The Effect of Post-Processing in Stop-Move Detection of GPS Data: A Preliminary Study

Eun-Kyeong Kim*,**,[1], Elena Ebert*, Robert Weibel*

* Department of Geography, University of Zurich, Switzerland
** University Research Priority Program (URPP) 'Dynamics of Healthy Aging', University of Zurich, Switzerland

Abstract. Stop-move detection has been an essential step to construct semantic trajectories and extract meaningful activity sequences of moving objects. Detecting stop and move segments accurately is critical because errors occurred in stop-move detection can be propagated and amplified in later steps in trajectory data analysis. In particular, post-processing that merges or discards the detected stop-move segments can make an impact on the accuracy and characteristics of detected stops and moves. Although many stop-move detection algorithms exist and new methods are continuously proposed in the field, studies on comparing the performance of the stop-move detection methods are still scarce.

In this study, we evaluated the effect of post-processing in stop-move detection with four selected existing stop-move detection algorithms—CandidateStops, SOC, POSMIT, and MBGP—in two input-data scenarios: (1) original data and (2) sampled data. The detected stops were assessed by two quantitative measures that quantify the accuracy at different levels of aggregation in space and time: (1) accuracy based on individual data points (i.e., F-measure) and (2) the shape of detected stops (i.e., shape compactness). With the case study, we found that the impact of post-processing on the detection results can vary by a selected algorithm and input data sparsity. The results can potentially provide insights into how to adopt and maneuver the stop-move detection methods for GPS data analysis.

Keywords. Stop detection, GPS data analysis, Semantic trajectory, Algorithm comparison.

[1] Corresponding author, eun-kyeong.kim@uzh.ch

LBS 2021
16th Conference on Location Based Services
24-25 November 2021 Online

Published in "Proceedings of the 16th International Conference on Location Based Services (LBS 2021)", edited by Anahid Basiri, Georg Gartner and Haosheng Huang, LBS 2021, 24-25 November 2021, Glasgow, UK/online.

https://doi.org/10.34726/1759 |

1. Introduction

Trajectory segmentation, particularly stop-move detection, has been an essential step to construct semantic trajectories and extract meaningful activity sequences of moving objects [1]. Detecting stop and move segments accurately is critical to infer further semantics of segments—activities in stops and transportation modes of moves in the case of human subjects—, because errors occurring in stop-move detection can be propagated and amplified in later steps in trajectory data analysis.

Stop-move detection methods for GPS data often consist of two phases: stop-move detection, in which initial stop-move segments are detected; post-processing, in which the detected segments are merged or discarded upon criteria. Some stop-move detection algorithms inherently integrate such post-processing procedures, e.g., SMUoT – Zhao et al. [2]. Some other approaches conduct additional post-processing after applying an existing algorithm, e.g., Fillekes et al. [3]. As a part of trajectory segmentation, such post-processing can make an impact on the accuracy and characteristics of detected stops and moves. Although many stop-move detection algorithms exist and new methods are continuously proposed in the field, studies on comparing the performance of the stop-move detection methods are still scarce. Hence, this study evaluates the effect of post-processing in stop-move detection for GPS data. First, we select four existing stop-move detection algorithms—CandidateStops, SOC, POSMIT, and MBGP and analyze the post-processing effect in two input-data scenarios, (1) original GPS data and (2) sampled data, by quantifying the accuracy at two levels of spatiotemporal aggregation.

2. Data and Methods

2.1. Data Collection and Labelling

GPS and Accelerometer (ACC) datasets analysed in this study were collected from 161 participants for a 30-day period in the *Mobility, Activity and Social Interaction Study* (MOASIS) [4]. Stop-move detection is based solely on GPS data. The sampling rate of GPS data is 1 Hz and that of ACC is 50 Hz.

To evaluate the accuracy, labelled GPS data were constructed based on ACC and GPS data by manually identifying stops via the interactive visualization tool for data labelling task. Out of 161 participants, 38 participants with relatively more GPS points were sampled and 90 study days with more GPS data points (30 days for each of Tuesday, Thursday, and Sunday) were extracted, in order to construct balanced labelled data across weekdays and participants. The GPS points were labelled as a stop upon the following criteria:

velocity less than 1 m/s [5]; acceleration between -0.3 m/s^2 and 0.3 m/s^2 [5]; stop duration at least 10 minutes.

