Assessment Method for Residual Value of Lead-acid Batteries Based on PAM Clustering Algorithm

Xuesong Feng\textsuperscript{a}, Xiaokun Zhang\textsuperscript{b} and Yong Xiang\textsuperscript{*}

School of Materials and Energy, University of Electronic Science and Technology of China, Chengdu, China

\textsuperscript{*}Corresponding author e-mail: xyg@uestc.edu.cn, \textsuperscript{a}zilch.fxs@hotmail.com, \textsuperscript{b}zxk@uestc.edu.cn

Abstract. As the residual value of the lead-acid batteries is not effectively evaluated in the current scraping and recycling processes of the lead-acid batteries, the partition around medoids (PAM) clustering algorithm is adopted, and the scraped lead-acid batteries are classified with the changes in the temperature and charge-discharge occurring when the lead-acid batteries are in service as the characteristic parameters. Besides, the validity of the algorithm is validated using 300 lead-acid battery packs to be scrapped at the communication base station, from which the results showed that the residual value of the lead-acid batteries can be effectively distinguished by the partition around medoids algorithm according to the temperature and charge-discharge characteristic parameters.

1. Introduction

Lots of lead-acid batteries (LABs) are scrapped every year in the world. In 2014, 2,400,000t waste lead was generated from scrapping of LABs in China[1]. However, these batteries were uniformly recycled at a low price due to the lack of efficient residual value assessment methods, but the fact was that the residual capacity of these scrapped LABs was different. Batteries with higher residual capacity can be activated and repaired for reuse[2]. Only batteries with lower residual capacity need to be directly sent to a recycling factory. Therefore, accurate and proper classification of the LABs with different residual capacities can not only improve the economic value of the LABs, but also reduce the transitional scraping of the LABs to the natural environment.

2. Assessment method for residual capacity of LABs

As LABs are generally used as a backup power source and the discharge (deep discharge) frequency is very low, the direct reference data for assessing the residual capacity of the LABs are limited. Currently, there are three conventional methods for assessing the residual capacity of the LABs. 1. Natural discharge method: When the battery has natural discharge (relative to manual discharge, it refers to normal discharge of the batteries caused by grid fault) (over 80% DOD) for a short time before recycling, so as to directly obtain the real-time capacity of the batteries from the battery management software platform; 2. Discharge test method: When the batteries are recycled, special discharge equipment is used for deep discharge and record the real-time capacity of the batteries; 3. Impedance test method[3]: When the battery is recycled, an impedance tester is used to test the
internal resistance of the batteries, and the real-time capacity of the batteries is assessed based on experience. These three methods have their respective advantages and disadvantages, as shown in Table 1:

**Table 1. The advantages and disadvantages of the assessment method**

| Method               | Advantages                                      | Disadvantages                                      |
|----------------------|-------------------------------------------------|---------------------------------------------------|
| Natural discharge    | No auxiliary equipment, low costs and high accuracy| Low probability of deep battery discharge and poor applicability |
| method               |                                                 |                                                   |
| Manual discharge test| Highest accuracy                                | Complex operation, heavy time consumption and high expenses |
| test method           |                                                 |                                                   |
| Manual impedance test| Fast evaluation and easy operation              | Low accuracy                                       |

It can be seen that it is impossible to achieve balance for these three methods in evaluation speed, operation difficulty, accuracy and economic efficiency. In this paper, the partition around medoids (PAM) clustering algorithm would be used to analyze the changes in ambient temperature and charge-discharge occurring when the LABs are in service, and a method is established for assessing the residual capacity of the LABs so as to achieve a balance between the above four characteristics.

3. Impact of changes in ambient temperature and charge-discharge on the attenuation of LABs

![Figure 1. The temperature distribution (a) and the charge and discharge current distribution (b) of a sample communication base station.](image-url)
After the LABs are put into operation, the aging is mainly affected by temperature and charge-discharge. Figure 1 shows the temperature change curve and charge-discharge curve of the LABs at the communication base station during one-year operation. The back-up battery pack of the base station is usually consist of 24 single LABs, so the figure 1(a) adopt thermodynamic diagram to show the temperature distribution, it’s clearly represent the temperature changed a lot during one-year. The figure 1(b) shows the charge and discharge current distribution, it indicates the frequency of the charge and discharge is very low.

