Scoring mechanism and journey quality detection based on statistical property of vehicle accelerometer data

Arijit Chowdhury
TCS Innovation Labs
Tata Consultancy Services
1B Eco space, New Town, Kolkata - 700156
Phone: +918274907402
arijit.chowdhury2@tcs.com

Tapas Chakravarty
TCS Innovation Labs
Tata Consultancy Services
1B Eco space, New Town, Kolkata - 700156
Phone: +919231000286
tapas.chakravarty@tcs.com

P. Balamuralidhar
TCS Innovation Labs
Tata Consultancy Services
unit IV, no 96 Avhilash SDC, EPIP Industrial estate Whitefield Bangalore
Phone: +919448089556
balamurali.p@tcs.com

ABSTRACT

In this paper, authors have presented a model based classification of the trips of vehicle by analyzing the vehicle vertical vibration measurements by using telematics for intelligent transport system. During the motion of the vehicle, road-vehicle-driver interactions create vertical movements of the vehicle. This vertical movements and vertical accelerations are harmful for shock absorbers and tires of a vehicle. Authors have proposed a data collection and statistical analysis method to obtain relationship between the observed data (using MINE tool and MIC [1]) and server side analytics for the classification of trip. Using an on-board diagnostic system (in our case 'ConnectPort® X5 R' manufactured by M/s Digi International Inc. is used) or smartphone accelerometer reading for vertical axis with location (GPS), velocity and time are captured and processed. Also at a mini segment (small section of a journey defined later in the paper) level, the proposed method will be used to classify vehicle vibration as normal, low and high taking into consideration the velocity of the vehicle. Regression equation is used to identify the relationship between Jerk energy (JE) i.e. rate of change of acceleration and average speed of the vehicle. That analysis is further used to identify low, normal and high JE values for a velocity and classify journey at a mini segment level.

1. INTRODUCTION

Cyber Physical Systems consist of using digital computing power to control, use and/or monitor physical components in real time. With the advancement made in digital technology and communication systems, it is possible to implement model based component to monitor a physical process and its state. With the power of computation available in today's smartphone [2] it is possible to monitor sensor data and transmit them over wireless channel to remote server for further processing. We propose a model based framework for data collection, processing and analysis with reusable components. Proposed methods are robust and computationally efficient and simple.

In the proposed model as described in figure 1, we combine the experimental test bed, and mathematical models to generate inference about the physical process. Then statistical inference is used to further interpret data and develop a scoring mechanism for a journey.

We have used model based analysis to identify patterns in collected dataset and then used statistical analysis to predict and quantify nature of a vehicle journey. Collecting data for entire trip and analyzing them in real time will greatly enable maintenance of vehicle and monitoring driver's performance for of a large set of vehicles. Thus this setup will be greatly useful for large number of car/bus.

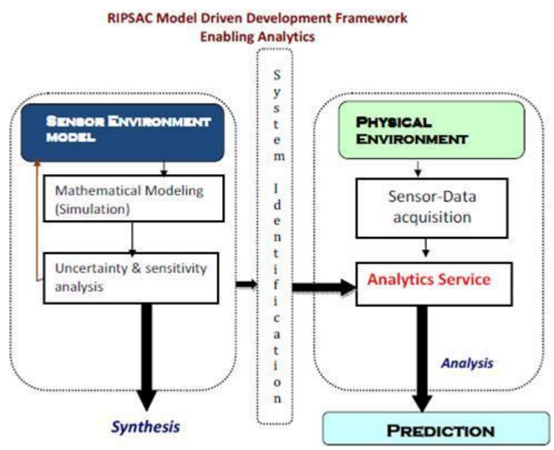

figure 1a

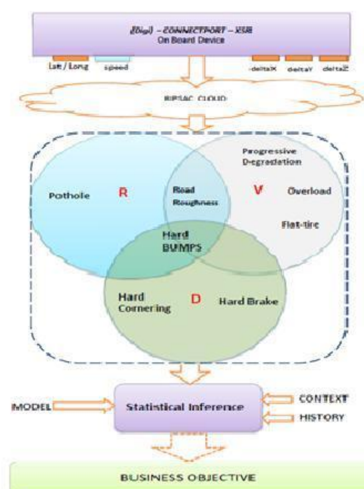

figure 1b

Figure 1: architecture of the system

2. PREVIOUS WORK

This work continues in the line of [3] which use dedicated device to capture 3 axis accelerometer and GPS data. Further [3] analyses captured data to find driving pattern. In [2] use of smartphone as internet connected sensor is used for data collection and analysis.

