Temporal Dependency Analysis in Vehicle Instantaneous Energy Consumption Estimation Using Attention Weight

Shun Sakurai[1], **Ryo Nishida**[2], **Masaki Onishi**[2]
[1]Hokkaido University, Japan
[2]National Institute of Advanced Industrial Science and Technology, Japan
sakurai.shun.r1@elms.hokudai.ac.jp,
{ryo.nishida,onishi-masaki}@aist.go.jp

Abstract

In this study, we investigate the problem of the estimating vehicle instantaneous energy consumption using driving information, road gradient information, and vehicle information. An existing estimation model employs the long short-term memory (LSTM) to extract temporal features from these information, and this model shows high estimation accuracy. However, the estimation accuracy can decrease when the length of the input temporal information is increased. Thus, there is a limit to the length of the temporal information that must be input even though past driving information contributes to improving estimation accuracy. Therefore, we propose a Transformer encoder-based model to estimate vehicle instantaneous energy consumption. In the LSTM-based model, vanishing gradient occurs with each propagation because feature extraction is performed sequentially using a recursive structure. However, the proposed model enables a rigorous representation of the dependency relationship by preventing time-dependent information decay in the data because it uses an attention mechanism that computes the similarity between each time point directly. We use an open dataset which is a large-scale database for energy use of real-world vehicles, to verify the estimation accuracy of each model. Experiment results show that the LSTM-based model performs better than the Transformer-based model and the further in the past the driving information is, the smaller the attention weight becomes.

Introduction

In this paper, we address the problem of estimating vehicle energy consumption. Recently, the issue of global warming has been discussed widely; however, it has only gained recognition and shows no signs of recovery; the global average sea level and temperature continue to increase (Callery 2023b,a). A possible reason for this is that the concentration of greenhouse gases in the air, which causes global warming, is increasing year annually. Globally, the transportation sector accounts for approximately 20% of CO_2 emissions (Ritchie 2020), which is a major greenhouse gas. Accordingly, there is an urgent need to understand the amount of carbon dioxide emitted from gasoline-powered vehicles and to switch from gasoline-powered to electric vehicles (EV). However, EVs face several problems, e.g., short cruising range, long recharging time, and inadequate recharging infrastructure. Thus, it is essential to consider and estimate electricity consumption effectively when driving an EV.

Models to estimate vehicle energy consumption are referred to as emission models, and such models can be classified as macroscopic, mesoscopic, and microscopic models. Macroscopic models are used to represent a comprehensive view of energy consumption over a large area, e.g., cities and regions. This process requires a small amount of data and incurs low computational costs; however, it is inaccurate because it uses aggregate statistics, e.g., average speeds and traffic densities to determine overall traffic trends. Representative examples of macroscopic models include MOBILE6.2, which was developed by the U.S. Environmental Protection Agency, and COPERT, which was developed by the European Union (United States Environmental Protection Agency 2023b; Ntziachristos et al. 2009). Mesoscopic models divide a wide area into fixed sections. Here, energy consumption is estimated using the distance, average speed, and section entrance speed (Hanabusa et al. 2011). Finally, microscopic models represent individual vehicles or drivers, and these models can be used to estimate energy consumption per second based on vehicle driving information such as vehicle speed, acceleration, and road gradient (United States Environmental Protection Agency 2023a). However, microscopic models are computationally expensive because they require a large amount of data. Compared to macroscopic and mesoscopic models, microscopic models are more accurate due to their ability to be configured in detail. Therefore, in this study, we focus on microscopic models because they consider the different behaviors and characteristics of each vehicle in the traffic flow and can provide estimates that are closer to actual traffic conditions.

Note that microscopic emission models can be further divided into equation-based models and learning models. Equation-based models utilize approximate or variant formulas based on physical knowledge to estimate energy consumption (George, Scora and Matthew, Barth 2023). However, such models cannot respond to environmental characteristics or noise such as traffic conditions, weather conditions and the measurement error because they use the same equation for all input information. Thus, it is difficult to implement equation-based models into intelligent transport systems due to their severe limitations, e.g., excessive adjustments to the input information are required in advance. In contrast, learning models estimate energy consumption by weighting the input information using machine learning methods. They are more accurate than equation-based models because they can represent complex relationships between the input information, e.g., the relationship between speed and acceleration and energy consumption (Yingjiu, Wenshan, and Shifeng 2021). In addition, a model that uses long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997) has been proposed. This model uses driving information from the past to the estimated time (Jia, Zhang, and Chen 2022). For example, if 5 s of the temporal driving information is used, input driving information for every second from 4 s before the estimated time to the time. It has been reported that the estimation accuracy of this model is higher than that of other models, which only handles driving information at the estimated time.

