On the Quest for Representative Behavioral Datasets: Mobility and Content Demand

Guangshuo CHEN, Sahar HOTEIT, Aline C. Viana, Marco FIORE

{guangshuo.chen, sahar.hoteit, aline.viana}@inria.fr, marco.fiore@ieiit.cnr.it

1. Objectives

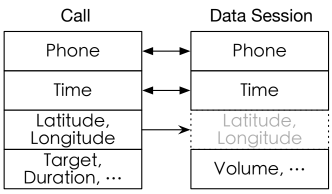
Understand correlations between human mobility and data demand.
Objectives:

- Extract fully-featured dataset from cellular traces;
- Construct mobility: identify Home locations, and estimate trajectories;
- Characterize user behaviors in terms of time, space and volume (data).

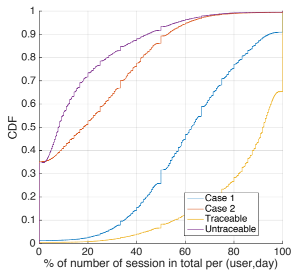
2. Dataset Description

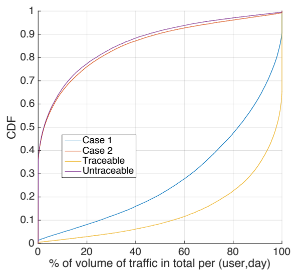
17,366 subscribers are extracted by applying a series of filters on cellular traces, consisting of 2,398,392 calls and 954,737 sessions in 4 weeks.

Session Location Estimation
Using call detail records, we infer users' locations for data sessions.

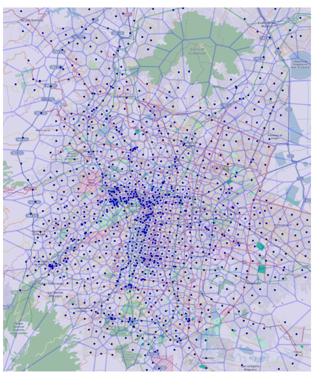


Based on trajectory estimation, at least 70% of sessions (80% of volume) are traceable for 80% of subscribers.





Decomposition of Cells in Mexico City

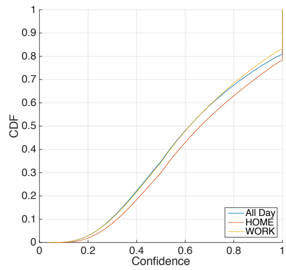


Cellular Cells (Voronoi)

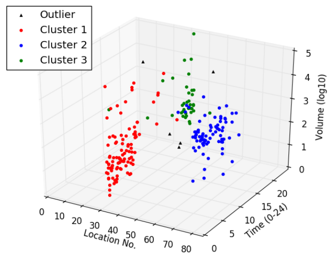
Population Density

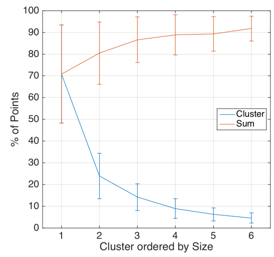
3. Analysis

Home Location Identification
For each user, the most visited cell between 10pm and 7am is identified as his/her *HOME*.



Clustering Sessions
Sessions (space, time, volume) are clustered by DBScan.

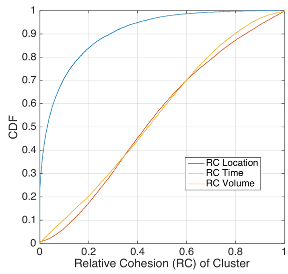


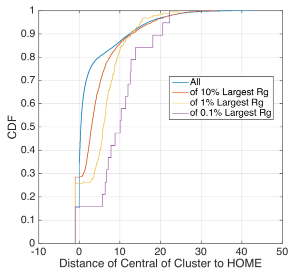


For each cluster, **Relative Cohesion** is calculated as a function of time, space or volume, respectively.

$$RC^{(*)} = \frac{\sum_{p \in C} dist^{(*)}(p, \mathbf{c})^2}{\sum_{p \in C} dist(p, \mathbf{c})^2} \quad (1)$$

$$RC^{(loc)} + RC^{(time)} + RC^{(vol)} = 1 \quad (2)$$





4. Future Works

- Estimate trajectories' incomplete information;
- Model volume demand;
- Characterize content demand;
- Link and predict mobility (important locations) and volume/content demand simultaneously.

This work was supported by the EU FP7 ERANET program under grant CHIST-ERA-2012 MACACO.