A Social Approach for Fuel Consumption Prediction

Chien-Ming Tseng, Sohan Dsouza, Chi-Kin Chau

Masdar Institute of Science and Technology, United Arab Emirates
{ctseng,sdsouza,ckchau}@masdar.ac.ae

ABSTRACT

Fuel consumption in automobiles depends on a number of factors: speed, terrain, engine load and driving behavior are just some of them. Accurate estimation of future fuel consumption is vital for not only the scheduling of refueling stops, but also for the choice of routes for the budget- and/or environmentally-conscious. In this study, we demonstrate a least-squares regression model on data we gathered in an on-road, multi-driver experiment to predict the fuel consumption for a given driver along a route on which they have not previously traveled, given their driving data along other routes, and the modeling of driving routes using other drivers' data along them. Using this technique, we can eliminate the assumptions and extrapolations in trip fuel consumption estimation, replacing them with socially-enhanced modeling based on real-life historical data on a diverse set of drivers, vehicles and routes.

1. INTRODUCTION

As automotive information systems read and process more data from vehicles and their environments, so too have increased the consumer expectations of useful information from these systems, often in real time. Among these is the estimation of future fuel consumption, which is typically combined with known information about fuel tank capacity and provided to drivers in the form of a Distance-to-Empty (DTE) readout. Accurate prediction of fuel consumption, and thereby DTE, is vital in allowing drivers to know not only when they need to refuel, but also the fuel consumption along different possible routes.

[4] describes the use of a linear regression technique to predict DTE for electric vehicles. The technique can be adapted for use with internal combustion vehicles to predict their fuel consumption along routes they have driven, based on modeled driving behavior and driving conditions. However, all drivers will not have driven along all possible routes. One driver A may be interested in finding out how much fuel she is likely to consume along a route X she has never driven previously. However, if other drivers B and C have driven route X, along with other routes Y and Z also driven by A, we can use the aforementioned technique to model the differences among the routes, and the differences among the drivers' fuel consumption patterns on the same routes. Combining these models allows us to predict the fuel consumption of driver A over X.

2. METHODOLOGY

We adapt the least-squares regression used in [4] to estimate DTE in an electrical vehicle for internal combustion vehicles. The drivers' running fuel consumption is computed using the formula provided in [3], with the engine data as inputs, for each slice of sampling time, and added up over a trip to give the total consumption. Fuel consumption is affected by different variables, e.g. driving behavior, change in car types, average speed and traffic conditions. Training the regression model using the trip data from each driver, we can estimate the DTE for the same drivers in real time. We then extend this single-driver modeling technique to the scenario of multiple drivers driving along a set of routes. In order to estimate the fuel consumption in different conditions, the multivariate regression model [2] is used. The regression model for a route i is

$$F_i(D_j) = \beta_{i0} + \beta_{i1}\chi_{i1} + \beta_{i2}\chi_{i2} + \ldots + \beta_{im}\chi_{im} \tag{1}$$

Where β_{ik} is a set of m unknown coefficients that are determined from the historical data (i.e. the training set). The variable $\chi)$ is the measureable data from the OBD installed in the car, and the response variable, F_i, is the fuel consumption in a particular route i given the driving data D_j. Solving β_i in Equation 1:

$$\beta_i = (\chi^T\chi)^{-1}\chi^T F_i \tag{2}$$

The value χ is computed from

$$\chi = \begin{bmatrix} 1 & \Delta T_a(r_i, D_1) & V_{ave}(r_i, D_1) & I_t(r_i, D_1) & D_c(r_i, D_1) \\ 1 & \Delta T_a(r_i, D_2) & V_{ave}(r_i, D_2) & I_t(r_i, D_2) & D_c(r_i, D_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \Delta T_a(r_i, D_j) & V_{ave}(r_i, D_j) & I_t(r_i, D_j) & D_c(r_i, D_j) \end{bmatrix} \tag{3}$$

Where $T_a(r_i, D_j)$ denotes the ambient temperature of route i for the historical data, included to consider that the ambient temperature will affect engine load via the heater or the air conditioner. $V_{ave}(r_i, D_j)$ denotes the average speed of driving in route i , since different speeds will cause different fuel consumption in the route. $I_t(r_i, D_j)$ denotes the total idle time in route i ; we assume that different traffic condition results in different idle time in the route. $D_c(r_i, D_j)$ denotes the driver and the displacement of the car, since the fuel consumption rate is different for different car types. Here, j is the total number of the historical data points.

Say driver k never drove in route n before, and we want to estimate the fuel consumption $F_i(D_j)$ of that driver in that route given the ambient temperature, the average speed, and the idle time. We can establish the relationship between the route n and routes $1 \ldots r$ using the multivariate regression model shown in Equation 4. The regression model $F_r(D_{1\ldots m}$ is the training data set. Since we want to use the data of driver k in a different route to determine their fuel consumption in the route they have never driven, the training data set should be $F_r(D_{1\ldots m}, k \in \{1 \ldots m\}$. However, since F_n does not include the date of driver k, the regression model has to be trained without the data of driver k, which is shown in Equation 4. Also note that the number of drivers m should be greater than the number of routes r in order to avoid ill-conditioning of the regression model.