Microscopic emission models that employ LSTM to incorporate temporal information have demonstrated high estimation accuracy; however, the accuracy may decrease when the length of the temporal information handled is increased. Thus, models are considered to perform best when the input length of the temporal information is a certain value even though accurate energy consumption estimation requires past driving information as well as the driving information of the estimated time.

Therefore, in this study, we investigate the necessity of information from the distant past when estimating the vehicle instantaneous energy consumption. For this purpose, we calculate attention using an emission model with a Transformer encoder, and we analyze the attention weight to verify the dependence between each observed time. In LSTM, the farther in the past the information is, the more attenuated the amount of information becomes. It is because that LSTM performs the feature extraction in a sequential manner while propagating information. On the other hand, attention can express strict dependencies between arbitrary data. It is because that there is no information decay due to the model structure as it calculates the similarity between each time directly using matrix calculations. Thus, by visualizing attention weight, we can analyze the information and times that are important to the model. Note that, temporal analysis, which we deal with, refers to the analysis of the contribution of each time to the instantaneous energy consumption estimation using temporal information from the past to the estimated time. Thus, that temporal analyzes in this study differ from general analyses, which is a temporal analysis of patterns and trends in data to predict future values.

Our contributions can be summarized as follows:

- We propose the Transformer encoder-based model, which is able to obtain more precise dependencies between each input temporal information and extract temporal features more exactly.
- Experiments reconfirmed that the past temporal information are required accurate estimation of energy consumption and that the estimation accuracy may decrease when the input length of temporal information is increased.
- We investigate how past temporal information contributes to the estimation of instantaneous energy consumption by analyzing the attention weight of the Transformer encoder-based model.

Related Work

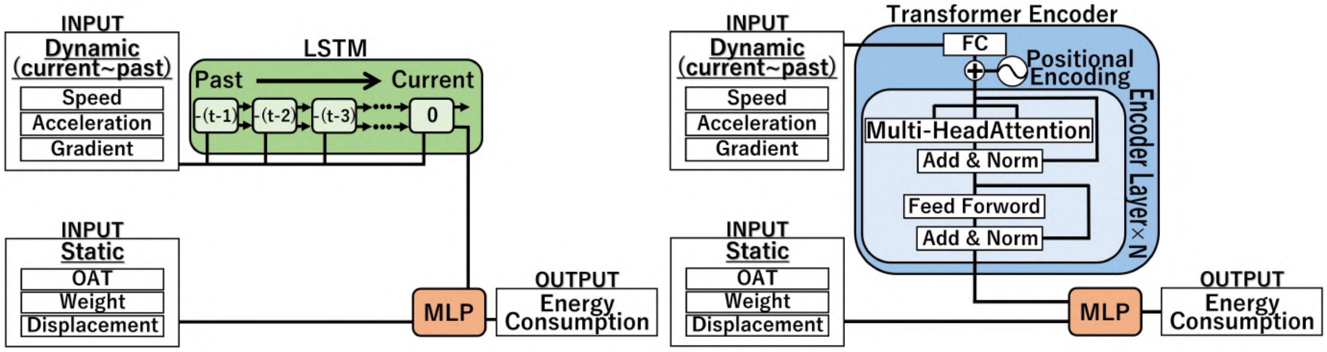
In the following, we describe learning-based emission models that handle temporal information.

The emission model proposed by Jia et al. has demonstrated high accuracy because it is a learning model that utilizes LSTM to process temporal information (Jia, Zhang, and Chen 2022). This model divides the input information into three parts, i.e., static information, dynamic information at the estimated time, and dynamic information in the past. The static information includes the day of the week, weather, and fuel type, and the dynamic information at the estimated time includes speed, acceleration, and driving environment. Finally, the historical dynamic information includes speed and acceleration. Note that each of parts is processed separately, and LSTM is employed to process the historical dynamic information. LSTM employs a gating mechanism and there are input gates, forget gates, and output gates, each of which controls the flow of information by generating values ranging from 0-1 through a sigmoid function. By discarding information in this manner, only relevant information is propagated. This process is repeated sequentially, beginning from past information up to the estimated time information, in order to output information as it takes into account temporalities. Then, the energy consumption is estimated by weighting the results of processing each of the three parts of the process.

