ISSN 2004-2965 Energy Proceedings, Vol. 24, 2021

International Conference on Applied Energy 2021
Nov. 29 - Dec. 5, 2021, Thailand/Virtual
Paper ID: 391

A Comprehensive Data-driven Fault Diagnosis Method for Electric Vehicles

Xiang Zhang[1], Zekun Zhang[1], Zhiqiang Wu[1], Zhaosheng Zhang[1], Ni Lin[1*]

1 Collaborative Innovation Center for Electric Vehicles in Beijing & National Engineering Laboratory for Electric Vehicles, Beijing Institute of Technology, Beijing 100081, China.

2 Advanced Technology Research Institute, Beijing Institute of Technology, Jinan 250300, China

ABSTRACT

Recent years have witnessed a transition in energy structure, where large number of electronic devices and systems are introduced in multiple fields including industry, academia, commerce, and so forth. Safe and efficient operations of these systems are critical to ensure productivity as well as to avoid hazard, which poses strict demands on fault diagnosis. Most of traditional methods tend to focus on algorithm design or certain types of hardware or software flaws under given operational conditions, thus not suitable for modern electronic systems that may suffer from a variety of different faults. In this paper, on top of our previous work on big data, a systematic way of fault diagnosis in electric vehicles is put forward, which covers data processing, feature extraction, model-based diagnosis, and model fusion. The proposed method is trained and validated using data from real-world electric vehicles, which are representative examples of modern complex systems. Results show a diagnosis accuracy over 95% can be achieved with a comprehensive consideration of fault modes under a variety of operational scenarios. The proposed algorithm can also be used to indicate key features leading to faults so that system level upgrade can be performed accordingly. The design criteria and idea of the algorithm is also adaptable to other systems or applications with minor changes.

Keywords: electric vehicles, fault diagnosis, big data, traffic electrification

1. INTRODUCTION

Electric devices and systems have played significant roles in industry, academia, commerce, as well as other related fields. Most of these systems are expected to be functional for years. However, due to manufacture imperfection and harsh working conditions, hardware failures are frequently reported, threating property or even human life.

To avoid this problem, considerable amount of efforts have been paid, leading to a variety of fault diagnosis methods such as model-based method, quantitative method, and history-based method. Most of these methods tend to focus on certain features of the target systems or require fundamental understanding that is not available in most cases. Recent developments in big data and artificial intelligence provide new possibilities for fault diagnosis, where hidden relationships among usage patterns, environmental conditions and design flaws can be clarified. These methods have been pervasively applied and in this work we focus our analysis on electric vehicles (EVs) due to their increasing market share as well as complex design and application scenarios. Early research work has been reported in [1] where Zhao developed an algorithm for fault diagnosis on battery and vehicle levels. Sun [2] introduced information entropy to detect faulty battery cells. Later research has been extended to thermal runaway [3] and mechanical failures [4].

Unfortunately, most of these frameworks are established based on knowledge of algorithm design instead of fundamental understanding of issues in fault diagnosis. In this work, we propose a systematic method for fault diagnosis in EVs that covers data processing, feature extraction, model-based diagnosis,

Selection and peer-review under responsibility of the scientific committee of the 13_{th} Int. Conf. on Applied Energy (ICAE2021).

