sccer mobility
Swiss Competence Center for Energy Research
Efficient Technologies and Systems for Mobility

In cooperation with the CTI
Energy funding programme
Swiss Competence Centers for Energy Research
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra
Swiss Confederation
Commission for Technology and Innovation CTI

Data Completeness Assessment for GNSS-assisted Travel Surveys

Since the late 1990s, GNSS-assisted movement tracking technologies have gradually replaced traditional travel survey approaches, such as face-to-face interviews, mailout/mail-back or telephone surveys. A practical problem, however, is posed by insufficient data quality, especially data incompleteness. In fact, detecting and quantifying recording gaps during the recording period, which occur e.g., as a result of the cold start problem at the start of movement, bad signal reception, participants leaving the device switched off, or other technological problems, is critical to avoid erroneous results of subsequent analyses. In this work, we propose a method for completeness assessment of movement trajectory datasets which were recorded in the course of a travel survey by detecting and quantifying recording gaps. We apply the method to an exemplary dataset collected from 79 participants over 18 weeks, and briefly discuss the results.

David Jonietz, Dominik Bucher, Martin Raubal
Institute of Cartography and Geoinformation
ETH Zürich
jonietzd@ethz.ch, dobucher@ethz.ch, mraubal@ethz.ch

Introduction

The use of Global Navigation Satellite System (GNSS)-assisted methods for travel surveys has gained much interest due to its strengths, such as a relatively high accuracy in recording time and position, low cost, and less problems with regards to trip-misreporting by respondents (Shen and Stopher 2017). Recording gaps, however, which occur frequently and for various reasons, may lead to erroneous or even heavily biased results of subsequent analyses such as the calculation of general statistics such as the average duration and length of trips, the modal split or frequently visited places (Hecker et al. 2010). In this study, we propose a method for detecting these gaps in movement trajectory data, and assessing the data completeness for individual survey participants. We apply our method to a dataset collected from 79 participants over 18 weeks in Switzerland.

Movement Data

Raw movement data obtained from GNSS typically consists of a series of chronologically ordered x, y-coordinate pairs enriched with a time stamp (Zheng (2016) . In the course of typical preprocessing, the movement data is typically structured into meaningful units (cf. Schönfelder and Axhausen 2010). Here, we use:

- **track points**: recorded positions at the fundamental level
- **trip legs**: aggregated track points based on the used transport mode
- **trips**: consist of one or more legs, and describe the journey from one 'activity' to another.
- **stay points**: locations where a participant spent longer than a certain time span; can qualify as an activity if they represent an actual destination of travel (e.g., work, home or a shop)

Concept –Types of Recording Gaps

As a first step, we distinguish between different types of recording gaps:

- **Temporal gaps**: the duration with no recorded data between the time stamp of the last trip leg or stay point and the first recorded time stamp of the subsequent trip leg or stay point where the spatial deviance between the position of the two points is smaller than an expected GPS error (e.g., 250 meters).
- **Spatio-temporal gaps**: gaps for which the spatial distance between the last track point of the former, and the first track point of the latter trip leg or stay point is larger than an expected GPS error

This distinction is motivated by the fact that in the first case, chances are high that no mobility behavior has been missed during the recording gap, whereas in the second case, the user has certainly changed position.

Concept – Completeness Assessment

Based on both types of gaps, the data completeness for each participant during the current time interval can then be evaluated based on several index values:

Index	Explanation
$pct\_dur\_gap_{total}$	percentage of temporal and spatio-temporal gaps with regards to the entire recording period
$pct\_dur\_gap_{spatial}$	percentage of spatio-temporal gaps with regards to the entire recording period
$avg\_dur\_gap_{spatial}$	average duration of spatio-temporal gap
$pct\_dist\_gap_{spatial}$	percentage of spatio-temporal gaps with regards to the entire recorded movement distance
$days\_gap_{spatial}$	number of days with spatio-temporal gap
$time\_gap_{spatial}$	number of spatio-temporal gaps in certain time intervals during the day

Tab. 1: Index Values

Case Study

We applied the concept to an exemplary subset of the dataset collected as part of the *GoEco!* project. Figures 1-3 show exemplary results on the aggregated as well as the user level.

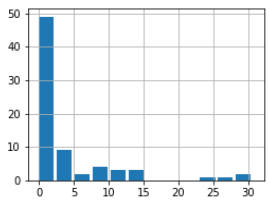

Fig. 1: $pct\_dur\_gap_{total}$ for all users

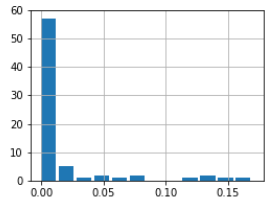

Fig. 2: $pct\_dist\_gap_{spatial}$ for all users

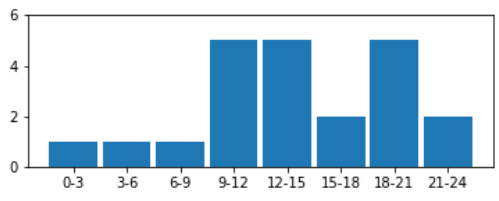

Fig. 3: $time\_gap_{spatial}$ for one user

Expected Impact

Identifying recording gaps and assessing data completeness is critical for ensuring the validity of GNSS-assisted travel surveys. The proposed method distinguishes between temporal and spatio-temporal recording gaps, since we expect the latter to have a more distorting effect on the analytic process. Using a range of index values, we can assess data completeness at the level of the individual survey participant, and potentially inform analysts in their decision to e.g., exclude individual persons from the analysis or apply an imputation method. Further, by incorporating the temporal dimension, more detailed decisions are possible, such as selecting certain times of day, days or weeks where data completeness is sufficient, or testing for temporal randomness or observable patterns of missing recordings.

References

Hecker D, Stange H, Korner C, May M (2010) Sample bias due to missing data in mobility surveys. In 2010 IEEE International Conference on Data Mining Workshops, pp. 241–248.

Schönfelder S, Axhausen KW (2010) Urban rythms and travel behavior: spatial and temporal phenomena of daily travel. Ashgate Publishing Ltd.

Shen L, Stopher, PR (2017) Review of GPS travel survey and GPS data-processing methods. Transport Reviews, 34(3): 1–19.

Zheng Y (2016) Trajectory data mining: An overview. ACM Trans. Intell. Syst. Technol. 6(3): 29(1)-29(41)

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich