Feature Valuation Toward Improved State Estimation for Automotive Lithium-ion Battery

Yanzhi Wang
Department of Industrial Engineering and Management
Peking University
Beijing, China
yanzhiwang@pku.edu.cn

Jianxiao Wang
National Engineering Laboratory for Big Data Analysis and Applications
Peking University
Beijing, China
wang-jx@pku.edu.cn

Jie Song
Department of Industrial Engineering and Management
Peking University
Beijing, China
songjie@coe.pku.edu.cn

2024 IEEE Industry Applications Society Annual Meeting (IAS) | 979-8-3503-7271-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/IAS55788.2024.11023788

***Abstract*—In the rapidly evolving fields of energy storage and big data, data-driven models for estimating battery states have become increasingly prevalent. However, the accuracy of these models is greatly affected by the quality of feature data used in training, especially in real-world lithium-ion battery scenarios where data diversity and quality significantly vary from laboratory settings. Our paper introduces a feature valuation framework tailored to data-driven predictions, focusing on evaluating the averaged marginal improvement of features on the model's performance. Experimental findings demonstrate a prediction error of 4.07% with only Battery and Power System, HVAC System, and Temperature Control and Monitoring features, yielding a 0.8% and 2.2% accuracy enhancement over the full dataset and BPS alone, respectively.**

***Index Terms*—electric vehicle, feature valuation, lithium-ion battery, shapley value, state estimation**

I. Introduction

The advent of electric vehicles and clean energy has precipitated a fundamental transformation in carbon emissions, marking a revolutionary shift in contemporary energy systems [1]. Lithium-ion batteries, as a vital energy storage technology, are undergoing rapid development due to the increasing popularity of electric vehicles [2]. Nevertheless, in recent years, the significance of managing battery states has been underscored by a gradual uptick in incidents pertaining to battery depletion and, more critically, battery safety [3]. Consequently, effective battery management hinges on the reliable identification of the battery's current state, facilitating interventions that are crucial for maintaining safety and efficiency. Among various factors, precise estimation of the State of Charge (SoC) of a lithium-ion battery is paramount for ensuring safe operation, prolonging its lifespan, and optimizing its overall performance [4].

However, predicting the SoC of batteries presents a significant challenge due to the highly nonlinear relationships among internal physical quantities and the limited observability of the battery's internal state. While external characteristics of the cell, such as voltage and current, yield important data, they do not consistently reflect the SoC. For example, the cell voltage primarily correlates with the lithium-ion concentration at the surface of the electrodes, while the SoC is contingent upon the average lithium-ion concentration within the electrodes [5]. Furthermore, the accuracy of SoC estimations is impacted by its complex responsiveness to environmental factors, particularly temperature variations [6].

Mechanism-based approaches to battery modeling predominantly encompass two main categories: physics-based electrochemical models and electrical equivalent circuit models. Electrochemical models, including the single-particle model and pseudo-two-dimensional model, derive battery state estimations through differential equations that delineate electrochemical reactions, ion diffusion, and electrolyte conductivity [7], [8]. Nonetheless, the precision of these models is critically contingent upon the accurate determination of internal parameters such as lithium-ion diffusion rates, electrode reaction kinetics, and electrolyte conductivity. Given that these parameters are frequently unknown and require optimization-based estimation, potential inaccuracies in parameter determination inherently constrain the fidelity of physics-based electrochemical models in practical scenarios [9]. Circuit models abstract the physical representation of battery internals using electrical components [10]. However, these approaches do not circumvent the necessity of estimating internal battery parameters like internal resistance and capacitance. Furthermore, the variability of these parameters over the battery's lifespan poses additional challenges to the predictability and accuracy of such models.

