State-of-Health Estimation for Li-ion batteries using Process Mining

1st Yunseo Kim
dept. of Energy and Chemical Engineering
Ulsan National Institute of Science and Technology
Ulsan, Korea
yunseokim@unist.ac.kr

1st Yeonsu Kim
dept. of Industrial Engineering
Ulsan National Institute of Science and Technology
Ulsan, Korea
yeon17@unist.ac.kr

***Abstract*—The reliability and safety of Li-ion batteries (LIBs) are crucial issues for optimizing and developing the batteries. However, accurately estimating the state-of-health (SoH) for LIBs remains challenging. This is because that the battery degradation is complex and ambiguous. In this paper, we aimed to precisely estimate the SoH of LIBs using process mining techniques for feature selection. Process mining, a technique used to extract and analyze data from actual processes, enhancing the efficiency of business processes, was applied to select meaningful features from data generated during the battery degradation. The method with process mining showed outstanding performance with lower RMSE than without that.**

***Index Terms*—State of Health (SoH), Li-ion battery (LIB), Process Mining (PM), Machine Learning (ML)**

I. Introduction

Li-ion batteries (LIBs), with high energy density, long lifetime, and low cost, have been widely applied as energy sources for electric vehicles (EVs) [1]. However, LIBs degrade with usage and time, resulting in capacity fade. In response to this problem, precisely estimating SoH for LIB has been required to optimize and develop the batteries. To conduct this study, there are several issues due to the complex degradation mechanism of the batteries [2].

Feature Selection is the process of selecting the features of the data to be used for model construction. The data used to assess battery health can vary widely and can be vast. However, not all data can have an equal impact on health assessments. Therefore, it is important to select important features to improve the accuracy of the model and increase computational efficiency.

Process mining is a technology that extracts and analyzes data from real processes to help you understand processes. It can be used not only in the business field but also in various fields such as the operation process of the battery. It can be used to analyze the data generated during the operation of the battery to select an effective feature and apply it to the state evaluation.

To improve battery state estimation accuracy, feature selection was performed using process mining. This allowed us to more accurately select the functions used to estimate battery status. By taking advantage of process mining's explanatory possibilities, we selected clear features and eliminated redundancy between functions. This increased the explanatory power and efficiency of the model. Although battery and process mining are completely different fields, converging each other's technologies allowed us to see new perspectives and usefulness. This led to new discoveries and innovations. Batteries charged and discharged under more complex and varied conditions can be difficult to offer the same level of solution. This suggests difficulties in generalization and suggests that model development for more diverse conditions is needed in the future.

In section 1 and 2, we first introduce the background of this study and related works. In section 3, we describe the dataset for estimating SoH and the method used. Then, in section 4, we analyze the result and conclude it with the process of evaluation and discussion.

II. Related Works

Fei, Z. et al. (2021) [3] described the basic structure for predicting SoH using ML. They selected useful features by eliminating redundance and irrelevance and showed the effect of feature selection. In addition, Yang, D. et al. (2018) [4] extracted four features from the charge voltage curve and applied Gaussian Process Regression (GPR) to estimate SoH for LIBs. They presented the outstanding performance of GPR than other ML models.

Academic fields such as economics, engineering, life, and natural sciences can benefit greatly from identifying anomalies in processes through process mining or by supporting predictive analysis of what is measured. In Ziolkowski et al. (2022) [5], raw detected time series data were used in a simulated seasonal coastal upstream system using marine science data.

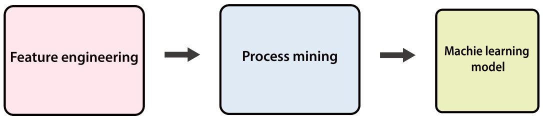


Fig. 1. Flow of Method

This presents a new methodology for identifying anomalies in the process.

In many cases, event data generated during process execution occurs at a much lower level of abstraction, and in some cases may be continuous, such as continuous sensor data. Koschmider et al. (2020) [6] presents a framework for discovering activity and process models from event location sensor data. This framework has the flexibility to apply to any dataset from raw sensor data.

It appears there hasn't been prior research focusing on continuous battery data. Therefore, our study aims to pave the way as the pioneering research in this domain.

