EVS32 Symposium

Lyon, France, May 19 - 22, 2019

Fault Diagnosis and Prognostics of EV Power Battery Based on Data Mining

Luping Wang [1,2)] Peng Jin[1,2,3)] Jiaxing He [1,2)] Jianyu Liu [1,2)] Zhuo Zhang [1,2)]

1) North China University of Technology,

5 Jinyuanzhuang Road, Shijingshan District, Beijing, China

2) Collaborative Innovation Center of Electric Vehicles in Beijing, Beijing, China

3) Beijing Institute of Technology

5 South Zhongguancun Street, Haidian District, Beijing, China(E-mail: jpzy216@163.com)

Executive Summary

Based on data mining to diagnose and predict battery faults, the influence of internal reactions of battery can be ignored, and the impending faults of batteries can be warned in advance. In this paper, PCA dimensionality reduction and K-means cluster are used to process the cell voltage, find the faulty cells, and use the basic principle of least squares support vector machine algorithm (LS-SVM) to extract the voltage data of the battery as a learning sample, and then predict battery state by building predictive models according to training data. Therefore, preventing the battery pack from being decayed due to faulty cells, the safety performance of the vehicle is improved, and the safety of the user's life and property is ensured.

1 Introduction

In recent years, EV accidents have occurred frequently. In the context of big data, how to make full use of data to analyze battery operating conditions and make fault predictions is an urgent problem to be solved. In this paper, the faulty battery data is mined from the historical big data of electric vehicle running time. The cell voltage is processed by PCA dimensionality reduction and k-means cluster analysis to find out the faulty battery cells, and the cause of the fault is analyzed. By the basic principle of the support vector machine algorithm (LS-SVM), the voltage of the EV battery pack is extracted as a learning sample, and a predictive model is established through training to predict the state trend and fault prediction, thereby improving the safety of the EV.

2 Fault Diagnosis and Prognostics Methods

2.1 Fault diagnosis methods

The fault data source of this paper is the cells voltage value at fault time of an electric vehicle, and the sampling interval is 10s, and the data sample set is obtained. Then use PCA dimension reduction to project high-dimensional data into lower-dimensional space, and then use K-means algorithm to cluster data.

2.1.1 PCA

Principal Component Analysis (PCA)[1] is a statistical analysis method that grasps the main contradictions of things. It can analyze the main influencing factors from multiple things, simplify complex problems. Dimension reduction steps from m dimension sample to n dimension:

1) Standardization of raw data;
2) Calculating the covariance matrix, and find the eigenvalues;
3) Extracting the eigenvector W corresponding to the n-dimensional eigenvalue;
4) Converting each sample in the sample set into a new sample;
5) Get the output sample set.

2.1.2 K-means clustering

The K-means clustering algorithm[2] is a representative of a typical prototype-based objective function clustering method, which is to find the optimal classification of a certain initial cluster center vector, so that the evaluation index is the smallest. K-means can directly cluster data without marking it in advance, Its basic idea is that for a sample set, K cluster centers are randomly given initially, and the samples to be classified are divided into clusters according to the nearest neighbor principle. Then the centroid of each cluster is recalculated according to the average method, so that the new cluster center is determined and iterated until the sum of square error (E) of the function is the smallest. Set the object collection to D={x_1, x_2, …, x_n}, xt={x_{i1}, x_{i2}, …, x_{it}}, the Euclidean distance formula for sample xi to sample xj is:

$$d(x_i,x_j)=[(x_{i1}-x_{j1})^2+(x_{i2}-x_{j2})^2+.....+(x_{in}-x_{jn})^2] \tag{1}$$

The sum of squared errors is:

$$E=\sum_{i=1}^{k}\sum_{x\in C_1} dist(c_i,x)^2 \tag{2}$$

Where k represents k clusters; c_i is the number of center point; *dist* is the Euclidean distance.

2.2 State prognostics method

In this paper, the voltage of the battery is taken as the basis of fault warning, and the least squares support vector machine (LS-SVM) is used to train the data and establish the prediction model. The LS-SVM proposed by Suykens et al.[3] is an extension of the standard support vector machine(SVM), which simplifies the calculation, and has the advantages of fast convergence speed and high precision.

