A Participatory Sensing Approach for Personalized Distance-to-empty Prediction and Green Telematics

Chien-Ming Tseng[†], Chi-Kin Chau[†], Sohan Dsouza[†] and Erik Wilhelm[‡]

[†]Masdar Institute of Science and Technology, UAE
[‡]Singapore University of Technology and Design, Singapore
{ctseng, ckchau, sdsouza}@masdar.ac.ae, erikwilhelm@sutd.edu.sg

ABSTRACT

Participatory sensing is an emerging concept that integrates crowd-sourced data collection and knowledge discovery of collective behavior. Capitalizing on the advent of abundant sensors and information collection systems in near-future vehicles, we develop a participatory sensing based system and its methodologies for driving energy efficiency applications. Distance-to-empty (DTE) is the distance an electric or internal-combustion engine (ICE) vehicle can reach before its energy/fuel is exhausted, which is determined by a variety of uncertain factors, such as driving behavior, terrain, types of road, traffic, and vehicle specification. Green telematics aims to optimize the route selection with lower energy consumption. In this paper, we explore an effective approach that integrates the vehicle data gathered from participatory sensing to provide more accurate personalized DTE prediction and green telematics. Our approach relies on extracting the driver/vehicle/route dependent features and discovering correlations from collective driving data. We also present concrete case studies of our results, such as (1) DTE prediction for EVs based on the data of ICE vehicles, (2) classification and recommendations of energy-efficient driving behavior, and (3) route-level energy consumption geo-fencing and planning.

Categories and Subject Descriptors

J.0 [**Computer Applications**]: General; K.4 [**Computing Milieux**]: Computers and Society

Keywords

Participatory sensing; Distance-to-empty; Green telematics

1. INTRODUCTION

While the in-vehicle information systems are increasingly sophisticated, the information presented in vehicles is not always accurate. One of the major features is *distance-to-empty* (DTE), which is the distance an electric or internal-combustion engine (ICE) vehicle can reach before its energy/fuel is exhausted. DTE is determined by a variety of factors, such as driver behavior, terrain, types of road, and traffic, as well as the vehicle specification (e.g., electric or gasoline power, tank capacity, engine load, vehicle weight).

The conventional approach of DTE prediction employed by car manufacturers is based on the projection of past long-term average vehicle energy efficiency (i.e., the total consumed energy/fuel over the total travelled distance). If there is perfect knowledge about the vehicle, driving behavior and the route to travel, future energy efficiency can be estimated with high accuracy. However, there are numerous uncertain factors, which make accurate prediction challenging.

On the other hand, the option of *green telematics* is being enabled by telematics providers, which aims to optimize the route selection decisions with lower fuel/energy consumption [5]. Green telematics systems are critical to optimal planning for recharging/refueling and fleet management. Like DTE prediction, green telematics is affected by the uncertain factors, such as driving behavior and routes.

These driving energy efficiency applications can be enhanced by exploring the historic data from other drivers. *Participatory sensing* is an emerging concept that integrates crowd-sourced data collection and knowledge discovery of collective behavior, enabling a variety of novel applications for pervasive computing systems [3]. The vehicles are becoming a vital platform for participatory sensing. First, there is a rise of advanced in-vehicle information systems, equipped with network connectivity and processing power, which can be turned into mobile sensing and information collection systems. Second, the wide availability of smartphones carried by passengers can extend the computing and sensing abilities of vehicles. Third, there are abundant off-the-shelf and after-market products for vehicle diagnostic tools to gather driving data and vehicle information. A participatory sensing platform for vehicles has been critical for a number of intelligent transportation applications (e.g., Google Map, Waze, real-time traffic alerts). Waze is a participatory App for traffic monitoring, which however does not have driving energy efficiency features.

In this paper, we focus on the applications of participatory sensing related to driving energy efficiency. In particular, there are a few example applications as follows.

1. *Vehicle Variable Applications*: Range anxiety is a critical issue for EVs, and accurate DTE prediction is highly desirable [12]. Since there are far many more ICE vehicles on the road than EVs, one can harness the data collected from ICE vehicles to improve the accuracy of DTE prediction for EVs. In particular, the data of taxis and buses on regular routes can shed light on the energy consumption of other vehicles.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
e-Energy'15, July 14–17, 2015, Bangalore, India.
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3609-3/15/07 ...$15.00.
DOI: http://dx.doi.org/10.1145/2768510.2768530.

2. *Driver Variable Applications*: With the diverse data collected from various drivers, one can compare the driving behavior among drivers. By identifying the driver-specific characteristics from the data, one can analyze the driving behavior and make energy efficient driving recommendations. The insights gathered from driving behavior analytics are also relevant to safe driving and auto insurance, to the benefit of consumers (e.g., pay-as-you-drive plan [6]).
3. *Route Variable Applications*: Extensive green telematics can be enabled to compare different route options according to energy/fuel consumption. Route-level energy consumption geo-fencing can be constructed for EVs. Moreover, one can also optimize the route selection in order to refuel at cheaper gas stations.

In spite of the promising potential, there are several challenges of developing an effective participatory sensing system for driving energy efficiency applications:

1. *Automation*: With the availability of large amount of the data, an automated system is required to process and analyze data, with minimal human assistance.
2. *Flexibility*: The analytics synthesized from the data should be reusable to diverse applications, without the need for re-processing.
3. *Errors and Noise*: There exist considerable errors and noise in the participatory sensing data (e.g., due to synchronization, mechanic dumping, loss of data, differences in data sources). The system needs to resolve the inconsistency and maximize the integrality.
4. *Scalability and Efficiency*: Many applications require real-time processing, and the system needs to deliver the results efficiently in a scalable manner.

In this paper, we explore an effective approach that integrates the vehicle data gathered from participatory sensing to provide more accurate personalized applications. Our approach relies on extracting the driver/vehicle/route dependent features and discovering correlations from collective driving data. Furthermore, we present concrete case studies that utilize our results in diverse related applications, such as (1) DTE prediction for EVs based on the data of ICE vehicles, (2) classification and recommendations of energy-efficient driving behavior, and (3) route-level energy consumption geo-fencing and planning.

Outline: We present the related work in Sec. 2. The system framework is presented in Sec. 3. The methodology of energy consumption model and correlation discovery is presented in Sec. 4. We performed empirical evaluation on our methodology on electric and ICE vehicles, with results discussed in Sec. 5. Finally, the case studies that utilize our results are discussed in Sec. 6.

2. RELATED WORK

The problem of estimating DTE has been the subject of a significant number of academic publications, and is also a feature which vehicle manufacturers have been including in production vehicles for over two decades [2]. At its most basic level, the DTE, or range remaining can be estimated by observing the mean energy use over short and long distances as the authors describe in [12]. The same authors proceed to describe a system for estimating future travel profile using a Monte Carlo approach, which is a critical step in determining remaining energy. Estimates of stochastic effects which may impact travel velocity and acceleration profiles can be crowd-sourced for identifying traffic congestion [4].

The second step is using an accurate vehicle model to take the travel profile generated and turn it into future energy demands. Such model-based estimation can be performed as described in [7] for EVs. It is also possible for fuel consumption data shared between vehicles to be used without underlying physical profile and vehicle modeling to predict the energy consumption for a given route [5], although this approach is sensitive to much uncertainty and necessitates rigorous machine learning approaches – something which these authors have investigated for control applications, but which can easily be applied to the problem at the center of this work [11]. Early efforts at using social network participation to identify areas of fuel use reduction have been published [8]. Another participatory sensing system for improving fuel efficiency has been proposed in [10], and a simpler model for predicting fuel consumption across a driver-route matrix was proposed in [14].

