Battery Grouping with Time Series Clustering based on Features

Jiayun Yang, Guojin Ma, Mingyu Gao, Zhiwei He*
Department of Electronics and Information
Hangzhou Dianzi University
Hangzhou, China
zwhe@hdu.edu.cn

***Abstract*----Battery grouping is important to the performance of the whole battery pack. In this paper, we use time series clustering algorithm for battery grouping. The proposed method uses the battery discharge curve to complete the grouping. First, we extract several kinds of characteristics of the discharge sequence. The similarities between batteries are then computed according to these features. Finally, a fast search and find of the density peak is utilized to cluster these features to accomplish the battery grouping. Experimental results show that the proposed battery grouping method is effective.**

Keywords—-battery grouping, time series, fast search and find the density peak

I. Introduction

With increasing current awareness of environmental issues there has been considerable interest in low-carbon travel, including the development of new electric vehicles [3]. The performance of the batteries, who are the power source of electric vehicles, will affect the performance of the whole car. But the capacity of a single battery cell in a car cannot provide the needed power for an electric vehicle, so the power battery needs to be in the form of a battery pack to ensure that the electric vehicle has enough power [7].

However, if batteries with inconsistent performance are combined into a group, some batteries in the battery pack can be easily over-charged or over-discharged, thus this will affect the performance of the entire battery pack and shorten its life. So it is very important to select a suitable grouping method. Now there are three main battery grouping methods: single parameter based grouping, multi parameters based grouping and dynamic characteristics based grouping.

In general, terminal voltage, internal resistance, static capacity, self-discharge rate and other parameters of the battery can be used as the basis of battery matching. Single parameter based grouping describes the battery from one aspect, that is to say, it select one of the battery parameters as the grouping basis, for example, two batteries with similar static capacity are grouped together. But the grouping effect of this method is usually poor [10]. The reason is that single feature cannot represent the whole performance of the battery. So there is a certain error. Multi-parameters based grouping select more parameters as the basis of grouping, which is better than the single parameter based grouping, but it does not take into account the changes in the parameters of the batteries at work, so there are some defects. Dynamic characteristics based grouping method is based on battery charging and discharging curves and takes full account of the internal characteristics of each battery, so the grouping effect is better. So in this paper, we choose dynamic characteristic based grouping method.

We select the discharge curve of the battery to cluster, classify the curve with higher similarity into a class. So as to achieve the purpose of battery grouping. The battery discharge curve is an ordered multidimensional data set, so the battery grouping based on the discharge sequence can actually be viewed as a time series clustering problem. And the clustering algorithm based on time series can be divided into two aspects: 1, time series clustering based on raw data, we can regard the time series as static data, and use the Dynamic Time Warping (DTW) [6] distance, the Euclidean distance, the correlation coefficient etc., to measure the similarity between data. .then use the clustering algorithm to complete grouping .2, time series clustering based on the feature [8-9]. By reducing time series dimensionality, that is to extract the feature of time series as the basis of similarity measure, and then use clustering algorithm complete grouping.

People may also consider clustering on raw data and use DTW to measure the similarity between time series [4]. Unfortunately, the time series of battery discharge has the characteristics of unequal length and high dimension. If we use the time series clustering based on the raw data, the computational complexity will be large. And we can only find the similarity of the sequence surfaces, not to touch its intrinsic mechanism, Therefore, in this paper, we use the time series clustering based on feature, first, we extract feature of the battery discharge sequence, The extracted feature represent the sequence, and use Euclidean distance measure the similarity between the two batteries, then use Fast searching and finding density peak to complete grouping [1].

