****
技術論文
****
Journal of Japan Society of Energy and Resources, Vol. 44, No. 1

Estimating Vehicular Fuel Consumption and CO_2 Emissions by Machine Learning Using Only Speed and Acceleration

Rahul Maroju*, Shoma Nishimura**, Ziyang Wang*, Ryuji Matsuhashi*

(Received date: June 16, 2022, Accepted date: December 12, 2022)

Abstract

Road transportation has a good share in the global carbon dioxide emissions and models for estimating the vehicular emissions using physical parameters help to understand and potentially reduce them. However, such models at a regional level are usually insensitive to driving dynamics, as they are based on the average speed of the vehicles. The existing models considering the instantaneous speed and acceleration also use other factors like weather, vehicle parameters, etc., which involves many measurements. Furthermore, the estimation is done at large time periods of the order of several seconds. In this work, a real-time time series data is used to develop a model using only the vehicular speed and acceleration. It is based on a novel technique of using windows of the driving dynamics captured in a very short period, making some assumptions. The optimal drive features that influence the fuel consumption have been estimated using many machine learning regression models, validated, and compared. Among them, a multi-layer perceptron resulted the highest cross-validation of 0.64 using only the window of speeds, which is concluded to be reasonably good for practical estimation. Finally, these models are aimed to be applied in real applications based on J-credit and eco-driving.

Keywords : CO_2 emissions, fuel consumption, driving dynamics, regression, machine learning

1. Introduction

Road transportation is a key contributor of emissions across the world, as it has a significant hand in the economic development of every country. Among the various sectors contributing to CO_2 emissions in the world, road transport is responsible for a good share of 16%. [1]

The effects of the emissions on the environment and human health are well-known. Thus, there is a strong and urgent necessity to reduce the emissions as much as possible. This is mainly for two reasons – to meet the global targets to mitigate the impacts of climate change, and a reduction in the emissions by reducing the use of fuel will help to save the precious fossil resources. Though there have already been many measures taken like technological advancement and policies to control emissions, the increasing number of vehicles in most parts of the world makes it a big challenge.

This study is a contribution to the existing models of estimating CO_2 emissions using only the vehicular driving dynamics of speed and acceleration. A novel technique based on using windows of the driving dynamics has been presented, which will be used with different machine learning models. Finally, the optimal set of drive features that influence the fuel consumption have been identified. The main motivation for this approach comes from the fact that the navigation data measurement in vehicles is becoming ubiquitous. At a regional level, providing a reasonable estimate of the CO_2 emissions using this data saves the requirement for sensors. Such an estimation would be very important for many reasons, primarily in understanding the effects of transport emissions for framing policies to achieve a net carbon neutral society. At an individual level, this can help the users track their driving behavior. Moreover, these models can be integrated into traffic network simulators to better understand the impacts of traffic policies.

2. Background

Among the models for estimating CO_2 emissions, some widely used ones at a regional level have been studied and mentioned below. The Air Pollution Model (TAPM) developed by the CSIRO Marine and Atmospheric Research is a widely used model for predicting large scale emissions. However, such a model doesn't account for small scale perturbations in emissions from vehicles, as it does not consider the driving dynamics. [2] Another model is the MOBILE 6.2 developed by the US EPA (Environment Protection Agency) and European based COPERT model. [3] They are based on the average speed, hence exclude the driving dynamics.

This motivates to develop models where the emissions can be

Corresponding author: Ryuji Matsuhashi,
E-mail: matu@enesys.t.u-tokyo.ac.jp
*The University of Tokyo
7 Chome-3-1 Hongo, Bunkyo-ku, Tokyo 113-8654, Japan
**AIOI Nissay Dowa Insurance Co. Ltd.
1-28-1 Ebisu, Shibuya-ku, Tokyo 150-8488, Japan

This paper is written based on the proceedings of the 41st Annual Meeting of the Japan Society of Energy and Resources.

estimated based on the driving modes, such as the effects of ramps, intersections, and traffic controls. The instantaneous speed and acceleration can be considered as the key parameters in capturing this information.

Among such models based on speed and acceleration, some models are based on VSP (vehicle specific power), a non-linear function of velocity and acceleration. [4] The emission rate is given by a lookup table for a range of VSPs. Some models are based on engine parameters like efficiency, power, friction factor, etc. [5]

Oduro et al. [6] proposed a multiple linear regression model for a particular type of vehicles for driving styles in Australia. They have reported a significant linear relationship between emissions and speed, acceleration and speed have a greater impact. Ahn et al. [7] proposed a polynomial and hybrid models using instantaneous speed and acceleration, and this model has been used in several other works. Also, the fuel consumption estimation is estimated over larger periods, spanning over the order of several seconds, hence the macroscopic relationship between the driving dynamics and fuel consumed is observed.

From these works, it has been observed that there are very few models for estimating CO_2 emissions based on only the vehicle's speed and acceleration. The models on a large scale are based on average speed, which excludes the driving dynamics. On the other hand, there are models based on many parameters which could pose some difficulties to be measured for a large-scale practical application.

3. Models

A brief overview of some common supervised regression models used in this study is given below.

3.1 Linear regression

Linear regression works on the principle of fitting a linear combination of transformations of input features to minimize the sum of squared error with the target response for all the data points. The optimal weights are obtained, characterizing the model.

