International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

Multiple Linear Regression using TensorFlow Predicting Fuel Consumption

Preeti Bamane[1] Mangal Patil[2]

[1]PG Student, Dept. of Electronics Engineering, Bharati Vidyapeeth College of Engineering Pune, Maharashtra, India
[2]Assistant Professor, Dept. of Electronics Engineering, Bharati Vidyapeeth College of Engineering Pune, Maharashtra, India

---

Abstract - *The estimation of fuel consumption is a sensitive topic for the evaluation in the automotive field. The estimate of fuel consumption is a sensitive topic for the assessment and development of Driving Assistance Systems, and an appropriate estimation is an essential job. In this paper, we proposed Multiple Linear Regression (MLR) - the most popular and frequently used statistical technique for prediction. To show the feasibility of the prediction of the model, an experiment carried out on open-source Hacking, and Countermeasure Research Lab (HCR Lab)'s driving dataset. The model developed and experimented using two optimizers "gradient descent" and "adam" and comparative analysis achieved. Our outcomes reveal that the simplicity of the model's structure and the few controlled variables present a decent prediction for fuel consumption.*

Key Words: **Fuel Consumption, MLR, Multicollinearity, TensorFlow.**

1. INTRODUCTION

Nowadays, the effect of fossil fuel consumption in the field of the transport system estimated in the range of 20% to 40%, thus it also leads to the influence of atmospheric pollution [1]. Furthermore, fuel fraudulent activities in the fleet management sector also increase in recent years, impacting the owners' economies where the fuel price is relatively high. In this sense, the accurate prediction of fuel consumption is indispensable in enhancing the fuel economy of the vehicle and preventing falsified movements in fleet management. The selection of any model input factors is of a prior task to improve the output forecasting performance. Therefore, it is vital to comprehend the factors that contribute to fuel consumption. In the advancement of automotive technology, an engineer can capture high resolution, multivariate time series dataset related to vehicle speed, engine conditions, fuel consumption, etc.

This paper focusses on the prediction of fuel consumption using regression analysis. The regression is the supervised machine learning empirical analysis, often used for predicting future values from past values. In linear regression analysis, there are two categories – Simple linear regression and multiple linear regression. As fuel consumption is dependent on more than one variable, our model developed and tested using multiple linear regression. The model is kept simple and implemented using a Google-developed TensorFlow platform for machine learning.

Rest of the paper is organized as follows. Section II discusses the MLR method. Section III and IV provides a detailed description of the data source and data analysis, respectively. Section V, focusing on model creation using TensorFlow. Section VI discusses the results, and finally, the paper concluded in Section VII.

2. MLR

The MLR is one of the traditional statistical model to express relationships between response and controlled variables [2, 3]. The MLR regression model for prediction is exemplified by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \varepsilon \tag{1}$$

Where y is the response variable, x_i $(i = 1,2,3, \dots, k)$ are the independent explanatory variables, β_i $(i = 1,2,3, \dots, k)$ are the regression coefficients, and ε is the residual error.

3. DATA SOURCE

Our methodology accomplished on the open source unique and highly valuable Hacking and Countermeasure Research Lab (HCR Lab)'s [4] driving dataset. The HCR Lab's accumulated a dataset by driving a vehicle about 46 km in 23 hours and covered a distance between Korea University and SANGAM World Cup Stadium. The driving dataset retrieved from the CAN bus network using OBD port II standard (On-Board Diagnostics).

4. DATA ANALYSIS

This paper considers the rigorous analysis of the driving dataset variables by using the ordinary least squares (OLS) analysis to eliminate the multicollinearity between variables [5]. We found that the selection of parameters influencing fuel consumption more active on the driving data evaluation.

The selection of input parameters for the MLR model is crucial to enhance the accuracy of forecasting. Here, five different parameters, namely, Accelerator, Throttle, Cylinder, Engine speed, Flywheel torque, Road gradient, and Vehicle speed, are chosen.

It is necessary to determine the presence of interaction between selected parameters to remove the multicollinearity. The OLS is used for the elimination of P-value >0.05 to remove multicollinearity. If the P-value is less than 0.05, the parameters have to be accepted, while in the other case (i.e., the P-value is higher than 0.05), the parameters have to be rejected. In the sense of the variance inflation factor (VIF) higher than 5 indicates a multicollinearity problem [6, 7].