Heran et al. proposed an emission model that uses a Transformer to process temporal information (Shen et al. 2022). This model determines the point at which the vehicle must next decelerate and the distance required for deceleration based on relevant location and road information, e.g., the speed limit. When the distance to the target point is sufficient, the Transformer is used to estimate the next speed based on the information on previous speeds. When the distance is insufficient, the speed is estimated using a Markov chain Monte Carlo method. Here, a velocity trajectory is created by repeating these steps. Using it and the vehicle information, the traction force is calculated and the energy consumption is estimated. However, the model proposed by Heran et al. is a mesoscopic emission model that calculates energy consumption over a fixed section. In addition, the energy consumption calculations implemented in the model are based on approximate formulas even though it utilizes the Transformer to predict future speeds from past speeds and estimate energy consumption.

Emission Models

Here, we describe the structure of the emission model. In this study, we introduce emission models that use dynamic information $\mathbf{X}_t$ and static information $\mathbf{z}$ as the input information to estimate the instantaneous energy consumption of a vehicle y_t as a function satisfying y_t=f($\mathbf{X}_t$,$\mathbf{z}$). The dynamic information comprises $\mathbf{X}_t$, which is a sequence of $\mathbf{x}_t$ containing three pieces of driving information at the estimated time t, i.e., speed v, acceleration a, and road gradient g, arranged each second from the past $t-(l-1)$ to the estimated



(a) LSTM-based model

(b) Transformer encoder-based model

Figure 1: Architecture of emission model

time t ($\mathbf{x}_t$=$\{v_t,\, a_t,\, g_t\}$, $\mathbf{X}_t$=$\{\mathbf{x}_{t-(l-1)}$,......, $\mathbf{x}_{t-1}$, $\mathbf{x}_t\}$, where l is the temporal length of the input). The static information $\mathbf{z}$ comprises the outside air temperature, vehicle weight, and vehicle displacement. The temporal features extracted from the dynamic information are merged with the static information and weighted to estimate the instantaneous energy consumption. In this study, as shown in Figure 1, we use a LSTM-based model (a) and a Transformer encoder-based model (b) to extract the temporal features from the dynamic information.

LSTM

As shown in Figure 1(a), the LSTM-based model processes the dynamic information from the past to the estimated time t using LSTM and weights the output and static information using the multilayer perceptron (MLP) to estimate energy consumption. The LSTM-based model differs from the model proposed by Jia et al in that the dynamic information from the past to the estimated time t is input to LSTM rather than only the past dynamic information being input to LSTM (Jia, Zhang, and Chen 2022). After LSTM processes the temporal information, the final state of the hidden layer is passed to the MLP, which processes it together with the static information to estimate the energy consumption. Note that the vanishing gradient occurs with each propagation and the information decays in proportion to the distance between the data because LSTM performs feature extraction in sequence using a recursive structure.

Transformer encoder

As shown in Figure 1(b), this model structure is similar to that of the LSTM-based model, except for the processing of dynamic information. The Transformer encoder-based model uses the Transformer encoder to extract temporal features from dynamic information. The Transformer is based on an attention mechanism (Vaswani et al. 2017). The Transformer encoder performs positional encoding on the input and passes it through an encoder layer multiple times to reflect the temporal nature of the data. Here, the data representing only the estimated time are extracted from the output of the last encoder layer and then these data and static information are weighted using the MLP to estimate the energy consumption. Dynamic information consisting of speed v, acceleration a, and road gradient g, arranged each second from the past $t-(l-1)$ to the estimated time t is processed as follows. First, positional encoding is performed to add ordinal information, i.e., temporal information to the input information because the order of the input cannot be represented by the structure of the model alone. Then, the encoder layer employs the multi-head attention mechanism to process the temporal information. The multi-head attention has several different attention heads $\{H_i \mid i = 1, \ldots, U\}$, where U is the number of heads and can extract various feature patterns. First, data with three different roles, Query, Key, and Value, are created by multiplying the input by different weights. Query specifies the information to focus on, Key expresses the relevance and similarity to other times, and Value expresses the information at each time. The similarity between each time is calculated by the inner product of Query and Key, and the normalized result is referred to as attention weight. The inner product of attention weight and Value produces data that reflect the information at other times. There is no information decay depending on the time of the data because the similarity between each time is calculated directly. Thus, we can obtain more precise dependencies between each data and extract temporal features more exactly. Note that the Transformer typically includes a decoder, which calculates the dependencies between the output of the encoder and the past output of the decoder to derive the output at the estimated time. However, in this study, we utilize only the encoder because we employ the Transformer to extract the temporal features from the input information.

