, such as DT(Decision tree), RF(Random forest), SVM(Support vector machine), KNN (k-nearest neighbor) and so on. Different models and algorithms are suitable for different scenarios. In this paper, KNN, DT and RF models are mainly used to analyze the quality of vehicles and predict whether the quality of vehicles is qualified or not through different characteristics of vehicles.

KNN is a stable and effective classification algorithm based on statistics. KNN algorithm has outstanding advantages: clear concept, simple implementation, good robustness and good classification effect. For a given unknown sample, calculate its similarity with each training sample and sort it, take the first k training samples, and judge the category of the unknown sample according to the category of the k samples, that is, the category of the k samples with the largest number of samples is the unknown sample.

Many DT implementations are based on the principle of information theory. In the process of DT formation, the most important part is the choice of split attributes. A common method is to calculate the information gain. The principle of information gain comes from information theory, which is the expected entropy reduction caused by using an attribute to divide the training set. The formation of DT is to recursively split each node in the data set until all categories of nodes belong to the same class or there are no redundant attributes to divide the training sample set.

Let $X_1, X_2, \cdots, X_n$ be the signal sent by the source. Before receiving $X_i$, the receiver's uncertainty about the signal sent by the source is defined as the self-information $I(X_i)$ of the information symbol, that is, $I(X_i) = -\log_2 P(X_i)$, where $P(X_i)$ is the probability that the source sends $X_i$.

Self-information can only reflect the uncertainty of symbols, and information entropy can be used to measure the uncertainty of the whole source $X$, which is defined as follows:

$$H(X) = -\sum_{i=1}^{n} P(X_i)\log_2(X_i) \tag{5}$$

For imbalanced classification, the simplest and most effective method may be the data level method, which directly preprocesses the data samples in the early stage to balance the data samples. For example, random downsampling and random oversampling methods reduce the number of negative samples by randomly removing negative samples and increase the number of positive samples by randomly copying positive samples to achieve the purpose of balancing samples.

RF is a classifier with multiple DTs as the basis, and the final prediction result depends on the voting of each DT's prediction result, with the class with more votes as the final prediction result. The advantage of combined classifier is that its results are more accurate than each individual classifier. In addition, when RF is divided each time, the accuracy of selecting a single or a small number of features is better than that of multiple features. Therefore, RF is also very effective on large data sets. RF also potentially gives an assessment of the importance of features.

## 3. EXPERIMENTAL ANALYSIS

There is a complex relationship between the evaluation indexes involved in automobile braking performance testing and the parameters such as automobile model and axle load distribution. This relationship is manifested in a large number of automobile braking testing records as a qualitative correlation between the indexes. Quantitatively "mining" this relationship is to reflect the quantitative association rules between the indexes and parameters. Through the previous data cleaning, integration, transformation, reduction and dispersion, we can now mine association rules from the processed data.

In this paper, a data set about automobile quality prediction is downloaded from the data warehouse website, and each data corresponds to a label, which indicates the quality level of the automobile. The labels of the whole data set have four categories. In this paper, the five-cross validation method is used to carry out the experiment.

The calculation result of power balance is shown in Figure 2. It can be seen that at the highest speed of the fourth gear, the corresponding power is 80.31kw, and at this time, more reserve power leads to higher fuel consumption. At the highest speed of the fifth gear, the corresponding power is 67.1kw. At this time, the fuel economy is very good with less reserve power.

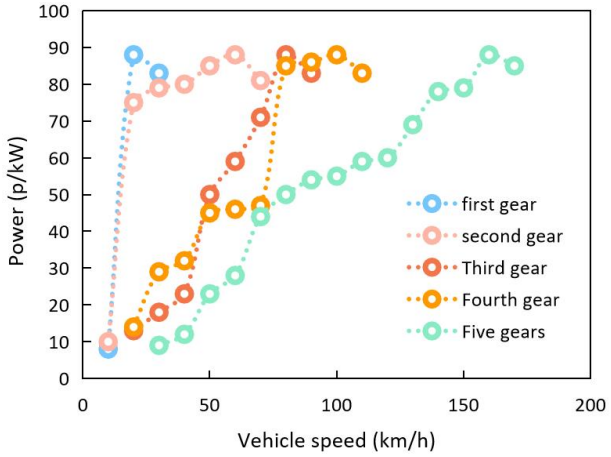


Figure 2. Power balance calculation

Next, we use KNN discrete test data to mine, mainly to find the correlation between braking performance and technical parameters of the car itself. Similarly, set the corresponding input and output parameters, minimum support, minimum confidence and other parameters. See Table 1 for the specific mining structure.