2.2. Stop-Move Detection Algorithms and Post-Processing

To assess the post-progressing effect on stops detected by a variety of stop-move detection algorithms, four stop-move detection algorithms were selected: CandidateStops [5], SOC (Sequence Oriented Clustering) [6], POSMIT (PrObability of Stops and Moves In Trajectories) [7], and MBGP (stop detection method by Montoliu, Blom and Gatica-Perez) [8]. On one hand, the selected algorithms meet the criteria of a good algorithm—maximum parsimony, ease of understanding, and high performance—to different degrees: lower (CandidateStops), moderate (MBGP), and highest (SOC; POSMIT) levels. SOC and POSMIT are expected to outperform CandidateStops and MBGP in stop detection. On the other hand, the selected algorithms implement different key measures and built-in post-processing procedures (Table 1).

Algorithm	**CandidateStops**	**SOC**	**POSMIT**	**MBGP**
Expected Usefulness	Low	High	High	Moderate
Key Measure	Density	Density	Probability	Density
Post-Processing	None	Merging stops	None	Merging stops

Table 1. Algorithmic Characteristics and Built-in Post-Processing of Selected Algorithms.

While CandidateStops and POSMIT do not have inherent post-processing in the algorithm, SOC and MBGP recognize stops too close in time and space as a single stop, using input parameters, i.e., maximum time gap and distance between subsequent GPS points. On top of the built-in post-processing of SOC and MBGP, we applied an additional post-processing procedure to detected stops by each of the four algorithms with the following rules:

- All moves shorter than 3 minutes were removed and classified as noise;
- Two stops are merged if the last GPS point of the preceding stop and the first GPS point of the following stop are within 150-meter distance and 1-hour time interval.

2.3. Evaluation Measures

The detected stops are assessed by two quantitative measures that quantify the accuracy at different levels of aggregation in space and time: (1) accuracy based on individual data points (i.e., F-measure based on confusion matrix [9]) and (2) the shape of detected stops (i.e., shape compactness based on the area and longest axis of a stop [10]).

2.4. Input-Data Scenarios

To observe how the post-processing effect changes over different input datasets, four algorithms are evaluated in two input-data scenarios: (1) original GPS data and (2) sampled data. For the second scenario, the GPS points were sampled with the rate of 1/60 Hz (1 point every minute).

3. Results

3.1. Accuracy at a GPS Data Point Level

The stop classification accuracy is compared at an individual GPS point level for the stops detected with vs. without applying post-processing for each algorithm. Each plot is drawn for each input-data scenario (Figure 1).

