Inferring Commuting Routes and Transportation Modes from Call Detail Records

Joel Pires[1] , Aldina Piedade[1], Marco Veloso[1], Santi Phithakkitnukoon[2,*], Zbigniew Smoreda[3], Carlos Bento[1,*]

[1]Center for Informatics and Systems of the University of Coimbra (CISUC), Coimbra, Portugal
[2] Dept. of Computer Engineering, Faculty of Engineering, Chiang Mai University, Thailand
[3]Sociology and Economics of Networks and Services Department, Orange Labs, Issy-les-Moulineaux, France

Abstract. We can infer mobility patterns from opportunistic data like Call Detail Records. However, considering challenges like low spatial resolution and temporal sparsity of the calls, it is appropriate to subsample and prepare rigorously the dataset. We then propose an optimized methodology to infer commuting patterns like the origin/destination trips and the respective travel modes adopted. We examined the commuting patterns of 1500 users, along fourteen months of communications, in three different Portuguese cities – Lisbon, Porto and Coimbra. Coimbra is a small city where the private car remains the main travel mode choice while the other two cities are larger and have multiple options of travel modes. The results obtained were posteriorly validated by two Portuguese censuses.

Keywords: Call Detail Records, Commuting Routes, Transportation Modes.

1 Methodology

This study assumes to be a continuation of the work of Pires et al. [1] and constitutes an effort in using Call Detail Records (CDRs) to infer commuting patterns like travel times, travel mode choices and commuting routes. It makes use of a CDR dataset that comprises 14 months of communications (from 2nd April of 2006 to 30th June of 2007) across 308 portuguese municipalities – a total of 435 701 911 records of approximately 18% of the Portuguese population. Each one of these records have the following fields: caller's and callee's IDs, location of caller's cellular tower, location of callee's cellular tower, duration and timestamp of the call. The dataset is very extensive and a process of subsampling was needed in order to shrink it without detrimenting its representativeness. For that we implemented the selection methodology of users introduced by Pires et al. [1]. From the remaining users of the selection process, we still had an excessive number of users considering the intensiveness of computational power needed for the inference of commuting patterns. Therefore, we focused on three Portuguese cities - Lisbon, Porto and Coimbra – and selected the 500 users with the highest average number of calls per day (a total sample of 1500 users). The Google Directions API has an algorithm that offers directions of movement given the geocoordinates of an origin and

* Corresponding author

a destination and given a preferable travel mode. Origin and destination, in this case, are home and workplace, previously inferred using a validated methodology [2]. For every user and every commuting type of trip (home to workplace or the reverse direction) we made API calls for the following travel modes: driving, walking, bicycling, bus, train, subway or tram. After an interpolation of the route points of each given route, we implemented a technique that assigns to each possible route a score. This score is inversely proportional to the probability of the route itself being the correct one. It is based on the multiplication of a *time penalty* by a *distance score*. The first variable is a penalty that assesses how bigger is a previously estimated commuting travel time than the duration of the Google API possible route. The second variable assesses how distant is the Google API possible route to the users' *intermediate towers*. Intermediate towers are cellular towers that were activated during the period between the timestamps of the pair of calls from which we inferred the minimal commuting travel time. So, the correct commuting route (and corresponding mode choice) is the one with the lower score.

2 Results

Obtained results, namely those related with the distributions of the percentages of the different travel modes on a city – like what can be seen in Figure 1 - were crossed with census data used as ground truth and the results were very promising. It was also possible to infer percentages of use of private, public, unimodal and multimodal modes, which is vital to take decisions to go in the direction of a more sustainable mobility in urban areas. Enables also public transport operators to understand the regions where the demand for transport is greater and, consequently, optimize the offered solutions.

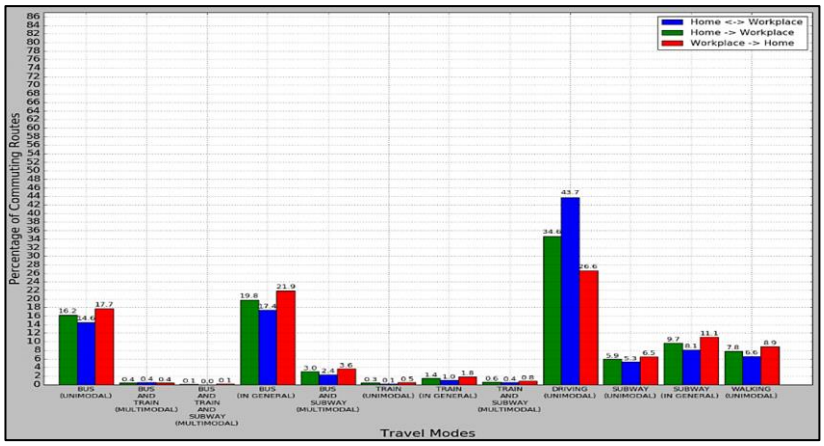

Fig. 1. Distribution of the percentages of the different travel modes in Lisbon

References

1. S. Phithakkitnukoon, Z. Smoreda, and P. Olivier, "Socio-geography of human mobility: A study using longitudinal mobile phone data," *PLoS One*, vol. 7, no. 6, pp. 1–9, 2012.
2. J. Pires, A. Piedade, S. Phithakkitnukoon, M. Veloso, and C. Bento, "How the Quality of Call Detail Records Influences the Detection of Commuting Patterns", 2019.