With the emergence of pay as you drive (PAYD) model for car insurance purpose which links driving behavior to insurance premium ways of obtaining driving pattern from different sensor data is gaining prominence. Johnson and Trivedi [4] discusses the driving style classification as non-aggressive (typical) and aggressive. Then use data from accelerometer and other sensor for driving style analysis from road safety point of view. Further Handel in [5] describes a way to find gearshift in car using accelerometer data. Gear change serves as an important parameter in determining driving pattern. Liaw [6] uses Fuzzy logic for driving pattern analysis. Based on above research work we propose a simple algorithm which gives driving scores for small section of trips to get insight into how the car was driven across an entire journey, as presented in section 4.

3. METHODOLOGY AND SYSTEM

'ConnectPort® X5 R' manufactured by M/s Digi International Inc is used to collect acceleration data. Smartphones can also be used for that purpose. Sampling frequency of 20 Hz is used to capture real data to test our algorithm. In the device itself python script is deployed to collect data. Jerk Energy (JE) as defined below is computed in the device itself and send processed data with location, time and to device cloud for further investigation.

Let us assume that a_1, a_2,…, a_n be the consecutive discrete acceleration samples in vertical direction (parallel to earth's gravity) at time t_1, t_2,..,t_n where $\Delta t = t_n - t_{n-1}$ for uniform sampling rate. Then jerk, i.e. rate of change of acceleration (m/sec3) is defined as [7], [8]:

$$J_i = {(a_{i+1} - a_i)}\big/{\Delta t} \quad \forall 1 \le i \le n-1 \tag{1}$$

Thus, Jerk Energy (JE) is $JE_s = J_{s1}^2 + J_{s2}^2 + \ldots + J_{s19}^2$

Our code collects the data for running bus then sends them to remote server for further processing. In the file we have 19 Jerk Energy values with one location (GPS) and average velocity and time stamp. Below is sample of 1 line of such data

368.92,400.53,...,1255.27; lat : 12.9514386667; lon : 80.1150148333; speed : 28.684; Time :2013-06-03T9:13:5

Each trip will have a file with 180 such lines of data. High JE values are harmful for a vehicle. Thus we have designed scoring mechanism to be sensitive to high value of JE.

4. ANALYSIS AND SCORING MECHANISM

We analyze captured processed data by using statistical methods and then develop scoring mechanism to assign score to a trip which is indicative of quality of the journey. Since Jerk Energy (JE) is indicative of bumpiness of the ride and gives indication of how the vehicle is driven it is used as the parameter to be observed and analyzed for deriving driving behavior. Dependence of JE on velocity is analyzed using MINE tool as described in below section.

4.1 Statistical analysis of dependency

Now using the above mentioned setup in section 3, we have collected data for trips of the bus for many months. Then to identify velocity dependence of jerk energy, we have used MINE tool. MINE (maximal information-based nonparametric exploration) statistics is used not only to identify relationships in data sets and characterize them. MINE tool computes MIC (maximal information coefficient), which is a measure of two-variable dependence. MIC is used to investigate relationship between different observed variables (Jerk Energy, Average Speed). MIC always lies between 0 and 1. High MIC values indicate that the pair of variables are related. It is proved that for two dependent variables. MIC value will converge to 1 with increasing sampling size. Thus higher MIC indicated dependency between two variables. We give sample data of MIC in table 1.

Table 1. Calculated MIC for average speed (X) and JE (Y) for 9th April, 2013

X	Y	MIC
Speed	Median of JE	0.79

It is intuitive that there may be dependency between the jerk energies and average speeds. We calculated statistics like mean, median and maximum on the set of 19 jerk energies to get one representative value of the jerk energies to make a pair with the corresponding average speed. It was found that maximum information was coming out while the median JE and average speed were being considered as a pair of variables and it was lowest while maximum JE and average speed were being considered as a pair. For each day we calculated a regression equation. This is done on all data for the months of April, May, June. From that we calculated the average regression equation for data points available. Below we give a plot for Velocity vs Jerk Energy data for a typical trip and also the fitted regression line.

