SMART
Singapore-MIT Alliance for Research and Technology

FUTURE URBAN MOBILITY

Comparing Home Detection Methods Through Multi-Source Geo-Social Data

Iva Bojic, Alexander Belyi, Yang Xu
Carlo Ratti

Senseable City Lab, SMART Center, Singapore
Senseable City Lab, Massachusetts Institute of Technology, Cambridge, MA, US

I. Motivation

The proliferation of big geo-social data has provided unprecedented opportunities to the understanding of human mobility patterns. A lot of related work is dedicated towards labeling/inferring **individual home location** from these data to facilitate a variety of research purposes.

However, there has been limited efforts on **comparing** these home detection methods, or **validating** the results using supplementary data. An improved understanding of the robustness of these home detection methods is not only important to human mobility research, but also to applications which rely on high-fidelity user location profiles.

II. Comparing Home Detection Methods

Two datasets are used to compare 5 different home detection methods:

Bank Card Transaction of Banco Bilbao Vizcaya Argentaria (**BBVA**) in Spain during 2011

Geo-tagged Photos/Videos from **Flickr** (from 2005 to 2014)

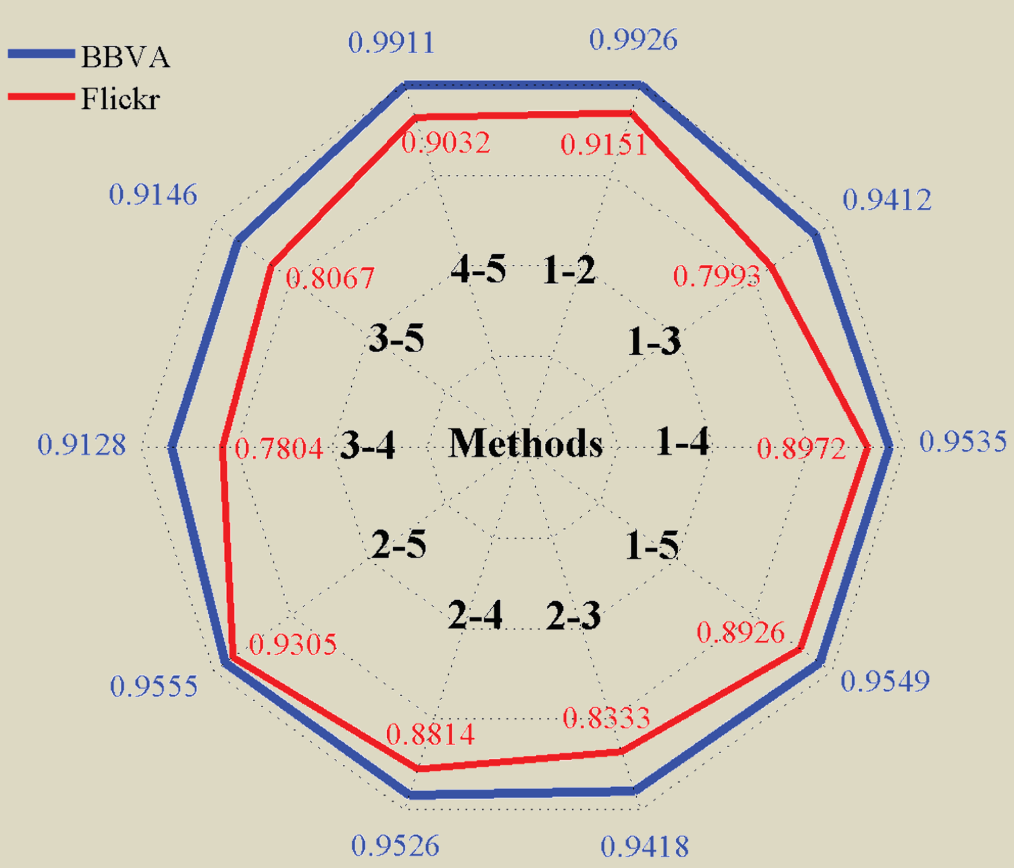
BBVA | flickr

1.2 Million | **141.7** Million Transactions | **2.5** k | **2.4** Million Photos/Videos

The 5 home detection methods estimate individual home location as:

1. place where a user/customer took/made the maximal number of photographs/videos/transactions.
2. place where a user/customer spend the maximal number of active days (i.e., days when a user made at least one activity).
3. place with the maximal timespan between the first and last phone/video/transaction.
4. place where a user/customer took/made the maximal number of photographs/videos/transactions from 7 PM to 7 AM.
5. place where a user/customer spend the maximal number of active days from 7 PM to 7 AM.

For each dataset, we compare each of the two methods by examining the predictions which produce the same home location estimate (scale from 0 to 1):



BBVA dataset appears to be more robust to different home definition methods as the results differ only from 1% to 9 % unlike in case of **Flickr** that are in range of 10%–20% [1].

Different datasets are not equally susceptible when applying different home detection methods. It is of vital importance to further validate these prediction results using "ground truth" data.

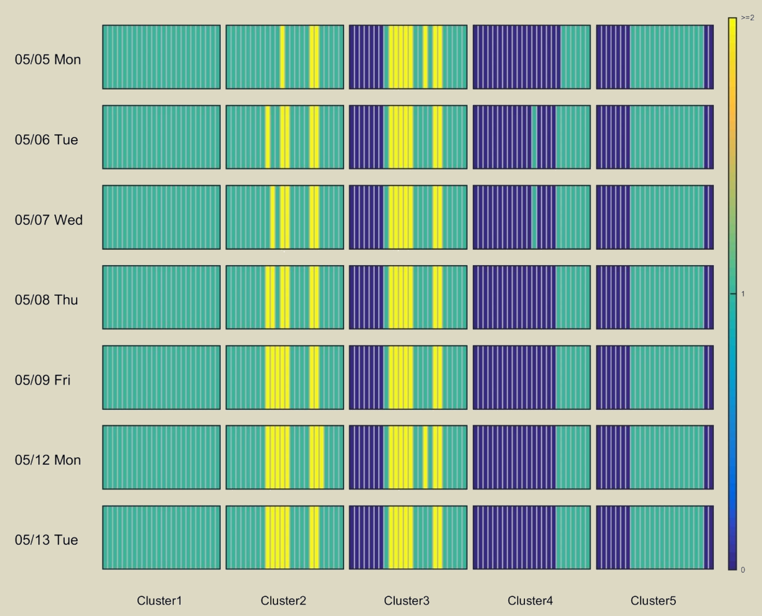
III. From BBVA/Flickr To Phone Data

Call detail records (CDRs), which are passively collected when phone users engage in communication activities (e.g., initiate or receive call/message), have been widely used to infer individual home locations.

Actively tracked mobile phone data [2] generally can offer a better spatiotemporal coverage of user activity than CDRs. This type of data can be used to derive "ground truth" of individual home locations, which can be further used to validate results produced by CDR-based home detection methods.

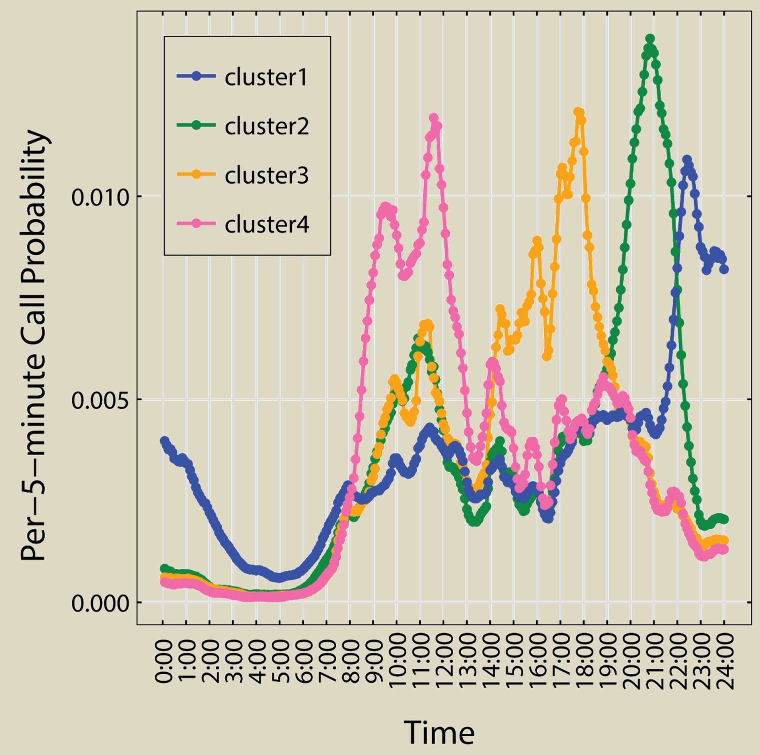
We use an actively tracked phone dataset collected in Tianjin, China to **derive "ground truth" of individual home location**. We then use another CDR dataset (Shenzhen, China) to derive the empirical temporal calling patterns of individuals. The calling probabilities are then used to generate **synthetic CDRs** for Tianjin's dataset (i.e., to mimic the calling patterns of people in Tianjin).

Actively Tracked Mobile Phone Data (Tianjin, China) → Cluster Users Based on Activity Level → User Activity Patterns



→ Derive Individual Home Location → **"Ground Truth" of Individual Home Location** → Validation

Call Detail Records (Shenzhen, China) → Cluster Users Based on Calling Patterns → Empirical Calling Probabilities



→ **Synthetic CDR Data for Users in Tianjin, China** → CDR-based Home Detection Methods → **Estimated Individual Home Location** → Validation

IV. Summary & Ongoing Work

1. The choice of home detection method should be done carefully with respect to the characteristics of the dataset being considered.
2. Regarding the phone-based study, we are currently deriving the "ground truth" and CDR-based home location estimates to perform the validation.

V. References

1. Bojic, Iva, et al. "**Choosing the right home location definition method for the given dataset.**" *International Conference on Social Informatics*. Springer International Publishing, 2015.
2. Xu, Yang, et al. "**Another Tale of Two Cities: Understanding Human Activity Space Using Actively Tracked Cellphone Location Data.**" Annals of the American Association of Geographers 106.2 (2016): 489-502.