Health Evaluation Method of Supercapacitor Based on Data Mining

Ziwei Fang1, Shuli Liu1 and Yanming Zhao1,2*

1School of Information and Electrical Engineering, Hunan University of Science and Technology, Xiangtan, Hunan, 411201, China
2School of Engineering Research Center of Hunan Province for the Mining and Utilization of Wind Turbines Operation Data, Hunan University of Science and Technology, Xiangtan, Hunan, 411201, China
*Corresponding author’s e-mail: ymzhao@hnust.edu.cn

Abstract. With the wide application of super capacitor energy storage system as energy storage device on urban rail trains, the problem of its safe operation has attracted great attention. In order to ensure the reliability of the supercapacitor energy storage system, it is necessary to evaluate its health effectively. In view of the fact that the existing model-based supercapacitor evaluation methods are difficult to establish a reasonable and accurate model and the parameter estimation is complex, a data-driven supercapacitor health evaluation method is proposed. In this method, based on the data of voltage, current, temperature and external characteristic parameters of charge and discharge of supercapacitor, the weighted Euclidean distance formula is introduced, the mixed weighted Euclidean distance is calculated, and the comprehensive evaluation index of supercapacitor health state is obtained. It can quickly and effectively evaluate the health status of supercapacitors.

1. Introduction
With the rapid development of urban rail transit technology, supercapacitors have gradually replaced traditional batteries such as lead-acid batteries as the main power source of urban rail trains with the advantages of high energy density, long life and long cycle life[1-2]. In the process of application as an energy storage system, supercapacitors need a large number of series-parallel combinations to form supercapacitor modules, and supercapacitor cells are inconsistent because of technology, materials and other factors. It will lead to the phenomenon of overcharge and discharge of the supercapacitor module or monomer, which will lead to the degradation or even serious damage of the supercapacitor cell performance and can't work properly, and the degradation of the monomer performance will affect the energy storage characteristics of the whole energy storage system[3]. Therefore, in order to improve the safety and reliability of the whole energy storage system, the health status of the supercapacitor must be evaluated in order to maintain or replace the aging supercapacitor in time, which is of great significance for prolonging the service life of the supercapacitor module and ensuring the safe operation of urban rail trains.

At present, domestic and foreign scholars have carried out a lot of research on the evaluation of the health status of supercapacitors. The common method is model-based. A Eddahech[4] based on the RLC equivalent circuit model of supercapacitors, proposed a recurrence with time-varying forgetting factor Least-squares parameter identification method, using voltage and current data as input, online
estimation of dynamic supercapacitor equivalent circuit parameters, and using supercapacitor equivalent internal resistance as SOH index. Compared with the results of IEC standards and electrochemical impedance spectroscopy, this method had good performance. T Dimitri et al [5] based on the exponential degradation law, used least squares fitting to determine the model parameters, established a supercapacitor capacity degradation prediction model, and used the charge transferred by the supercapacitor and the duration of the temperature stress higher than rated as the model input, and modified the output capacitance value of the model to adapt to different stress conditions. H Chaoui et al [6-7] proposed an online life estimation method for supercapacitors, which detected the charge and discharge current and voltage of supercapacitors, and used Lyapunov's adaptive theory to identify the equivalent resistance and equivalent capacitance of the capacitor online, which guaranteed its convergence and stability. This type of method can estimate the supercapacitor parameters, but it depends on the accuracy of the equivalent model, and it is difficult to establish a unified supercapacitor equivalent circuit model for different application backgrounds. The method based on data mining has the advantages of application flexibility and model-free, which avoids the difficulty of modeling in the above-mentioned model-based method. Therefore, this paper adopts a data mining method and proposes a supercapacitor health state based on hybrid weighted Euclidean distance. Firstly, the evaluation method uses the change data of external characteristic parameters such as current, voltage, temperature and power during the charging and discharging process of the supercapacitor to construct the external characteristic attribute matrix of the supercapacitor, and uses the CRITIC method to calculate the weight of each characteristic attribute in the hybrid attribute matrix. Then it calculates the weighted Euclidean distance between the attribute matrices, and obtains the comprehensive evaluation index of the supercapacitor health state, which can quickly and effectively evaluate the current health status of the supercapacitor.

2. Hybrid weighted Euclidean distance calculation method

2.1. Weight calculation

CRITIC method is an objective weight weighting method proposed by Diakoulaki. Its basic idea is to obtain weights by using data. The greater the standard deviation, the greater the fluctuation and the higher the attribute weight; the data correlation is measured by the correlation coefficient. If there is a strong positive correlation between the two attributes, it means that the smaller the correlation is and the lower the attribute weight will be [8-9]. According to the CRITIC method, the objective weight of each attribute is calculated as follows.

2.1.1. Attribute data normalization. A dataset with m attributes, each containing n data, whose attribute matrix is expressed as:

\[
X = \begin{bmatrix}
P_1, P_2, \ldots, P_m \\
x_{11} & x_{12} & \cdots & x_{1m} \\
x_{21} & x_{22} & \cdots & x_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
x_{n1} & x_{n2} & \cdots & x_{nm}
\end{bmatrix}
\]

(1)

Where \( P_j \) is the \( j \) attribute, \( P_j = (x_{1j}, x_{2j}, \ldots, x_{nj})^T \); \( x_{ij} \) is the \( i \)-th data value of the \( j \)-th attribute, \( i = 1, 2, \ldots, n, j = 1, 2, \ldots, m \).

Due to the different dimension of different attributes, the numerical value varies greatly. If the calculation with dimension is used, the result will be greatly affected, so dimensionless normalization is required. Generally, attributes can be divided into benefit-type and cost-type. Benefit-type attributes refer to the index that the larger the attribute value, the better, and the cost-type attribute refers to the index that the smaller the attribute value the better. For different attribute types, the non-dimensional
normalization method is also different. For benefit-type attributes and cost-type attributes, they are normalized according to formula (2) and formula (3) respectively.

\[
x'_j = \left( x_j - x_{j_{\text{min}}} \right) / \left( x_{j_{\text{max}}} - x_{j_{\text{min}}} \right)
\]

(2)

\[
x'_j = \left( x_{j_{\text{max}}} - x_j \right) / \left( x_{j_{\text{max}}} - x_{j_{\text{min}}} \right)
\]

(3)

Due to the different dimension of different attributes, the numerical value varies greatly. If the calculation with dimension is used, the result will be greatly affected, so dimensionless normalization is required. Generally, attributes can be divided into benefit-type and cost-type. Benefit-type attributes refer to the index that the larger the attribute value, the better, and the cost-type attribute refers to the index that the smaller the attribute value the better. For different attribute types, the non-dimensional normalization method is also different. For benefit-type attributes and cost-type attributes, they are normalized according to formula (2) and formula (3) respectively.

2.1.2. Attribute standard deviation calculation. The volatility of the attribute is measured by the standard deviation, and the standard deviation of the attribute \( P_j \) is:

\[
S_j = \left( \sum_{i=1}^{n} \left( x'_j - \frac{1}{n} \sum_{i=1}^{n} x'_j \right)^2 \right)^{1/2}
\]

(4)

2.1.3. Calculation of attribute correlation coefficient. The correlation of attributes is measured by the correlation coefficient. The correlation coefficient between attributes \( P_j \) and \( P_k \) is:

\[
R_{jk} = \sum_{i=1}^{n} \left( 1 - \frac{Cov(P_i, P_j)}{\sqrt{Var(P_i)Var(P_j)}} \right)^{1/2}
\]

(5)

2.1.4. Attribute data information calculation. Set \( C_j \) represent the amount of information contained in the \( j \)-th evaluation index, and \( C_j \) can be expressed as:

\[
C_j = S_j \times R_j
\]

(6)

The larger \( C_j \) is, the greater the role of attribute \( P_j \) in the whole evaluation system is, and more weight should be assigned to \( P_j \).

2.1.5. Attribute objective weight calculation. The objective weight of attribute \( P_j \) is:

\[
\omega_j = C_j \left( \sum_{j=1}^{m} C_j \right)^{-1}
\]

(7)

From this, we can get that the weight vector is \( \omega = [\omega_1, \omega_2, \ldots, \omega_m] \), and \( \sum_{j=1}^{m} \omega_j = 1 \).