Experiment

In this study, we investigate how past temporal information contributes to the estimation of instantaneous energy consumption by analyzing the learned attention weight of the Transformer encoder-based model. The length of the temporal information input to the model is considered to be limited even though driving information of several steps in the past contributes to the accuracy of energy consumption estimation. The recursive structure of LSTM causes information to decay in proportion to the distance between the data. In contrast, the Transformer does not cause the information decay dependent on the distance between the data because the similarity is calculated between arbitrary data using the attention mechanism. In this study, we employed the Transformer encoder-based model to examine the temporal dependence between each time by analyzing the attention weight. The experimental dataset, model training setup, evaluation metrics, and experimental results are described in the following sections.

Dataset

In this experiment, we used the open dataset called the Extended Vehicle Energy Dataset (eVED) (Zhang et al. 2022). The eVED is an extension of the Vehicle Energy Dataset (VED) (Oh, Leblanc, and Peng 2022). The data in the VED were collected by monitoring 383 vehicles in Michigan, USA, for approximately one year (from November 2017 to November 2018). The measured information includes driving information, vehicle information, and environmental information. The dataset consists of four energy types of vehicles, i.e., gasoline, hybrid, plug-in hybrid, and EV. However, only the EV data were used in this experiment. A total of 284,057s of EV data measured over 501 trips was split into training, validation, and test data at a ratio of 3:1:1. The model inputs were the temporal information including speed, acceleration, and road gradient, as well as static information, including the outside air temperature, vehicle weight, and displacement. The temporal information include dynamic information of each second from the past time $t-(l-1)$ to the estimated time t and experiments are conducted with four temporal lengths l (5, 10, 15 and 20). The output was the energy consumption.

Baseline

In addition to the LSTM-based and Transformer encoder-based models, we use the MLP-based model that does not input the past dynamic information as a comparison method to investigate the necessity of processing the temporal information. The MLP-based model outputs energy consumption using dynamic information including speed, acceleration, and road gradient at the estimated time t only, as well as static information including outside temperature, vehicle weight, and vehicle displacement.

Model Settings

These experiments were conducted using three models; the MLP-based model, which processes information only at the estimated time, the LSTM-based model, which processes information from the past to the estimated time using LSTM, and the Transformer encoder-based model, which processes information from the past to the estimated time using the Transformer encoder. The adjustable parameters and fixed parameters of each model are described as follows. The hyperparameters that are commonly tuned for all models include the number and size of the hidden layers in the MLP, the learning rate, and the L2 regularization parameter. In addition, the LSTM-based model and the Transformer encoder-based model have hyperparameters for the number and size of the hidden layers in the temporal processing part, which were also adjusted in this experiment. In addition, the batch size for all models was fixed at 64, the number of dimensions and heads U of the Transformer encoder-based model were set to 512 and 8, respectively, and these were not adjusted in this experiment. The mean square error (MSE) was used as the training loss function and the L2 regularization term was employed to prevent overfitting. The loss function is calculated as follows:

$$\mathcal{L}_{MSE} = \frac{1}{(\sum_{i=1}^{M} T_i)} \sum_{i=1}^{M} \sum_{t=1}^{T_i} (y_{i,t} - \hat{y}_{i,t})^2 \quad (1)$$

$$\mathcal{L} = \mathcal{L}_{MSE} + \frac{\lambda}{2M} \sum_{i=1} |w_i|^2 \quad (2)$$

where M is the number of trips, T_i is the travel time of trip i, $y_{i,t}$ is the actual energy consumption of trip i at time t, $\hat{y}_{i,t}$ is the predicted value, w_i is the ith weight of the network, and λ is the L2 regularization parameter. Models were trained on the training data, and the hyperparameters were adjusted on the validation data to minimize the loss function.

