

PAPER • OPEN ACCESS

## Accessing and constructing driving data to develop fuel consumption forecast model

To cite this article: Rei-Jo Yamashita *et al* 2018 *IOP Conf. Ser.: Earth Environ. Sci.* **113** 012217

View the [article online](#) for updates and enhancements.

You may also like

- [Acquisition, Modeling, and Evaluating Method of Driving Behavior Based on OBD-II: A Literature Survey](#)  
Galih Hermawan and Emir Husni
- [Eco-driving behavior tendency among Indonesian people: a preliminary study](#)  
Rida Zuraida and Dimitrius Widjaja
- [The effect of plug-in hybrid electric vehicle charging on fuel consumption and tail-pipe emissions](#)  
Ahmet Mandev, Patrick Plötz and Frances Sprei

### ECS Toyota Young Investigator Fellowship

For young professionals and scholars pursuing research in batteries, fuel cells and hydrogen, and future sustainable technologies.

At least one \$50,000 fellowship is available annually.  
More than \$1.4 million awarded since 2015!



Application deadline: January 31, 2023



TOYOTA

**Learn more. Apply today!**

# Accessing and constructing driving data to develop fuel consumption forecast model

Rei-Jo Yamashita<sup>1,a</sup>, Hsiu-Hsen Yao<sup>2,b</sup>, Shih-Wei Hung<sup>2</sup>, Acquah Hackman<sup>2,c</sup>

<sup>1</sup> Shigun Research Institute Corporation, Ibaraki, Japan

<sup>2</sup> Dept. of Computer Science and Engineering Yuan-Ze University, Taoyuan, Taiwan

<sup>a</sup> ryama@shigun.co.jp, <sup>b</sup> csyao2015@gmail.com, <sup>c</sup> acquah.hackman@utg.edu.gm

**Abstract.** In this study, we develop a forecasting models, to estimate fuel consumption based on the driving behavior, in which vehicles and routes are known. First, the driving data are collected via telematics and OBDII. Then, the driving fuel consumption formula is used to calculate the estimate fuel consumption, and driving behavior indicators are generated for analysis. Based on statistical analysis method, the driving fuel consumption forecasting model is constructed.

Some field experiment results were done in this study to generate hundreds of driving behavior indicators. Based on data mining approach, the Pearson coefficient correlation analysis is used to filter highly fuel consumption related DBIs. Only highly correlated DBI will be used in the model. These DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. We then use K-means cluster analysis to group to the driver class and the route class. Finally, more than 12 aggregate models are generated by those highly correlated DBIs, using the neural network model and regression analysis. Based on Mean Absolute Percentage Error (MAPE) to evaluate from the developed AMs. The best MAPE values among these AM is below 5%.

## 1. Introduction

The rise of the global awareness of the energy crisis becomes more important, and how to save the driving fuel consumption becomes a big issue. In order to study the features of driving fuel consumption, we consider three kinds of related factors: vehicle's factors, routes' factors, and driving behavior factors. Based on our pilot study, we found that driving behaviors are highly correlated to driving fuel consumption while one fixes vehicle factors and route factors. Thus, this study will focus on developing a driving fuel consumption forecast model based on driving behavior factors.

In this study, a forecasting models, to estimate fuel consumption is developed based on the driving behavior, in which vehicles and routes are known. First, the driving data are collected via telematics and OBDII. Then, the driving fuel consumption formula is used to calculate the estimate fuel consumption, and driving behavior indicators(DBI) are generated for analysis. Based on statistical analysis method, the driving fuel consumption forecasting model is constructed.

Some field experiment results were done in this study to generate hundreds of driving behavior indicators. Based on data mining approach, the Pearson coefficient correlation analysis is used to filter highly fuel consumption related DBIs. Only highly correlated DBI will be used in the model. These DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. We then use K-means cluster analysis to group to the driver class and the route class. Finally, more than 12 aggregate models(AM) are generated by those highly correlated DBIs, using the



neural network model and regression analysis. Based on Mean Absolute Percentage Error (MAPE) to evaluate from the developed AMs. The best MAPE values among these AM is below 5%.

Telematics and OBDII interface in vehicle are used in this study to collect driving behavior and fuel consumption data. The data from telematics include the following information, such as GPS, driving time, speed, 3 dimensional G-sensor data, angular velocity, etc.; and the data collected via OBDII consist of rpm, tp, and throttle nozzle, etc. This study follows the formula developed by research results from Industrial Technology Research Institute(itri) to do fuel consumption calculation.