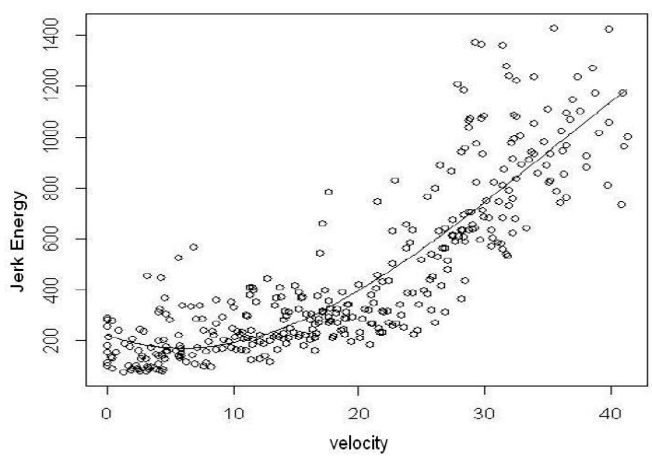

Figure 2: Jerk energy plot and fitted curve

The equation of the regression line is

$$v = 238.53 - 4.42 * u + 0.56 * u^2 \quad \& \hat{\sigma} = 151.1 \tag{2}$$

Thus we have established a relationship of Jerk energy and average speed. This is used further in the following subsection to give a score to a journey.

4.2 Scoring algorithm for vehicle journey

Jerk Energy damages the vehicle also serves as an indicator of quality of the journey in terms of bumpiness and comfort, impact on car suspension system. Now for a trip data each line of data file as described above makes a mini segment. For each mini segment we have 19 Jerk Energy (JE) value. Median of these JE values is chosen as representative JE of the mini segment denoted by JR_R. This value is real data for mini segment. Using the above equation we compute ideal predicted value for the velocity (JE_P). As we know the standard deviation is 151.1 we calculate score for a mini segment as score = (JE _R – JE_P)/151.1.

Now for each file we have 180 such scores. Then we calculate overall score for the journey by using the below mentioned algorithm.

1. Compute the no. of values ≥ 0.5 and < 1 this is stored as *S(0.5)*
2. Compute the no of values ≥ 1 this is stored as *S(1)*
3. Then we calculate score by using the following formulae

Fraction of time the JE value is in the range [.5, 1) denoted by

$$f(0.5) = S(0.5)/180$$

Fraction of time the JE value is greater than 1 denoted by

$$f(1) = S(1)/180$$

Journey Quality Score (JQS) = $\sqrt{f(0.5)^2 + f(1)^2 / 2}$

JQS serves as the measure of bad driving for a trip. Theoretically JQS falls in the range [0, 1]. But mostly it is good if value is less than 0.2. It is normally expected that for a good journey JE values will fall mostly in range. The algorithm classifies values with limit $\mu + 0.5 * \sigma$ and $\mu + \sigma$ as decision boundaries.

Below we give a sample analysis done on 3rd June data. As each file have 180 lines of data we have 180 mini segments for all the trips we give the plot of scores.

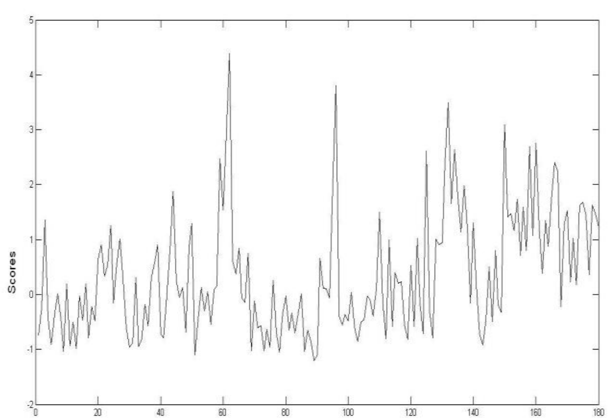

Figure 3: June 3 scores for mini segment

For this journey *S(.5) =20* , *S(1) = 48* and JQS = 0.204. As clearly visible from the graph that mostly real values of JE are higher than predicted value and also lot of values fall outside the range ($\mu + \sigma$). This journey receives a high score which means that around 20% of the journey is with high Jerk Energy. Below we present scores for 2 other days.

