Journal of the Korean Society of Manufacturing Technology Engineers 27:2 (2018) 92~98

https://doi.org/10.7735/ksmte.2018.27.2.92

J. Korean Soc. Manuf. Technol. Eng.
ISSN 2508-5107(Online) / ISSN 2508-5093(Print)

Development of Cluster Analysis System based on the Result of Secondary Battery Performance Evaluation

Sang Hun Ok[a], Chang Woo Ban[b], Min Young Park[b], Dong Young Jang[b*]

[a] Korea Electronics-Machinery Convergence Technology Institute, 232, Gongneung-ro, Nowon-gu, Seoul 01811, Korea
[b] Seoul National University of Science and Technology, 232, Gongneung-ro, Nowon-gu, Seoul 01811, Korea

ARTICLE INFO

Article history:

Received	17	March	2018
Revised	3	April	2018
Accepted	5	April	2018

Keywords:
Secondary battery
Cluster Analysis
Control and Measurement
Evaluation

ABSTRACT

The demand for battery modules for vehicle driving is increasing rapidly. However, many battery modules are not economically viable because they are either discarded or chemically recycled for use as stationary batteries. In this research, we analyze the residual performance of rechargeable batteries to determine the reuse suitability and propose a cluster analysis technique to minimize performance degradation owing to performance variation. The system constructed from the research is expected to demonstrate its value in supportings the utilization of limited resources economically because it directly intervenes in preventing and responding to problems that occur because of the performance error in the reuse of secondary batteries.

1. Introduction

Hybrid electric vehicle (HEVs), plug-in hybrid vehicles (PHEVs), and electric vehicle (EVs) have emerged as eco-friendly transportation vehicles in recent years and their regulations and support have been actively improved. Recently, following HEVs and PHEVs, EVs capable of operating freely from noise and harmful substances have been attracting attention. However, the distribution rate is delayed compared to the necessity and excellence. The biggest obstacle is the cost of the vehicle-driven rechargeable batteries, which are used in HEVs, PHEVs and EVs. The batteries account for a large portion of the vehicle cost.

Battery cell research is currently active; however, it does not lead to the market reaction in a short period of time. As the waste batteries discharged from 2018 have a state of health (SOH) of approximately 70-80 %, the waste batteries can be reused before recycling them by chemical treatment[1].

As the energy density requirement for the stationary battery module used in the energy storage system (ESS) is lower than that of the battery module for the driving vehicle, the reusability is high, and demonstration projects are being actively conducted at the national and enterprise levels. Vehicle-driving battery modules need to have a different preprocessing process to cope with problems that may occur when the unit and the combination of reuse are different owing to different operating environments, operating conditions, and driver's habits. The imbalance in battery performance leads to the performance degradation of the energy system. The battery management system (BMS) implements the imbalance compensation; however, it is inefficient, limited in scope, and hardware dependent[2].

Therefore, the present study evaluates the residual performance of a cell to be discarded from a battery module

* Corresponding author. Tel.: +82-2-970-6450
Fax: +82-2-6499-1883
E-mail address: dyjang@seoultech.ac.kr (Dong Young Jang).

for driving a vehicle and performs an analysis according to an index that constitutes a cluster according to the demand attribute of a reuse application. Hence, the goal is to fundamentally eliminate the problematic factors arising from battery imbalance.

2. Measurement System Integration

2.1 Measurement and Control System Construction

The control and measurement unit can be roughly divided into three parts. These three important parts consist of a controller capable of applying a power source or an electronic load to a battery according to a desired profile, a meter capable of observing the reaction to obtain a measured value, and a physical switching device that can switch power, electronic load injection, switch the probe to another sample, and perform measurements repeatedly. The details of this requirement are summarized in Table 1.

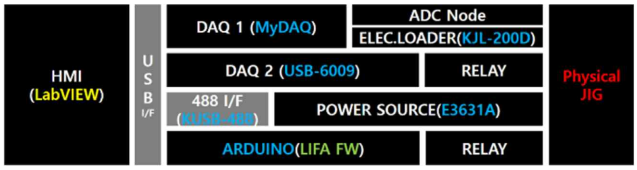


Fig. 1 Schematic of integrated hardware for secondary battery evaluation

According to the requirements above, the hardware is as shown in Fig. 1.

Here, the probe and jig, which interconnect between the control and measurement system and the sample, are manufactured exclusively to secure the secondary battery measurement node. The built-in dedicated probe minimizes the contact resistance by forming a contact pressure firmly along the polarity of the standard cell. Therefore, not only can the measured value be secured; but various electronic loads are applied as well to check the response of the battery such that the potential loss is minimized.

Table 1 Requirements of measurement and control system and devices

Device / Channel	Purpose	Remarks
E3631A / GPIB	Control Charging Voltage	GPIB-USB-HS
	Control Charging Current	
	Monitor Charging Voltage	
	Monitor Charging Current	
	Monitor Battery Voltage(O.L.)	
NI MyDAQ / AO0	Control Electric Load Discharging Current	KJL-200D
NI MyDAQ / AI0	Monitor Electric Load Voltage	
NI MyDAQ / AI1	Monitor Electric Load Current	
NI USB 6009 / DO0	Control 1st Relay for O.L. S.W.1	NER-11343 (8Ch. Relay)
NI USB 6009 / DO1	Control 2nd Relay for O.L. S.W.2	
NI USB 6009 / DO2	Control 3rd Relay for O.L. S.W.3	
NI USB 6009 / DO3	Control 4th Relay for O.L. S.W.4	
NI USB 6009 / DO4	Control 5th Relay for O.L. S.W.5	
NI USB 6009 / DO5	Control 6th Relay for O.L. S.W.6	
NI USB 6009 / DO6	Control 7th Relay for O.L. S.W.7	
NI USB 6009 / DO7	Control 8th Relay for O.L. S.W.8	
Arduino Uno R3 / DO0	Control 1st Relay for Char./Dischar. S.W.9	NER-11343 (8Ch. Relay)
Arduino Uno R3 / DO1	Control 2nd Relay for Char./Dischar. S.W.10	
Arduino Uno R3 / DO2	Control 3rd Relay for Char./Dischar. S.W.11	
Arduino Uno R3 / DO3	Control 4th Relay for Char./Dischar. S.W.12	
Arduino Uno R3 / DO4	Control 5th Relay for Char./Dischar. S.W13.	
Arduino Uno R3 / DO5	Control 6th Relay for Char./Dischar. S.W.14	
Arduino Uno R3 / DO6	Control 7th Relay for Char./Dischar. S.W.15	
Arduino Uno R3 / DO7	Control 8th Relay for Char./Dischar. S.W.16	

The hardware part of the above mentioned configuration is shown in Fig. 2. A total of 48 physical contacts are repeatedly switched according to the design sequence during the test. Therefore, prior to this measurement test, the channel resistance and contact resistance (sensitive elements that can affect the power injection and measurement value) must be checked. Subsequently, the calibration to ensure precision, reproducibility, and accuracy was conducted.

The hardware architecture and its operation, and the software architecture that performs the designed measurement and analysis techniques are shown in Fig. 3. The control and measurement test results through interfacing heterogeneous equipment are immediately reflected in the cluster analysis and the resulting derivation function, such that post-data management can be smoothly implemented. The code employing the state machine system design that adopts the main-state and the sub-state concept to process the multi-state transition condition is characterized.