Evaluation Metrics

The trained models were evaluated on the test data. In this study, four evaluation metrics were used to evaluate the models, the root mean square error (RMSE), the mean absolute error (MAE), the R^2 value, and the mean absolute percentage error per trip ($\mathrm{MAPE}_{\mathrm{trip}}$). Note that the $\mathrm{MAPE}_{\mathrm{trip}}$ is the mean value of the MAPE calculated for each trip. These evaluation metric are calculated as follows:

$$\mathrm{RMSE} = \sqrt{\frac{1}{(\sum_{i=1}^{M} T_i)} \sum_{i=1}^{M} \sum_{t=1}^{T_i} (y_{i,t} - \hat{y}_{i,t})^2} \quad (3)$$

$$\mathrm{MAE} = \frac{1}{(\sum_{i=1}^{M} T_i)} \sum_{i=1}^{M} \sum_{t=1}^{T_i} |y_{i,t} - \hat{y}_{i,t}| \quad (4)$$

$$\mathrm{R}^2 = 1 - \frac{\sum_{i=1}^{M} \sum_{t=1}^{T_i} (y_{i,t} - \hat{y}_{i,t})^2}{\sum_{i=1}^{M} \sum_{t=1}^{T_i} (y_{i,t} - \bar{y})^2} \quad (5)$$

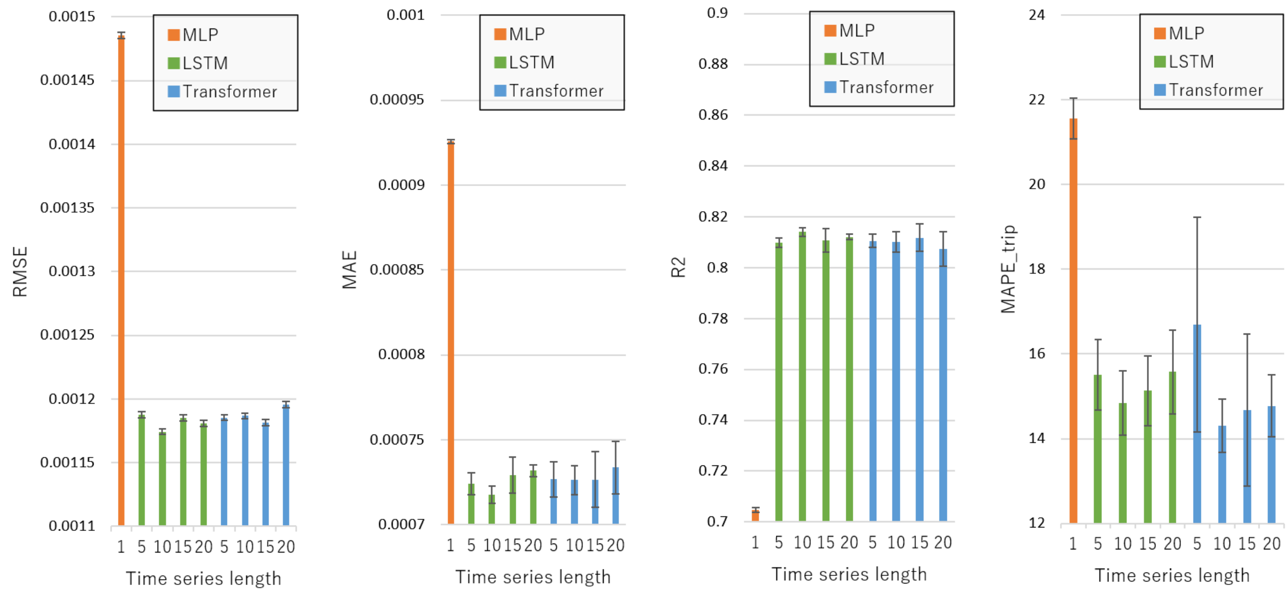
$$\mathrm{MAPE}_{trip} = \frac{100}{M} \sum_{i=1}^{M} \left| \frac{\sum_{t=1}^{T_i} (y_{i,t} - \hat{y}_{i,t})}{\sum_{t=1}^{T_i} y_{i,t}} \right| \quad (6)$$

where $\bar{y}$ is the average actual energy consumption of each trip.

Results

In this section, we describe the evaluation performance comparison of each models, and visualize attention weights of the Transformer encoder-based model to analyze how past temporal information contributes to the vehicle energy consumption estimation.

