Identify Hard stop and momentary stop using the vehicle trajectory dataset.

Kathan Thakkar, Yax Prajapati, Harsh Pandya, Dhruvi Rajput
Ahmedabad University, kathan.t@ahduni.edu, yax.p@ahduni.edu.in, harsh.p9@ahduni.edu.in, dhruvi.r@ahduni.edu.in

***Abstract*- This paper proposes a methodology for detecting parked vehicles (hard stops) and vehicles stopped at traffic signals (momentary stops) in vehicle trajectory data obtained from autonomous drones (AU drones) using DBSCAN (Density-Based Spatial Clustering of Applications with Noise). The dataset includes track information, x and y coordinates, frame numbers, and velocity, providing a comprehensive basis for analysis. By applying DBSCAN to cluster trajectory points based on spatial proximity and velocity, we identify clusters indicative of stops. Parked vehicles, characterized by prolonged periods of minimal movement, and vehicles stopped at traffic signals, representing brief halts in vehicular motion, are distinguished. Our approach offers a scalable solution for analyzing vehicular behavior in diverse environments, facilitating proactive interventions to enhance safety and optimize transportation systems. Through rigorous experimentation, we demonstrate the effectiveness of our methodology in accurately identifying parked vehicles and vehicles stopped at traffic signals, contributing to advancements in vehicular trajectory analysis and intelligent transportation systems.**

***keywords*- data cleaning, feature selection, DBSCAN algorithm, clustering, parameter optimization, epsilon, minimum sample size, training, testing, analysis, clusters, range, performance evaluation.**

Introduction

In the realm of data analysis, the journey from raw data to actionable insights involves several pivotal stages, each crucial for extracting meaningful information. This paper delves into the intricacies of this process, emphasizing the importance of feature selection, the application of the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, and the optimization of its key parameters. Our study encompasses meticulous data-cleaning procedures aimed at refining datasets for accuracy and consistency. Subsequently, we explore the significance of feature selection, particularly in isolating essential variables vital for robust analysis. With a focus on DBSCAN, we investigate its effectiveness in uncovering inherent structures within datasets, highlighting its suitability for tasks such as clustering. Integral to this exploration is the examination of parameter optimization, specifically the adjustment of epsilon (ε) and minimum sample size, which profoundly impact the algorithm's performance. Through systematic testing, training, and analysis, we discern the intricate interplay between these parameters and clustering outcomes, providing valuable insights for informed decision-making in data analysis endeavors.

Methodology

In our pursuit of extracting meaningful insights from vehicular data, we initiated our methodology with a meticulous process of feature extraction and dataset creation. Pertinent features including track ID, coordinates (x and y), and velocity were selected to tailor a new dataset optimized for subsequent analysis. To ensure data integrity and mitigate bias, redundant data points with identical values across track ID, coordinates, and velocity were systematically eliminated.

Subsequently, the prepared dataset was imported into our analysis environment utilizing the Pandas library in Python, enabling efficient data manipulation and exploration. This laid the groundwork for the application of advanced clustering techniques.

Our clustering endeavors were anchored by the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, recognized for its capacity to identify clusters of arbitrary shapes in spatial data. Through iterative testing, different combinations of epsilon and minimum sample size parameters were systematically explored, crucial for DBSCAN's efficacy in delineating clusters within the dataset. This iterative process allowed us to discern optimal parameter values conducive to meaningful clustering outcomes.

Parameter optimization constituted a pivotal stage of our methodology, aimed at distinguishing between hard and soft stops within the vehicular data. Leveraging insights from clustering results, we identified parameter values that facilitated effective discrimination between these stop events. These selected parameters formed the groundwork for subsequent analyses aimed at characterizing vehicular behaviors.

To enhance our understanding of soft stops, a detailed examination of vehicle behavior during these instances was conducted. Frame numbers from video footage corresponding to soft stops were extracted, and the velocity and acceleration patterns of the vehicles during these periods were analyzed. Utilizing Gaussian distribution, velocity, and acceleration ranges characteristic of soft stops were derived, leveraging the mean and standard deviation. Subsequently, the data were labeled based on these derived ranges, and our data was labeled accordingly, allowing the model to better identify soft stops within the dataset.

