Transportation Research Part C 83 (2017) 134–145

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Data-driven fuel consumption estimation: A multivariate adaptive regression spline approach

Yuche Chen *, Lei Zhu, Jeffrey Gonder, Stanley Young, Kevin Walkowicz

National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, United States

A R T I C L E   I N F O

Article history:
Received 13 April 2017
Received in revised form 3 August 2017
Accepted 3 August 2017

Keywords:
Data-driven analytics
Fuel consumption estimation
Multivariate adaptive regression spline
Eco-routing

A B S T R A C T

Providing guidance and information to drivers to help them make fuel-efficient route choices remains an important and effective strategy in the near term to reduce fuel consumption from the transportation sector. One key component in implementing this strategy is a fuel-consumption estimation model. In this paper, we developed a mesoscopic fuel consumption estimation model that can be implemented into an eco-routing system. Our proposed model presents a framework that utilizes large-scale, real-world driving data, clusters road links by free-flow speed and fits one statistical model for each of cluster. This model includes predicting variables that were rarely or never considered before, such as free-flow speed and number of lanes. We applied the model to a real-world driving data set based on a global positioning system travel survey in the Philadelphia-Camden-Trenton metropolitan area. Results from the statistical analyses indicate that the independent variables we chose influence the fuel consumption rates of vehicles. But the magnitude and direction of the influences are dependent on the type of road links, specifically free-flow speeds of links. A statistical diagnostic is conducted to ensure the validity of the models and results. Although the real-world driving data we used to develop statistical relationships are specific to one region, the framework we developed can be easily adjusted and used to explore the fuel consumption relationship in other regions.

1. Introduction

The transportation sector is a big energy consumer and one of the largest greenhouse gas (GHG) emissions contributors. Governments around the world are taking steps to address the energy and GHG emissions problems caused by transportation (Chen and Fan, 2013, 2014; Zhang et al., 2016; Jiang et al., 2016a, 2016b). Clearly, a portfolio of strategies should be employed by the transportation sector to mitigate GHG emissions and dependence on fossil fuels (Morrison and Chen, 2011; Chen et al., 2017a, 2017b; Jiang et al., 2015; Yi and Bauer, 2017). Although renewable sources of transportation fuels and alternative fuel vehicles are playing roles in this process, reducing fuel consumption and GHG emissions of the existing fleet, where more than 90% are conventional internal combustion engine vehicles, remains an important and effective approach in the near term (Hu and Chen, 2016; Chen and Meier, 2016; Jiang et al., 2014; Chen and Borken-Kleefeld, 2016).

One particular area of interest is to provide guidance to drivers so that they can achieve better fuel efficiency during driving. This is broadly known as eco-driving, which includes two types of tactics. One is to offer microscopic operational tips to drivers (such as maintaining steady speed, smoothing acceleration) to achieve better fuel efficiency. Many studies looked

* Corresponding author.
E-mail address: Yuche.Chen@nrel.gov (Y. Chen).

http://dx.doi.org/10.1016/j.trc.2017.08.003
0968-090X/

into this tactic (Asadi and Vahidi, 2011; Manzie et al., 2007; Wan et al., 2016; Ubiergo and Jin, 2016), and it was reported that on average a 5–15% fuel efficiency improvement could be obtained (Onoda, 2009). Although the potential fuel savings are intriguing, it requires drivers to closely follow microscopic operational tips, which is not feasible and can potentially cause safety concerns (Barkenbus, 2010). But with emerging technologies, such as vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I), they make better eco-driving strategies available and address the safety concerns. For example, Wan et al., 2016 proposed a Speed Advisory System (SAS), which utilizes communications between vehicles to reduce idling time at red lights and achieve fuel minimization driving. Ubiergo and Jin (2016) presented a hierarchical eco-driving strategy based on traffic signal control and V2I communications with the purpose of reducing travel delays and saving fuels by smoothing speed trajectory. The second often overlooked tactic is to guide drivers to choose fuel-efficient routes. The route choice can take into consideration road conditions and traffic conditions, which are predominant factors in determining driving cycle and fuel efficiency. It was claimed that once the route is selected, the aforementioned microscopic tactics seems to have a relatively small influence on fuel efficiency (Nie and Li, 2013). In addition, emerging technologies provide better data to make accurate and real-time fuel consumption estimation for routing optimization. For example, Zulkefli et al. (2017) established a hardware-in-the-loop system to assess fuel consumption of connected and autonomous vehicle with accurate tracking of vehicle speed and measuring of engine operating condition. Zheng and Liu (2017) developed a volume estimation method using connected vehicle trajectory data to estimate traffic volume at signalized intersection. The volume estimation can later be used to estimate travel delay and fuel wasting on idling. Mo et al. (2017) utilized license plate recognition data and systematic LPR data-mending method to estimate vehicle speed profile. And there are other studies that investigated the accuracy of network-level speed and volume data (Kim and Coifman, 2014). Although the above data are not fuel consumption data directly, they can be used to accurately proxy fuel data. And, timely and accurately informing fuel consumption estimations of different routes can help drivers choosing eco-routing options and save fuel and GHG emissions.

