

Proposal of Evaluation Metric for Machine Learning Model for Battery Voltage Estimation

Takuma KAWAHARA¹Yuki SATO²Koji SATO²

ABSTRACT

In order to evaluate a machine learning model that estimates the voltage of a lithium-ion battery, we built a new accuracy evaluation metric based on the Arrhenius law, and tested it to see if it could be used to judge the relative merits of machine learning models. Two types of machine learning model were selected. One was a type that performed discrete estimates in regions where data are not available, while the other was capable of performing continuous estimation. Then, discrete temperature data based on floating car data collected from electric vehicles were generated, and the models were trained and evaluated. The results confirmed that the commonly used evaluation metric for evaluating deviation from test data showed no clear differences in evaluation results between the two model types. On the other hand, the authors' new evaluation metric based on the Arrhenius law showed clearer differences in the evaluations of the two models. Implementing this metric made it possible to evaluate a machine learning model even with limited collected data, thereby advancing the practical application of battery voltage estimation technology using machine learning models.

1. Introduction

Aiming to achieve carbon neutrality by 2050, Honda is developing battery electric vehicles (hereinafter, BEV) and fuel cell vehicles. Lithium-ion batteries (hereinafter, batteries) are widely used for energy storage in BEV⁽¹⁾. It is known that normally, the overpotential of a battery (the difference between closed and open circuit voltages) changes according to usage environment factors, such as temperature and energy demand from the vehicle, and the amount of energy that can be output varies^{(2), (3)}. Being able to estimate with high precision the overpotential of a battery under given conditions is believed to enable highly accurate estimates of the BEV cruising range in all environments. This in turn will reduce the risk of power depletion and helps eliminate customer anxiety about running out of power.

In recent years, it has become possible to collect large volumes of real-world data (floating car data) from running vehicles, and research is being actively pursued on battery voltage estimation technology that utilizes battery big data and machine learning⁽⁴⁾⁻⁽⁶⁾. As an example of prior research, one study reported research in which an algorithm was created for estimating voltages using deep learning, based on battery time-series data collected from electric buses⁽⁴⁾.

Shuangqi Li et al. tested voltage estimation accuracy with test data created based on the time-series data. The result was that the voltage could be estimated with a mean absolute percentage error (MAPE) of 2.42%. Another example of prior research is a study in which a battery charge/discharge cycle test was performed using a charge/discharge test device, and an algorithm was created for estimating battery voltages and temperatures with machine learning based on the collected data⁽⁵⁾. Manh-Kien Tran et al. used 30% of the data as test data and tested the accuracy. They reported as a result that the decision tree model was able to estimate voltage and temperature with a 0.99 coefficient of determination. The authors are focusing on machine learning as a technology that can reproduce with high accuracy the non-linear voltage behavior of batteries affected by multiple factors.

However, to actually apply this technology, it is necessary to fully determine how batteries are used in the market. For example, battery usage conditions differ depending on power unit types, such as BEV, hybrids, and plug-in hybrids, and vehicle properties, such as those of city commuters or long-distance vehicles. An inability to acquire sufficient training data is also a consideration. Due to such differences in battery usage conditions, it is believed that estimation accuracy will differ depending on

¹ Innovative Research Excellence Power Unit & Energy

E-mail: takuma_kawahara@jp.honda

² TechnoPro, Inc. TechnoPro Design Company

the type of machine learning algorithm⁽⁷⁾. For this purpose, it is necessary to select an appropriate machine learning method corresponding to the battery usage data, which differ depending on the vehicle characteristics and market.

In this study, a method is tried that uses electrochemical knowledge for assessing the machine learning models. The paper proposes a method for assessing battery voltage estimation machine learning models, using the Arrhenius law, a commonly known law about chemical reactions.

2. Issues and Approach

2.1. Data Related Issues

Unlike physical models that calculate estimated values deductively, machine learning modelling learns data tendencies inductively based on collected data, and then calculates estimated values. Therefore, estimation in the absence of collected data may become less accurate.

As one example, Fig. 1 shows the results when comparing how machine learning models perform estimates in regions where data are not available. The following four types of machine learning models are being compared. The explanatory variable is x and the target variable is y .

- Random Forest Regressor (RF)
- K-Nearest Neighbors Regressor (KNN)
- Gaussian Process Regressor -RBF kernel (GP)
- Support Vector Regressor -RBF kernel (SVR)

Here, the RF and KNN models perform discrete estimates in regions where data are not available, while the GP and SVR models are capable of performing continuous estimation.

