Exploratory Analysis on Los Angeles Road Data

USC Machine Learning Team

9/10/2017

Section 1: Synopsis

The objective of this file is to perform exploratory analysis on accelerometer and GPS data collected on October 15th, 2017 in Kenya. More specifically, this file aims to explore the relationship between traveling speed and vertical accelration when the vehicle hits speed bumps.

Section 2: Data Loading

Before data processign, we need to load the raw dataset from **kenya__oct__15__data__labeled.csv**. We can count the number of rows in the raw dataset, and take a look at the first 10 rows.

```
if (!exists("Kenya.raw")) {
    Kenya.raw <- read.csv("./kenya_oct_15_data_labeled.csv")
}
print(nrow(Kenya.raw))
```

```
## [1] 15238
```

```
head(Kenya.raw, 10)
```

```
##    X Batch_id                    time Longitude  Latitude      x      y
## 1  0       73 2017-10-15 13:43:02.008  36.76371 -1.297734  0.000  0.002
## 2  1       73 2017-10-15 13:43:02.019  36.76371 -1.297734  0.000  0.000
## 3  2       73 2017-10-15 13:43:02.120  36.76371 -1.297734  0.000 -0.001
## 4  3       73 2017-10-15 13:43:02.224  36.76371 -1.297734  0.001 -0.002
## 5  4       73 2017-10-15 13:43:02.327  36.76371 -1.297734 -0.001  0.002
## 6  5       73 2017-10-15 13:43:02.431  36.76371 -1.297734  0.000  0.001
## 7  6       73 2017-10-15 13:43:02.535  36.76371 -1.297734  0.000 -0.002
## 8  7       73 2017-10-15 13:43:02.639  36.76371 -1.297734  0.000  0.000
## 9  8       73 2017-10-15 13:43:02.742  36.76371 -1.297734  0.001  0.001
## 10 9       73 2017-10-15 13:43:02.846  36.76371 -1.297734  0.000 -0.001
##         z velocity x_raw         y_raw       z_raw label
## 1  -0.006        0 0.000  0.2067574821 -0.62027245     0
## 2  -0.001        0 0.000  0.2067574821 -0.63127245     0
## 3   0.006        0 0.000  0.1057574821 -0.02527245     0
## 4  -0.005        0 0.104 -0.1022425179 -0.54527245     0
## 5   0.001        0 0.001  0.1037574821 -0.44227245     0
## 6   0.003        0 0.001  0.2077574821 -0.13027245     0
## 7  -0.004        0 0.001 -0.0002425179 -0.54627245     0
## 8   0.005        0 0.001 -0.0002425179 -0.02627245     0
## 9  -0.001        0 0.104  0.1027574821 -0.12927245     0
## 10 -0.003        0 0.104 -0.0012425179 -0.44127245     0
```