Through a large number of driving data collection, this study generated a fuel consumption predict model based on driving behavior data. More than one hundred driving behavior indicators(DBIs) are generated in the pilot study and through the Pearson correlation coefficient analysis, only the highly fuel consumption correlated DBIs will be filtered out. These DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. The cluster analysis, then apply K-means method to group various classes of driver clusters and route clusters. Finally, in order to select the excellent effect model to forecast the fuel consumption through the driving behaviors, a large number of aggregate models are generated. Applying data mining approach, the study found out the best fuel consumption predict model by calculating Mean Absolute Percentage Error (MAPE). The approach selects both the neural network method and regression analysis method for these machine learning tools.

Sec 2 discusses the research background and some related studies; and in Sec 3, the research framework is addressed, including data collection method, DBIs generation criteria, the pilot study design, and ANOVA analysis schemes. Finally, in Sec 4, some study results show low MAPE features of the developed fuel consumption forecast models based on various driving behavior indicators.

## 2. Background and Related Study

Evans, L., "Driver behavior effects on fuel consumption in urban driving," Human Factors research, proposed to allow some of the participants in the ordinary traffic signal on the road driving, and hope that the driver can minimize travel time, and this The study found that a 1% reduction in travel time would cost 1.1% less than normal driving while driving, but it was also found that it would be difficult for the driver to reduce travel time because the study found that traffic was Signal or road conditions, so to effectively reduce the fuel consumption of the problem is so complicated[6].

L. Evans, R. Herman and T. Lam, "Multivariate Analysis of Traffic Factor Related to Fuel Consumption in Urban Driving," proposed an experiment in the Detroit metropolitan area to study driving for vehicle fuel consumption influences. Detailed records of the speed of travel, acceleration and fuel consumption, and multi-variable statistical techniques to analyze the results found that the average time per unit distance to explain the oil consumption is the most important factor[10].

Liu Xiang-Yu defines the calculation of the five indicators related to the fuel consumption of the driving behavior (step on the brake index, step on the throttle index, the degree of speed change, the left and right speed is too fast, the loan speed is slow and the high speed time ratio), and establish an energy saving information The platform provides a variety of reports to allow users to learn fuel-efficient knowledge[1].

Tsai-Zong Xian use statistical methods to develop the threshold of braking and acceleration, the use of two standard deviations as accelerated (step on the throttle), negative two standard deviation as a brake, as a factor for fuel consumption. As the study of the use of refueling jump and mileage to do the assessment, there will be some errors, so this study mainly use the computer to get the fuel value for analysis[3].

Lu-Jing Wen Proposed for fuel consumption between different roads, to find out the indicators of high fuel consumption between roads, and to establish a model based on the neural network[4].

Huang-Shi Wei use driving road factors to find a high correlation for fuel consumption factors to do fuel consumption prediction model of the establishment of the use of nerve and complex return to the construction of fuel consumption forecast model[5].

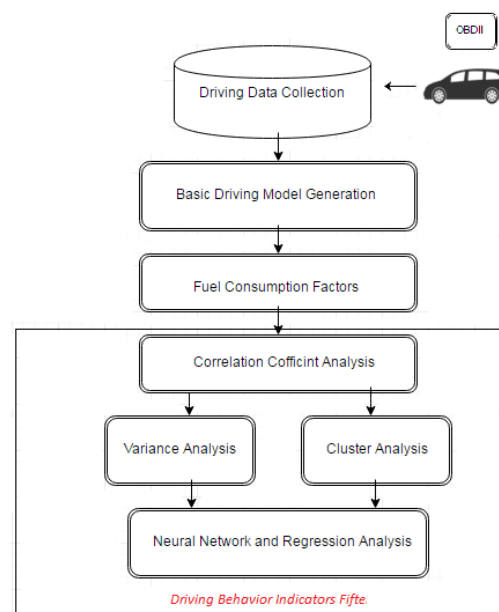
### 3. Research Schemes

The main purpose of this study is to develop some fuel consumption predict model based on the driving behavior factors; and to explore whether these factors are highly related to fuel consumption. The research process consists of six stages. First, candidate driving behavior indicators are generated for analysis purpose. Secondly, driving fuel consumption formula is developed. After that, the data mining process has been done, in which the training data provide the original DBIs. Only highly fuel consumption correlated DBIs, developed by Pearson correlation coefficient analysis, can be used in the final model. Then, the cluster analysis may generate various driver clusters and different route clusters. Finally, based on neural network model and regression analysis, the final fuel consumption forecasting model is selected.