Given these challenges, employing data-driven methods for SoC estimation emerges as a potentially effective approach. In contrast, these approaches utilize comprehensive datasets from lithium-ion battery scenarios and employ machine learning algorithms, treating the battery as a 'black box' to identify intricate correlations and patterns among features, thereby markedly enhancing the precision of battery state predictions. Contrasted with mechanism-based approaches, data-driven methods excel in processing large, complex datasets, extracting meaningful insights without the need for detailed physical modelling, thereby offering efficiency in model development and updating. Utilizing a vast array of real-world electric

This work is supported by National Natural Science Foundation of China (No. 52277092), Chinese Association of Science and Technology Young Elite Scientists Sponsorship Program (No. YESS20210227) and Peking University Ordos Research Institute of Energy. *(Corresponding author: J. Wang)*

vehicle operating data, including SoC, ambient temperature, and driving data, researchers develop a lithium-ion battery charging capacity prediction model based on the tree-based algorithm [11]. The model's predictions, integrated with a fault determination mechanism, are effectively validated against the charging records of actual electric vehicles. Furthermore, researchers employed dropout techniques to build robust neural networks, enabling them to discern the laws of SoC prediction from various perspectives, encompassing recorded SoC and battery data, vehicle driving information, as well as environmental details [12].

Although data-driven methods excel in addressing nonlinear issues and internal parameters estimation challenges in lithium-ion batteries, their performance is contingent upon the quality of the training dataset [9]. Consequently, the challenge of effectively evaluating and selecting precise feature-based data to enhance accuracy of state estimation represents a burgeoning frontier in the realm of data-driven methodologies. Several scholars have evaluated data sample quality and introduced a valuation framework across diverse data scenarios [13]. The effectiveness of selecting high-quality data has been substantiated, illustrating that assigning greater weight to premium datasets considerably reduces prediction inaccuracies [14]. Nonetheless, these investigations overlook the crucial aspect of data feature selection for battery state prediction, an essential consideration in unraveling the predictive model's "black box" to identify critical predictive features. Researchers extract six statistical features from the voltage relaxation curve for predicting battery capacity and discover that the optimal estimation results are achieved by utilizing only three metrics—variance, skewness, and maximum as inputs [15]. This suggests that an increase in data features does not always lead to improved performance in assessment. However, the diversity and richness of data from real electric vehicles, including factors such as weather, traffic conditions, and driving behavior, exert both deterministic and random effects on the batteries, extending beyond the scope of laboratory data [12]. The significance of these features for lithium-ion batteries state prediction must be accurately evaluated to facilitate improvements in the data-driven methodology, thereby realizing enhanced economic and safety benefits.

Crucially, the data continuously collected by sensors in diverse electric vehicles constitutes significant assets. Yet, the research community has not established a comprehensive framework for effectively appraising the value of these data, especially regarding their contribution to the precise assessment of the actual state of lithium-ion batteries. Identifying and differentiating high-value from low-value features in this realm is still an area that remains uncharted. To address it, this study makes the following two contributions:

1) Focusing on the SoC prediction of lithium-ion batteries in electric vehicles, we propose a feature valuation framework that assesses both the impact of batteries and their associated environmental and on-vehicle data features based on data-driven method.

2) Fairness-based feature valuation is employed in a practical electric vehicle travel data, and the interpretability of evaluation hierarchy is rigorously affirmed through addition experiments.

The flowchart detailing the feature evaluation process for a specific electric vehicle is presented in Fig. 1. And steps *Feature Collection* and *Model Selection* are thoroughly detailed in Section II, whereas the *Feature Valuation process* is comprehensively addressed in Section III.

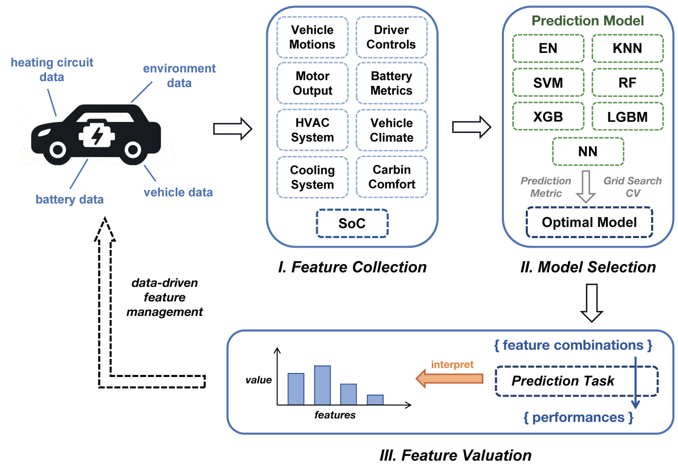

Fig. 1. Features Valuation based on SoC Prediction.