III. Method

The overarching process in figure 1 unfolds as follows:
1. Initiating with feature engineering, which involves identifying and selecting relevant features for the analysis.
2. Next, the implementation of process mining takes place. This step includes analyzing and visualizing the data to uncover patterns, bottlenecks, and inefficiencies in the process.
3. Finally, the utilization of machine learning models comes into play. These models are used to make predictions, classify data, or identify anomalies based on the patterns and insights extracted from the process mining stage.

By following this step-by-step process, we can gain valuable insights, optimize their operations, and make data-driven decisions.

A. Dataset

To successfully conduct this study, we utilized the dataset obtained from Sandia National Labs (SNL). These datasets served as the foundation for our research and analysis. For our experimentation, we specifically focused on commercial 18650 LFP cells that were cycled to 80% of their maximum capacity. To ensure accuracy and reliability, we conducted a thorough capacity check. This check involved subjecting the cells to three charge/discharge cycles, ranging from 0% to 100% state of charge (SOC), at a rate of 0.5C. By employing this rigorous testing methodology, we were able to gather comprehensive data and draw meaningful conclusions from our study.

START
A
B
C
END

Fig. 2. Example of PM

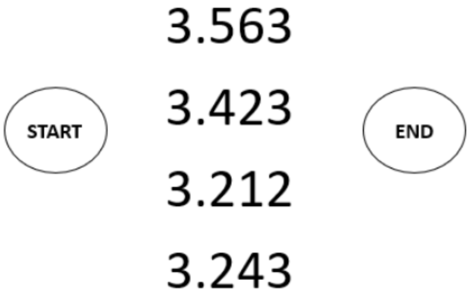


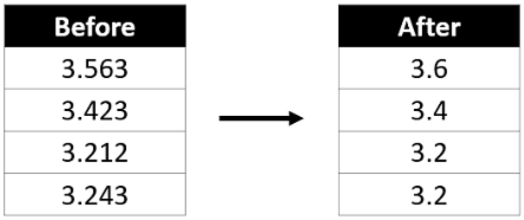
Fig. 3. Example of battery data

B. Process Mining Technique

In Figure 2 and 3, Since the existing PM deals with segmented data, there is a limitation that continuous battery degradation data cannot be applied. Therefore, we discretized the numbers by rounding off the first digit of the decimal point like Figure 4.

The original battery degradation data, which lacks evident procedures and continuity, posed challenges for the application of process mining. Therefore, to apply process mining to battery degradation data, we transformed 18 extracted features from the battery degradation dataset into an event log format. Among them, two features with a large NaN value were excluded.

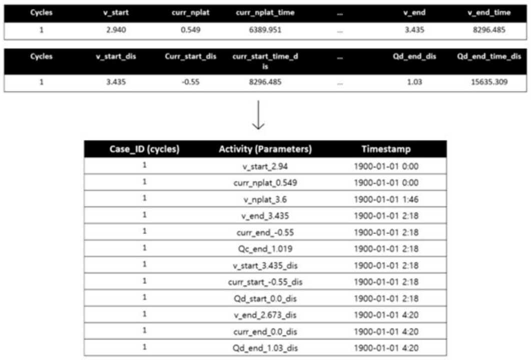
As depicted in Figure 5, each cycle is recognized as a single case ID. Considering the presence of duplicate feature names in both charging and discharging, and leveraging the time information inherent to each feature, we generated timestamps. These timestamps were created by converting time information into seconds and assigning arbitrary dates. Each feature's information is transformed into activities. Only features with existing timestamps are considered, resulting in a total of 12



Before
3.563
3.423
3.212
3.243

After
3.6
3.4
3.2
3.2

Fig. 4. Example of discretization



Cycles	v_start	curr_nplat	curr_nplat_time	...	v_end	v_end_time
1	2.940	0.549	6389.951	...	3.435	8296.485

Cycles	v_start_dis	Curr_start_dis	curr_start_time_dis	...	Qd_end_dis	Qd_end_time_dis
1	3.435	-0.55	8296.485	...	1.03	15635.309