Paul Ruetschi [4] proposed that the corrosion rate of the polar plates and water evaporation rate would increase at high temperature, thereby reducing battery life; V. Svoboda et al. [5] Proposed that fast charging would reduce the battery life; Noman Bashir et al. [6] Proposed that poor charging, time to previous full charge and the minimum charge since the previous full charge were the key factors affecting battery life. Therefore, the temperature time series data and charge-discharge time series data were categorized into five dimensions in this paper: the duration of high temperature above 40℃, the load current, the cumulative discharge duration, the cumulative discharge frequency, and the cumulative charge duration.

4. Establishment of clustering algorithm

The data can be divided into meaningful clusters based on the data characteristics through clustering analysis[7] and can be used to assess the residual capacity of lead-acid batteries. The characteristic parameters affecting the lead-acid battery life described in Section 3 can be used to classify the LABs with different residual capacity.

4.1. PAM algorithm

The PAM algorithm is a common cluster analysis algorithm. As a technology based on representative objects, an actual object is selected to represent the cluster. A representative object is used for each cluster, and the remaining objects are assigned to the cluster in which the most similar representative object exists[8]. This algorithm is a strongly robust clustering algorithm, and is especially suitable for application scenarios where noise and outliers exist.

Specifically, in the given sample space $D = \{x_1, x_2, ..., x_n\}$, the clusters $C = \{C_1, C_2, ..., C_k\}$ are divided by the PAM algorithm to minimize the square errors:

$$E = \sum_{i=1}^{k} \sum_{x \in C_i} \text{dist}(x, o_i)$$

In which, $x$ is a point in the sample space; $o_i$ means the representative object of the $C_i$ cluster; $\text{dist}(x, o_i)$ denotes the distance between points. For the PAM algorithm, all attempts are not made through constant iteration $o_i$ until the clustering quality is not improved through any iteration.

4.2. Modeling

First, build an object $x$. According to Section 3, set the object format $x = \{x_{HTD}, x_{LC}, x_{DCD}, x_{DCR}, x_{CCD}\}$ consisting of 5 dimensions. $x_{HTD}$ Refers to the number of days when the temperature is higher than 40℃; $x_{LC}$ is the load current; $x_{DCD}$ denotes the cumulative duration of discharge; $x_{DCR}$ is the cumulative discharge frequency; $x_{CCD}$ means the cumulative discharge duration. Then, set up the algorithm process. The flow of the PAM algorithm is as follows:
Table 2. The PAM algorithm

| Step | Description |
|------|-------------|
| 1    | Randomly select k objects from D as the initial representative objects |
| 2    | Repeat |
| 3    | Assign each remaining object to the cluster represented by the nearest representative object |
| 4    | Randomly select one non-representative object \( O_{random} \) |
| 5    | Calculate the total cost S for replacing the representative object \( o_i \) with \( O_{random} \) |
| 6    | if S<0, then replace \( o_i \) with \( O_{random} \), and form a new cluster represented by k representatives |
| 7    | Until S no longer improved |

Figure 2 graphically shows the calculation process of the algorithm.

---

5. Test and validation

5.1. Test method

In this paper, real operation data of 300 battery packs to be scrapped at the communication base station were selected from the battery monitoring platform of China Tower (it is the world’s largest communication infrastructure operator, and has over 2 million communication base stations as the study object, and these battery packs have the following characteristics: 1. The actual capacity of each battery had been verified by the discharge test method; 2. The data were intercepted for over one years, and natural discharge (over 80%DOD) occurred before one year. The purpose was to conveniently validate the quality of the algorithm based on the known residual capacity of the battery. Table 3 lists some battery operation data by object.
Table 3. The data of the objects

| Object | $x_{HID}$ | $x_{LC}$  | $x_{DCD}$ | $x_{DCT}$ | $x_{CCD}$ |
|--------|-----------|-----------|-----------|-----------|-----------|
| $x_1$  | 43        | -76.29    | 63.95     | 31        | 115.76    |
| $x_2$  | 23        | -64.85    | 30.45     | 17        | 34.43     |
| $x_3$  | 60        | -36.82    | 39.42     | 18        | 36.48     |
| ...    | ...       | ...       | ...       | ...       | ...       |
| $x_{289}$ | 21     | -81.77    | 31.51     | 25        | 54.37     |
| $x_{289}$ | 58     | -46       | 57.29     | 17        | 26.39     |
| $x_{300}$ | 75     | -78.55    | 35.39     | 26        | 62.78     |