Some field experiment results were done in this study to generate hundreds of driving behavior indicators. Based on data mining approach, the Pearson coefficient correlation analysis is used to filter highly fuel consumption related DBIs. Only highly correlated DBI will be used in the model. These DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. We then use K-means cluster analysis to group to the driver class and the route class. Finally, more than 12 aggregate models(AM) are generated by those highly correlated DBIs, using the neural network model and regression analysis. Based on Mean Absolute Percentage Error (MAPE) to evaluate from the developed AMs.

Telematics and OBDII data are used to collect driving behavior(including step-on-the-brakes, step-on-the-throttle, and steering-wheel-handle) and fuel consumption data. The data from telematics include GPS data, driving time, speed, 3 dimensional G-sensor data, angular velocity, etc.; and the data collected via OBDII consist of distance, time, round-per-minute, throttle-opening, and throttle-nozzle,...etc. This study follows the formula developed by research results from Industrial Technology Research Institute(ITRI) to do fuel consumption calculation.

In the pilot study, the following five driving behavior indicators are developed to check the correlation of driving fuel consumption.



**Fig. 3-1** Research Flow

#### Driving Behavior Indicators

1. SAA/D: *Sum of Absolute Values of Acceleration Per Kilometer*  $\Sigma_t | \text{Acceleration}(t) |$  /Total Distance

2. CAT/D: *Count of the Seconds of High Acceleration Per Kilometer*  $\text{Count}_t ( | \text{Acceleration}(t) | > \text{Threshold} ) / \text{Total Distance}$
3. SADS/D : *Sum of Absolute Values of Speed Change Per Kilometer*  $\sum_t | \text{Speed}(t+1) - \text{Speed}(t) | / \text{Total Distance}$
4. CDST/D: *Count of the Seconds of Speed Change Per Kilometer*  $\text{Count}_t ( | \text{Speed}(t+1) - \text{Speed}(t) | > \text{Threshold} ) / \text{Total Distance}$
5. SRT/D: *Sum of Round-Per-Minute \* Throttle-oPening Per Kilometer*  $\sum_t (\text{rpm}(t) * \text{tp}(t)) / \text{Total Distance}$

Using the data from telematics and the OBDII information, according to the speed grouping and the establishment of driving behavior derivative variables to carry out fuel consumption analysis, to explore the sections and the times of the fuel consumption changes and regression analysis using a variety of fuel consumption estimation formula and the establishment of phase corresponding analysis model, so that users can move by their own driving to estimate traffic fuel consumption.

Through a large number of driving data collection, this study generated a fuel consumption predict model based on driving behavior data. More than one hundred driving behavior indicators are generated in the pilot study and through the Pearson correlation coefficient analysis, only the highly fuel consumption correlated DBIs will be filtered out. These DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. The cluster analysis, then apply K-means method to group various classes of driver clusters and route clusters. Finally, in order to select the excellent effect model to forecast the fuel consumption through the driving behaviors, a large number of aggregate models are generated. Applying data mining approach, the study found out the best fuel consumption predict model by calculating Mean Absolute Percentage Error (MAPE). The approach selects both the neural network method and regression analysis method for these machine learning tools.

Applying data mining method to find the fuel consumption correlated models, the data is divided into training data and testing data. The training data is used to build fuel consumption forecast model, and the training data is used to find out the best fuel consumption prediction model by calculating MAPE values from these training data.

#### 4. Analysis of Variance Between Driving Behavior and Fuel Consumption

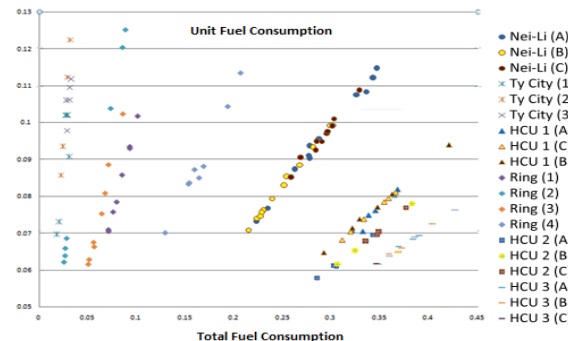
In this study more than one hundred driving behavior indicators are first generated. Through the Pearson correlation coefficient analysis, only 13 highly fuel consumption correlated DBIs are remained. These 13 DBIs are divided into four classes: speed class, acceleration class, Left/Right/U-turn class and the other category. The cluster analysis, then apply K-means method to group into various driver clusters and route clusters. Finally, in order to forecast the fuel consumption through the driving behaviors, a large number of aggregate models(AM) are generated. The study developed some fuel consumption predict model and by calculating Mean Absolute Percentage Error (MAPE) to find the best. The approach selects both the neural network method and regression analysis method.