3.2 Support vector regression

A support vector machine works with the objective to find a hyperplane in a transformed space, using a function called *kernel* that distinctly classifies all the data points. The data points closest to the hyperplane on either side of it are called the support vectors.

In the method of support vector regression (SVR), two decision boundaries are created within a threshold distance around the hyperplane, and the best fit hyperplane is the one that has the maximum number of points. This hyperplane will then serve as the mapping to estimate the target values for any point. The most important parameters to fit involves the choice of the kernel and the epsilon for the threshold distance.

3.3 Random Forest regression

Decision tree models are based on *recursive partitioning* – starting from the root, each node can be split into child nodes. These nodes can then be further split and they themselves become parent nodes of their resulting child nodes. At each spilt, a criterion of information gain is maximized. Decision tree learns the structure of the tree, features used at each node and the threshold parameters for each decision

This way the entire space is split into regions based on a decision tree. For a decision tree regression, the optimal value of a predictive variable within a region is the average of the values of the training responses in that region.

Random forest is an ensemble decision tree-based model which uses the technique of *bootstrap aggregating* or *bagging*. *Bootstrapping* is a technique of generating a set of samples from of the same size by selecting with replacement. For example, if the data set is $\{x_1, x_2, x_3\}$, some bootstrapped sets are $\{x_1, x_3, x_1\}, \{x_2, x_2, x_2\}$, etc. A collection of decision trees each is trained with a bootstrapped set of data. Each tree is trained using only a random subsampled set of features. The final estimate of the random forest is the aggregated output of each of the tree, which is the average estimate of all these decision trees in our case.

3.4 Extra trees regression

The extra trees model is very similar to a random forest model, but it fits several randomized decision trees generally using the entire dataset and then uses the average of the estimates of the estimators. The main difference comes from the fact that a random forest chooses the optimum split whereas the extra trees regressor chooses it randomly. Thus, this fits the model much faster even for a large number of trees.

Random forest and extra trees are the most popular ensemble methods using decision trees.

3.5 Multi-layer perceptron

Multi-layer perceptron (MLP) model is one of the most popular deep learning models with a potential to characterize non-linear patterns. It consists of an input and output layer with many hidden layers in between them, each layer containing many nodes. It is trained using the principle of back-propagation where the weights at each layer are trained to minimize the loss.

4. Data and method

The research content is explained in six parts – problem formulation, data and processing, correlation analysis, input features, model validation and simulations.

4.1 Problem formulation

Our objective is to estimate the amount of fuel consumption for every few seconds using the driving dynamics of a vehicle. The choice of this period is set to two seconds for a reasonable real-time fuel consumption estimation. Later, the corresponding CO_2 emissions can be easily calculated using some direct conversion factors for the type of the vehicle. According to [8], for complete combustion, about 99% of the carbon in the burnt fuel is emitted as CO_2, thus the CO_2 emitted, and fuel consumed are greatly correlated.

We had set some assumptions for our analysis. First, a type of the vehicle is fixed – a typical passenger car. Next, the vehicle is assumed to be in normal running condition, hence several factors i.e., the effects of age, etc., are ignored. The variables influencing the emissions can be classified into six broad categories, as travel-related, weather-related, vehicle-related, roadway-related, traffic-related, and driver-related factors. Among them, we only use the travel-related information in form of speeds and ignore the other variables assuming they are fixed.

4.2 Data and processing

A real-time experiment of a car ride is performed in Japan for a total duration of 30 minutes, and two time series datasets along with their corresponding time stamps are obtained. One of them is the drive data containing the Global Positioning System (GPS) parameters of the vehicle are measured by attaching a telematic tag during the ride which sends time-to-time data to an application via Bluetooth. Another sensor is attached to measure the fuel consumed periodically during each instant.

The two datasets are obtained at different sampling rates, the drive data has 15 samples per second, and the fuel consumption is obtained at 20 samples per second at different time stamps. To fit a model, we need to merge these datasets to a single dataset of the lower rate, i.e., the sampling time $\Delta t = 1/15\ sec$. This is done by computing the fuel consumption (y_n) at the time stamps of the speeds using linear interpolation. It should be noted that smoother interpolation techniques like spline interpolation results in negative values of fuel consumption at some instances making the interpolated estimation of the fuel consumption invalid.

From the drive data, only the vehicle speed (v_n) is used throughout this analysis and the acceleration (a_n) is calculated using the basic definitions in (1), (2) and (3).

Original data: $v_n, y_n, n = 1, 2, \ldots$

$$\text{Backward-difference acceleration } \boldsymbol{a_n^{(b)}} = \frac{\boldsymbol{v_n - v_{n-1}}}{\boldsymbol{\Delta t}} \quad (1)$$

$$\text{Central-difference acceleration } a_n^{(c)} = \frac{v_{n+1} - v_{n-1}}{2\,\Delta t} \quad (2)$$

$$\text{Non-negative acceleration } \tilde{a}_n = \begin{cases} a_n, & a_n > 0 \\ 0, & otherwise \end{cases} \quad (3)$$

A new feature non-negative acceleration is defined as the acceleration (backward-difference or central-difference) when it is positive and zero otherwise. The motivation for using this comes from the observation that during a drive, a positive acceleration consumes fuel, whereas a non-abrupt deceleration consumes little or almost no fuel. So, a non-negative acceleration may serve as an indicator more directly related to the fuel consumption than the typical acceleration.

