

# テレマティクスを用いた自動車からの CO<sub>2</sub>排出量の推定と削減可能性の評価

## Estimation and Reduction Potential Evaluation of Vehicle CO<sub>2</sub> Emissions using Telematics

前 匡 鴻 \* · 王 子 洋 \* · 西 村 将 麻 \*\* · 松 橋 隆 治 \*  
Masahiro Mae      Ziyang Wang      Shoma Nishimura      Ryuji Matsuhashi

### **Abstract**

CO<sub>2</sub> emissions from gas-powered vehicles have a large impact on global warming. The aim of this paper is to develop a method for estimating CO<sub>2</sub> emissions from driving data for various vehicles using machine learning. CO<sub>2</sub> emissions from gas-powered vehicles vary depending on various factors such as vehicle type and driving pattern. The developed approach uses actual driving records and features of the vehicles constituting driving resistance. Machine learning based on random forest regression enables better estimation of CO<sub>2</sub> emissions compared to conventional multiple linear regression. The result implies the possibility of reducing CO<sub>2</sub> emissions by eco-driving with smooth acceleration and deceleration.

**Key words** : CO<sub>2</sub> emission, vehicle fuel consumption, estimation, machine learning, driving data

### 1. 序論

二酸化炭素 (CO<sub>2</sub>) に代表される温室効果ガスの排出量の削減は、日本における 2050 年のカーボンニュートラル社会の実現<sup>1)</sup>に向けて推進していく必要がある。現代の日本において、運輸部門から排出される CO<sub>2</sub> は全体の 2 割程度を占めており、その多くが自動車から排出されている。2021 年度における日本の CO<sub>2</sub> 排出量<sup>2)</sup>においては、全体の 10 億 6400 万トンの 7.7% に当たる 8191 万トンの CO<sub>2</sub> が自家用乗用車から排出されている。特に、自家用乗用車は旅客に利用される他の飛行機やバスや鉄道と比較して、輸送量当たりの CO<sub>2</sub> の排出量が最も大きく、自家用乗用車から排出される CO<sub>2</sub> の削減の必要性が高まっている。

自家用乗用車から排出される CO<sub>2</sub> の削減を進めていく上で、まず、自動車からの CO<sub>2</sub> 排出量を正確に推定する必要がある。自動車からの CO<sub>2</sub> 排出量は車種、車体の大きさ、車両重量、エンジンの種類、運転特性、季節など、多様な特徴量に依存しており、それらの特徴量と自動車からの CO<sub>2</sub> 排出量の関係性を全て第一原理からモデリングすることは、日本全国に 4000 万台存在する自家用乗用車<sup>3)</sup>の 1 台 1 台において、それぞれ用途やドライバーが異なることから容易ではない。したがって、自動車からの CO<sub>2</sub> 排出量

の推定において重要な特徴量を選択し、それらの特徴量含む限られたデータから汎用的な推定が可能なモデルを構築することが有効である。

自動車の走行データを取得可能なテレマティクスの普及に伴い、実際の走行データをビックデータとして活用した自動車からの CO<sub>2</sub> 排出量の推定手法の開発が進められている。自動車からの CO<sub>2</sub> 排出量推定について、線形回帰、サポートベクターマシン、エクストラツリー、ランダムフォレスト、多層パーセプトロン、深層学習などの教師あり学習を用いた推定モデルが複数提案されている<sup>4)5)6)7)</sup>。実際の走行データを用いた機械学習においては、取得可能なデータの種類やデータの大きさから適切な学習手法を選択する必要がある。特に、自家用乗用車の CO<sub>2</sub> 排出量推定においては、タクシーなどの車種がある程度限定される CO<sub>2</sub> 排出量推定<sup>8)</sup>と比較して、多様な車種に対する CO<sub>2</sub> 排出量推定モデルを構築する必要があり、学習に用いなかった車種や今後販売される未知の車種にも対応可能な汎化された推定モデルの構築が必要となる。また、走行データも途切れない時系列データが取得可能であるとは限らず、月毎の運転記録をイベントベースで記録したマンスリレポートのみが取得可能である場合なども存在するため、時系列解析に依存しない汎用性の高い CO<sub>2</sub> 排出量推定手法が必要である。

Corresponding author; Masahiro Mae, E-mail: mmae@ieec.org

\* 東京大学大学院工学系研究科電気系工学専攻

〒113-0033 東京都文京区本郷 7-3-1

\*\* あいおいニッセイ同和損害保険株式会社

〒150-8488 東京都渋谷区恵比寿 1-28-1

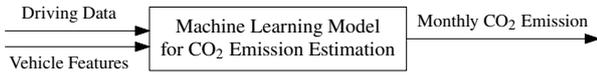


Fig. 1 Model of monthly CO<sub>2</sub> emission estimation using driving data and vehicle features.