II. Time Series Clustering Based On Features

A. Feature Extraction Of Time Series

Usually the time series has a number of features,, each of which characterizes one aspect of time series. In this paper, we describes the time series from three aspects [2]: time series statistical distribution, Fourier spectrum conversion and the length of time series. Statistical characteristics must be considered in the analysis of many time series, it descripts global structure of time series, including the mean, variance, skewness and kurtosis etc. Fourier spectral transform converts time series from time domain to frequency domain, and extracts

978-1-5090-1412-5/17/$31.00 ©2017 IEEE

transformed Fourier coefficients as frequency domain features of time series. Because of the different battery performance is not the same, the discharge time is also different, and so the time series collected is not equal, the total time of each time series long as one characteristic of clustering.

1) Statistical Characteristics

In this paper, we uses kurtosis, skewness, mean and variance of the time series as the statistical characteristics of sequence. Mean and variance are used to describe the center and deviation of data. Kurtosis is a description of shape steepness degree of the overall distribution of all the values. The sequence is compared with the normal distribution, the kurtosis of 0 means that the shape steepness degree of data distribution is same with the normal distribution. Kurtosis is greater than 0, indicating that the data distribution is more steep than the normal distribution, less than 0 data distribution than the normal distribution is relatively flat, The calculation formula is:

$$\text{Kurtosis} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^4/SD^4 - 3 \quad (1)$$

Skewness is the description of symmetry of the data distribution. Skewness greater than 0 indicates that the data is a right deviation, otherwise is a left deviation. Its formula is:

$$\text{Skewness} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^3/SD^3 \quad (2)$$

The SD is the standard deviation

2) Fourier Transform Coefficients

The Fourier transform transforms difficult-to-handle time-domain signals into easily-analyzed frequency-domain signals. According to Parseval's theory, the time-domain energy function is the same as the frequency-domain energy spectrum function. The high-frequency part of a signal is not important, so most of the energy in the frequency domain is concentrated on the first few coefficients, Therefore, in this paper, we select the first two coefficients of Fourier transform as our global characteristics.

$$Y_f = \frac{1}{\sqrt{n}}\sum_{i=1}^{T} Y_t \exp\left(-i\frac{2\pi}{n}ft\right), \quad f = 1,2,\ldots,T \quad (3)$$

Y_f (f = 1, 2,⋯ T) is the Fourier coefficient and Y_t is the time domain information

3) Total Length Of Time

In this paper, the battery whose initial voltage is 4.2 V discharge at constant current, when the voltage drops to 2.7V, the battery stops discharging. The performance of each battery is not consistent, the discharge time will be different, and so the dimension of the time series is different. So we will use the total length of the time series as one of the characteristics.

B. Data Standardization

As described above, we extract the seven features of the time series. As shown below, Figure 1 shows the two battery discharge curves, as well as the two curves extracted eigenvalues.

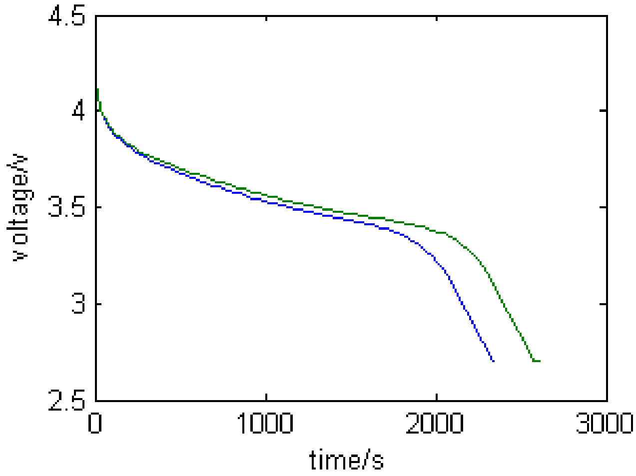

Fig. 1. The two discharge curves

274.7112	8.9389	3.4774	0.2843	-0.5746	3.7405	2340
303.4545	9.5747	3.4880	0.2824	-0.6869	3.9557	2580

Each feature has a different order of magnitude, if the extracted features are used directly for measuring the similarity, the contribution of each feature for distance is different, and so it will affect the accuracy of similarity. To eliminate this effect, we need to standardize the extracted features so that each feature is between 0 and 1. The normalization formula is shown as follows:

$$X = (x - min)/(max - min) \quad (4)$$

Where *max* the maximum value of the data, and min is the minimum value of the data.