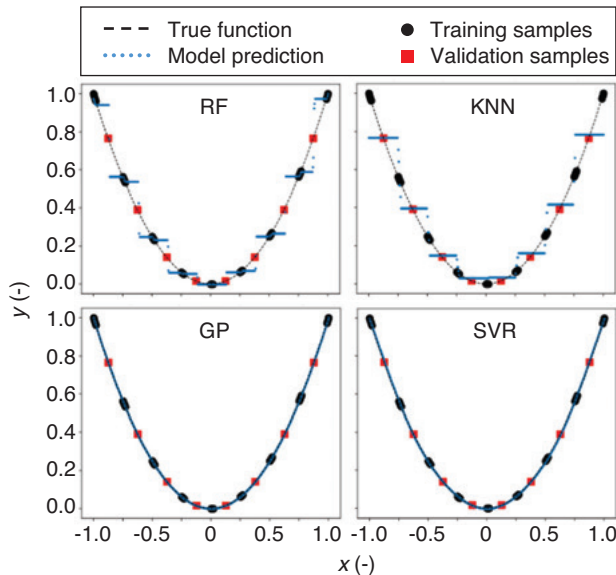


Fig. 1 Comparison of estimation results of machine learning models when data are given discretely

The dashed line in Fig. 1 shows the true trend $y=x^2$ while the dotted line shows the machine learning model's estimated values. The circles are the training data; the squares show the validation data for hyperparameter tuning. Both are discretely sampled from the true trend. The results showed that for both machine learning models, when data were available, estimated values were close to the true trend. On the other hand, in conditions where learning and validation data were unavailable, it was confirmed that estimates produced by the RF and KNN models deviated from the true trend. In the RF model, node branches along the decision tree are generated with reference to the training data to learn the data tendency. Therefore, it is a property of the model that when given discrete training data, the estimated values are discrete. The KNN model is an algorithm that calculates estimated values based on neighboring data within the training data. Therefore, like RF, it is a property of the model that when given discrete training data, the estimated values are discrete. It follows that in cases where the true trend changes continuously, the RF and KNN models tend to deviate from the true trend where data are unavailable. With the GP and SVR models, it was possible to confirm that even within ranges where validation data are unavailable, estimations do not deviate from the true trend. It is believed that this is because the property of GP and SVR models that estimates values continuously is effective in learning continuous true trends. It follows from this result that, in order to test whether a machine learning model can learn a true trend, validation with discrete data alone is insufficient, and comprehensive validation after continuously changing the explanatory variables is required.

2.2. Approach

As described in the foregoing section, in order to apply battery voltage estimation technology using machine learning models, it is necessary to address the issue of continuously and comprehensively assessing the machine learning model.

As an answer to this issue, one possible approach is to comprehensively collect floating car data in all conditions and perform comprehensive machine learning model validation. However, the conditions that can be obtained from the floating car data depend greatly on the vehicle usage environment. For example, temperature data depends on the climate in which the vehicle is driven, and there is no guarantee that continuous and comprehensive data can be obtained encompassing low to high temperatures. Therefore, it is possible that, to collect the necessary and sufficient data, the data would need to be gathered from a large number of vehicles over a long period of time. However, from a practical point of view this is not feasible, so another validation approach is needed.

To address the above issue, in this paper the authors propose a new evaluation metric drawing on electrochemical knowledge. With this evaluation metric,

continuous and comprehensive data are generated virtually and input into the machine learning model, and the response to the input data of the outputted values is evaluated. When the input data are generated virtually, there are no true values corresponding to the model's output values, so it is not readily possible to quantify the deviation between the output values and true values. However, the values output in response to changes in the input data can be expressed as a trend using an electrochemical theoretical formula. In the proposed approach, a quantitative evaluation metric is calculated by comparing the output value response of the model to changes in the input data and the theoretical formula.

In this study, discrete temperature data was generated from collected floating car data, then four types of machine learning models (RF, KNN, GP, and SVR) were trained and evaluated. Then, tests were conducted to see if it was possible to determine differences between RF and KNN models that perform discrete estimations and GP and SVR that can perform continuous estimations, when the training data are provided discretely.