This work differentiates from the previous work in several aspects: (1) We explicitly consider the features related to the driver, vehicle and route dependence in the energy consumption model. (2) We explore the correlations in driving data for personalized applications of individual drivers. (3) We present a unified study to diverse applications related to driving energy efficiency.

3. SYSTEM FRAMEWORK

We developed a system framework (depicted in Fig. 1) that consists of a number of components for data sensing and collection, data analytics, and information processing.

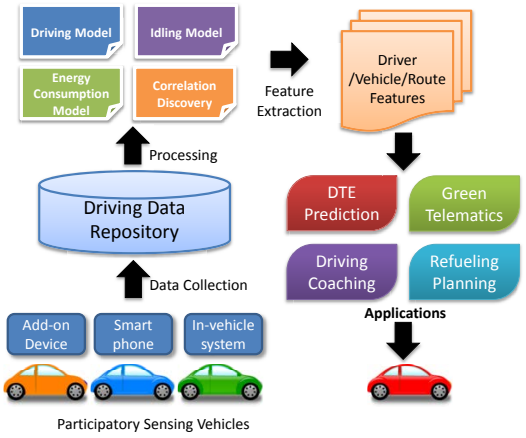


Figure 1: System framework.

3.1 Data Collection

The data sensing and collection system supports a range of methods. One method was based on a smartphone App. We developed an App for Android phone paired over a Bluetooth connection with a standard ELM327 dongle (depicted in Fig. 2), which is widely available online and in automotive electronic stores. The dongle plugs into the OBD port, and the ELM327 protocol is used by the App to query for data on specific engine and other vehicle parameters. The OBD

data, supplemented with data from the phone's own positioning and accelerometer sensors, are accumulated within the device and, if configured and connected for this purpose, uploaded to a specific data server.

Another method is to rely on a tailor-made sensor-equipped electronic unit that plugs directly into the OBD port to draw power and query similar data, uploading the data over the cellular network. In future, when in-vehicle computers become more open to legal software customization, it would also be possible to develop user-level programs that can be installed on these computers and query data through APIs and use either the onboard connectivity or a paired mobile data device to support participatory sensing.

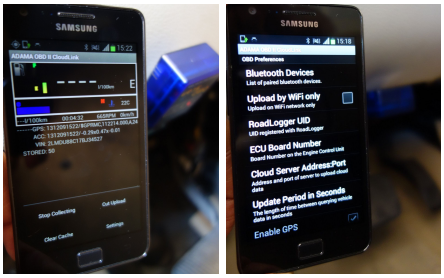


Figure 2: Android app and ELM327 dongle

3.2 Data Processing and Analytics

The data processing and analytics system receives data from the participatory sensing devices, stores it in databases, and processes it with analytic models for the applications. In a companion project, we develop CloudThink platform [1] as an open system for storing the data uploaded over Internet connections from the various in-vehicle data-gathering applications and devices, with a data access server allowing secure and traceable API service in multiple formats, and a web server to permit data visualization.

4. METHODOLOGY

In this section, we present a unified energy/fuel consumption model for both electric and ICE vehicles.

4.1 Basic Concepts

In order to perform the participatory sensing and subsequent estimation of DTE, features of the model must be defined. For this work, three categories of features important for the estimation will be described. The first relates vehicle speed and idle duration to the typical traffic conditions experienced on a road segment. The second enables the type of driver to be defined and parameterized. The third describes the physical characteristics of the vehicle which is the subject of the identification. The following subsections provide more details on how the model was constructed.

1. *Route*: The level of traffic which is expected for a road segment depends on a plethora of factors, including time of day, day of year, special events, etc., and is generally difficult to predict. For this work, the type of routes was used as the primary explanatory variable and was divided into three categories: small public or private roads with urban traffic, lower capacity "urban" highways, and higher capacity freeways.
2. *Driver*: The parameterization of the driver was enabled through the black-box statistical approach chosen for the model described in the next section. Different drivers are expected to behave according to preferences for stop/start accelerations, aggression in various scenarios, propensity for hypermiling, etc.
3. *Vehicle*: The extrinsic parameters capturing vehicle characteristics are only the weight of the vehicle. All other characteristics such as power, wheelbase, top speed are parametrized implicitly in the regression coefficients identified during the modeling process. This parameter is selected to be universal to both electric and ICE vehicles.

4.2 Energy Consumption Model

A tuple of driver-vehicle-route combination is denoted by $(\mathsf{D},\mathsf{V},\mathsf{R})$. The driving data repository is consisted of the data sets as a subset of $\{(\mathsf{D},\mathsf{V},\mathsf{R})\}$.

We divide a route R by a set of n segments[1]. The set of segments are denoted by $\{\mathsf{R}^i\}_{i=1}^n$.

The total energy consumption E of driver D with vehicle V on route R is given by:

$$E_{\mathsf{D},\mathsf{V},\mathsf{R}} = \sum_{i=1}^{n} (E^{\mathrm{drv}}_{\mathsf{D},\mathsf{V},\mathsf{R}^i} + E^{\mathrm{idl}}_{\mathsf{D},\mathsf{V},\mathsf{R}^i}) \quad (1)$$

where $E^{\mathrm{drv}}_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ is the driving energy consumption and $E^{\mathrm{idl}}_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ is the idling energy consumption of segment R^i. In this paper, we denote a symbol by $\hat{}$ for its estimated quantity. For example, $\hat{E}_{\mathsf{D},\mathsf{V},\mathsf{R}}$ denotes the estimated energy consumption.

4.2.1 Driving Energy Consumption Model

Considering a particular segment, the driving energy consumption of a moving vehicle E^{drv} has unit in liter or kWh. Here, we drop the subscript $_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ for brevity.

While there are more sophisticated approaches of estimating the driving energy consumption by detailed modelling of vehicle mechanics, this paper utilizes a black-box approach without the detailed knowledge of vehicle mechanics. This approach aims to maximize the applicability on a wide range of scenarios arising from participatory sensing.

In this paper, we estimate $\hat{E}^{\mathrm{drv}}$ by a linear equation:

$$\begin{aligned}\hat{E}^{\mathrm{drv}} &= \alpha_1\bar{v} + \alpha_2\bar{v}^2 + \vec{\alpha}_3\vec{d} + \vec{\alpha}_4\vec{a} + \vec{\alpha}_5\vec{g} \\ &\quad + \alpha_6\ell + \alpha_7 m + c \end{aligned} \quad (2)$$

where

- $\bar{v}$ is the continuous average speed (i.e., the average speed without idling);
- $\vec{d} = (\tau_d, \mu_d, \sigma_d)$ is the deceleration tuple, where
  - τ_d is the total duration of deceleration;
  - μ_d is the mean deceleration (i.e., the sum of deceleration values divided by the deceleration duration);
  - σ_d is the standard deviation of deceleration;
- $\vec{a}$ is the acceleration tuple (similar to the deceleration tuple);

[1]The length of each segment we choose currently is 100m. But we remark that it is possible to have dynamically variable lengths according to the road environments.

- $\vec{g} = (\bar{g}_x, \bar{g}_y, \bar{g}_z)$ is the gyroscope tuple, where
  - $\bar{g}_x, \bar{g}_y, \bar{g}_z$ are the mean absolute measurement values in $x/y/z$-axis along the moving vehicle;
- ℓ is the auxiliary load of idling, which the baseline reading when the vehicle is not moving;
- m is the vehicle weight;
- $\alpha_1, ..., \alpha_8, c$ are coefficients.