↓

Case_ID (cycles)	Activity (Parameters)	Timestamp
1	v_start_2.94	1900-01-01 0:00
1	curr_nplat_0.549	1900-01-01 0:00
1	v_nplat_3.6	1900-01-01 1:46
1	v_end_3.435	1900-01-01 2:18
1	curr_end_-0.55	1900-01-01 2:18
1	Qc_end_1.019	1900-01-01 2:18
1	v_start_3.435_dis	1900-01-01 2:18
1	curr_start_-0.55_dis	1900-01-01 2:18
1	Qd_start_0.0_dis	1900-01-01 2:18
1	v_end_2.673_dis	1900-01-01 4:20
1	curr_end_0.0_dis	1900-01-01 4:20
1	Qd_end_1.03_dis	1900-01-01 4:20

Fig. 5. Data Tramsformation

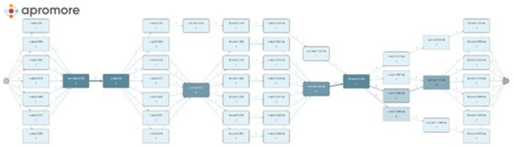


Fig. 6. Process Map

features transformed into activities for each case ID.

This is represented as a detailed and comprehensive process map using apromore, which is considered to be one of the most widely used and highly regarded software tools for process discovery. The process map, as depicted in Figure 6, provides a visual representation that effectively captures and illustrates the various steps and components involved in the process. With the help of apromore, organizations can gain valuable insights and a deeper understanding of their processes, enabling them to identify areas for improvement and make informed decisions to optimize their operations.

As a result of representing the characteristics of battery charging/discharging in the process map, a total of five major transactions were identified. Five dark-colored boxes mean frequently occurring activity. Therefore, in order to solve redundancy and irrelevance, major feature selection was performed by excluding curr_nplat, v_nplat, curr_end in the charging process and curr_start and Qd_start in the discharging process.

1) Ridge Regression (Ridge): Ridge regression is a variant of linear regression, with L2 regulation added to reduce the complexity of the model. This model is effective in solving multicollinearity problems and preventing overfitting. The performance was evaluated by applying Ridge regression to the battery condition evaluation. Changes in prediction results using Ridge regression can be seen through Figure 8. The RMSE in Ridge without PM case is 0.001426047 and with PM case is 0.001261.