First, we discuss the effect of temporal information on accuracy based on the experimental results in terms of each evaluation metric for each of the models. Figure 2 plots the values of the MLP-based, LSTM-based, and the Transformer encoder-based models. Here, the LSTM-based and



(a) RMSE (↓) (b) MAE (↓) (c) R^2 (↑) (d) $\mathrm{MAPE}_{\mathrm{trip}}$ (↓)

Figure 2: Metric results for each model

Transformer encoder-based models have four different temporal lengths l (5, 10, 15, and 20), which are listed from left to right. As shown in this figure, the MLP-based model exhibits inferior results compared to the other models in terms of all metrics, which indicates that the past information are required accurate estimation of energy consumption. In addition, Figure 2 shows that the LSTM-based and Transformer encoder-based models are more accurate when the temporal length was set to 10 or 15 than when the temporal length was set to 20. These results indicate that there is an appropriate length of time required for estimation because the accuracy may not be improved even if the temporal length is increased. Thus, we also compared the temporal length that each model can handle. The values of each metric indicate that the LSTM-based model is the most accurate when the temporal length is set to 10, and the Transformer encoder-based model is most accurate when the temporal length is set to 10 or 15. These results show that the Transformer can handle longer temporal lengths than LSTM when estimating the vehicle instantaneous energy consumption. This is because LSTM causes vanishing gradient due to sequential computation, in addition to the fact that the relationship with the information at the estimated time t decreases as the time of information moves away from the estimated time t.

Note that the LSTM-based model performs better than the Transformer encoder-based model when looking closely at the estimation accuracy of Figure 2. For example, comparing the LSTM-based and the Transformer encoder-based model's four evaluation metrics values when the input length of the temporal information is 10, the LSTM-based model performs better in three of the evaluation metrics except MAPE_{trip}. The results suggest that information control using LSTM gates is superior to the Transformer's attention mechanism in terms of handling temporal information when estimating vehicle energy consumption. The measured energy consumption per second and the LSTM-based model's prediction are shown in Figure 3. As can be seen, the values estimated by the LSTM-based model are similar to the measured values. Note that we used EV data in this experiment, which take both positive and negative values, where negative values indicate regenerative energy generated during braking and positive values indicate energy consumed.

Second, we describe the verification of the temporal dependencies between the estimated time t and all times by visualizing the attention weight of all test data. Figure 4 shows the attention weight at each time for the entire 59,718 s of test data obtained with a temporal length of 20. The color bar is assigned a color according to the value of the attention weight, where red and blue indicate high and low values, respectively, and time indicates the relative time starting from the estimated time t. As shown in Figure 4, the attention weight value is smaller for the past information, which indicates that there is little relationship between information at the estimated time and in the distant past. In addition, it suggests that the LSTM-based and Transformer encoder-based models are suitable for temporal lengths of 10-15 because the attention weight is the lowest between 10-15 s. The relationship between the information in the distant past and that at the estimated time is low; however there is a large amount of information with attention weight greater than 0. This is because the information in the distant past is not strongly related to the estimation; however, there is a temporal relationship between the information in the distant past and that at the estimated time. In addition, the weights of highly relevant information are decayed by the weights of less rele-

