View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by AIS Electronic Library (AISeL)

Association for Information Systems
AIS Electronic Library (AISeL)

MWAIS 2019 Proceedings

Midwest (MWAIS)

Spring 5-21-2019

Comparison of Data Mining and Mathematical Models for Estimating Fuel Consumption of Passenger Vehicles

Mehmet Sevkli
American University of Middle East, mehmet.sevkli@aum.edu.kw

Aise Zulal Sevkli
Denison University, sevklia@denison.edu

Oslem Cosgun
Dakota State University, ozlem.cosgun@dsu.edu

Follow this and additional works at: https://aisel.aisnet.org/mwais2019

Recommended Citation

Sevkli, Mehmet; Sevkli, Aise Zulal; and Cosgun, Oslem, "Comparison of Data Mining and Mathematical Models for Estimating Fuel Consumption of Passenger Vehicles" (2019). *MWAIS 2019 Proceedings*. 34.
https://aisel.aisnet.org/mwais2019/34

This material is brought to you by the Midwest (MWAIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in MWAIS 2019 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

Comparison of Data Mining and Mathematical Models for Estimating Fuel Consumption of Passenger Vehicles

Mehmet Sevkli
American University of Middle East, Industrial Engineering, Kuwait,
mehmet.sevkli@aum.edu.kw

Aise Zulal Sevkli
Denison University, Computer Science, Granville, OH, USA,
sevklia@denison.edu

Ozlem Cosgun
Dakota State University, College of Business and Information Systems, Madison, SD, USA,
ozlem.cosgun@dsu.edu

ABSTRACT

A number of analytical models have been described in the literature to estimate the fuel consumption of vehicles, most of which require a wide range of vehicle and trip related parameters as input data, which might limit the practical applicability of these models if such data were not readily available. To overcome this drawback, this study describes the development of three data mining models to estimate fuel consumption of a vehicle, including linear regression, artificial neural network and support vector machines. The paper presents comparison results with five instantaneous fuel consumption models from the literature using real data collected from three passenger vehicles on three routes. The results indicate that while the prediction accuracy of the instantaneous fuel consumption models varies across the data sets, those obtained by the regression models are significantly better and more robust against changes in input data.

Keywords

Fuel consumption models, Support vector machine, Artificial neural network, Linear regression

INTRODUCTION

Transportation uses more than one third of global energy [Ahn and Rakha, 2008; Rakha et al. 2011] that is obtained from petroleum products. Fuel consumption in vehicles results in greenhouse gas emissions, in particular CO2, with detrimental effects on the environment. The combustion of gasoline and diesel in vehicle engines is the second largest source of CO2 emissions, the amount of which is proportional to the amount of fuel consumed by the engine. The rate at which fuel is consumed by an engine depends on a wide range of factors, including those that are vehicle specific (such as the engine speed or vehicle mass) and those that are trip specific (such as rolling resistance or road gradient). A number of fuel consumption models have been developed in the literature that differ from one another in the way in which they incorporate these factors into the calculation.

This paper contributes to the literature by proposing data mining models for use in estimating fuel consumption of passenger vehicles and shows that they provide more accurate estimations as compared to some of the available models in the literature. The data mining models we develop range from simple linear regression to more sophisticated models, such as artificial neural networks and support vector machines. We also present real data collected from a passenger vehicle using existing on-board diagnostics (OBD2), a Bluetooth interface and a smartphone. This data is used in the numerical comparison of five existing consumption models and three data mining models.

The rest of the paper is structured as follows. In Section 2, we provide a brief overview of the relevant literature. Section 3 describes the experimental set-up and process of data collection. Numerical results are presented in Section 4 and conclusions are given in Section 5.