C. Similarity Measurement

There are many methods to measure the similarity of data, such as DTW, information entropy, correlation distance, etc. In this paper, we use Euclidean distance.

D. Clustering based on Fast Search And Find of The Density Peak

The core idea of the clustering algorithm in this paper is to describe the clustering center. Clustering center has the following two characteristics: First, it has large local density, that is, it is surrounded by the point whose density is less. Second, the distance with other data points whose density is greater is relatively large. Set the dataset which is clustering $S = \{X_i\}_{i=1}^{N}, I_s = \{1,2,\ldots,N\}$ and $d_{ij} = dist(X_i, X_j)$ represents the distance between the data points X_i and X_j, we can use two variables to describe for each data point, the local density ρ_i and distanceδ_i, ρ_i is calculated as follows.

$$\rho_i = \sum_{j \in I_s \backslash \{i\}} \times (d_{ij} - d_c) \quad (5)$$

$$\times(x) = \begin{cases} 1, x < 0; \\ 0, x \geq 0; \end{cases} \quad (6)$$

The parameter d_c>0 is the cutoff distance. Generally, the principle of selecting d_c is that the average number of neighbors of each data point is about 1-2% of the total number of data points, the neighbor is the data points that are no more than d_c away from it , ρ_i represents the number of data points in the dataset S that distance with X_i is less than d_c.

The distance δ_i

Let $\{q_i\}_{i=1}^N$ denote a descending order subscript sequence of $\{\rho_i\}_{i=1}^N$

And it defined

$$\delta_{q_i} = \begin{cases} min\left\{d_{q_i q_j}\right\}, i \geq 2, i > j; \\ max\left\{\delta_{q_j}\right\}, i = 1, j \geq 2; \end{cases} \tag{7}$$

By definition, we know that δ_i is the minimum distance between X_i and the point whose local density is greater thanX_i, for the point whose density is largest, we choose the maximum value of the distance δ in the remaining points. After that, we can choose the data points with large ρ and δ as the cluster center according to the decision graph corresponding to (ρ, δ), For the point that is not cluster center, We classify it and the cluster center which is closest to it as a class

However, the method which select cluster center by decision graph contains some subjective factors, some people may think that the number of cluster center should be 6, some people may feel that should be 8, so we need a way to determine the number of clustering. We can define a variable γ_i, considering the value of ρ and δ

$$\gamma_i = \rho_i \delta_i \tag{8}$$

The point whose γ_i is larger, is more likely to be a clustering center, we first arrange γ_i in a descending order.

$$\gamma_{x_1} \geq \gamma_{x_2} \geq \cdots \geq \gamma_{x_N} \tag{9}$$

And then plot x_i as the abscissa and γ_i as the ordinate. There will be an obvious jump in the transition from the point which is not cluster center to cluster center, so as to determine the number of clusters. However, in order to eliminate the impact of dimension. The ρ and the δ need to be standardized before calculating

For the point which is not cluster center, we classify it and the point whose density is larger and is closest to it as a class. So clustering complete.

More details on the algorithm can be found in [1].