According to the P-value and VIF factor shown in Table 1, we notice that the highest P-value on the Road gradient is 0.182, and the highest VIF factor on the Accelerator is 14 reveals the existence of multicollinearity.

Table 1 Statistical Analysis obtained from OLS summary

Parameters	Standard error	t-value	P-value	VIF Factor
Accelerator	3.012	13.640	0.000	**14.5**
Throttle	0.715	8.870	0.000	4.3
Cylinder	0.868	3.264	0.001	1.2
Engine speed	0.028	17.938	0.000	5.1
Vehicle speed	0.563	-10.770	0.000	2.8
Flywheel torque	0.597	5.583	0.000	6.1
Road gradient	4.267	1.336	**0.182**	1.1

Consequently, Road gradient and Accelerator are removed from the dataset. Once again P-value and VIF factor are computed for remaining parameters, and its results are demonstrated in Table 2. Finally, five independent explanatory variables that utmost influence the fuel consumption are chosen such as Throttle, Cylinder, Engine speed, Vehicle speed, and Flywheel torque.

Table 2 Statistical Analysis obtained from OLS summary

Parameters	Standard error	t-value	P-value	VIF Factor
Throttle	0.581	20.347	0.000	2.8
Cylinder	0.868	3.264	0.001	1.2
Engine speed	0.032	24.715	0.000	3.5
Vehicle speed	0.966	-8.233	0.000	2.8
Flywheel torque	0.350	16.205	0.000	2.3

5. MODEL CREATION

The influence of TensorFlow software platform facilitates to solve MLR machine learning problems effectively and efficiently. In the training phase, explicit prediction of future values is estimated using weights and bias.

5.1 Training Model

TensorFlow involves ‘Placeholders’ and ‘Variables’ for creating an early stage of the model parameters [8].

- Placeholders are consumed to allocate a place in the memory where values can be saved. They are essentially used to feed data into the tensor graph.
- Variable is used to store values that adapt after every step of training phase to enhance the accuracy. During initialization random values are allocated and it varies as the model attempts to fit the precise predictions.

In the model, two placeholders namely x and y along with to feed feature matrix and label matrix, X_train and Y_train, are used, respectively. Variables weights and biases namely W and B, respectively. In TensorFlow session, sustained each input and output data variables using feed dictionary function and iterates with 2000 number of Epochs.

5.2 Tensor Graph

TensorBoard is a visualization toolkit of TensorFlow platform that profiling the workflow of the trained algorithm. Figure 1 below demonstrated the core graph for the entire TensorFlow MLR Algorithm.