このように、たしかに先行研究において実際の走行データを用いた機械学習による CO<sub>2</sub> 排出量推定に関する研究は複数行われているが、月間の走行データを用いた車種に対して汎化した機械学習による CO<sub>2</sub> 排出量推定モデルの構築については提案されていない。本論文では、車両の走行抵抗のダイナミクスに着目した自動車の種類に対して汎化した車両特徴量と実際の走行データを用いた機械学習モデルによる CO<sub>2</sub> 排出量の推定手法を提案する。

本論文の貢献は次の通りである。

- 車両の走行抵抗に基づく特徴量の選択による自動車の種類に対する汎化
- 車両特徴量と走行データを用いた機械学習モデルによる CO<sub>2</sub> 排出量の推定

本論文の構成は次の通りである。第 2 節で、自動車からの CO<sub>2</sub> 排出量推定に対する本論文の問題設定について述べる。第 3 節で、機械学習を用いた自動車からの CO<sub>2</sub> 排出量推定モデル構築の手法について説明する。第 4 節で、テレマティクスから取得される実際の走行データを用いた CO<sub>2</sub> 排出量推定の性能評価を行う。第 5 節で、結論を示して本論文のまとめとする。

## 2. 問題設定

本論文では、自動車からの CO<sub>2</sub> 排出量の推定性能を向上させることを目的とする。Fig. 1 に、本論文で扱う自動車からの CO<sub>2</sub> 排出量推定モデルを示す。自動車からの CO<sub>2</sub> 排出量の推定には、車両データと走行データと特徴量として用いる。車両データは、車種の情報をそのまま用いた場合、学習に用いていない車種に対する推定の汎化が難しいという問題がある。したがって、車両データは学習に使用する限られた車種だけでなく、学習に用いなかった車種や今後販売される未知の車種にも対応可能な特徴量を選択する必要がある。本論文で想定する走行データは、測定期間内で連続的な時系列の運転データではなく、月間の運転特性をカウントしたものであるため、線形性が低いという問題点がある。したがって、本論文で想定する非線形な走行データを特徴量として用いることのできる推定手法が必要である。以上の考察から、CO<sub>2</sub> 排出量を推定するモデルを構築するにあたり、次の要求事項を満たす必要がある。

- 多様な車種に対応可能な CO<sub>2</sub> 排出量の推定
- 運転特性の走行データに基づく CO<sub>2</sub> 排出量の推定

Table 1 Symbols of vehicle dynamics model.

Symbol	Description	Unit
$F$	Driving Force	N
$v$	Velocity	m/s
$a$	Acceleration	m/s <sup>2</sup>
$g$	Gravitational Acceleration	m/s <sup>2</sup>
$m$	Vehicle Mass	kg
$A$	Frontal Area	m <sup>2</sup>
$\rho$	Air Density	km/m <sup>3</sup>
$r$	Wheel Radius	m
$B_r$	Viscous Damping Coefficient	Nms/rad
$C_r$	Rolling Resistance Constant	-
$C_d$	Air Resistance Coefficient	N/kN

## 3. 機械学習を用いた自動車からの CO<sub>2</sub> 排出量推定モデルの構築

本節では、機械学習を用いた自動車からの CO<sub>2</sub> 排出量推定モデル構築の手法について説明する。走行抵抗に基づく車両特徴量の選択による自動車の種類に対する推定性能の汎化と、テレマティクスから取得可能な月間の走行データを特徴量として選択することによる実際の運転に合わせた CO<sub>2</sub> 排出量推定を可能とする。これらの特徴量を用いた機械学習を、ランダムフォレスト回帰を用いた CO<sub>2</sub> 排出量推定モデルにより行うことで、従来の多重線形回帰の推定手法よりも高い推定精度を実現する。