3. Validation Steps

Figure 2 summarizes the validation steps in this study. The overall analysis consists of two steps, namely, Step 1: Model Training and Step 2: Model Evaluation. In Step 1, the data required for training the machine learning models are generated based on battery floating car data collected

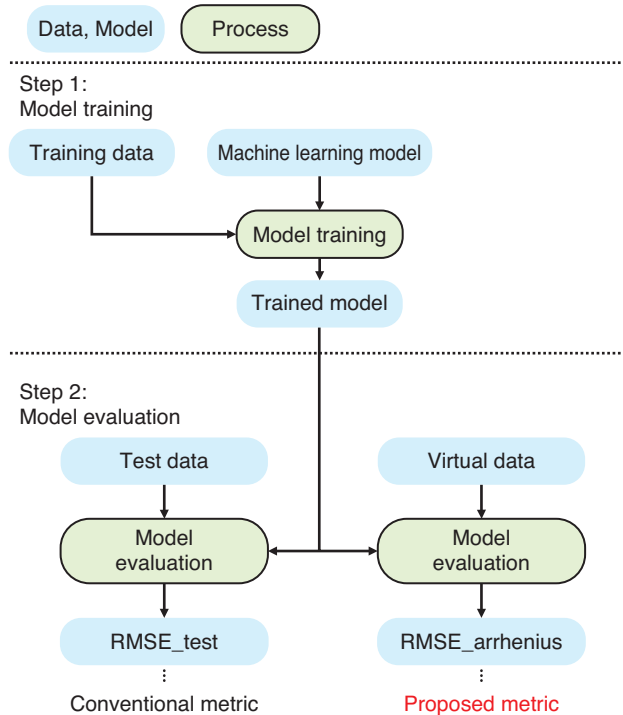


Fig. 2 Flowchart of entire evaluation in this study

from vehicles. Then, the four machine learning models (RF, KNN, GP, and SVR) are trained. In Step 2, the accuracy of the trained machine learning models is quantified with both standard evaluation metrics and the proposed new evaluation metric.

In this paper, Chapter 4 gives the details of Step 1. Chapter 5 provides details about Step 2 and the two evaluation metrics. Chapter 6 describes the validation results and discusses the differences between the two evaluation metrics. Finally, Chapter 7 is the conclusion of this study.

4. Model Training

Step 1 consists of the following processes.

- (1) Generating the training data
- (2) Training the machine learning models

This chapter describes each process in detail.

4.1. Generating the Training Data

In this study, battery floating car data (21,765,370 records) from plug-in hybrid vehicles were used to generate training data. Data were collected from vehicles that were being operated within 30 days of their first use, with little change in battery performance from initial levels. Collected data categories were battery overpotential, temperature, current, and open circuit voltage. The purpose of this study is to find an evaluation metric that can appropriately judge the relative merits of machine learning models trained with discrete data. Therefore, discrete data were generated based on the above data and used for training the models. In this study, discretization was performed with respect to temperature. At each of the following temperatures, 500 data records were randomly extracted within a ± 0.5 deg. C range: -20, -10, 0, 10, 20, 30, 40 (all deg. C). The extracted data temperature distribution is as shown in Fig. 3.

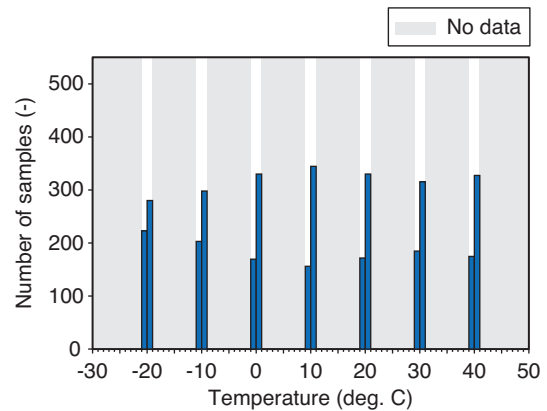


Fig. 3 Frequency histogram for temperature of training data

Table 1 Optimized hyperparameters of machine learning models

ML model	Hyperparameters
RF	n_estimators: 10; max_depth: 50
KNN	n_neighbors: 6; weights: uniform
GP	kernel function: RBF length_scale: 200; length_scale_bounds: fixed
SVR	kernel function: RBF epsilon: 0.01

4.2. Training the Machine Learning Models

The configuration of the battery overpotential estimation model is shown in Fig. 4. Battery overpotential is estimated using data gathered from vehicles and a machine learning model. The three explanatory variables of the machine learning model are the open circuit voltage, current, and temperature. The target and explanatory variables are both time-synchronized data. Four types of machine learning models were selected for use from the available supervised regression models. RF and KNN were used as the models that perform discrete estimates in regions where data are not available. GP and SVR were used as the models that are capable of continuous estimation even in regions where data are not available. The hyperparameters of the machine learning models were determined after optimization using the Hold-out Method on the training data. The optimized parameters are shown in Table 1. For other hyperparameters not shown in Table 1, the default scikit-learn 1.0.1 parameters were used⁽⁸⁾.