Remarks:

1. The deceleration tuple is critical to capture the energy consumption for EVs in the presence of regenerative braking, by which the vehicle's kinetic energy is converted to energy storage, during braking.
2. The deceleration duration τ_d can capture the dynamic of deceleration. For example, when a driver gradually releases gas pedal from high speed, the mean deceleration μ_d is small with a long deceleration duration, on the other hand, when a driver decelerates his vehicle for a short brake (e.g., avoiding speed camera), μ_d is also small with a short deceleration duration. In both cases, the mean deceleration will be similar, but the duration of deceleration can distinguish the difference.
3. The gyroscope tuple $\vec{g}$ can capture the curvature of a route. Normally, travelling a curve route consumes more energy than a straight route, in spite of equal distance. The gyroscope provides important information about the environments. The gyroscope may be available in smartphones.
4. The auxiliary load of idling ℓ is the *calculated engine load* when the vehicle is not moving, which is often available from On Board Diagnostic (OBD) port. The auxiliary load of idling provides the baseline energy consumption of a stationary vehicle, which is more affected by factors such as in-vehicle air-conditioning and stereo systems.

4.2.2 Idling Energy Consumption Model

Similarly, we rely on a black-box approach to estimate the idling energy consumption. Considering a particular segment, we estimate the idling energy consumption $\hat{E}^{\text{idl}}$ by a linear equation:

$$\hat{E}^{\text{idl}} = \beta_1 \tau^{\text{idl}} + \beta_2 \ell \tag{3}$$

where

- τ^{idl} is the total idle duration,
- ℓ is the auxiliary load of idling.

Here, we drop the subscript $_{\mathsf{D,V,R}^i}$ for brevity. The auxiliary load of idling ℓ provides the baseline energy consumption of a stationary vehicle.

4.3 Estimation of Coefficients

The coefficients $(\alpha_1, ..., \alpha_7)$, (β_1, β_2) and c in the linear equations (Eqns. (2)-(3)) can be estimated by the standard regression method, if provided a sufficiently large data set of driving data $(\bar{v}, \vec{d}, \vec{a}, \vec{g}, \ell, m, \tau^{\text{idl}})$ and energy consumption data $(\hat{E}^{\text{drv}}, \hat{E}^{\text{idl}})$. We assume that each driver-vehicle pair (D, V) has collected sufficient historic personal driving data, and the coefficients can be estimated in a-priori manner.

One notable advantage of regression method is that it is less susceptible to random noise, which can arise from various sources (e.g., due to synchronization, mechanic dumping, inaccurate measurements).

4.4 Dependence and Features

By participatory sensing, drivers can share their driving data (e.g., coefficients of Eqns. (2)-(3)) with each other. Next, we explore an effective approach that integrates the participatory sensing data to personalized applications.

In fact, the linear equations (Eqns. (2)-(3)) provide a convenient way to extract the features that are related to driver, vehicle, and route dependence. In Table 1, we heuristically assign the dependence of each parameter, according to the major effects from the driver, vehicle or route.

For the coefficients, it is assumed that their dependence is complementary to that of the respective parameters. For example, the average speed $\bar{v}$ is more likely affected by the driver and route, while to a less extent by the type of vehicle. Hence, coefficient α_1 is considered to be vehicle-dependent, such that the product $\alpha_1 \bar{v}$ will be specific to a particular tuple $(\mathsf{D}, \mathsf{V}, \mathsf{R}^i)$. For convenience, we assume that c is driver-dependent.

Parameters	Driver-dependent	Vehicle-dependent	Route-dependent
$\bar{v}, \vec{d}, \vec{a}, \vec{g}$	✓		✓
ℓ	✓	✓	
m		✓	
τ^{idl}			✓
$\alpha_1, ..., \vec{\alpha}_5$		✓	
α_6			✓
α_7	✓		✓
c	✓		
β_1	✓	✓	
β_2			✓

Table 1: Dependence of parameters and coefficients.

In light of dependence, we can specify each quantity by the respective subscripts when referring to a particular tuple $(\mathsf{D}, \mathsf{V}, \mathsf{R}^i)$. For example, we write $\bar{v}_{(\mathsf{D,V})}$ and $\alpha_{1(\mathsf{V})}$.

We denote a matrix of energy consumptions by $\mathbf{E} = (E_{\mathsf{D,V,R}^i})$. Each $E_{\mathsf{D,V,R}^i}$ can be computed by a linear equation:

$$E_{\mathsf{D,V,R}^i} = (\mathbf{c}_{\mathsf{V}}, \mathbf{c}_{\mathsf{R}^i}, \mathbf{c}_{\mathsf{D,R}^i}, \mathbf{c}_{\mathsf{D,V}}) \cdot (\mathbf{x}_{\mathsf{D,R}^i}, \mathbf{x}_{\mathsf{D,V}}, \mathbf{x}_{\mathsf{V}}, \mathbf{x}_{\mathsf{R}^i})^T + c \tag{4}$$

where

$$\mathbf{c}_{\mathsf{V}} = (\alpha_{1(\mathsf{V})}, \alpha_{2(\mathsf{V})}, \vec{\alpha}_{3(\mathsf{V})}, \vec{\alpha}_{4(\mathsf{V})}, \vec{\alpha}_{5(\mathsf{V})}) \tag{5}$$

$$\mathbf{x}_{\mathsf{D,R}^i} = (\bar{v}_{\mathsf{D,R}^i}, \bar{v}^2_{\mathsf{D,R}^i}, \vec{d}_{\mathsf{D,R}^i}, \vec{a}_{\mathsf{D,R}^i}, \vec{g}_{\mathsf{D,R}^i}) \tag{6}$$

$$\mathbf{c}_{\mathsf{R}^i} = (\alpha_{6(\mathsf{R}^i)}, \beta_{2(\mathsf{R}^i)}) \tag{7}$$

$$\mathbf{x}_{\mathsf{D,V}} = (\ell_{\mathsf{D,V}}, \ell_{\mathsf{D,V}}) \tag{8}$$

$$\mathbf{c}_{\mathsf{D,R}^i} = (\alpha_{7(\mathsf{D,R}^i)}) \tag{9}$$

$$\mathbf{x}_{\mathsf{V}} = (m_{\mathsf{V}}) \tag{10}$$

$$\mathbf{c}_{\mathsf{D,V}} = (\beta_{1(\mathsf{D,V})}) \tag{11}$$

$$\mathbf{x}_{\mathsf{R}^i} = (\tau^{\text{idl}}_{\mathsf{R}^i}) \tag{12}$$

To conveniently align with dependence, we refer the inputs as *features*, when they are grouped by the driver, vehicle or

route dependence. The features we refer to are as follows.

- *Driver dependent features*: $\mathbf{x}_{\mathsf{D},\mathsf{R}^i}$, $\mathbf{x}_{\mathsf{D},\mathsf{V}}$, $\mathbf{c}_{\mathsf{D},\mathsf{R}^i}$, $\mathbf{c}_{\mathsf{D},\mathsf{V}}$.
- *Vehicle dependent features*: $\mathbf{c}_{\mathsf{V}}$, $\mathbf{x}_{\mathsf{D},\mathsf{V}}$, $\mathbf{x}_{\mathsf{V}}$, $\mathbf{c}_{\mathsf{D},\mathsf{V}}$.
- *Route dependent features*: $\mathbf{x}_{\mathsf{D},\mathsf{R}^i}$, $\mathbf{c}_{\mathsf{R}^i}$, $\mathbf{c}_{\mathsf{D},\mathsf{R}^i}$, $\mathbf{x}_{\mathsf{R}^i}$.

A number of applications can be enabled by flexibly substituting the proper features. We present several examples as follows.

