A Hybrid Ensemble Architecture for Improved State of Health Predictions of Lithium-ion batteries using Stacked Generalization with Bagging and Boosting

M.J.M. Ashik Rasul, Jonghoon Kim
Energy Storage Conversion Laboratory, Chungnam National University

ABSTRACT

This paper proposes a novel hybrid ensemble architecture integrating bagging and boosting techniques via stacked generalization for robust lithium-ion battery state of health (SOH) predictions. The ensemble combines Random Forest (bagging) and Gradient Boost (boosting) as base learners, fusing their predictions using support vector regression as the meta-learner. By synergizing these diverse ensemble methods, the approach captures complex degradation patterns while alleviating individual model limitations. Evaluations on the NASA battery aging dataset demonstrate the hybrid ensemble's effectiveness, outperforming constituent ensembles and baseline regressions across error metrics. Four interpretable features are extracted from charge curve data, with relevance validated via correlation analysis. Leveraging support vector regression provides advantages in handling non-linearities and robust regression performance. Moreover, the interpretable meta-learner elucidates the relative importance of base models, explaining the ensemble's decision-making process. This architecture shows promise for reliable SOH estimation, enabling optimal battery management and utilization across diverse applications.

1. Introduction

Lithium-ion batteries are widely deployed in various applications, including electric vehicles and renewable energy systems. However, their performance degrades over time due to complex electrochemical processes, necessitating accurate state of health (SOH) estimation for optimal battery management and maintenance. Machine learning techniques have shown promise in addressing this challenge by leveraging historical battery data to model the intricate relationships between operating conditions and battery degradation[1].

Ensemble learning methods[2], which combine multiple base models, have demonstrated superior predictive performance compared to individual models. Bagging (e.g., Random Forest) builds diverse base models on different subsets of the data, reducing variance and overfitting. In contrast, boosting (e.g., Gradient Boost) sequentially trains weak models, focusing on instances misclassified by previous models, effectively reducing bias. While both bagging and boosting have been explored individually for SOH prediction, their integration within a hybrid ensemble architecture remains unexplored. This study proposes a novel approach that combines the strengths of these techniques through stacked generalization, where the predictions from Random Forest and Gradient Boost are fused using support vector regression as the meta-learner.

2. Methodology

The proposed hybrid ensemble approach as indicated in Fig. 1. combines the strengths of bagging and boosting techniques through stacked generalization for robust SOH predictions. The real-world battery dataset from the NASA Ames PCoE was utilized, containing comprehensive charge curve data for feature extraction. Four features were extracted: two statistical features (mean and standard deviation of voltage values) and two time-domain features (constant-current charging time and constant-voltage charging time). These features were validated using correlation coefficients to ensure relevance for SOH prediction.

The base models employed were Random Forest (RF) and Gradient Boost (GB), representing bagging and boosting techniques, respectively. RF introduces randomness through bootstrap sampling and feature subsampling, reducing overfitting. GB iteratively trains weak learners (decision trees) by focusing on previously misclassified instances, adaptively combining them into a strong predictor. The predictions from RF and GB models were used as input features for a support vector regression (SVR) meta-learner. This meta-learner learned the optimal combination of base model predictions fusing their outputs into a single robust prediction. Nested cross-validation was employed, with the outer loop for meta-learner training and the inner loop for the base model training, ensuring generalization and mitigating overfitting.

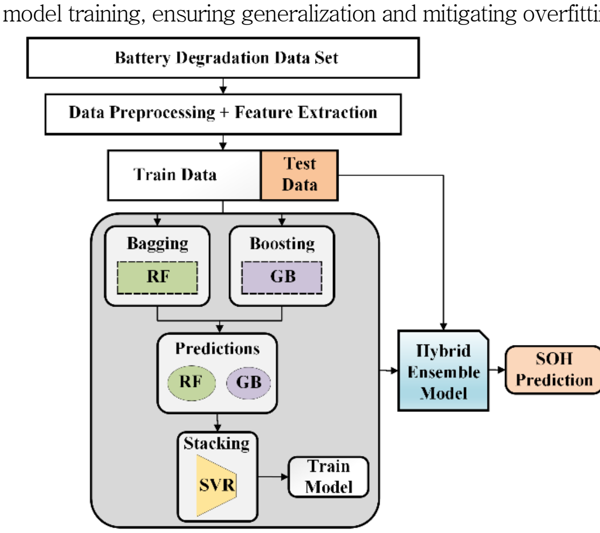

Fig.1 Framework of the proposed hybrid ensemble architecture

3. Results and Discussion

The proposed hybrid ensemble approach, combining RF and GB through stacked generalization, was evaluated using the NASA battery dataset. The battery capacity prediction was performed starting from cycle number 100 for the batteries in the dataset. Battery B0005 was employed as the training dataset for all models. The performance was assessed using three metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) value, calculated for capacity estimation on two representative batteries, B0006 and B0007, and summarized in Table 1.