Given the important role of eco-routing in saving transportation fuel consumption, in this study, we developed a fuel consumption estimation model to be implemented into an eco-routing guidance system. The emerging technologies provide possibilities to implement better eco-routing strategy. But still, methodology for fuel estimation in eco-routing application is a knowledge gap in literature, particularly, on how to utilize newly available data to accurately estimate fuel consumption. Most of recent literature focused on designing the control strategies, rather than fuel estimation method. There are three types of fuel consumption estimation models: macroscopic, microscopic, and mesoscopic. Macroscopic models (such as Motor Vehicle Emission Simulator [MOVES[1]] lookup table approach) estimate fuel consumption through look-up tables, which are easy to use but could not demonstrate accuracy (EPA, 2010). Microscopic models (such as VT-Micro, Comprehensive Modal Emissions Model [CMEM]) can accurately estimate fuel consumption at the same time stamp (usually each second) as simulated speed trajectories, but traffic simulations are complex and time demanding, thus not suitable for eco-routing systems, which require generated fuel estimations promptly for route choice purpose (Brooker et al., 2015; Nagel and Scheicher, 1994; Rakha et al., 2004). Mesoscopic models combine the advantages of the other two models. They apply large-scale real-world driving data to a microscopic model to obtain accurate fuel estimations and establish a statistical relationship between fuel consumption and influencing factors (such as speed and acceleration) which can be used to promptly evaluate fuel consumption of routes without conducting traffic simulation.

Our proposed model distinguishes itself from existing mesoscopic models by (1) developing a framework to utilize large-scale real-world driving data, (2) applying a clustering prediction algorithm for a better- fitting performance, and (3) considering predicting variables not considered before, such as link free flow speed, number of lanes, etc. Specifically, we adopted a Multivariable Adaptive Regression Spline (MARS) (Friedman, 1991) approach to optimally determine clusters of free-flow speed and fit one regression curve for each cluster. Although the real driving data we used to develop a statistical relationship are specific to one region, the framework we developed can be easily adjusted and utilized to explore fuel consumption relationship in other regions.

The rest of the paper is organized as follows. Section 2 reviews relevant literature on fuel consumption estimation models. Section 3 describes the proposed fuel consumption estimation model and Section 4 discusses the data applied to the model. Results, statistical diagnostic, and comparison with other models are presented in Section 5. Section 6 concludes the paper and discusses directions and future applications.

2. Literature review

In this section, we summarized the characteristics of three types of fuel consumption models, namely macroscopic, microscopic, and mesoscopic. The three types of models differ based on how vehicle activities and fuel consumptions are aggregated over time and space.

Macroscopic models typically estimate vehicle fuel consumption rate based on factors such as average travel speed, vehicle type, and model year. Usually, the estimation relationship is in the format of a lookup or mapping table, such as the tables in some major energy and transportation emission inventory models (Annual Energy Outlook (EIA, 2016), MOBILE6 (EPA,

[1] Note that MOVES is considered to be a multi-resolution model, i.e., macroscopic, mesoscopic and microscopic, based on the approach users choose within MOVES model. Here, the macroscopic model is specifically referring to the look-up table approach to find fuel consumption/GHG emission rate at given speed range.