The regression problem of LS-SVM can be described as follows:

$$\mathrm{P}\quad : \min_{\omega,\mathrm{b},\mathrm{e}} J_p(\omega,\mathrm{e})=\frac{1}{2}\omega^T\omega+\frac{1}{2}\gamma\sum_{\mathrm{i}=1}^{N}\mathrm{e}_{\mathrm{i}}^{2} \tag{3}$$

$$such\ that \quad y_{\mathrm{i}}=\omega^T\varphi(x_{\mathrm{i}})+b+e_{\mathrm{i}},\quad \mathrm{i}=1,...,N \tag{4}$$

Where the weight vector $\omega\in R^{\mathrm{n}}$. While $\mathrm{b}\in R$ is the threshold, e is a slack variable, its significance is to introduce outliers in the support vector, and e is included in the final optimization goal. $\gamma>0$ is the weight of the outliers.

Solve the LS-SVM regression estimation function:

$$y(x) = \sum_{i=1}^{N} a_i y_i K(x, x_i) + b \tag{5}$$

Where $K(x, x_i) = \varphi(x)\varphi(x_i)$ is a kernel function. This paper uses a Gaussian radial basis kernel function (RBF-Kernel), the formula is as follows:

$$K(x, x_i) = \exp\left(-\frac{\| x - x_i \|^2}{2\sigma^2}\right) \tag{6}$$

3 Results and Discussion

3.1 Fault diagnosis of cells

3.1.1 Battery pack fault diagnosis

Fig.1 is a figure of 99 cells voltage data, which includes overvoltage fault data. Fig.3 is a figure of 63 cells voltage data, which includes undervoltage fault data. The two-dimensional figure shown in Fig. 2 is obtained by the PCA dimensionality reduction and K-means clustering methods for 101 individual voltage data points obtained during power battery overvoltage alarm, which shows the normal voltage data will gather together, and the faulty cells have been separated further, and the 28, 36 and 45 cells are overvoltage faulty cells. Fig. 4 shows the faulty cells have been separated, and the 26 and 47 cells are undervoltage faulty cells.

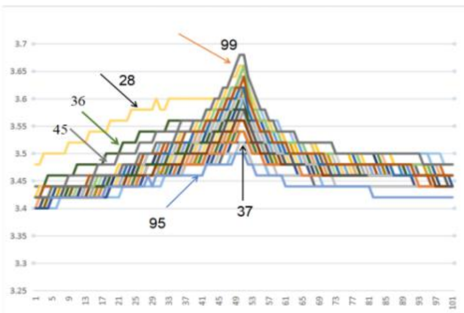


Figure 1: 99 voltage cells

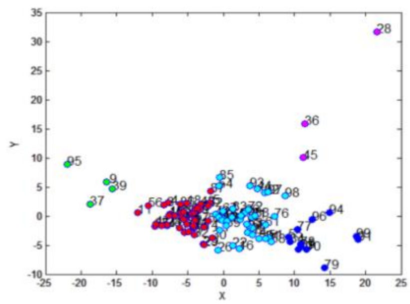


Figure 2: Overvoltage data processed

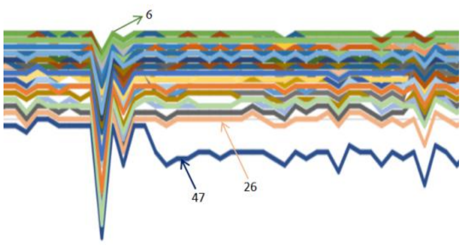


Figure 3: 63 voltage cells

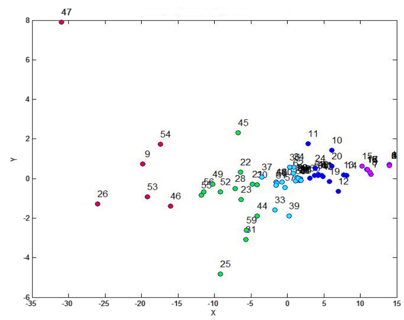


Figure 4: Undervoltage data processed

3.1.2 Cells fault diagnosis

Fig.5 is a two-dimensional result of PCA dimensionality reduction and K-means clustering of the individual voltage datas obtained during the power battery undervoltage alarm. It can be easily seen from that the 96 and 97 are undervoltage cells. Due to the low voltage of the the 96 and 97 is significantly different from other voltage cells, a fault alarm is caused, and the specific cause of the failure can only be determined by disassembling the battery pack.