### 3.1 走行抵抗に基づく車両特徴量の選択

本論文では、車種に依存しない CO<sub>2</sub> 排出量推定モデルを構築するために、自動車の燃費に直接影響する走行抵抗のダイナミクスに着目し、機械学習に用いる特徴量の選択を行う。車両の走行抵抗は次式によりモデル化される<sup>8)</sup>。

$$F = ma + mgC_r + 4v \frac{B_r}{r^2} + \frac{1}{2} Av^2 \rho C_d \quad (1)$$

ただし、各変数はそれぞれ Table 1 の物理量を表す。右辺の各項は左から順に加速抵抗、転がり抵抗、粘性抵抗、空気抵抗を表す。

駆動力  $F$  は、燃料消費量や CO<sub>2</sub> 排出量と繋がりがあることから、本論文の問題設定では目的変数に関連する。速度  $v$  と加速度  $a$  は走行データに繋がりがあることから、本論文では車両特徴量としては用いない。したがって、車両重量  $m$  と前方投影面積  $A$  を車両特徴量として用いることで、機械学習により重力加速度  $g$ 、空気密度  $\rho$ 、タイヤ半径  $r$ 、粘性弾性係数  $B_r$ 、転がり抵抗係数  $C_r$ 、空気抵抗係数  $C_d$  に相当する要素を学習することとなる。また、エンジンタイプも燃費に大きく影響する特徴量であることから、内燃機関車 (InternalCombustionVehicle : ICV)、または、ハイブリッド車 (HybridElectricVehicle : HEV) の 2 種類を区別して車両特徴量として扱う。

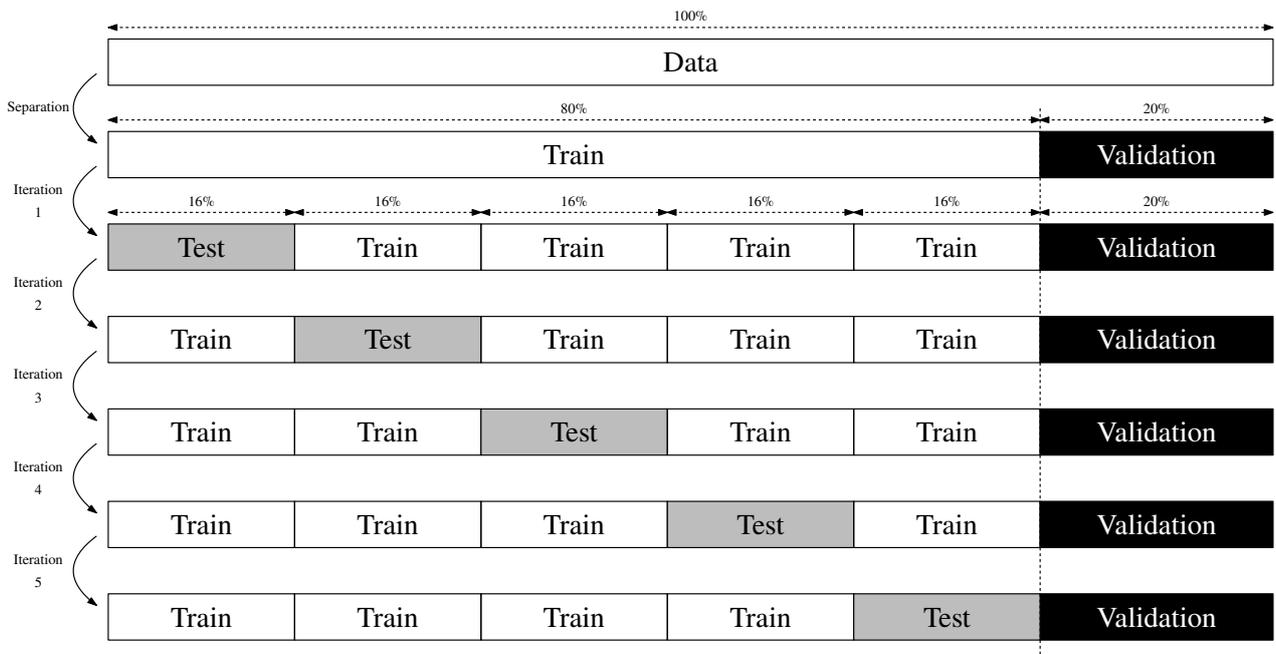


Fig. 2 5-fold cross-validation utilizing a grid search technique in random forest regression.

Table 2 Features obtained from telematics.

Feature	Unit
Target Year and Month	-
Start Date of Relevant Month	-
End Date of Relevant Month	-
Monthly Driving Time	s
Monthly Driving Distance	m
Speeding on Expressway	times
Sudden Acceleration on Expressway	times
Sudden Braking on Expressway	times
Driving Distance on Expressway	m
Speeding on National Highway	times
Sudden Acceleration on National Highway	times
Sudden Braking on National Highway	times
Driving Distance on National Highway	m
Speeding on Local Road	times
Sudden Acceleration on Local Road	times
Sudden Braking on Local Road	times
Driving Distance on Local Road	m
Safe Driving Score	%

### 3.2 走行データに基づく運転特徴量の選択

本論文では、テレマティクスより取得可能な月間の走行データを用いて機械学習を行う。走行データにおける特徴量の一覧を Table 2 に示す。特徴量は、月間の測定期間、走行時間、走行距離、高速道・国道・一般道における速度超過・急加速・急減速の回数、そして、それらの走行特性から計算される安全運転スコア<sup>9)</sup>により構成される。走行データは月別に測定され、同じ車種に対しても複数のユーザーからデータを取得している。

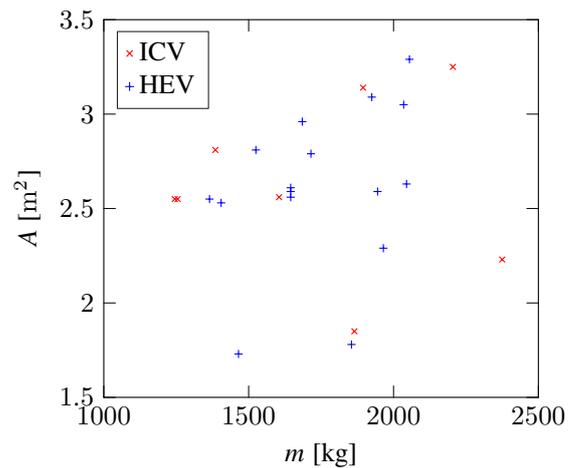


Fig. 3 Distribution of vehicle mass and frontal area of 24 vehicles in driving data. ICV denotes Internal Combustion Vehicle and HEV denotes Hybrid Electric Vehicle, respectively.

### 3.3 ランダムフォレスト回帰を用いた CO<sub>2</sub> 排出量推定モデル

走行抵抗に基づく車両特徴量と走行データに基づく運転特徴量を合わせて用いることで、機械学習により自動車からの CO<sub>2</sub> 排出量推定モデルを構築する。本論文では、各説明変数の線形和で推定される多重線形回帰モデルに加え、教師あり機械学習に分類されるランダムフォレスト回帰モデルを CO<sub>2</sub> 排出量推定に用いる。CO<sub>2</sub> 排出量は走行抵抗や燃費から決定される物理量であるが、本論文で用いる Table 2 のテレマティクスより取得可能な走行データは月間のデータとなっており、また、走行特性は速度超過・急加速・

Table 3 Estimation performance of monthly CO<sub>2</sub> emission and computation time of each machine learning method.

Method	R <sup>2</sup> [-]	RMSE [kg]	MAE [kg]	Computation Time
Multiple Linear Regression	0.873769	30.8296	19.8524	< 1 second
Random Forest Regression	0.980831	12.0140	7.55227	> 4 hour

急減速の回数をカウントしたデータになっている。したがって、それらの説明変数と CO<sub>2</sub> 排出量の関係は線形でなく、それらの説明変数を用いる場合はランダムフォレスト回帰に代表される非線形の回帰手法が有効であると考えられる。ランダムフォレスト回帰は、学習のために大量のデータが必要であるが汎化性能が高いという特徴があり、多様な車種や走行データに適応した CO<sub>2</sub> 排出量推定を行う上で適していると考えられる。

学習においては、全データをシャッフルした後、80%を学習用のデータとして用いて、残りの 20%を評価用のデータとして用いる。ランダムフォレスト回帰の過学習を防ぐために Fig. 2 に示すようにグリッドサーチによる 5 分割交差検証を行う。学習用のデータは 5 分割され、そのうち 4 つを学習、1 つをテストに用いる学習を 5 回繰り返し、5 回の学習の平均値を学習結果とする。

#### 4. テレマティクスの走行データを用いた CO<sub>2</sub> 排出量推定の評価

本節では、テレマティクスから取得される実際の走行データを用いた CO<sub>2</sub> 排出量推定の性能評価を行う。機械学習に用いる車両の走行データと計算条件を述べた上で、多重線形回帰とランダムフォレスト回帰の学習結果を比較して推定性能の評価を行う。