E. Discharge Curve Clustering Based on Density Peak Fast Searching And Finding of The Density Peak

- Extracting feature of the obtained battery discharge sequence. Each sequence is represented by the extracted features
- The extracted features are normalized.
- The parameter t∈ (0,1) is chosen to determine the cutoff distance d_c
- Calculating the Euclidean distance d_{ij} between each data point
- The distance d_c is determined. The distance d_{ij} (i<j) which is calculated in the previous step is arranged in descending order. The sequence is $d_1 \leq d_2 \leq \cdots \leq d_M$, $M = \frac{1}{2}N(N-1)$, and $d_c = d_{f(Mt)}$, the f (Mt) indicate rounded to the nearest integer of Mt.
- Calculating the local density $\{\rho_i\}_{i=1}^N$,The calculated $\{\rho_i\}_{i=1}^N$ is sorted in descending order to generate the subscript sequence $\{q_i\}_{i=1}^N$
- Calculating $\{\delta_i\}_{i=1}^N$ and $\{n_i\}_{i=1}^N$, n_i represents the data subscript of the point who is closest to X_i and local densities are larger than X_i
- Determine the cluster center $\left\{m_j\right\}_{j=1}^{n_c}$, and initialize the data point categorization attribute tag $\{c_i\}_{i=1}^N$,

$$c_i = \begin{cases} k, \text{If } x_i \text{ is the cluster center, it belongs to the Kth cluster.} \\ \qquad -1, \quad otherwise \end{cases}$$

- The points that are not clustering centers are classified, and its categorization attribute tag c_i=c_{n_i}

F. Cluster Curve Clustering Based on Hierarchical Clustering

- Calculating the DTW between the battery discharge curves to get the distance matrix.
- Each object is classified into a class, and data is classified as N classes.
- Specify the number of clustering.
- Locating the two classes that the distance is shortest and merge them into a single class.
- Recalculating the distance between the new class and all the old classes
- Repeat steps 4 and 5 above until the number of clusters is C to stop clustering.

III. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, we select 160 batteries, whose initial voltage is 4.2V and discharge at constant current, when the voltage drop to 2.7V, stopping discharging and getting the battery discharge curve. And we need to extract the features on the curve, then using fast search and find of density peak clustering algorithm to complete with grouping. We choose t = 0.04, and get the clustering center decision graph of (ρ_i, δ_i)) as shown in Fig 2. The abscissa is the local density ρ_i , and the ordinate is the distance δ_i. The point which has greater ρ_i andδ_i, is more likely to be the cluster center. So we can choose the four points which is marked in the graph as the cluster center according to the cluster number decision graph Fig 3.

