Data Imputation Framework for Cloud-Connected Battery Management Systems

Umar Jamil, Faiz Majeed, Miyoung Lee, Seunghyun Lee, Jiwon Park, Jonghoon Kim*

*Chungnam National University, South Korea

umarjamil10000@gmail.com, faizmajeed07@gmail.com, yoy0307@naver.com, tmdgus7072@gmail.com, pjkmi8830@naver.com, *whdgns0422@cnu.ac.kr

Abstract

Cloud-connected battery management systems generate massive volumes of operational data from electric vehicles and energy storage systems, but data quality degradation from sensor faults, communication failures, and noise measurement severely compromises monitoring accuracy and safety. This paper presents an intelligent data cleaning framework combining deep learning-based outlier detection and LSTM-based data restoration with feature fusion for cloud-based battery management. The proposed method employs temporal feature analysis to accurately detect dirty samples including outliers, noise-polluted data, and missing values, while a feature fusion approach combining temporal and model-based features enables precise data reconstruction. Validation using real electric bus operation data from a cloud-based monitoring platform demonstrates that the framework achieves 93.3% detection rate for noise-polluted samples, 100% detection for missing values, and maintains restoration accuracy of 98.97% for noise-polluted data and 97.89% for missing data.

Keywords: Battery management system, Cloud data, Data restoring, Long Short-Term Memory (LSTM)

Ⅰ. Introduction

The proliferation of electric vehicles and large-scale energy storage systems has generated unprecedented volumes of battery operational data. Cloud-based battery management systems (BMS) leverage advanced computing infrastructure to provide intelligent monitoring, predictive maintenance, and fleet-level optimization. However, the quality of collected battery data is frequently compromised by multiple factors including sensor malfunctions, wireless communication failures, electromagnetic interference, and network packet loss [1]. These data quality issues manifest as outliers, noise pollution, and missing values, which severely degrade the performance of state estimation algorithms, fault diagnosis systems, and predictive models. Traditional data cleaning methods such as cluster analysis, support vector machines (SVM), and recurrent neural networks (RNN) suffer from high misdiagnosis rates and limited reconstruction accuracy [2]. The fundamental challenge lies in distinguishing genuine battery behavior anomalies from measurement artifacts while accurately restoring corrupted data without introducing systematic biases [3].

The proposed method presents a comprehensive intelligent data cleaning framework with the following key contributions 1) Development of an LSTM-based data quality assessment model that analyzes temporal features to accurately detect outliers, noise-polluted samples, and missing values with significantly reduced misdiagnosis compared to conventional methods. 2) Proposal of a feature fusion approach combining temporal features from LSTM networks and model-based features from electrochemical relationships to achieve superior data reconstruction accuracy. 3) Integration framework for end-edged cloud architecture enabling practical deployment in cloud-based vehicle battery management platforms.

Ⅱ. Method

Cloud-based BMS architectures adopt end-edge-cloud computing paradigms to enable scalable data processing and intelligent decision-making. The Cyber Hierarchy and Interactional Network (CHAIN) framework provides multi-scale insights by integrating real-time monitoring at the edge layer, intelligent processing at the fog layer, and advanced analytics at the cloud layer. The LSTM network processes battery time-series data to extract temporal features capturing sequential dependencies. Fig. 1 presents the flow chart of the proposed algorithm. For a data window X_t = [x_{t-w}, ..., x_t], the LSTM cell states evolve according to Eq. 1-5.

$$f(t)=\sigma\left(W_f\cdot\left[h(t),x(t)\right]+b_f\right) \quad (1)$$

$$i(t)=\sigma\left(W_i\cdot\left[h(t),x(t)\right]+b_i\right) \quad (2)$$

$$o(t)=\sigma\left(W_o\cdot\left[h(t),x(t)\right]+b_o\right) \quad (3)$$

$$C(t)=f(t)\circ C(t-1)+i(t)\odot\tanh\left(W_c\cdot\left[h(t-1),x(t)\right]+b_c\right) \quad (4)$$

$$h(t)=o(t)\odot\tan\left(C(t)\right) \quad (5)$$

Where *f(t)*, *i(t)*, *o(t)* are forget, input, and output gates, *C(t)* is the cell state, *h(t)* is the hidden state, σ is the sigmoid function, and ⊙ denotes element-wise multiplication. The outlier detection score combines three criteria in Eq. 6.

$$S_{outlier}\left(x(t)\right)=\alpha\left(S_{statistical}\right)+\beta\left(S_{temporal}\right)+\gamma\left(S_{consistency}\right) \quad (6)$$

Where α =0.4, β =0.3, ɣ =0.3 are optimized weights. The statistical score measures deviation of local distribution, temporal score evaluates consistency, and consistency scores checks physical plausibility.