2.2. Mixed weighted Euclidean distance calculation

The contribution degree of different attributes of the same evaluation object is different, and the same attribute of different evaluation objects also different. The weighted Euclidean distance is used to overcome the contribution of different attributes to the result, and the mixed attribute matrix of the
evaluated object is adopted. To calculate the weight, it effectively reflects the difference of the same attribute of different evaluation objects, thereby improving the accuracy of the evaluation results. For evaluation objects \( A \) and \( B \), their attribute matrices are respectively \( A, B \subset X \), the mixed attribute matrix is \( Y = \begin{bmatrix} A \\ B \end{bmatrix} \), the normalized attribute matrix of \( A \) and \( B \) is \( A = (a_{i})_{nm}, B = (b_{i})_{nm} \), the mixed weighted Euclidean distance between attribute matrix \( A \) and \( B \) is expressed as \[ D(A,B,\omega) = \left( \sum_{j=1}^{m} \omega_{j} \left( \sum_{i=1}^{n} (a_{ij} - b_{ij})^{2} \right) \right)^{1/2} \] (8)

In the formula, the \( \omega = [\omega_{1}, \omega_{2}, \ldots, \omega_{n}] \) is calculated by the mixed attribute matrix \( Y \) according to formula (2)-formula (7).

3. Health evaluation method of supercapacitor based on mixed weighted Euclidean distance

The failure of supercapacitor is generally caused by the aging of supercapacitor module itself, which changes the electrode, electrolyte and other supercapacitor module components in terms of physical and chemical properties, but the change of its properties is difficult to measure. In order to describe the failure of supercapacitor, the change of external characteristic parameters such as voltage, current, temperature and residual electricity can be used as characteristic parameters. Because the external parameters of different capacitance aging are different, the importance of each parameter for aging evaluation is not constant for different capacitor monomers. Therefore, this paper introduces the mixed weighted Euclidean distance to build the failure diagnosis index system, the basic idea is to use the factory capacitance as the calibration capacitance, to construct the dynamic attribute matrix from the sample to be tested to the calibration capacitance, to calculate the weight of the characteristic attribute in the dynamic attribute matrix by CRITIC, as the attribute weight in the weighted Euclidean distance, and to measure it by the mixed weighted Euclidean distance between the sample to be tested and the calibrated capacitor. Finally, the evaluation result of the capacitance to be measured is determined, based on the mixed weighted Euclidean distance health evaluation process, as shown in figure 1.

![Flow chart of health status evaluation based on hybrid weighted Euclidean distance](image)

Figure 1. Flow chart of health status evaluation based on hybrid weighted Euclidean distance.
3.1. Evaluation threshold determination
A number of fault capacitance samples and a number of capacitors with a capacity loss of less than 20% are obtained as training data, and a new capacitance data is used as calibration data. The formula (2)-(8) is used to calculate the weighted Euclidean distance from each fault capacitor sample to the new capacitor, and its average value is used as the upper threshold of evaluation $D_{OL}$. The weighted Euclidean distance from each capacitor with capacity loss less than 20% to the new capacitor is calculated, and the average value is used as the lower threshold of evaluation $D_N$.

3.2. Evaluation index quantification
Use the hybrid weighted Euclidean distance to define the comprehensive evaluation index of the super capacitor health status as follows:

$$PHM = \left(1 - \frac{D_N}{D_{OL}}\right)^{-1} \times 100\%$$ (9)

Where, $D_S$ is the weighted Euclidean distance between the evaluated capacitor and the factory capacitor calculated by formula (2)-(8). Table 1 shows the Health status of the super capacitor. When the value of $PHM$ is 80%-100%, the health status of supercapacitor is excellent. When the value of $PHM$ is 60%-80%, the health status of supercapacitor is good. When the value of $PHM$ is 40%-60%, the health status of supercapacitor is general. When the value of $PHM$ is 20%-60%, the health status of supercapacitor is poor, and its capacity has been greatly reduced, which needs to be paid more attention. When the value of $PHM$ is below 20%, the capacitor needs to be replaced in time.

Table 1. Health status pros and cons classification table of Supercapacitor.

| Serial number | PHM (%) | State of health |
|---------------|---------|----------------|
| A             | (80,100] | Excellent      |
| B             | (60,80]  | Good           |
| C             | (40,60]  | General        |
| D             | (20,40]  | Poor           |
| F             | (0,20]   | Failure        |

4. Experiment analysis

4.1. Experimental platform construction and data acquisition
In this paper, eight supercapacitor cells with different aging stages are selected for study. The monomer chooses the rated voltage of BCAP0350E270T11350F, produced by Maxwell. In order to measure the experimental data needed for health evaluation, build an experimental test platform, as shown in figure 2. The EBC-A10H tester is used to test the charge and discharge of the supercapacitor cell, and the charge and discharge current, voltage and charge and discharge quantity are collected. At the same time, the surface temperature of the supercapacitor cell is collected by using DS18B20 temperature sensor and temperature acquisition board.

