Data-Driven Classification of Vehicle Driving Behavior in Mixed Traffic Using Car-Following Trajectories

THE UNIVERSITY OF ARIZONA

THE UNIVERSITY OF ARIZONA COLLEGE OF ENGINEERING
Center for Applied Transportation Sciences

Jun Zhao, Mingfeng Shang, Gabriel Geffen, Yao-Jan Wu

I. Introduction

- **Automated Vehicles (AVs):** Transforming transportation, improving string stability and traffic throughput, reducing fuel consumption and emissions.
- **Electric Vehicles (EVs):** Increasing adoption, distinct powertrain characteristics affecting traffic dynamics.
- **Challenges:** Limited understanding of the interplay between vehicle powertrain types and operation modes in real-world mixed traffic.
- **Objective:** Classify vehicles in mixed traffic using two approaches: single-stage models that directly categorize the data and multi-stage models that perform classification in sequential steps.

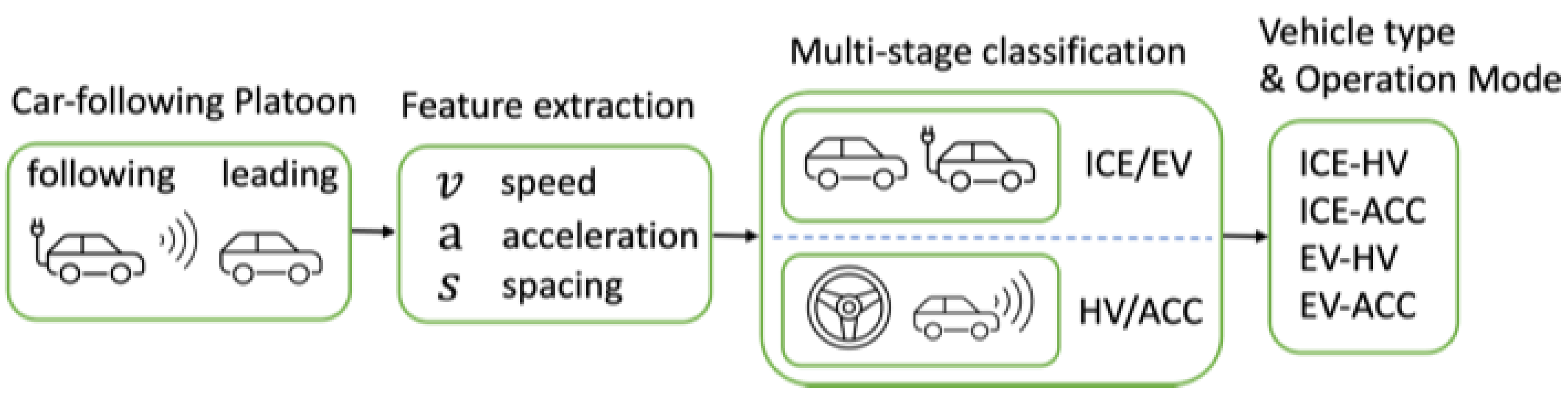
II. Modeling

A. Single-stage classification model

- Description: Performing classification in one step by directly mapping input data to output labels.
- Advantages: Simplicity and efficiency.
- Challenges: May underperform with complex or hierarchical data patterns.
- Models: Random Forest (RF), XGBoost, Support Vector Machine (SVM), Logistic Regression (LR).

B. Multi-stage classification model

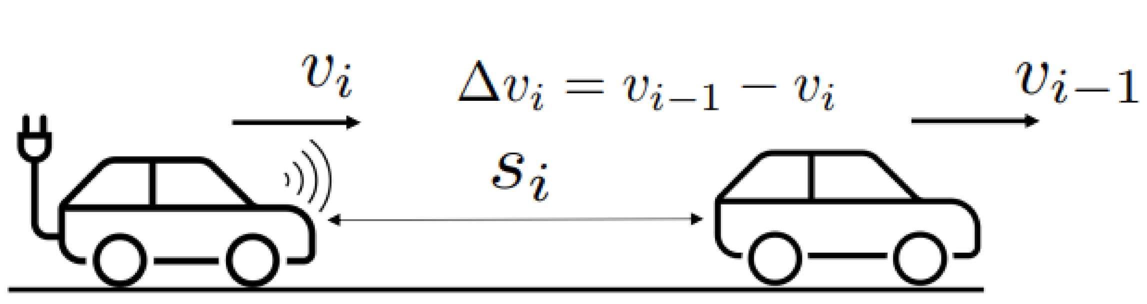
- Description: Decomposing the classification task into sequential subtasks for improved accuracy and interpretability.
- Approach: Two-stage model first classifies operation mode (ACC vs. Human-driven), then vehicle type (EV vs. ICE).
- Advantages: Improving interpretability and potentially higher accuracy.
- Challenges: Increasing latency and potential error propagation.



