bandicoot: a toolbox for mobile phone metadata

Yves-Alexandre de Montjoye, Alex Pentland, Luc Rocher, Julien Hendrickx
(Collaboration with MIT)

Abstract − bandicoot is an open-source Python toolbox to help data scientists analyze mobile phone metadata. With only a few lines of code, bandicoot loads your datasets, visualizes your data, performs analyses, and exports the results.

Every time we send or received a text or a phone call, our mobile phones generate metadata: who we call, at what time, for how long, and from where. Collected at large scale, they have been used to design transportation systems, planning disaster responses, and fight epidemics. [1] While the use of machine learning algorithms on mobile phone metadata has been evolving fast, it currently lacks the standardization needed to thrive. Numerous crucial implementation choices are often lost from one research paper to another making it hard to replicate results, to quantify the impact of new methods, and to transfer knowledge.

We have introduced *bandicoot*, an open-source Python toolbox, to solve these issues. bandicoot extracts more than 160 robust behavioral features from mobile phone metadata, and focuses on making it easy for researchers and practitioners to load metadata and compute robust features from them.

bandicoot indicators fall into three categories (see Figure 1):

1. **Individual indicators** (e.g. percent of nocturnal interactions, time it takes someone to answer text messages) describe an individual's phone usage.
2. **Spatial indicators** (e.g. entropy of visited antennas, radius of gyration) describe mobility patterns.
3. **Social network indicators** (e.g. clustering coefficient, assortativity) describe individuals' social network and compare their behaviors with those of their contacts.

Emphasis is put on correctness and consistency through numerous unit tests covering 91% of the source code, domain-specific code detecting incorrect entries, and reporting variables to assess data quality or potential data issues.
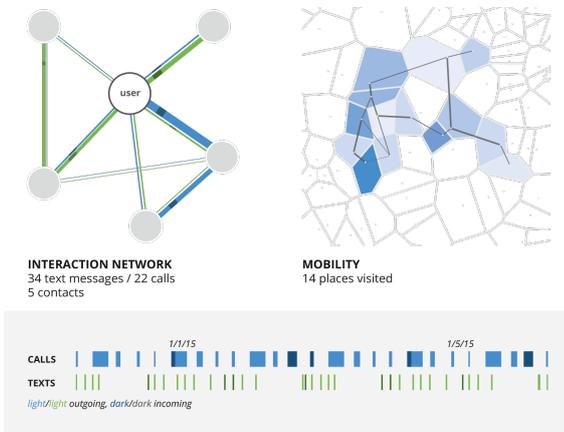

INTERACTION NETWORK
34 text messages / 22 calls
5 contacts

MOBILITY
14 places visited

CALLS   1/1/15   1/5/15

TEXTS

light/light outgoing, dark/dark incoming

Figure 1: bandicoot uses individual, spatial, and social features.

bandicoot is currently used at large scale experiments by carriers (e.g. Orange, Telenor), NGOs, and international organizations (World Bank).

bandicoot's behavioral indicators were, for instance, used to predict users' personality traits [3] - the Big Five indicators (neuroticism, extraversion, etc.) - resulting in accuracies up to 70% better than random. Similar methodology has recently been used to predict the gender of users in low and medium income countries

Figure. 2 (reproduced from [2]) shows that a training set of 10.000 people is enough to reach an accuracy ranging from 74,3% to 88,4%.
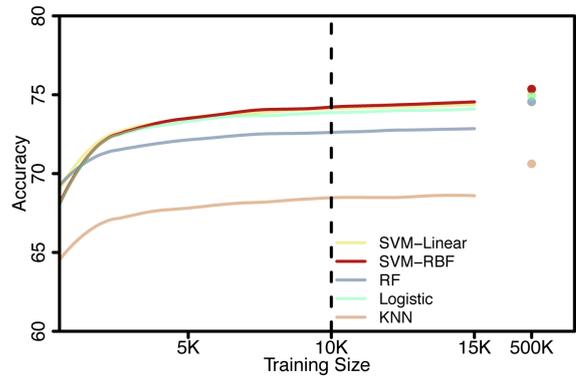

Figure 2: Gender prediction reaches 74,3% accuracy with a training size of 10,000 people.

During the "Data for Development" challenge, an international research challenge [4] using a massive anonymized dataset provided by telecommunication company Orange, bandicoot was used to address socio-economic development question in Ivory Coast. Best contributions explored disease containment for country-wise epidemics, social divisions, or optimization of public transportation.

References

[1] Blondel, V. D., Decuyper, A., Krings, G. (2015). "A survey of results on mobile phone datasets analysis". EPJ Data Science, 4(1), 1-55.

[2] Jahani, E., Sundsøy, P. R., Bjelland, J., Iqbal, A., Pentland, A., de Montjoye, Y. A. "Predicting Gender from Mobile Phone Metadata".

[3] de Montjoye, Y. A., Quoidbach, J., Robic, F., Pentland, A. S. (2013). "Predicting personality using novel mobile phone-based metrics". In Social computing, behavioral-cultural modeling and prediction (pp. 48-55). Springer Berlin Heidelberg.

[4] Sundsøy, P., Bjelland, J., Iqbal, A. M., de Montjoye, Y. A. (2014). "Big Data-Driven Marketing: How machine learning outperforms marketers' gut-feeling". In Social Computing, Behavioral-Cultural Modeling and Prediction (pp. 367-374). Springer International Publishing.

[5] de Montjoye, Y. A., Smoreda, Z., Trinquart, R., Ziemlicki, C., Blondel, V. D. (2014). "D4D-Senegal: the second mobile phone data for development challenge". arXiv preprint arXiv:1407.4885.