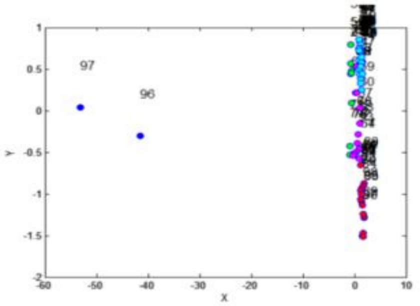


Figure 5: Cells voltage under undervoltage

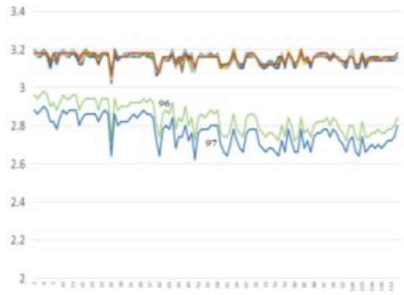


Figure 6: Undervoltage cells are 96 and 97

3.1.3 Cells difference fault diagnosis

Due to the variability in the production of power batteries, the single cells produced in the same batch still have performance differences (capacity、 internal resistance and so on), during the long-term use process, various performance differences will be revealed, and the battery difference is reflected in the external cell voltage.

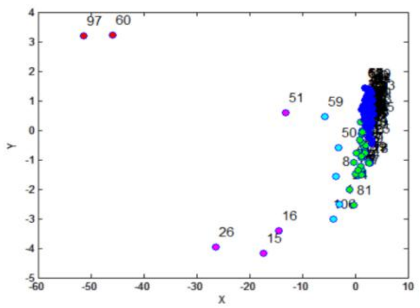


Figure 7: Differential fault cells

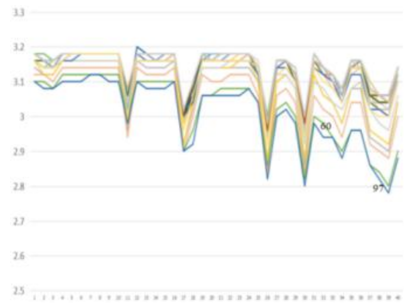


Figure 8: Differential fault cells are 60 and 97

Fig.7 shows the two-dimensional effect of the individual voltage datas acquired during the differential alarm of power battery after PCA dimensionality reduction and K-means clustering. It can be seen from Fig.7 and Fig.8 that the 60 and 97 are differential fault cells. At the same time, it is necessary to focus on the single cells of No. 15, No. 16, No. 26 and No. 51 in the later stage.

3.1.4 Battery voltage fault diagnosis within one week

In Fig.9 are the clustering figures obtained after the datas of the battery cells in the battery pack for one week of the failure of the power battery by K-means. Among them, the data of normal battery cells are generally gathered at the bottom of the figure, and the faulty data is substantially free on the rightmost side and above, and if it is free outside for several days, the faulty battery can be determined. For example, the 53 and 94 are free outside for six days, which can be judged as faulty cells.