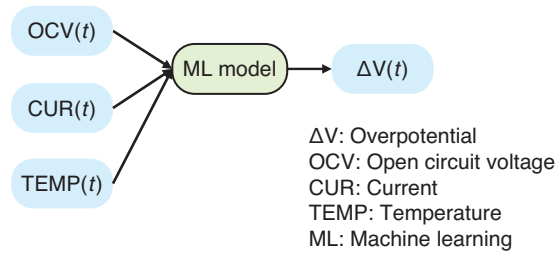


Fig. 4 Structure of machine learning model

5. Model Evaluation

In Step 2, in order to test the difference between the commonly used evaluation metric and the proposed new evaluation metric, the trained models' accuracy on the following two metrics was evaluated.

- (1) Accuracy with respect to test data
- (2) Accuracy with respect to the theoretical formula that describes overpotential

This chapter describes each evaluation metric in detail.

5.1. Accuracy with Respect to Test Data

Accuracy with respect to test data was quantified using the process shown in Fig. 5. This evaluation metric is commonly used for evaluating the accuracy of machine learning models. First, the test data explanatory variables are input into the trained model and estimated values are obtained. Then, the estimated values and test data target variables were compared, and the estimation deviation calculated. The estimation deviation was quantified with the root mean square error (RMSE) shown in Eq. (1) below.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - y_i)^2} \quad (1)$$

Here, n is the total number of records, f is the estimated value of the machine learning model, and y is the test data target variable. This evaluation metric is shown below as "RMSE_test."

As with the training data, the test data were extracted from the floating car data. At each of the following temperatures, 500 data records were randomly extracted within a ± 0.5 deg. C range: -20, -10, 0, 10, 20, 30, 40 (all deg. C) (Fig. 6). However, they were extracted without any repetitions of the data extracted for training data.

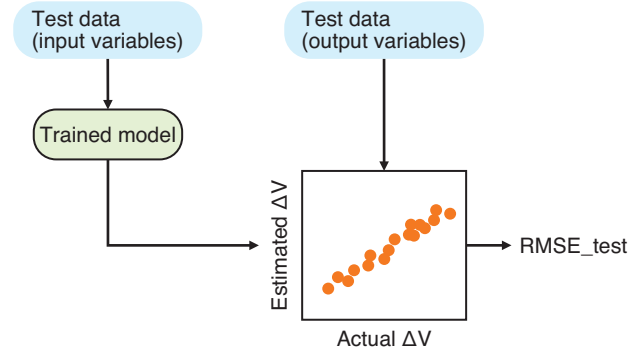


Fig. 5 Flow chart of evaluation of RMSE_test

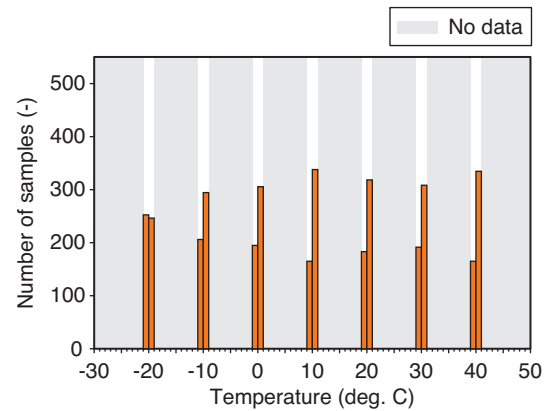


Fig. 6 Frequency histogram for temperature of test data

5.2. Accuracy with Respect to the Theoretical Formula that Describes Overpotential

Accuracy with respect to the theoretical formula that describes overpotential was quantified using the process shown in Fig. 7. This evaluation metric is the proposed new evaluation metric, the aim of which is to continuously and comprehensively evaluate machine learning models. This study targeted the temperature dependence of overpotential, which is relatively easy to determine with a theoretical formula.

