Proposal for Semester Project

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

Semester:	FS24
Data:	GPX data related to running workouts measured with a Garmin watch
Title:	Identification of patterns and trends in running training performance
Student 1:	Daniel Cellerino
Student 2:	Sophie Blatter

Abstract

In this project, we will analyze three running sessions executed by Daniel on the same route at different times of the year or in different years. The goal is to identify clusters of route segments that exhibit similar performances. This will allow us to uncover distinct patterns or behaviors during the run, understanding which parts of the route most significantly impact Daniel's performance and on which segments he can focus his training to improve his overall performance. This is achieved by performing similarity analysis using Dynamic Time Warping (DTW) to compare the trajectories of the runs, taking into account the different sampling frequencies and the number of recorded points. Subsequently, we will apply a Monte Carlo simulation to assess the robustness of the obtained results.

Research Questions

Research Question 1: What are the recurring patterns in the performances of three running sessions executed on the same route at different times of the year?

Research Question 2: How can the identified patterns in the running performances be utilized to improve performance?

Results / products

Research Question 1:

For the first research question we anticipate that Daniel's performance will vary significantly across different terrain levels. Specifically, we expect:

- Daniel will likely maintain a more consistent and higher average speed on flat sections, with minimal time deviation from the average across different runs. We expect that Daniel will get faster over time, since he gets more experienced in running.

- On moderate terrain, there may be a slight reduction in speed compared to flat sections, with some variability depending on specific conditions during each run.

- The steep sections are expected to show the greatest variability in speed and the most significant time deviations. In these sections, physical conditions could be of significant importance and have a greater impact on the variability of the three analyzed runs.

We anticipate to identify specific sections where Daniel can focus his training to improve performance.

Research question 2:

- For the second research question, we expect that Daniel can improve his performance by focusing on better preparation for uphill segments of the run.

Data

We will use movement data recorded by Daniel using a Garmin watch and uploaded to Strava over several years. At the same time, it will be necessary to ensure that the data, in addition to containing geographic coordinates, also include elevation values above sea level. If this information is not present in the GPX file downloaded from Strava, it will be necessary to extract it from external sources.

Analytical concepts

In our semester project, we will primarily use analytical concepts related to trajectory analysis. We will use conceptual movement spaces with discrete properties and adopt the trajectory modeling approach based on the concept of three-dimensional network space. This will allow us to understand the patterns and behaviors in Daniel's movements along the designated route. Additionally, we will use additional spatial analysis methods such as map matching to associate trajectory points with road segments and slope calculation to assess the difficulty of different sections of the route.

R concepts

1. GPX Data Importation

- Installation and loading of necessary libraries.
- Importation of GPX files for three running sessions (sf).

2. Normalization and Data Preprocessing

- Changing the coordinate system and calculating average speed between points (sf, dplyr).
- Adding elevation data and calculating slope (elevatr).

3. Similarity Analysis with DTW

- Calculation of DTW distances between the three running sessions (dtw).
- Normalization and clustering using DTW distances.

4. Monte Carlo Simulation on Results

- Adding noise to the results.
- Running Monte Carlo simulation and analyzing the results.

Additionally, we may explore other spatial analysis methods as needed during the project, such as map matching algorithms or stop-and-move detection techniques

Risk analysis

We need to decide whether to also consider the downhill sections of the run. In that case, the speed decreases compared to running on flat terrain.

It is likely that due to the GPS accuracy, it will not be possible to measure the sinuosity of the path since the measurement intervals are irregular and the watch's accuracy in optimal conditions is 3-5 meters.

Questions?

How can we work with data that has been measured at irregular intervals (DTW)? How can we add altitude above sea level to our points?