Figure 2. Experimental of charge-discharge test platform of supercapacitor.
4.2. Examples of health evaluation

The newly manufactured supercapacitors with excellent performance are used as calibration capacitors, and several aging supercapacitors are used as training capacitors. $C_1$-$C_8$ is a capacitor to be evaluated. Through the health evaluation step of supercapacitor backup power supply based on mixed weighted Euclidean distance proposed in this paper, the specific results are shown in Table 2. From Table 2, it can be seen that the health status of supercapacitors from bad to good is $C_4$<$C_8$<$C_6$<$C_5$<$C_3$<$C_2$<$C_7$<$C_1$. $C_4$ and $C_8$ aging are the most serious and invalidated, $C_5$ and $C_6$ aging are serious and poor, $C_3$ is in general health, $C_2$ and $C_7$ are in good health, and $C_4$ is excellent.

| Evaluation grade | $PHM$ (%) | State of health |
|------------------|-----------|----------------|
| 1                | 100       | Excellent      |
| 2                | 61.8      | Good           |
| 3                | 44.6      | General        |
| 4                | 0         | Failure        |
| 5                | 34.5      | Poor           |
| 6                | 32.1      | Poor           |
| 7                | 63.6      | Good           |
| 8                | 8.1       | Failure        |

5. Conclusion

In view of the shortcomings of the existing supercapacitor health evaluation methods based on the supercapacitor equivalent circuit model, this paper uses the data clustering mining method based on the supercapacitor charge and discharge voltage, current, temperature, electricity and other external characteristic parameter data. A health evaluation method of supercapacitor based on mixed weighted Euclidean distance is proposed to effectively avoid establishing a reasonable and accurate equivalent circuit model based on the principle of complex supercapacitor.

Acknowledgments

2020 Hunan Postgraduate Research Innovation Project (CX20200998)

References

[1] XIN Y, YANG Z, LIN F, et al. (2019) Estimation of the health status of urban rail transit supercapacitors based on parameter feedback. Journal of Electrical Engineering, 34: 396-404.
[2] MA L, LIAO W. (2015) Research on the control strategy of supercapacitor energy storage on urban rail trains. Transactions of the Chinese Society of Electrical Engineering, 30: 63-68.
[3] ZHOU Y, WANG Y, WANG K, et al. (2020) Hybrid genetic algorithm method for efficient and robust evaluation of remaining useful life of supercapacitors. Applied Energy, 4: 114169.1-15.
[4] EDDAHCH A, AYADI M, BRIAT O, et al. (2013) Online Parameter Identification for Real-Time Supercapacitor Performance Estimation in Automotive Applications. International Journal of Electrical Power & Energy Systems, 10: 162-167.
[5] TORREGROSOSSA D, PAOLONE M. (2016) Modelling of current and temperature effects on supercapacitors ageing. Part II: State-of-Health assessment. Journal of Energy Storage, 4955: 95-101.
[6] CHAOUI H, GUALOUS H. (2017) Online Lifetime Estimation of Supercapacitors[J]. IEEE Transactions on Power Electronics, 9:7199-7206.
[7] OUKAOUR, AMRANE, GUALOUS, et al. (2016) Online System Identification for Lifetime Diagnostic of Supercapacitors with Guaranteed Stability. IEEE transactions on control systems technology: A publication of the IEEE Control Systems Society, 6:2094-2102.
[8] YANG G, CUI Y, XUAN Y, et al. (2018) Operation quality evaluation of measurement equipment based on CRITIC and ideal point method. Power system protection and control, 24: 62-69.
[9] ZHANG L, ZHANG X. (2015) Weighted clustering method based on improved CRITIC method[J]. Statistics and Decision, 18: 65-68.
[10] ZHU L. (2016) An Improvement of Weighted Euclidean Distance Clustering Algorithm[J]. Computer and Digital Engineering, 3: 421-424.