

The potential of mobile phone data for spatial and temporal observation of urban human mobility

Silvia Goldman Ber Kapel¹, Jorge Rady de Almeida Jr²

¹Engenharia de Computação – Escola Politécnica da Universidade de São Paulo (USP)
São Paulo, SP – Brazil

²Engenharia de Computação – Escola Politécnica da Universidade de São Paulo (USP)
São Paulo, SP – Brazil

sgbkapel@usp.br, jorgerady@usp.br

***Abstract.** The evolution of humanity is increasingly moving towards urban life; at the same time, mobile phones now occupy an extremely important place, supplying population data of unparalleled coverage. These data have provided unprecedented opportunities to understand human mobility in the urban environment, and are of great value to urban human mobility research. We here aim to provide an initial outline of the potential of aggregated mobile phone dataset analysis in the context of urban human mobility, the types of data sources and applications observed in some recent studies, discuss the strengths and weaknesses by putting together a set of recommendations, which may be useful for future works.*

1. Introduction

The evolution of humanity has been increasingly moving towards urban life. According to the United Nations, 68 percent of the world's population will live in urban areas by 2050 (Nations et al. 2018). In this scenario, the demand for smart city solutions is growing in response to current challenges in urban life, and urban human mobility is an important research area. One of the key drivers of smart city solutions, particularly in urban human mobility, is mobile phone data analysis.

Personal mobile phone devices occupy an extremely important place, with population data of unparalleled coverage, being of great value for urban human mobility research (Naboulsi et al. 2016)(Calabrese, Ferrari, and Blondel 2014). The effective handling of Big Data generated by these personal mobile devices requires special techniques and models, which will in fact add the capacity to gain new knowledge to achieve the desired results for both cities and their citizens. The challenges are diverse, from data privacy, data source scope definition, how to treat it, the different techniques for analysis and others (Calabrese et al. 2014). However, aggregated mobile phone data analysis usually has no privacy issues, is less complex and has a great potential.

We here aim to provide a summary overview of the potential of aggregated mobile phone dataset analysis in urban human mobility area, the types of data sources, techniques and applications observed in some recent studies, bringing together a set of best practices that can be useful for future work. The paper is organized as follows. Section 2 revises the main works in the study of urban human mobility data mining and explores several applications. Section 3 presents an overview of the urban mobility data

analytics process and explores the mobile phone users' datasets. Section 4 presents the study in which we discuss some case references on how mobile phone data is applied to spatial and temporal observation of urban mobility, exploring the methods used and applicability. In Section 5 we show the main results of our study and discuss challenges and opportunities. Finally, Section 6 concludes the paper with our final remarks.

2. Related Work

Advances and convergence of information and communication technologies have revolutionized people's way of life (Naphade et al. 2011). Accordingly, the high and growing number of mobile subscribers makes mobile personal communication technology one of the most successful innovations of recent times. An increasing number of people depend entirely on their mobile devices, not only for work, but also for their personal life. Mobile subscribers today represent a large share of the world's population; their mobile phones are always interacting with the telecommunication network, generating georeferenced traffic and events.

In this scenario, the digital footprints generated by mobile phone users have rapidly emerged as a primary source of knowledge about human mobility (Huang, Cheng, and Weibel 2019), at a minimal cost, representing an important opportunity for knowledge extraction in the area of urban mobility. Understanding human mobility is key to many urban-related applications such as urban planning (De Nadai et al. 2016), demographics studies (Pappalardo et al. 2015), transportation (Huang et al. 2019), estimating migratory flows, crowd management (Celes, Boukerche, and Loureiro 2019), epidemic modeling and energy demand forecasting (Selvarajoo, Schlapfer, and Tan 2018).

Urban mobility solutions involve two large blocks, urban sensing and urban data analytics. Urban sensing allows obtaining data from those digital footprints, and urban analytics, understanding the city dynamics (Celes et al. 2019). There are several methods and techniques on how to collect, process and analyze all these data, explored in some researches such as (Calabrese et al. 2014)(Zhao et al. 2016)(Naboulsi et al. 2016)

3. Urban Sensing and Urban Data Analytics

The data-driven process for urban mobility solution can be summarized in five key steps, illustrated in Figure 1.

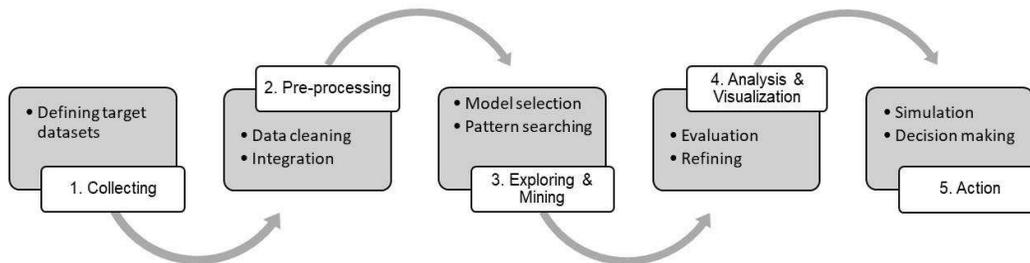


Figure 1. Urban Mobility Data-Driven Analytics Process

Initially, raw data from target data sources are collected and submitted to preprocessing for cleaning, summarization and integration. Subsequently, a model selection and pattern search support the exploring and mining step. Analysis and

visualization are essential to evaluate and to refine the model, gaining insights. The final step allows discovering knowledge from the collected data, assisting decision-making, planning or simulation.

3.1. Urban Mobility Datasets

There is a variety of mobile phone users' datasets containing rich knowledge about locations and mobility, helping address many urban challenges. The main mobile phone users datasets are cellular network records, social media records, proximity records and positioning records (Celes et al. 2019), each one having advantages and limitations.

In this context, cellular network records emerge as a valuable main data source largely used in urban mobility works, presenting several advantages, such as coverage of large areas and volume of users, and no additional costs or infrastructure for data collection (Celes et al. 2019). Cellular network records contain timestamped and geo-referenced logs on each voice call, texting and Internet activity of every serviced customer. There are two types of data records (Calabrese et al. 2014) :

1) Call Detail Records (CDRs): A CDR contains the details of a phone call or SMS. In general, a CDR consists of origin and destination phone numbers, a timestamp, the duration (of calls), the communication type (call or SMS), ID of the base station or cell tower (BTS/cell) involved in the call.

2) Internet Protocol Detail Records (IPDRs): An IPDR contains details of Internet usage. This typically consists of mobile phone ID, timestamp, number of bytes transferred, the website visited, and ID of the BTS/cell the phone connects to.

4. Case Study

In this section we show some case references regarding how aggregated mobile phone data, cellular network records, is applied to spatial and temporal observation of urban human mobility, exploring the data sources, techniques and methods used. As our interest is discussing the potentialities of mobile phone data in urban human mobility applications, in our case references exploration we focus on data sources and results achieved. We also observe their reported challenges.

We selected four references in three different urban-related application areas: Urban planning (Ríos and Muñoz 2017)(De Nadai et al. 2016), Energy forecasting (Selvarajoo et al. 2018) and Crowd Management (Celes et al. 2019). (De Nadai et al. 2016) explore mobile phone data to extract information about human activity, and to combine such data with land use and socio-demographic information to test the four conditions that promote "life" in a city according Jacobs theories, in four great Italian cities. (Ríos and Muñoz 2017) analyze mobile phone call records to detect land use patterns over urban areas using latent semantic topic models in Santiago, Chile. (Selvarajoo et al. 2018) propose a novel electrical load forecasting method using mobile phone data. The cell phone records are used to map the time-varying population distribution in the Trentino region, Italy. (Celes et al. 2019) show the benefits of using mobile phone call records combined with social media and Point of Interest (PoI) datasets to detect the occurrence of crowd situations in Milan, Italy.

