Research on Multi-Source Heterogeneous Data Fusion Technology of New Energy Vehicles Under the New Four Modernizations

Fan Zhang*, Jing Yang, Chuan Sun, Xin Guo, Tiantian Wan
China Automotive Research (Tianjin) Automotive Information Consulting Co., Ltd., Tianjin 300000, China

*Corresponding author e-mail: zhangfan2019@catarc.ac.cn

Abstract. In order to better adapt to the development trend of the city, new energy vehicles are widely used to fit the development concept of green transportation. As the focus of smart city development, big data is indispensable for its application in new energy vehicle guarantees, especially in maintenance and machinery guarantees. In the operation and maintenance of new energy vehicles, the application of multi-source heterogeneous data technology is extremely important. Based on this research background, the paper introduces and constructs the new energy vehicle fault feature vector of the new energy vehicle, gives a multi-source time domain frequency domain data fusion new energy vehicle fault diagnosis method, and uses the neural network to give the basic probability distribution. The evidence theory fuses the signals of each sensor to get the diagnosis result.

Keywords: Multi-source heterogeneous data technology, new energy vehicle, data model processing, new energy vehicle fault diagnosis.

1. Introduction
At present, most of the research on the reliability of new energy vehicles at home and abroad is the experimental research done by enterprises under the mandatory requirements of government regulations, that is, recording failures and performing failure mode analysis to prepare for improvement and optimization, and at the same time prove the reliability of their products Meet the standards and regulations, but the in-depth research on the distribution of vehicle failures is very lacking. In particular, most domestic research on new energy vehicles focuses on the development of key components and systems, and there is not much systematic research on vehicle reliability [1]. This research focuses on the use and maintenance of new energy vehicles of the majority of car owners. Based on the collection of a large number of multi-source and heterogeneous data on the failure of new energy vehicles, a statistical analysis theory of vehicle model failure laws is formed. At the same time, the paper presents a multi-source time-domain and frequency-domain data fusion new energy vehicle fault diagnosis method, using neural network to give the basic probability distribution, evidence theory fusion of the signals of each sensor to obtain the diagnosis result.
2. New energy vehicle fault feature vector

2.1. New energy vehicle service failure time characteristic quantity

The time-domain parameters commonly used for fault diagnosis of new energy vehicles are divided into dimensional parameters and non-dimensional parameters. Dimensional parameters mainly include peak value, root mean square value, absolute mean value, and variance [2]. The peak value is the maximum instantaneous value that the signal may appear, expressed as

\[ x_p = \max[x(t)] \]  \hspace{1cm} (1)

The root mean square value reflects the power of the signal and is defined as

\[ x_{\text{rms}} = \sqrt{\int_{-\infty}^{\infty} x^2 p(x)dx} \]  \hspace{1cm} (2)

Can be approximately calculated as

\[ x_{\text{rms}} = \sqrt{\frac{1}{T} \int_{0}^{T} x^2(t)dt} \]  \hspace{1cm} (3)

The absolute mean is defined as

\[ x_{av} = \int_{-\infty}^{\infty} |x| p(x)dx \]  \hspace{1cm} (4)

Can be approximately calculated as

\[ x_{av} = \frac{1}{T} \int_{0}^{T} |x(t)|dt \]  \hspace{1cm} (5)

Variance describes the fluctuation degree of the signal deviating from the central trend, it is the dynamic component of the signal, and the variance is defined as

\[ \sigma_x^2 = \int_{-\infty}^{\infty} (x - \mu_x)^2 p(x)dx \]  \hspace{1cm} (6)

In the formula, \( \mu_x \) represents the mean value. The above formula can be approximately calculated as

\[ \sigma_x^2 = \frac{1}{T} \int_{0}^{T} (x(t) - \mu_x)^2 dt \]  \hspace{1cm} (7)

Normalize them to obtain dimensionless digital characteristic parameters: waveform index \( S \), peak index \( C \), impulse index \( I \), Yu Degree \( L \) and kurtosis index \( K \), etc., respectively, are defined as follows

\[ S = \frac{x_{\text{rms}}}{x_{av}} \]  \hspace{1cm} (8)

\[ C = \frac{x_p}{x_{\text{rms}}} \]  \hspace{1cm} (9)

\[ I = \frac{x_p}{x_{av}} \]  \hspace{1cm} (10)

\[ L = \frac{x_p}{x_e} \]  \hspace{1cm} (11)

\[ K = \frac{\int_{-\infty}^{\infty} x^4 p(x)dx}{\left[ \int_{-\infty}^{\infty} x^2 p(x)dx \right]^2} \]  \hspace{1cm} (12)

Where \( x_e \) is the square root amplitude.