In addition to refining our approach to soft stops, a strategy to accurately distinguish between hard stops and soft stops was implemented. Recognizing the potential for false negatives in identifying hard stops, all tracks were iterated through, and the duration of vehicle halts was assessed. If a vehicle remained stationary for a duration longer than observed during soft stops, it was labeled as a hard stop, thus minimizing the risk of misclassification.

The efficacy and robustness of our methodology were subjected to rigorous evaluation to ascertain its reliability and validity. Employing both quantitative metrics and qualitative assessments, we validated the effectiveness of our clustering approach and the identified thresholds for distinguishing between hard and soft stops.

Ultimately, the culmination of our methodology involved the selection of final parameters deemed optimal for subsequent analyses. These parameters, underpinned by empirical evidence and rigorous experimentation, formed the cornerstone for further investigations into vehicular behavior and performance, enriching our understanding of driving dynamics and facilitating informed decision-making in related domains.

Results

In our investigation of hard stops, characterized by instances where vehicles come to a hard stop, we focused on utilizing key features including track ID, x and y coordinates, and velocity. Employing a stringent criterion where velocity is less than 1, and setting the parameters of epsilon to 100 and minimum samples to 29, we successfully delineated clusters indicative of hard stop events. The resultant clusters exhibited reasonable coherence, with each cluster representing a distinct range of coordinates within which vehicles experienced long halts. This approach yielded some insights into the spatial distribution of hard stops.

Conversely, the identification of soft stops, where vehicles slow dow and don't halt abruptly, presented a more intricate challenge. Incorporating an additional feature, frames, alongside the aforementioned features, we aimed to capture the nuanced dynamics of soft stop events. However, despite exhaustive experimentation involving varying values of epsilon and minimum samples, we encountered difficulties in achieving reasonable clustering outcomes. The inherent variability in the velocity range indicative of soft stops, falling between 5 and 10 units, and setting the parameters of epsilon to 100 and minimum samples to 100, posed challenges in delineating cohesive clusters. As of the present analysis, we have not yet identified parameter configurations that yield clusters reliably representing soft stop events within the context of moving traffic. This highlights the complexity inherent in discerning subtle variations in vehicular behavior and underscores the need for further refinement in our approach to soft stop detection.

In our analysis, we observed that Cluster 12 (orange-colored) appeared to be a false positive. This cluster comprised vehicles that exhibited behaviors aligning with soft stops; however, further examination revealed that these instances were primarily vehicles heading toward parking spots and pausing at entry gates. Our method categorized these halts as soft stops based on assumptions derived from previous calculations of the dataset.

These results underscore the divergent challenges associated with identifying hard and soft stop events in vehicular data, reflecting the multifaceted nature of driving dynamics. While our methodology demonstrates efficacy in capturing abrupt halts characteristic of hard stops, further refinement is warranted to effectively discern the nuanced patterns indicative of soft stop events. Continued exploration and refinement of clustering techniques, alongside the incorporation of additional contextual information, hold promise for enhancing our understanding of stop events in dynamic traffic scenarios.