As the objective is to estimate the fuel consumption every 2 seconds, which corresponds to 30 samples, we use the aggregated data of the speeds and the fuel consumption as the total fuel consumed in that duration. That means, if the (Input, Target) of the original data is in the form (v_n, y_n), then the transformed data will be in the form $\left([v_n, v_{n+1}, \cdots, v_{n+29}], z_n = \sum_{i=0}^{29} y_{n+i}\right)$.

4.3 Correlation analysis

During a ride, there is an empirical relationship between the fuel consumption and acceleration. The more the vehicle accelerates, the more fuel it consumed to drive faster. To observe this dependence mathematically, the Pearson correlation coefficient is calculated between these two series of data by shifting the fuel consumption for a range of indices. A positive shift implies making the future values of the series as the current values by the corresponding index and similarly a negative shift implies making the past values as the current values by the index.

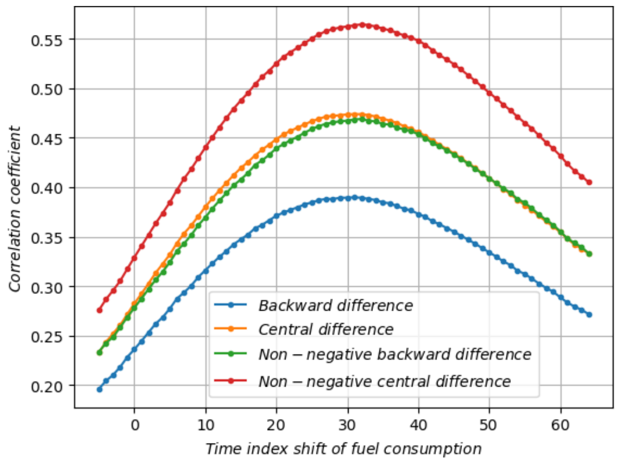

Fig.1 Correlation between various forms of accelerations and fuel consumption

Fig.1 shows the correlation coefficients for different forms of accelerations defined earlier. We observe that the highest correlation occurs at an index of 32, which means the current fuel consumption seem to depend on some past accelerations, and in general, the driving dynamics. Central difference acceleration is far more correlated to the fuel consumption than the backward difference. Also, the use of non-negative accelerations has been justified due to their significantly larger correlation coefficient.

Next, we see the effect of applying moving average on this correlation. A moving averaged series is obtained by replacing the corresponding data point with the average of some data points around it in a window. Different sizes of this moving average window have been tested and the corresponding maximum correlation coefficient has been recorded. From **Fig.2**, we observe that the correlation increases significantly even for a small MA window, and there has been a consistent increase in the correlation by increasing the order of the moving averaging, finally followed by saturation.

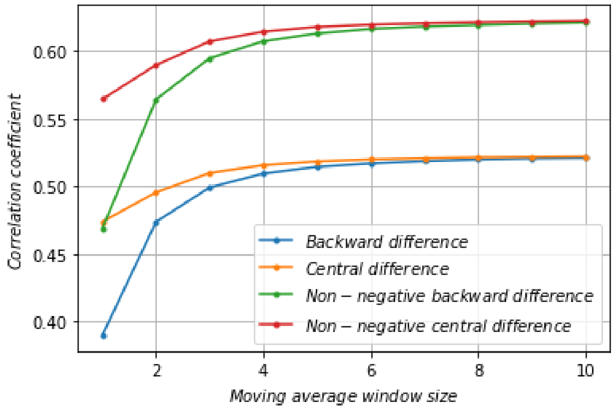

Fig.2 Maximum correlation between various moving averaged accelerations and fuel consumption

We choose an MA window size of 5 for some tests, as the increase in the correlation coefficient is less than 1% for higher order. Again, central difference acceleration always demonstrated better correlations, and hence throughout this analysis, we shall use it by default, and acceleration implies central difference acceleration.

This analysis provides some evidence for the dependence of the fuel consumption at an instant on the past accelerations, or the driving dynamics in general due to a peak in the correlation. However, the correlation coefficient measures the linear similarity between the two data series, and the relation between the fuel consumption and acceleration is not linear in general. Hence, many non-linear models will be used in this analysis, but the key takeaway from this analysis is to use the information of the past vehicle drive parameters to estimate the current fuel consumption.

4.4 Input Features

From the correlation analysis, we inferred that the fuel consumption may depend on some of the past driving dynamics of the vehicle. However, as we have a variety of information of the past – speed, acceleration, and many variants, the question remains which of these can be used to fit a model. We present a method to systematically use the past information for various cases to obtain which of this information serve the best for the estimation. For this, we introduce the concept of a drive window. To estimate the fuel consumption z_n, a window of drive data can be used as the input features, which is a set of consecutive points.