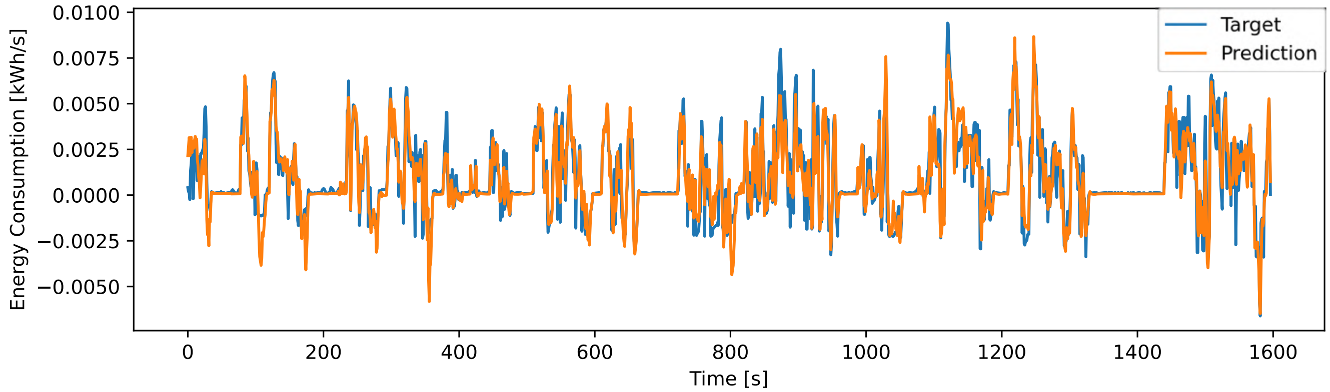


Figure 3: Correct and predicted values

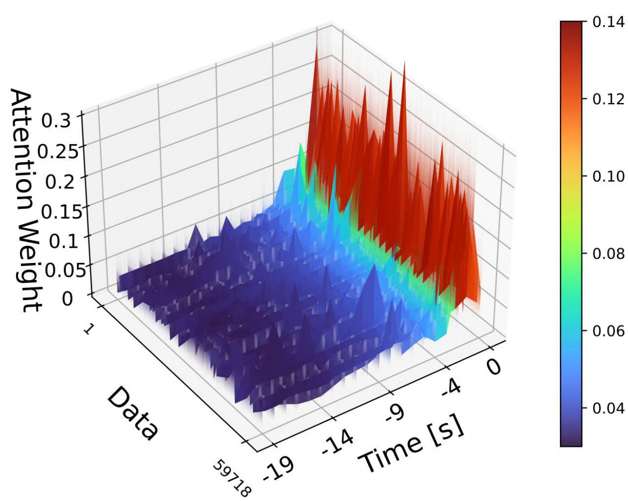


Figure 4: Attention weight

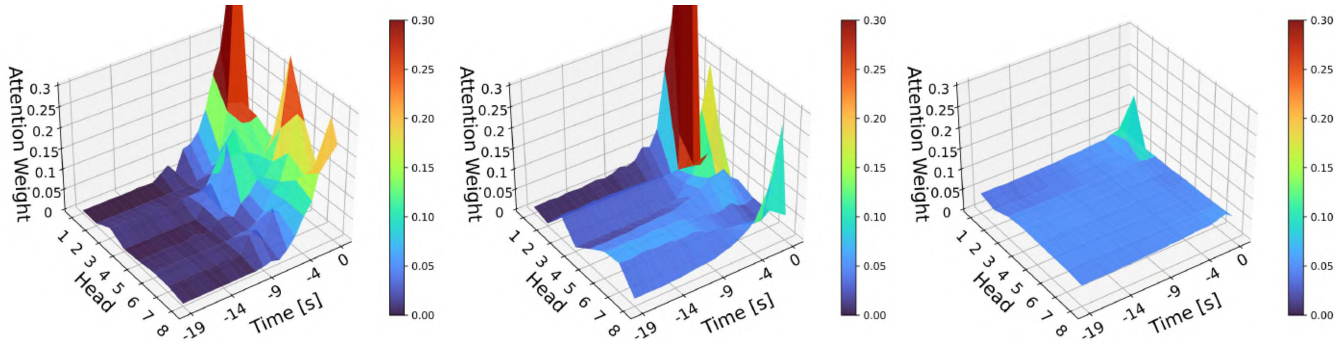
vant information because the attention weight is derived using normalization. This is what causes the loss in accuracy when the appropriate temporal length is exceeded, and it is clear that there is an upper limit to the temporal length that should be input when estimating energy consumption.

Further, we describe the results of visualizing the attention weight for each driving situation, e.g., acceleration , deceleration and steady states, and verifying the dependency between the attention weight and driving situations. Figure 5 shows the attention weight of each head in the deceleration (a), the steady state (b), the acceleration state (c), which were extracted from the test data. As we can observe, the shape of the attention weight distribution differs depending on the driving conditions; however, there is a decreasing trend toward past time, which is the same as shown in Figure 4. The weight of the deceleration state is greater closer to the estimated time, and the weight of the acceleration state tends to decrease; however, the difference in the weights is small for information at any time. Thus, it can be seen that the time that is highly related to the estimated time differs depending on the given driving situation. The steady state has distribution features for both acceleration and deceleration weights, each represented by separating the heads to be expressed. Specifically, heads $H_1, ..., H_4$ have a distribution similar to the deceleration state, and heads $H_5, ..., H_8$ have a distribution similar to the acceleration state. These results demonstrate that the Transformer encoder can express various dependencies between the driving information at the estimated time and the past driving information using the multi-head attention mechanism. On the other hand, these results indicate that information about the distant past is less necessary because the attention weight tends to decrease under all driving conditions.

Conclusion

In this study, we investigated the necessity of information from the distant past when estimating the vehicle instantaneous energy consumption. For this purpose, we proposed the Transformer encoder-based model, and analyzed the attention weight to verify the dependence between dynamic information at each time. As a result, we found that the input length of the temporal information when the model performs best exists even though using past dynamic information improves accuracy in estimating vehicle instantaneous energy consumption. In addition, the Transformer encoder-based model was able to use the multi-head attention mechanism to allow for various representations; however, the LSTM-based model performs better than The Transformer encoder-based model. This experimental results indicated that the LSTM's gates mechanism is superior to the Transformer's attention mechanism in terms of the temporal feature extraction in the estimation of vehicle instantaneous energy consumption.