For the evaluation, after continuously and comprehensively generating explanatory variables data, the generated data were input into the trained model and the estimated values were obtained. Next, the inputted virtual temperature values were plotted along the horizontal axis, and the model's estimated values along the vertical axis. With the results, fitting was performed with a function derived from the Arrhenius equation, and deviation quantified by RMSE. This evaluation metric is shown below as "RMSE_arrhenius."

The virtual data were generated by continuously changing the conditions within the expected range of usage conditions of the battery. In concrete terms, data were generated under the following patterns (671 patterns at each temperature).

- Current: 0.5 to 1.5 C, at 0.1 C intervals (11 patterns)
- Open circuit voltage: 3.5 to 4.1 V, at 0.01 V intervals (61 patterns)

Temperature conditions were generated under the following patterns.

- Temperature: -20 to 40 deg. C, at 1 deg. C intervals (61 patterns)

The generated data temperature distribution is shown in Fig. 8. A comparison with the test data temperature distribution of Fig. 6 confirms that the data are continuous and comprehensive. The machine learning model's average estimated value for all patterns at each temperature was

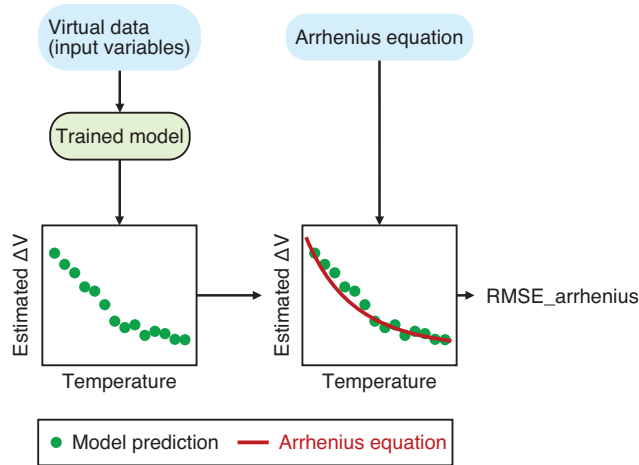


Fig. 7 Flow chart of evaluation of RMSE_arrhenius

used as the representative value for each temperature.

The Arrhenius equation was used as the theoretical equation for evaluating the temperature dependence of the machine learning model estimated values. It is known that within the internal resistances of batteries, the temperature dependence of the resistance ensuing from interfacial charge transfer resistance and slowing of ion diffusion speed follows the Arrhenius equation⁽⁹⁾. On the other hand, the electrical resistance of the current flowing through the collector and conductive material in the electrodes does not follow the Arrhenius equation. Taking the temperature dependence of these resistances into account, the temperature dependence of overpotential is considered to follow Eq. (2) below, which adds a constant term to the Arrhenius equation.

$$\Delta V = A \exp\left(\frac{-E_a}{RT}\right) + \Delta V_0 \quad (2)$$

Here, ΔV is the overpotential, R is the gas constant, T is the temperature, E_a is the activation energy, A is a temperature-independent coefficient, ΔV_0 is a constant that expresses overpotential that does not follow the Arrhenius equation. The optimization parameters when fitting the temperature dependent estimated values of the machine learning model with Eq. (2) are A , E_a , and ΔV_0 .

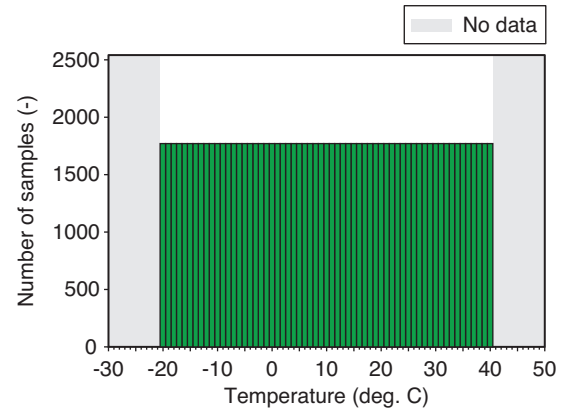


Fig. 8 Frequency histogram for temperature of generated data

6. Results and Observations

Figure 9 shows the evaluation results of the four machine learning models with the two types of evaluation metric used to determine if it is possible to judge the relative merits of the RF and KNN models, which perform discrete estimates in regions where data are not available, and GP and SVR models, which perform continuous estimation even in regions where data are not available. In generating training data and test data, a random pattern (random