2.2. New energy vehicle failure frequency domain characteristics of new energy vehicle failure

New energy vehicle fault frequency domain characteristic parameters are selected as the percentage of signal energy in each decomposition frequency band to the total energy. The fault signal is demodulated by the wavelet cluster envelope, and the demodulated signal is decomposed by a 3-layer wavelet packet using db20 wavelet, thus forming a frequency band on the scale 3. The wavelet packet decomposition
The coefficients of each wavelet packet are reconstructed, the signals of each frequency band are extracted, and the energy $E_{k}(k = 0,1,2,L,7)$ and total energy $E$ of each frequency band signal are calculated to obtain the percentage of the signal energy $E_{k}$ of each decomposed frequency band to the total energy $E$.

### 2.3. Feature vector selection

The paper selects time-domain characteristic parameters: waveform index $S$, peak index $C$, impulse index $I$, ridge $L$ and kurtosis index $K$. The new energy vehicle fault frequency domain characteristic parameters select the signal energy of 7 decomposition frequency bands as a percentage of the total energy, and 12 characteristic quantities constitute a characteristic vector.

### 3. Multi-source time domain and frequency domain data fusion new energy vehicle fault diagnosis system

Multi-source time-domain and frequency-domain data fusion new energy vehicle fault diagnosis system includes three modules: data-level fusion module, feature-level local diagnosis module, and decision-level D-S evidence theory fusion diagnosis module [3]. The principle of this system is shown in Figure 2.
3.1. Data-level fusion module
The data-level fusion module mainly performs multi-sensor data collection and feature extraction. In order to diagnose the fault of the new energy vehicle, the required information is obtained from the state detection of the diagnosed object through multiple sensors, and the signal is input to the new energy vehicle fault diagnosis computer through the relevant conversion circuit.

3.2. Feature-level local diagnosis module
There are many feature-level local diagnosis methods. Neural networks, Bayesian theory, D-S evidence theory, etc. are commonly used methods. Here, a BP neural network with the same structure is used, each neural network is the most basic three-layer BP algorithm, and the network error is set to 0.005. The disadvantages of BP network are slow convergence speed and easy to fall into local minima [4].

3.3. The fusion diagnosis module of decision-level D-S evidence theory
This module takes the basic probability distribution of different states given by the feature-level parallel local diagnosis module as input, and uses the evidence combination formula in D-S evidence theory to fuse the local diagnosis results to obtain the final diagnosis result.

In a recognition framework $\Theta$ of proposition $A$, there is a set function $m:2^{\Theta} \rightarrow [0,1]$ that satisfies
\[
\begin{align*}
\sum_{A \subseteq \Theta} m(A) &= 1 \\
m(\emptyset) &= 0
\end{align*}
\]  

(13)

Then $m(A)$ is called the mass function of $A$ on the frame $\Theta$, also called the basic probability distribution, which indicates the degree of accurate trust in $A$.

For proposition $A$, its trust function is defined as
\[
Bel(A) = \sum_{B \subseteq A} m(B), \forall A \subseteq \Theta
\]  

(14)

The likelihood function is defined as
\[
Bel(A) = \sum_{B \subseteq A} m(B), \forall A \subseteq \Theta
\]  

(15)
Figure 3: Description of evidence interval in D-S evidence theory

Figure 3 shows the description of the evidence interval in D-S evidence theory. When the system makes a decision, it will select a value in interval $(P(A), Bel(A))$ as the final reliability of the proposition, and the highest reliability among all candidate propositions is the decision result. For the same proposition, different decision rules will produce different reliability. If $A \subseteq \Theta$ and $m(A) > 0$, then $A$ is called the focal element, and $m_1, m_2, \cdots, m_n$ is the basic credibility distribution on the same recognition frame $\Theta$, the focal elements are $A_1, A_2, \cdots, A_k$ and $B_1, B_2, \cdots, B_n$ respectively, and

$$K = \sum_{A \in \Theta \setminus \{\phi\}, k=1}^{d} m(A) \cdot m(B) \cdot m(C) \cdot \cdots < 1$$

$$\gamma(A) = \sum_{A \in \Theta \setminus \{\phi\}, k=1}^{d} m(A) \cdot m_2(B) \cdot m_3(C) \cdot \cdots$$