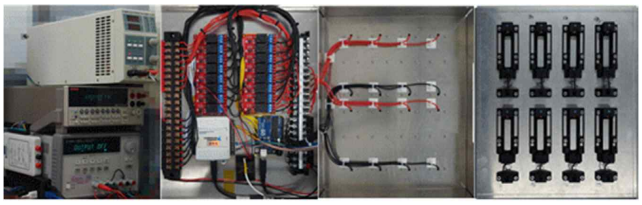


Fig. 2 Integrated hardware for secondary battery evaluation

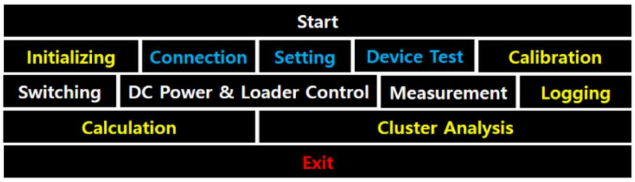


Fig. 3 Schematic of integrated software for secondary battery evaluation

Fig. 4 Integrated front panel for secondary battery evaluation

The front panel of the final determined system is shown in Fig. 4. In the right panel, marking can be given for labeling meaningful texts such as sample information to distinguish data, or cluster analysis and DC profile-related parameters can be set. Indicators for directing on-site operations to prepare the measurement process or for monitoring key parameters in real time during the test are also arranged based on the set parameter relevance. All of the above mentioned measurement data and group index results can be found in the chart at the top left, and the charts showing results from cluster analysis and the detailed cluster characteristics are arranged at the bottom of the chart. The state of the battery, the measurement situation, and cluster analysis results were used to intuitively grasp the gauge and two-dimensional (2D) scatter plot(auto scaled).

2.2 Analysis Process Construction

This measurement and analysis process starts with preprocessing to ensure the regularity of the secondary battery's state of charge (SOC). As samples, standard cells ICR18650 B4 (B4) and ICR18650 ME1 (ME1) were selectively used according to the situation.

A 2D scatter plotting of the measurement results of each sample can intuitively estimate the SOH. Among the measured values of the sample, the response characteristic to the positive DC pulse is reflected on the horizontal axis while the response characteristic to the negative DC pulse is reflected on the vertical axis. In general, the trend is in the form of a proportional function.

However, if the performance factor of the battery is not revealed in response to the positive DC pulse, a unit step function form coexists in the latter half (Fig. 5, left). This phenomenon occurs because some samples are measured in a band that does not reflect the SOH properly and act as a

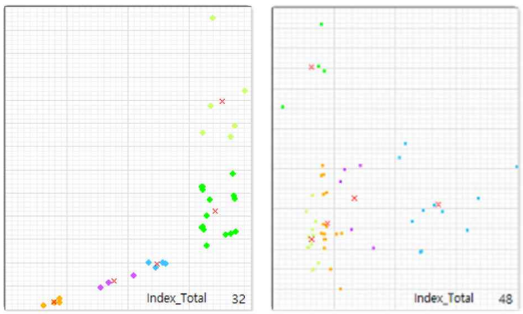


Fig. 5 Distribution of battery measurement results before voltage matching (left), and after voltage matching (right)

factor to distort the diagnosis results. This is because the secondary battery of high SOC immediately enters the constant voltage (CV) mode. Therefore, testing must be performed after ensuring the regularity through preprocessing with voltage matching. It can be confirmed that the data of the SOC normalized by the voltage matching is applied to the 2D scatter plot by reflecting the performance index as it is, by adding the same model sample of two groups including the existing sample part (Fig. 5, right).

2.3 System Performance Evaluation

This measurement and analysis process was performed in accordance with the devised DC profile technique. The electrochemical impedance spectroscopy (EIS) technique can also be used to measure the battery usage environment; however it is not useful in the systems covered in this study[3]. In terms of nondestructive inspection, expensive equipment is required, which is not economical and is sensitive to noise. It is disadvantageous in realizing measurement technology and reproducibility. Based on the limitations of the hybrid pulse power characterization (HPPC) and direct current internal resistance (DCIR) techniques, the newly developed DC profile technique can measure the SOH and available capacity by observing the load response characteristics in a short period of time. The data obtained in the process are shown in Fig. 6.

A pattern for the idle state utilization was set up after the battery was loaded. The open circuit voltage (OCV) measurement and the DC pulse injection have sequential access to the module unit samples; therefore the sample outside the area of interest is idle for the next measurement. The available capacity estimation is performed by measuring the DC pulse power and the OCV high resolution. By calculating the resolution of the instrument used in this system, it takes approximately 300 s to examine one module consisting of eight cells; consequently, the usable capacity can be calculated by allowing 0.5 % error. By examining the

Fig. 6 Response at developed DC profile on B4, ME1(marked as a green window) models

economical efficiency and extending the inspection time by 100 %, the resolution is improved to more than 200 %; therefore, the available capacity variation in the sample can be evaluated.