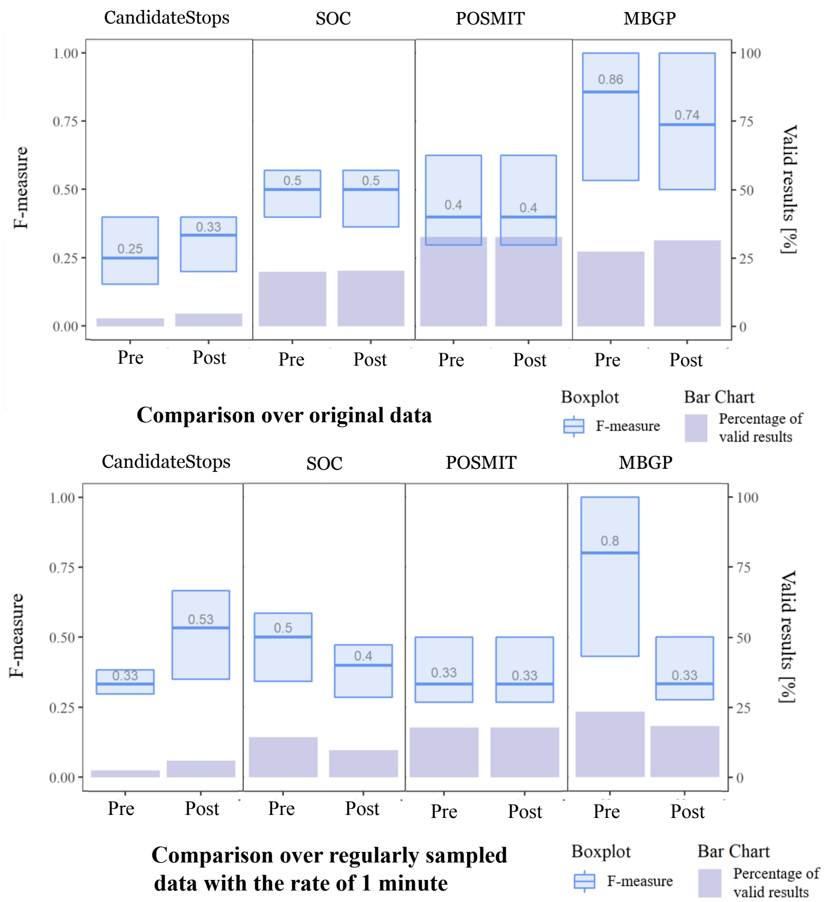

Figure 1. F-measure and percentage of valid results for stop detection accuracy measurement at the GPS point level with vs. without applying post-processing: Comparison over original data (upper) and comparison over regularly sampled data with the rate of 1 minute (lower).

3.2. Shape Compactness at an Individual Stop Level

The shape compactness of the detected stops was evaluated at an aggregated level of an individual stop with vs. without applying post-processing for each algorithm. Each table represents the results based on each input-data scenario (Table 2, Table 3). The larger shape compactness value indicates more circular shape.

Algorithm	Process	Min.	Q1	Median	Q3	Max.	Mean
CandidateStops	Pre	$2.46*10^{-17}$	$6.57*10^{-15}$	$1.91*10^{-15}$	$8.01*10^{-15}$	$5.88*10^{-11}$	$8.04*10^{-14}$
	Post	$3.82*10^{-17}$	$3.87*10^{-11}$	$9.30*10^{-09}$	$1.34*10^{-06}$	$3.38*10^{-03}$	$3.75*10^{-05}$
SOC	Pre	$8.82*10^{-10}$	$7.84*10^{-09}$	$2.55*10^{-08}$	$1.47*10^{-07}$	$2.57*10^{-03}$	$2.03*10^{-05}$
	Post	$8.82*10^{-10}$	$7.33*10^{-09}$	$2.54*10^{-08}$	$2.06*10^{-07}$	$2.57*10^{-03}$	$2.17*10^{-05}$
POSMIT	Pre	$1.39*10^{-08}$	$4.43*10^{-06}$	$4.09*10^{-04}$	$1.72*10^{-02}$	$1.13*10^{01}$	$7.48*10^{-01}$
	Post	$1.39*10^{-08}$	$4.43*10^{-06}$	$4.09*10^{-04}$	$1.72*10^{-02}$	$1.13*10^{01}$	$7.48*10^{-01}$
MBGP	Pre	$2.76*10^{-11}$	$6.55*10^{-08}$	$1.23*10^{-07}$	$2.11*10^{-07}$	$5.35*10^{-06}$	$1.77*10^{-07}$
	Post	$2.76*10^{-11}$	$7.29*10^{-08}$	$1.46*10^{-07}$	$3.18*10^{-07}$	$8.62*10^{00}$	$9.30*10^{-02}$
Labelled data		$1.99*10^{-11}$	$6.72*10^{-10}$	$3.08*10^{-09}$	$7.79*10^{-09}$	$8.71*10^{-07}$	$6.97*10^{-08}$

Table 2. Shape compactness of detected stops based on original data (Scenario 1).

Algorithm	Process	Min.	Q1	Median	Q3	Max.	Mean
CandidateStops	Pre	-	-	-	-	-	-
	Post	$4.44*10^{-13}$	$2.06*10^{-12}$	$1.08*10^{-11}$	$4.63*10^{-05}$	$1.12*10^{-01}$	$9.78*10^{-03}$
SOC	Pre	$5.54*10^{-11}$	$2.38*10^{-09}$	$9.36*10^{-09}$	$3.12*10^{-08}$	$1.39*10^{-04}$	$1.22*10^{-06}$
	Post	$3.07*10^{-10}$	$6.17*10^{-09}$	$5.39*10^{-08}$	$3.55*10^{-04}$	$5.10*10^{00}$	$1.31*10^{-01}$
POSMIT	Pre	$6.26*10^{-09}$	$4.10*10^{-06}$	$3.27*10^{-04}$	$1.47*10^{-02}$	$1.12*10^{01}$	$6.05*10^{-01}$
	Post	$6.26*10^{-09}$	$4.10*10^{-06}$	$3.27*10^{-04}$	$1.47*10^{-02}$	$1.12*10^{01}$	$6.05*10^{-01}$
MBGP	Pre	$9.38*10^{-11}$	$1.11*10^{-08}$	$5.06*10^{-08}$	$1.02*10^{-07}$	$8.26*10^{-07}$	$8.02*10^{-08}$
	Post	$9.01*10^{-10}$	$2.53*10^{-07}$	$2.63*10^{-05}$	$3.56*10^{-03}$	$1.06*10^{01}$	$4.23*10^{-01}$
Labelled data		$1.99*10^{-11}$	$6.72*10^{-10}$	$3.08*10^{-09}$	$7.79*10^{-09}$	$8.71*10^{-07}$	$6.97*10^{-08}$

Table 3. Shape compactness of detected stops based on regularly sampled data with the rate of 1 minute (Scenario 2).