II. Data-driven State of Charge Prediction

The dataset employ in our study encompasses 72 real driving trips using a BMW i3 (60 Ah), comprising data on the environment, vehicle, battery, and heating circuit [16]. We transform the data from each trip into panel data at 10-second intervals.

A. Characterization of Input Features

For the purpose of SoC prediction, we identified 42 candidate features. It is important to acknowledge that our feature selection process is inherently constrained by the available dataset. Consequently, we have categorized these features according to their intrinsic characteristics to facilitate a more coherent understanding and to enhance the interpretability of the feature valuation process. These features were systematically classified into eight distinct categories based on the originating sensors, as detailed below:

- **Vehicle Motion Characteristics (VMC):** Velocity, Elevation, Longitudinal Acceleration.
- **Driving Inputn (DI):** Throttle, Regenerative Braking Signal.
- **Electric Motor Characteristics (EMC):** Motor Torque.
- **Battery and Power System (BPS):** Battery Voltage, Battery Current, Battery Temperature, max. Battery Temperature.
- **HVAC System (HVAC):** Heating Power CAN, Heating Power LIN, Requested Heating Power, AirCon Power, Heater Signal, Heater Voltage, Heater Current.

- **Temperature Control and Monitoring (TCM):** Ambient Temperature, Coolant Temperature Heatercore, Requested Coolant Temperature, Coolant Temperature Inlet, Heat Exchanger Temperature, Cabin Temperature Sensor, Ambient Temperature Sensor.
- **Cooling System (CS):** Coolant Volume Flow, Temperature Coolant Heater Inlet, Temperature Coolant Heater Outlet, Temperature Heat Exchanger Outlet.
- **Interior Environmental Temperature (IET):** Temperature Defrost lateral left, Temperature Defrost lateral right, Temperature Defrost central, Temperature Defrost central left, Temperature Defrost central right, Temperature Footweel Driver, Temperature Footweel Co-Driver, Temperature Feetvent Co-Driver, Temperature Feetvent Driver, Temperature Head Co-Driver, Temperature Head Driver, Temperature Vent central right, Temperature Vent central left, Temperature Vent right.

Our subsequent valuations of features in automotive lithium-ion batteries are predicated on this categorization.

B. Data-driven Model Candidates

Given that the primary contribution of our article is the introduction of a framework for feature valuation, we concentrate on the variations in SoC prediction resulting from the utilization of different features. Therefore, in the process of model selection, we opt for learning-based models that demonstrate strong generalization capabilities and have established maturity in application as our candidates. In the realm of data-driven battery condition assessment methods, eXtreme Gradient Boosting (XGB), Support Vector Machine (SVM), and Neural Networks (NN) have shown promising results [15], [17], [18]. In this study, we additionally utilize Elastic Net (EN), K-Nearest Neighbors (KNN), Random Forest (RF), and LightGBM (LGBM) as further candidate prediction models. These approaches is aimed at mining numerical laws from the large-scale and diverse datasets collected by electric vehicles, employing methodologies that span linear, unsupervised, integrated, and efficient learning perspectives.