We implemented K-means, K-medians, fuzzy C-means (FCM), Gaussian mixture models, and vector quantization (VQ) as cluster analysis techniques. In the standard algorithm of K-means, the variance in cluster is

$$V_s = \sum_{i=1}^{k} \sum_{j \in S_i} |x_j - \mu_i|^2$$

V_s : Variance in cluster
S_i : Group of point in one cluster
μ_i : Centroid of ith cluster

by computing the Euclidean distance from each cluster to μ_i and distributing the data by finding clusters closest to the data, a group of point in one cluster to minimize distribution V_s, should

$$S_i^{(t)} = \left\{ x_p : |x_p - \mu_i^{(t)|^2} \le |x_p - \mu_j^{(t)}|^2 \forall j, 1 \le j \le k \right\}$$

be calculated. μ_i is the center of gravity of each cluster element.

When the calculation is repeated, the center of gravity of each cluster element is defined again by

$$\mu_i^{(t+1)} = \frac{1}{|S_i^{(t)}|}$$

When the cluster is not varied, the analysis is completed. K-medians is a modified algorithm of the K-means technique, and the center of gravity of the cluster is obtained by the data median. When sorted according to the size of the data, the center of the sequence is called the median, and the center of gravity of the cluster is determined according to this principle. FCM is one of the fuzzy cluster analysis techniques. The algorithm almost similar to the K-means method; however, unlike the hard cluster analysis that clearly determines which cluster the data belongs to, it belongs to a soft cluster analysis that weigh the population to which the data belongs. n finite sets of data $X = \{x_1, ..., x_n\}$ are the finite c-arrays $C = \{c_1, ..., c_c\}$ and the partition matrix is $W = \omega_{ij} \in [0,1]$, $i = 1, ..., n$, $j = 1, ..., c$. When the degree of each element belonging to the cluster c_j is denoted by ω_{ij}, the objective function is developed according to the goal of

minimizing the FCM value[4].

$$\underset{c}{argmin}\sum_{i=1}^{n}\sum_{j=1}^{c}\omega_{ij}^{m}\left|\left|x_i-c_j\right|\right|^2$$

$$\left(where,\ \omega_{ij}=\frac{1}{\Sigma_{k=1}^{c}\left(\frac{\left|\left|x_i-c_j\right|\right|}{\left|\left|x_i-c_k\right|\right|}\right)^{\frac{2}{m-1}}}\right)$$

VQ is divided into codebook generation, encoding, and decoding. As this is a supervised learning technique, the process of dividing into the number of predetermined clusters proceeds[5]. The Gaussian mixture models are Gaussian models that represent the natural phenomena in a mixed model and minimize errors in data distribution. In the parameter estimation process, the calculation is performed to determine the number of samples of the expectation-maximization process belonging to which Gaussian; when the convergence condition is satisfied, the analysis is completed[6].

In the cluster analysis, the B4 model without additional processing was used as the sample. We set the stopping criteria condition (tolerance, maximum iteration) as a control variable. The scale has 14 modules, totaling 112 cells, and 10 times analysis using a technique performed with 5 class sets. The results of the method are compared with each other focusing on data cohesion, deviation, and reproducibility. The results are summarized in Fig. 7. As the FCM technique provides constant data for all trials, the standard deviation is 0, which is most advantageous for reproducibility and the VQ technique is disadvantageous. The data related to the internal resistance distribution directly related to the cell output density show the best results in the K-means technique. The difference between the standard deviations of the x and y data is due to the distribution of the measurement data. As shown, this result alone reveals a much denser distribution characteristic for the y axis of the raw data. Using these characteristics, a selective weight can be assigned to the sample measurement result data according to the property of the secondary battery reuse application, which means that it can be reflected directly in the result of the cluster analysis technique because it is revealed in the standard deviation data.