LITERATURE REVIEW

Much work has been done on the modeling of fuel consumption for vehicles. We will not present a detailed exposition of the existing work but instead refer the interested reader to the review papers (Demir et al. 2011,2014) as well as (Capraz et al. 2016) where detailed descriptions of fuel consumption models can be found. In summary, these models tend to come in two flavors: (1) macroscopic models, suitable for modeling wide-area emissions and require aggregate parameters, and (2) microscopic models, which are instantaneous models described to provide a second-by-second basis fuel consumption of a moving vehicle and are to be used mostly for short-trips. The former class includes models that are simpler and with less number of input parameters than the latter.

Recent years have seen the application of artificial neural networks (ANNs) in automotive engineering, including forecasting engine characteristic under different working conditions. Wu and Liu (2011) studied a predictive system for car fuel consumption using a back propagation neural network and a radial basis function neural network (Wu and Liu, 2012). The engine style, weight of car, vehicle type and transmission system type were used as input parameters for training the neural network training and forecasting fuel consumption forecasting procedure in both (Wu and Liu, 2011, 2012). Togun and Baysec (2010a) presented an artificial neural network model and genetic programming model (Togun and Baysec, 2010b) to predict the torque and fuel consumption of a gasoline engine. Numerous studies have been undertaken to predict the performance and exhaust emission characteristics of internal combustion engines by using ANNs (Parlak et al. 2006; Alonso et al. 2007; Sayin et al. 2007). In (Rahimi-Ajdadi et al. 2014), the tractor fuel consumption was estimated on the basis of engine speed, throttle and load conditions, chassis type, total tested weight by using ANN. The authors of (Kakaee et al.

2015) combined ANN and multi-objective optimization was to reduce fuel consumption of and emissions from a heavy-duty diesel engine. These prediction results demonstrated the proposed system using the neural network is effective and the performance is satisfactory in terms of fuel consumption prediction.

There exist studies featuring alternative ways to model transportation CO2 emissions. For example, Velázquez-Martínez et al. (2015) presented a statistical-mathematical methodology that allows companies to reduce transportation CO2 emissions when delivering in regions with a variety of road and traffic conditions. Wong et al. (2016) employed online time-sequence incremental and decremental least squares support vector machines for engine air-fuel ratio prediction. To the best of our knowledge, there has not been any other work featuring the use of support vector machines for predicting engine related performance measures, such as fuel consumption and emissions.

REAL TIME DATA COLLECTION

All the numerical experiments conducted in this study use real-data with three different vehicles, the names of which are shown in the first column of Table 1. The rest of the columns defines further specifications. The data were collected when the drivers were in the normal course.

	Fuel Type	**Engine Displacement (ml)**	**Frontal Surface Area (m^2)**	**Vehicle Weight (kg)**	**Trip Distance (km)**	**Average Speed (km/h)**
Toyota Corolla	Diesel	1396	2.09	1425	139.23	140
Kia Ceed	Diesel	1582	2.25	1800	432.42	100
BMW X3	Diesel	1980	2.66	2610	361.81	120

Table 1. Vehicle properties

Data from the vehicles was collected in real-time using existing on-board diagnostics (OBD2), a Bluetooth interface and a smartphone. An OBD2 protocol allows accessing the vehicle's Engine Control Unit (ECU) easily through a Bluetooth OBD2 connector. Instant data are obtained via Torque Pro [22] that uses the OBD2 connector. Many recently manufactured vehicles support the use of OBD2. The diagnostic socket should be fitted to the vehicle using an adapter plug prior to instant data collection. In order to capture the parameters needed by the fuel consumption models via Torque Pro, a Bluetooth adaptor is plugged into the OBD2 socket of the vehicle, through which parameters such as the vehicle weight and engine displacement are input into the Torque Pro application on the smart phone. The three vehicles were driven on three different routes and their latitude and longitude coordinates of the origins and destinations were saved.