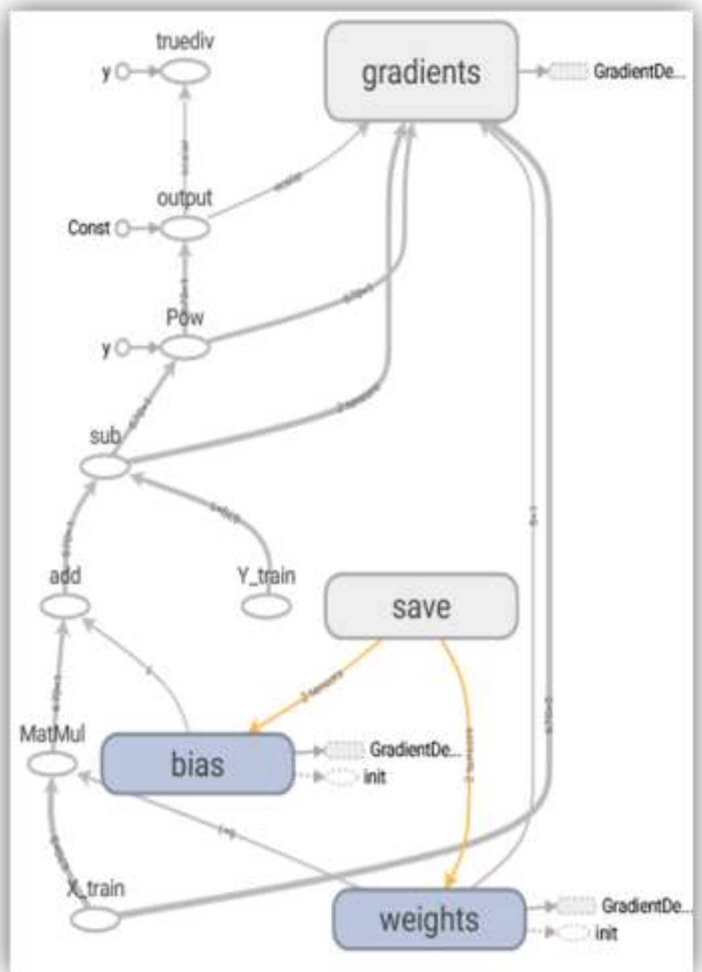

Fig. 1 Tensor Board Graph

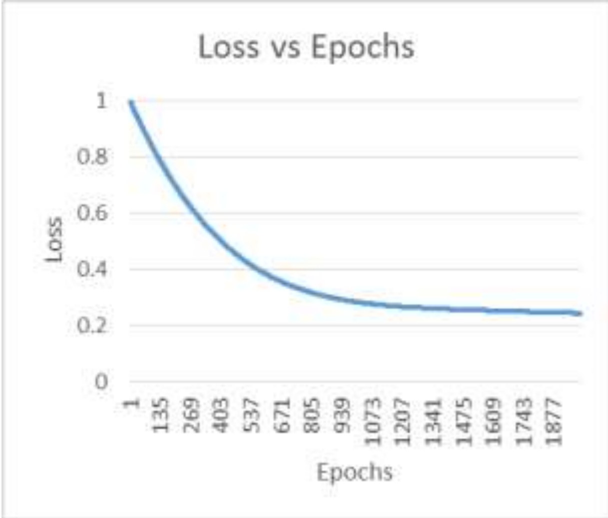

Fig. 2 Gradient Descent optimizer

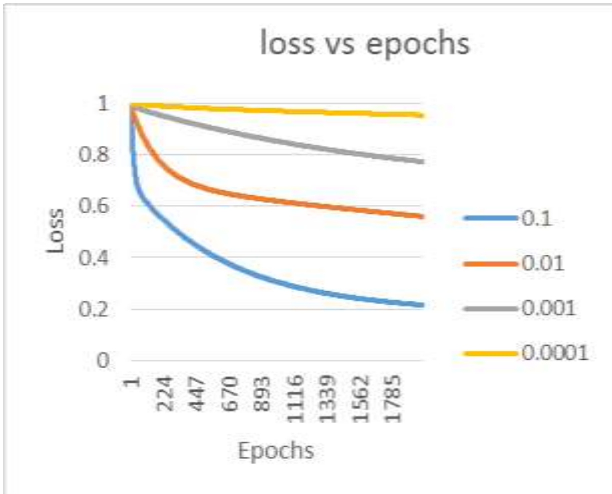

Fig. 3 Adam optimizer

6. RESULT AND DISCUSSIONS

Moreover, TensorFlow takes benefit of the differentiable property of the functions by running two proficient optimizers' namely gradient descent and optimizers to learn the parameters. Gradient descent declines the cost function in the context of steepest descent and converges in reflective time ensured by the proper selection of learning rate. The loss provides an approach of how dexterous the trained model is to depict the data in the dataset.

The 1000 samples from HCRL driving data is divided into two parts. The training datasets with 67% of the observations are used to train the model. The remaining 33% is used to test the model prediction, and then compare the predicted values with the real costs. Figure 2 and 3 illustrates the plot of two MLR training errors using gradient descent optimizer and Adam optimizer. It can be seen that from figure 2, the learning rate of 0.1 is better suited for describing the data over 2000 epochs.

The graphical representation between the actual and predicted value of fuel consumption is represented in the graph given below. From the figure 4, it is observed that the MLR method for prediction of fuel consumption achieved closer values between actual and predicted fuel consumption values.