Ⅲ. Results and discussion

The proposed data cleaning and restoration framework demonstrates robust performance for voltage and current in cloud datasets. Fig. 2, algorithm accurately detects missing data in both channels, achieving a 100% detection rate (59/59 samples), which confirms effectiveness of missing-data identification strategy. Although no random noise instances were present in evaluated dataset, with detection rates of 24.1% (19/79) for voltage and 49.4% (39/79) for current, indicating greater sensitivity to abrupt current anomalies under dynamic operating conditions. Data restoration results are illustrated in Fig. 3 and Fig. 4 for voltage and current, respectively. For voltage restoration algorithm achieves high accuracy, with missing data MAE of 0.0149V and MAPE of 0.46%, while noisy data restoration results in MAE of 0.1293V and MAPE of 4.07%. The restored voltage signals remain within physically realistic natural discharge profile, demonstrating reliable reconstruction performance. Current restoration (Fig. 4) is more challenging due to rapid signal fluctuations; however, framework reconstructs missing current data with an MAE of 0.3149A, indicating strong interpolation capability. MAPE becomes infinite because of near-zero reference values, making MAE an appropriate evaluation metric for current signals. Outlier restoration results further highlight adaptability to load variations. Fig. 5 presents error detected outliers, missing noisy data for voltage and current. Finally, Fig. 6 presents the frequency distribution of restoration errors, further validating proposed approach for cloud-connected BMS.

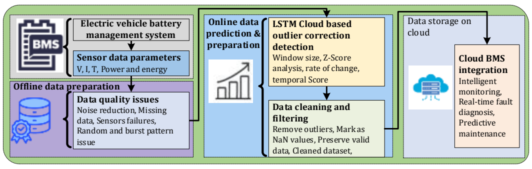

Fig. 1 Flow chart of the proposed method.

(a) (b)

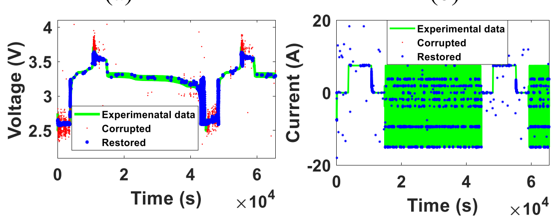

Fig. 2 (a)Voltage and (b) current original data in comparison with outliers' detections along with noise.

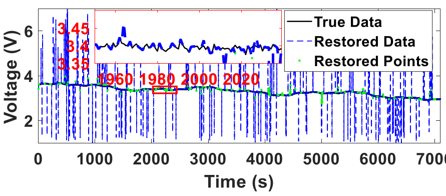

Fig. 3 Voltage restoration for missing data with LSTM

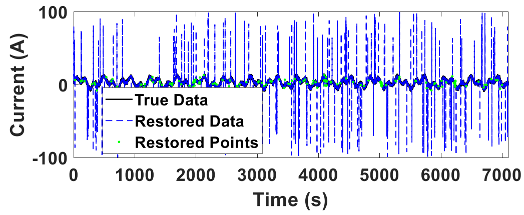

Fig. 4 Current restoration for the missing data through LSTM in cloud.

(a) (b)

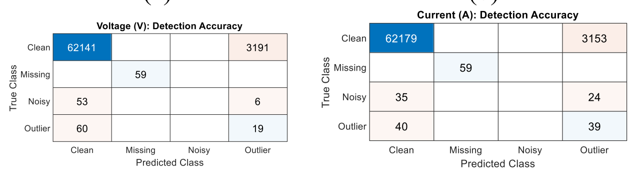

Fig. 5 Errors detected for outliers, missing and noisy data for a) voltage and b) current data.

(a) (b)

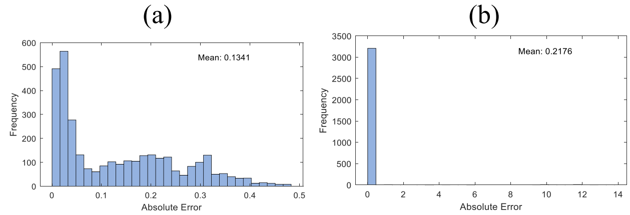

Fig. 6 Distribution of frequency error across different (a)voltage; (b) current samples.

IV. Conclusion

This paper presents an intelligent feedback-based continuous learning framework for cloud-based battery management systems, achieving 93.3% detection accuracy for noise-polluted data and achieving restoration accuracy exceeding 98.89%. The proposed approach enables reliable cloud data restoration for the BMS applications.

ACKNOWLEDGMENT

This research was supported by SNS (Development of LFP Battery SOC/SOH estimation algorithm) and the Korea Institute of Energy Technology Evaluation and Planning (KETEP) funded by the Korean government (Ministry of Trade, Industry and Energy) in 2024 (RS-2024-00398346, Training of technical workforce for O&M and asset management based on ESS big data).

REFERENCES

[1] Cavus, M., Dissanayake, D. and Bell, M., 2025. Next generation of electric vehicles: AI-driven approaches for predictive maintenance and battery management. Energies, 18(5), p.1041.

[2] Mulpuri, S.K., Sah, B. and Kumar, P., 2025. An intelligent battery management system (BMS) with end-edge-cloud connectivity– a perspective. Sustainable Energy & Fuels, 9(5), pp.1142-1159.

[3] Li, X., Wang, Q., Xu, C., Wu, Y. and Li, L., 2024. Survey of Lithium-Ion Battery Anomaly Detection Methods in Electric Vehicles. IEEE Transactions on Transportation Electrification.