Spatial and Temporal Characterization of Travel Patterns in

Ritu Patel, Sarthak Siddhpura, Moin Vinchhi, Vrunda Patel, Vishv Boda
Ahmedabad University, Ahmedabad, Gujarat 380009, India

I. Research Contributions by Paper

- Proposed a systematic framework for clustering and classifying vehicle trajectories using spatial and temporal features, bypassing the need for map-matching by directly processing raw trajectory data.
- Introduced an LCS-based similarity measure that quantifies route overlap between trajectories, using travel distance instead of match count to handle varying sampling rates, tested on New York City SHRP2 data.
- Extended the DBSCAN algorithm with LCS-based distance to identify spatially distinct traffic stream clusters, producing five major clusters from 1000 trajectories with parameters $\epsilon = 0.4$ and MinTrs = 14.
- Developed the concept of Cluster Representative Subsequences (CRSs) via hierarchical agglomerative clustering of pairwise LCSs, enabling efficient representation of cluster-shared road segments.
- Demonstrated trajectory classification by assigning new trajectories to clusters based on LCS similarity with CRSs, achieving spatial pattern detection without network topology knowledge.
- Applied the framework to analyze temporal traffic patterns, clustering daily time series into day groups (e.g., Sundays vs. weekdays for Cluster 1), enhancing travel time reliability analysis.
- Validated the approach using 1000 SHRP2 trajectories from May 4, 2010, in NYC, identifying key traffic streams (e.g., Manhattan-Long Island, JFK routes) and their temporal variations.

II. Key Insights From Paper

The paper by Kim and Mahmassani provides a robust foundation for our research on identifying abnormal driving behavior through trajectory datasets, leveraging its LCS-based similarity measure and clustering framework. The Longest Common Subsequence (LCS) approach, which quantifies spatial overlap between trajectories using travel distance rather than mere point matches, aligns seamlessly with our goal of defining "normal" behavior on road patches probabilistically. By assuming drivers typically follow similar routes on a given patch, we can adapt LCS to establish a baseline of normalcy—trajectories with high LCS similarity (e.g., simSeq $\geq$ 0.6) to a reference cluster could represent typical behavior, while those with low similarity (e.g., simSeq < 0.3) might indicate anomalies like erratic detours or lane deviations. The paper's experiment with NYC data, showing distinct similarity values (e.g., 0.63 for diverging, 0.99 for overlapping trajectories), suggests a statistical approach: compute LCS distances for a training set of trajectories on a patch, model their distribution (e.g., Gaussian), and set a threshold (e.g., mean - 2σ) to flag abnormalities. This method's robustness to varying sampling rates—demonstrated by adjusting for uneven intervals in SHRP2 data—ensures applicability to real-world GPS trajectories, where noise and inconsistencies are common, making it a practical tool for our binary classifier development.

The integration of DBSCAN with LCS-based distance offers a scalable clustering strategy that can enhance anomaly detection efficiency, a critical insight for our project. By grouping trajectories into spatially coherent clusters (e.g., five clusters from 1000 trajectories with $\epsilon = 0.4$), the paper identifies major traffic streams without requiring predefined road network information, a feature we can exploit to segment road patches dynamically. For our classifier, this implies a two-stage process: first, cluster trajectories on a patch to define normal behavior clusters (e.g., core trajectories with MinTrs $\geq$ 14), then classify new trajectories as normal if they fall within an ϵ-neighborhood (e.g., distSeq ≤ 0.4) of a cluster's core, or abnormal if labeled as noise by DBSCAN. The probabilistic angle emerges from analyzing cluster density—core trajectories represent high-probability normal routes, while noise trajectories (e.g., those with insufficient neighbors) signal potential anomalies. The paper's finding that clusters capture directional flows (e.g., Manhattan to Long Island vs. reverse) suggests incorporating spatial-temporal features like directionality into our model, using statistical tests (e.g., chi-square) to validate cluster consistency and detect deviations indicative of abnormal behavior, such as sudden direction changes.

Finally, the Cluster Representative Subsequence (CRS) concept and temporal analysis provide a framework for refining our classifier's interpretability and adaptability. CRSs, derived by merging overlapping LCSs, distill dense road segments shared by cluster members, offering a succinct representation of normal routes (e.g., JFK airport access roads in Cluster 3). For our research, CRSs could serve as probabilistic templates: a new trajectory's LCS similarity to a CRS (e.g., simSeq > 0) assigns it to a normal cluster, while low similarity flags it as abnormal, with p-values from a similarity distribution providing confidence levels. The paper's temporal clustering of daily patterns (e.g., Sundays vs. weekdays in Cluster 1) inspires a dynamic classifier that adjusts normalcy thresholds based on time-specific norms—abnormal behavior might differ between rush hours and weekends. Tested on 80 daily samples, this approach highlights how spatial clusters reveal temporal variations, suggesting we extend our binary classifier with a temporal dimension (e.g., K-means on time series of LCS scores) to capture context-dependent anomalies like speeding during off-peak hours. Together, these insights enable a data-driven, statistically grounded classifier that leverages trajectory clustering for precise abnormal behavior detection.

III. Unaddressed Issues and Assumptions

- **Unaddressed Issues:**
  - Lacks explicit anomaly detection methodology, focusing on clustering rather than classifying abnormal vs. normal trajectories, limiting direct applicability to our binary classifier.
  - Ignores speed and acceleration features in LCS, potentially missing temporal anomalies (e.g., aggressive driving) critical for driving behavior analysis.
  - Does not address scalability for real-time processing of large-scale trajectory streams, a practical concern for dynamic anomaly detection.
  - Limited evaluation of CRS robustness against noise or rare events, leaving uncertainty in its effectiveness for outlier detection in diverse datasets.
- **Assumptions Made:**
  - Assumes route overlap (LCS) sufficiently captures similarity, overlooking behavioral differences (e.g., speeding) within overlapping trajectories.
  - Assumes short sampling intervals (2-3 seconds) allow straight-line approximations, risking inaccuracies with sparse or erratic GPS data.
  - Assumes spatial proximity alone defines relatedness, neglecting contextual factors like traffic conditions or road type that may alter normal behavior.
  - Assumes DBSCAN parameters (ϵ, MinTrs) can be intuitively tuned, potentially leading to inconsistent clustering across varied road patches.

IV. Motivation Behind Choosing This Paper

This paper was chosen for its innovative LCS-based trajectory clustering framework, which directly supports our research on abnormal driving behavior detection using probabilistic-statistical methods. The LCS similarity measure, robust to sampling rate variations, provides a spatial-temporal foundation to define normal behavior on road patches, aligning with our assumption of similar driver behavior. The DBSCAN extension and CRS generation offer scalable tools to cluster trajectories and extract representative patterns, enabling a binary classifier that flags anomalies as deviations from statistically defined norms. Tested on real NYC data, the framework's ability to uncover spatial and temporal patterns without map-matching inspires a data-driven approach to enhance road safety through precise, context-aware anomaly detection.

References