Virtual samples were generated based on random values to verify the cluster analysis performance by arranging the 2D scatter plot area to the area where no actual sample measurement data was distributed. In addition, as there is no practical limitation in determining the size of a virtual sample, the cluster analysis algorithm can be verified in various angles. In particular, the distribution characteristics can be simulated, such that the pattern as well as the uniform distribution can be test. As many samples are available, the stopping criteria condition was looser than the actual sample analysis condition. In the simulation process, the scattering tendency according to the number of classes in the analysis parameters was observed. However, when the value exceeded eight, the number of classes was fixed to eight because the tendency was similar to the geometric similarity. The results of analyzing 10000 virtual samples by the FCM technique (Fig. 8, left), and the results of analyzing 5000 virtual samples by the VQ technique (Fig. 8, right) can also be visually confirmed to be divided into eight classes. Our simulation results show no definite correlations between the number of classes and the sample size for 1000-10000 samples.

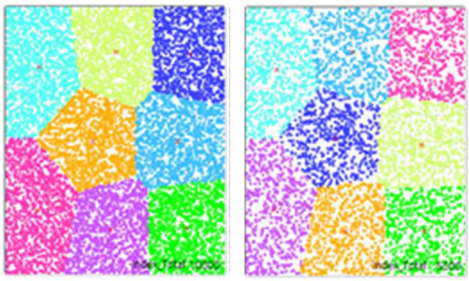
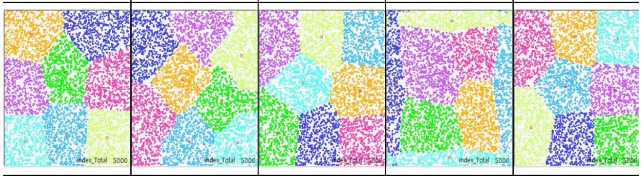


Fig. 8 Visualized cluster analysis result comparison on huge virtual specimens

	K-Means	K-Medians	Fuzzy C-Means Clustering	Gaussian Mixture Models	Vector Quantization
X_Stdev_Stdev	0.003631	0.006208	0.000000	0.003416	0.009257
Y_Stdev_Stdev	0.000302	0.000431	0.000000	0.001280	0.000410
Diff_Stdev_Stdev	0.003329	0.005777	0.000000	0.002136	0.008847

Fig. 7 Class data summary according to cluster analysis method

Table 2 Virtual data scattering according to cluster analysis method for intuitive evaluation

k-means	k-medians	FCM	Gaussian mixture models	VQ
				

Therefore, only 5000 virtual samples were analyzed.

To first examine the economics, we summarized the compute time required to analyze 5000 virtual samples for each analysis technique without using the computational resources that are input into the analysis technique. The VQ technique required stable analysis time with very low standard deviation (approximately 9.7) even after repeated execution, and the average value was also good. However, the analysis time of the FCM and Gaussian mixture models was 15734.9 s and the standard deviation was the largest at 3645.1. Further, we found that it was unsuitable for quasi-uniform distribution samples. In addition, the characteristics of small- and medium-scale samples were found to be different from those of the analysis. As the implementation of FCM was repeated, unrecognized deviations from the actual sample data analysis were revealed and the numerical value was very small. However, we assumed that the standard deviation of 0 was a coincidence or the scale of the deviation was too small to be recognized. Next, the techniques are compared by that the quality of the analysis result can be measured intuitively through the distribution in the analysis result. Table 2 summarizes the results of the analysis of 5000 specimens with quasi-uniform distribution to evaluate the geometrical composition of each technique. The results are summarized as images in Table 2. For the samples included in the ideal class, the standard deviation should be zero owing to the matching of all the performance data; however, an ideal analysis result would have a regular polygon shape sharing the neighbor side because of the practical constraint in the virtual samples of the similar distribution. Some analytical results show that each class has a polygon shape that shares neighbors. At this time, the quality of the analysis result can be intuitively evaluated through the geometrical composition of each class. The boundaries of the sides are clear and the closer the sides are in the polygon, the closer it is to the ideal result. The results

