Variability of people's activity space using GPS-based trajectory data

Giovanny Moncayo-Unda[1,2]
Marc Van Droogenbroeck[3]
Ismaïl Saadi[4]
Mario Cools[2,5,6]

Abstract: *The increasing popularity, usage, and ubiquity of mobile devices, as their easy access to global positioning systems (GPS), has promoted new research approaches in the transportation field. The easy way to collect data allows to extract and classify episodes of people's daily activity based on their location data, without additional information. This study presents a spatial activity analysis of a group of people from Quito, most of the Central University of Ecuador. The spatial variability analysis of the trajectories generated by GPS tracking will allow us to identify common daily, weekly, or monthly travel-behavior patterns. The method, developed in Python and R, gives all potential for initial data processing related to trajectory and spatial information from participants, respectively, to be analyzed later with demographic data.*

Keywords: Activity space variability, GPS, scikit-mobility, Python, ANOVA, R

1. Introduction

The daily mobile devices' and smartphones' use allows collecting mobility data based on global positioning systems (GPS) safely, efficiently, and inexpensively (Korpilo et al., 2017). Numerous approaches have been developed to research and understand the retrieval, detection, and classification of activities (Usyukov, 2017), trip information (Gong et al., 2014), and travel mode based on the location history of the travelers (Wu et al., 2016). Many of these methodologies used mobile devices for this purpose (Ferrer & Ruiz, 2014).

At first, the researchers studied the spatial variability of people from a political point of view and considering some analytical advantages of this analysis (Jones & Clarke, 1988). Then, due to the constant growth of the world population, the analysis focused on the times people use to traverse trajectories in general (van Wee et al., 2006) or spend on public transport (Durán-Hormazábal & Tirachini, 2016). Similarly, research related to the analysis of inter- and intra-individual variability in the same trajectory (Gadermann & Zumbo, 2007), individual travel behaviour (Gallotti, et al., 2015) or patterns related to travel between home and work (Zhou et al., 2021) became of special interest.

In this paper, we present a conceptual data framework that analyses GPS traces collected from mobile devices. Using the scikit-mobility package (Pappalardo et al., 2019) from python and the aspace package (Buliung & Remmel, 2008) from R, we analyze the variability of the activity spaces among people living in Quito, Ecuador, and relate the variability with socio-demographic information.

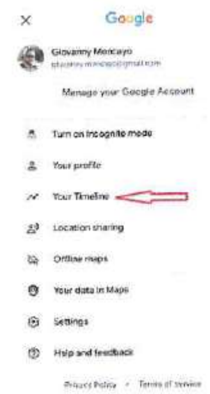
[1] Central University of Ecuador, Faculty of Engineering and Applied Sciences, Ecuador
[2] University of Liège, Local Environment Management & Analysis (LEMA), UEE, Belgium
[3] University of Liège, Montefiore Institute of Electrical Engineering and Computer Science, Belgium
[4] Université Gustave Eiffel, IFSTTAR, COSYS-GRETTIA, Marne-la-Vallée, France
[5] KULeuven Campus Brussels, Department of Information Management, Simulation and Modeling, Belgium
[6] Hasselt University, Faculty of Business Economics, Belgium

2. Methodology

2.1. Data Collection

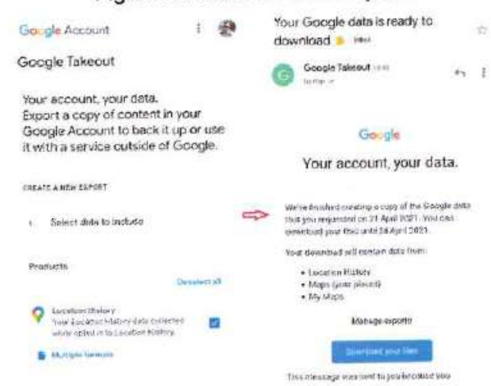
One hundred seventy voluntary people of Quito-Ecuador (82% university students, 32% females, mean age between 20 and 29 years), mostly from the Central University of Ecuador, collected GPS data using the Google Maps application at the beginning of 2020. All participants checked if the **Timeline** option in Google Maps app was activated or each participant activated it on their mobile device after receiving a very brief explanation of how to do it and the risks/benefits that it involves. The mentioned option is shown in Figure 1.

Figure 1 : Timeline activation



After at least three weeks of data collection, each participant requested a copy of their data from Google and proceeded to submit it via email for this research. The timeline data can be requested only by the owner of the Gmail account. See Figure 2.

Figure 2 : Timeline data request



2.2. Dataset Definition

The dataset construction consists of **2 parts:**

The first, developed in Python, consists of decrypting the location history file for each participant (about 2.5 million GPS location records (latitude, longitude)), and data cleaning and filtering using the scikit-mobility package to identify the points of interest within each day, trajectory and trip.

The second part, developed in R, uses the aspace package to estimate centrographic statistics and computational geometries for the spatial point patterns like the standard deviation ellipse (SDE), which characterizes the dispersion of point observations along two orthogonal axes. We calculate the SDE for each participant based on all the identified points within a week.

The resulting dataset contains the indicators presented in Table 1:

Field	Type	Description
id	Factor	Identifier of each participant
Y	Factor	Year when data was collected
M	Factor	Month when data was collected
W	Factor	Week when data was collected
CENTRE.x	Number	X-coordinate of the center (latitude)
CENTRE.x	Number	Y-coordinate of the center (longitude)
Area.sde	Number	Area of the SDE

Table 1 : Dataset

3. Results

3.1. Variability analysis

As a preliminary result for this paper, we conduct a within-subjects analysis of the SDE area variability for each week. For the data from each individual, we use one-way ANOVA analysis. We calculate whether or not the area of the SDE is the same across the different weeks of data that were available for that person. The ANOVA resuls show that for 56% of the participants, no statistically significant differences exist between their activity space (in terms of area) ($p>0.05$), and correspondingly for 44% of the participants, the inter-weekly differences in the area are significant.

