A Social Approach for Predicting Distance-to-Empty in Vehicles

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

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

ABSTRACT

Distance-to-Empty (DTE) in vehicles depends on several uncertain factors, such as speed, terrain, traffic and driving behavior. Accurate estimation of DTE is vital for not only the scheduling for refueling, but also for the choice of routes for the budget- and/or environmentally-conscious. Traditional approaches often rely on a single driver's personal history. In this paper, we explore a social approach by using other drivers' data to predict the fuel consumption for a given driver along a new route that is not traveled previously. We develop a least-squares regression model and corroborate the performance empirically by an on-road, multi-driver experiment. Our results can enable a new kind of social platform for trip planning based on the shared data among drivers.

1. INTRODUCTION

While in-vehicle information systems are increasingly sophisticated, the information presented from vehicles is not always accurate. One of the major features is Distance-to-Empty (DTE) or alternatively, the fuel consumption for the remaining journey, which are hindered by several uncertain factors, such as speed, terrain, traffic and driving behavior, as well as the intrinsic characteristics of vehicles (e.g., fuel tank capacity, engine load). 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. In addition, the estimation of DTE is useful for scheduling of refueling, which can optimize the waiting time in the gasoline station.

Previous work relies on using a single driver's personal history for prediction. In contrast, we focus on a social approach by using other drivers' data to predict the DTE for a given driver along a new route. For example, if drivers A, B and C have driven a set of routes X and Y, and only drivers B and C have driven a route Z, we will be able to obtain an estimation for driver A and route Z based on the differences among the routes, and the differences among the drivers' fuel consumption patterns on the same routes.

e-Energy'14, June 11–13, 2014, Cambridge, UK.
ACM 978-1-4503-2819-7/14/06.
http://dx.doi.org/10.1145/2602044.2602073.

To this end, we build on the linear regression approach developed by [4] using a single driver, and extend it to consider multi-driver settings.

2. METHODOLOGY

We adapt the least-squares regression used in [4] to estimate DTE for an electrical vehicle, which also applies to internal combustion vehicles. We first describe the single-driver approach. The drivers' running fuel consumption is computed using a 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. 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 a multi-driver setting. The multivariate regression model in [2] is a blackbox approach, without the detailed knowledge of driving conditions. The regression model estimates the fuel consumption for a route i by

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

where (β_{ik}) is a set of (m) unknown coefficients that are determined from the historical data (i.e., the training set). The variables (χ) are the measurable data obtained from the vehicle (e.g., speed, engine parameters), and the response variable, (F_i), is the fuel consumption in a particular route (i) given the data of driver (D_j). Solving (β_i) in Eqn. 1:

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

We focus on internal combustion vehicles. The variables χ we employ are listed as follows:

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

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 vehicle, since the fuel consumption rate is different for different vehicle types. Here, (j) is the total number of the historical data points.

We next describe the multi-driver extension. Say driver (a) never drove in route (x) before, and we want to estimate the fuel consumption ($F_x(D_a)$) 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 (x) and routes ($1 \ldots r$) using the multivariate regression model shown in Eqn. 3. 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 ($F_x$) does not include the data of driver (a), the regression model has to be trained without the data of driver (a), resulting in Eqn. 3. Also note that the number of drivers (m) should be greater than the number of routes (r) in order to avoid the ill-condition of the regression model.

$$F_x(D_{1\ldots m}) = \gamma_0 + \gamma_1 F_1(D_{1\ldots m}) + \ldots + \gamma_r F_r(D_{1\ldots m}) \quad (3)$$

where $a \notin \{1 \ldots m\}$ and ($F_i(D_{1\ldots m})$) 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 Eqn. 1:

$$\gamma_i = (\overline{F}^T \overline{F})^{-1} \overline{F}^T F_n \quad (4)$$

where

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

Since we have $F_{1\ldots r}(D_{1\ldots m}), k \in \{1 \ldots m\}$, $F_{1\ldots r}(D_k)$ can be determined and substituted into Eqn. 3 to compute $F_x(D_a)$.[1]

3. EXPERIMENT

We carried out an experiment to corroborate the performance of our approach empirically. The data from three vehicles of different classes were gathered. Our data collection apparatus consisted of Bluetooth ELM327 dongles plugged into the vehicles' onboard diagnostic (OBD) ports and paired with drivers' smartphones and a app is developed for collection and upload of OBD data from the vehicles. [2]

[1] If we use only historical data of the driver to predict her fuel consumption, requires dividing the route into many segments based on the characteristic, and is complicated and needs a complete knowledge of fuel consumption of different type of route. On the other hand, social approach allows us to find the fuel consumption between different routes, some of them are very difficult to be separated into segments for analyzing. For the social approach, distance, terrain and traffic data over routes are part of the input of the model, but the model itself is constructed based on how other drivers are affected by these factors at the time they were driving the routes. If something like road blockage or construction affects traffic for a significantly long time, it will also be reflected by the models since the models are periodically updated. Google Maps also provides estimation of traffic and time to destination (TTD), which has already included social approach. Although using a better estimation of TTD may give a better estimation of average speed, it still require a complicated model to take into account of the nuances of a new route with the same average speed, idle time and other conditions.

[2] 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. The OBD data are mass air-flow, manifold absolute pressure, intake air temperature and engines' RPM, which are then utilized to compute the fuel consumption rate. Furthermore, the geolocation data, accelerometer readings and device identification from the smartphone are also recorded for driving behavior parameters of the model.

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 Fig. 1. We then applied multi-driver approach 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.

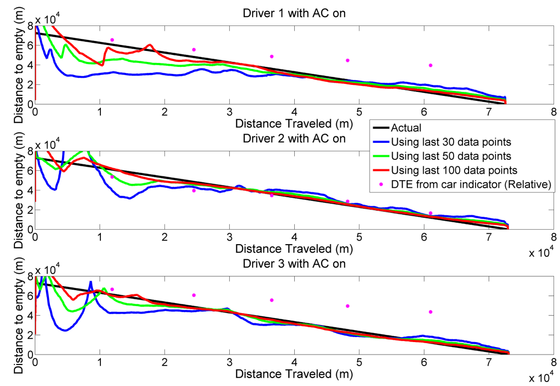

Figure 1: DTE estimated with different-sized terms of fuel intensity (p_{long}). Red dots are estimates given by in-vehicle displays.

Route	Driver A	Driver B	Driver C
1	1.71 (10.0%) 1.70 (12.1%)	1.23 (17.5%) 1.18 (19.4%)	1.77 (18.8%) 1.84 (18.6%)
2	0.88 (9.6%) 0.88 (14.6%)	0.61 (10.9%) 0.59 (22.1%)	0.81 (22.5%) 0.75 (30.7%)
3	0.67 (10.2%) 0.65 (4.8%)	0.57 (7.8%) 0.40 (23.2%)	0.73 (8.7%) 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 vehicle 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. We will also build a social platform for trip planning based on the shared data among drivers.

6. REFERENCES

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

[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`.

[4] L. Rodgers, E. Wilhelm, and D. Frey. Conventional and novel methods for estimating an electric vehicleâĂŹs âĂŸdistance to emptyâĂŹ. In *Proceedings of the ASME 2013*, Portland, Oregon, USA, 2013.