Detecting Unstructured Events Through Multi-Source Data Analysis

Patrizia Sulis, Ed Manley

Centre for Advanced Spatial Analysis, University College London.

Introduction.

There is increasing interest in using urban data analysis for different applications, particularly in relation to urban management, transport optimisation and fields related to cities becoming 'smarter'. Data used in the analyses comes from different sources and involves various domains, from transportation engineering to urban planning.

Three forms of data in particular have been of interest within the research community for a number of years. The first one includes data sourced from what are called 'smart cards' (for ticketing purposes): mainly used for calculating commuting flows and tracking users trajectories (Munizaga and Palma 2012; Hasan et al. 2012), it also has some applications for urban zoning and activity detection (Roth et al. 2010; Zhong et al. 2014). The second type concerns mobile phone data, generally aggregated and anonymised by providers, used mainly for detecting urban dynamics and activities because of the fine granularity, considering mobile phone as a proxy for each single user. Research focus has been similar from pioneering works till recent ones (Ratti et al. 2006; Candia et al. 2008; Calabrese et al. 2011; Hasan et al. 2012). Finally, the third one involves social media data, which use increased in latest years mainly because of the users' growth and its availability through API. Spatial applications are particularly relevant when data attributes include location information, as seen in Twitter (Hawelka et al. 2014), or Foursquare (Cranshaw et al. 2012) data.

In this paper we will discuss how a deeper understanding of urban dynamics can be yielded through the joint analysis of all three datasets. Specifically we seek to develop a platform for the identification of events, with a view to informing urban planning and management.[1]

In: A. Comber, B. Bucher, S. Ivanovic (eds.): Proceedings of the 3rd AGILE Phd School, Champs sur Marne, France, 15-17-September-2015, published at http://ceur-ws.org

Problem origins and relevance.

Data mentioned above have strengths and limitations that must considered in the analysis and evaluation of results (table 1.1).

Table 1.1

type of data	strengths	limitations
smart card	• good spatial precision • use of device is widely diffused	• discontinuous data • potentially multiple users per card
mobile phone	• continuous data • use of device is widely diffused	• spatial precision varies depending on data structure and content, • uneven density of antennas and data • potentially multiple devices per user
social media	• easily accessible • spatially located • some content is context rich	• small percentage of geo-tagged data • few users, some demographic categories are under represented

Furthermore, current research using data analysis presents common limitations, regardless of the nature of data involved, that can be summarised in three main points.

1. With few exceptions (Calabrese et al. 2010; Traag et al. 2011; Pereira et al. 2013), current research focuses mainly on big scale, regular and routine flows and activity dynamics, such as those involving home-to-work commuting in the city. Furthermore, aside from some works using Twitter data, the interest is not specifically on anomalies and variations (event detection), rather on the flows in themselves. Therefore many urban phenomena are disregarded by this research field. Specifically, two kinds of phenomena are neglected: small and medium scale events, and unplanned, spontaneous events. They represent a majority happening constantly in the urban space, and one of the most difficult to catch and predict, either with traditional or alternative urbanism tools, because of their transient nature.

2. The majority of current research uses data collected from a single source. Few works use data sourced from different domains (Calabrese and Ratti 2006; Zhang 2014), usually in an attempt to validate results reached with one single-sourced database (Lenormand et al. 2014).

However, the potential of using multi-source datasets involves different aspects:

a) data collected from various technological networks may fill gaps (spatially and temporally), and overcome or reduce limitations introduced by single-source databases;

b) getting results from multiple data analyses opens to the possibility of cross validating results obtained from single-source data for the same space-time context, contributing to identify potential events more reliably.

3. Finally, it is worth to consider how very little research involving urban data analysis focuses on direct urban planning applications. Urban data analysis has strong potential for building models that describe human dynamics, and be relevant for scenario-building approach or planning evaluation. There are already some small, interesting attempts in current research (Calabrese et al. 2010; Ferrari et al. 2014): simulation models may be built to predict realistic evolutions of complex urban phenomena, starting from real-time data collected from a portion of the day.

Therefore, the use of urban data represent an unprecedented opportunity to acquire valuable information about urban phenomena, describe them from a quantitative perspective, and inform spatial planning processes in a successful way.

The main research questions at the basis of this research are the following:

1. Is it possible to detect small scale, unplanned and irregular urban phenomena through urban data analysis?
2. Can the combined use of multi-source data in the analysis improve the possibilities of event detections in the physical urban space?
3. How can a reliable quantitative description of urban phenomena improve the process of urban planning?

