ReadMe: Vehicle Classification using GPS Data

Swadesh Vaibhav, Vipin Baswan

November 2019

1 Code Introduction

The code is divided into 4 sections:

1. **Mounting necessary files:** Run this- code to have the environment set up with all the necessary files.
2. **Importing necessary prerequisites:** Run this piece of code to import the necessary libraries and download the dependencies.
3. **Preprocessing:** Run this piece of code preprocess the data. Please refer to the code for block-by-block details. Also, please note that ceratin changes to the names of the files may be required as per the situation.
4. **Training:** Run this code for training the model on the preprocessed data. Again, please refer to the comments of the code for more details.s

2 Requirements

Before running the code, please install Anaconda. Open the code using Jupyter Notebook. Then follow the commands as written in the comments for each cell to execute the code cell-by-cell. The code has the following dependencies:

1. scikit-learn (*conda install -c anaconda scikit-learn*)
2. xgboost (*conda install -c conda-forge xgboost*)
3. matplotlib (*conda install -c conda-forge matplotlib*)
4. glob (*sudo pip install glob2*)
5. keras (*conda install -c anaconda scikit-learn*)
6. pyproj (*conda install -c conda-forge pyproj*)
7. pandas (*conda install -c anaconda pandas*)
8. numpy (*conda install -c anaconda numpy*)

3 Running the code

The code can be run by: Using Jupyter Notebook: Press *Shift + Enter* to run the cell.

4 Use Case - Running the model on a given database

4.1 Data collection and Preprocessing

The data has been provided by MapMyIndia, collected using the FastTrackerz service. The data contains multiple files, each of which contains the GPS coordinates of different vehicles IDs in a given time interval.

1. All the data corresponding to each vehicle was put into one file. Garbage values and useless information were removed from these files.
2. The coordinate values of the vehicles were used to calculate the average speed, average acceleration, variance of speed and average of acceleration for each vehicle.
3. This preprocessed data was pushed into one single dataframe.

4.2 Testing using pre-trained models

The eXtreme Gradient Boost model and Random Forest model were used to get the output of the given data. The results of the two models have been saved in two different files.