As an amalgamation of Lasso and Ridge regression characteristics, EN is particularly apt for handling data with high-dimensional features and strong inter-feature correlations [19]. The methodology for its parameter β estimation is delineated as follows:

$$\underset{\beta}{\operatorname{argmin}}\left(\|y-X\beta\|^2+\lambda_2\|\beta\|^2+\lambda_1\|\beta\|_1\right) \tag{1}$$

The model incorporates the residual sum of squares, the $L2$ regularization term from Ridge regression (λ_2), and the $L1$ regularization term from Lasso (λ_1). These components collectively reduce over-fitting, with Lasso's term promoting sparsity in the regression coefficients. Besides, KNN predicts based on the proximity of data points to its nearest neighbors, effectively adapting to nonlinear battery-related data with complex patterns. In addition, RF enhances prediction accuracy and robustness by constructing multiple decision trees and aggregating their outcomes. This method is especially effective with large-scale, multivariate data, bolstering the model's resilience to outliers and noise [20]. XGB and LGBM optimizes its objective function $\mathcal{L}$ as gradient boosting tree:

$$\mathcal{L}^{(t)}=\sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)}+f_t\left(\mathbf{x}_i\right)\right)+\Omega\left(f_t\right) \tag{2}$$

which combines a loss function l and a regularization term Ω, enabling the model to enhance predictive accuracy while also controlling model complexity. LGBM stands out by its selective data focus and streamlined processing, enhancing efficiency in both training speed and memory usage for large-scale, high-dimensional datasets such as battery state prediction in electric vehicles [21]. Our NN model employs a multi-layer perceptron, which is particularly advantageous for SoC predictions due to its ability to capture complex, non-linear relationships, thereby enhancing accuracy and robustness..

In our experimental setup, diverse features are utilized as inputs with SoC as the output for model training. Mean Absolute Error (MAE) is selected to measure the accuracy of SoC prediction, and the expression is as follows:

$$\mathrm{MAE}=\frac{1}{n}\sum_{i=1}^{n}\left|y_i-\hat{y}_i\right| \tag{3}$$

We adopt a 5-fold cross-validation method to assess predictive accuracy and employ grid search for hyper-parameter optimization across models. Based on MAE, we evaluate and select the most suitable model for subsequent feature valuation, ensuring optimal predictive efficacy.

III. Feature Evaluation Framework

The variety of data-driven models used in assessing the state of lithium-ion battery vehicles presents challenges to existing model-based feature importance algorithms. For instance, while advanced boosting tree models like XGB and LGBM have proven technique on evaluating features based on their frequency in decision tree splits and their contribution to model performance, the most suitable model might belong to a different class of learning algorithms [21], [22]. In response to this, our study proposes a comprehensive and generalized feature valuation framework, designed to accommodate multiple data-driven model types by using precision-oriented method.

Let $X=\{X_1, X_2, \ldots, X_n\}$ represent the set of time series datasets collected from n sensors in an electric vehicle, where each X_i denotes a distinct feature (set) with corresponding data. Define the panel data domain P contains subsets of the feature set X. If considering all possible feature combinations, P can be defined as the power set of X, i.e., $P=\mathcal{P}(X)$. The value function ϕ is defined as $\phi: P \rightarrow \mathbb{R}$, mapping each feature subset (element of P) to a real number, indicative of the value or performance of that feature combination. In the SoC prediction, value is conceptualized as the performance of the featured dataset derived from cross-validation, where we employ the negative MAE as metric. As shown in III of Fig. 1, our key objective is to interpret this mapping relationship on ϕ, thereby elucidating the value of each feature set X_i in the prediction process.

A. Leave-One-Out Method

Leave-One-Out (LOO) refers to determining the contribution of a feature's presence or absence to the predictive performance of the remaining feature subset. This approach highlights the marginal utility of a quantitative feature across the entire set, as expressed below:

$$LOO\left(X_i\right) = \phi\left(X\right) - \phi(X \backslash \left\{X_i\right\}) \tag{4}$$

where $X \backslash \{X_i\}$ represents the complement of X_i in the set. This difference in prediction performance, denoted as $LOO\left(X_i\right)$, indicates the impact of feature X_i on the overall model performance. A significant feature value calculated by LOO suggests that X_i substantially influences the model's predictive accuracy trained by the entire set X.

B. Shapley Value Method

Upon noting that LOO assesses marginal improvement of the performance in data-driven model across the complete feature set, we aim to broaden this approach by including every feature subsets as a baseline. The concept of Shapley value, derived from game theory, is applied for benefit allocation, computing the weighted average utility of a player in relation to all subsets of its exclusion. When this concept is adapted for feature valuation, the formula is as follows:

$$\Psi\left(X_i\right) = \sum_{S \subseteq X \backslash \{X_i\}} \frac{|S|!(n-|S|-1)!}{n!} \left(\phi\left(S \cup \{X_i\}\right) - \phi(S)\right) \tag{5}$$

Here, $\Psi\left(X_i\right)$ is the Shapley value of feature X_i, S is a subset of features excluding X_i, n is the total number of features, and $\phi\left(S \cup \{X_i\}\right) - \phi(S)$ represents the incremental value of adding feature X_i to the subset S. Through the weighted summation of averaging coefficients, the average marginal enhancement attributable to a single feature across all subsets is ascertainable. Theoretically, the efficiency property of this method dictates that the aggregate of the values for all features should equal the total value derived from employing all features, minus the value attained when no features are used. That is expressed as:

$$\sum_{i=1}^{n} \Psi\left(X_i\right) = \phi(X) - \phi(\emptyset) \tag{6}$$

This ensures that the collective contributions of all features are deterministically quantified and equitably allocated among them, thereby demonstrating a level of referential integrity and fairness. Consequently, the valuation of features is entirely derived from the enhancements realized through the data-driven model. Therefore, the meaningfulness of comparing high and low values across different features is enhanced due to the limitations imposed on the total number of feature values, offering advantages in our feature valuation.

C. Insights for Revenue Sharing

In the realm of energy big data, synergies among various interrelated industries facilitate the ongoing integration of data elements, optimizing the modeling of energy systems. Similarly, in the context of electric vehicles, when data collection devices sourced from various vendors converge, they collectively contribute to a more precise estimation of the battery state. More precise predictions of the SoC enable better and more accurate battery capacity and health management, which in turn can significantly reduce the costs associated with power usage and battery maintenance. Should more precise predictions of battery state result in heightened deterministic gains, the sharing strategy of these gains can be effectively guided by feature valuation. If data for each feature set X_i is sourced from sensors supplied by different vendors i, then the revenue distribution for each vendor r_i can be expressed as follows:

$$r_i = \frac{w_i}{\sum_{j=i}^{n} w_j} * \left(G\left(X\right) - G\left(\emptyset\right)\right) \tag{7}$$

where G is the empirical function delineating the benefits derived from data in accurately predicting battery condition, and w_i is regarded as a coefficient associated with the feature value, taking into account realistic allocations.

IV. Results

A. Model Selection for SoC Prediction

We initially implement a grid search coupled with cross-validation to identify the model's optimal hyperparameters. For EN, the regularization scale is set between 0.001 to 10.0 with an $l1$ ratio of 0.2 to 0.8; the regularization coefficient for SVM ranges from 1 to 100; the KNN model's k-value is chosen between 3 to 7; for the boosting tree model (XGB & LGBM), the estimator and max depth are set within the ranges of 100 to 800 and 3 to 10, respectively. The NN is configured as a fully connected network with 2-4 hidden layers, each containing 100-150 nodes, and includes the option to select either the ReLU or tanh activation function. The best predictions MAE of the final optimized parameters are detailed in table I.

TABLE I
MAE across different data-driven models

Model	MAE (%)
EN	5.28
SVM	5.17
KNN	14.44
RF	7.39
XGB	5.70
LGBM	5.44
NN	**4.87**

The optimal model is neural network (NN), characterized by an activation function setting of ReLU and the hidden layer sizes of (150, 150, 150), achieved the lowest MAE among all models at 4.87%. Neural networks are frequently viewed as 'black boxes', largely attributed to their intricate, multi-level nonlinear operations and a multitude of internal parameters. This complexity hinders the comprehension of their capacity for precise SoC prediction through the analysis of extensive feature data. Such opacity further emphasizes the need for our forthcoming analysis on the importance of features.

B. Feature Valuation for better SoC Prediction

Building on the feature classification methodology outlined in II-A, we employ eight feature sets (n=8) as the unit for feature valuation. We use negative MAE as the cross-validation metric to calculate the predictive value in a feature-based sampling. Recognizing the inherent randomness in neural network training, we conducted the experiment eight times to bolster its reliability. The amalgamated outcomes of these iterations are depicted in Fig. 2, where they are differentiated based on the two methods employed: LOO and Shapley Value (SV).

Our analysis reveals that the feature value distribution calculated by SV is more polarized compared to LOO, as indicated by the larger absolute values at both extremes. For battery-related feature sets BPS, both methods recognize their predictive importance for SoC, with SV attributing a higher degree of importance as 10.3% than LOO as 8.2%. However, there are notable discrepancies in the value rankings for other features between the two methods. For instance, in the case of VMC, the feature set associated with vehicle driving, LOO assigns a marginally higher value (1.5%) than SV (0.6%); conversely, the importance of HVAC is significantly more pronounced in SV than LOO, ranking second only to BPS in SV. Other features such as DI, EMS, and CS show negligible absolute value ($\leq 0.2\%$) under both the LOO and SV computational frameworks, leading us to consider these sets collectively in subsequent analyses.

Given that both LOO and SV are predicated on marginal utility, the emergence of negative values for the two feature sets describing temperature (IET & TCM). Among these, IET, which prioritize the comfort of vehicle occupants by focusing on cabin temperatures, shows are consistently identified as potential disruptors (-0.3%) in accurately predicting SoC. Moreover, for features emphasizing the overall monitoring and control of operational temperatures, particularly engine cooling and ambient conditions, the TCM feature in SV exhibited a markedly negative impact on battery state prediction. This effect, as quantified by SV as -1.3%, was significantly more pronounced than the relatively minor impact observed in LOO as -0.1%.

Furthermore, our findings indicate that SV exhibits considerably less fluctuation than LOO in feature valuation, suggesting a lower susceptibility to the inherent randomness of the neural network training and testing. This robustness in SV stems from its averaging computation process, lending greater consistency and credibility to its results compared to LOO.

C. Validation through Feature Filtering

Based on the stability analysis of the results, we rank the features in descending order according to their Shapley values. Starting with an empty set, features are incrementally added in this order to calculate the predictive utility for the SoC. To minimize the impact of less influential feature sets on the regularity of the results, we grouped three feature sets with negligible values (DI/EMC/CS) into one category.

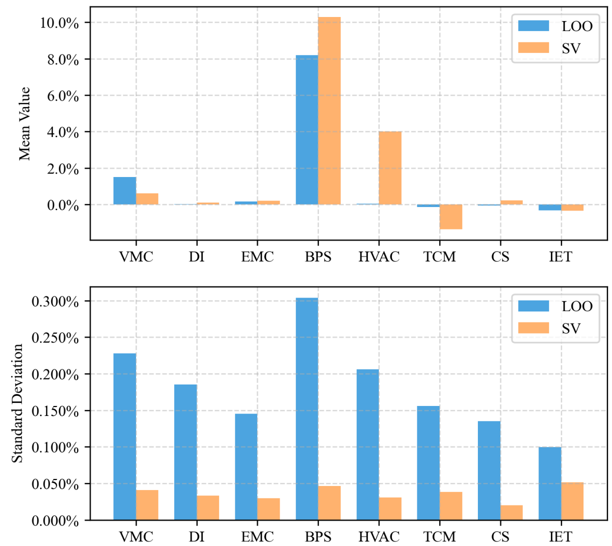

Fig. 2. The mean and standard deviations of different feature value using LOO and Shapley Value (SV).

Consequently, a total of six groups are utilized in the feature addition experiments, with the outcomes presented in Fig. 3.

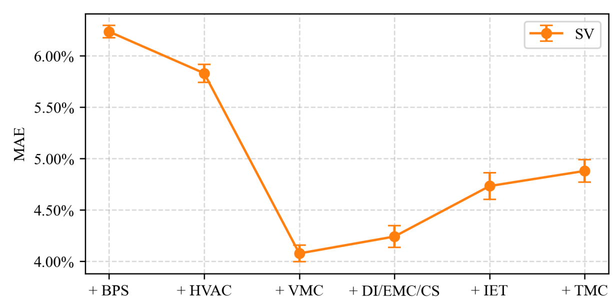

Fig. 3. Variations in MAE with the Addition of Feature Groups

We observe that employing only the battery-related feature set BPS limits the SoC prediction error to 6.24%. The accuracy further improves with the inclusion of HVAC and VMC features, dropping to a minimum error of 4.07% once both are added. However, integrating feature sets with minimal contributions led to the introduction of redundant features, resulting in a slight increase in prediction error, averaging at 4.23%. The addition of IET and TMC features, which hold negative feature values, adversely impacted the training and decision-making processes of the prediction model, ultimately elevating the error to that of the full data set's prediction performance, increase MAE in 0.8% compared to the lowest.

After arranging the LOO-computed feature values in descending order and identifying the point of optimal prediction performance, we compare the MAE at various stages: with no features added which using only the label mean, employing only battery-related BPS features, and with all features included. Our analysis reveals that SV (4.07%) outperforms LOO (4.22%) in identifying the optimal feature subset, in-

TABLE II
MAE IN DIFFERENT FEATURE VALUATION METHOD

Method	MAE (%)
no feature	18.61
only BPS	6.24
all features	4.88
LOO (lowest)	4.22
SV (lowest)	**4.07**

dicating that former hase a more accurate assessment of the value of features. In addition, a more efficient selection of the appropriate data subset for training can markedly enhance SoC prediction performance. As a result, in electric vehicle driving scenarios, incorporating data from the HVAC system and vehicle driving records alongside the Battery and Power System data yields the best SoC prediction, averaging a meaningful 2.2% improvement in accuracy over using only battery data. This strategic feature management — the more precise utilization of feature data — resulted in a 0.80% increase in predictive utility.

V. CONCLUSION

In this study, we have delineated the features pertinent to predicting the State of Charge (SoC) of lithium-ion batteries in electric vehicles. Taking into account data collected from 8 major categories of electric vehicle features from actual driving trips, we observe that the Shapley Value (SV) calculation yielded superior differentiation and stability in feature values than Leave-One-Out (LOO). The significance of the battery-related feature BPS is consistently recognized. Other features such as HVAC is accorded higher importance in SV, whereas the LOO calculation indicated a greater contribution from the vehicle power feature VMC. Additionally, both methods concurr on the negligible contribution of DI, EMC, and CS features. Concerning the negative impact of temperature-related features, SV more prominently underscored the TCM's potential to disrupt accurate SoC prediction, compared to the similar evaluations of the IET feature in both SV and LOO.

Through experiments involving the addition of features based on value-based ordering, we discover that the SV approach achieves the lowest Mean Absolute Error (MAE) in feature management than LOO. Specifically, a prediction error of 4.07% was attained using only the BPS, HVAC, and TMC features, representing a 0.8% and 2.2% improvement in prediction accuracy compared to using the full dataset and employing only the BPS feature, respectively. This highlights the significance of feature management in estimating the state of lithium-ion batteries, and offers crucial insights into the roles of HVAC and vehicle driving state in SoC evaluation, which also sheds light on the negative impact of temperature variables on the training of data-driven models.

Certain research areas for further exploration are: 1. Deeper analysis of the interplay between feature values in the context of battery principles; 2. Adjustment of the Shapley value for diverse data scenarios, including scenarios with high feature correlation or intricate model training.