We define a drive window, or simply a window of speeds (v) of size M and center index k in (4).

$$\overline{\boldsymbol{v}}_{n,k}^{M} = \begin{cases} \left[v_{\left(n-k-\frac{M-1}{2}\right)}, \cdots, v_{(n-k)}, \cdots, v_{\left(n-k+\frac{M-1}{2}\right)}\right], \text{if } M \text{ is odd} \\ \left[v_{\left(n-k-\frac{M}{2}\right)}, \cdots, v_{(n-k)}, \cdots, v_{\left(n-k+\frac{M}{2}-1\right)}\right], \text{otherwise} \end{cases} \tag{4}$$

Similarly, we can define a window of acceleration. The idea is to use the drive information in this window. Some useful choices of the features are defined below.

Only speeds: $\overline{\boldsymbol{v}}_{n,k}^{M} = [\cdots,\ v_{(n-k-1)},\ v_{(n-k)},\ v_{(n-k+1)},\ \cdots]$

Center speed and accelerations:

$\left[v_{(n-k)}, \overline{\boldsymbol{a}}_{n,k}^{M}\right] = \left[v_{(n-k)} | \cdots, a_{(n-k-1)}, a_{(n-k)}, a_{(n-k+1)}, \cdots\right]$

Using a window of accelerations and one of the speeds, all the other speeds in the window can be approximately reconstructed.

Center speed and non-negative accelerations: $\left[v_{(n-k)}, \overline{\tilde{\boldsymbol{a}}}_{n,k}^{M}\right]$

This is the same as the previous case but using non-negative accelerations instead of normal accelerations.

Mean speed and accelerations: $\left[\frac{1}{M}\sum_{i\in\mathcal{M}} v_{(n-k+i)}, \overline{\boldsymbol{a}}_{n,k}^{M}\right]$

where $\mathcal{M}$ is the range of indices such that the mean of all the speeds in the drive window is used.

These are some of the basic sets of features, and in a similar fashion many other variants of these feature sets have been used. For each set, a grid search with window size (M) and center index (k) is performed to obtain an optimal set of drive parameters.

4.5 Model validation

The model is validated using the technique of k-fold cross-validation (CV), where the entire data is split into k parts or folds without shuffling. The model is fitted on k-1 folds and tested on the remaining fold. This way, every data point will be tested in one of the splits. R^2 score is used as the performance metric, as it is a popular regression metric. The final score of the model is the mean score for all the splits.

We maintain the order of the data samples, as this is a time series regression. We rely on the assumption that the data in each block is independent of the others, and thus the corresponding estimations use information only from that block thus making the validation meaningful. [9]

4.6 Simulations

As we aim to estimate the aggregated fuel consumed during the entire duration of 2 seconds, M must be at least 30, as we need to use at least that period of the vehicle drive data. Otherwise, some portion of the drive information remains unused thus making the model impractical. So, a choice of range is decided as $M = 30, 31, \cdots, 45$. This corresponds to sliding windows from 2 seconds to 2.5 seconds of drive information.

The list of center indices has been chosen in a range of the past driving dynamics as $k = 20, 19, \cdots, -5$ based on experiments. We use a 5-fold cross-validation technique for evaluating the performance for each case. Some popular regression models have been fitted and compared.

5. Results and discussion

5.1 Results using different regression analysis

Some regression models have been fitted using the center speed and trajectory of acceleration as the input features and the cross-validation (CV) scores have been calculated for a range of window sizes and window center indices mentioned earlier.

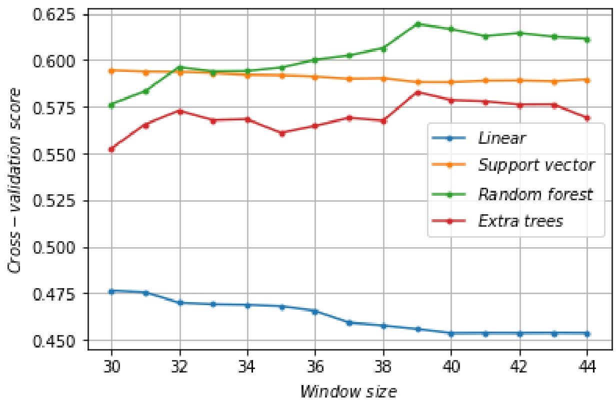

Fig.3 Maximum cross-validation score for each window size using various regression models

Fig.3 shows the CV score for each window size, using the optimal window center index, which resulted the maximum score. Among these models, random forest performs better than the other models for most of the window sizes.

Support vector regression (SVR) and extra trees provide a decent peak score close to 0.60 and 0.58 respectively. As SVR works based on a very different method than tree-based approach, it does not detect an optimal window size, but instead the score decreases gradually and consistently by increasing the window size. SVR is fitted using different kernels – linear, polynomial and a radial basis function (RBF), among which RBF gives the best score, which is shown in the figure. Extra trees regression is not particularly helpful for our case, as the training time is very less even with a random forest. However, it is reported only to show that it could detect a peak score at a window size of 39 showing further evidence for that size as the optimal window size.

5.2 Random Forest regression analysis

In summary, using a decision tree-based ensemble model – random forest for estimating the fuel consumption using the vehicle trajectory would be apt. Intuitively, it works using a series of conditional cases – did the speed increase or decrease which is equivalent to whether the acceleration is positive or negative and allocates a fuel consumption value from the decision space based on the corresponding vehicle trajectory.

Hence, the random forest model is tested in detail with many cases of the feature sets. Initially, the number of estimators (trees) in the model is set to 30, and later this has been increased.