1. To predict $E_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ by a different driver $(\mathsf{D}',\mathsf{V},\mathsf{R}^i)$, we can use the following equation:

$$\hat{E}_{\mathsf{D},\mathsf{V},\mathsf{R}^i} = (\mathbf{c}_{\mathsf{V}}, \mathbf{c}_{\mathsf{R}^i}, \mathbf{c}_{\mathsf{D}',\mathsf{R}^i}, \mathbf{c}_{\mathsf{D}',\mathsf{V}}) \cdot (\mathbf{x}_{\mathsf{D}',\mathsf{R}^i}, \mathbf{x}_{\mathsf{D}',\mathsf{V}}, \mathbf{x}_{\mathsf{V}}, \mathbf{x}_{\mathsf{R}^i})^T + c' \tag{13}$$

2. To predict $E_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ by a different vehicle $(\mathsf{D},\mathsf{V}',\mathsf{R}^i)$, we can use the following equation:

$$\hat{E}_{\mathsf{D},\mathsf{V},\mathsf{R}^i} = (\mathbf{c}_{\mathsf{V}}, \mathbf{c}_{\mathsf{R}^i}, \mathbf{c}_{\mathsf{D},\mathsf{R}^i}, \mathbf{c}_{\mathsf{D},\mathsf{V}'}) \cdot (\mathbf{x}_{\mathsf{D},\mathsf{R}^i}, \mathbf{x}_{\mathsf{D},\mathsf{V}'}, \mathbf{x}_{\mathsf{V}}, \mathbf{x}_{\mathsf{R}^i})^T + c \tag{14}$$

3. To predict $E_{\mathsf{D},\mathsf{V},\mathsf{R}^i}$ by a different route $(\mathsf{D},\mathsf{V},\mathsf{R}'^i)$, we can use the following equation:

$$\hat{E}_{\mathsf{D},\mathsf{V},\mathsf{R}^i} = (\mathbf{c}_{\mathsf{V}}, \mathbf{c}_{\mathsf{R}'^i}, \mathbf{c}_{\mathsf{D},\mathsf{R}'^i}, \mathbf{c}_{\mathsf{D},\mathsf{V}}) \cdot (\mathbf{x}_{\mathsf{D},\mathsf{R}'^i}, \mathbf{x}_{\mathsf{D},\mathsf{V}}, \mathbf{x}_{\mathsf{V}}, \mathbf{x}_{\mathsf{R}'^i})^T + c \tag{15}$$

4.5 Correlation Discovery in Driving Data Set

Let all the data sets in the repository be $\{\mathsf{D},\mathsf{V},\mathsf{R}^i\}$. Note that not every tuple is present in the repository. Next, we seek to identify the correlations in $\{\mathsf{D},\mathsf{V},\mathsf{R}^i\}$, such that one can identify the most similar $(\mathsf{D}',\mathsf{V},\mathsf{R}^i)$ to predict the energy consumption of $(\mathsf{D},\mathsf{V},\mathsf{R}^i)$.

In this paper, we characterize the correlation in the driving data between a pair (D,V) and $(\mathsf{D}',\mathsf{V}')$ with the same route using the concept of *dynamic time warping* (DTW) [13]. DTW is a widely used concept for determining the similarity among time series, and identifying the corresponding similar regions between two time series. DTW has been used in many applications, such as speech recognition, gesture recognition, robotics and bioinformatics.

The basic idea of DTW is to determine an optimal alignment between two time series. Consider two time series $X = (x[t])_{t=1}^{n_X}$ and $Y = (y[t])_{t=1}^{n_Y}$ of lengths n_X and n_Y respectively. A *warp path* is defined as $W = (w[k])_{k=1}^{n_W}$, where the k^{th} element is $w_k = (i,j)$, such that i is an index from time series $x[t]$ and j is an index from time series $y[t]$. n_W is the length of the warp path W, such that $\max(n_X, n_Y) \le n_W < n_X + n_Y$.

The warp path W is subject to the following constraints:

1. $w[1] = (1,1)$;
2. $w[n_W] = (n_X, n_Y)$;
3. if $w[k] = (i,j)$ and $w[k+1] = (i',j')$, then $i \le i' \le i+1$ and $j \le j' \le j+1$.

An optimal warp path (illustrated in Fig. 3) is the one with the minimum distance $\mathsf{dist}(W^*)$, defined by:

$$\mathsf{dist}(W^*) = \arg\min_{w} \sum_{k=1}^{n_W} d(w[k]) \tag{16}$$

where $d(w[k])$ is the distance of the coordinates (i,j) of the k^{th} element in W.

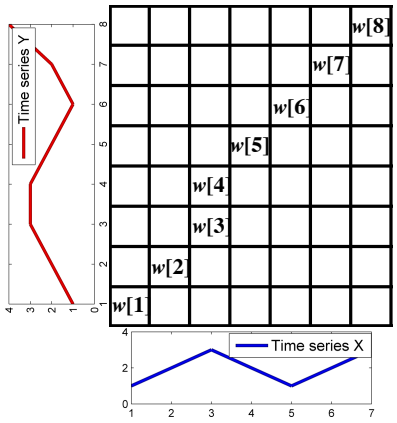


Figure 3: An illustration of warp path

A simple approach to determine an optimal warp path between two time series is to use dynamic programming. But there are other more efficient algorithms with linear running time [13].

Let $v_{\mathsf{D},\mathsf{V},\mathsf{R}^i}[t]$ be the time series of speed profile for tuple $(\mathsf{D},\mathsf{V},\mathsf{R}^i)$. For each pair of $(\mathsf{D},\mathsf{V},\mathsf{R}^i)$ and $(\mathsf{D}',\mathsf{V}',\mathsf{R}^i)$, let

$$\chi^{\mathsf{R}^i}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')} = \mathsf{dist}(W^*) \tag{17}$$

where W^* is the minimum-distance warp path between the time series $v_{\mathsf{D},\mathsf{V},\mathsf{R}^i}[t]$ and $v_{\mathsf{D}',\mathsf{V}',\mathsf{R}^i}[t]$.

Let $\mathscr{R}(\mathsf{D},\mathsf{V})$ be a set of route segments that have speed profiles recorded with (D,V). Namely, if $\mathsf{R}^i \in \mathscr{R}(\mathsf{D},\mathsf{V})$, then the speed profile $v_{\mathsf{D},\mathsf{V},\mathsf{R}^i}[t]$ exists in the repository.

Define a correlation metric between each pair of (D,V) and $(\mathsf{D}',\mathsf{V}')$ by the average minimum warp path distance over all recorded segments:

$$\bar{\chi}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')} = \frac{\sum_{\mathsf{R}^i \in \mathscr{R}(\mathsf{D},\mathsf{V}) \cap \mathscr{R}(\mathsf{D}',\mathsf{V}')} \chi^{\mathsf{R}^i}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')}}{|\mathscr{R}(\mathsf{D},\mathsf{V}) \cap \mathscr{R}(\mathsf{D}',\mathsf{V}')|} \tag{18}$$

Note that $\bar{\chi}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')} = \infty$, if $\mathscr{R}(\mathsf{D},\mathsf{V}) \cap \mathscr{R}(\mathsf{D}',\mathsf{V}') = \varnothing$.

In this paper, we will use $\bar{\chi}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')}$ to characterize the similarity between each pair of (D,V) and $(\mathsf{D}',\mathsf{V}')$. We find the tuple $(\mathsf{D}',\mathsf{V}')$ with the smallest value $\bar{\chi}_{(\mathsf{D},\mathsf{V}),(\mathsf{D}',\mathsf{V}')}$ to estimate energy consumption of (D,V). We also employ k-nearest neighbors (k-NN) clustering to find the k most similar speed profiles of (D,V).

