Analyzing road damage with accelerometer data

Tejas Kotha IMT2016112
Shreyas Gupta IMT2016122
Ayush Saraswat IMT2016123
Tanishq Gupta IMT2016126

Introduction

Severity of damage in roads is really important for taking right measures to fix it. A road can be deemed bad for several reasons

- Potholes
- Uneven coverage of old potholes with cement or road tar mixture
- Muddy and rocky settlement on the road

problem statement

We will try to get some data from the roads and see if we can infer the condition of the roads from the data collected. It is relatively easy to get proper data about the severity of the roads.

Data collection

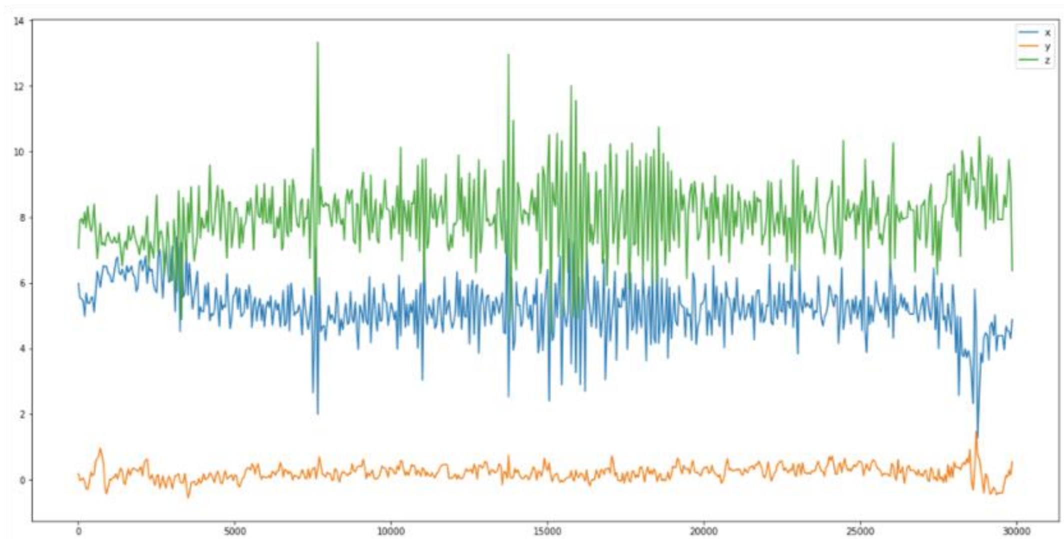
We will be using accelerometer data from the smartphones to get signals which will hopefully give us a better picture about the condition of the roads.

The reason why accelerometer seems like the ideal sensor to measure severity of road is because the accelerometer can detect sudden jerks and changes on the road because the sudden change would bring in some acceleration in that direction. The data collected from the accelerometer is inaccurate because the sensitivity of the accelerometer in most smartphones is very high. There are a lot of fluctuations with the slightest movement of the smartphone.