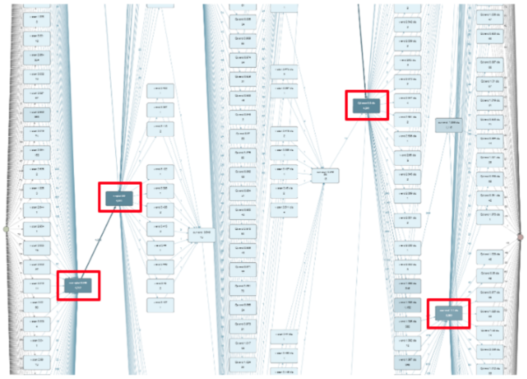


Fig. 7. Excluded Features in Map

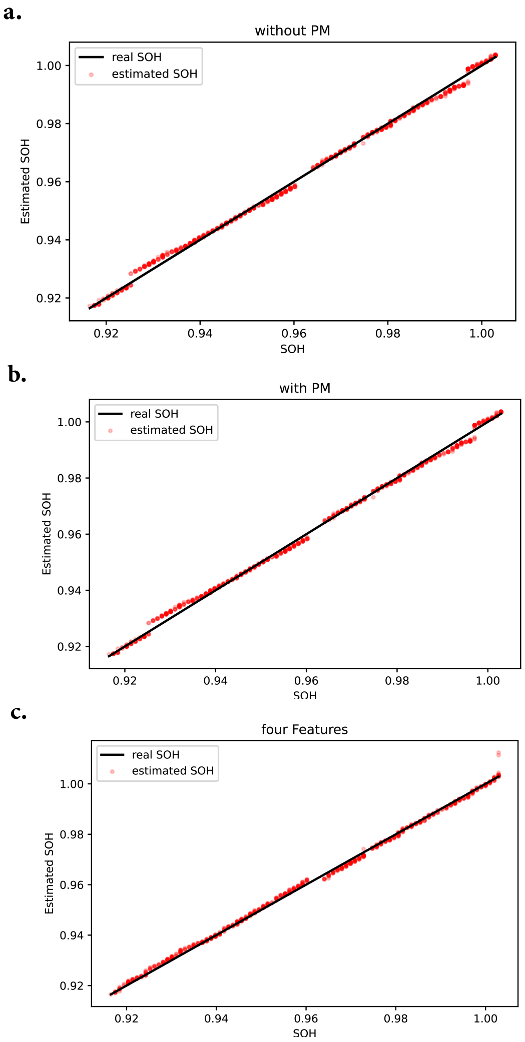


Fig. 8. SoH of Ridge Regression

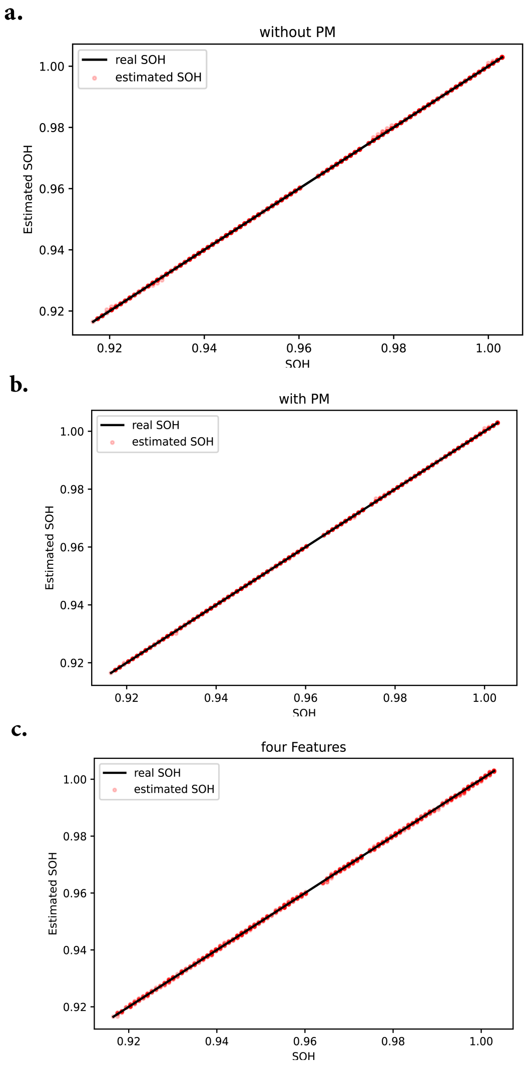


Fig. 9. SoH of Gaussian Process Regression

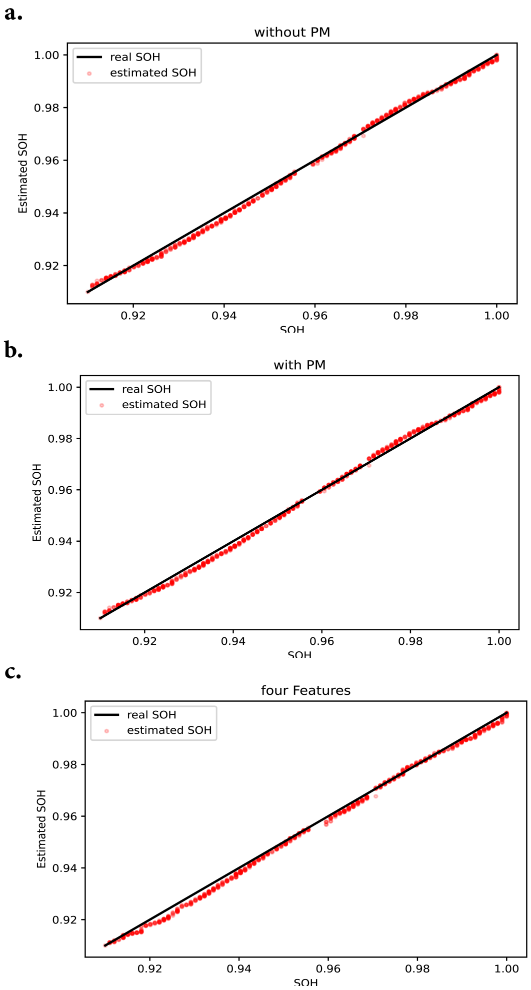


Fig. 10. SoH of Support Vector Regression

2) Gaussian Process Regression (GPR): GPR is a Gaussian process regression analysis, which is useful for regression modeling on complex data. The model makes predictions considering the uncertainty of the data and works flexibly on high-dimensional datasets. GPR modeling based on battery charge/discharge data can provide more sophisticated and uncertainty-considered predictions compared to linear regression. Changes in prediction results using Gaussian Process Regression can be seen through Figure 9. The RMSE in GPR without PM case is 0.000637772 and with PM case is 0.000364199.

3) Support Vector Regression (SVR): Support Vector Regression (SVR) is one of the regression analysis techniques used in machine learning, applying the idea of Support Vector Machine (SVM) to the regression problem. SVR determines prediction lines in a way that maximizes margins between data points, which allows it to model nonlinear relationships of data. It uses kernel tricks to map data into high-dimensional spaces to identify complex patterns and form outliers-resistant models. SVR provides strong predictive power even in small datasets and can be utilized in a variety of fields. Changes in prediction results using Support Vector Regression can be seen through Figure 10. The RMSE in Ridge without PM case is 0.003461954 and with PM case is 0.001505678.