##### 4.1 学習条件

本論文では、機械学習に用いるデータとして、Table 2 のテレマティクスより取得可能な月間の走行データ、各車両の車両重量、前方投影面積、エンジンタイプを説明変数、月間 CO<sub>2</sub> 排出量  $E_{CO_2}$  [kg] を目的変数として、正規化した値を用いて学習を行う。学習に用いる月間 CO<sub>2</sub> 排出量は、テレマティクスのデータとは別に取得した月間燃料消費量  $C_{gasoline}$  [L] のデータから次のように計算される。

$$E_{CO_2} = 2.3 \times C_{gasoline} \quad (2)$$

Fig. 3 に、走行データに用いた車種のエンジンタイプごとの車両重量、前方投影面積の分布を示す。Fig. 3 より、車両重量と前方投影面積との間に強い相関はなく、これらの変の間に多重共線性の可能性は低いと考えられる。また、内燃機関車とハイブリッド車のデータがともに広範囲な変数空

間に散らばっていることは、変数空間内の内挿性能の高さに繋がる。Fig. 3 に示す 24 車種について、合計  $n = 673248$  個の月間の走行データを学習に用いる。機械学習には、Python 3<sup>10)</sup> の Scikit-learn<sup>11)</sup> を用いる。平均二乗誤差を損失関数として学習を行う。多重線形回帰においては、過学習の可能性が低いので、 $k$ -分割交差検証は行わず、80%のデータを一括で用いて学習し、残りの 20%のデータを用いて評価を行う。ランダムフォレスト回帰において、Scikit-learn の関数 (RandomForestRegressor) に与えるハイパーパラメータは、次のように経験的に設定し、これらの組み合わせに対して総当たりに学習を行う。

- 'n\_estimators': [200,800,1400]
- 'max\_features': ['auto','sqrt']
- 'max\_depth': [20,30,40]

ただし、各ハイパーパラメータは次のような特徴を持つことに注意されたい。n\_estimators は決定木の数であり、大きいほど表現力が高まるが同時に計算コストも高まる。max\_features は 1 つの決定木が分岐に使用する特徴量の数であり、決定木の乱数性が決定される。max\_depth は決定木の深さの最大値であり、大きいほど表現力が高まるが過学習の可能性も高まる。機械学習の計算には、Intel Xeon Gold 6258R の CPU を 2 つ、NVIDIA Quadro RTX8000 の GPU を 2 つ、768 GB の RAM を搭載するサーバを用いる。

##### 4.2 学習結果

自動車からの月間 CO<sub>2</sub> 排出量推定に関して、多重線形回帰とランダムフォレスト回帰を用いた機械学習を行った。目的変数  $y$ 、推定値  $\hat{y}$ 、平均値  $\bar{y}$  に対して、次式で定義される決定係数 R<sup>2</sup>、二乗平均平方根誤差 (Root Mean Square Error : RMSE)、平均絶対誤差 (Mean Absolute Error : MAE) において、推定精度の評価を行った。

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (4)$$

$$MAE(y, \hat{y}) = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (5)$$

学習結果における月間 CO<sub>2</sub> 排出量推定精度と各学習に必要な計算時間の比較を Table 3 に示す。ランダムフォレスト回帰における学習後の推定モデルにおいて、総当たりしたハイパーパラメータの内、次の組み合わせにおいて最も良い推定結果が得られた。

- ‘n\_estimators’: 1400
- ‘max\_features’: ‘auto’
- ‘max\_depth’: 40

### 4.3 推定精度の評価

学習結果における多重線形回帰とランダムフォレスト回帰のそれぞれの学習結果に関する考察を述べる。

多重線形回帰は、決定係数  $R^2=0.87$  となり、ある程度高い決定係数を実現できている上に、計算時間が 1 秒以下で非常に短いことがわかる。多重線形回帰は、モデルがシンプルであることから、 $t$  値や  $p$  値による各特徴量の寄与度の評価や学習結果の解釈が容易で、過学習を防ぎやすいということも利点として挙げられる。

ランダムフォレスト回帰は、決定係数  $R^2 = 0.98$  となり、非常に高い決定係数を実現できているが、計算時間が 4 時間以上と比較的長いことがわかる。ランダムフォレスト回帰そのものがアンサンブル学習に基づいていることに加え、 $k$ -分割交差検証を用いることで過学習を防ぎ、汎化性能の向上に努めている。