##### 4.1 Experiments

When design experiments, for different sections and different driving experiments and analysis, designed the following experiment. There are more lane and traffic lights on the campus around the road and there are more traffic areas of the main road; at the same time, participate in the driving test must be completed on the five roads at least once the driving record. At the end of the experiment, suburban road line 1 completed 12 experiments, suburban road line 2 completed 18 experiments, suburban road line 3 to complete 12 experiments, the campus road around the completion of 33 experiments, and the main road completed 9 times experiment.

In each experiment, vehicles' types are fixed, and in different routes, several times of driving data are collected, one may find in Fig. 4-1, even fix vehicle factors and route factors, the driving fuel

consumption are variant. Different driving behavior will cause the variant distribution of fuel consumption. In the same route, the maximum difference of fuel consumption can be more than double.



**Fig. 4-1** Fuel Consumption/Distance vs. Fuel Consumption

#### 4.2 Correlation Coefficient Analysis

Based on the analysis of the relationships between fuel consumption and the driving behavior indicators selected by the previous section and the Pearson correlation coefficient of the unit fuel consumption is used to filter the original driving behavior indicators, which are more than 87 indicators.

After the correlation analysis, it can be seen that the DBIs of the acceleration and the speed are the most before and after the index, and the index of the turning is less. Therefore, it can be seen that the speed and the acceleration are obviously related to the unit fuel consumption high.

In the speed class: low speed, average speed, speed average speed for the highly relevant, and variations of speed for the high degree of correlation;

In the acceleration class: more than threshold, jerk, total acceleration, accelerator more than threshold for very high correlation ; In the turning class only vector acceleration to achieve a high degree of correlation, the other are only common relevance.

**Table 4-1** Fuel Consumption and Driving Behavior(DB) Correlation Coefficient Analysis  
(A= Acceleration Class, S=Speed Class, T=Turning Class)

ID	DB Indicator	Class	R <sup>2</sup>
01	Count Amount of Accelerator more than threshold	A	0.89
02	Sum of Absolute Value of Acceleration Change Rate	A	0.88
03	Sum of Absolute Value of Acceleration	A	0.88
04	Count Amount of Braking more than threshold	A	0.83
05	Sum of Low Speed Values	S	0.89
06	Mean Speed Values	S	0.82
07	Sum of Driving Average Speed Values	S	0.84
08	Sum of Variations of Speed	S	null
09	Sum of High Speed Values	S	0.75
10	Sum of Vector Acceleration	T	0.84
11	Sum of Curvature	T	0.68
12	Sum of Count Amount of Sharp Turn	T	0.58
13	Sum of Absolute Value of Horizontal Acceleration Change Rate	T	0.53

#### 4.3 Forecasting Fuel Consumption Based on Driving Behavior by Neural Network Analysis

The fuel consumption index found in the previous section is divided into the speed-related fuel consumption index, the acceleration-related fuel consumption index, the turning fuel consumption index, the variance analysis road type fuel consumption index, the variance analysis total fuel consumption index and the whole fuel consumption index , These six categories of indicators were



placed into the class of neural network model to predict fuel consumption, to compare the six types of indicators of fuel consumption differences.

The comparison method uses the mean absolute percentage error (MAPE) and the mean square error root (MSE) to compare the degree of deviation between the complex and the neural network in the actual value and the predicted value. The smaller the value, the higher the accuracy of the model prediction, the closer the MAPE is to 0, the better the estimation effect. The predictive performance formula, when the MAPE value is less than 10% on behalf of its ability to predict high accuracy, 10 to 20% is good, 20 to 50% is reasonable, and more than 50% is not correct.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|x_i - x_1|}{x_i} \times 100$$

MAPE = Mean Absolute Percentage Error

This study is to use this kind of neural network model to construct the oil prediction model, and the indexes are regarded as the independent variables, and the unit fuel consumption is the variable. All the experimental data are divided into training data, Number of posts: 27 pen, the use of the hidden layer of the neural network as a layer (as shown below), and the final output for the predicted unit fuel consumption.

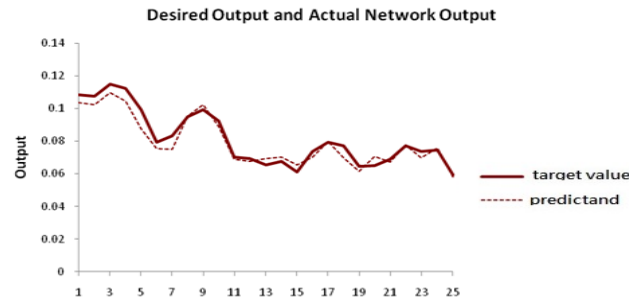
In the fuel consumption forecast model for rate is very large, which maybe affected by the road, and some of the error rate is also large, maybe the reason for driving alone, so in this section will be driving and road as a category of information into the fuel consumption forecast model to predict fuel consumption model.