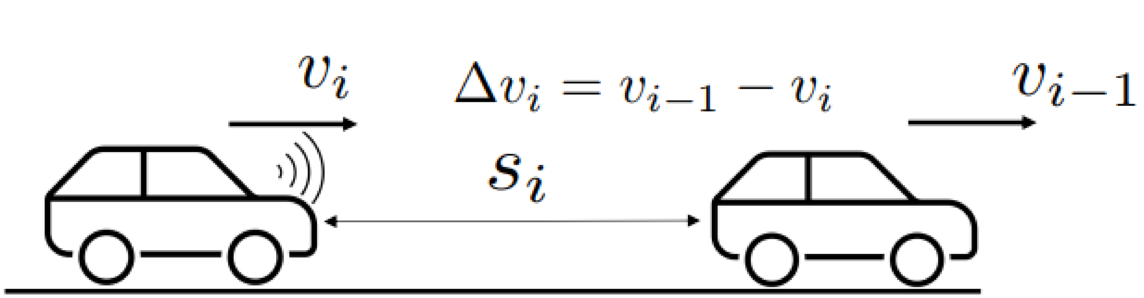
III. Data

A. Data Processing

- Objective: Analyze interaction dynamics between two vehicles in a car-following scenario.
- Method: Decompose vehicle platoon into multiple two-vehicle pairs.
- Variables:
  - Velocity: v_i (following vehicle) and v_{i-1} (lead vehicle).
  - Inter-vehicle spacing (IVS): S_i , distance between vehicles.
  - Relative velocity: $\Delta v_i = v_{i-1} - v_i$, speed difference.
  - Acceleration: a_i (following vehicle) and a_{i-1} (lead vehicle).
- Input Vector: $X_i = (v_i, v_{i-1}, a_i, a_{i-1}, \Delta v_i)$



(a) EV-ACC vehicle and its lead vehicle.



(b) ICE-ACC vehicle and its lead vehicle.

B. Data Descriptive Analysis

- Objective: Investigate differences in car-following behaviors across four categories: EV-ACC, ICE-ACC, EV-HV, and ICE-HV.
- Class Imbalance: Addressed using class weighting strategies.
- Findings:
  - EVs:
    - Less oscillatory behavior and smoother spacing
    - Follow lead vehicles more closely than ICE vehicles.
  - ACC-equipped vehicles:
    - Smoother car-following dynamics.
  - Human-driven vehicles :
    - Keep longer following distances and show more oscillatory behavior than ACC-equipped vehicles.

IV. Results

A. Single-stage classification models

- Random Forest: Highest accuracy (74%) and F1 Score (0.70).
- XGBoost, SVM, Logistic Regression: Comparatively lower performance.

Model	Accuracy	F1 Score
Random Forest	74%	0.70
XGBoost	69%	0.66
SVM	67%	0.68
Logistic Regression	60%	0.60

B. Multi-stage classification models

- Sequence Operation Mode-Vehicle Type: First classify operation mode, then vehicle type.
  - Stage 1: XGBoost achieved 87% accuracy.
  - Stage 2: Random Forest achieved 86% accuracy.
  - Overall: 75% accuracy and F1 Score of 0.73.
- Sequence Vehicle Type-Operation Mode: First classify vehicle type, then operation mode.
  - Similar performance: Slightly lower accuracy and F1 Score.

C. Comparisons between the single-stage and the multi-stage model

- Single-stage vs. Multi-stage:
  - Single-stage: Simplicity and fewer tuning steps.
  - Multi-stage: Better interpretability and slightly higher performance.
  - Conclusion: Both approaches are effective; choice depends on the need for simplicity vs. interpretability.

Stage 1 Model	Stage 1 Accuracy	Stage 2 Model	Stage 2 Accuracy	Total Accuracy	Total F1 Score
Random Forest	85%	Random Forest	86%	73%	0.69
		XGBoost	83%	70%	0.66
		SVM	85%	72%	0.68
		Logistic Regression	84%	71%	0.66
XGBoost	**87%**	**Random Forest**	**86%**	**75%**	**0.73**
		XGBoost	81%	71%	0.68
		SVM	83%	73%	0.70
		Logistic Regression	83%	73%	0.69
SVM	84%	Random Forest	86%	72%	0.65
		XGBoost	82%	68%	0.61
		SVM	86%	72%	0.65
		Logistic Regression	83%	69%	0.61
Logistic Regression	85%	Random Forest	86%	73%	0.68
		XGBoost	81%	68%	0.63
		SVM	85%	72%	0.67
		Logistic Regression	84%	71%	0.65

V. Concluding Remarks

- Impact: Vehicle automation and electrification significantly influence traffic dynamics.
- Framework: Effective in classifying vehicle types and operation modes using real-world trajectory data.
- Future Work: Incorporate synthetic and data-driven trajectories to enhance model robustness and practical applicability

NOT OK TO SHARE ON SOCIAL MEDIA