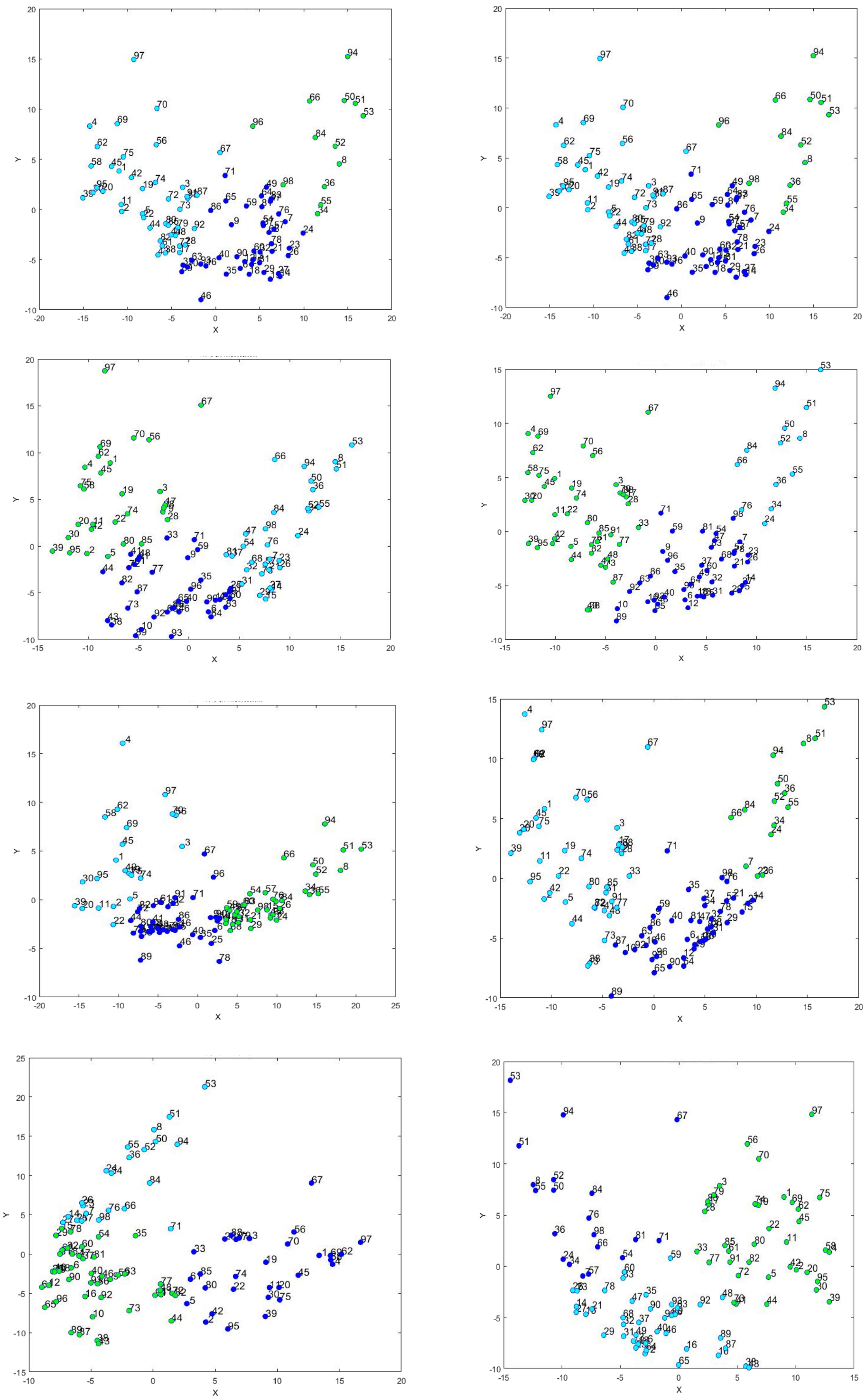


Figure 9: Cells voltage data processing figure within one week

3.2 Battery pack voltage prognostics

3.2.1 Method validation

In the process of battery packet voltage prediction, the optimal support vector machine training parameters are used first, and then the sample set is used to train the support vector machine. Finally, the predicted value is obtained by using the trained model. The actual forecasting method steps are as follows: 1) Select the overvoltage fault datas to form the voltage sequence and transform it into training samples; 2) Normalize the training samples; 3) Optimize the parameters by grid method; 4) Verify the rationality of parameter selection by K-fold cross-validation method; 5) Training the regression model with pre-processed training samples; 6) Predicting the voltage sequence using the regression model.

In this paper, the voltage of the battery is taken as the basis of fault warning, and training data by predicting the normal voltage and fault voltage of the battery by the LS-SVM, and then the prediction model is established. Fig.10 and Fig.11 are normal voltage prediction and normal voltage prediction error figures, respectively[4].

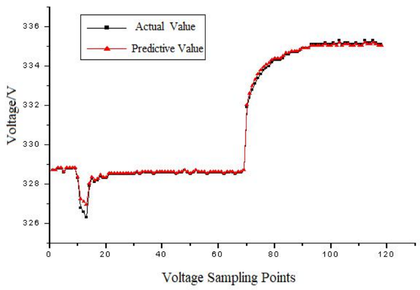


Figure 10: LS-SVM normal voltage prediction

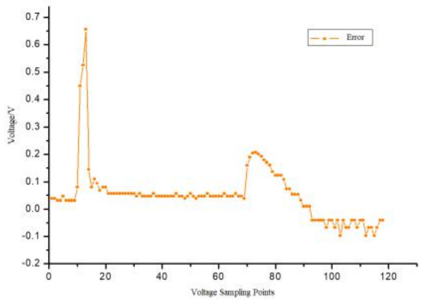


Figure 11: LS-SVM normal voltage prediction error

Fig.12 and Fig.13 are the LS-SVM fault voltage prediction and LS-SVM fault voltage prediction error figures, respectively. The voltage fault data has a distinct spike. When the total voltage exceeds the critical value (the 50th sampling point), the battery management system will alarm.

