Analyzing urban vitality patterns with topological data analysis

Krasen Samardzhiev*[1] and Daniel Arribas-Bel†[1]

[1]Department of Geography, University of Liverpool

January 31, 2019

Summary

Urban researchers have found various uses for new forms of data such as twitter data and mobile phone data, by adapting old methods or adapting new ones from fields such as machine learning. Topological data analysis is a promising new data analysis paradigm, that tries to quantify the qualitative properties of the data, such as its shape. This paper uses these new methods to quantitatively analyze and group areas that exhibit similar patterns of urban vitality within the province of Milan.

KEYWORDS: urban analytics, urban vitality, new forms of data, clustering, topological data analysis.

1 Introduction

Mobile phone calls, text messages or apps enable the capture of large amounts of data that can be stored for later analysis. When combined with spatio-temporal information they can be of interest to researchers looking to answer questions about the urban environment. This data can lead to new insights, but can also present new challenges. In order to meet these challenges researchers adopt new methodologies from fields such as machine learning. Recently, a new field of data analysis - topological data analysis (TDA) is finding applications in various domains. What distinguishes it from machine learning and traditional statistical methods is its focus on more qualitative properties of the data, such as the 'shape' of the data (Carlsson, 2014). The goal of this paper is to demonstrate the usefulness of this new data analysis paradigm by using TDA techniques and methods to quantitatively analyze and group areas that exhibit similar patterns of urban vitality, measured though mobile phone activity data, within the province of Milan.

New forms of data such as mobile phone activity or twitter activity have proven useful for gaining insight into various urban processes and city dynamics. Papers by Sung et al. (2013), De Nadai et al. (2016) and Sulis et al. (2018) use geo-tagged pedestrian counts, mobile phone and twitter

*K.Samardzhiev@liv.ac.uk

†D.Arribas-Bel@liverpool.ac.uk

activity data, respectively, as proxies for urban vitality. The common goal of these papers is to quantitatively test whether the theory of urban diversity presented in Jacobs (1972) holds in the context of non-american cities. Capturing the various patterns, if they exist, of urban vitality within a city can help with this type of quantitative modelling and can bring fresh insights and new classifications. For example, Louail et al. (2015) use the spatial structure of activity hotspots to classify different cities, Sharma et al. (2015) use mobile phone activity data alongside image data to create a geo-demographic classification of a city and Arribas-Bel and Tranos (2018) explore the fast and slow dynamics that the areas within a city exhibit.

Arribas-Bel (2014) identify several properties and challenges when working with data of this type. Some of these are the accidental nature of the data, i.e. the original purpose for gathering it was not research, and the data's temporal and spatial resolution. The challenges that stem from these properties are - the accuracy and coverage of the data, the skills needed to access, store and efficiently process the data, as well as dealing with noise and missing values and the fact that traditional methods might struggle due to the volume or data type. Researchers have met these challenges by adapting old methods for the new forms of data (Arribas-Bel and Tranos, 2018) or by directly using methods from fields such as machine learning, for example Arribas-Bel and Schmidt (2013).

The new emergent field of topological data analysis or TDA also tries to address these challenges. TDA methods extract insights from the data that other methods miss, by focusing on the 'shape' of the data (Carlsson, 2014). The most widely used measure of 'data shape' is the number of clusters or 'connected components. TDA techniques can expand on this and measure higher level concepts such as the number of holes or spheres (Ghrist, 2008) or relationships between the clusters (Singh et al., 2007). This information is useful since it provides insights into the underlying generating processes. For example, when transformed into delay embeddings, signals form a shape in 'phase' space and topological methods can be used to detect and classify them. This has proven useful in various domains and for various purposes (Emrani et al., 2014) (Pereira and De Mello, 2015) (Seversky et al., 2016). TDA techniques can also be used to to detect changes in behaviour of the signals as in Gidea and Katz (2018) and Piangerelli et al. (2018). Successful applications in other domains using similar data and the inherent properties of TDA techniques, such as robustness to noise (Cohen-Steiner et al., 2007) make them good candidates for the task of identifying similar patterns in mobile phone activity data.

2 Data

For this paper mobile phone activity data will be used as a proxy for urban vitality. The specific dataset used is Telecom Italia hourly phone activity data from the Milan province in Italy Italia (2015). The data spans two months from November to December 2011. The province is split into a grid of 10000 squares and each square has associated with it its geographic location and hourly internet values , as well as incoming and outgoing SMS and call values. These values are proportional to the SMS, text and internet usage at that location. The data is also split by country of origin and destination country. For the purposes of this paper the data is cleaned of missing values, origin-destination data is ignored and all SMS, call and internet values are combined into a single mobile phone activity value.