seed) was repeated 32 times for training and validation. Also, the average of 32 RMSE repeats are arranged and shown for each machine learning model in Table 2. From these results, it was confirmed in the RMSE_test that the difference between the RF and KNN variance (average of 20.8 mV) and GP and SVR variance (average of 20.4 mV) was too small to enable judging of relative merits. On the other hand, with the proposed new metric, RMSE_arrhenius, the RF and KNN variance (average of 12.3 mV) was calculated at 10 times that of the GP and SVR variance (average of 1.2 mV), confirming that the difference between the two could be judged clearly.

Figure 10 shows the results when comparing the machine learning model predicted values with the test data actual values. All the results conform to a single random pattern. Based on the results, no characteristic tendency could be confirmed in any model in the test data estimation, and so it was judged that there was no difference between each model. It is thought that the reason for this is that because the temperature range of the test data is the same as the temperature range of the trained model, all the models are appropriately learning the trends of the data.

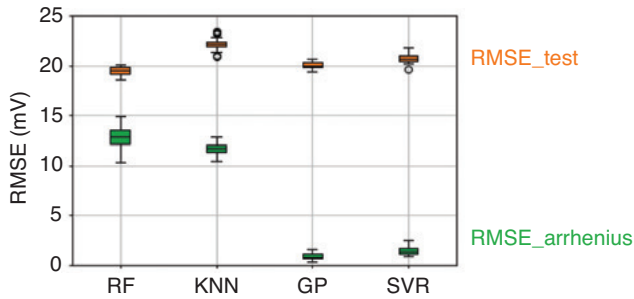


Fig. 9 Evaluation results of machine learning models

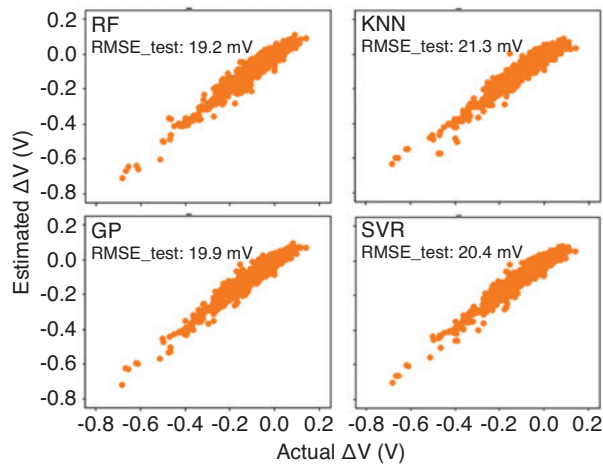


Fig. 10 Comparison of estimations of machine learning models with actual values of test data

Table 2 Evaluation results of machine learning models

ML model	RMSE_test	RMSE_arrhenius
RF	Average: 20.8 Standard deviation: 1.4	Average: 12.3 Standard deviation: 1.1
KNN		
GP	Average: 20.4 Standard deviation: 0.5	Average: 1.2 Standard deviation: 0.5
SVR		

(Unit: mV)

Figure 11 shows the results of comparing the machine learning models' estimated temperature dependent values with Eq. (2). Here, the dots show the estimation results obtained from virtual data input into trained models, shown as average values per 1 deg. C. The solid line is the fitting of the above average values, from Eq. (2). The result confirms that with the RF and KNN models, the estimated values changed discontinuously with respect to temperature changes compared to the GP and SVR models, and in relative terms, clearly deviate from the theoretical values predicted by Eq. (2). On the other hand, with the GP and SVR models, it was confirmed that the estimated values changed continuously with respect to temperature changes, and closely tracked Eq. (2) in temperature ranges from low to high.

From this result, it was confirmed that with the conventional evaluation metric (RMSE_test), it was not possible to judge the relative levels of accuracy in regions where data are not available, but with the proposed new metric (RMSE_arrhenius), it was possible to clearly judge the relative merits.