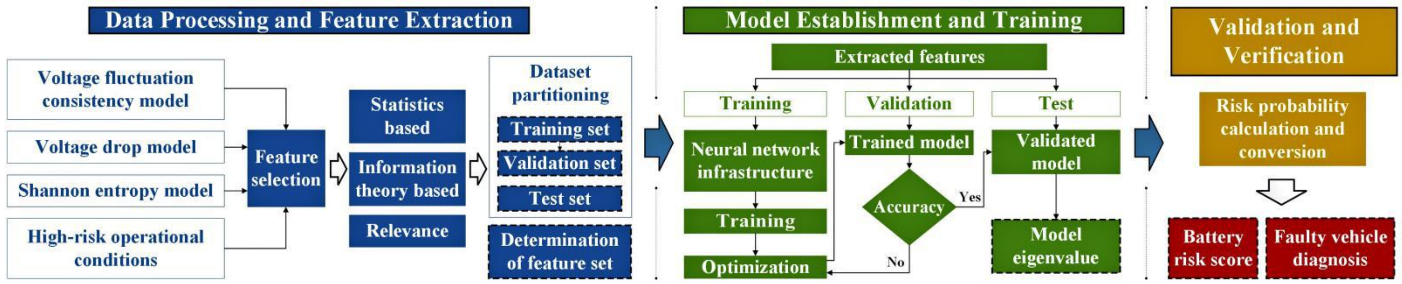

Fig. 1. Overall structure of the proposed algorithm

and model fusion. Feature extraction and model-based diagnosis are done through joint considerations of battery level and vehicle level statistical features, risky operational scenarios, and outputs from three different models that reveal potential hazards. On top of this, model fusion is performed through neural network to combine the pre-processed data such that the resulting output has less uncertainty. Self-correction is then introduced to optimize the algorithm towards specific

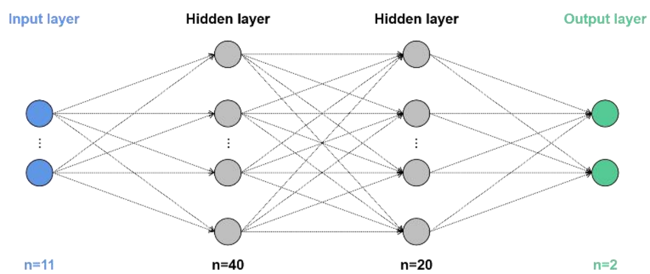

Fig. 2. Neural network infrastructure in the proposed algorithm

application scenario. Finally, validation and verification are done using data from 500 EVs, and a diagnosis accuracy over 95% can be expected.

The rest of the paper is organized as follows. Section 2 focuses on data pre-processing and feature extraction. The proposed model is established and trained accordingly in Section 3, followed by validation and verification results in Section 4. In the end, Section 5 concludes the paper.

2. DATA PROCESSING AND FEATURE EXTRACTION

2.1 Data source and pre-processing

Data used in this paper is collected through the National Big Data Platform of Electric Vehicles in China, which collects data from EVs in a real-time fashion. By July 2021, over 5 million EVs from 317 manufactures have been connected in [5]. In this work, for model training and validation purposes, data of more than one thousand vehicles are selected with a joint consideration of vehicle healthy state, usage history, location, data quality, as well as a few other factors. Pre-processing is performed to compensate mistakes from data collection and to extract useful information for later model training and algorithm design while filtering out other items such as VIN number and so forth.

2.2 Feature extraction and screening

Feature extraction is performed after data pre-processing to reveal key characteristics that may be potentially linked with faults in an EV. Some features can be obvious and easily derivable through simple calculations whereas others are outputs from more complicated models or related to extreme working scenarios. In this work, three models are adopted, including Shannon entropy model, volatility detection model, and voltage drop consistency model. Entropy model detects hidden issues from the perspective of information, while the other two models focus on voltage changes. Both model eigenvalues and secondary indicators, such as life cycle anomaly rate index and number of monomers, are included. Extreme working scenarios refer to abusive usage behaviors or environmental conditions that have been reported to trigger critical damage or speed up performance degradation in EVs. In this way, 82 features are selected for further screening.

From the perspective of algorithm design, adopting 82 features is not practical mainly due to complexity. To solve this problem, screening is performed based on how much information each feature carries and relevance among features. The amount of information from a certain feature can be quantitively described by calculating variance of it. According to information theory, larger variance means more information and vice versa [6]. Relevance, on the other hand, shows correlation among features and is calculated as Pearson correlation coefficient [7]. To ease analysis, we arbitrarily set boundaries of variance and at Pearson correlation coefficient at 0.1 and 0.8, respectively. Optimization of these parameters will be included as part of our future work.

Table I. Three types of selected features and definitions

Type of selected feature	Definition
Output from Shannon entropy model	For faulty vehicles detected by Shannon entropy model, record maximum, mean, and variance after processed through Z score.
Processed data based on outputs from upstream models	Average error rate of battery cells indicated by Shannon entropy model, volatility detection model, and voltage drop consistency model.
Risky operational scenarios	Fast charging in high SOC range, SOC≥90.0%, Charging rate≥0.5C.
Risky operational scenarios	Operation under high temperature and high SOC scenarios, SOC≥60.0%, environment temperature ≥30.0℃.
Risky operational scenarios	Frequent high-power charging and discharging, charging rate≥0.5C.

Data normalization is then introduced to handle the problem where dimensions and dimensional units are different among the features. Specifically, Z-score standardization is adopted, which processes the mean and standard deviation of the original data so that it can be fitted into a standard normal distribution [8]. The reason we choose normal distribution is to reduce the amount of calculation time needed for neural network in following steps to ensure the proposed algorithm feasible for real-time applications. After screening, three types of features are selected as inputs for model training as listed in Table I.

3. MODEL ESTABLISHMENT AND TRAINING

3.1 Neural network infrastructure

The core of the prediction fusion method is to carry out a second round of training on the outputs from each upstream model to obtain a better results of faulty vehicle diagnosis than using any of these models individually. Based on this idea, in this work, a neural network is established on top of Shannon entropy model, volatility detection model, and voltage drop consistency model. Since no analysis shall be performed in time domine, a neural network with linear layer combination would be an optimal trade-off between accuracy and complexity.

The overall architecture of the neural network built in this paper is shown in Fig. 1, which is composed of an input layer, two hidden layers and an output layer. The selected features from previous step are adopted as model inputs. In training process, input, output and hidden layers allow the neural network to forget or write new information into memory cells. Nonlinear transformation is carried out between the hidden layer and the output layer using Sigmoid function, given as

$$Sigmoid\ (x) = \frac{1}{1+e^{-x}}$$

where x is the result from previous layer. Sigmoid function is used as the activation function of neural network. Its main functionality is to map variables into a specific value between 0 and 1 and to output the weight values of fault and normal vehicles. At the end of each forward pass, prediction result of the neural network is compared with a label calculated based on real-world EVs, and the cross entropy loss function is introduced to optimize the model, expressed as

$$H(p,q) = \sum_i p(i) \cdot \log(\frac{1}{q(i)})$$

where p is the real distribution, q denotes the fitted distribution, and i the possible value. After each round of training, the model is upgraded in order to minimize output from the cross entropy function so that fitting result q can be as close to the real result p as possible.

3.2 Output as risk score

Although the Sigmoid function can output weight values from 0 to 1, due to the nature of function itself, even for a binary classification problem, summation of weight values could hardly be 1, which fails to meet the requirements from the perspective of probability. Therefore, probability conversion is performed to obtain the probability of each vehicle to be faulty, which is then converted into a score for better presentation. In this work, Softmax function is introduced to perform probability conversion, expressed as

$$S_i = \frac{e^k}{\sum_j e^j}$$

where k represents the value of the ith element. In this model, k is the weight value of 0 and 1, and $\sum_j e^j$ represents the sum of ownership weight values. In other words, it is the ratio of the index of this element to the sum of all element indexes. In this way, a risk score can be calculated, which is roughly inverse proportional to output of the Sigmoid function. As part of model establishment, a score of 67 is selected in this work as the global optimization result for validation and verification.

4. RESULTS AND DISCUSSION

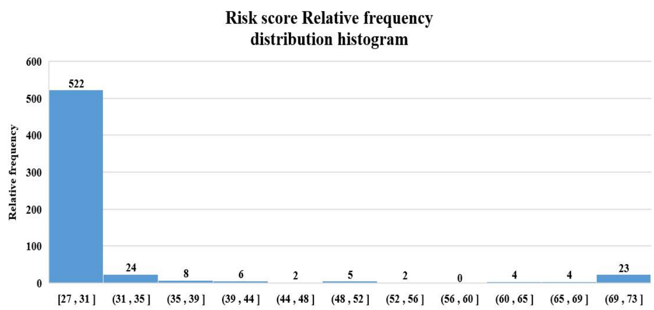

Fig. 3. Risk score from the proposed algorithm

To verify accuracy of the proposed algorithm, usage history data from 600 EVs are used. Vehicles are chosen randomly from several major manufactures, among which 21 are faulty whereas the rest 579 are functioning properly. Output from the proposed algorithm is illustrated in Fig. 3 and the result for faulty vehicle identification is given in Fig. 4. As can be seen, there is a clear boundary between faulty vehicles and