4.1. Data Sources

This subsection explores the data sources used by each case reference. Table 1 shows the data source summary, including mobile phone data characteristics and additional data sources used. In Table 1, we can see that, in general, additional data sources are required to perform the analysis.

Table 1. Data Source description

Main Data Source (mobile phone data)						Additional Data Sources	
Mobile Network Data Source	Type		CDR/IPDR				Granularity (minutes)
	Individual	Aggregated	Call	SMS	Internet		
Telecom Italia Mobile (TIM)		●			■	60	OpenStreetMap (OSM): detailed maps; Census data: people and buildings; Urban ATLAS: land use data; ISTAT: logistic facilities; Foursquare: visited venues
Major telco company in Chile	●		■			---	City maps
Telecom Italia Mobile (TIM)		●	■	■	■	10	Electric load data: line x ampere values with a time resolution of 10 minutes; Census data: census population information at the municipality level
Telecom Italia Mobile (TIM)		●	■	■	■	10	Social media data: geolocated Twitter data; PoI datasets

In terms of data volume, (De Nadai et al. 2016) use TIM customers and roaming customers mobile phone call records of the six Italian cities of Bologna, Florence, Milan, Palermo, Rome, and Turin, from February to October 2014. (Ríos and Muñoz 2017) use 880,000,000 calls by about 3 million customers collected in Santiago, Chile. Despite individual data collection, the dataset is aggregated into initial pre-processing. (Selvarajoo et al. 2018) use aggregated data of 6,575 grid cells covering the Trentino region, Italy. (Celes et al. 2019) use aggregated data of 1,000 grid cell covering the city of Milan, Italy.

4.2. Processing and Analysis

This subsection presents an overview of main processing and analysis methods used, summarized in Table 2.

Table 2. Processing and Analysis overview

Paper Reference	Mobile phone data records Processing			Analysis
	Individual - Summarization	Filtering	Aggregated - Summarization	
(De Nadai et al., 2016)			●	Regression model - Ordinary Least Squares (OLS)
(Ríos & Muñoz, 2017)	●			Generative statistical method adapted - Latent Dirichlet Allocation (LDA)
(Selvarajoo et al., 2018)			●	Linear regression model on log-transformed data; ARIMA model for forecasting
(Celes et al., 2019)			●	Time series model; Anomaly detection method - Seasonal Hybrid ESD (S-H-ESD) Semantic analysis - NLP2 named-entity recognition method

5. Results and Discussion

This section discusses the results from the case references. Table 3 summarizes the strengths and weaknesses of the works studied, mainly regarding the following aspects: data types used; if individual records are requested to perform the analysis; whether data over a long period (at least multiple months) are needed; if data privacy is an issue; whether special pre-processing techniques are requested (e.g., big data volume processing); if noise data could affect the results (e.g., data collection from users who are just crossing a few areas and not belonging to the desired dataset); and whether fined location precision is requested.

Table 3. Strengths and weaknesses summary

Urban human mobility application	Mobile phone data characteristics					Individual records needed ?	Long period needed ?	Privacy issues ?	Pre-processing issues ?	Noise issues ?	Fined location precision needed ?	
	Individual	Aggregated	Call	SMS	Internet Granularity (minutes)							
Population activity		✓			✓	60	No	Recommended	No	No	No	No
Land functional use	✓	✓	✓				No	Recommended	Yes	Yes	No	No
Energy demand forecasting		✓	✓	✓	✓	10	No	Recommended	No	No	Yes	No
Crowd detection		✓	✓	✓	✓	10	No	No	No	No	No	Yes

As can be seen from Table 3, all the case references work well with aggregated mobile phone data; no individual data is required. Furthermore, the aggregated data of mobile phones do not have privacy issues and there are no pre-processing issues regarding large volumes of data. Apart from crowd detection, for the other three cases, analyzing long period datasets is recommended to avoid seasonal biases and, in addition, it is also important to detect special holidays or vacation periods included in the dataset. Sometimes we may have noise data in our dataset, which can be discovered early in the process or only in advanced steps. In this case, specific techniques could be required to filter these data. In general, fined location precision is not required except for crowd detection, in which any improvement in location accuracy is relevant for analysis. In this situation, one option is to add geolocated social media records to data sources (Celes et al. 2019).

