**Car Project**

**Rahman Saeed Khorsandi**

**Monday, July 17, 2016**

**Motivation**

Race car data usually is acquired for these purposes:

1. To analysis racecar function
2. To analysis driver performance
3. To use for racecar development and improvement

In racing, one of the most important questions is that how fast are you and speed data is one the best channels to answer this question. However, it is not rescannable to use only one channel. There are other helpful channels such as braking, acceleration, driver behavior and performance and so on. In fact, there are three main categories for data analysis of racecars, powertrain analysis (RPM, oil pressure, water temperature), chassis analysis (lateral and longitudinal G-forces, wheel speed) and driver (steering angle, throttle, brake pressure).

For an example, to analysis speed, the comparison between different laps can be useful which helps to see how a setup or component change affects the performance on a particular sector of the lap. In addition, driver performance on different laps, or performance between different drivers can be analyzed. Figure 1 shows the variation of speed in different laps for data set. There is a large variation and low mean in lap 22 which is because of outgoing time. However, the speeds in laps 23, 24 and 25 are similar to each other. There is big decrease of speed in the last lap. Figure 2 shows the map and laps of the data set. It seems from this figure that the GPS data is noisy or maybe the GPS was not calibrated and if we can use google map information, we can clean the data. A few more plot are shown in the appendix. All in all, speed is what really matters when it comes to racing, and throttle and RPM are directly connected to the development of speed and using of steering angle together with throttle angle can give useful information.

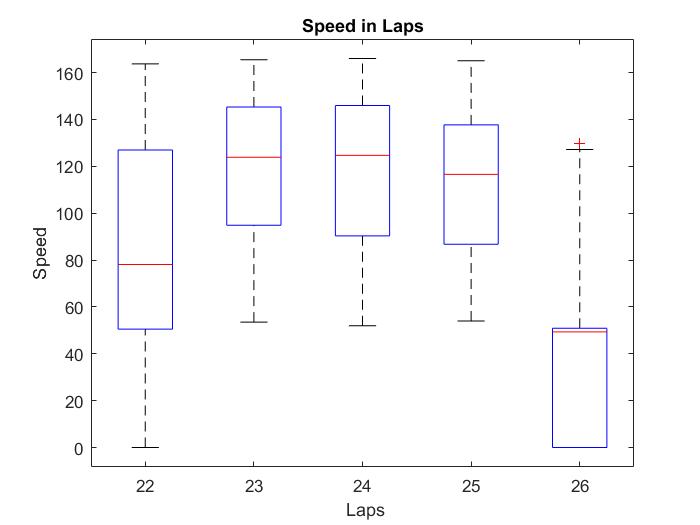


Figure 1: Speed variation in different laps



Figure 2: Map of the lap

**Steps:**

1. Studying of racecar variables (channels) and searching for information about them and how they are related to each other.
2. Import data set into MATLAB, clean and normalize it.
3. Basic analysis of data in Matlab such as normailization, mean, plot and so on. However, processing in MATLAB is really slow and it is not the perfect software for big data.
4. Using Cloudera. FTP the data to CHD (Cloudera distributed Hadoop) local file system using filezilla. I did it using VM VirtualBox in windows. Importing data from local file system to HDFS. Using spark-sell in CDH. Creating RDD (resilient distributed dataset) based on input data.

**Main idea:**

The main idea in this project is to find the correlation between bio data and car telemetry data in race cars. In this data set, ECG and EMG signals of the driver are available and it will be helpful to know how bio data and telemetry data of the car affect each other and maybe we can use these information to improve the car or train drivers.