Among these features, using center speed and the trajectory of acceleration give the best score of 0.619. Using central difference acceleration is always better than using the backward difference acceleration. Using moving average of acceleration and non-negative acceleration slightly reduced the scores. However, using acceleration all the scores are in a similar range of around 0.6.

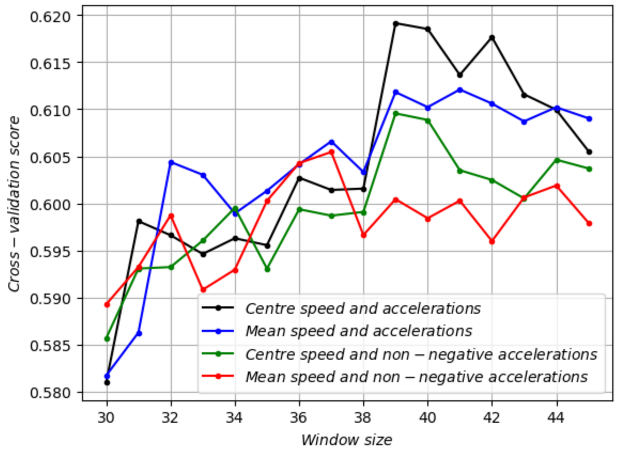

Fig.4 Maximum cross-validation score for each window size using a random forest model with 30 estimators for different input features

Fig.4 and **Table 1** summarizes the maximum scores along with the optimal window size and center index for various kind of features. For each case, a grid search is performed for different window sizes and center indices. Once the optimal window is estimated, the number of estimators in the random forest is varied from 20 to 200 in steps of 10 and among these, the highest score has been reported in the table.

Table 1 Cross-validation scores and the optimal window size and center index using a random forest with 30 estimators using various kinds of features

Input features	Score	Size	Center index
Only speeds	0.529	36	13
All speeds and accelerations	0.615	36	13
Center speed and accelerations	0.619*	39	12
Center speed and non-negative accelerations	0.610	39	12
Center speed and MA (5) accelerations	0.604	41	12
Center speed and MA (5) non-negative accelerations	0.610	36	8
Center speed and backward difference accelerations	0.585	42	11
Mean speed and accelerations	0.612	41	12
Mean speed and non-negative accelerations	0.611	37	11
Mean speed and MA (5) accelerations	0.601	41	14
Mean speed and MA (5) non-negative accelerations	0.606	40	8
Mean speed and backward difference accelerations	0.580	43	11

*: Best score, acceleration implies central-difference acceleration

The benefit of using non-negative acceleration seem to appear only at lower window sizes. For most of the other sizes, including the optimal window size of 39, it gave a score lower than normal acceleration. Thus, using normal acceleration captures better information for estimation than non-negative acceleration.

There is a great consistency in the optimal windows for all these cases. Each unit of the window size and center index corresponds to 1/15 seconds, and hence the optimal duration of the drive data used is very similar for all cases.

5.3 Multi- layer perceptron analysis

Next a multi-layer perceptron is fitted. This model is mentioned separately from the others as it gives the best score when the trajectory of speeds is given directly as the input features instead of any acceleration. For the same range of window size (M) and center indices (k), a grid search of the optimal parameters for model for each case using a single hidden layer have been tested as mentioned in **Table 2**.

Table 2 Choice of parameters for MLP

Parameters	Tested choices
Hidden layer sizes	$10, 15, 20, \cdots, 80$
Solver	Adam, SGD
$alpha$	$0.01, 0.05, 0.1, 0.5$
Learning rate	Constant, Adaptive

After this search, the best cross-validation score was 0.642 and we found the optimal parameters to be

$$M^* = 38, k^* = 9, Hidden\ layer\ size^* = 35,$$
$$Solver^*: Adam, alpha^* = 0.5$$

Next, a multi-layer perceptron with two layers have been tested with various hidden layers and $alpha$, but it did not improve the performance. Thus, we fix a single layer model, instead of multiple layers. Later, using a single layer perceptron for this optimal window, the $alpha$ has been searched more finer in the range 0.1 to 1.0 in steps of 0.1. The optimal $alpha$ is found to be 0.6, and this is used as the final best model. Thus, clearly this model is the best among the regression models tested. Using an Adam optimizer, the learning rate of constant or adaptive does not apply. Using accelerations, the best score for $M = 39$ was 0.618 obtained at hidden layer size of 50 and $alpha$ of 15. Even the variants of acceleration have shown lower scores than using just speeds, and hence not analyzed any further.