Table 2. Table showing scores and JQS for different days

Date	S(0.5)	S(1)	JQS
10 June	16	41	0.173
2 April	8	14	0.0633

From the data it is clear that 2nd April journey was better so it received low score. Below we give comparison plot for 2nd April and 3rd June journey.

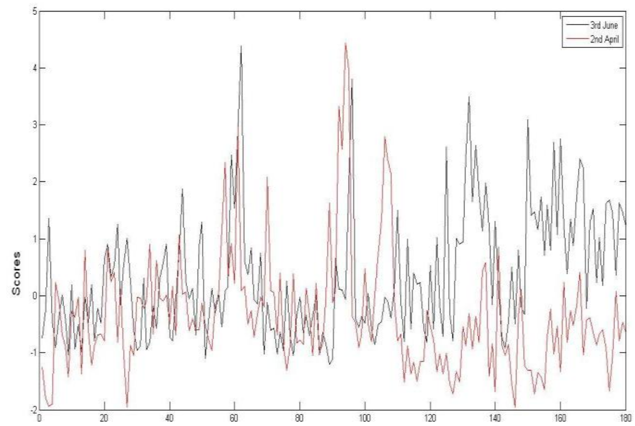

Figure 4: June 3 and April 2 scores plot for comparison

Clearly it is visible from figure 4 that April 2 journey have low scores and low JQS. Hence using easy computation we can identify bad driving using proposed architecture. Previous methods for finding driving patterns presented in [3, 4, 5, 6, and 8] assign a single score for a trip. Approach presented here differs in the aspect that scores are given to mini segments. Then a JQS is obtained in a two level process. Hence using this method one can get insights into small section of trips as well as overall trip. JQS serves as a real time monitoring of journey as well as for a collection of vehicle it serves as a measure to identify bad cars or bad driver.

5. SUMMARY

The statistical analysis and algorithm to derive score are able to capture the relationship of Jerk energy and average speed. Using that author are able to successfully establish a scoring mechanism for monitoring a vehicle and using proposed architecture large scale data collection and maintenance of large no of vehicle will be made possible. This algorithm can serve as analytics as a service as well as for insurance premium calculation for pay as you drive model.

6. REFERENCES

[1] David N. Reshef, Yakir A. Reshef, Hilary K. Finucane, Sharon R. Grossman, Gilean McVean, Peter J. Turnbaugh, Eric S. Lander, Michael Mitzenmacher, Pardis C. Sabeti, Detecting Novel Associations in Large Data Sets, Science 16 December 2011: vol. 334 no. 6062 pp. 1518-1524, DOI: 10.1126/science.1205438

[2] A. Ghose, P. Biswas, C. Bhaumik, M. Sharma, A. Pal, and A. Jha. Road condition monitoring and alert application: Using

in-vehicle smartphone as internet-connected sensor. In Pervasive Computing and Communications Workshops.

[3] T. Toledo, O. Musicant, and T. Lotan. In-vehicle data recorders for monitoring and feedback on drivers behavior. Transportation Research Part C: Emerging Technologies, 16(3):320–331, 2008.

[4] D. A. Johnson and Mohan M. Trivedi, Driving Style Recognition Using a Smartphone as a Sensor Platform, 2011 14th International IEEE Conference on Intelligent Transportation Systems Washington, DC, USA. October 5-7, 2011

[5] P. Handel, Discounted Least-Squares Gearshift Detection Using Accelerometer Data, IEEE Transactions on Instrumentation and Measurement, Volume 58, Issue 12, Dec. 2009.

[6] B. Y. Liaw. Fuzzy logic based driving pattern recognition for driv ing cycle analysis. Journal of Asian Electric Vehicles, 2(1):551–556, 2004.

[7] T Chakravarty , A Chowdhury, A Ghose and C Bhaumik, P. Balamuralidhar, Statistical analysis of road-vehicle-driver interaction as an enabler to designing behavioural models. International Conference on Mathematical Modeling in Physical Sciences, IC-MSQUARE 2013.

[8] Tapas Chakravarty, Avik Ghose, Chirabrata Bhaumik, Arijit Chowdhury, MobiDriveScore – A System for Mobile Sensor Based Driving Analysis, A Risk Assessment Model for Improving One’s Driving. 7th International Conference on Sensing Technology (ICST 2013), December 3 to 5, 2013, Wellington, New Zealand.