Statistical analysis of road-vehicle-driver interaction as an enabler to designing behavioural models

T Chakravarty¹, A Chowdhury¹, A Ghose¹ and C Bhaumik¹, P. Balamuralidhar²
¹TCS Innovation Labs, BIPL, Salt Lake, Sector-V, Kolkata, India
²TCS Innovation Labs, Abhilash, Plot-96, EPIP, Whitefield, Bangalore

E-mail: tapas.chakravarty@tcs.com

Abstract. Telematics form an important technology enabler for intelligent transportation systems. By deploying on-board diagnostic devices, the signatures of vehicle vibration along with its location and time are recorded. Detailed analyses of the collected signatures offer deep insights into the state of the objects under study. Towards that objective, we carried out experiments by deploying telematics device in one of the office bus that ferries employees to office and back. Data is being collected from 3-axis accelerometer, GPS, speed and the time for all the journeys. In this paper, we present initial results of the above exercise by applying statistical methods to derive information through systematic analysis of the data collected over four months. It is demonstrated that the higher order derivative of the measured Z axis acceleration samples display the properties Weibull distribution when the time axis is replaced by the amplitude of such processed acceleration data. Such an observation offers us a method to predict future behaviour where deviations from prediction are classified as context-based aberrations or progressive degradation of the system. In addition we capture the relationship between speed of the vehicle and median of the jerk energy samples using regression analysis. Such results offer an opportunity to develop a robust method to model road-vehicle interaction thereby enabling us to predict such like driving behaviour and condition based maintenance etc.

1. Introduction

Telematics is defined as a technology that relates to the collection of moving vehicle data and transporting these over telecom network to a central server for further analysis. Apart from a sensor enabled dedicated device, some emerging applications demand use of Smartphones also [1]. Insurers analyse these data in order to identify a given driving behaviour in terms of potential risk induced. The analysis must lead to quantitative figures that are used for underwriting purpose. On a different context, fleet managers may use such telematics application to identify safe driving behaviour on daily basis; in fact, fleet managers may like to identify aberrations (on a particular trip) with respect to a baseline behaviour where the baseline behaviour is classified based upon a-priori knowledge; often captured as part of blind profiling. Analysis of the data being collected on daily basis leads to the ability to create predictive models about the three-way interaction between road, vehicle and driver. This paper presents some initial results of such exercise being carried out on an office bus. In the following sections, the test setup is explained along with the observations and computed inference.

¹ To whom any correspondence should be addressed.
2. Experiment and observation

In this section we discuss the setup and method used for acquisition and analysis of the physical signatures related to the road-vehicle interaction. Towards such experimentation, we deployed ‘ConnectPort® X5 R’ (a telematics gateway device for Vehicle Area Network from M/s Digi International Inc.) in the office bus meant for the associates transporting to TCS Siruseri (India) campus. This device collects location (GPS), speed, time and accelerometer data. A python script is implemented to collect the 3-axis raw acceleration data, pre-process it before sending the processed data to a device cloud over cellular network. The bus travels 5-days a week and a large data set has been collected over last four months. It is also to be noticed that the route and the timings remain identical for the given bus.

2.1 Data pre-processing

The typical trip duration is approximately 90 minutes. To capture meaningful signatures of the road-vehicle interaction, it is needed to measure the vibrations at sampling rate of 20Hz or higher. Thus the raw data size for a single trip alone becomes significantly large thereby resulting in high cellular charges. To mitigate this issue, we pre-processed the raw data by computing jerk energy (JE) for every 20 acceleration samples and sending only the JE values to device cloud.

Let us assume that $a_1, a_2, \ldots, a_n$ be the consecutive discrete acceleration samples at time $t_1, t_2, \ldots, t_n$ where $\Delta t = t_n - t_{n-1}$ for uniform sampling rate. Then jerk, i.e. rate of change of acceleration (m/sec$^3$) is defined as:

$$J_i = \frac{a_{i+1} - a_i}{\Delta t} \quad \forall 1 \leq i \leq n-1$$

(1)

Thus, ‘jerk energy’ is

$$JE_i = J_{i1}^2 + J_{i2}^2 + \ldots + J_{i19}^2$$

(2)

where, ‘s’ represents a time-window of 1 sec. The JE computation takes place inside the telematics device itself. Subsequently, the computed JE values are uploaded. These JE data are being collected on regular basis. Further analysis is carried out for trip level data set as well as aggregated (over week & month) data; in order to identify any behavioral patterns that are unique to the given interaction between the three entities namely rough road, vehicle and the driver. In our analysis, we utilized the statistical tool called MINE (maximal information-based nonparametric exploration) [3,4] to evaluate relationship between the known parameters and measured variables like JE. In the tool MINE, ‘maximal information coefficient (MIC)’ is a measure of ‘two variable dependence’. Its value ranges between 0 (no dependency) and 1 (dependent). In the collected data, we have 19 JE values that correspond to one location and one speed attribute. Intuitively it is suspected that the jerk energies corresponding to vertical vibration (+Z axis) and the average speed must be related. Statistical measures like mean, median and maximum of 19 jerk energy samples are calculated to get one representation signature in order to pair with the corresponding average speed. It was found that value of maximum information coefficient (MIC) was highest for the pair of ‘median JE’ and ‘average speed’. A representative result is presented in table 1.