RESULTS

In this section, we present numerical results obtained by five fuel consumption models from the literature and the three data mining models developed in this paper. These results are not only used to test the accuracy of the models against real data, but also compare the prediction accuracy of eight models. The former category consists of five microscopic models, including an instantaneous fuel consumption model (IFCM) and a four-mode elemental fuel consumption model (FMEFCM) both described by Bowyer et al. (1985), physical emission rate estimator (PERE) model presented by Nam and Giannelli (2005) a comprehensive modal emission model (CMEM) for heavy-goods vehicles described by Scora and Barth [20] and Barth and Boriboonsomsin (2008), and vehicle specific power (VSP) proposed by Jimenez-Palacios (1998). These models were explained in detail by Capraz et al. (2016).

As for the data mining models, we test linear regression (LR), neural networks (NN) and support vector regression (SVR) compared with each other and against the five mathematical models above.

Numerical Results Obtained by the Fuel Consumption Models

The fuel consumption on the three routes was estimated via the five fuel consumption models by providing the input data. The results are given in Table 2, where the second column presents the actual amount of fuel consumed (in liters) on the three routes and the rest of the columns present the predicted amounts for each of the five models.

Models	Real-time data	IFCM	FMEFCM	PERE	CMEM	VSP
Toyota-Corolla	15.946	29.904	15.100	13.301	12.541	8.462
Kia-Ceed	30.175	54.534	33.227	25.696	29.901	17.244
BMW-X3	35.012	63.347	37.094	32.675	36.734	14.527

Table 2. Real time data and models' calculations (total consumption)

Based on the results presented in Table 2, we calculate the mean absolute percentage error (MAPE) and see that FMEFCM, PERE and CMEM provide comparatively better estimates of total fuel consumption than the others. FMEFCM yields the minimum average MAPE out of the five models.

Numerical Results Obtained with the Data Mining Models

In order to predict the fuel consumptions with the regression models, speed, acceleration, engine RPM, volumetric efficiency, EGR commanded and slope of the road are used as input variables for the full regression model. When all the variables are used in analysis, the correlation between real data and predicted data is quite high. However, although all variables are used in the analysis, not all of them are controllable. Given that speed and acceleration are controllable by driver and the slope is dependent on route selection, further analysis was conducted with these three parameters alone. This applies to the three regression models tested here.

In developing the three data mining models, and prior to applying them, the data was first normalized. The least squares method was used to find parameters for linear regression (LR). For the artificial neural network (ANN), a one hidden layer architecture is used with 10 neurons in the hidden layer. The sigmoid activation function (Cybenko. 1989) is used within the hidden layer and a linear activation function is used for the output layer. The Levenberg–Marquardt algorithm (Levenberg, 1944; Marquardt, 1963) is implemented to find the weights which will minimize the least square error. In developing the SVR, the epsilon-SVR is applied with radial basis kernel function (Basak et al. 2007). Both the epsilon value in the loss function and gamma value in radial basis kernel function is specified as 0.1, and the cost parameter C is set as 1.

Random subsampling was used for testing the performance of data mining models. In particular, a randomly selected 80% of the data is used for training and 20% for testing for each run. The procedure is repeated for 1000 times. The average and standard deviation (in parenthesis) of the R values are given in Table 4 for each statistical model, separately for training and for testing. The results presented under columns 3–5 pertain to the full regression (with all variables), whereas columns 6–8 relate to the smaller regression with only the three independent variables, namely speed, acceleration and slope.

		All Variables			Speed, Acceleration and Slope		
Models		LR	ANN	SVR	LR	ANN	SVR
Toyota-Corolla	Training	0.89 (0.005)	0.96 (0.007)	0.97 (0.001)	0.85 (0.006)	0.93 (0.007)	0.93 (0.002)
	Testing	0.89 (0.019)	0.95 (0.021)	0.96 (0.007)	0.85 (0.022)	0.91 (0.029)	0.92 (0.013)
Kia-Ceed	Training	0.78 (0.014)	0.92 (0.014)	0.94 (0.002)	0.75 (0.015)	0.90 (0.013)	0.91 (0.004)
	Testing	0.76 (0.056)	0.89 (0.048)	0.92 (0.014)	0.74 (0.057)	0.88 (0.033)	0.90 (0.018)
BMW-X3	Training	0.90 (0.004)	0.93 (0.010)	0.95 (0.002)	0.76 (0.011)	0.80 (0.019)	0.80 (0.006)

