Modelling Urban Flows: Spatial Effects in Origin-Destination data

Lenka Hasova[1], Levi John Wolf[1]

[1] Department of Geography, University of Bristol

January 17, 2020

Summary

This paper is an introduction of preliminary research for Ph.D. thesis 'Machine Learning methods for Urban Flows: spatial effects in Origin-Destinations'. It examines the biggest challenges in the theory of Gravity models and Spatial Interaction models for urban flow prediction, including the nature of the human spatial decision-making. It also outlines the biggest concerns with the methodological approaches for predicting urban flows and presents the future steps of this research.

KEYWORDS: spatial interaction, gravity, distance-decay, urban flow, origin-destination

1. Introduction

Spatial interaction modelling (SIM) also known as gravity models are typically applied to capture the spatial effects in urban flow. Today, understanding urban flows is increasingly important within an urban world (Kourtit *et al.*, 2014). Whilst the theoretical basis for SIM was developed as early as the 20th century (Wilson, 1971), SIM methods was limited by data availability and computational power (Poot *et al.*, 2016; Pappalardo and Simini, 2018; Guzmán, 2018). In the last decade in an era of "big data", SIM methods have been developed in the areas of migration flows, trade flows, commuter flows and other areas(Ren *et al.*, 2014; Robinson and Dilkina, 2017; Barbosa *et al.*, 2018). However, simple statistical approaches fail to capture the inherently non-linear and complex nature of urban flows. More recently, advances in Machine Learning (ML) methods and Artificial Intelligence (AI) represents an opportunity to capture these complex representations (Openshaw, 1988). Whilst statistical methods have been expanded to consider spatial non-stationarity, there is a limited number of ML methods that consider this.

2. Background,

Human mobility, spatial behaviours and the spatial decision-making processes has been critically discussed in the literature for decades. For example, scientists began to question the predictability of these decisions of individuals labeled "Economic Man" (Wolpert, 1964). They explored the role of behavioural and cognitive environment in shaping the spatial knowledge and decision process of individuals (Cadwallader, 1975). They found that an individual's perception of the environment is not always related to their behaviour within that environment but there are many more aspects that are difficult to capture by qualitative research alone.

One of the major quantitative findings of urban spatial decision processes is that human travel patterns are far more complex than simple frequency of trips or distances(Hanson and Hanson, 1981). The study found several significant variables, including socio-economic status, family roles and the dimensionality of the trip (single-destination, double-destination or multi-destination trips). In particular, the significance of the gender role transformed the image of the "Economic Man" to one where individual choices are more sporadic than once perceived. Whilst Golledge and Stimson (1996) agree with this statement but they argue that some rules still apply. In particular individuals are influenced by the economic and social structures created by society but at the same time, individuals want to minimize their space-time utility. The question that remains is to what degree is the spatial decision-making process random? and whether these social structures are explicitly or implicitly imposed (Song *et al.*, 2010).

Following Newton's laws of gravity, a base framework for SIM and entropy maximization was developed by Wilson (1970). Whilst this base framework is still extensively used to study urban flows, the concept of space is often misspecified (Fotheringham, 1983). In fact, already Newton himself acknowledged that his theory is not without a flaw (Gates, 2019). The misspecification of space is often linked to the distance decay-function, where either the dimensionality dilemma occurs (Fotheringham, 1983), or it violates the first law of Geography from Tobler (2004), (Chen, 2015).

Whilst Anselin (1988) captured these spatial effects, through Spatial Autoregressive models, these alone cannot capture spatial interactions. Because spatial interaction has a two-dimensional relationship based on pairs of areas or units, which needs more complex solutions such as introduction of spillover effects (LeSage *et al.*, 2007; LeSage and Pace, 2008).