ACKNOWLEDGMENT

REFERENCES

[1] J. Ma, H. Kong, J. Wang, H. Zhong, B. Li, J. Song, and D. M. Kammen, "Carbon-neutral pathway to mitigating transport-power grid cross-sector effects," *The Innovation*, p. 100611, 2024.
[2] A. Manthiram, "A reflection on lithium-ion battery cathode chemistry," *Nature communications*, vol. 11, no. 1, p. 1550, 2020.
[3] X. Feng, M. Ouyang, X. Liu, L. Lu, Y. Xia, and X. He, "Thermal runaway mechanism of lithium ion battery for electric vehicles: A review," *Energy storage materials*, vol. 10, pp. 246–267, 2018.
[4] Y. Wang, Z. Chen, and C. Zhang, "On-line remaining energy prediction: A case study in embedded battery management system," *Applied Energy*, vol. 194, pp. 688–695, 2017.
[5] G. L. Plett, *Battery management systems, Volume II: Equivalent-circuit methods*. Artech House, 2015.
[6] X. Hu, D. Cao, and B. Egardt, "Condition monitoring in advanced battery management systems: Moving horizon estimation using a reduced electrochemical model," *IEEE/ASME Transactions on Mechatronics*, vol. 23, no. 1, pp. 167–178, 2017.
[7] J. Li, K. Adewuyi, N. Lotfi, R. G. Landers, and J. Park, "A single particle model with chemical/mechanical degradation physics for lithium ion battery state of health (soh) estimation," *Applied energy*, vol. 212, pp. 1178–1190, 2018.
[8] A. Jokar, B. Rajabloo, M. Désilets, and M. Lacroix, "Review of simplified pseudo-two-dimensional models of lithium-ion batteries," *Journal of Power Sources*, vol. 327, pp. 44–55, 2016.
[9] Y. Wang, J. Tian, Z. Sun, L. Wang, R. Xu, M. Li, and Z. Chen, "A comprehensive review of battery modeling and state estimation approaches for advanced battery management systems," *Renewable and Sustainable Energy Reviews*, vol. 131, p. 110015, 2020.
[10] S. Nejad, D. Gladwin, and D. Stone, "A systematic review of lumped-parameter equivalent circuit models for real-time estimation of lithium-ion battery states," *Journal of Power Sources*, vol. 316, pp. 183–196, 2016.
[11] Z. Wang, C. Song, L. Zhang, Y. Zhao, P. Liu, and D. G. Dorrell, "A data-driven method for battery charging capacity abnormality diagnosis in electric vehicle applications," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 990–999, 2021.
[12] R. Li, H. Wang, H. Dai, J. Hong, G. Tong, and X. Chen, "Accurate state of charge prediction for real-world battery systems using a novel dual-dropout-based neural network," *Energy*, vol. 250, p. 123853, 2022.
[13] Y. Wang, J. Wang, F. Gao, and J. Song, "Unveiling value patterns via deep reinforcement learning in heterogeneous data analytics," *Patterns*, vol. 5, no. 5, 2024.
[14] Y. Wang and J. Song, "Dissecting renewable uncertainty via deconstructive analysis-based data valuation," *IEEE Transactions on Industry Applications*, 2024.
[15] J. Zhu, Y. Wang, Y. Huang, R. Bhushan Gopaluni, Y. Cao, M. Heere, M. J. Mühlbauer, L. Mereacre, H. Dai, X. Liu *et al.*, "Data-driven capacity estimation of commercial lithium-ion batteries from voltage relaxation," *Nature communications*, vol. 13, no. 1, p. 2261, 2022.
[16] M. Steinstraeter, J. Buberger, and D. Trifonov, "Battery and heating data in real driving cycles," 2020. [Online]. Available: https://dx.doi.org/10.21227/6jr9-5235
[17] J. Wei, G. Dong, and Z. Chen, "Remaining useful life prediction and state of health diagnosis for lithium-ion batteries using particle filter and support vector regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 7, pp. 5634–5643, 2017.
[18] T. Weigert, Q. Tian, and K. Lian, "State-of-charge prediction of batteries and battery–supercapacitor hybrids using artificial neural networks," *Journal of Power Sources*, vol. 196, no. 8, pp. 4061–4066, 2011.
[19] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 67, no. 2, pp. 301–320, 2005.
[20] T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer, 2009, vol. 2.
[21] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.
[22] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.