C. Comparative methodology

The purpose of Figure 11 is to illustrate the variation in voltage over a specific time period. In part (a) of the figure, the recorded voltage changes are shown for cycles 1, 21, 41, 61, 81, 101, 121, and 141. By analyzing the graph's schematic features, we can extract four values that are relevant for machine learning: F1 (CC duration), F2 (CV duration), F3 (slope), and F4 (vertical). It is important to note that the data presented in this figure is sourced from reference [4].

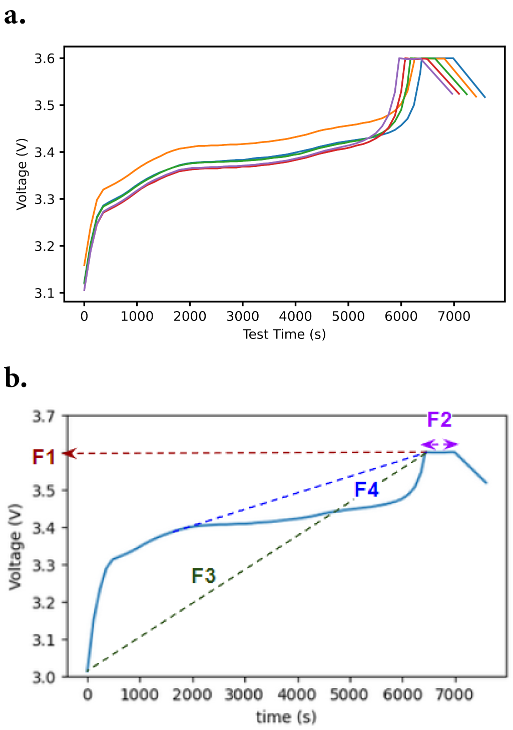


Fig. 11. Comparative Feature Selection

TABLE I
PERFORMANCE OF MACHINE LEARNING

	Ridge	GPR	SVR
without PM	0.001424	0.000638	0.003462
with PM	**0.001261**	**0.000364**	**0.001506**
4 Features	0.001591	0.000826	0.001701

IV. Result

In table I and figure 12, we used Gaussian Process Regression (GPR), and Ridge Regression (Ridge), Support Vector Regression (SVR) as regression models to compare performance. Root Mean Squared Error (RMSE) was used as a performance indicator. In process mining-based prediction, RMSE showed significantly lower values than RMSE of the existing methodology. Although the method using the conventional method of '4 features' showed better results with a slight difference in ridge regression, other machine learning models showed excellent results. Through this, we proved that the process mining-based prediction method we presented was valid.

