Ahmedabad University

Weekly Report - 1

Identify Hard stop and momentary stop using vehicle trajectory dataset

Team Name: **The Overfitters**

Name	Enrolment Number
Jinil Savaj	AU2240159
Jay Raval	AU2240151
Meet Suthar	AU2240198
Karan Prajapati	AU2240161
Vishesh Bhatia	AU2240027

Research on Stop Detection in Trajectory Data Analysis

In this week, we explored various methodologies for identifying stops in vehicle trajectory datasets. Our research primarily focuses on enhancing the accuracy of stop detection through clustering techniques and dynamic environmental context analysis. The aim is to develop an efficient model that can distinguish between different types of stops, such as hard stops and momentary stops, using spatio-temporal data.

Methodologies Explored:

1. RandomForest-Based Stop Detection Approach:

Inspired by *A Cluster-Driven Classification Approach to Truck Stop Location Identification Using Passive GPS Data* (Patel et al., 2022), this approach classifies stop types using a **RandomForestClassifier**. The model builds on techniques from freight movement analysis to distinguish between different stop purposes with high accuracy.

Key Features:

- RandomForestClassifier is chosen for its superior performance in stop classification.
- The model achieves high overall accuracy in correctly classifying all stops.
- Classifies stops into primary (e.g., delivery/pickup) and secondary (e.g., refueling/resting) categories.

Feature Selection:

- Distance traveled, velocity, acceleration, and speed thresholds to detect stop patterns.
- Temporal attributes such as stop duration and time of day.
- Proximity to POIs (businesses, fueling stations, rest areas) for contextual stop classification.
- Rolling window velocity averages to differentiate momentary stops from extended stops.
- Binary stop indicators to enhance classification accuracy.

By leveraging the RandomForest Model and spatial-temporal analysis, this approach enhances the understanding of mobility behaviors.

2. Clustering-Based Stop Detection:

We analyzed multiple clustering algorithms to determine their effectiveness in identifying vehicle stops.

DBSCAN Approach:

- **Advantages:**
  - Suitable for arbitrary cluster shapes and varying stop densities.
  - Automatically detects stop regions based on density without predefining the number of clusters.
  - Robust against noise, making it ideal for real-world vehicle movement data.
- **Implementation:**
  - Data cleaning and preprocessing.
  - Clustering based on velocity condition ($v = 0$).
  - Classification of stops based on velocity and acceleration variations.
  - Identifying high-density stop regions for further analysis.

K-Means Approach:

- **Challenges:**
  - Requires predefining the number of clusters (K).
  - Assumes spherical cluster shapes, which may not always represent real-world stop distributions.
  - Sensitive to outliers, which can distort results in trajectory data.
- **Potential Use:**
  - Can be useful in structured environments where the number of stops is known.

Machine Learning for Stop Classification:

- **Decision Trees & Gradient Boosting:** Investigate their potential for distinguishing between different stop types by capturing non-linear relationships.
- **Ensemble Learning Approaches:** Explore hybrid models combining multiple classifiers to enhance classification accuracy.
- **Real-Time Data Integration:** Incorporate geotagged social media data, traffic conditions, and road sensor data to refine stop detection insights.

Comparison of Methods:

Method	Pros	Cons
DBSCAN	No need to specify clusters, handles noise well	Sensitive to parameter tuning
K-Means	Fast and computationally efficient	Requires predefined clusters, sensitive to outliers
RandomForest Classifier	Considers dynamic environmental factors	Computationally intensive

Challenges and Next Steps:

- **Feature Exploration:** Investigate key features that could improve stop classification accuracy.
- **Dataset Identification:** Research and evaluate datasets that align with project objectives.
- **Methodology Assessment:** Analyze various classification approaches to determine the most effective techniques.

Conclusion

Our exploration highlights **DBSCAN** as a suitable clustering technique due to its flexibility and noise-handling capabilities. However, the **RandomForestClassifier** approach presents a novel perspective by incorporating environmental context, which can significantly improve detection accuracy. Future work will focus on refining feature selection and integrating and comparing machine learning models to enhance stop classification performance.

Bibliography

[1] "Difference between K-Means and DBScan Clustering," GeeksforGeeks, Jul. 19, 2020. https://www.geeksforgeeks.org/difference-between-k-means-and-dbscan-clustering/

[2] D. Birant and A. Kut, “ST-DBSCAN: An algorithm for clustering spatial–temporal data,” Data & Knowledge Engineering, vol. 60, no. 1, pp. 208–221, Jan. 2007, doi: https://doi.org/10.1016/j.datak.2006.01.013.

[3] D. Chang, Y. Ma, and X. Ding, “Time Series Clustering Based on Singularity,” International Journal of Computers, Communications & Control (IJCCC), vol. 12, no. 6, p. 790, Dec. 2017, doi: https://doi.org/10.15837/ijccc.2017.6.3002.

[4] V. Patel, M. Maleki, Mehdi Kargar, J. Chen, and H. Maoh, “A cluster-driven classification approach to truck stop location identification using passive GPS data,” Journal of Geographical Systems, vol. 24, no. 4, pp. 1–21, Jun. 2022, doi: https://doi.org/10.1007/s10109-022-00380-y.

[5] Negin Masnabadi, Farhad Hosseinali, and Zahra Bahramian, “Developing a spatial and temporal density-based clustering algorithm to extract stop locations from the user’s trajectory,” Journal of Geospatial Information Technology, vol. 9, no. 2, pp. 105–128, Oct. 2021, doi: https://doi.org/10.52547/jgit.9.2.105.