多重線形回帰とランダムフォレスト回帰のどちらの手法も、その一般的な特性から学習に用いた変数空間の範囲内に対する内挿性能は高いことが見込まれるが、外挿性能に関しては高くないことが知られている。そのため、データセットに含まれない未知の車種や新たな走行データが、学習に利用したデータの変数空間の範囲内に含まれるように、ある程度幅広い車種や走行データでの学習が行われていることが望ましいと言える。

また、J-クレジット制度<sup>12)</sup>などへの実社会応用の観点では、多重線形回帰とランダムフォレスト回帰において決定係数の高さや計算時間にトレードオフの関係があるため、推定した CO<sub>2</sub> 排出量の具体的な利用方法や、推定と学習データ更新の頻度に合わせて、適切な推定手法が選択されることが望ましいと言える。

## 5. 結論

テレマティクスの走行データに基づいた運転特徴量を機械学習に用いることで、走行特性に基づく自動車からの CO<sub>2</sub> 排出量を推定することができる。本論文では、テレマティクスの走行データに基づく運転特徴量と走行抵抗に基づく車両特徴量を用いた機械学習による自動車からの CO<sub>2</sub> 排出量の推定手法を開発した。開発した CO<sub>2</sub> 排出量推定モデ

ルでは、車種の情報を用いず、走行抵抗に基づいて車両重量と前方投影面積を車両特徴量として用いることにより、車種に依存しない CO<sub>2</sub> 排出量の推定性能の汎化を実現した。これにより、機械学習に用いていない車種や、将来発売される新しい車種に対する CO<sub>2</sub> 排出量の推定性能の劣化を低減することができる。本手法を用いることにより、車種に汎化された走行特性に基づく自動車からの CO<sub>2</sub> 排出量の推定が可能となり、テレマティクスの走行データを活用した CO<sub>2</sub> 排出量の評価や、J-クレジット制度による温室効果ガスの排出削減量の取引への応用が期待される。

### 参考文献

- 1) 日本経済新聞：菅首相の所信表明演説全文 (2020).  
<https://www.nikkei.com/article/DGXMZO65465240W0A021C2000000>
- 2) 国土交通省：運輸部門における二酸化炭素排出量 (2023).  
[https://www.mlit.go.jp/sogoseisaku/environment/sosei\\_environment\\_tk\\_000007.html](https://www.mlit.go.jp/sogoseisaku/environment/sosei_environment_tk_000007.html)
- 3) 国土交通省：数字で見る自動車 2021 (2021).  
[https://www.mlit.go.jp/jidosha/jidosha\\_fr1\\_000069.html](https://www.mlit.go.jp/jidosha/jidosha_fr1_000069.html)
- 4) Schoen, A., Byerly, A., Hendrix, B., Bagwe, R. M., Santos, dos E. C. and Ben Miled, Z.: A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles, IEEE Transactions on Vehicular Technology, Vol. 68, No. 7, pp. 6343–6351 (2019).
- 5) Ghahramani, M. and Pilla, F.: Analysis of Carbon Dioxide Emissions From Road Transport Using Taxi Trips, IEEE Access, Vol. 9, pp. 98573–98580 (2021).
- 6) Maroju, R., Nishimura, S., Wang, Z. and Matsushashi, R.: Estimating Vehicular Fuel Consumption and CO<sub>2</sub> Emissions by Machine Learning Using Only Speed and Acceleration, Journal of Japan Society of Energy and Resources, Vol. 44, No. 1, pp. 30–38 (2023).
- 7) Sai Manvitha, M., Vani Pujitha, M., Hari Prasad, N. and Yashitha Anju, B.: A Predictive Analysis on CO<sub>2</sub> Emissions in Automobiles using Machine Learning Techniques, in 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), pp. 394–401, IEEE (2023).
- 8) Hattori, M., Shimizu, O., Nagai, S., Fujimoto, H., Sato, K., Takeda, Y. and Nagashio, T.: Quadrant Dynamic Programming for Optimizing Velocity of Ecological Adaptive Cruise Control, IEEE/ASME Transactions on Mechatronics, Vol. 27, No. 3, pp. 1533–1544 (2022).

- 9) あいおいニッセイ同和損保：タフ・つながるクルマの保険 (2018).  
<https://www.aioinissaydowa.co.jp/personal/product/tough/tsunagaru>
- 10) Python Software Foundation: Python 3.7.13 (2022).  
<https://www.python.org>
- 11) Cournapeau, D.: scikit-learn 1.0.2 (2021).  
<https://scikit-learn.org>
- 12) 経済産業省：J-クレジット制度 (2013).  
<https://japancredit.go.jp>