	Testing	0.90 (0.015)	0.91 (0.037)	0.92 (0.011)	0.75 (0.043)	0.78 (0.044)	0.79 (0.026)

Table 3. Results of Data mining Models

As seen from the Table 3, SVR provides the best results for all three vehicles as compared with ANN and linear regression. SVR outperforms other models in terms of both accuracy and consistency, as indicated by the average and standard deviation of correlation coefficient (R) values. Even when only speed, acceleration and slope are used as input values, SVR delivers better results.

In the next set of experiments, we will present more general results instead of constructing a model for each kind of vehicle. These models are constructed by considering additional input variables such as frontal area, engine displacements, and mass of vehicle. The new model can be applied all kinds of vehicles. Table 4 shows result of the new model for each statistical model. In the table, "all variables" indicate that the independent variables are those used in the full regression model as well as the additional variables above. On the other hand, "extended controllable variables" include vehicle characteristics (frontal area, vehicle weight and engine displacement) as well as the slope, instantaneous acceleration and speed.

	All Variables			Extended Controllable Variables		
	LR	NN	SVR	LR	NN	SVR
Training	0.87 (0.004)	0.95 (0.005)	0.96 (0.001)	0.81 (0.006)	0.90 (0.006)	0.90 (0.002)
Testing	0.86 (0.016)	0.94 (0.019)	0.95 (0.005)	0.80 (0.022)	0.88 (0.024)	0.90 (0.009)

Table 4. Results of a generalized statistical model for all vehicles

Similar to Table 3, Table 4 shows that SVR outperforms other methods in both cases, where in the first case all variables are used as input, and in the second in addition to. ANN also provides competitive results, however SVR yields more robust results especially for testing data which can be seen from standard deviations of the R values.

CONCLUSIONS

In this study, we numerically compared five fuel consumption models from the literature using real-time data collected from three vehicles. We also developed, tested and compared three data mining models (LR, ANN and SVR) against the real-time data. The results suggest that the data mining models can be applied easily for predicting fuel consumption of vehicles, and the results indicate that they are superior to the existing fuel consumption models in terms of prediction accuracy.

ACKNOWLEDGEMENT

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK) under Project no. 114E421.

REFERENCES

1. Ahn, K. and Rakha, H. (2008) The effects of route choice decisions on vehicle energy consumption and emissions, *Transport Res D-Tr E*, 13,151-167.

2. Alonso, J.M., Alvarruiz, F., Desantes, J.M., Hernández, L. and Moltó G. (2007) Combining neural networks and genetic algorithms to predict and reduce diesel engine emissions, *IEEE T Evolut Comput*, 11,1, 46-55.

3. Barth, M. and Boriboonsomsin, K. (2008) Real-world CO2 impacts of traffic congestion, *Transport Res Rec*, 2058, 1,163–171.

4. Basak, D., Pal, S. and Patranabis, D.C. (2007) Support vector regression, *Neural Information Processing-Letters*, 11, 10, 203-224.

5. Bowyer, D.P., Biggs, D.C. and Akçelik, R. (1985) Guide to fuel consumption analysis for urban traffic management, *Australian Road Research Board Transport Research*, 32.