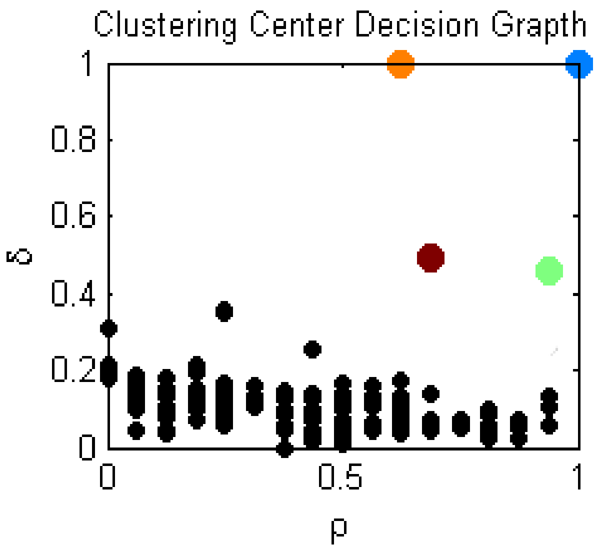

Fig. 2. Clustering Center Decision Grapth

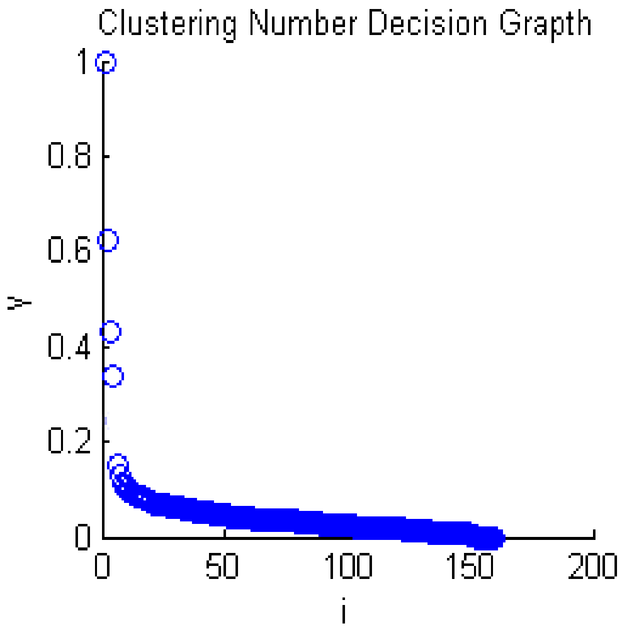

Fig. 3. Clustering Number Decision Graph

The decision graph of number of clustering as shown in Fig 3. The $\gamma = \rho\delta$, and γ is in descending order, That is to say $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_i$, and then plot i as the abscissa and γ_i as the ordinate. According to Fig.3 we can see that from the fourth point to the fifth point has a significant jump, the fifth point is closer to the sixth point, so we can determine that the optimal number of clusters is 4, So in Fig 2, we choose four points with larger ρ_i and δ_i as cluster centers, For the point that is not cluster center, We classify it and the point whose density is larger and is closest to it as a class. The clustering results are shown in Fig 4. The t represents the discharge time of the battery, v represents the voltage of the battery when the discharge, the same color discharge curve classified as a class. From the figure we can see that the algorithm classified similar discharge curve as a class.

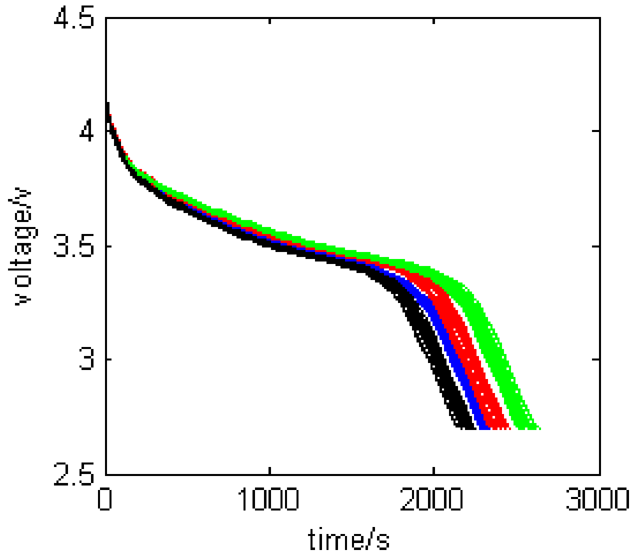

Fig. 4. The Result of Time Series Clustering Based On Feature

At the same time, we use the traditional hierarchical clustering algorithm to cluster the discharge curves. Because the obtained discharge curve has the characteristic of unequal length, Euclidean distance, correlation distance etc., cannot be used. Therefore, we choose the DTW to measure the similarity between the batteries, and use hierarchical clustering algorithm to complete the grouping. We choose the number of clusters to be 4, and use the longest distance method to measure the distance between clusters, the results are shown in Fig 5.