Section 3: Data Processing

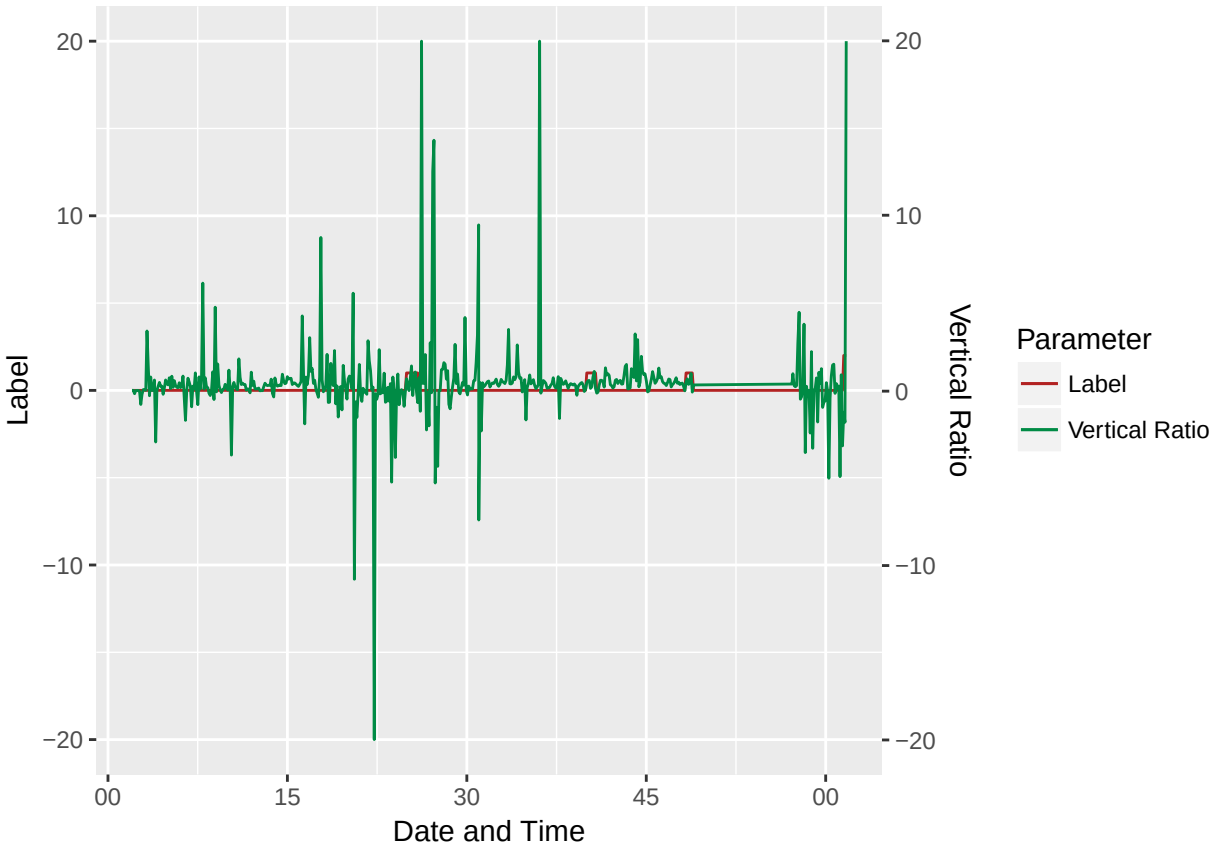
```
for (i in 1:nrow(Kenya.raw)) {
    if (Kenya.raw$velocity[i]== 0) {
        Kenya.raw$x_ratio_speed[i] = 0
    }
    else {
        Kenya.raw$x_ratio_speed[i] = Kenya.raw$x_raw[i] / Kenya.raw$velocity[i]
    }
    if (Kenya.raw$z_raw[i]== 0) {
        Kenya.raw$x_ratio_z[i] = 0
    }
    else {
        Kenya.raw$x_ratio_z[i] = Kenya.raw$x_raw[i] / Kenya.raw$z_raw[i]
        if (Kenya.raw$x_ratio_z[i] > 20) {
            Kenya.raw$x_ratio_z[i] = 20
        }
        if (Kenya.raw$x_ratio_z[i] < -20) {
            Kenya.raw$x_ratio_z[i] = -20
        }
    }
}
write.csv(Kenya.raw, file = "./kenya_oct_15_data_processed.csv")
Kenya.raw$time <- as.POSIXct(Kenya.raw$time, format="%Y-%m-%d %H:%M:%OS")
Kenya.73 = Kenya.raw[Kenya.raw$Batch_id == 73, ]
Kenya.75 = Kenya.raw[Kenya.raw$Batch_id == 75, ]
```

Section 4: Exploratory Data Analysis

```
# data visualization
require(ggplot2)
```

```
## Loading required package: ggplot2
```

```
Kenya.73.plot <- ggplot(Kenya.73[1: 500, ], aes(x = time))
Kenya.73.plot <- Kenya.73.plot + geom_line(aes(y = label, colour = "Label"))
Kenya.73.plot <- Kenya.73.plot + geom_line(aes(y = x_ratio_z, colour = "Vertical Ratio"))
#Kenya.73.plot <- Kenya.73.plot + geom_line(aes(y = x_raw, colour = "Vertical Acceleration"))
#Kenya.73.plot <- Kenya.73.plot + geom_line(aes(y = velocity, colour = "Traveling Speed"))
Kenya.73.plot <- Kenya.73.plot + scale_y_continuous(sec.axis = sec_axis(~.*1, name = "Vertical Ratio"))
Kenya.73.plot <- Kenya.73.plot + scale_colour_manual(values = c("firebrick", "springgreen4"))
Kenya.73.plot <- Kenya.73.plot + labs(y = "Label", x = "Date and Time", colour = "Parameter")
print(Kenya.73.plot)
```



We can see from the above graph **Los Angeles Road Data Session #5: Time-Series Display of Z Acceleration** that there are three shaded fractions which strikes out. The overall similar pattern in these three shaded areas is the anormaly in the reading of Z acceleration. According to the test driver Ernest, in the shared areas **Speedbump #1 & #2** and **Speedbump #3 & #4**, the accelerometer read an anormaly followed by a much larger anormaly, because Ernest was stepping on the gas. Cross-referencing with the Speed graph. We find that speed might be a factor of the scale of change to Z acceleration when the vehicle hits a speedbump.

Section 5: Recommendations

Based on the findings from the Section 4, We think it will be a good start to model a logistic regression on Z acceleration to identify speedbumps. If this method shows promise, we can apply it on potholes as well.