In addition, in some geographies, the human behavior of mobile phone use is changing; a larger number of people talk less and use more data. Focusing on Internet activity is a good option as it allows the passive reconstruction of people's mobility: even in the absence of direct user activity (e.g., making / receiving a call, receiving / sending an SMS), mobile phones can be tracked as they will likely be connected to the Internet for background traffic and push notifications (De Nadai et al. 2016).

Finally, mobile phone data is an important data source in the area of urban human mobility research. The initial step of defining the most appropriate data sources for the study to be undertaken is a key success point in gaining the desired knowledge. Due to the small number of studies evaluated, this work provides only an initial outline of recommendations on which characteristics of the data sources perform best.

6. Conclusion

We presented an initial overview into the potential of aggregated mobile phone data analysis for spatial and temporal observation of urban human mobility. We summarized

some recent work on reference tables that provide detailed data source information as well as the methods and techniques used. We also provide an overview of key challenges and propose some recommendations for future studies. This is a work in progress and can be complemented by exploring other recent works in this field, as well as works on various types of urban mobility applications.

References

- Calabrese, Francesco, Laura Ferrari, and Vincent D. Blondel. 2014. "Urban Sensing Using Mobile Phone Network Data: A Survey of Research." *ACM Computing Surveys* 47(2):1–20.
- Celes, Clayson, Azzedine Boukerche, and Antonio A. F. Loureiro. 2019. "Crowd Management: A New Challenge for Urban Big Data Analytics." *IEEE Communications Magazine* 57(4):20–25.
- Huang, Haosheng, Yi Cheng, and Robert Weibel. 2019. "Transport Mode Detection Based on Mobile Phone Network Data: A Systematic Review." *Transportation Research Part C: Emerging Technologies* 101:297–312.
- Naboulsi, Diala, Marco Fiore, Stephane Ribot, and Razvan Stanica. 2016. "Large-Scale Mobile Traffic Analysis: A Survey." *IEEE Communications Surveys & Tutorials* 18(1):124–61.
- De Nadai, Marco, Jacopo Staiano, Roberto Larcher, Nicu Sebe, Daniele Quercia, and Bruno Lepri. 2016. "The Death and Life of Great Italian Cities." Pp. 413–23 in *Proceedings of the 25th International Conference on World Wide Web - WWW '16*. New York, New York, USA: ACM Press.
- Naphade, Milind, Guruduth Banavar, Colin Harrison, Jurij Paraszczak, and Robert Morris. 2011. "Smarter Cities and Their Innovation Challenges." *Computer* 44(6):32–39.
- Nations, United, Department of Economic, Social Affairs, and Population Division. 2018. *World Urbanization Prospects The 2018 Revision*.
- Pappalardo, Luca, Dino Pedreschi, Zbigniew Smoreda, and Fosca Giannotti. 2015. "Using Big Data to Study the Link between Human Mobility and Socio-Economic Development." Pp. 871–78 in *2015 IEEE International Conference on Big Data (Big Data)*. IEEE.
- Ríos, Sebastián A. and Ricardo Muñoz. 2017. "Land Use Detection with Cell Phone Data Using Topic Models: Case Santiago, Chile." *Computers, Environment and Urban Systems* 61:39–48.
- Selvarajoo, Stefan, Markus Schlapfer, and Rui Tan. 2018. "Urban Electric Load Forecasting with Mobile Phone Location Data." Pp. 1–5 in *2018 Asian Conference on Energy, Power and Transportation Electrification (ACEPT)*. IEEE.
- Zhao, Kai, Sasu Tarkoma, Siyuan Liu, and Huy Vo. 2016. "Urban Human Mobility Data Mining: An Overview." Pp. 1911–20 in *2016 IEEE International Conference on Big Data (Big Data)*. IEEE.