Before running the algorithm, set the preset number of clusters to be 3 based on general experience, corresponding to 3 categories with physical significance: usable, LABs, LABs with repair potential and truly scrapped LABs. Then, run the PAM clustering algorithm in the Python development environment so as to obtain the results, because the object had 5 dimensions in total and the data were visualized based on the hue, size and shape. It can be seen that the battery packs were divided into 3 clusters, and the cluster marks were defined as 1–3.

![Figure 3. The visualizing of the object data (a) and the PAM operated result (b)](image)

5.2. Validation

As the premise of two deep discharges had been set at the time of selecting the lead-acid battery sample data set, a new reference model can be defined by dividing the capacity difference between the two deep discharges, and the validity of the cluster was measured based on the correspondence degree of the PAM model and the reference model.

Specifically, the capacity differences between the two deep discharges of each object were also divided as per 3 clusters to obtain one reference model $C^* = \{C^*_1, C^*_2, C^*_3\}$. Given that $\lambda$ and $\lambda^*$ respectively represent the cluster mark vectors corresponding to $C$ and $C^*$. According to the Jaccard coefficient [9] calculation formula which was used to measure the validity of the clusters and the similarities, the cluster validity of the PAM algorithm was measured as follows:

$$JC = \frac{a}{a + b + c} = 0.87$$  (2)
It can be seen that the calculation results of the PAM and the reference model were highly similar, which meant that the battery packs with different attenuation were screened out by the PAM algorithm.

6. Summary

In this paper, a method of screening out obsolete LABs with high recycling value using the PAM algorithm was developed. Due to the temperature changes of the LABs and the charge-discharge difference of the in-service systems, the residual capacity of obsolete LABs was different. In order to increase the recycling value of the LABs, PAM clustering algorithm was used to categorize and classify the LABs with temperature and current distribution as characteristic indicators. Finally, with the operating data of 300 real battery packs of the communication base station as the samples, the calculation results were obtained, and the validity of the algorithm was validated by using the Jaccard coefficient for measuring the cluster similarity. The results demonstrated that the presented PAM algorithm is effective to distinguish the LABs with different residual capacity according to the characteristic indicators including their temperature distribution and charge-discharge behavior.

References

[1] Tian, Xi, et al. "The lead-acid battery industry in China: outlook for production and recycling." Waste Management & Research 33.11 (2015): 986 - 994.
[2] Singh, Anupama, and P. B. Karandikar. "A Broad Review on Desulfation of Lead-Acid Battery for ElectricHybrid Vehicle." Microsystem Technologies-Micro-and Nanosystems-Information Storage and Processing Systems, vol. 23, no. 6, 2017, pp. 2263 – 2273.
[3] Kwiecien, Monika, et al. "Determination of soh of lead-acid batteries by electrochemical impedance spectroscopy." Applied Sciences 8.6 (2018): 873.
[4] Ruetschi, Paul. “Aging Mechanisms and Service Life of Lead–acid Batteries.” Journal of Power Sources, vol. 127, no. 1, 2004, pp. 33 - 44.
[5] Svoboda, Vojt̆ĕch, H. Doering, and J. Garche. "The influence of fast charging on the performance of VRLA batteries." Journal of Power Sources 144.1 (2005): 244 - 254.
[6] Bashir, Noman, et al. “Lifetime Maximization of Lead-Acid Batteries in Small Scale UPS and Distributed Generation Systems.” 2017 IEEE Manchester PowerTech, 2017, pp. 1 - 6.
[7] Tan, Pang-Ning, et al. Introduction to Data Mining. 2005.
[8] Han, Jiawei, et al. Data Mining: Concepts and Techniques. 2011.
[9] Jaccard, Paul. “THE DISTRIBUTION OF THE FLORA IN THE ALPINE ZONE.1.” New Phytologist, vol. 11, no. 2, 1912, pp. 37 – 50.