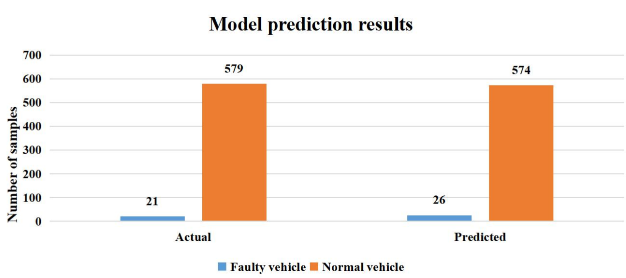

Fig. 4. Faulty vehicle identification result

normal ones as score of fault ones tend to allocate between 69-73 whereas most of normal vehicles get a score in the range of 27 to 31. The huge margin between these two ranges actually ensure robustness of the algorithm and makes it possible to monitor deterioration of a vehicle as its score increase overtime.

According to the score, 20 out of 21 faulty vehicles are successfully identified, indicating the ability of the proposed algorithm to distinguish normal and faulty vehicles. Due to lack of accident report, it is unable to come to a conclusion whether the proposed algorithm has actually misjudged one faulty vehicle. It is possible that the vehicle become a faulty one because random events such as traffic accident, which is beyond the scope of this work. Meanwhile, 6 vehicles are identified as faulty but still functioning. Unfortunately, we cannot come to a conclusion whether this is caused by imperfect algorithm design, or these vehicles are actually on high risk level. Continuous tracking and analysis are required for an uncontroversial result. Overall, the proposed algorithm has successfully identify most faulty vehicles with an accuracy rate over 95%.

5. CONCLUSION

In this work, a systematic method for fault diagnosis has been put forward, which jointly considered 82 possible factors that may lead to critical damage in an EV. Shannon entropy model, volatility detection model, and voltage drop consistency model are included to derive hidden information from measured I-V curve while EVs are in use. On top of these, a neural network is established based on the idea of model fusion to obtain a better diagnosis result than using any of the three models individually. Validation is performed using data from 600 real-world EVs, and an accuracy over 95% is achievable. The design principle of the proposed algorithm can be extended to fault diagnosis in complex systems other than EVs.

REFERENCE

[1] Yang Zhao, Zhenpo Wang, Lei Zhang, Fault and defect diagnosis of battery for electric vehicles based on big data analysis methods. Applied Energy. 10.1016/j.apenergy.2017.05.139.

[2] Zhenyu Sun, Peng Liu, Zhenpo Wang, Real-time Fault Diagnosis Method of Battery System Based on Shannon Entropy, Energy Procedia, Volume 105, 2017, Pages 2354-2359, ISSN 1876-6102.

[3] Ajay Raghavan, Peter Kiesel, Lars Wilko Sommer, et al. Embedded fiber-optic Sensing for Accurate Internal Monitoring of Cell State in Advanced Battery Management Systems Part 1: Cell Embedding Method and Performance [J]. Journal of Power Sources, 2017, 341: 466-473.

[4] Xiangdong Kong, Yuejiu Zheng, Minggao Ouyang, Languang Lu, Jianqiu Li, Zhendong Zhang, Fault diagnosis and quantitative analysis of micro-short circuits for lithium-ion batteries in battery packs, Journal of Power Sources, Volume 395, 2018, Pages 358-368, ISSN 0378-7753.

[5] Zhenpo Wang, Jichao Hong, Peng Liu, Lei Zhang, Voltage fault diagnosis and prognosis of battery systems based on entropy and Z-score for electric vehicles, Applied Energy, Volume 196, 2017, Pages 289-302, ISSN 0306-2619.

[6] Jingwen Song, Pengfei Wei, Marcos A. Valdebenito, Matthias Faes, Michael Beer, Data-driven and active learning of variance-based sensitivity indices with Bayesian probabilistic integration, Mechanical Systems

and Signal Processing, Volume 163, 2022, 108106, ISSN 0888-3270.
[7] Imane Jebli, Fatima-Zahra Belouadha, Mohammed Issam Kabbaj, Amine Tilioua, Prediction of solar energy guided by pearson correlation using machine learning, Energy, Volume 224, 2021, 120109, ISSN0360-5442.
[8] Qiao Xue, Guang Li, Yuanjian Zhang, Shiquan Shen, Zheng Chen, Yonggang Liu, Fault diagnosis and abnormality detection of lithium-ion battery packs based on statistical distribution, Journal of Power Sources, Volume 482, 2021, 228964, ISSN 0378-7753.