6. Capraz, A.G., Togun, N., Sevkli, M., Özel, P., Bektaş, T. and Şevkli, A.Z. Analysis and Validation of Several Fuel Consumption Models for Automobiles Using Real-Time Data, *Transportation Research Part D* (Under Review)
7. Cybenko, G. (1989) Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems*, 2, 4, 303-314.
8. Demir, E., Bektaş, T. and Laporte, G. (2011) A comparative analysis of several vehicle emission models for road freight transportation, *Transport Res D-Tr E*, 16, 5, 347–357.
9. Demir, E., Bektaş, T. and Laporte, G. (2014) A review of recent research on green road freight transportation, *Eur J Oper Res*, 237, 3, 775–793.
10. Jimenez-Palacios, J.L. (1998) Understanding and quantifying motor vehicle emissions with vehicle specific power and TILDAS remote sensing, *PhD thesis*, USA: Massachusetts Institute of Technology.
11. Kakaee, A.H., Rahnama, P., Paykani, A. and Mashadi, B. (2015) Combining artificial neural network and multi-objective optimization to reduce a heavy duty diesel engine emissions and fuel consumption, *J Cent South Univ*, 22, 4235-4245.
12. Levenberg, K. (1944) A Method for the Solution of Certain Non-Linear Problems in Least Squares, *Quarterly of Applied Mathematics*, 2, 164–168.
13. Marquardt, D. (1963) An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *SIAM Journal on Applied Mathematics*, 11, 2, 431–441.
14. Nam, E.K. and Giannelli, R.A. (2005) Fuel consumption modeling of conventional and advanced technology vehicles in the physical emission rate estimator (PERE), *Technical report*, Washington, USA, <http://www3.epa.gov/otaq/models/ngm/420p05001.pdf> [accessed 15.12.2015]
15. Parlak, A., Islamoglu. Y., Yasar, H. and Egrisogut, A. (2006) Application of artificial neural network to predict specific fuel consumption and exhaust temperature for a Diesel engine, *Appl Therm Eng*, 26, 824-828.
16. Rahimi-Ajdadi, F. and Abbaspour-Gilandeh, Y. (2014) Artificial neural network and stepwise multiple range regression methods for prediction of tractor fuel consumption, *Measurement*, 44, 2104-2111.
17. Rakha, H., Ahn, K., Moran, K., Saerens, B. and Bulck, E. (2011) Virginia tech comprehensive power-based fuel consumption model: Model development and testing, *Transport Res D-Tr E*, 16,7, 492–503.
18. Sayin, C., Ertunc, H.M., Hosoz, M., Kilicaslan, I. and Canakci, M. (2007) Performance and exhaust emissions of a gasoline engine using artificial neural network, *Appl Therm Eng*, 27, 46-54.
19. Scojra, G. and Barth, M. (2006) Comprehensive modal emission model (CMEM), version 3.01: User's guide. *Technical report*, University of California, USA. <http://www.cert.ucr.edu/cmem/docs/CMEM_User_Guide_v3.01d.pdf> [accessed 15.12.2015].
20. Togun, N.K. and Baysec, S. (2010a) Prediction of torque and specific fuel consumption of a gasoline engine by using artificial neural networks, *Appl Energy*, 87, 349-355.
21. Togun, N. and Baysec, S. (2010b) Genetic programming approach to predict torque and brake specific fuel consumption of a gasoline engine, *Appl Energy*, 87, 3401-3408.
22. Torque Pro (OBD 2 & Car), (2015) <https://play.google.com/store/apps/details?id=org.prowl.torque&hl=en> [accessed 15.12.2015].
23. Velázquez-Martínez, J.C., Fransoo, J.C., Blanco, E.E. and Valenzuela-Ocana, K.B. (2016) A new statistical method of assigning vehicles to delivery areas for CO2 emissions reduction, *Transport Res D-Tr E,* 43, 133-144.
24. Wong, P.K., Wong, H.C. and Vong, C.M. (2016) Online time-sequence incremental and decremental least squares support vector machines for engine air-ratio prediction, *Int J Engine Res*, 1-13.
25. Wu, J.D. and Liu, J.C. (2011) Development of a predictive system for car fuel consumption using an artificial neural network, *Expert Syst Appl*, 38, 4967-4971.
26. Wu, J.D. and Liu, J.C. (2012) A forecasting system for car fuel consumption using a radial basis function neural network, *Expert Syst Appl*, 10, 1883-1888.