

Proposal for Semester Project

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

Semester:	FS24
Data:	Walking activity recorded with Strava
Title:	Differentiation of walking patterns
Student 1:	Saskia Gianola
Student 2:	Sarah Wirth

Abstract

Based on walking data recorded with Strava we want to derive the purpose of the movement. Using the attributes sinuosity, speed and number of stops we want to find out whether an activity belongs to a commuting path, is recreational or belongs to running an errand for example grocery shopping. We want to test if the trajectories differ enough to derive walking patterns. As we recorded the data ourselves, we can add semantic information and test whether the classification is possible.

Research Questions

Is it possible to derive the type of activity from movement data considering the attributes sinuosity, speed and visited locations? Which type of analytical concept performs best in differentiating between the different activities? Optional: Do the concepts perform differently for different people.

Results / products

We expect to be able to distinguish between different types of walking-activity (commuting path, recreational, running errands) as we expect the trajectories to differ regarding sinuosity, speed and number of stops. As a result we aim to have an estimation of the performance of different analytical concepts at the task of differentiating between walking-activities. This could be presented as misclassification tables for each analytical concept, in which we assess which concept succeeded best at discerning the different walking-activities. For visualization we aim to create maps showing the different groups of trajectories.

Data

We use the activity data, which Saskia documented with Strava. The Strava- data is suited for our project because of its high spatial and temporal resolution. If more trajectories are required, the following options are possible: Option A: only work with the data from 12 trajectories Option B: Saskia documents more trajectories. -> suitable if we only want to use data from one person Option C: Sarah documents more trajectories. -> suitable if we want to use data from two different people and want to see whether the same analytical concepts differ in their capability of differing the type of walking- activities.

Analytical concepts

We intend to use similarity measures to find similar trajectories. Furthermore, we want to extract characteristics of the trajectories to determine sinuosity, speed and number of stops for each trajectory. Then we want to classify the trajectories based on these characteristics individually using similarity measures. For improved classification, we could get information about starting- and endpoints by including the distances to certain infrastructures (train station, bus stops, retail stores). Last, we want to visualize the paths and groups of paths.

R concepts

We want to use the following packages: (“XML”), (“leaflet”), (“sf”), (“tmap”), (“ggplot2”), (“tidyverse”), (“readr”)

We further intend to use different similarity measures (like the ones previously used in the course) as well as other similarity measures based on derivatives of the time stamped coordinates like speed, sinuosity and number of stops, to group our trajectories. We also aim to write our own functions to facilitate certain steps, make our project more immune to errors and more adaptable to changes in our data.

Risk analysis

Our analysis depends on whether our data contains enough trajectories. Therefore we might need more Data (Options as described in “Data”). Adding data from another person might also add a challenge as we would need to distinguish between walking-characteristics for each person. Our plan B is going for travel-mode detection using ca. 2 months of Google-Timeline data, which has rather big temporal intervals, which would be another risk for our plan B.

Questions?

How many trajectories do we need, to have a sufficient amount of data for our project? Which option to get more trajectories (“Data”) should we use? As the task of differentiating between different walking activities seems difficult, we are not sure about the workload of comparing the capability of the different analytical concepts for the movement- data of different people.