| Variable | MIC (strength) | MIC-p^2 (nonlinearity) | MAS (non-monotonicity) |
|----------|----------------|------------------------|------------------------|
| speed    | Median of JE   | 0.74                   | 0.16                   | 0.06                   |

Table 1: Measure of relationship between average speed (X) and JE (Y) for 4th April, 2013

The MIC value leads us to establish functional relationship between speed and JE. Based on high MIC score for the collected real data, we fitted regression equation for average speed and JE. Figure 1 is an example scatter plot for jerk energy vs average speed of one trip. Included in the scatter plot, is the fitted curve representing the mathematical relationship between the above two variables. From
figure 1, it is clear that even at very low speeds, the measured JE is about 200 whereas the system noise (in terms of JE) in the telematics device is measured to be approximately 50. Thus, there exists an offset value in JE that is entirely a particular signature of the vehicle itself (boarders alighting is one possible cause). This result is used later, when we attempt to measure vehicle degradation. In figure 1, we observe multiple JE values for any small band of speed variation. This is attributed to different road roughness as well as driving behaviour. After deriving suitable regression lines for each day of the three months; April to June, we derived a monthly equation of predicting JE based on the values of vehicle speed. We found out that most of the fitted regression lines were quadratic in nature. After collecting all the coefficients (a, b, c), we took Trimmed Mean for each of them to obtain the monthly coefficients. This is given as:

\[ v = 238.53 - 4.42u + 0.56u^2 \quad \text{&} \quad \sigma^2 = 151.1 \]  

(3)

2.2 Hazard rate

Analysis of JE data shows that ‘x = 1/JE’ follows Weibull distribution with shape parameter k >1. From the observed data, we can derive the scale parameter \( \lambda \) and shape parameter \( k \). Additional observations on different vehicles inform us that these two parameters vary primarily on driving behavior and are weakly-persistent on vehicle condition. For Weibull distribution the failure rate \( h(x) \) (or hazard rate) is given as

\[ h(x) = \frac{k}{\lambda} \left( \frac{x}{\lambda} \right)^{k-1} \]  

(4)

From the acquired data acquired for a given trip, k and \( \lambda \) and \( h(x) \) are computed. Traditionally, Weibull distribution is used to measure mean-time-to-failure. But we have replaced time ‘t’ by ‘x’ = inverse of jerk energy; thus \( h(x) \) represents the probability of the vehicle system failing when subjected to a small jerk, having survived large initial ‘jerks’. We characterize the above described vehicle at \( x= 1/300 \) where 300 is the nominal JE value for a vehicle at low speeds from the above regression model. Figure 2 shows the variation of scale & shape parameters from measured +Z axis accelerometer data on consecutive days.

![Figure 1: Regression equation between u (speed in Km/hr) and JE (m²/sec⁶)](image1)

![Figure 2: Variation of Weibull parameters computed from 3 months data](image2)

3. Results

The setup described as above is used to calculate hazard rate (HR) for each trip: journey to office and back. In figure 3, the measured hazard rate for the bus is presented for two months. The mean HR (of monthly samples) & the upper/lower variability levels can potentially display a degradation of the suspension-damper system of the vehicle. In the previous work, Chakravarty et al [5] presented the
method to model complex road-vehicle interaction. The given method is also used to simulate HR for a large vehicle when interacting with different road types; from very good to bad. The results are displayed in figure 4. Here, HR values are computed for speeds of 20km/hr and 30km/hr. These results are indicative only. When HR is measured for a real drive, such computed HR can be compared with these indications to estimate the road type. It is known that the ISO proposed geometric mean of spectral densities for different road grades increases exponentially with road type [6]. From the values of hazard function it is clear that road roughness attribute is correctly captured by hazard function. Now, higher HR value also denotes vehicle suspension degradation. When a vehicle continues to traverse same road on a daily basis, an out-of-range value of HR can be directly associated with the nature of driving for that trip. Similarly, it is possible to classify potentially hazardous driving by comparing the measured HR with the peers on the same road.

5. Concluding remarks
In this paper, methods of analysing the physical signatures obtained from complex road-vehicle-driver interaction are presented. It is shown that the vertical vibration signatures of a moving vehicle offer strong insights into the nature of such interaction. It is possible to build predictive models of such interaction so that behavioural aberrations can be identified even for a smaller data sample.

Reference

[1] Ghose A, Biswas P, Bhaumik C, Sharma M, Pal A, Jha A 2012 *IEEE Int. Conf. on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, vol., no., pp.489,491
[2] Bilal S, Pal A, Srinivasarengan K, Balanmuridhar P, 2012 *IEEE 6th Int. Conf. on Sensing Technology (ICST)*, Kolkata p 8-12
[3] http://www.exploredata.net/
[4] http://www.sciencemag.org/content/suppl/2011/12/14/334.6062.1518.DC1
[5] Chakravarty T, Srinivasarengan K, Roy S, Bilal S, Balamuralidhar P, 2013 *IOP Journal of Physics: Conference Series*, Vol. 410(2013) 012065
[6] Wong J, 2001 *Theory of Ground Vehicles* (Wiley-Interscience)