EXAMPLE 1. *In Fig. 4, the speed profiles of three drivers* $(\mathsf{D}_1,\mathsf{V}_1),(\mathsf{D}_2,\mathsf{V}_2),(\mathsf{D}_3,\mathsf{V}_3)$ *on the same route* R^1 *are plotted. Table 2 shows the minimum warp path distances between driver* $(\mathsf{D}_1,\mathsf{V}_1)$ *and other drivers. We observe that smaller minimum warp path distance indeed shows closer similarity in the speed profiles;* $(\mathsf{D}_1,\mathsf{V}_1)$ *is more similar with* $(\mathsf{D}_2,\mathsf{V}_2)$ *than* $(\mathsf{D}_3,\mathsf{V}_3)$.

	Minimum warp path distance
$\chi^{\mathsf{R}^1}_{(\mathsf{D}_1,\mathsf{V}_1),(\mathsf{D}_2,\mathsf{V}_2)}$	1.1385
$\chi^{\mathsf{R}^1}_{(\mathsf{D}_1,\mathsf{V}_1),(\mathsf{D}_3,\mathsf{V}_3)}$	1.3883

Table 2: Minimum warp path distance

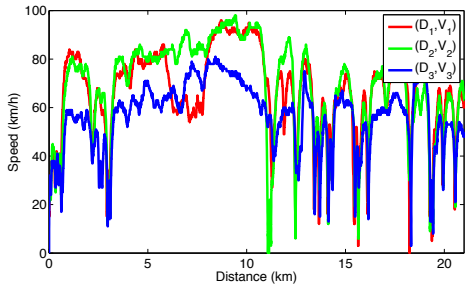


Figure 4: Speed profiles of three drivers on the same route.

5. EXPERIMENTS

We performed experiments to empirically evaluate the performance of our approach. The observations are reported in this section.

5.1 Experiment Setup

5.1.1 Routes

We select a range of different roads, as depicted in Fig. 5 and summarized in Table 3. R_1 has light traffic, consisting of both highway and suburban roads. R_2 has heavier traffic with traffic lights, and thus results in the much longer idle time seen in Table 3. R_3 and R_4 are similar in terms of traffic, with no need to stop. However, R_3 has a number of traffic signs and speed bumps, which force drivers to slow down, whereas R_4 has no restriction.

Route	Average Route Length (KM)	Average Driving Duration (sec)	Average Idle Duration (sec)
R_1	21.49	1356.4	10.4
R_2	10.36	761.8	398.6
R_3	6.78	412	0
R_4	8.18	450.8	0

Table 3: The routes in the experiments.

5.1.2 Vehicles

Four vehicles are used in the experiments, as summarized in Table 4, which consist of three ICE vehicles and one EV. The pictures of the vehicles are shown in Fig. 6.

Vehicle	Maker	Model	Year	Type	Displacement
V_1	Toyota	Yaris	2014	ICE	1.5
V_2	Hyundai	Veloster	2014	ICE	1.6
V_3	Nissan	LEAF	2013	EV	NA
V_4	BMW	650i	2014	ICE	5.0

Table 4: The vehicles in the experiments.

5.1.3 Data Collection

For ICE vehicles, we collected data through Bluetooth ELM327 dongles connected to the vehicles' onboard diagnostic (OBD) ports and paired with a smartphone App. The OBD data collected include mass air-flow, manifold absolute pressure, intake air temperature and engines' RPM, which

(a) Toyota Yaris (b) Hyundai Veloster

(c) Nissan LEAF (d) BMW 650i

Figure 6: The vehicles in the experiments.

are then utilized to compute the fuel consumption rate. Furthermore, the geo-location data, accelerometer and gyroscope measurements from the smartphone are also recorded. For EV (i.e., Nissan LEAF), we rely on Android App called "LEAF Spy Pro" to collect the EV's data. The sample rate of our smartphone App is 2 Hz, whereas the one for LEAF spy pro is 0.25 Hz.

5.2 Data Sets

Fig. 7 shows some sample driving data. Fig. 7a shows the speed profiles of the EV and an ICE vehicle by the same driver on the same route. We can observe that both vehicles stopped several times due to traffic lights. Fig. 7b shows the energy consumption profiles of EV and ICE vehicles. Although the speed profiles are similar, there are remarkable differences in the energy consumption profiles. Notably, the energy consumption level decreases when the EV slows down, because of the regenerative braking that can convert kinetic energy into stored energy in battery.

Figs. 7c-7d show the driving data collected from the vehicles. The gyroscope data is used to determine the vehicle movement and orientation, which is gathered from smartphone. However, the gyroscope data is required to be aligned with vehicle's moving direction in order to obtain the corrected orientation. We use an automatic alignment algorithm to infer the vehicle's reference orientation [9, 15].

5.3 Estimation Errors

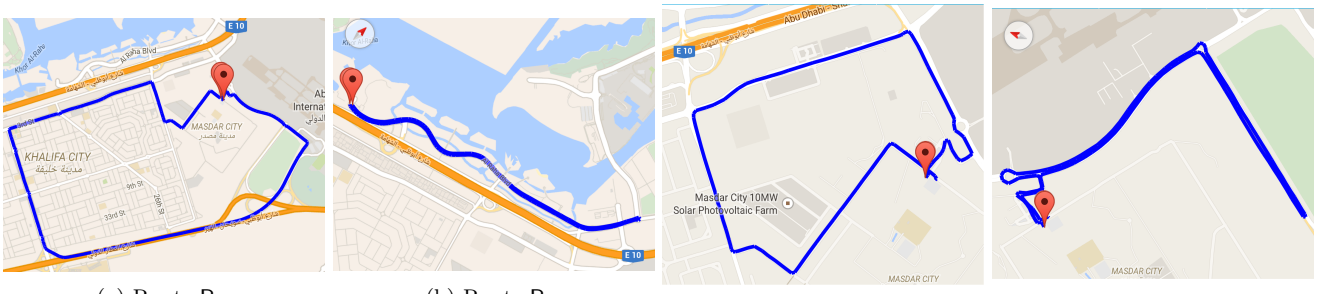
To evaluate the accuracy of estimation, we measure the deviation error of the estimated energy consumption by the per segment error:

$$\varepsilon^i = \frac{|(E^{\text{drv}}_{\text{D,V,R}^i} + E^{\text{idl}}_{\text{D,V,R}^i}) - (\hat{E}^{\text{drv}}_{\text{D,V,R}^i} + \hat{E}^{\text{idl}}_{\text{D,V,R}^i})|}{E^{\text{drv}}_{\text{D,V,R}^i} + E^{\text{idl}}_{\text{D,V,R}^i}} \tag{19}$$

and the accumulative error:

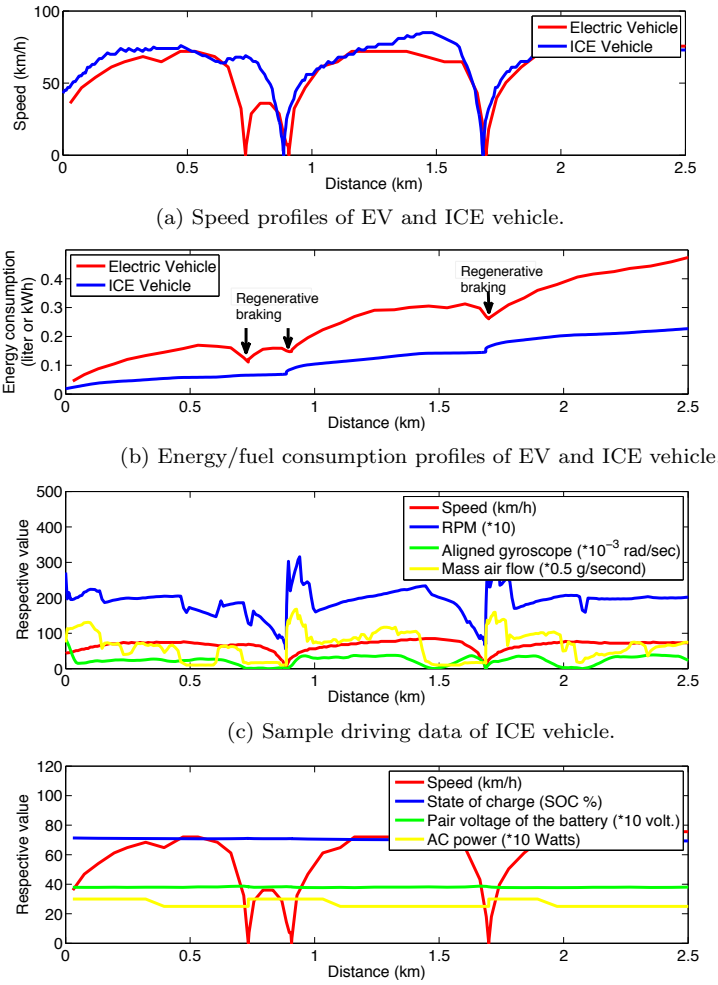
$$\varepsilon^{\text{acc}} = \frac{|E_{\text{D,V,R}} - \hat{E}_{\text{D,V,R}}|}{E_{\text{D,V,R}}} \tag{20}$$

As an example, we plot the distribution of per-segment errors of one vehicle collected from three rounds of experiments in Fig. 8. It shows that the mean error is about 3.5% and the distribution is close to a normal distribution. Since we are interested in the energy consumption of the overall trip, the accumulative error is more relevant. Fig. 9



(a) Route R_1. (b) Route R_2. (c) Route R_3. (d) Route R_4.

Figure 5: The routes in the experiments.



(a) Speed profiles of EV and ICE vehicle.

(b) Energy/fuel consumption profiles of EV and ICE vehicle.

(c) Sample driving data of ICE vehicle.

(d) Sample driving data of EV.

Figure 7: Sample driving data.

shows the accumulative error against travelled distance. We observe that the accumulative error fluctuates against travelled distance, but there is a general trend of decreasing error. This is due to the fact that the positive and negative deviations can offset each other over a longer distance. Therefore, the accumulative error reaches a lower value after a longer distance.

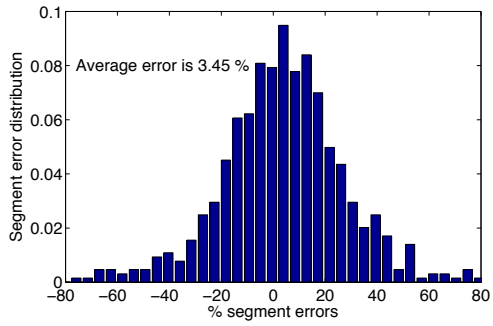


Figure 8: Distribution of per segment errors.

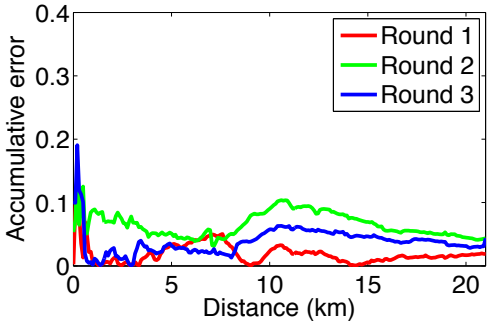


Figure 9: Accumulative error against travelled distance.

5.4 Energy Consumption Model Validation

We first use the data sets of R_1 and R_2 to obtain the coefficients of Eqns. (2)-(3). Next, we examine the accuracy of estimating the energy consumption using Eqns. (2)-(3) against the actual energy consumption. The respective accumulative error averaged over three rounds of experiments are shown in Table 5 and Fig. 10. We observe that the accumulative errors range from 0.1% to around 5%, which is sufficient to practical DTE prediction. In Fig. 11, we use other drivers' driving data to predict other drivers' energy consumption in the same route. The initial error is relatively large due to several reasons: 1) the prediction can overestimate or underestimate the accurate DTE, which creates short-term fluctuations. The prediction error will be offset by the positive and negative deviations in a long term. Therefore, the accumulative error can reach a lower value after a longer distance. 2) For vehicle D_3 (LEAF), the sample rate is relative low, therefore, the initial error is larger

than others', because the driving behavior cannot be fully captured when sample rate is low. We are working on increasing the sample rate of the LEAF which will improve the initial error.

Route	Vehicle	Driver	Average accumulative error
R_1	V_1	D_1	3.6%
R_1	V_2	D_2	0.1%
R_1	V_3	D_2	3.2%
R_1	V_3	D_3	2.0%
R_1	V_4	D_4	0.1%
R_2	V_1	D_1	4.4%
R_2	V_2	D_2	3.7%
R_2	V_3	D_2	4.7%
R_2	V_3	D_3	5.4%
R_2	V_4	D_4	5.1%

Table 5: Estimation errors for route-vehicle-driver tuples.

5.5 Correlation Discovery

Next, we obtain the correlation metrics $\bar{\chi}_{(D,V),(D',V')}$ in Table 6. The two top-most similar pairs are highlighted in bold in each row. For example, (D_1, V_1) has a small value with (D_2, V_2) and (D_3, V_3). We use k-nearest neighbor method to choose the k most similar (D', V') for each (D, V), and use their respective driving data to estimate the energy consumption of (D, V).

$\bar{\chi}_{(D,V),(D',V')}$	(D_1, V_1)	(D_2, V_2)	(D_3, V_3)	(D_2, V_3)	(D_4, V_4)
(D_1, V_1)	0	**1.0799**	**1.0214**	1.1659	1.2319
(D_2, V_2)	**1.0799**	0	1.1022	**0.9872**	1.1895
(D_3, V_3)	**1.0214**	**1.1022**	0	1.1423	1.4211
(D_2, V_3)	1.1659	**0.9872**	**1.1423**	0	1.2169
(D_4, V_4)	1.2319	**1.1895**	1.4211	**1.2169**	0

Table 6: Correlation metrics.

