Ahmedabad University

CSE523 Machine Learning

Weekly Report

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

Submitted to faculty: Mehul Raval

Date of Submission: 29-03-24

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

1. Introduction:

In our quest to enhance the efficacy of stop detection within vehicle trajectory datasets, our group is considering alternative approaches beyond the traditional adaptation of the DBSCAN algorithm. While DBSCAN offers a solid foundation for clustering-based stop detection, exploring other algorithms such as k-Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), and Random Forest presents an opportunity to broaden our analytical toolkit. This report delves into the rationale behind this shift in approach and provides a comparative analysis of the performance of these alternative algorithms in identifying and classifying vehicle stops.

2. Motivation for Change:

While DBSCAN has shown promise in identifying clusters within spatial data, its suitability for precise stop detection, especially in the context of distinguishing between hard stops and momentary stops, has been questioned. Our group recognizes the need for a more nuanced and adaptable approach to stop detection, one that leverages the strengths of various machine learning algorithms tailored to the intricacies of vehicle trajectory data. By exploring alternative algorithms, we aim to identify the most suitable technique or combination of techniques for robust stop detection, with implications for improving road safety and traffic management strategies.

3. Proposed Approach

3.1. k-Nearest Neighbors (KNN):

KNN is a simple yet powerful algorithm that classifies data points based on the majority class of their nearest neighbors. In the context of vehicle stop detection, KNN can be trained on labeled trajectory data to predict the likelihood of a data point representing a stop based on the characteristics of its neighboring points. We will explore the impact of different distance metrics and neighborhood sizes on the performance of KNN in accurately identifying stops.

3.2. Gaussian Mixture Models (GMM):

GMM is a probabilistic model that represents the distribution of data points as a mixture of multiple Gaussian distributions. By fitting GMM to vehicle trajectory data, we can model the underlying distribution of stops and non-stops, enabling us to probabilistically classify data points as stops or non-stops. We will investigate the optimal number of components and covariance structures to achieve reliable stop detection using GMM.

3.3. Random Forest:

Random Forest is an ensemble learning method that constructs a multitude of decision trees and combines their predictions to improve accuracy and robustness. In the context of vehicle stop detection, Random Forest can be trained on features extracted from trajectory data to classify data points as stops or non-stops. We will explore feature selection techniques and ensemble configurations to optimize the performance of Random Forest for stop detection tasks.

Future work:

We will conduct a comparative analysis of the performance of DBSCAN, KNN, GMM, and Random Forest in terms of their accuracy, precision, recall, and computational efficiency for vehicle stop detection tasks. We will evaluate the algorithms using labeled ground truth data and a range of evaluation metrics to provide insights into their strengths and weaknesses in different stop detection scenarios.