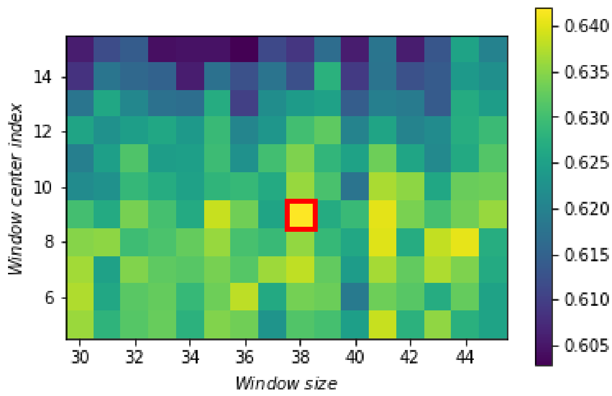

Fig.5 Cross-validation score using only speeds with a multi-layer perceptron model using one layer

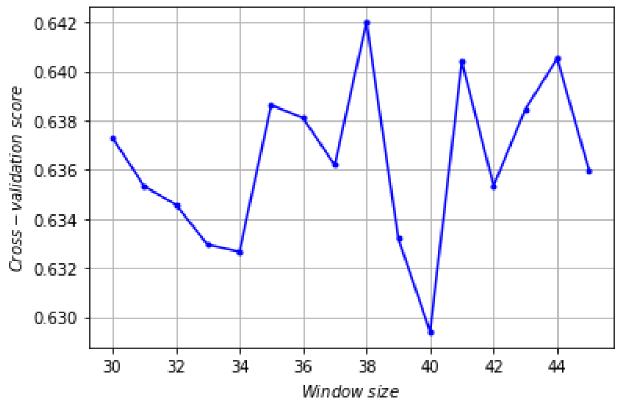

Fig.6 Maximum cross-validation scores using an MLP model for each window size

Fig.5 shows a 2-D plot of the cross-validation scores for each window size and center index. The window parameters with the highest cross-validation score is highlighted with a red square. For most of the window sizes, there is a unique index where the score reaches a local maximum, and that point is identified and used in **Fig.6** which shows the maximum cross-validation score for each window size.

Interpreting the optimal window choice in time indices, to estimate the fuel consumption at an instant n, we use the speeds in the window duration $[n-9-38/2, n-9+38/2-1] = [n-28, n+9]$. This means, in a physical sense, to estimate the fuel consumption in the time interval $[t, t+2]$, we used the information of speeds in the period $[t-28/15, t+9/15] = [t-1.867, t+0.6]$, where t is in seconds. This shift in the window of the dependence could possibly be due to the actual physics of the way the fuel is consumed to drive the vehicle at a particular speed. This method has helped to identify the optimal delay of this dependence with strong evidence. There is a remarkable similarity in the optimal window using the two approaches of random forest and an MLP model.

Finally, we see the prediction plots of the fuel consumption using the best MLP model for each of the five splits of the cross-validation in **Fig.7**.

It can be inferred that the estimate deviates greatly from the actual consumption in some cases. The rough shape of the consumption has been predicted well, but the magnitude of it has deviated, and mostly underpredicted. This could possibly be due to two reasons. First, we ignored the effects of several other parameters which contribute to the fuel consumption. This model can be easily adopted with any additional parameters too which can be measured, appending them to the list of input features.