Speed Factor Model takes the Driving Average Speed, Average Speed, Low Speed and High Speed (Speed Class) as input, Acceleration Factor Model for Total Acceleration, Accelerator more than threshold and braking more than threshold (Acceleration Class) as input; The First Class Factor Model takes Enter Speed and Low Speed, Total Acceleration, and Vector Acceleration as input. The Second Class Factor Model is the input of Driving Average Speed, High Speed, Braking more than Threshold, Step throttle more than threshold with Jerk as input; All Factors Model for the First Class Factor Model and Second Class Factor Model as input, the best results, the error rate for the above six inputs for the lowest, as shown below.

**Table 4-2** Neural Network Model  
(C= whether Category or not)

Model ID	Model Name	Category	MAPE
N12	All DBI Factor Model	C	4.69
N08	Acceleration Factor Model	C	5.15
N11	Second Class Factor Model	C	5.18
N13	All Factors(Group Category) Model	C	5.24
N10	First Class Factor Model	C	5.25
N06	All DBI Factor Model	NC	5.54
N05	Second Class Factor Model	NC	5.57
N04	First Class Factor Model	NC	5.96
N02	Acceleration Factor Model	NC	6.85
N07	Speed Factor Model	C	7.18
N09	Turning Factor Model	C	8.3

N01	Speed Factor Model	NC	10.84
N03	Turning Factor Model	NC	14.55



**Fig. 4-2** Actual Results and Predict Results based on Neural Network Model-All Factors(MAPE : 4.69%)

## 5. Conclusion

To get the best forecast unit fuel consumption, one may use the fuel consumption factor plus the category model, and if you want to expand to other roads and other driving should use the "fuel consumption factor plus classification group" will come to use only fuel consumption factor to predict the model. Okay The fuel consumption model is chosen to have a significant influence on the driving and the factors that have significant influence on the road are used to predict the fuel consumption model.

In the absence of categorical data, the best model of the neural network is used as input for all fuel consumption relationships, while the complex regression analysis is a high factor in the choice of fuel consumption factors with significant effects on driving and roads.

## References

- [1] H.Y, Liu. 2010. The Study of Automobile Energy Reduction Digital Teaching System. Master's thesis. Taoyuan: Yuan Ze University Information Engineering
- [2] C.C, Wu. 2010. A Study of Driving Characteristics of Vehicles Movement Master's thesis. Taoyuan: Yuan Ze University Information Engineering
- [3] Z.X, Tsai. 2008. A study of oil consuming factors through driving behavior and personality using digital tachograph and questionnaire. Master's thesis. Taipei : Taipei University Department of Statistics
- [4] J.W, Lu. 2010. Applying data mining techniques on the fuel consumption analysis model of driving a car. Master's thesis. Taoyuan: Yuan Ze University Information Engineering
- [5] S.W, Huang. 2011. A study of fuel consumption analysis by using the driving record. Master's thesis. Taoyuan: Yuan Ze University Information Engineering
- [6] Evans,L,"Driver behavior effects on fuel consumption in urban driving", Human Factors, Vo1.21, No.4, pp.389-398, 1979
- [7] Itakura, "Driver Modeling Based on Driving Behavior and its Evaluation in Driver Identification", Proc. IEEE, vol. 95, no. 2, pp. 427-437., 2007
- [8] M.F. Chang, R. Herman, "An Attempt to Characterize Traffic in Metropolitan Areas" Transportation Science, Vol.12, No.1, 1978
- [9] E. Ericsson, "Independent Driving Pattern Factors and their Influence on Fuel-Use and Exhaust Emission Factors" Transportation Research Part D, Vol.6, No.5, 2001, pp.325-345.
- [10] L. Evans, R. Herman and T. Lam, "Multivariate Analysis of Traffic Factor Related to Fuel Consumption in Urban Driving" Transportation Science, Vol.10, No.2, May 1976
- [11] L. Evans, R. Herman, "Automobile Fuel Economy on Fixed Urban Driving Schedule" Transportation Science, Vol.12, No.2, May 1978



- [12] J. N. Hooker, “Optimal Driving for Single-Vehicle Fuel Economy” Transportation Research Part A, Vol.22, No.3, 1988, pp.183-201.
- [13] R. Herman, S. Ardekani, “The Influence of Stops on Vehicle Fuel Consumption in Urban Traffic” Transportation Science, Vol.19, No.1, 1985