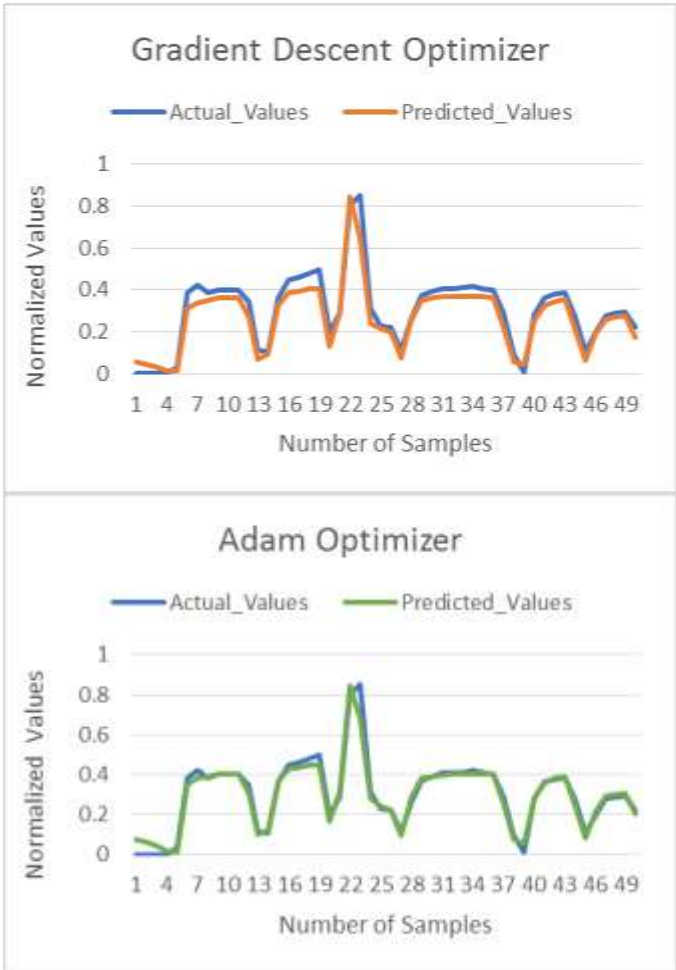

Fig. 4 Prediction Result

This model produces an R2 score of 0.934 and Root Mean Square Error (RMSE) of 0.048769. The R-square coefficient measured the goodness of fit, whereas RMSE measured the differences between values predicted by the model, and the values actually estimated.

7. CONCLUSION

This paper uses MLR to predict fuel consumption based on HCRL driving data. The goal of this work is to determine the parameters which most influence fuel consumption. We take advantage of TensorFlow to create the MLR model. We compared the performance of prediction on Gradient descent and Adam optimizers and found that Adam optimizer is more effective than others. The simulation results show that the relative simplicity of the proposed model's structure with the few input variables provides reasonable outcome. Here, a better estimate is obtained when the values of fuel consumption are normalized in a specific range. The use of prediction and regression helps us to find errors and improve the accuracy of the model. Based on the result, MLR has an acceptable accuracy with R-square 93.4% and RMSE of 0.048769.

Conflicts of Interest

The authors have no conflicts of interest.

REFERENCES

[1] Cartenì, A. Urban Sustainable Mobility. Part 1: Rationality in Transport Planning, 2014.

[2] Azadi, S., Karimi-Jashni, A.: Verifying the performance of artificial neural network and multiple linearregression in predicting the mean seasonal municipal solid waste generation rate: A case study of farsprovince. Iran. Waste Manag.48, 14–23 (2016)

[3] Chong, S.S., Aziz, A.A., Harun, S.W., Arof, H., Shamshirband, S.: Application of multiple linearregression, central composite design, and ANFIS models in dye concentration measurement and pre-diction using plastic optical fiber sensor. Measurement74, 78–86 (2015).

[4] F. Perrotta, T. Parry, and L. C. Neves, "Application of machine learning for fuel consumption modelling of trucks," in Big Data (Big Data), 2017 IEEE International Conference on. IEEE, 2017, pp. 3810–3815.

[5] Ghysels, E., Qian, H.: Estimating MIDAS regressions via OLS with a polynomial parameter profiling.Econom. Stat.9, 1–6 (2018)

[6] Kleinbaum, D.G., Kupper, L.L., Muller, K.E.: Applied Regression Analysis and Other MultivariableMethods. PWS Publishing Co., Boston (1988)

[7] Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C.: Selection and vali-dation of parameters in multiple linear and principal component regressions. Environ. Model. Softw.23(1), 50–55 (2008)

[8] "Module: tf | TensorFlow," TensorFlow. [Online] https://www.tensorflow.org/api_docs/python/tf Last checked on May 29, 2020.