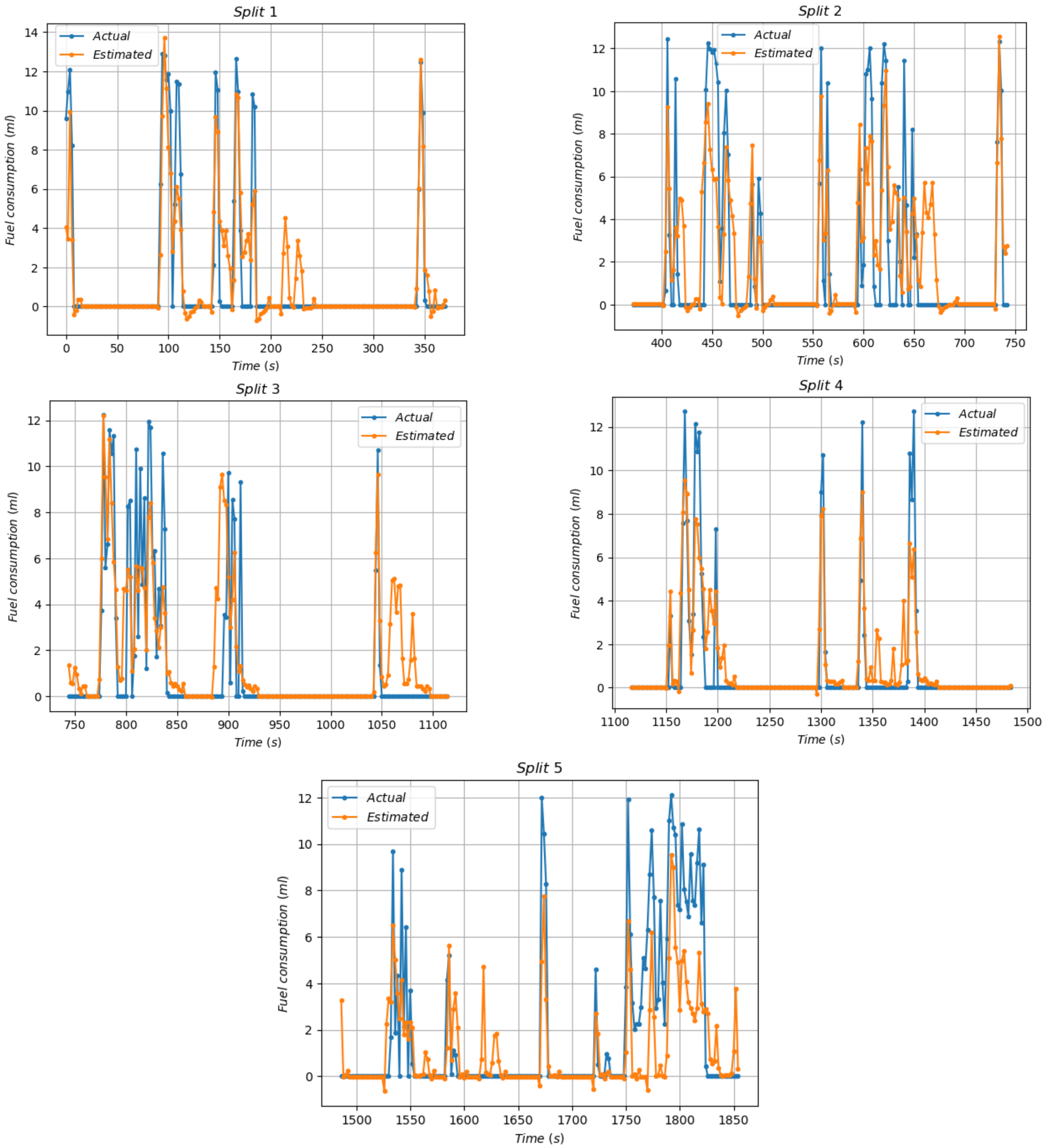

Fig.7 Actual and estimated fuel consumption using the best MLP model for each cross-validation split

Another reason is that the training data was not sufficiently large to observe various driving characteristics. Splits 1, 3 and 4 involved relatively smoother rides with little change in acceleration, making the estimation more reliable. On the other hand, splits 2 and 5 have many abrupt changes in the acceleration. Hence, training with more such driving dynamics certainly helps to increase the estimation score.

6. Applications

There are two practical applications where such a model can be potentially used– J-credit and eco-driving.

6.1 J-Credit

The J-Credit (Japan Credit) Scheme [10] is designed to certify the amount of GHG (greenhouse gas) emissions reduced and removed by sinks within Japan. Under this scheme, the government certifies the amount of greenhouse gas emissions (such as CO_2) reduced or removed by sinks through efforts to introduce energy-saving devices and manage forests, as "credit". This scheme was created by expansively integrating the Domestic Credit Scheme and the Offset Credit (J-VER; Japan's verified emissions reduction) Scheme. It is administered by the central government.

Credits created under the scheme can be used for various purposes, such as achieving the targets of the Nippon Keidanren's Commitment to a Low Carbon Society, and carbon offset.

A way in which the proposed model can be benefited using J-Credit is discussed. First, a region and duration of the estimation analysis is decided. The driving trends of a statistically significant number of vehicles is measured during the entire duration in this region and the corresponding CO_2 emissions are estimated using the model. This can be extrapolated to estimate the total vehicular emissions in that region making some assumptions. If these vehicles are to be replaced with non-emitting vehicles like electric vehicles, the total amount of emissions saved can be calculated and credited using this scheme.

6.2 Eco-driving

As mentioned earlier, these days the navigation data is often measured during transportation. Giving an estimate of the emissions using this data helps users to better understand the impact of the ride and encourages them to lower their emissions. There are many general rules for making a ride with relatively lesser emission, and it is summarized in [11].

(1) Adopt a driving style avoiding unnecessary accelerations and decelerations (braking) as this consumes more fuel.