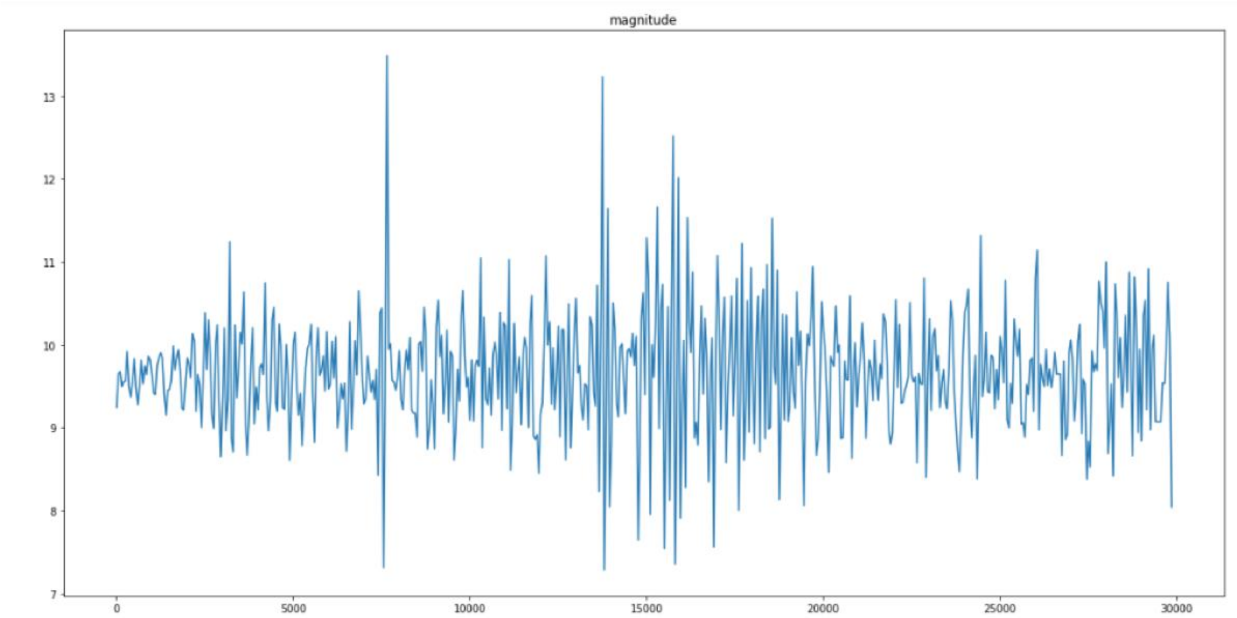
The data was collected while riding a scooty using SensorLab app in and around Electronic City. The dataset contains signal data from 14 different roads like ISBR road, Wipro village road, BSNL road etc. The ranks for the roads were calculated by analyzing images and taking the average of all team members’ verdict about the road. The ranking is on a scale of 1-10 with 10 being a really good road.

Approach

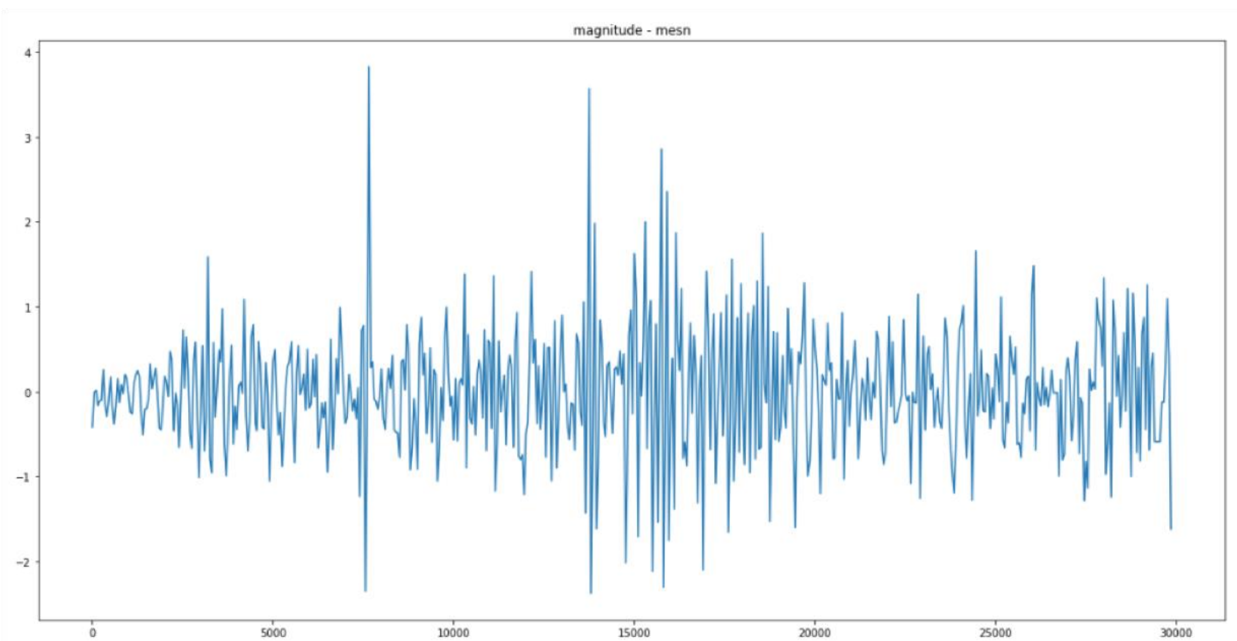
- The data from the accelerometer comes from three different axes - X, Y and Z.



