

Proposal for Semester Project

Patterns & Trends in Environmental Data / Computational Movement Analysis
Geo 880

Semester:	FS24
Data:	Daily trajectories
Title:	Detecting Speed and Route Deviations in Biking Data
Student 1:	Tanja Falasca
Student 2:	Tatiana Blumberg

Abstract

In this work, we will examine biking GPS data. The focus of this project is to analyse speed and route deviations considering the environment and time in which the tracking was conducted. The analysis will be conducted in the R environment. Depending on the results, we will try to identify possible reasons behind these deviations.

Research Questions

1. Are there any speed deviations, and if so, what are they and why do they occur? (Consider different speeds on various trips or fluctuations during a single trip, and explore the potential reasons.)
2. Are there any differences in the routes preferred at different times of the day or in different directions (e.g., outbound and return trips)?

Results / products

We expect to find speed deviations, possibly due to traffic conditions (e.g. traffic lights or increased traffic during rush hours). We also expect differences between the outbound and return trips, as well as deviations between routes taken at different times of the day.

Data

We are going to work with trajectory data and use the biking data from it. The data was collected with the Strava app mainly in Karlsruhe, Germany. The additional data we are going to use is the OpenStreetMap of the Karlsruhe city, which will be taken from an open internet resource. During the work on the project, we might consider using some other open data.

Analytical concepts

We will analyse our data from the perspective of constrained Euclidean space, as bicyclists are constrained by the streets in their movement. We will use the Lagrangian movement perspective, as the GPS fixes are saved instantly and are not connected to any static checkpoints.

R concepts

We will mainly use the following R concepts, functions and packages: 1. Data Import and Cleaning: readr, dplyr 2. Spatial Data Handling: sf 3. Data Visualization: ggplot2, mapview 4. Integration with OSM Data: osmdata, Osmar 5. Speed and Route Analysis: Custom functions for detecting deviations, clustering algorithms, and network analysis methods

Risk analysis

The first challenge is the potential presence of outliers in the travel data - data collected while using other modes of transport (e.g., car, train). As this is not suited for our research questions, we will filter this data out during preprocessing. The second challenge is that the initial dataset might not be large enough for robust analysis. Our solution would be to extract data over a longer period of time from the Strava app.

Questions?

1. How to match raw GPS data with the bikeway?
2. What potential indicator for speed deviation should we consider? (e.g., semantic indicators such as station origin/destination, operational times)
3. How to match the open street data with the biking data? (this might be needed for identifying the traffic lights and the road junctions)

References (APA 7)

Jankowska, M. M., Yang, J., Luo, N., Spoon, C., & Benmarhnia, T. (2023). Accounting for space, time, and behavior using GPS derived dynamic measures of environmental exposure. *Health and Place/Health & Place (Online)*, 79, 102706. <https://doi.org/10.1016/j.healthplace.2021.102706>