6. CASE STUDIES

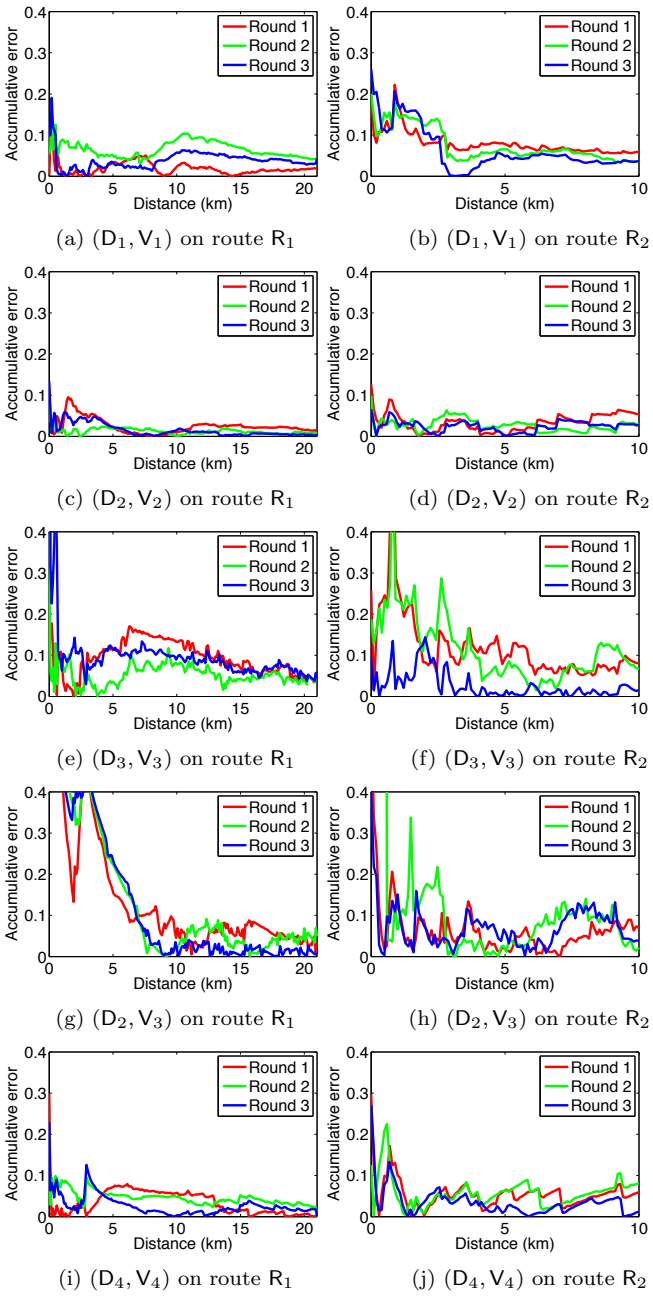
In this section, we apply our results to some concrete case studies: (1) DTE prediction, (2) characterization of driving behavior, (3) energy-efficient driving recommendations, and (4) route-level energy consumption geo-fencing and refueling planning.

6.1 DTE Prediction

Based on the correlation metrics, we can estimate the energy consumption using the driving data by the most similar driver-vehicle pairs. In particular, one can use the data of ICE vehicles (V_1, V_2, V_4) to predict the energy consumption of the EV (V_3).

In the following, five scenarios are considered:

1. Predicting (D_1, V_1) using estimators (D_2, V_2),(D_3, V_3)
2. Predicting (D_2, V_2) using estimators (D_1, V_1),(D_2, V_3)
3. Predicting (D_3, V_3) using estimators (D_1, V_1),(D_2, V_2)
4. Predicting (D_2, V_3) using estimators (D_2, V_2),(D_3, V_3)
5. Predicting (D_4, V_4) using estimators (D_2, V_2),(D_2, V_3)



(a) (D_1, V_1) on route R_1 (b) (D_1, V_1) on route R_2

(c) (D_2, V_2) on route R_1 (d) (D_2, V_2) on route R_2

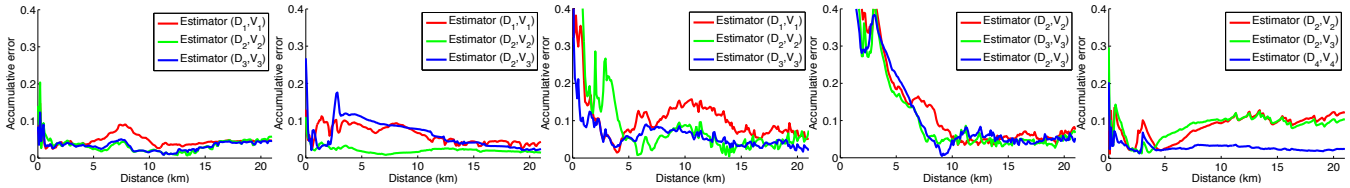
(e) (D_3, V_3) on route R_1 (f) (D_3, V_3) on route R_2

(g) (D_2, V_3) on route R_1 (h) (D_2, V_3) on route R_2

(i) (D_4, V_4) on route R_1 (j) (D_4, V_4) on route R_2

Figure 10: Estimation errors against traveled distance.

The respective accumulative errors for route R_1 are shown in Table 7, along with the prediction using its own data. Three rounds of data for each driver-vehicle pair are averaged and then used to predict others. For each driver-vehicle pair, we use the two most similar driver-vehicle pairs, as well as its own data. The diagonal entries of the Table 7 are the predictions based on the own data. Except driver (D_4), the estimation errors are observed to be low. Driver (D_4) drives much more aggressively compared to other drivers, and hence, its correlation metrics have larger values. Therefore, the estimation errors are larger for (D_4) using other drivers.



(a) (D_1, V_1) (b) (D_2, V_2) (c) (D_3, V_3) (d) (D_2, V_3) (e) (D_4, V_4)

Figure 11: Estimation errors against traveled distance.

Error	Estimator (D_1, V_1)	(D_2, V_2)	(D_3, V_3)	(D_2, V_3)	(D_4, V_4)
(D_1, V_1)	4.7%	5.8%	4.6%	..	..
(D_2, V_2)	4.3%	1.4%	..	2.4%	..
(D_3, V_3)	5.2%	7.4%	2.0%	..	..
(D_2, V_3)	..	5.9%	6.7%	5.3%	..
(D_4, V_4)	..	12.1%	..	10.3%	2.5%

Table 7: Estimation errors.

6.1.1 *DTE Prediction for EV*

In particular, we study the performance of DTE prediction for EV (i.e., Nissan LEAF). The data collected from EV includes:

1. State of charge (SOC), denoted by $\mathcal{S}$, indicates how much electricity remains in the battery.
2. Initial capacity of the battery, denoted by $\mathcal{B}_A$.
3. Battery pack voltage when driving, denoted by $\mathcal{B}_V$.

The remaining energy ($\Delta\mathcal{E}^t$) in battery at time t is given by:

$$\Delta\mathcal{E}^t = \mathcal{S}^t \times \mathcal{B}_A \times \mathcal{B}_V^t \tag{21}$$

If the future average power intensity ($\bar{\mathcal{P}}$) is known, then DTE is given by:

$$\text{DTE} = \frac{\Delta\mathcal{E}^t}{\bar{\mathcal{P}}} \tag{22}$$

We especially compare the DTE prediction based on our energy consumption model with the on board DTE meter on Nissan LEAF, which is captured by a camera mounted over the dashboard. Fig. 12 shows our estimated DTE and on board DTE meter readings in the experiment. The reference line in the Fig. 12 denotes the true DTE as the distance goes. We observe that our approach gives a considerably more accurate prediction than the onboard meter.

6.2 Classification of Driving Behavior

The correlation metrics between every pair of drivers can provide a distance matrix. One can apply standard clustering techniques on the set of driver-vehicle pairs in such a metric space. The clustered data set can classify energy-efficient driving behavior. Fig. 13 shows the clustered data set based on Table 6. We observe that energy-efficient driver-vehicle pairs (colored in green) tend to stay closer in the correlation metric space. Likewise, one can also obtain the cluster of energy-inefficient driver-vehicle pairs, and a complete classification of driving behavior.

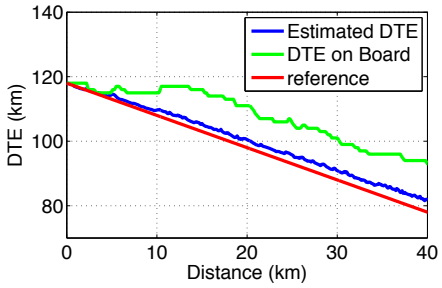


Figure 12: Comparison between our estimated DTE and onboard DTE meter reading.