3 Methods

After preprocessing the data, at the most abstract level the main goal is to cluster 10000 time series based on the topological properties, each time series exhibits. To do this first, first the temporal data each for square is transformed into a 'delayed embeddings' representation similar to Emrani et al. (2014) and Pereira and De Mello (2015). Then its topological 'shape' will be calculated using persistent homology (Carlsson, 2014). Previous research (Pereira and De Mello, 2015), (Emrani et al., 2014), (Gidea and Katz, 2018), (Seversky et al., 2016), (Piangerelli et al., 2018) shows that transforming the data in this way and quantifying its shape, captures information inherent to the data generating process. Lastly the distances between the topological shapes will be calculated based on Bubenik and Dłotko (2017), for all pairs of grid squares and the distance matrix will be used for clustering. In order to show which insights come uniquely from the use of TDA methods, the results will be compared with more traditional time series clustering methods.

References

Arribas-Bel, D. (2014). Accidental, open and everywhere: Emerging data sources for the understanding of cities. *Applied Geography*, 49:45 – 53. The New Urban World.

Arribas-Bel, D. and Schmidt, C. R. (2013). Planning and Design. 40:362–371.

Arribas-Bel, D. and Tranos, E. (2018). Characterizing the spatial structure(s) of cities on the fly: The space-time calendar. *Geographical Analysis*, 50(2):162–181.

Bubenik, P. and Dłotko, P. (2017). A persistence landscapes toolbox for topological statistics. *Journal of Symbolic Computation*, 78:91–114.

Carlsson, G. (2014). Topological pattern recognition for point cloud data. *Acta Numerica*, 23:289368.

Cohen-Steiner, D., Edelsbrunner, H., and Harer, J. (2007). Stability of persistence diagrams. *Discrete and Computational Geometry*, 37(1):103–120.

De Nadai, M., Staiano, J., Larcher, R., Sebe, N., Quercia, D., and Lepri, B. (2016). The death and life of great italian cities: A mobile phone data perspective. In *Proceedings of the 25th International Conference on World Wide Web*, WWW '16, pages 413–423, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.

Emrani, S., Chintakunta, H., and Krim, H. (2014). Real time detection of harmonic structure: A case for topological signal analysis. *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, (3):3445–3449.

Ghrist, R. (2008). Barcodes: the Persistent Topology of Data. *BULLETIN (New Series) OF THE AMERICAN MATHEMATICAL SOCIETY*, 45(107):61–75.

Gidea, M. and Katz, Y. (2018). Topological data analysis of financial time series: Landscapes of crashes. *Physica A: Statistical Mechanics and its Applications*, 491:820–834.

Italia, T. (2015). Telecommunications - sms, call, internet - tn.

Jacobs, J. (1972). *The death and life of great American cities / Jane Jacobs.* Penguin Harmondsworth.

Louail, T., Lenormand, M., Cantu Ros, O. G., Picornell, M., Herranz, R., Frias-Martinez, E., Ramasco, J. J., and Barthelemy, M. (2015). From mobile phone data to the spatial structure of cities. *Scientific Reports*, 4(1):5276.

Pereira, C. M. and De Mello, R. F. (2015). Persistent homology for time series and spatial data clustering. *Expert Systems with Applications*, 42(15-16):6026–6038.

Piangerelli, M., Rucco, M., Tesei, L., and Merelli, E. (2018). Topological classifier for detecting the emergence of epileptic seizures. *BMC Research Notes*, 11(1):392.

Seversky, L. M., Davis, S., and Berger, M. (2016). On Time-Series Topological Data Analysis: New Data and Opportunities. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, pages 1014–1022.

Sharma, M., Francis, K., and Ghosh, H. (2015). Unsupervised Geo-Demographic Classification of City-Area Using Multimodal Multimedia Data. In *2015 IEEE International Conference on Multimedia Big Data*, pages 240–243. IEEE.

Singh, G., Mémoli, F., and Carlsson, G. (2007). Topological Methods for the Analysis of High Dimensional Data Sets and {3D} Object Recognition. *Eurograph. Sympos. Point-Based Graphics*, pages 91–100.

Sulis, P., Manley, E., Zhong, C., and Batty, M. (2018). Using mobility data as proxy for measuring urban vitality. *JOURNAL OF SPATIAL INFORMATION SCIENCE Number*, 16:137–162.

Sung, H.-G., Go, D.-H., and Choi, C. G. (2013). Evidence of Jacobs's street life in the great Seoul city: Identifying the association of physical environment with walking activity on streets. *Cities*, 35:164–173.