A Novel Approach to Route Similarity Measures for Shared Mobility Matching Systems

Antoniu Negrea

2025

Contents

Chapter 1

Modeling the Experimental Part

This chapter rigorously describes the data used, the planned experiments, the mathematical modeling of the similarity measures, the algorithms compared, and the validation methods. The goal is to demonstrably prove, in a reproducible and analytically supported manner, that the proposed approach brings improvements over existing methods found in the literature.

1.1 The Dataset

1.1.1 Representation of Urban Routes

Each route is modeled as an ordered list of GPS points:

$$R = \{(lat_1, lon_1), \ldots, (lat_n, lon_n)\}.$$

The data sources are:

- simplified urban road network
- artificially simulated routes for controlled scenarios

To reduce complexity, the GPS points are projected onto a discretized network $G = (V, E)$, where V are intersections and E are segments.

1.1.2 User Profiles

Each user is associated with a triplet:

$$U_i = (o_i, d_i, t_i),$$

where o_i is the origin, d_i the destination, and t_i the temporal interval (time window).

1.2 Simplified Similarity Measures

The proposed methodology uses two measures that are easy to implement:

1.2.1 Geometric Similarity

For two routes R_1, R_2:

$$S_{geo}(R_1, R_2) = 1 - \frac{1}{|R_1|} \sum_{p \in R_1} \min_{q \in R_2} d(p, q),$$

where $d(p, q)$ is the Haversine distance.

1.2.2 Segment Overlap

$$S_{overlap}(R_1, R_2) = \frac{|R_1 \cap R_2|}{\max(|R_1|, |R_2|)}.$$

1.2.3 The Final Similarity Function

A simple linear function:

$$S_{final}(R_1, R_2) = \alpha S_{geo}(R_1, R_2) + \beta S_{overlap}(R_1, R_2),$$

where $\alpha + \beta = 1$.

1.3 Matching Algorithms

1.3.1 Static Matching — Partition Merging

The objective is to group users such that the cost is minimized:

$$\text{cost}(G) = \sum_{i,j \in G} (1 - S_{final}(R_i, R_j)).$$

1.3.2 Dynamic Matching — Greedy

For a new request:

$$\Delta\text{cost} = \text{cost}(G \cup \{U_k\}) - \text{cost}(G).$$

The request is allocated to the group with the minimum Δcost.

1.4 Proposed Experiments

1.4.1 Experiment 1: Static vs. Greedy

The comparison between the two algorithms (Static Partition Merging and Dynamic Greedy) uses two principal metrics:

1. **Efficiency** (Travel Gain): The reduction in total travel distance.

$$\text{TravelGain} = \frac{D_{solo} - D_{shared}}{D_{solo}}.$$

2. **Equity** (Maximum Relative Detour, MRD): The highest proportional increase in travel distance/time experienced by any single rider in a shared group. This assesses the fairness of the solution.

1.4.2 Experiment 2: Impact of Parameters α, β

The influence of weights on matching quality is analyzed using both the **Travel Gain** and **MRD** metrics to assess the trade-off between efficiency and equity.

1.4.3 Experiment 3: Scalability

We measure:

- runtime;
- memory used.

1.5 Validation Methods

Validation is exclusively numerical:

1.5.1 Internal Validation

Repeated simulations with artificially generated data.

1.5.2 External Validation

Comparison of results with:

- Xia & Curtin (2019) – spatial model;
- Duan (2018) – partition merging;
- Sun (2023) – greedy.

1.6 Conclusion

The experimental model is simplified, reproducible, and easy to implement. It allows for the evaluation of similarity functions and matching algorithms.