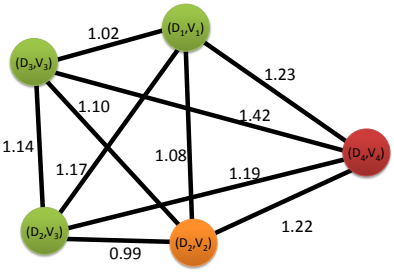


Figure 13: Clustering of driver-vehicle pairs based on correlation metrics. Green nodes are the energy-efficient driver, whereas the red node is energy-inefficient driver.

6.3 Driving Recommendations

Our energy consumption model can be applied to provide recommendations of energy-efficient driving. Assuming other parameters are constant (e.g., acceleration, deceleration, and gyroscope data), the energy consumption is only dependent on the vehicle speed in our model. For example, Fig. 14 shows the energy intensity of several driver-vehicle pairs at different speeds. It is observed that there is a minimum point for ICE vehicles (namely, the least energy-consuming speed), and an energy intensity increasing with speed for EV (because electric motors operate more efficiency at low speeds). For ICE vehicles, we can recommend that the driver to maintain at the least energy-consuming speed obtained from the model. For EV, the optimal speed will be the smallest possible speed that can arrive at the destination before a deadline.

6.4 Geo-fencing and Planning

Geo-fencing depicts the geographical range before the energy of vehicle is exhausted. Traditionally, geo-fencing is estimated only considering the geographical distance. With the information from our model, more detailed route-level

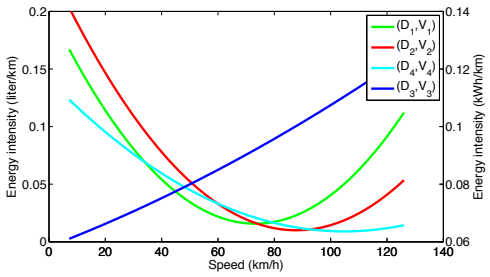


Figure 14: Energy intensity of vehicles at different speed according to our energy consumption model.

energy consumption geo-fencing can be constructed, for example, as seen in Fig. 15. The route-level energy consumption geo-fencing is constructed in the following manner. We first obtain the map data from OpenStreetMap (OSM). We employ our model to estimate the energy consumption at each point along a route. The least energy consumption required to reach a particular point can be estimated by A* algorithm, considering multiple route alternatives. The least energy consumption will be visualized on top of OSM data to provide route-level energy consumption geo-fencing. A critical application energy consumption geo-fencing is to enable informed decisions for refueling.

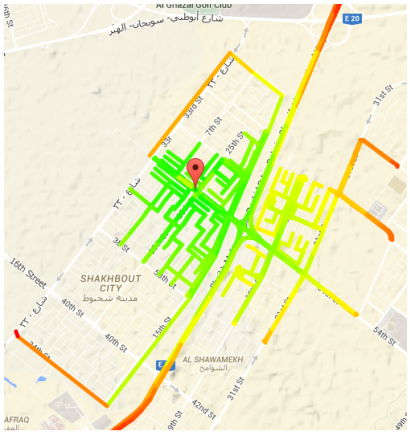


Figure 15: Route-level energy consumption geo-fencing.

7. CONCLUSION AND FUTURE WORK

An effective approach has been proposed to integrates the vehicle data gathered from diverse drivers and vehicles for personalized applications of improving driving energy efficiency. Our system provides a unifying approach for both ICE vehicles and EVs. The advantages include identifying the features according to the driver, vehicle and route dependence. The processed data can flexibly support diverse applications, such as DTE prediction, green telematics and energy-efficient route planning, and classification of energy efficient driving behavior.

Future work will include integration of extensive data from large-scale datasets such as those to be available from CloudThink [1] as well as from expanded GIS databases(e.g. the geometric of the roads). We are working to integrate our methodology to open platforms (e.g. OpenStreetMap). Moreover, hybrid vehicles or plug-in hybrid vehicles are not considered in this work, because of the lack of proper software to interoperate the proprietary data about battery state in these vehicles. Proper software will be sought in future to allow further experiments on these vehicles.

8. ACKNOWLEDGEMENT

We would like to thank the reviewers and the shepherd (Dr. Sarvapali D.(Gopal) Ramchurn) for helpful comments and suggestions.

9. REFERENCES

[1] CloudThink. `http://cloud-think.com/`, 2015. [Online; accessed 19-February-2015].

[2] Michael J Burke, Nick Sarafopoulos, and Viet Q To. Electronic system and method for calculating distance to empty for motorized vehicles, 1994. US Patent 5,301,113.

[3] Andrew T. Campbell, Shane B. Eisenman, Nicholas D. Lane, Emiliano Miluzzo, Ronald A. Peterson, Hong Lu, Xiao Zheng, Mirco Musolesi, Kristóf Fodor, and Gahng-Seop Ahn. The rise of people-centric sensing. *IEEE Internet Computing*, 12(4):12–21, July 2008.

[4] S. Dornbush and A. Joshi. StreetSmart traffic: Discovering and disseminating automobile congestion using VANET's. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 11–15, April 2007.

[5] Raghu K. Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F. Abdelzaher. Greengps: A participatory sensing fuel-efficient maps application. In *Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services*, MobiSys '10, pages 151–164. ACM, 2010.

[6] Torsten J. Gerpott and Sabrina Berg. Explaining customers' willingness to use mobile network-based pay-as-you-drive insurances. *Int. J. Mob. Commun.*, 11(5):485–512, October 2013.

[7] J.G. Hayes, R.P.R. de Oliveira, S. Vaughan, and M.G. Egan. Simplified electric vehicle power train models and range estimation. In *Vehicle Power and Propulsion Conference (VPPC), 2011 IEEE*, pages 1–5, Sept 2011.

[8] Xiping Hu, Victor C.M. Leung, Kevin Garmen Li, Edmond Kong, Haochen Zhang, Nambiar Shruti Surendrakumar, and Peyman TalebiFard. Social drive: A crowdsourcing-based vehicular social networking system for green transportation. In *Proceedings of the Third ACM International Symposium on Design and Analysis of Intelligent Vehicular Networks and Applications*, DIVANet '13, pages 85–92. ACM, 2013.

[9] Prashanth Mohan, Venkata N. Padmanabhan, and Ramachandran Ramjee. Nericell: Rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM Conference on Embedded Network Sensor Systems*, SenSys '08, pages 323–336. ACM, 2008.

[10] Austin Louis Oehlerking. StreetSmart: modeling vehicle fuel consumption with mobile phone sensor data through a participatory sensing framework (master's thesis), 2011.

[11] Jungme Park, Zhihang Chen, L. Kiliaris, M.L. Kuang, M.A. Masrur, A.M. Phillips, and Y.L. Murphey. Intelligent vehicle power control based on machine learning of optimal control parameters and prediction of road type and traffic congestion. *Vehicular Technology, IEEE Transactions on*, 58(9):4741–4756, Nov 2009.

[12] L Rodgers, E Wilhelm, and D Frey. Conventional and novel methods for estimating an electric vehicle's "distance to empty". In *Proc. of the ASME 2013 International Conference on Advanced Vehicle Technologies*, 2013.

[13] Stan Salvador and Philip Chan. Toward accurate dynamic time warping in linear time and space. *Intell. Data Anal.*, 11(5):561–580, October 2007.

[14] Chien-Ming Tseng, Sohan Dsouza, and Chi-Kin Chau. A social approach for predicting distance-to-empty in vehicles. In *Proceedings of the 5th International Conference on Future Energy Systems*, e-Energy '14, pages 215–216. ACM, 2014.

[15] Yan Wang, Jie Yang, Hongbo Liu, Yingying Chen, Marco Gruteser, and Richard P. Martin. Sensing vehicle dynamics for determining driver phone use. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '13, pages 41–54. ACM, 2013.