The motivation for this research is the necessity to conduct an exploration of spatiotemporal variations at a finer grained level in comparison to what is done currently. Detecting daily and weekly differences at lower scales is particularly interesting for developing a deeper understanding of different urban areas and the relations among them. Furthermore, results of this analysis may be of paramount importance as reliable basis for following planning applications or post-design evaluations.

The objects of this research are what have been defined as ‘unstructured events’ (Mamei and Colonna 2015). Those are urban phenomena when it is not possible, totally or partially, to calculate the attendance of people because of the lack of ticketing or attendance data. They may be planned in advance, and use of space may not be strictly codified. Examples of these types of event are street and local markets, public demonstrations of different nature (political, social, etc.), disruptions or incidents, spontaneous gatherings and many more.

The main objective of the research is to detect different types of 'unstructured events' and describe them from a quantitative perspective, through a complementary analysis of data collected from different sources, specifically smart card, mobile phone and micro-blogging social media platform data. These sources are selected because containing relevant spatial and temporal attributes that are of primary importance for performing data analysis at the spatial level. The multi-source data approach is crucial for detecting even small and usually uncatchable elements, which by their peculiar nature produce insufficient data for analysis. Through these methods it may be possible to identify a higher volume of events, describing them from a quantitative perspective, and enriching perspectives yielded through singular data sources.

References.

Calabrese F, Ratti C (2006) Real time Rome. Networks and Communication Studies, 20(3-4): 247–258

Calabrese F, Pereira F C, Di Lorenzo G, Liu L, Ratti C (2010) The geography of taste: analyzing cell-phone mobility and social events. In Pervasive computing. Springer, pp 22–37.

Calabrese F, Colonna M, Lovisolo P, Parata D, Ratti C (2011) Real-time urban monitoring using cell phones: A case study in Rome. Intelligent Transportation Systems, IEEE Transactions on, 12(1): 141–151.

Candia J, González M C, Wang P, Schoenharl T, Madey G, Barabasi A L (2008) Uncovering individual and collective human dynamics from mobile phone records. Journal of Physics A: Mathematical and Theoretical, 41(22), 224015.

Cranshaw J, Schwartz R, Hong J I, Sadeh N M (2012) The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. In ICWSM. Retrieved from http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/download/4682/4967

Ferrari L, Mamei M, Colonna M (2014) Discovering events in the city via mobile network analysis. Journal of Ambient Intelligence and Humanized Computing, 5(3): 265–277.

Hasan S, Schneider C M, Ukkusuri S V, González M C (2012) Spatiotemporal patterns of urban human mobility. Retrieved from http://humnetlab.mit.edu/wordpress/wp-content/uploads/2010/10/JStat_Hasanpaper_2012.pdf

Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C (2014) Geo-located Twitter as proxy for global mobility patterns. Cartography and Geographic Information Science, 41(3): 260–271.

Lenormand M, Picornell M, Cantu-Ros O G, Tugores A, Louail T, Herranz R, Ramasco J J (2014) Cross-checking different sources of mobility information. arXiv Preprint arXiv:1404.0333. Retrieved from http://arxiv.org/abs/1404.0333

Mamei M, Colonna M (2015) Estimating Attendance From Cellular Network Data. arXiv Preprint arXiv:1504.07385. Retrieved from http://arxiv.org/abs/1504.07385

Munizaga M A, Palma C (2012) Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. Transportation Research Part C: Emerging Technologies 24: 9–18.

Pereira F C, Rodrigues F, Ben-Akiva M (2013) Using Data From the Web to Predict Public Transport Arrivals Under Special Events Scenarios. Journal of Intelligent Transportation Systems, 0(0): 1–16.

Ratti C, Williams S, Frenchman D, Pulselli R M (2006) Mobile landscapes: using location data from cell phones for urban analysis. Environment and Planning B Planning and Design, 33(5) 727. Ratti,

Roth C, Kang S M, Batty M, Barthélemy M, Perc M (2011) Structure of Urban Movements: Polycentric Activity and Entangled Hierarchical Flows. PLoS ONE, 6(1). Retrieved from http://www.complexcity.info/files/2011/06/batty-plos-one-2011.pdf

Traag V A, Browet A, Calabrese F, Morlot F (2011) Social Event Detection in Massive MobilePhone Data Using Probabilistic Location Inference . In 2011 IEEE Third International Conference on Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom): 625–628.

Zhang D, Huang J, Li Y, Zhang F, Xu C, He T (2014) Exploring Human Mobility with Multi-source Data at Extremely Large Metropolitan Scales. In Proceedings of the 20th Annual International Conference on Mobile Computing and Networking. New York, NY, USA: ACM, pp. 201–212.

Zhong C, Arisona S M, Huang X, Batty M, Schmitt G (2014) Detecting the dynamics of urban structure through spatial network analysis. International Journal of Geographical Information Science: 1–22.