**Appendix**

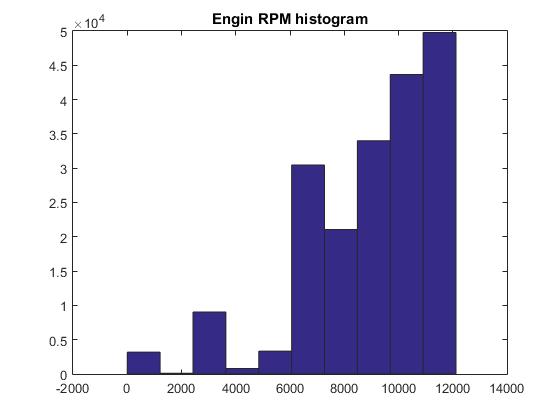


Figure 3: Histogram of Engin RPM

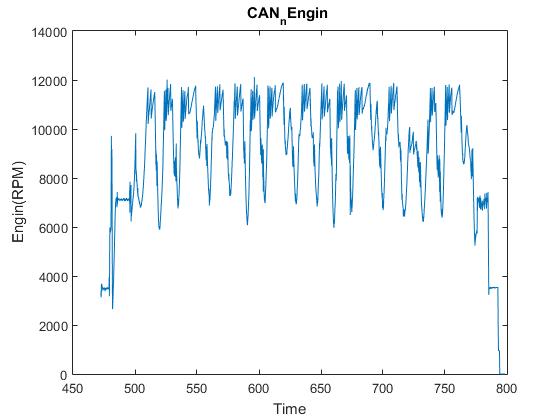


Figure 4: Engin RPM

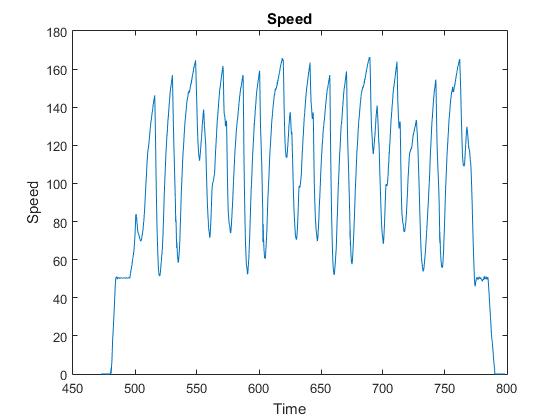


Figure 5: Speed channel

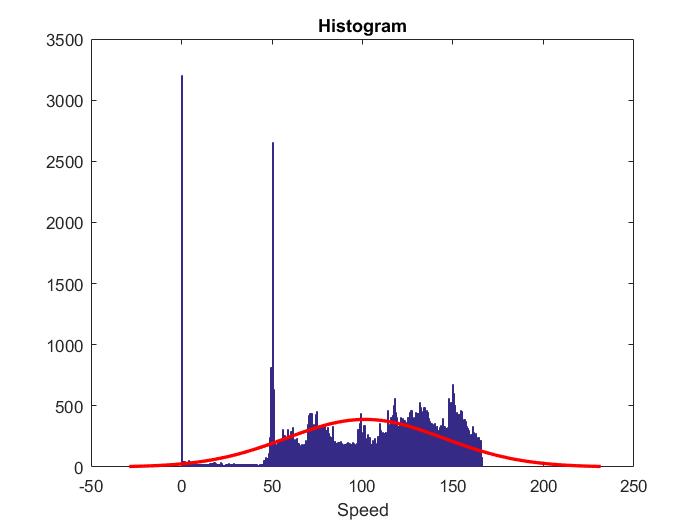


Figure 6: Histogram of Speed. A Gaussian (in red) is mapped on the speed to show the distribution of speed

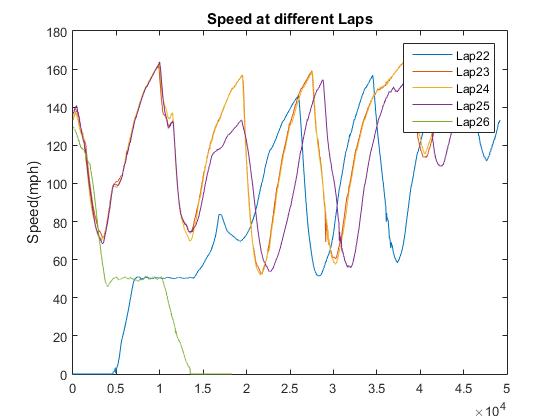


Figure 7: Speed in different Laps. There are 5 laps in data set 1. In Lap 22, speed starts from 0 which is because of outing time. The last lap, lap 26, ends in 0 which means the car stops in this lap.

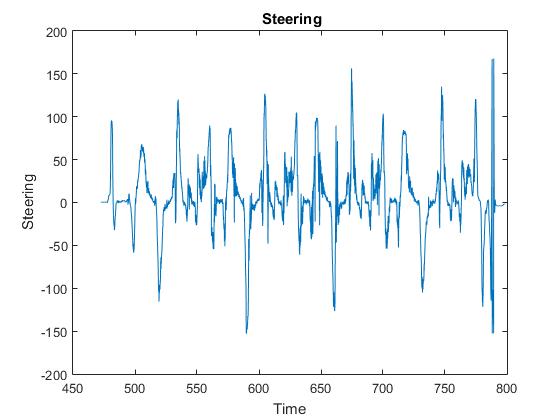


Figure 8: Steering angle. It seems to be noisy and there are outliers which should be removed.