4. Discussion

The evaluation at two aggregation levels implies that the impact of post-processing in stop-move detection varies by base stop-move detection algorithms as well as input data traits. In summary, the major findings are:

- At a data point level, the accuracy of stop-move detection without post-processing is the lowest with CandidateStops and the highest with MBGP, although SOC and POSMIT were expected to perform the best.

- At a data point level, post-processing largely improves the accuracy for CandidateStops, but makes little impact on the outputs of POSMIT, and worsens the accuracy for MBGP and SOC, especially for the sparse input data with low sampling rates.
- At an individual stop level, post-processing tends to change detected stops into more circular-shaped ones, with the highest impact on the outputs of CandidateStops and weaker impacts on the results of SOC and POSMIT. The shape compactness of the detected stops from SOC and MBGP without post-processing is similar to that of the labelled data.

References

[1] Y. Zheng, "Trajectory Data Mining: An Overview," *ACM Trans. Intell. Syst. Technol.*, vol. 6, no. 3, pp. 1–41, 2015, doi: 10.1145/2743025.

[2] Z. Zhao, L. Yin, S.-L. Shaw, Z. Fang, X. Yang, and F. Zhang, "Identifying stops from mobile phone location data by introducing uncertain segments," *Trans. GIS*, vol. 22, no. 4, pp. 958–974, Aug. 2018, doi: 10.1111/tgis.12332.

[3] M. P. Fillekes, E. Giannouli, E.-K. Kim, W. Zijlstra, and R. Weibel, "Towards a comprehensive set of GPS-based indicators reflecting the multidimensional nature of daily mobility for applications in health and aging research," *Int. J. Health Geogr.*, vol. 18, no. 1, p. 17, Dec. 2019, doi: 10.1186/s12942-019-0181-0.

[4] C. Roecke, M. Katana, M. Fillekes, M. Martin, and R. Weibel, "Mobility, physical activity and social interactions in the daily lives of healthy older adults: the MOASIS project," *Innov. Aging*, vol. 2, no. suppl_1, pp. 274–274, Nov. 2018, doi: 10.1093/geroni/igy023.1014.

[5] T. P. Nogueira, R. B. Braga, and H. Martin, "An ontology-based approach to represent trajectory characteristics," *Proc. - 5th Int. Conf. Comput. Geospatial Res. Appl. COM.Geo 2014*, pp. 102–107, 2014, doi: 10.1109/COM.Geo.2014.22.

[6] L. Xiang, M. Gao, and T. Wu, "Extracting stops from noisy trajectories: A sequence oriented clustering approach," *ISPRS Int. J. Geo-Information*, vol. 5, no. 29, pp. 1–18, 2016, doi: 10.3390/ijgi5030029.

[7] L. Bermingham and I. Lee, "A probabilistic stop and move classifier for noisy GPS trajectories," *Data Min. Knowl. Discov.*, vol. 32, no. 6, pp. 1634–1662, 2018, doi: 10.1007/s10618-018-0568-8.

[8] R. Montoliu, J. Blom, and D. Gatica-Perez, "Discovering places of interest in everyday life from smartphone data," *Multimed. Tools Appl.*, vol. 62, no. 1, pp. 179–207, 2013, doi: 10.1007/s11042-011-0982-z.

[9] B. Lamiroy and T. Sun, "Computing precision and recall with missing or uncertain ground truth," *9th Int. Work. GREC 2011, Seoul, Korea, Sept. 15-16, 2011, Revis. Sel. Pap.*, vol. Springer, no. 7423, pp. 149–162, 2013, doi: 10.1007/978-3-642-36824-0_15.

[10] A. M. Maceachren, "Compactness of Geographic Shape: Comparison and Evaluation of Measures," *Geogr. Ann. Ser. B, Hum. Geogr.*, vol. 67, no. 1, pp. 53–67, Apr. 1985, doi: 10.1080/04353684.1985.11879515.