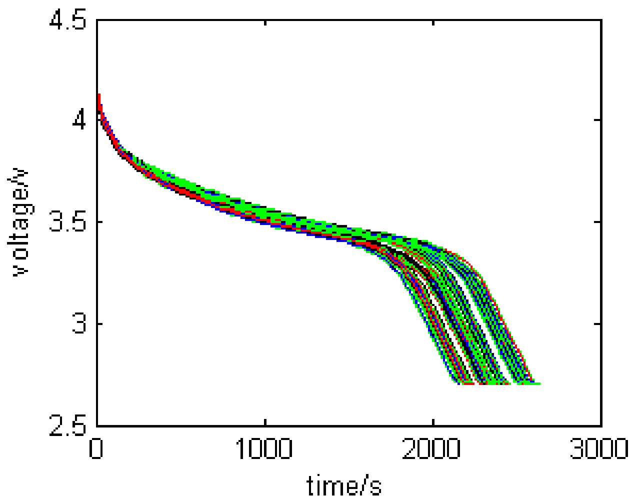

Fig. 5. The Result of Time Series Clustering Based On Raw Data

It is obvious that the clustering effect is not ideal. In this paper ,we select the same type battery to test, so the trend of the battery discharge curve is similar, If you use Dynamic Time Warp to measure the similarity between cells, it does not take into account the time factor of the battery discharge curve and cannot highlight the differences between the battery. So there is a certain error in measuring distance, resulting in the results of the battery group is not ideal. The method of feature extraction in this paper can not only improve the accuracy of the similarity but also reduce the complexity of the algorithm. We use the fast search and find the density of the peak to complete cluster, whose computed complexity is lower, and does not need to specify the number of clusters and cluster centers, so it has certain advantages.

IV. SUMMARY

Battery grouping has a wide range of applications. In this paper, we propose the battery grouping technology which is based on the battery discharge curve, Firstly, we will discharge the battery at the 1A current, when the voltage reaches 2.7V to stop discharging, and getting the time series of battery discharging. Secondly, we extract the time domain and frequency domain characteristics of each time series, including the kurtosis, skewness and Fourier transform coefficients of the sequence and so on. And then using the extracted features to measure the similarity between the batteries. Finally, we use fast search and find the density of the peak clustering algorithm to complete the battery grouping. Experimental results show that the proposed clustering algorithm is effective.

In this paper, we use the clustering algorithm of fast searching and finding density peaks to complete the matching. Compared with other clustering algorithms, such as K-means, hierarchical clustering, this algorithm does not need to initialize the clustering center or the number of clusters, which is determined by the decision diagram. So human intervention is less, the accuracy rate will be relatively high.

ACKNOWLEDGEMENT

This work was supported in part by Nature Science Foundation of China under Grant 61471151, U1609216, and 61671194.

REFERENCES

[1] Alex Rodriguez , Alessandro Laio . "Clustering by fast search and find of density peaks". Science 2014,344(6191):1492-1496

[2] Sun xu. "Analysis and Application of the Global Characteristics Cluster". Statistical Thinktank 2009,(3):55-59

[3] SONG Yonghua, YANG Yuexi, HU Zechun." Present Status and Development Trend of Batteries for Electric Vehicles". Power System Technology.2011,35(4):1-7

[4] LI Mei, Bai Fenglan . "Analysis of Similarity of DNA Sequences Based on Dynamic Time ." Journal of Biomathematics.2009,24(2):374-378

[5] Dhillon I S, Guan Y, Kulis B. Kernel k-means: spectral clustering and normalized cuts. Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2004: 551-556.

[6] Begum N, Ulanova L, Wang J, et al. Accelerating Dynamic Time Warping Clustering with a Novel Admissible Pruning Strategy. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2015: 49-58.

[7] Wang, T.; Tseng, K.J.; Zhao, J.; Wei, Z. Thermal investigation of lithium-ion battery module with different cell arrangement structures and forced air-cooling strategies. Appl. Energy 2014, 134:229–238.

[8] Song Ci ,Pei Tao. "Research Progress of Time Series Clustering Method Based On Characteristics". Progress In Geography, 2012,31(10):1307-1317

[9] Zhou Qian, Wu Tiejun. "Trend feature extraction method based on important points in time series". Journal of Zhejiang University.2007, 41(11):1782-1787

[10] WANG Yongchen，NI Jiangfeng，WANG Haibo，GAO Lijun，HU Daozhong，WANG Zidong." Sorting methods of lithium ion batteries consistency" Energy Storage Science and Technology.2013,2(5):522-527