Table 1 Battery capacity estimation accuracy

Cell	Model	Accuracy Metric		
		MAE	RMSE	R^2
B0006	RF	0.03183	0.04912	0.6143
	GB	0.02985	0.04914	0.6140
	Hybrid Model	0.01015	0.01182	0.9777
B0007	RF	0.03461	0.03511	0.4630
	GB	0.03214	0.03638	0.4236
	Hybrid Model	0.00324	0.00391	0.9933

For battery B0006, as depicted in Fig. 2, the hybrid ensemble model's predictions closely followed the actual capacity degradation trend, effectively capturing sharp drops and fluctuations observed in the real data. It outperformed the individual RF and GB models. Similarly, for battery B0007, as illustrated in Fig. 3, the hybrid ensemble model demonstrated superior performance in tracking the overall capacity degradation pattern compared to the individual models. The improved accuracy can be attributed to the synergistic combination of bagging and boosting techniques through stacked generalization.

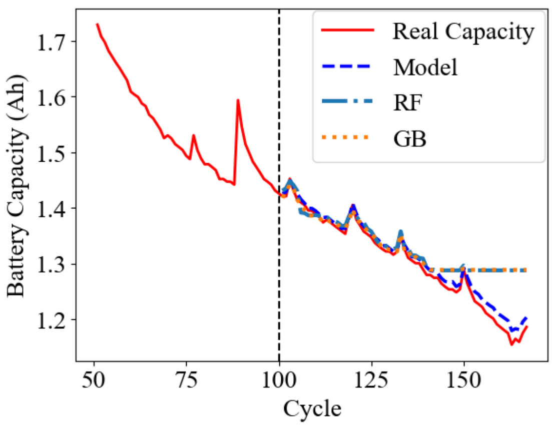

Fig.2 Battery capacity predictions - B0006

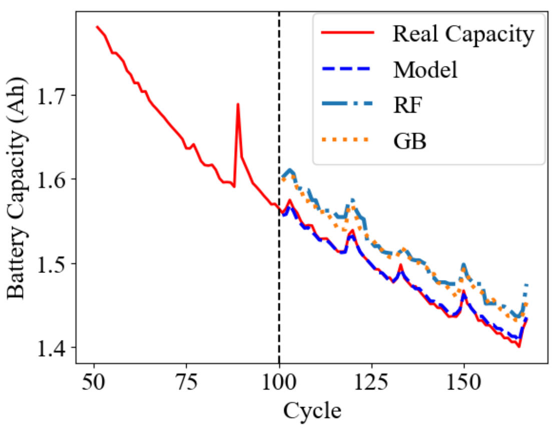

Fig.3 Battery capacity predictions - B0007

Employing SVR as the meta-learner offered key benefits, including handling non-linear relationships, high-dimensionality, and robust regression performance. The SVR model optimally combined base model predictions, leveraging their complementary strengths while mitigating weaknesses. Moreover, the interpretable nature of the SVR meta-learner provided insights into the relative importance of base models, elucidating the ensemble's decision-making process. While promising results were demonstrated on the NASA dataset, further evaluation across diverse battery datasets is necessary to validate the generalizability and robustness of the proposed method.

4. Conclusion

This study presented a novel hybrid ensemble architecture that synergistically integrates bagging and boosting techniques through stacked generalization for robust SOH predictions in lithium-ion batteries. By leveraging RF (bagging) and GB (boosting) as complimentary base models and fusing their predictions using SVR as the meta-learner, the proposed approach effectively captured diverse degradation patterns while mitigating individual model weaknesses. Evaluations on the NASA battery dataset demonstrated the hybrid ensemble model's superior predictive performance over individual baseline methods. The interpretable support vector regression meta-learner provided insights into the relative importance of base models, aiding ensemble decision comprehension. While showing promise on this dataset, further validation across diverse battery data is recommended to establish the method's generalizability and robustness for reliable SOH estimation and optimal battery management.

This research was supported by the Korea Evaluation Institute of Industrial Technology (No. 200116167, Development of Battery Safety Diagnosis System (BDS) SoC that predicts the internal state, explosion risk, remaining useful life, and replacement timing of electric vehicle batteries) and G2Power (Reflecting reuse battery diagnostic technology solution and UBESS design).

References

[1] X. Sui, S. He, S. B. Vilsen, J. Meng, R. Teodorescu, and D. - I. Stroe, “A review of non-probabilistic machine learning-based state of health estimation techniques for Lithium-ion battery,” Appl. Energy, vol. 300,2021

[2] X. Sui, S. He, and R. Teodorescu, “Small-sample-learning-based lithium-Ion Batteries Health Assessment: An optimized ensemble framework,” IEEE Transactions on Industry Applications, pp. 1–14, Jan. 2024.