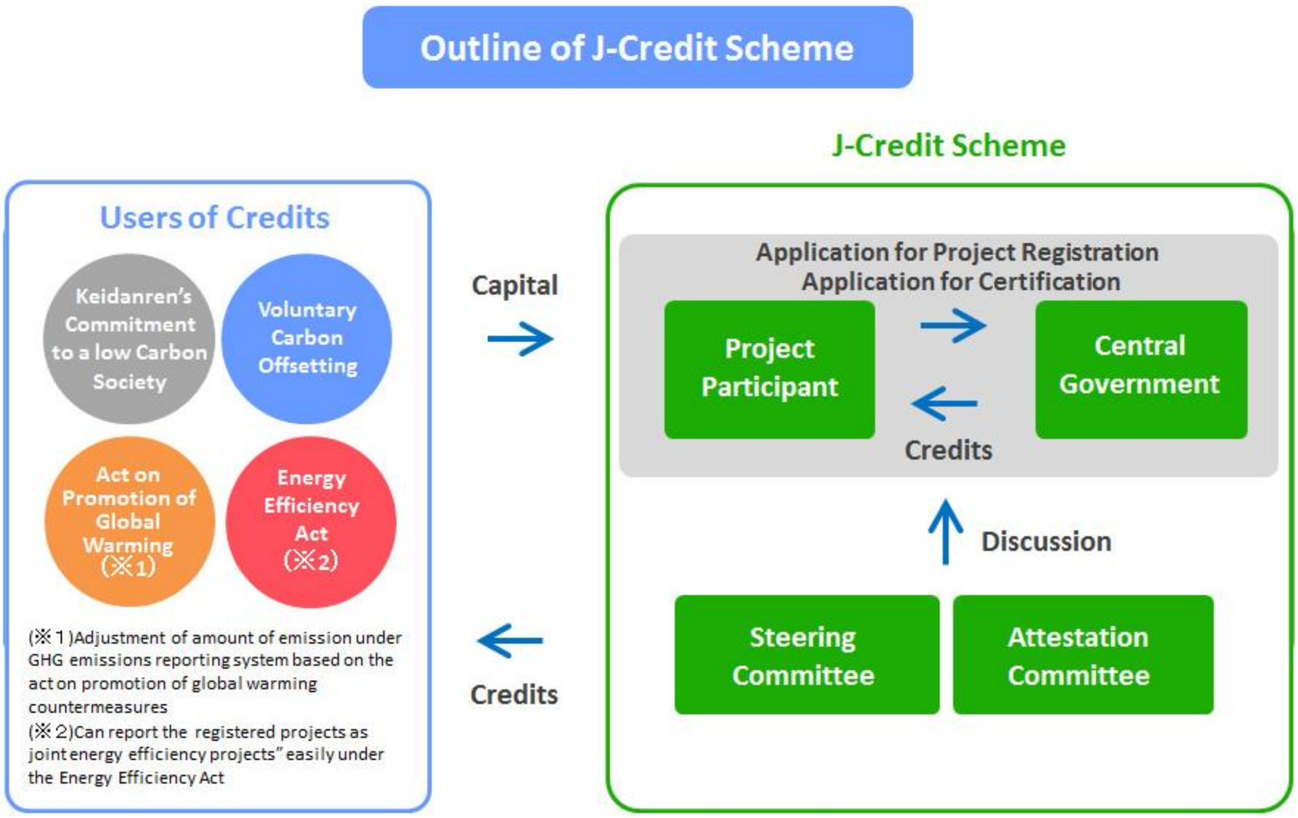

Fig.8 Outline of J-Credit scheme [10]

(2) Use the engine as efficiently as possible. As the engine load increases, the engine efficiency increases. Decreasing the speed decreases the loss due to internal friction. Thus, a combination of high loads and low speed consumes lesser fuel for the same power supplied by the engine.

An “eco index” can be formulated based on the emissions during the ride. The more the emissions in the ride, the lesser the value of the eco index. Using this index, standards on the emissions can be set and encouraged to be followed.

In this way, the proposed model can promote a convenient way to monitor and encourage drivers to make their ride more sustainable. A display on the vehicles can show the real-time emissions during the ride so that the drivers can be more aware of their profile of emissions.

7. Conclusion

The motivation for developing models for estimating the CO_2 emissions is clear not only for understanding the impacts of the emissions on the environment but also provide a scope to reduce them. The vehicle speed and acceleration have a major impact on the CO_2 emissions through the fuel consumption, and thus the driving dynamics can be used as features to estimate them. A method has been presented using a drive window for obtaining the time window of the driving dynamics which the fuel consumption largely depends on. Some machine learning models have been applied and the multi-layer perceptron is the best model which resulted a reasonably good R^2 score for practical fuel consumption estimation. The advantage of this method is that it can be extended to any other type of models and using more features influencing the fuel consumption and the results can be compared. We can use this model to gain a J-Credit and promote eco-driving. The estimations at a regional level can be done in collaboration with the municipalities. In the future, many other factors can be considered to make the estimation more accurate. Also, the correlation between CO_2 reduction and safety drive will be studied.

References

1) International Organization of Motor Vehicle Manufacturers; https://www.oica.net/category/climate-change-and-co2/ (accessed 2022.11.16)

2) Hurley; “TAPM V4 Part1: Technical description” CSIRO Marine and Atmospheric Research Paper No.25.29pp, 2008.

3) COPERT; https://www.emisia.com/utilities/copert/ (accessed 2022.11.16)

4) Zhao et al.; Modeling of vehicle CO2 emissions and signal timing analysis at a signalized intersection considering fuel vehicles and electric vehicles; European Transport Research Review 13, 5 (2021); https://doi.org/10.1186/s12544-020-00466-y

5) Cappiello et al. "A statistical model of vehicle emissions and fuel consumption," Proceedings. The IEEE 5th International Conference on Intelligent Transportation Systems, 2002, pp. 801-809, doi: 10.1109/ITSC.2002.1041322.

6) Oduro et al. (2013); CO2 vehicular emission statistical analysis with instantaneous speed and acceleration as predictor variables. 2013 International Conference on Control, Automation and Information Sciences, ICCAIS 2013. 158-163. 10.1109/ICCAIS.2013.6720547.

7) Ahn et al.; Estimating vehicle fuel consumption and emissions based on instantaneous speed and acceleration levels, Journal of Transportation Engineering, 128(2)(2002), pp.182–190

8) H.C.Frey, et al.; “Methodology for Developing Modal Emission Rates for EPAs Multi-Scale Motor Vehicle & Equipment Emission System.” Publication EPA420-R-02-027, Prepared by North Carolina State University for US Environmental Protection Agency, Ann Arbor, 2002.

9) Christoph et al.; On the use of cross-validation for time series predictor evaluation, Information Sciences, Volume 191, 2012, Pages 192-213, ISSN 0020-0255, https://doi.org/10.1016/j.ins.2011.12.028.

10) J-Credit; https://japancredit.go.jp/ (accessed 2022.11.16)

11) Adriano et al. (2012); Driving style influence on car CO2 emissions.