For Hard Stops:

```
Cluster 0:
x range: 1602.0 - 3477.0
y range: 102.0 - 712.0

Cluster 1:
x range: 3241.0 - 3483.0
y range: 592.0 - 630.0

Cluster 2:
x range: 2555.0 - 3021.0
y range: 597.0 - 641.0

Cluster 3:
x range: 2423.0 - 2438.0
y range: 768.0 - 776.0

Cluster 4:
x range: 1278.0 - 1339.0
y range: 415.0 - 537.0

Cluster -1:
x range: 32.0 - 3733.0
y range: 181.0 - 1922.0

Cluster 5:
x range: 2134.0 - 2169.0
y range: 138.0 - 151.0
```

```
Cluster 6:
x range: 869.0 - 1076.0
y range: 1412.0 - 1606.0

Cluster 7:
x range: 542.0 - 568.0
y range: 1552.0 - 1564.0

Cluster 8:
x range: 1226.0 - 1233.0
y range: 728.0 - 734.0

Cluster 9:
x range: 1136.0 - 1325.0
y range: 1285.0 - 1398.0

Cluster 10:
x range: 2449.0 - 2480.0
y range: 1647.0 - 1655.0

Cluster 11:
x range: 3286.0 - 3324.0
y range: 849.0 - 857.0

Cluster 12:
x range: 908.0 - 1007.0
y range: 1259.0 - 1300.0

Cluster 13:
x range: 238.0 - 453.0
y range: 29.0 - 317.0

Cluster 14:
x range: 1579.0 - 1598.0
y range: 1337.0 - 1340.0
```

```
Cluster 6:
x range: 212.0 - 450.0
y range: 450.0 - 618.0

Cluster 7:
x range: 1372.0 - 1614.0
y range: 1393.0 - 1568.0

Cluster 8:
x range: 844.0 - 858.0
y range: 884.0 - 898.0

Cluster 9:
x range: 1445.0 - 1451.0
y range: 1274.0 - 1280.0

Cluster 10:
x range: 1734.0 - 1815.0
y range: 1303.0 - 1581.0

Cluster 11:
x range: 1921.0 - 1982.0
y range: 1305.0 - 1469.0

Cluster 12:
x range: 1556.0 - 1560.0
y range: 1281.0 - 1285.0

Cluster 13:
x range: 2050.0 - 2318.0
y range: 1249.0 - 1480.0

Cluster 14:
x range: 2253.0 - 2706.0
y range: 1282.0 - 1541.0

Cluster 15:
x range: 2446.0 - 2456.0
y range: 1298.0 - 1305.0
```

fig 1: Range of coordinates for each cluster for a hard stop

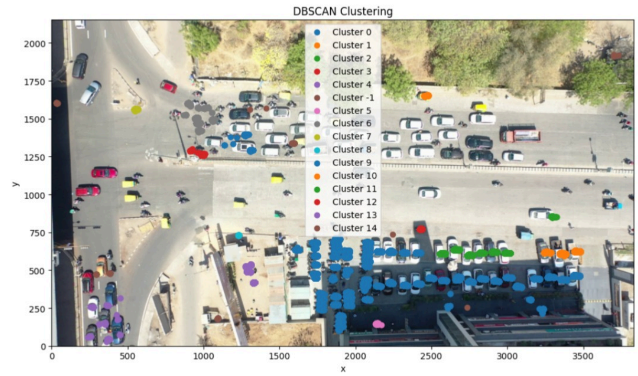

fig 2: Graphical representation of clusters For hard stop

```
Cluster -1:
x range: 36.0 - 3824.0
y range: 36.0 - 1862.0

Cluster 0:
x range: 1080.0 - 1563.0
y range: 756.0 - 907.0

Cluster 1:
x range: 893.0 - 1297.0
y range: 1319.0 - 1647.0

Cluster 2:
x range: 1767.0 - 1795.0
y range: 389.0 - 586.0

Cluster 4:
x range: 253.0 - 444.0
y range: 41.0 - 313.0

Cluster 3:
x range: 2149.0 - 2333.0
y range: 720.0 - 851.0

Cluster 5:
x range: 1115.0 - 1144.0
y range: 1269.0 - 1391.0
```

Fig 3: Range of coordinates for each cluster for soft stop

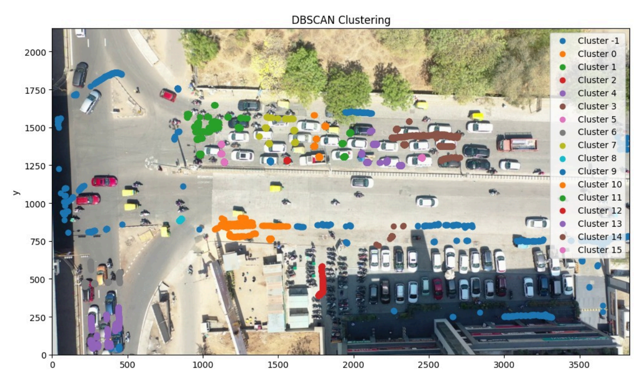

Fig 4: Graphical representation of clusters For soft stop

Discussion

In our pursuit of identifying hard and soft stop events within vehicular data, several approaches were considered to leverage clustering techniques, specifically utilizing DBSCAN. Here, we discuss three distinct methodologies proposed for this purpose:

- **Approach 1**: The initial step in Approach 1 involved data preprocessing, focusing on cleaning and removing irrelevant segments to streamline subsequent analyses. Subsequently, clustering was performed using DBSCAN, where the condition of zero velocity ($v = 0$) served as the primary criterion for identifying stop events. This facilitated the extraction of clusters representing spatial ranges where vehicles came to a halt. To distinguish between hard and soft stops within these clusters, additional parameters such as velocity and acceleration were examined over different time intervals. By assessing the ability of vehicles to traverse the identified clusters at varying speeds, we proposed a classification mechanism based on the feasibility of high-speed passage, indicative of soft stops, versus abrupt halts characteristic of hard stops.
- **Approach 2**: Similar to Approach 1, Approach 2 commenced with data cleaning to enhance data quality and relevance. Clustering using DBSCAN was then employed to identify spatial ranges where vehicles exhibited zero velocity, denoting stop conditions. After cluster identification, a novel aspect of this approach involved the utilization of time series algorithms to analyze vehicle velocity trends leading up to the identified stop locations. By scrutinizing the velocity profiles of vehicles before they entered the spatial ranges delineated by the clusters, we aimed to discern patterns indicative of hard or soft stop events. This approach leveraged temporal dynamics to provide deeper insights into stop event classification.
- **Approach 3**: In Approach 3, akin to the preceding methodologies, data cleaning constituted the preliminary step to refine the dataset. DBSCAN clustering was then applied to isolate spatial clusters representing stationary vehicles based on zero velocity conditions. Following cluster identification, an alternative strategy was proposed, focusing on analyzing the frequency of vehicle passages through the identified spatial ranges. By quantifying the frequency of vehicle movements within these clusters, we aimed to discern patterns indicative of stop event types. This approach offered a unique perspective, emphasizing the utilization of vehicular traffic flow dynamics to infer the nature of stop events.
- **Approach 4**: In this approach, we sought to leverage pre-clustered data to facilitate the clustering process. Initially, data extraction was performed from the time frames corresponding to soft stops for each vehicle in the dataset. Subsequently, a detailed analysis of the velocity and acceleration patterns were conducted, and using the Gaussian method, ranges for velocity and acceleration were determined. Within the dataset, vehicles exhibiting velocity and acceleration falling within these predetermined ranges were labeled as soft stops, while the remainder were labeled as hard stops. To minimize false positives, a stringent criterion was implemented wherein vehicles were only labeled as hard stops if they remained stationary at a particular location for a duration exceeding the typical soft stop duration. This meticulous approach aimed to enhance the accuracy of stop event classification, ensuring robustness in distinguishing between hard and soft stops within the vehicular dataset.

These proposed methodologies underscore the multifaceted nature of stop event identification and classification in vehicular data. Each approach presents distinct advantages and considerations, reflecting the complexity inherent in discerning subtle variations in vehicular behavior. Future research endeavors could involve the integration of these methodologies or the exploration of novel approaches to further enhance the accuracy and robustness of stop event classification in dynamic traffic environments.

Conclusion

In conclusion, our research investigated methodologies for identifying and classifying hard and soft stop events within vehicular data. Leveraging clustering techniques, particularly DBSCAN, alongside novel approaches such as time series analysis and traffic flow dynamics, we aimed to discern patterns indicative of stop event types. Through systematic experimentation and analysis, we observed the efficacy of different approaches in delineating spatial clusters representative of stop conditions. While each method presented unique advantages and considerations, our findings underscored the multifaceted nature of stop-event identification in dynamic traffic environments. Moving forward, continued exploration and refinement of these methodologies hold promise for enhancing our understanding of vehicular behavior and informing traffic management strategies for improved safety and efficiency on roadways.

References

[1] https://www.researchgate.net/publication/236002026_Spatio-Temporal_Clustering_Methods_Classification

[2] https://www.mdpi.com/2071-1050/13/2/690

Implementation

https://www.kaggle.com/code/yaxprajapati/notebook7562088143/edit