Table 1. Mining results

| evaluating indicator | | Correlation parameter | | Support (%) | Confidence (%) |
|---|---|---|---|---|---|
| name | attribute value | name | attribute value | | |
| Front axle brake balance | high | Axle load of front axle | high | 14.968 | 59.627 |
| | | Old and new degree | high | | |
| | | Axle load of front axle | middle | 14.25 | 50.685 |
| | | Old and new degree | low | | |
| Rear axle brake balance | middle | Axle load of front axle | middle | 20.633 | 57.592 |
| | | Old and new degree | middle | | |
| | | Axle load of front axle | high | 21.979 | 59.833 |
| | | Old and new degree | high | | |

The braking balance of automobile is mainly related to the old and new degree of automobile and axle load. As can be seen from Table 1, whether the front axle braking balance or the rear axle braking balance, it is required that the old and new degree of the car is at an advanced level. And the corresponding axial load is not very large. This can be understood from two aspects. On the one hand, the brake balance of the car will not change much in the early stage, but with the growth of the car age, the brake balance will have problems before the brake rate; On the other hand, the greater the axle load of a car, the more likely it is to have problems in its braking balance.

Comparing the KNN, DT and RF models, in terms of accuracy, the comparison results of the three models are shown in Figure 3.

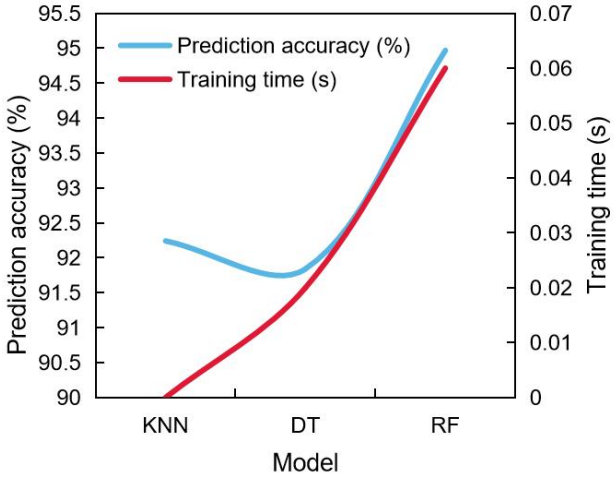


Figure 3. Comparison of training time and prediction accuracy of the model

DT has the lowest accuracy rate, KNN is in the middle, and RF has the highest accuracy rate, and the accuracy rate, recall rate and F value in the four categories are all high, and the pre-judgment result is better. Therefore, it can be concluded that RF is more suitable for this experiment than DT and KNN models and has better classification performance on data sets.

## 4. CONCLUSION

Car quality problems are constantly emerging, and complaints about car quality are increasing. Reliability is an important index to measure automobile quality. Only by comprehensively and systematically analyzing the application status of reliability technology in China's automobile industry can we find out the problems, prescribe the right medicine and improve the quality of domestic automobiles, which is of great practical significance to China's automobile industry. By studying the application of big data and its related technologies in the performance evaluation of production vehicles, this paper improves the accuracy and efficiency of performance evaluation of production vehicles, and provides new analytical means and methods for performance evaluation of production vehicles. The results show that RF is more suitable for this experiment than DT and KNN models, and has better classification performance on data sets.

## REFERENCES


[1] Qiao, W. D., A performance evaluation method of drive motor system for electric vehicles. Motor and Control Applications, 44(7), 7 (2017).

[2] Liu, C., Li, Y., Xia, L., Li, H., Li, B., Ma, X. F., Zhang, J., Comprehensive performance evaluation of electric vehicle charging stations. Journal of North China Electric Power University: Natural Science Edition, 44(1), 7 (2017).

[3] Wang, D., Design and performance evaluation of can bus in automobile control system. Small internal combustion engines and motorcycles, 047(004), 54-56 (2018).

[4] Yue, Y. N., Wu, Y., Design and implementation of special vehicle virtual test application system for comprehensive performance evaluation. Journal of Automotive Engineering, 9(1), 9 (2019).
[5] Niu, C. Y., Zeng, J., Xu, J. X., You, G. P., Performance test and evaluation of automobile equipped with automatic emergency braking system. Journal of Hebei University of Technology, 47(5), 7 (2018).
[6] Wang, D., Design and performance evaluation of can bus in automobile control system. Small internal combustion engines and motorcycles, (004), 047 (2018).
[7] Tu, A., Luo, X. H., Li, X., Han Z. J., Research on the shift quality evaluation system of automatic transmission vehicles China test, 43 (11), 5 (2017).
[8] Zhang, M., Long, R. X., Liu, P. J., He, Z., Research on the factors affecting the preparation time of automobile product recall Industrial Engineering and Management, 27 (4), 8 (2022).
[9] Wang, Y., Mi, H., Xu, H., Zang, J. B., Wang, H. Y., Carbon emission assessment and economic analysis of refrigerant recovery and regeneration Journal of Refrigeration, 42 (5), 8 (2021).
[10] Sun, Y., Construction of transaction service system for used motor vehicles based on price evaluation model. Information Technology, 42(12), 4 (2018).
[11] Hao, G., Jin, T., Evaluation of rolling bearing performance deterioration based on hidden Markov model. Journal of Chongqing Jiaotong University (Natural Science Edition), 40(02), 123-128 (2021).
[12] Xiong, J., Shi, J. H., Wan, H. S., Construction of comprehensive risk field model for people, vehicles and roads and evaluation of driving style. journal of transportation systems engineering and information technology, 21(6), 105-114 (2021).