BMS data science challenge

Fatima Kahil

Motivation

- Detection of road anomalies using accelerometer data
- Labelling of acceleromter data for road classification purposes

Outline

- Plot data against time

- Rely on Z-axis accelerometer data readings

- Smooth data (window size: compromise between smoothing noise and not missing any valuable information: for now 5 seconds)

- Detect anomalies (focus on pothole vs speedbump)

- Assign labels to data points

Data analysis

- Smoothing the data: averaging over 50 data points
- 50 data points: every 5 seconds
- Focusing on Z-axis acceleration
- Speed bump vs. pothole:

  - speed bump: sudden increase followed by sudden decrease

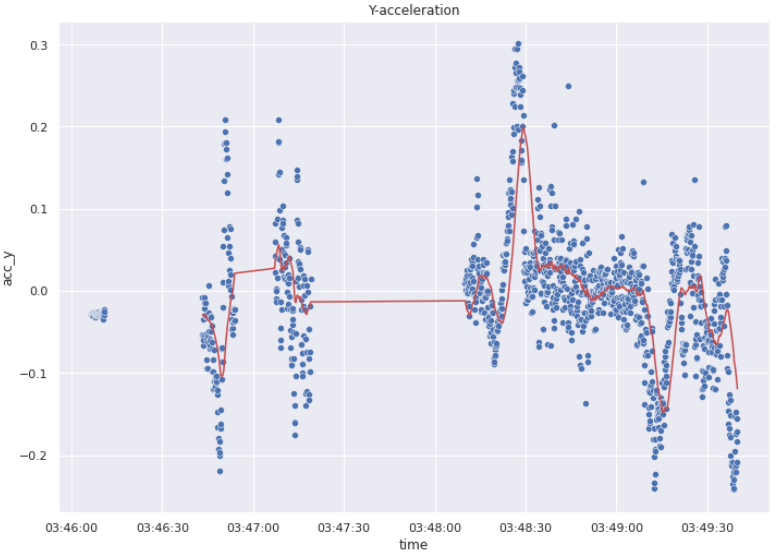
  - pothole: sudden decrease followed by sudden increase

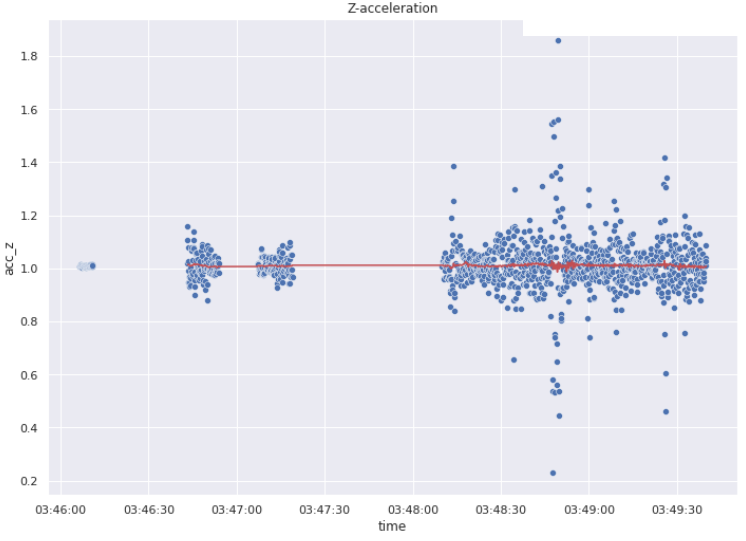
Results

A- Smoothing timseries: rolling window over 50 datapoints (every 5 seconds). Adding new columns to the data table with the mean of the numerical entries

road_type	acc_x_mean	acc_y_mean	acc_z_mean	speed_mean	latitude_mean	longitude_mean	heading_mean	road_speed_limit_mean
AL_ROAD_OF_MAJOR_IMPORTANCE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
AL_ROAD_OF_MAJOR_IMPORTANCE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
AL_ROAD_OF_MAJOR_IMPORTANCE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
AL_ROAD_OF_MAJOR_IMPORTANCE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
AL_ROAD_OF_MAJOR_IMPORTANCE	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
MAJOR_ROAD	-0.1100	-0.2009	1.0042	19.82	43.238388	-2.872232	170.088132	50.0
MAJOR_ROAD	-0.1074	-0.1940	1.0055	19.64	43.238383	-2.872231	171.071827	50.0
MAJOR_ROAD	-0.1063	-0.1919	1.0040	19.46	43.238379	-2.872230	172.027744	50.0
MAJOR_ROAD	-0.1051	-0.1882	1.0056	19.28	43.238374	-2.872230	172.979137	50.0
MAJOR_ROAD	-0.0990	-0.1821	1.0190	19.10	43.238370	-2.872229	173.917036	50.0


X-acceleration
acc_x
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
03:46:00
03:46:30
03:47:00
03:47:30
03:48:00
03:48:30
time


Y-acceleration
acc_y
0.3
0.2
0.1
0.0
−0.1
−0.2
03:46:00
03:46:30
03:47:00
03:47:30
03:48:00
03:48:30
03:49:00
03:49:30
time


Z-acceleration
acc_z
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
03:46:00
03:46:30
03:47:00
03:47:30
03:48:00
03:48:30
03:49:00
03:49:30
time

B- Z-axis accelerometer data:

- Pothole: correspond to Minimas

- Speedbumps: correspond to Maximas

- What is needed here is a threshold of the acceleration below which values are excluded

C- Labelling datapoints:

- Assigning data points around the minima or the maxima (plus/minus 25 data points) to the same label.

Problems:

- some of the Minimas and maximas are separated by only 10 seconds and labels are Overlapping: changing the rolling window size?

- Reason for the short separation between potholes and speedbumps : they are cracks? Another metric should be assigned to other types of anomalies depending on their Geometrical separation?