3.2. Logit regression model

With the previous classification analysis (participants with significant differences in the area of their weekly activity space were coded as 0, participants that had the same activity space in size coded as 1), a binary logistic regression model was fitted to determine the role of personal characteristics of the participants. The following factors were considered: professional status (university student or no), age, gender, residential location, vehicle possession, and habitual use of public transport.

The mathematical representation of the logit model is:

$$\log\left(\frac{y}{y-1}\right) = \beta_0 + \beta_1.x_1 + \beta_2.x_2 + \cdots + \beta_n.x_n \qquad (1)$$

Where:

y: *Dependent variable (ANOVA output)*

x_i: *Independent variables (Demographic inputs)*; $i = 1,2,\ldots,n$

β_i: *Estimates for the model (Coeficients)*; $i = 1,2,\ldots,n$

After running the model described in (1), Table 2 shows the regression's results (AIC = 122.1).

Coefficients	Estimate	p-value
University Student (Yes)	0.28490176	0.70174545
Age	0.10408447	0.15345720
Gender (Male)	-0.07503815	0.88900808
Residence Location (North)	1.07635405	0.13253509
Residence Location (South)	0.22955541	0.72565125
Residence Location (Valley)	2.60733205	**0.02691149**
Own Vehicle (Yes)	-1.13111965	0.19262269
Usual Transport Pattern (Public Transportation & others)	-0.66407078	0.53529439

Table 2: Logit regression model results

The results show that whether or not the size of the activity space differs from one week to another does not depend on the socio-demographic factors, nor vehicle possession or public transport use. The only factor that had a slight influence is the residential location, where persons residing in the Valley area had a higher probability of the same activity space than persons residing in the central area of Quito.

4. Discussion

Google Maps application has become a good ally for mobility research, due to its constant evolution, accurate is increasing in location data collection, and other research or marketing fields can use it. Accuracy in data collection is good when the device's GPS is active, and obviously, this accuracy is the best when the user uses the application to move from one place to another; however, this decreases the mobile device's battery power rapidly. Another factor in better accuracy of data collection is a permanent internet connection of the mobile device. Finally, as is well known, Google maintains strict confidentiality agreements for handling personal information, including the location data of each user, which is a big problem when researching with people's location data.

On the other hand, the ANOVA indicates that for 56% of the participants, no statistically significant differences in the trajectory area per week. We also show through the Logit model that socio-demographic factors of the individuals do not have a strong influence on spatial variability. It is important to consider increasing the amount of location data for the research; the model can consider other fields generated in SDE analysis. This knowledge could help governments plan and manage urban mobility and traffic in cities, such as in confinement times of cities.

5. Conclusion

In this paper, ANOVA and binomial logit regression demonstrate that the size of the variability spaces is not significantly different between weeks for the majority (56%) of participants and that socio-demographic information cannot be used for classifying activity spaces variability. Further research will focus on other parameters like the spatial overlap of the activity spaces to determine the stability of activity spaces.

Bibliography

Buliung, R. N., & Remmel, T. K., *Open source, spatial analysis, and activity-travel behaviour research: Capabilities of the aspace package.* **Journal of Geographical Systems**, *10*(2), 191–216, 2008
Durán-Hormazábal, E., & Tirachini, A., *Estimation of travel time variability for cars, buses, metro and door-to-door public transport trips in Santiago, Chile.* **Research in Transportation Economics**, *59*, 26–39, 2016
Ferrer, S., & Ruiz, T. *Mobile Technologies for Activity-Travel Data Collection and Analysis.* **Mobile Technologies for Activity-Travel Data Collection and Analysis, i**, 171–186, 2014
Gadermann, A. M., & Zumbo, B. D., *Investigating the intra-individual variability and trajectories of subjective well-being.* **Social Indicators Research**, *81*(1), 1–33, 2007
Gallotti, R., Bazzani, A., & Rambaldi, S., *Understanding the variability of daily travel-time expenditures using GPS trajectory data.* **EPJ Data Science**, *4*(1), 18, 2015
Gong, L., Morikawa, T., Yamamoto, T., & Sato, H., *Deriving Personal Trip Data from GPS Data: A Literature Review on the Existing Methodologies,* **Procedia-Social and Behavioral Sciences**, *138*, 557–565, 2014
Jones, P., & Clarke, M., *The significance and measurement of variability in travel behaviour.* **Transportation**, 15, 1988
Korpilo, S., Virtanen, T., & Lehvävirta, S., *Smartphone GPS tracking—Inexpensive and efficient data collection on recreational movement.* **Landscape and Urban Planning**, *157*, 608–617, 2017
Pappalardo, L., Simini, F., Barlacchi, G., & Pellungrini, R., *Scikit-mobility: a Python library for the analysis, generation and risk assessment of mobility data*, 2019
Usyukov, V., *Methodology for identifying activities from GPS data streams.* **Procedia Computer Science**, *109*(2016), 10–17, 2017
van Wee, B., Rietveld, P., & Meurs, H., *Is average daily travel time expenditure constant? In search of explanations for an increase in average travel time.* **Journal of Transport Geography**, *14*(2), 109–122, 2006
Wu, L., Yang, B., & Jing, P. *Travel mode detection based on GPS raw data collected by smartphones: A systematic review of the existing methodologies. Information (Switzerland).* **MDPI AG.**, 2016
Zhou, Y., Thill, J. C., Xu, Y., & Fang, Z., *Variability in individual home-work activity patterns.* **Journal of Transport Geography**, *90*, 102901, 2021