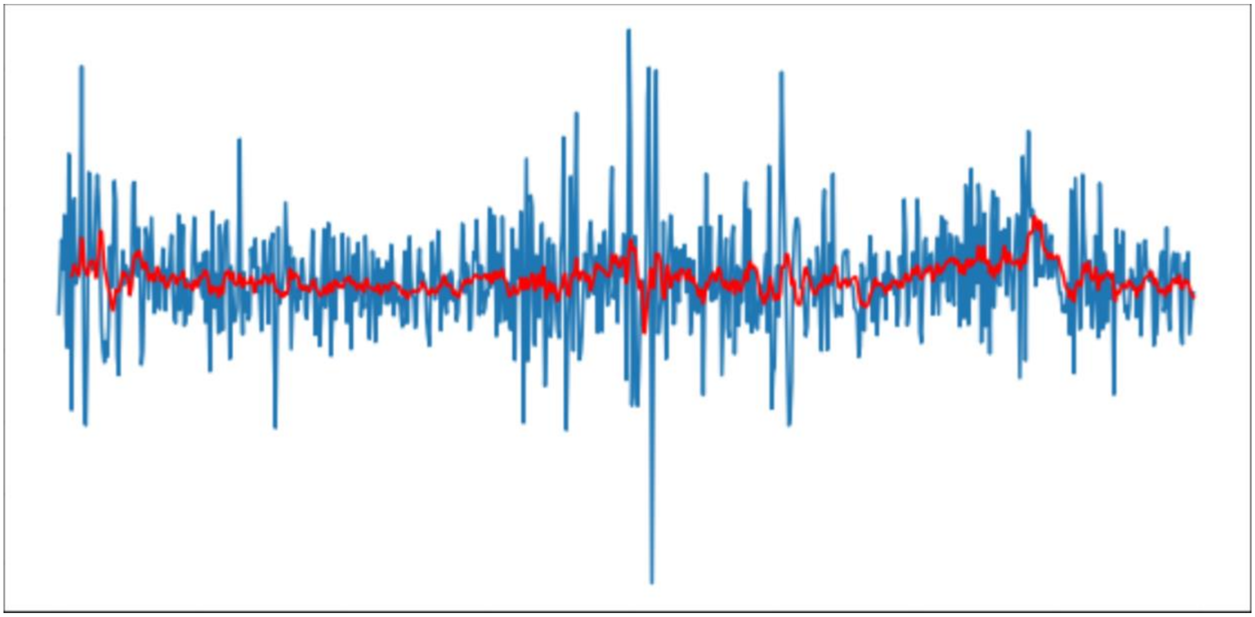
- We can identify the peaks and changes in the data by identifying the axis manually but we can take the magnitude of all the axis for a better insight.



- We trim out data length to a standard length for better comparison.
- We then remove the gravitational acceleration from our data by removing the average. This should work as ‘g’ is constant. We can see the plot move down by 10. This is because we are in fact removing g.

