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: 1-04-24

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

1. Introduction:

After careful consideration and extensive deliberation within our group, we have chosen to leverage the DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm for vehicle stop detection tasks. This decision was informed by a thorough evaluation of alternative approaches, including k-Nearest Neighbors (KNN), Gaussian Mixture Models (GMM), and Random Forest. In this report, we provide a rationale for our choice, highlighting the merits of DBSCAN compared to other algorithms, along with strong validation points to support its effectiveness in the context of vehicle stop detection.

2. Rationale for Choosing DBSCAN:

Merits of DBSCAN:

- Robustness to Noise: DBSCAN is inherently robust to noise and can effectively filter out outliers, making it well-suited for analyzing noisy trajectory data commonly encountered in real-world transportation datasets.
- Flexibility in Cluster Shape: Unlike traditional clustering algorithms that assume spherical clusters, DBSCAN can identify clusters of arbitrary shape, allowing it to capture the diverse spatial patterns of vehicle stops with greater accuracy.
- Parameter-Free Nature: DBSCAN does not require the specification of the number of clusters beforehand, eliminating the need for manual parameter tuning and making it more adaptable to different stop detection scenarios.
- Efficiency: DBSCAN's computational complexity is linear with respect to the size of the dataset, enabling efficient processing of large-scale trajectory data without compromising on performance.

Demerits of DBSCAN:

- Sensitivity to Parameters: While DBSCAN is parameter-free in terms of the number of clusters, it still requires tuning of the epsilon (ε) and minimum points (MinPts) parameters, which can impact the clustering results and may require domain-specific knowledge for optimal setting.

- Difficulty in Handling Varying Density: DBSCAN may struggle to accurately identify clusters in regions of varying density, particularly in areas with overlapping stops or rapidly changing traffic conditions.
- Limited Scalability to High-Dimensional Data: In scenarios where trajectory data is represented by high-dimensional feature vectors, DBSCAN's performance may degrade due to the curse of dimensionality, leading to increased computational overhead and memory usage.

3. Validation Points for DBSCAN:

To validate the effectiveness of DBSCAN for vehicle stop detection, we propose the following strong validation points:

- Ground Truth Comparison: Compare the stop clusters identified by DBSCAN against ground truth data obtained from manual annotation or reliable sensor measurements, quantifying the algorithm's accuracy, precision, and recall.
- Temporal Consistency: Analyze the temporal consistency of stop clusters detected by DBSCAN over different time intervals, assessing the algorithm's ability to capture persistent stop patterns while filtering out transient noise.
- Comparative Evaluation: Conduct a comparative evaluation of DBSCAN against alternative algorithms (e.g., KNN, GMM, Random Forest) using benchmark datasets and standardized evaluation metrics, highlighting DBSCAN's superiority in terms of accuracy, robustness, and computational efficiency.

4. Conclusion:

In conclusion, our group has chosen to adopt the DBSCAN algorithm for vehicle stop detection based on its robustness, flexibility, and efficiency compared to alternative approaches. By leveraging DBSCAN's strengths and addressing its limitations through careful parameter tuning and validation, we are confident in its ability to deliver accurate and reliable stop detection results.