Mining Geographic Data for Fuel Consumption Estimation

Vitor Daniel Ferreira da Cunha Ribeiro

Dissertation conducted under the supervision of *Prof. Ana Cristina Costa Aguiar*
and PhD Student *João Rodrigues*

31^{st} January, 2013

---

1. Motivation

Mobility has always been an important factor in human history and society today highly depends on it. Modern cities possess large and complex transportation networks, whose operating conditions sometimes preclude the minimization of human mobility associated costs, and contribute to an unsustainable growth of carbon emissions. Even in this technological era people live in, with space and mobile technology, the advantages or disadvantages of each mobility option is not always direct or available to a user in real time. In this scenario, sometimes a user is imbued with finding the best option with limited resources. Modern vehicles manufactured in the United States and Europe are required by law to be equipped with On Board Diagnostics (OBD) technology to enable emission control and testing. With an OBD device it is also possible to calculate fuel consumption. Still, if a person wants to know his fuel consumption, either he owns and understands the OBD technology, and has the availability and knowledge to install an OBD interface in his vehicle, or he uses a software, that normally requires user input. Furthermore, not all vehicles posses an interface with trip assessment information, and even if they do, a user has little chance of using that information in a collaborative fashion. In an effort to reduce both emissions and, mostly important, costs, since money is a global motivator, an innovative solution for fuel consumption estimation is proposed, also taking advantage of the opportunities generated by the growing market of mobile devices.

2. Goal

The goal of this thesis is to provide an innovative solution with an algorithm that enables estimation of fuel consumption from GPS data, present in every modern smartphone, using regression methods to discover the relationship between GPS and OBD gathered data. A large dataset of GPS and OBD data was gathered by various volunteers, using an Android mobile application as a gathering unit.

3. Work Description

In the development of this thesis it was necessary to study engine mechanics, the OBD standard, fuel consumption and emission models, GPS technology, data mining methodologies, mathematical and statistic methods, and regression and machine learning approaches. To gather data, an OBD module was developed to extend the MyDrivingDroid Android application. The following subsections detail the most important areas of research, technologies and methodologies present in the thesis.

3.1. Fuel Consumption Calculation

Fuel consumption is a ratio of fuel consumed per distance travelled. In most European vehicles it is normally represented as litres per 100 kilometres. The algorithm is aimed at four-stroke petrol and diesel engines. Although the basic designs of petrol and diesel engines are similar, the mechanics are different. A gasoline engine compresses its fuel and air charge and then initiates combustion by the use of a spark plug. A diesel engine just compresses air until the combustion chamber reaches a temperature for self-ignition to occur. While the OBD standard does not provide a fuel consumption parameter, it provides other values that enable its calculation. Since engine mechanics are different and the available OBD values depend on the vehicle, various formulas were devised.

3.2. MyDrivingDroid Bluetooth OBD Module

The MyDrivingDroid project developed at Instituto de Telecomunicações is aimed at data gathering taking advantage of an Android Smartphone's embedded sensors. A service was developed to connect with a Bluetooth OBD interface, plugged in a vehicle, and gather automotive data. The OBD communications follow the SAE J1979 and ISO 15031-5 protocols.

3.3. Deriving Acceleration And Steepness From The GPS

To obtain both acceleration and steepness from the gathered data a least square error method was used. For a given set of data, the least squares method obtains the best-fit curve that produces the minimum sum of the squared deviations. The acceleration in each second was derived from a least square error curve between the 3 surrounding speeds. The inclination least square algorithm uses more points. The number of points used satisfies a minimum radius from the original position and a minimum of 5 points required. This means that the radius gradually increases if the number of obtained points does not meet the required minimum.

3.4. OBD Data Interpolation

Since vehicles have different response rates it is not guaranteed that all the required OBD responses are available in all seconds. In order to prevent seconds in the middle of the data with no fuel consumption values, a simple solution was implemented which consisted in linearly interpolating the OBD response values with the GPS timestamps.

3.5. Regression

Exploratory data analysis has to be dealt with carefully at the risk of leading to the wrong conclusions. For a careful analysis the Velleman and Hoaglin guide of the four fundamental steps known as *'the four **R**'s of exploratory analysis'* is used. Since the goal is to obtain fuel consumption from GPS values, in the dataset obtained, fuel consumption serves as a dependent or response variable, and GPS velocity, acceleration and inclination as independent or stimulus variables. A linear model using the ordinary least squares (OLS) method to fit a line in a dataset was used, but since not all relationships are linear, a third order polynomial was experimented which yielded a better fit. In order to support this statement, both models were fitted in each vehicle's dataset and the dataset with all vehicles together, and the residuals were compared. Also for every tested fit, it was experimented with and without outliers to help measure the influence these 'distributional monsters' can have. In the end, it was concluded that fuel consumption can be approximately estimated with: $FuelConsumption \approx 10.1200 + 1.0980 * Inclination + 5.9010 * Acceleration - 0.8440 * Speed + 0.0354 * Speed^2 - 0.0003 * Speed^3$.

4. Conclusions

It was possible to conclude that some OBD applications provided useful information for OBD data handling and, alongside the analysed emission models, a valuable guideline for the implementation of the fuel consumption algorithm from OBD data. The ELM327 manual also proved fundamental for the implementation and comprehension of the OBD communication protocols. The models also allowed a better understanding of emissions and vehicle dynamics, led to the learning of products and projects that use them, and allowed to understand their importance and impact on a society that highly depends on mobility. The experience with the Android platform also proved to be challenging but fruitful, as the barriers that emerged resulted in research about the components like Bluetooth, and even different smartphones, and a closer look into the Android source code, which enabled an even bigger comprehension of this platform. Since there is a lot of focus on the GPS, there was also a lot of research into the space technology and its navigation and localization methods. The result was a regression model that provides fuel consumption in real time.