Table 3 Visualized VQ result on 10,000, 4-Gaussians virtual specimens

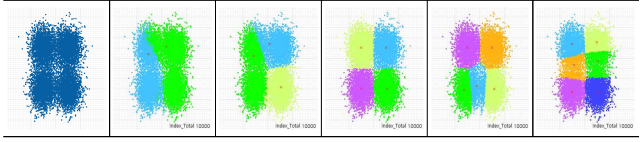
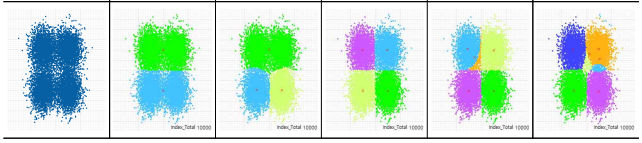
Before analysis	2 classes	3 classes	4 classes	5 classes	6 classes
					

Table 4 Visualized Gaussian mixture models analysis result on 10,000, 4-Gaussians virtual specimens

Before analysis	2 classes	3 classes	4 classes	5 classes	6 classes
					

of the Gaussian mixture model technique did not reveal meaningful class distributions.

Finally, the virtual sample data following the Gaussian distribution at 10000 are analyzed and the results are summarized in Table 3 and Table 4. Unlike quasi-uniform distribution samples, 4-Gaussians samples can identify four groups with noticeable cohesion. By implementing different analytical techniques, we evaluated the process of manipulating the number of classes and the results of the analysis comprehensively. The best results were obtained with the lowest sample performance variation in the class and the analysis was performed with four classes. This is a common feature regardless of the analysis technique, and the analysis time is generally short.

3. Conclusion

In the experimental stage performed by this analysis system, the measurement technique designed for the performance evaluation of the secondary cell was applied to the actual sample and the cluster analysis was performed with the measurement data. In addition, virtual samples were generated to verify the cluster analysis algorithm in various ways.

The available capacity estimation results using the measurement technique were evaluated for the resolution of the detection error in a specific SOC band. Consequently, effective numerical values can be obtained, and the resolution of the available capacity estimation can be improved if the

performance measurement resolution based on the sample response is low.

In the cluster analysis experimental stage, we could improve the system. When the combination of analytical algorithm and sample distribution characteristics is not appropriate, we found out that even if the number of required classes is small, it requires high computing resources. In addition, the results of the intuitive analysis on the visualized 2D scatter plot and the quantitative values of the class together show the inverse trend in the relationship between analytical performance and analysis time. When analyzing the characteristics of the sample, the cluster analysis technique and related parameters must be tuned.

The system constructed as a result of the research is expected to demonstrate its value in supportings the utilization of limited resources economically, because it directly intervenes in preventing and responding to problems that occur owing to the performance error in the reuse of secondary batteries.

Acknowledgement

This research was supported by the Ministry of Trade, Industry and Energy (MOTIE), KOREA, through the Education Support program for Creative and Industrial Convergence (Grant Number N0000717).

References

[1] Neubauer, J. S., Pesaran, A., 2011, The Ability of Battery Second Use Strategies to Impact Plug-in Electric Vehicle Prices and Serve Utility Energy Storage Applications, Journal of Power Sources, 196 10351-10358.

[2] Wikimedia Foundation, Inc. Viewed 14 Jan. 2018, Metadata: Wikipedia Is One of the Largest Reference Website <https://en.wikipedia.org/wiki/Battery_balancing>.

[3] Diard, J. P., Le Gorrec, B., Montella, C., 1998, EIS Study of Electrochemical Battery Discharge on Constant Load, Journal of Power Sources, 70 78-84.

[4] Wikimedia Foundation, Inc., Viewed 28 March 2018, Metadata: Wikipedia Is One of the Largest Reference Website <https://en.wikipedia.org/wiki/K-means_clustering>.

[5] Lee, B. H., Jang, W. S., Kang, H. I., 2007, An Watermarking Method Based on Singular Vector Decomposition and Vector Quantization Using Fuzzy C-Mean Clustering, Journal of Korean Institute of Intelligent Systems, 17:7 964-969.

[6] Wikimedia Foundation, Inc., Viewed 14 February 2018, Metadata: Wikipedia Is One of the Largest Reference Website <https://en.wikipedia.org/wiki/Mixture_model>.