Ahmedabad University

CSE523 Machine Learning

Weekly Report - 3

Project 5: Identify Hard stop and momentary stop using vehicle trajectory dataset.

Submitted to faculty: Mehul Raval

Date of Submission: 23-03-24

Roll No.	**Name of the Student**
AU2140040	Kathan Thakkar
AU2140171	Harsh Pandya
AU2140224	Dhruvi Rajput
AU2140230	Yax Prajapati

AIM OF THE WEEK:

Data set cleaning, adding features or columns as labels and acceleration, algorithm application, trying to find optimal values of the parameters of the DBSCAN and plotting the clusters on the map.

Introduction:

In this week's report, we present significant advancements in refining our data analysis pipeline, focusing on data cleaning, feature engineering, algorithm application, parameter optimization, and visualization techniques. Our primary objective was to improve the accuracy and effectiveness of our analysis by incorporating robust methodologies.

Data cleaning:

We began by meticulously cleaning our dataset, addressing missing values, outliers, and inconsistencies to ensure data accuracy and reliability. Additionally, we enriched the dataset by adding new features or columns, which serve as labels and capture essential characteristics for subsequent analysis.

Approach:

1. Derived acceleration and provided labels to each row using data processing and presumptions.
2. Found the optimal values of epsilon and min_data_points using grid search.
3. Made clustering for X,Y coordinates for labels.
4. Plotted the data using matplot library on the frame.
5. Checked for vehicles whose label is not equal to the above classified label and lies in the cluster coordinate ranges.
6. Found average velocity of the vehicles.
7. Checked if velocity lies in the presumption range.
8. From that, clusters with avg velocity lying as per the presumptions is declared as hard stop or soft stop.

Assumption Condition:

```
if row['Vxy_smoothed'] < 1 and -1 <= row['acceleration'] <= 1:
    return 'possible_hard'
elif 0 <= row['Vxy_smoothed'] <= 5 and row['acceleration'] < 0:
return 'possible_soft'
else:
return 'other'
```

Code Implementations:

1. Cluster Formation:

Using clustering algorithms such as DBSCAN, we segmented the dataset into clusters, designating certain clusters as potential hard or soft stops based on our assumptions regarding vehicle behavior.

2. Identifying Data Points within Stop Range:

For each cluster identified as a stop (either hard or soft), we determined the spatial range within which vehicles are considered to be stationary.This range was defined based on assumptions regarding vehicle velocity and acceleration during stops.

3. Calculating Average Velocity:

After delineating the stop ranges for each cluster, we computed the average velocity of vehicles within those ranges.This involved aggregating the velocity values of all vehicles within the stop range and computing their mean velocity.

4. Stop Classification:

Finally, we classified each stop based on predefined conditions related to smoothed velocity and acceleration values. If the conditions are met, the stop is classified as either a potential hard stop or a potential soft stop. Otherwise, it's labeled as 'other'.

Link of Code:
https://colab.research.google.com/drive/19KzRhxubpV95pOpMlBDKfbMIafIhPplq?usp=sharing

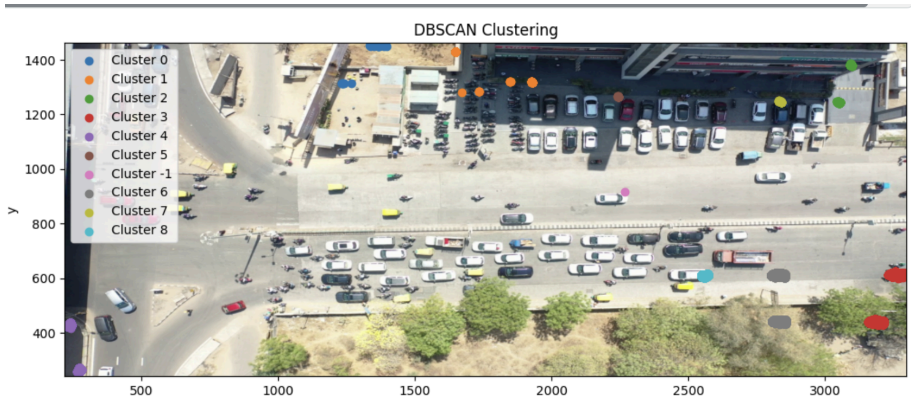
Future Directions:

Looking ahead, we plan to delve deeper into the analysis of clustered data and explore additional techniques for feature engineering and dimensionality reduction. Additionally, we aim to validate our findings through rigorous evaluation and comparison with ground truth data where applicable. By continuously refining our data analysis pipeline, we strive to unlock deeper insights and drive informed decision-making.

Inferences:

1. For hard stop

```
Cluster 0:
x range: 1233.0 - 1397.0
y range: 1313.0 - 1452.0

Cluster 1:
x range: 1648.0 - 1929.0
y range: 1282.0 - 1430.0

Cluster 2:
x range: 3044.0 - 3100.0
y range: 1241.0 - 1386.0

Cluster 3:
x range: 3160.0 - 3289.0
y range: 429.0 - 624.0

Cluster 4:
x range: 230.0 - 280.0
y range: 252.0 - 438.0

Cluster 5:
x range: 2243.0 - 2245.0
y range: 1264.0 - 1269.0

Cluster -1:
x range: 2267.0 - 2267.0
y range: 916.0 - 916.0

Cluster 6:
x range: 2804.0 - 2858.0
y range: 432.0 - 618.0

Cluster 7:
x range: 2832.0 - 2840.0
y range: 1245.0 - 1252.0

Cluster 8:
x range: 2548.0 - 2572.0
y range: 603.0 - 614.0
```



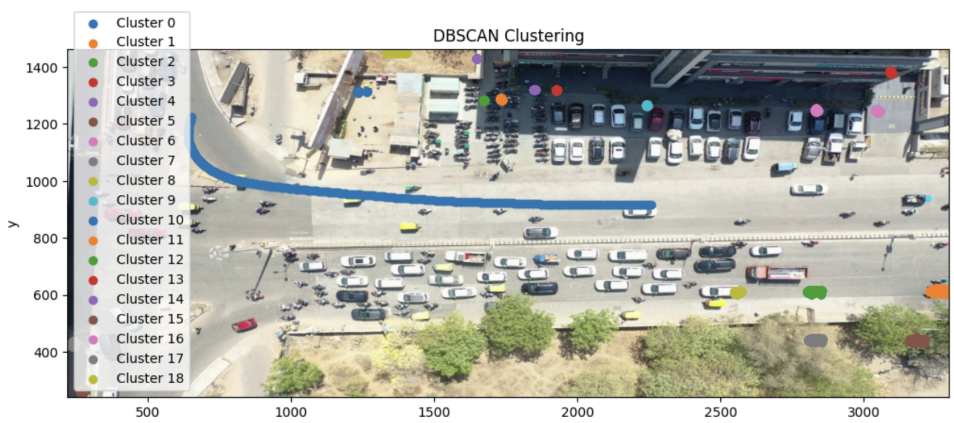
Avg velocity around hard stop cluster

```
Cluster 0: Average Vxy_smoothed = 0.7368108539059155
Cluster 1: Average Vxy_smoothed = 0.3649891806774388
Cluster 2: Average Vxy_smoothed = 0.08518269428032262
Cluster 3: Average Vxy_smoothed = 0.03279780196441385
Cluster 4: Average Vxy_smoothed = 0.06275889197243689
Cluster 5: Average Vxy_smoothed = 0.07505486153646998
Cluster 6: Average Vxy_smoothed = 0.03454859301040853
Cluster 7: Average Vxy_smoothed = 0.062351974912740735
Cluster 8: Average Vxy_smoothed = 0.04041313357572521
```

2. For soft stop

```
Cluster 0:
x range: 1234.0 - 1266.0
y range: 1312.0 - 1314.0

Cluster 1:
x range: 1733.0 - 1737.0
y range: 1284.0 - 1289.0

Cluster 2:
x range: 1672.0 - 1675.0
y range: 1280.0 - 1285.0

Cluster 3:
x range: 3091.0 - 3100.0
y range: 1378.0 - 1386.0

Cluster 4:
x range: 1648.0 - 1650.0
y range: 1428.0 - 1430.0

Cluster 5:
x range: 3160.0 - 3213.0
y range: 431.0 - 449.0

Cluster 6:
x range: 3041.0 - 3056.0
y range: 1241.0 - 1250.0

Cluster 7:
x range: 231.0 - 247.0
y range: 419.0 - 437.0

Cluster 8:
x range: 1333.0 - 1404.0
y range: 1449.0 - 1452.0

Cluster 9:
x range: 2242.0 - 2245.0
y range: 1264.0 - 1269.0

Cluster 10:
x range: 641.0 - 2258.0
y range: 915.0 - 1224.0

Cluster 11:
x range: 3233.0 - 3287.0
y range: 603.0 - 623.0
```

```
Cluster 12:
x range: 2805.0 - 2853.0
y range: 603.0 - 616.0

Cluster 13:
x range: 1927.0 - 1930.0
y range: 1318.0 - 1321.0

Cluster 14:
x range: 1849.0 - 1853.0
y range: 1319.0 - 1321.0

Cluster 15:
x range: 266.0 - 280.0
y range: 254.0 - 271.0

Cluster 16:
x range: 2830.0 - 2840.0
y range: 1245.0 - 1253.0

Cluster 17:
x range: 2809.0 - 2856.0
y range: 433.0 - 448.0

Cluster 18:
x range: 2548.0 - 2571.0
y range: 603.0 - 614.0
```



Avg velocity around soft stop cluster

```
Cluster 0: Average Vxy_smoothed = 0.41386813945723344
Cluster 1: Average Vxy_smoothed = 0.09506160255420254
Cluster 2: Average Vxy_smoothed = 0.6178402563994254
Cluster 3: Average Vxy_smoothed = 0.1017385432276178
Cluster 10: Average Vxy_smoothed = 24.536614398035816
Cluster 4: Average Vxy_smoothed = 0.04151470651923802
Cluster 8: Average Vxy_smoothed = 0.4322661647374758
Cluster 5: Average Vxy_smoothed = 0.03140541347192343
Cluster 6: Average Vxy_smoothed = 0.07405383760700782
Cluster 7: Average Vxy_smoothed = 0.06655915718120968
Cluster 9: Average Vxy_smoothed = 0.08500445554631642
Cluster 11: Average Vxy_smoothed = 0.028827517628754105
Cluster 12: Average Vxy_smoothed = 0.027874163792535038
Cluster 13: Average Vxy_smoothed = 0.08141902707953202
Cluster 14: Average Vxy_smoothed = 0.5324245750647666
Cluster 15: Average Vxy_smoothed = 0.056037853299871314
Cluster 16: Average Vxy_smoothed = 0.06566715925217212
Cluster 17: Average Vxy_smoothed = 0.03456695576698474
Cluster 18: Average Vxy_smoothed = 0.03758282059751457
```