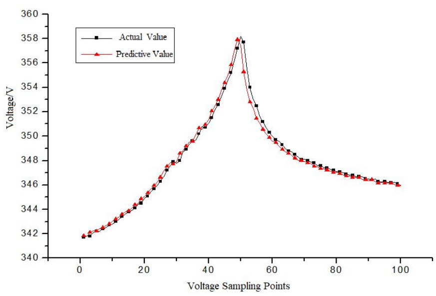


Figure 12: LS-SVM fault predicted and actual values

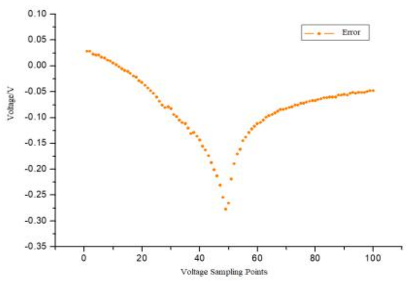


Figure 13: LS-SVM fault prediction error

The mean squared error (MSE) and the mean absolute percentage error (MAPE) are introduced as the indicators for evaluating the approximation ability and generalization error performance of the LS-SVM. The expressions for MSE and MAPE are as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{7}$$

$$MAPE = \sum_{t=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times \frac{100}{n} \tag{8}$$

Table 1 : LS-SVM normal and fault voltage operation results

	Normal Voltage Prediction	Fault Voltage Prediction
MSE	0.0136	0.0236
MAPE	0.011	0.0244

3.2.2 Battery pack voltage prognostics

It is verified that the prognostics method is feasible and then predicts the fault voltage data. D consecutive time series values are taken from the original training sample set, the first d-1 are used as the training set, and the second to the dth are used as the test set. In the past, d-5, d-10, and d-15 data were used as training data. Predicting the data at the last 5, 10, and 15 time points by LV-SVM, a comparison chart of actual and predicted values is obtained.

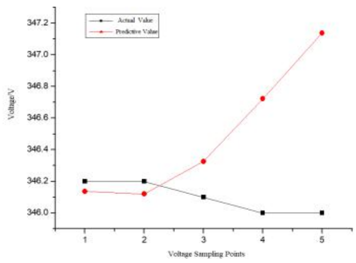


Figure 14: Real and predicted values for the last 5 time points

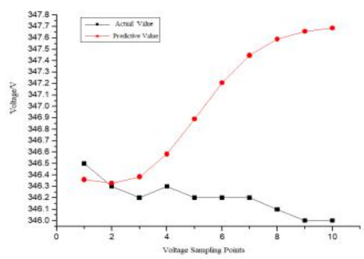


Figure 15: Real and predicted values for the last 10 time points

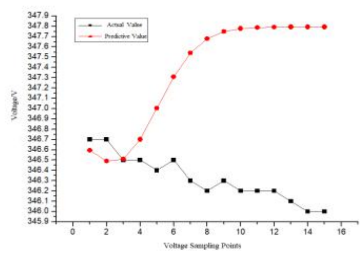


Figure 16: Real and predicted values for the last 15 time points

As shown in Fig.14, Fig.15 and Fig.16, the error in predicting to 1-4 time points is relatively small, usually around 0.2V, which can be neglected relative to the total voltage 300V .The value of the fourth time point of the predicted fault alarm is 357.58V, and the error is 0.106%. In practice, when the total voltage reaches its peak value (357.2V), a fault alarm will occur. The fault alarm can be predicted at the fourth time point (40s). The predicted value is basically consistent with the actual alarm value, which can be used as overvoltage fault prediction.

4 Conclusion

In this paper, a battery fault diagnosis and prognostics model based on data mining is constructed. Firstly, the faulty batteries in battery pack are analyzed by PCA dimensionality reduction and K-means clustering methods. Then the normal voltage and fault voltage are predicted by LS-SVM, which proves that the prognostics method is feasible. Finally, the predicted value of the test data point is compared with the actual alarm value, which proves that the values are basically consistent.

Reference

[1] ZHAO Gui-ru, *A Method of Dimensionality Reduction for Large Scale Data Using PCA*[J].Computer Knowledge and Technology, 2014, (8): 1835-1837.

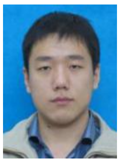
[2] Wang Qian, Wang Cheng, Feng Zhen-yuan, Ye Jin-feng, *Review of K-means clustering algorithm*[J]. Electronic Design Engineering, 2012, 20(7):21-24.

[3] Suykens J A K, Vandewalle J. *Least squares support vector machine classifiers*[J]. Neural Processing Letters, 1999, 9(3): 293−300.

[4] DAI Lin-chao, WU Lin-li, ZHAO Hai-na, LI Xun-ming. *Error prediction based on least squares support vector machines*[J]. Journal of Central South University: Science and Technology, 2009(S1).

Authors

Wang Luping (1994-), Female, Hebei Province, China, Post-graduate student, Main research direction: Battery Management System (BMS) and Lithium Battery Modeling. E-mail: 847725684@qq.com.



Jin Peng (1980-), Male, Main research direction: Battery Management System, Electric Drive System, CAN Bus Communication.