Since then, many different methodologies and alternative interaction models has been applied for modelling Urban flows. The Bayesian methods celebrate a great success in application for urban flows modelling, in part due to (LeSage, 2000; Smith and LeSage, 2004; LeSage and Llano, 2013; LeSage and Satici, 2016). However, comparison of Frequentists and Probabilistic approaches on spatial flows with spatial effects by Griffith *et al* (2017) shows that the Frequentists approach has less residual error and the distance decay parameters were very different. Also radiation models, introduced by Simini *et al.* (2012) could be one of the alternative models that could further succeed in Urban flow prediction. The models replace distance decay with a function of population size in a radiation distance between origin and destination. However, decreased predictive power as same as in the gravity models was observed in their application (Masucci *et al.*, 2013; Ren *et al.*, 2014; Robinson and Dilkina, 2017). Openshaw (1998) argued for a paradigm shift away from classic statistical approaches, to ML ones. He was one of the first to apply neural networks to model spatial interaction flows, with results demonstrating a two-fold improvement in model performance over classic statistical approaches. More recently, Robinson and Dilkina (2017) also applied a similar methodology which confirm the results of Openshaw (1998).

3. Methods

The aim of the research is to apply ML methodologies including Convolutional Neural Networks (CNNs) on urban population flow data to evaluate their performance with traditional SIMs. Specifically, we want to explore whether incorporating non-stationarity in several CNNs architectures better captures the local patterns of urban flows and improves model fit. Moreover, the research focuses on the spatial aspect of flow networks and correct definition of spatial in the modeling framework.

Following questions are going to be answered.

- What is the nature of the flow, how it relates to human nature and what role does the 'space' plays in their decision-making process?
- Where are the biggest gaps in the traditional methodology; how do we account for 'space' in gravity-based models?
- What are the advances that can be gathered from the recent research?
- Is ML an appropriate methodology for predicting 'spatial' urban flows?
- Do ML approaches better capture complex representations of 'spatial' urban flowss?
- What are the limitations of ML method for predicting the 'spatial' urban flows?

The research uses multiple open data sources to explore the diverse nature of dynamic urban flows. Firstly, NHS patient's movement data, within the Avon area, are used. Those represents the lower scale flow data with a strong spatial constraints and are the main dataset for this research. To compare the results with other flows, UK migration data are used. Those represents the high scale flows with a strong socio-economical constraint.

The NHS patient's movement data are Origin-Destination prescription flow data, where origin is a unique NHS prescriber (general practitioner, surgery, health centre), destination is a unique dispenser (pharmacy) and the flow volume is a count of prescriptions on the flux each month. Both, prescriber and dispenser are geolocated entities that can be connect to place and infrastructure. Using a gravity-based models can help us reveal the patterns of the patients decision-choice based movement. Furthermore, using ML methods will help to tackle the inconsistencies in the SIM theory, validate some of the conclusions made in the theory in recent years and increase the overall accuracy of the models.

There are several methodological challenges that needs to be overcome. Firstly, finding the most appropriate distance-decay function for the patients' flow is crucial for model building. Secondly, finding a correct model definition for an asymmetrical flow network of NHS patients. Lastly, training a neural network with spatial effects and interpreting the results.

4. Conclusion

Vast majority of the studies face serious overestimations sourced in the model's inability to capture the outliers. However, in any flow data, urban flows especially, it is the outliers we want to be able to predict with the highest accuracy. The outliers are tightly connected to the spatial and hierarchical aspects of the flow, but models accounting for space in flow data has been built on theories that violate the idea of space and hierarchy in them. Broido and Clauset (2019) proved that almost 95% of natural flow networks possess a hierarchical or spatial structure. This research represents a next step for the social, geographical and AI sciences towards better understanding of machine learning methods for urban flow and urban development. Moreover, understanding the patient's movements and spatial decision making, is crucial for efficient distribution of sources in health sector.

5. Limitations

Much deeper questions arises here. Firstly, the nature of the urban flow was proved to be complex and is influenced by the diverse nature of a human being and the socio-economical system of our society, subsequently, the scientists are still unsure about how much of the movement is predictable and how much remains uncertain. Perhaps to understand the dynamics of networks, links and the nature of network itself, we first must learn how the network evolves (Batty, 2003). Secondly, the traditional methods, although mathematically simplistic, fail to impress in urban flow modeling (Chen, 2015).