V. Conclusion

By extracting features using process mining, our study utilized the advantage of explanatory possibilities. The process mining-based feature selection we proposed showed superior performance compared to the existing method. Through this, we demonstrate that more accurate and meaningful features can be of great help in battery state evaluation. This is an important finding that can help better understand and predict the deterioration mechanism and state changes of real batteries. Sometimes, however, our approach can perform poorly when compared to conventional methods. This can be attributed to several factors, such as the characteristics of the data or the constraints that occurred during the modeling process. These results show that the process mining technique we used does not provide an optimal solution in all situations. Therefore, it should be recognized that our method may be limited under certain conditions.

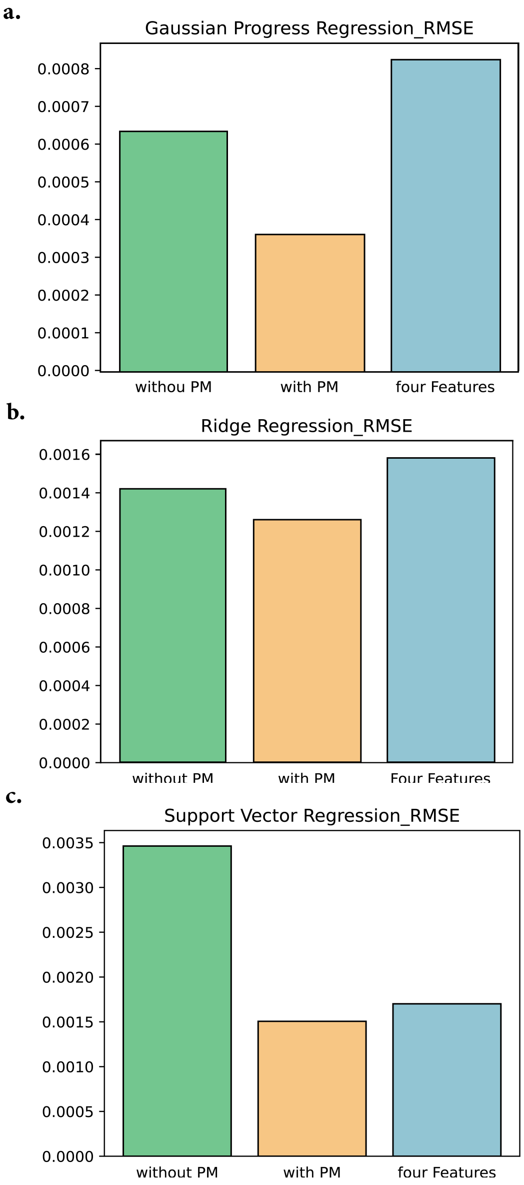


Fig. 12. RMSE of three ML models

This study confirmed the importance and potential of battery state evaluation using process mining. However, it is worth recognizing that our approach does not provide a perfect solution in all cases. As for the future research direction, it is important to explore the generalization potential of the results using more diverse datasets and various process mining techniques. In addition, improving performance through combination with existing methods or improved modeling techniques is also an important task. This study made it possible to understand more clearly the complexity and difficulty of predicting battery condition evaluation. This understanding will play an important role in developing practical solutions that can be applied in real industry and technology. Finally, this study demonstrates the applicability of process mining techniques in complex systems such as batteries, suggesting the possibility that these methodologies could be useful in other industries and technical fields. Therefore, it is expected that this methodology will be used to lead future technological development and innovation in industrial fields.

References

[1] Gandoman Foad H, et al (2019). Concept of reliability and safety assessment of lithium-ion batteries in electric vehicles: Basics, progress, and challenges. Applied Energy.

[2] Mikolajczak Celina, et al (2012). Lithium-ion batteries hazard and use assessment. Springer Science Business Media.

[3] Fei, Z., Yang, F., Tsui, K. L., Li, L., Zhang, Z. (2021). Early prediction of battery lifetime via a machine learning based framework. Energy, 225, 120205.

[4] Yang, D., Zhang, X., Pan, R., Wang, Y., Chen, Z. (2018). A novel Gaussian process regression model for state-of-health estimation of lithium-ion battery using charging curve. Journal of Power Sources, 384, 387-395.

[5] Ziolkowski, Tobias Koschmider, Agnes Schubert, René Renz, Matthias. (2022). Process Mining for Time Series Data.

[6] Koschmider, Agnes Janssen, Dominik Mannhardt, Felix. (2020). Framework for Process Discovery from Sensor Data.