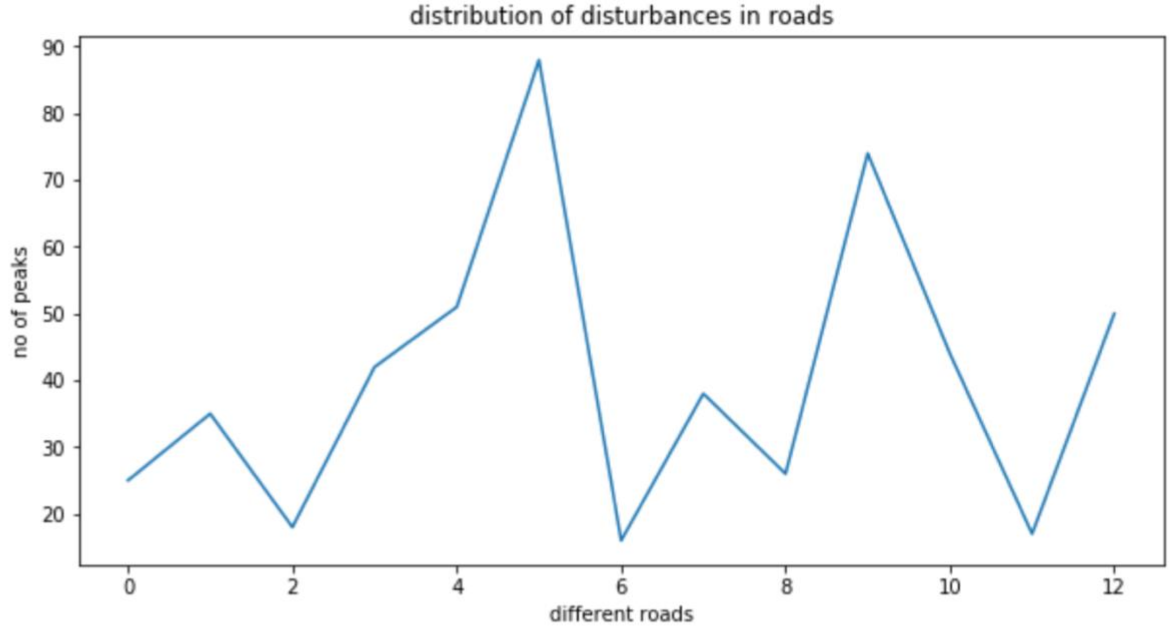


- We then apply a moving average filter to smoothen out the signal. This helps us identify the peaks easily as it will remove erroneous fluctuations. The red signal in the plot below is the smoothened signal.

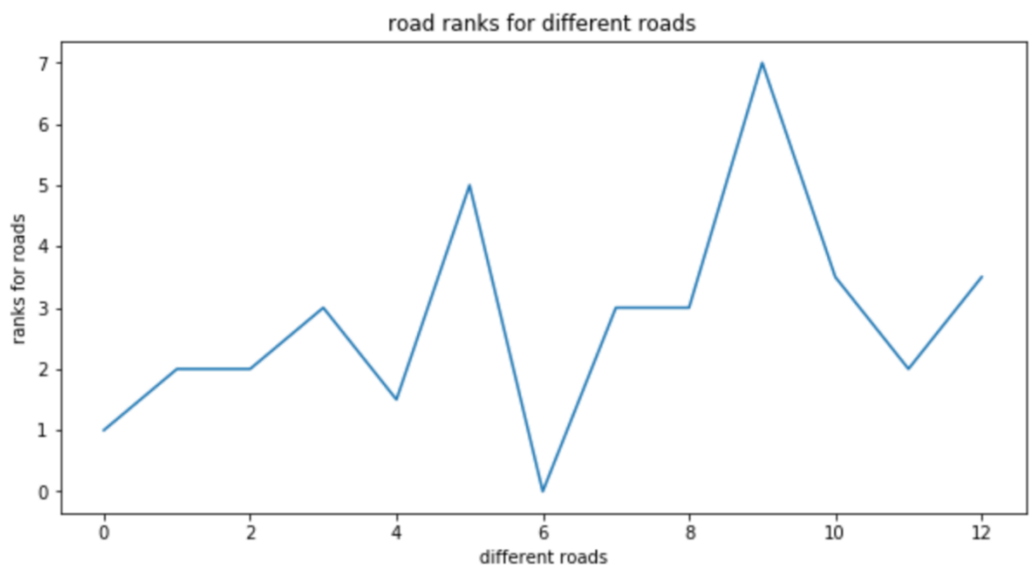


- After calculating the moving average values, we will observe some peaks in the signal data.
- We will then set a threshold level which is our reference point for peak detection. Any data point above this threshold value would be considered a pothole/broken road etc, meaning bad road conditions. This threshold value is max/2 and min/2 . Any value more than max/2 and less than min/2 are labeled as perfect peaks which will help us rank the roads based on the damage.

- Depending on the number of peak values which are above our threshold, we'll rank our roads.



We see that the peak distribution plot is very similar to the inverse of our road ranks (because the better the rank, the better the road)



So what we infer is that accelerometer data can properly capture the disturbances in the roads.

The threshold level and peak-count value to ranking can be adjusted to match the labels we assigned to our raw data but adjusting these values manually isn’t a very intelligent way of analyzing signals, so we’ll let a ML model like linear regression do it.

Training

We split our data into train and test data. The data contains name of the road and the corresponding rank. We use a basic regression model like LinearRegression to predict road ranks with the help of peak counts. On training, we get a logarithmic loss of around **0.56**, which is actually pretty decent given the mode of data collection.