Machine Learning could be effective in modeling flow networks; however, black box models won't allow us to see the relationships defined inside, and do we really don't want to know? Those are critical issues that cannot be answered by one research, neither by one scientist.

6. Acknowledgements

This research was supported by the Economic and Social Research Council.

7. Biography

Lenka Hasova is Ph.D. candidate in Advanced Quantitative Geography at University of Bristol. Her research examines the possibilities of Machine Learning application for modelling Urban flows, with major interest in correct interpretation of the spatial effects within the framework.

Levi J. Wolf is a senior lecturer in Quantitative Human Geography at University of Bristol, Alan Turing fellow and Centre for Spatial Data Science fellow. He works in spatial data science, building new methods and software to learn new things about social and natural processes.

References

Anselin, L. (1988) *Spatial Econometrics: Methods and Models*. Springer Science & Business Media.

Barbosa, H. *et al*. (2018) ‘Human Mobility: Models and Applications’. *Physics Reports*, 734, pp. 1–74. DOI: 10.1016/j.physrep.2018.01.001.

Batty, M. (2003) Available at: http//www.casa.ucl.ac.uk/working_papers/paper63.pdf.

Broido, A.D. and Clauset, A. (2019) ‘Scale-Free Networks Are Rare’. *Nature Communications*, 10(1), pp. 1–10. DOI: 10.1038/s41467-019-08746-5.

Cadwallader, M. (1975) ‘A Behavioral Model of Consumer Spatial Decision Making’. *Economic Geography*, 51(4), pp. 339–349. DOI: 10.2307/142918.

Chen, Y. (2015) ‘The Distance-Decay Function of Geographical Gravity Model: Power Law or Exponential Law?’ *Chaos, Solitons & Fractals*, 77, pp. 174–189. DOI: 10.1016/j.chaos.2015.05.022.

Fotheringham, A.S. (1983) ‘A New Set of Spatial-Interaction Models: The Theory of Competing Destinations’. *Environment and Planning A: Economy and Space*, 15, pp. 15–36.

‘Jim Gates: Supersymmetry, String Theory and Proving Einstein Right, Artificial Intelligence Podcast’. 5/71.

Golledge, R.G. and Stimson, R.J. (1996) *Spatial Behaviour, A Geographic Perspective*. New York: The Guildford Press.

Griffith, D.A., Fischer, M.M. and LeSage, J. (2017) ‘The Spatial Autocorrelation Problem in Spatial Interaction Modelling: A Comparison of Two Common Solutions’. *Letters in Spatial and Resource Sciences*, 10(1), pp. 75–86. DOI: 10.1007/s12076-016-0172-8.

Guzmán, L.F.S. (2018) ‘A Statistical Approach for Studying Urban Human Dynamics’. Available at: https://run.unl.pt/handle/10362/56709 (Accessed: 30 October 2019).

Hanson, S. and Hanson, P. (1981) ‘The Travel-Activity Patterns of Urban Residents: Dimensions and Relationships to Sociodemographic Characteristics’. *Economic Geography*, 57(4), pp. 332–347. DOI: 10.2307/144213.

Kourtit, K., Nijkamp, P. and Reid, N. (2014) ‘The New Urban World: Challenges and Policy’. *Applied Geography*, 49, pp. 1–3. DOI: 10.1016/j.apgeog.2014.01.007.

LeSage, J.P. (2000) ‘Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models’. *Geographical Analysis*, 32(1), pp. 19–35. DOI: 10.1111/j.1538-4632.2000.tb00413.x.

LeSage, J.P., Fischer, M.M. and Scherngell, T. (2007) ‘Knowledge Spillovers across Europe: Evidence from a Poisson Spatial Interaction Model with Spatial Effects*’. *Papers in Regional Science*, 86(3), pp. 393–421. DOI: 10.1111/j.1435-5957.2007.00125.x.