$$F_n(D_{1\ldots m}) = \gamma_0+\gamma_1 F_1(D_{1\ldots m}+\gamma_2 F_2(D_{1\ldots m}+\ldots+\gamma_r F_r(D_{1\ldots m} \tag{4}$$

for $k \notin \{1 \ldots m\}$

Where $F_i(D_{1\ldots m})$ denotes the fuel consumption of drivers $1 \ldots m$ in route i given ambient temperature, average speed and idle time. Solving for γ_i in Equation 1:

$$\gamma_i = (\overline{F}^T \overline{F})^{-1} \overline{F}^T F_n \tag{5}$$

Where

$$\overline{F} = \begin{bmatrix} 1 & F_1(D_1) & \ldots & F_r(D_1) \\ 1 & F_1(D_2) & \ldots & F_r(D_2) \\ \vdots & \vdots & \vdots & \vdots \\ 1 & F_1(D_m) & \ldots & F_r(D_m) \end{bmatrix}, F_n = \begin{bmatrix} F_n(D_1) \\ F_n(D_2) \\ \vdots \\ F_n(D_m) \end{bmatrix}$$

Since $F_{1\ldots r}(D_{1\ldots m}), k \in \{1 \ldots m\}$, $F_{1\ldots r}(D_k)$ can be determined and substituted into Equation 4 to compute $F_n(D_k)$.

3. EXPERIMENT

We designed an experiment in which we gathered data from three vehicles of different classes — SUV, sedan, and hatchback — driven along the same path simultaneously. Our data collection apparatus consisted of Bluetooth ELM327 dongles plugged into the vehicles' onboard diagnostic (OBD) ports and paired with drivers' smartphones, on which were installed a mobile app we developed for collection and upload of OBD data from the vehicles, along with geolocation data, accelerometer readings and device identification from the smartphone.

Three different cars with three different drivers were used in this experiment: A) Ford Fusion 2012, 4 cylinder, 2.5 L. B) Hyundai Veloster 2014, 4 cylinder, 1.6 L and C) Lincoln MKX 2007, 6 cylinder, 3.7 L. We chose a 36.3 kilometer-long triangular circuit for the experiment, split up into three segments (routes) of lengths 10km, . Each vehicle's driver was assigned a particular driving style: A) cautious, B) moderate, and C) aggressive. The data collection run consisted of two rounds of the circuit, which adds up to about 73 km. Hence, each route was covered twice.

4. RESULTS

Using linear regression for the prediction of DTE for each driver along the circuit, we demonstrated that this approach can outperform the vehicles' own DTE estimation models, as seen in Figure 1.

We then applied the extended modeling methodology to predict fuel consumption for each driver along each run along each route, using the data from them along other routes and other drivers along all routes. The resulting matrix of estimations, and their difference from actual consumption data, is shown in Table 1.

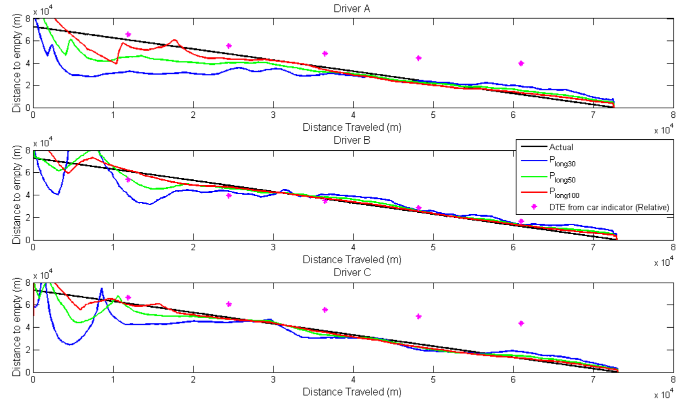

Figure 1: DTE estimated with different-sized terms of fuel intensity (p_{long}). Red dots are estimates given by in-vehicle displays.

Route	Driver A	Driver B	Driver C
1	1.71 (10.0%)	1.23 (17.5%)	1.77 (18.8%)
	1.70 (12.1%)	1.18 (19.4%)	1.84 (18.6%)
2	0.88 (9.6%)	0.61 (10.9%)	0.81 (22.5%)
	0.88 (14.6%)	0.59 (22.1%)	0.75 (30.7%)
3	0.67 (10.2%)	0.57 (7.8%)	0.73 (8.7%)
	0.65 (4.8%)	0.40 (23.2%)	0.73 (20.8%)

Table 1: Estimation of fuel consumption in litres for each run of each route by each driver, with error in parentheses

5. DISCUSSION AND FUTURE WORK

The accuracy of the prediction is limited by the number of training routes, which is in turn limited by the number of drivers. However, the data acquisition system we developed for use in this experiment has been linked to the CloudThink platform [1], which is being expanded to include a network of diverse automobile data acquisition devices. Apart from the data gathered by running more experiments of our own, we will be acquiring more data from this platform as its user base expands. More data from more drivers over the same routes, and in different climactic and traffic conditions, will help us address the paucity of data and the short, low-consumption trips in the experiments adversely affecting the accuracy of fuel consumption prediction in this particular demonstration.

6. REFERENCES

[1] CloudThink Consortium. Cloudthink homepage. Web page at `http://www.cloud-think.com`, 12 2013.

[2] R.J. Freund, W.J. Wilson, and P. Sa. *Regression Analysis*. Regression Analysis Series. Elsevier Science, 2006.

[3] Bruce D. Lighter. Map- and maf-based air/fuel flow calculator. Web page at `http://www.lightner.net/obd2guru/IMAP_AFcalc.html`, 10 2013.

[4] L. Rodgers, E. Wilhelm, and D. Frey. Conventional and novel methods for estimating an electric vehicles distance to empty. In *Proceedings of the ASME 2013 International Design Engineering Technical Conferences and the International Conference on Advanced Vehicle Technologies*, Portland, Oregon, USA, 2013.