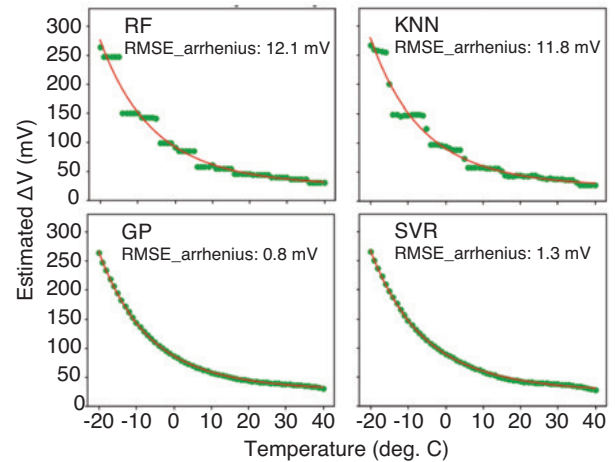


Fig. 11 Temperature dependence of machine learning model estimations compared to curves predicted from Arrhenius equation

7. Conclusion

In this paper, the authors have proposed a new evaluation metric based on an electrochemical theoretical formula (Arrhenius equation) for the evaluation of the battery voltage estimation models that use machine learning. In the study, discrete temperature data were generated from data collected from vehicles, then used to train and evaluate four types of machine learning models (RF, KNN, GP, and SVR). The result shows that with a conventional evaluation metric (RMSE targeting test data), it was not possible to clearly judge differences between RF and KNN models and GP and SVR models, but with the proposed new metric, it was possible to clearly judge the differences.

When using conventional evaluation metrics to evaluate a machine learning model, it is necessary to collect data from a large number of vehicles over a long time period, which poses issues in terms of practicality. However, the proposed evaluation metric makes it possible to evaluate a machine learning model even with limited collected data, and so this is believed to advance the practical application of battery voltage estimation technology using machine learning models. In this study, the metric was used to evaluate the temperature dependence of overpotential, which is easy to determine with a theoretical formula. However, if it possible to define a governing theoretical formula that can express the relevant trend, other variables (current, charge rate, etc.) are believed to be capable of use for the evaluation. In future studies, there are plans to look at dependence evaluations based on other variables, in addition to temperature dependence of overpotential.

References

- (1) Chian, T. Y., Wei, W. L. J., Ze, E. L. M., Ren, L. Z., Ping, Y. E., Abu Bakar, N. Z., Faizal, M., Sivakumar, S.: A Review on Recent Progress of Batteries for Electric Vehicles, *International Journal of Applied Engineering Research*, Vol. 14, No. 24, p. 4441-4461, (2019)
- (2) Vidal, C., Gross, O., Gu, R., Kollmeyer, P., Emadi, A.: xEV Li-Ion Battery Low-Temperature Effects—Review, *IEEE Transactions on Vehicular Technology*, Vol. 68, No. 5, p. 4560-4572, (2019)
- (3) Meyer, N., Whittal, I., Christenson, M., Loisel-Lapointe, A.: The Impact of Driving Cycle and Climate on Electrical Consumption & Range of Fully Electric Passenger Vehicles, *EVS26 - International Battery, Hybrid and Fuel Cell Electric Vehicle Symposium*, (2012)
- (4) Li, S., Li, J., He, H., Wang, H.: Lithium-ion battery modeling based on Big Data, *Energy Procedia*, Vol. 159, p. 168–173, (2019)
- (5) Tran, M.-K., Panchal, S., Chauhan, V., Brahmabhatt, N., Mevawalla, A., Fraser, R., Fowler, M.: Python-Based Scikit-Learn Machine Learning Models for Thermal and Electrical Performance Prediction of High-Capacity Lithium-Ion Battery, *Int. J. Energy Res.*, Vol. 46, Issue 2, p. 786-794, (2022)
- (6) Heinrich, F., Klapper, P., Pruckner, M.: A comprehensive study on battery electric modeling approaches based on machine learning, *Energy Informatics*, Vol. 4, Article No. 17, (2021)
- (7) National Institute of Advanced Industrial Science and Technology: *Machine Learning Quality Management Guideline*, 2nd Edition, (2022)
- (8) <https://scikit-learn.org/stable/modules/generated/sklearn.tree.DecisionTreeRegressor.html#sklearn.tree.DecisionTreeRegressor>, 2021/1/15
- (9) Rangarajan, S. P., Barsukov, Y., Mukherjee, P. P.: In Operando Impedance Based Diagnostics of Electrode Kinetics in Li-Ion Pouch Cells, *Journal of The Electrochemical Society*, Vol. 166, No. 10, A2131-A2141, (2019)

■ Author ■



Takuma KAWAHARA



Yuki SATO



Koji SATO