LeSage, J.P. and Llano, C. (2013) ‘A Spatial Interaction Model with Spatially Structured Origin and Destination Effects’. *Journal of Geographical Systems*, 15(3), pp. 265–289. DOI: 10.1007/s10109-013-0181-8.

LeSage, J.P. and Pace, R.K. (2008) ‘Spatial Econometric Modeling of Origin-Destination Flows*’. *Journal of Regional Science*, 48(5), pp. 941–967. DOI: 10.1111/j.1467-9787.2008.00573.x.

LeSage, J.P. and Satici, E. (2016) 'A Bayesian Spatial Interaction Model Variant of the Poisson Pseudo-Maximum Likelihood Estimator'. In Patuelli, R. and Arbia, G. (eds.) *Spatial Econometric Interaction Modelling*. Advances in Spatial Science. Cham: Springer International Publishing, pp. 121–143. DOI: 10.1007/978-3-319-30196-9_7.

Masucci, A.P. *et al.* (2013) 'Gravity vs Radiation Model: On the Importance of Scale and Heterogeneity in Commuting Flows'. *Physical Review E*, 88(2), p. 022812. DOI: 10.1103/PhysRevE.88.022812.

Openshaw, S. (1988) 'Building an Automated Modeling System to Explore a Universe of Spatial Interaction Models'. *Geographical Analysis*, 20(1), pp. 31–46. DOI: 10.1111/j.1538-4632.1988.tb00161.x.

Openshaw, S. (1998) 'Neural Network, Genetic, and Fuzzy Logic Models of Spatial Interaction'. *Environment and Planning A: Economy and Space*, 30(10), pp. 1857–1872. DOI: 10.1068/a301857.

Pappalardo, L. and Simini, F. (2018) 'Data-Driven Generation of Spatio-Temporal Routines in Human Mobility'. *Data Mining and Knowledge Discovery*, 32(3), pp. 787–829. DOI: 10.1007/s10618-017-0548-4.

Poot, J. *et al.* (2016) 'The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science'. pp. 63–86.

Ren, Y. *et al.* (2014) 'Predicting Commuter Flows in Spatial Networks Using a Radiation Model Based on Temporal Ranges'. *Nature Communications*, 5, p. 5347. DOI: 10.1038/ncomms6347.

Robinson, C. and Dilkina, B. (2017) 'A Machine Learning Approach to Modeling Human Migration'. *ArXiv:1711.05462 [Physics]*. Available at: http://arxiv.org/abs/1711.05462 (Accessed: 26 October 2019).

Simini, F. *et al.* (2012) 'A Universal Model for Mobility and Migration Patterns'. *Nature*, 484(7392), pp. 96–100. DOI: 10.1038/nature10856.

Smith, T.E. and LeSage, J.P. (2004) 'A BAYESIAN PROBIT MODEL WITH SPATIAL DEPENDENCIES'. In *Advances in Econometrics*. Bingley: Emerald (MCB UP), pp. 127–160. DOI: 10.1016/S0731-9053(04)18004-3.

Song, C. *et al.* (2010) 'Limits of Predictability in Human Mobility'. *Science*, 327(5968), pp. 1018–1021. DOI: 10.1126/science.1177170.

Tobler, W. (2004) 'On the First Law of Geography: A Reply'. *Annals of the Association of American Geographers*, 94(2), pp. 304–310. DOI: 10.1111/j.1467-8306.2004.09402009.x.

Wilson, A.G. (1971) 'A Family of Spatial Interaction Models, and Associated Developments'. *Environment and Planning A: Economy and Space*, 3(1), pp. 1–32. DOI: 10.1068/a030001.

Wilson, A.G. (1970) 'Inter-Regional Commodity Flows: Entropy Maximizing Approaches'. *Geographical Analysis*, 2(3), pp. 255–282. DOI: 10.1111/j.1538-4632.1970.tb00859.x.

Wolpert, J. (1964) 'THE DECISION PROCESS IN SPATIAL CONTEXT [1]'. *Annals of the Association of American Geographers*, 54(4), pp. 537–558. DOI: 10.1111/j.1467-8306.1964.tb01783.x.