We present two challenges for the future. The first is to reduce the size of the LSTM-based model without compromising its accuracy. The LSTM-based model is computationally expensive due to the sequential data-by-data processing of the temporal information from the past to the estimated time t. Thus, if the model is handled inside a microscopic transportation simulator, long simulation times and simultaneous simulation of multiple vehicles require a lightweight model. The second is the development of a physics-informed



(a) Deceleration (b) Steady (c) Acceleration

Figure 5: Evaluation value of each model

model. As explained in the introduction section, there are two types of microscopic emission models, i.e., equation-based models and learning models, and we plan to combine these model to realize improved estimation accuracy. Physics-informed models are expected to incorporate known physical laws, thereby enabling accurate predictions on limited data and reducing the risk of overfitting.

Acknowledgments

This paper is based on results obtained from a project, JPNP21028, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

References

Callery, S. 2023a. Global Temperature - Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/vital-signs/global-temperature/.

Callery, S. 2023b. Sea Level - Climate Change: Vital Signs of the Planet. https://climate.nasa.gov/vital-signs/sea-level/.

George, Scora and Matthew, Barth. 2023. Comprehensive Modal Emission Model (CMEM), Version 2.0 User ' s Guide. https://www.cert.ucr.edu/sites/default/files/2019-07/CMEM_User_Guide_v3.01d.pdf, last accessed on 05/15/2023.

Hanabusa, H.; Shiraishi, T.; Koide, K.; Horiguchi, R.; Tanaka, S.; Kuwahara, M.; Oguchi, T.; and Oneyama, H. 2011. Development of a Hybrid Traffic Simulation Framework for Environmental Evaluation for ITS Applications. In *18th ITS World CongressTransCoreITS AmericaERTICO - ITS EuropeITS Asia-Pacific*.

Hochreiter, S.; and Schmidhuber, J. 1997. Long short-term memory. *Neural computation*, 9(8): 1735–1780.

Jia, T.; Zhang, P.; and Chen, B. 2022. A Microscopic Model of Vehicle CO2 Emissions Based on Deep Learning - A Spatiotemporal Analysis of Taxicabs in Wuhan, China. *IEEE Transactions on Intelligent Transportation Systems*, 23(10): 18446–18455.

Ntziachristos, L.; Gkatzoflias, D.; Kouridis, C.; and Samaras, Z. 2009. COPERT: A European Road Transport Emission Inventory Model. In *Information Technologies in Environmental Engineering*, 491–504. Springer Berlin Heidelberg.

Oh, G.; Leblanc, D. J.; and Peng, H. 2022. Vehicle Energy Dataset (VED), A Large-Scale Dataset for Vehicle Energy Consumption Research. *IEEE Transactions on Intelligent Transportation Systems*, 23(4): 3302–3312.

Ritchie, H. 2020. Sector by sector: where do global greenhouse gas emissions come from? *Our World in Data*. Https://ourworldindata.org/ghg-emissions-by-sector.

Shen, H.; Wang, Z.; Zhou, X.; Lamantia, M.; Yang, K.; Chen, P.; and Wang, J. 2022. Electric Vehicle Velocity and Energy Consumption Predictions Using Transformer and Markov-Chain Monte Carlo. *IEEE Transactions on Transportation Electrification*, 8(3): 3836–3847.

United States Environmental Protection Agency. 2023a. United States Environmental Protection Agency. MOVES2014a user interface reference manual. https://19january2017snapshot.epa.gov/sites/production/files/2016-11/documents/420b16085.pdf, last accessed on 05/15/2023.

United States Environmental Protection Agency. 2023b. User's Guide to MOBILE6.1 and MOBILE6.2: Mobile Source Emission Model Factor-Final Version. https://nepis.epa.gov/Exe/ZyPURL.cgi?Dockey=P1001DSD.txt, last accessed on 05/09/2023.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yingjiu, P.; Wenshan, Z.; and Shifeng, N. 2021. Emission modeling for new-energy buses in real-world driving with a deep learning-based approach. *Atmospheric Pollution Research*, 12(10): 101195.

Zhang, S.; Fatih, D.; Abdulqadir, F.; Schwarz, T.; and Ma, X. 2022. Extended vehicle energy dataset (eVED): an enhanced large-scale dataset for deep learning on vehicle trip energy consumption.