Then, the synthesized mass function $m : 2^n \rightarrow [0,1]$ is as follows

$$m(A) = \begin{cases} 0, & A = \phi \\ \frac{\gamma(A)}{1 - K}, & A \neq \phi \end{cases}$$

Among them, $K$ is called the uncertainty factor, $\gamma(A)$ is called the influence factor of the mass function.

4. Experimental verification

For AMT system, drive motor system and power battery system fault diagnosis and fault-tolerant model simulation analysis, due to the numerous failure units of each assembly, one of the representative failures is selected for each assembly system for offline simulation analysis and research. The AMT system takes the gear selection mechanism/gear position sensor fault diagnosis and fault tolerance model simulation as an example. The paper selects the CAN bus data fragments collected by the remote data acquisition system during the operation of a hybrid bus, exports the data from the database and transfers the data to Workspace through MATLAB, as a fault-tolerant model for the gear selection mechanism/gear position sensor diagnosis the simulation input parameters. According to the input parameters, the gear position sensor feedback gear position between 0-84s is consistent with the gear position reflected by the speed ratio and corresponds to the TCU control gear selection signal [5]. No fault occurs during this period; after 84s, the TCU requires a change To the fifth gear, and the gear position sensor and the gear ratio reflect that the gear is in the fourth gear at this time, and the abnormal situation occurs continuously, it is considered that the gear selection mechanism has failed to execute the gear selection command, the actuator has malfunctioned and the malfunction occurs at 84s The flag
fault_shift_mechanism stepped from 0 to 1, and the gear position sensor fault flag output is 0, as shown in Figure 4 below [6].

![Shift selection mechanism/gear position sensor fault flag output](image)

**Figure 4.** Shift selection mechanism/gear position sensor fault flag output

It can be seen from Figure 4 that after a failure occurred in 85s, the input and output parameters of the simulation model were compared to prove that the shift process signal was shielded; after that, the vehicle kept the current gear during 84-95s; when the brake signal was simulated at 95s, the model outputs a step from 1 to 0 for the motor on state, and a step from 0 to 1 for the off state of the engine, which means that the fault-tolerant strategy is executed so that both the motor and the engine are turned off.

5. Conclusion

The new energy vehicle multi-source time-domain frequency-domain data fusion fault diagnosis method uses evidence theory for decision-level fusion, which improves the accuracy of diagnosis and reduces the uncertainty of diagnosis. The test results show that the fault diagnosis method is effective.

References

[1] Wang, Y., Zhang, Y., Qian, Z., Wang, S., Hu, Y., & Yin, B. Multi-source traffic data reconstruction using joint low-rank and fundamental diagram constraints. IEEE Intelligent Transportation Systems Magazine, 11(3) (2019) 221-234.

[2] Zhang, D., He, T., & Zhang, F. National-scale traffic model calibration in real time with multi-source incomplete data. ACM Transactions on Cyber-Physical Systems, 3(2) (2019) 1-26.

[3] Chen, S., & Hu, D. Parameterizing anthropogenic heat flux with an energy-consumption inventory and multi-source remote sensing data. Remote Sensing, 9(11) (2017) 1165-1176.

[4] Zhang, W., Zhang, Z., & Chao, H. C. Cooperative fog computing for dealing with big data in the internet of vehicles: Architecture and hierarchical resource management. IEEE Communications Magazine, 55(12) (2017) 60-67.

[5] Xie, X., Zhang, F., & Zhang, D. PrivateHunt: Multi-source data-driven dispatching in for-hire vehicle systems. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2(1) (2018) 1-26.

[6] You, L., Zhao, F., Cheah, L., Jeong, K., Zegras, P. C., & Ben-Akiva, M. A generic future mobility sensing system for travel data collection, management, fusion, and visualization. IEEE Transactions on Intelligent Transportation Systems, 21(10) (2019) 4149-4160.