Machine Learning for Near-Real Time Bus Ridership Prediction During "Extreme" Weather

Francisco Rowe[*1], Michael Mahony[†1] and Sui Tao[‡2]

[1]Department of Geography and Planning, University of Liverpool, Liverpool, UK
[2]Institute of Future Cities, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

November 16, 2021

Summary

Summary of no more than 100 words

Given an increasingly volatile climate, the relationship between weather and transit ridership has drawn increasing interest. However, challenges stemming from spatio-temporal dependency and non-stationarity have not been fully addressed in modelling and predicting transit ridership under the influence of weather conditions. Drawing on three-month smart card data in Brisbane, Australia, this research implements a suite of tree-based machine-learning algorithms, to model and predict near real-time bus ridership in relation to sudden change of weather conditions. The study confirms that there indeed exists a significant level of spatio-temporal variability of weather-ridership relationship, which produces equally dynamic patterns of prediction errors.

KEYWORDS: Public transport; Machine learning; Big data; Travel behaviour; Weather

1. Abstract

Global warming resulting from the carbon emissions from human activities has become one of the most critical challenges for humanity (Nordhaus, 2007; Moser, 2010). Human-induced warming has raised the global temperature by approximately 1°C compared to the pre-industrial era according to a report by the Intergovernmental Panel on Climate Change (IPCC) (Allen et al., 2018). The occurrence of extreme weather events, including extremely high and low temperatures, storms and hurricanes, has also increased in frequency since the 1950s (Stott, 2016). Such an increasingly versatile climate will considerably impact the natural and social environment, as such, potentially causing new issues and challenges for human societies (Urry, 2015; Allen et al., 2018).

In combating global warming, interventions to curb carbon emissions are critical (Nordhaus, 2007; Zhang and Cheng, 2009). Particular attention has been drawn to the transportation sector (Loo and Li, 2012; Metz, 2015). The global collective contribution of road and other types of transport to carbon emissions has been estimated to be over 14% (IPCC, 2014). Promoting greener and attractive public transit service while reducing excessive car usage has been stressed as a key strategy (Friman et al., 2013; Tao et al., 2019). However, the increasing variability of weather associated with climate change may cause disruptive effects on the operation and management of urban transit systems (Hofmann and O'Mahony, 2005; Miao et al., 2019), as well as demand for transit service (Böcker et al., 2013; Tao et al., 2018). How to ensure reliable and adequate transit service in response to sudden change of weather has attracted more concerns (Arana et al., 2014; Singhal et al., 2014).

A scrutiny of the transport literature reveals that real-time relationships between weather and public transport ridership has received increasing scholarly attention (e.g. Singhal et al., 2014; Wei et al., 2019). However, limited empirical research has simultaneously investigated the spatial and temporal

[*] F.Rowe-Gonzalez@liverpool.ac.uk
[†] M.R.Mahony@liverpool.ac.uk
[‡] S.Tao@cuhk.edu.hk

dynamics of transit use in relation to weather conditions, which may restrain our ability to better predict transit demand and inform real-time operation of transit service. Traditionally, time-use surveys and household travel diaries are used to study urban travel behaviour (Kitamura, 1988; Schlich and Axhausen, 2003). While valuable, these data sources are expensive, infrequent, offer low population coverage and become available with some delay following their gathering. Moreover, the spatial and temporal scales of such data are often limited due to high collection cost (Noland and Polak, 2002). New forms of data, such as smart card and mobile phone data have emerged as novel and useful sources to study human mobility. These data are generated at unprecedented temporal frequency and geographical granularity, enabling the analysis of human travel activity in real and near-real time scales. Another key constraint to our current understanding of real-time relationships between weather and public transport ridership is a dearth in the application of simple, affordable, rapidly customisable and accurate predictive models. Such models could be key to bridge empirical research and practice in the operation of transit service. Predictive machine learning approaches enabled by affordable computing power emerge as a promising candidate. Unlike traditional statistical modelling methods, machine learning approaches do not necessarily rely on a predefined manual model specification (James et al., 2013), and require limited human interventions in specifying the most appropriate model configuration for modelling urban traffic data associated with spatial and temporal dependence and non-stationarity (e.g. Zhou et al., 2021). Recently, more transport studies have opted for machine learning approaches to examine and model transport phenomena in a timely and efficient manner (Ma et al., 2015; Polson and Sokolov, 2017; Wang et al., 2018). However, the application of machine learning approach to model and predict transit ridership vis-à-vis weather conditions has been relatively limited.

This paper aims to address the identified gaps by using off-the-shelf machine learning models to produce near-real time, stop-level bus ridership predictions in response to sudden changes in weather conditions. To this end, we draw on a three-month smart card dataset of bus ridership, containing over 10 million observations allied with detailed weather measurements, trip length, calendar events, and built environment features to form an integrated spatio-temporal database. We train and assess three off-the-shelf tree-based machine learning algorithms, Random Forest (RF), eXtreme Gradient Boosting (XGBoost) and Tweedie XGBoost in their ability of short-term ridership prediction. We in particular assess the predictive model performance during periods of extreme temperature, rainfall, wind and humidity records.

Our results suggest relatively high accuracy for the three models. We also find marked spatio-temporal variability in prediction error across models pointing to complex relationships between stop-level bus ridership and local weather conditions. Tweedie XGBoost generates the most accurate predictions, providing narrower prediction error intervals in a wide set of scenarios, particularly during periods of heavy rain. The proposed model has the capacity to account for the skewed distribution of bus ridership, and generate accurate real-time predictions. We contend that the current study represents a necessary step towards further channelling big-data analysis into the real-time operation of transit service in the face of an increasingly volatile climatic context (e.g., reallocating bus services).

2. Acknowledgements

No funding was provided for this research. We would like to acknowledge the developers of the following R libraries used in our paper (order alphabethically): chron, doParallel, dplyr, ggcorrplot, ggmap, ggplot2, ggsn, ggthemes, glmnet, gstat, h2o, kableExtra, moments, osmdata, parallel, patchwork, raster, RColorBrewer, recommenderlab, rgdal, rgeos, runner, sf, showtext, sp, summarytools, tidyquant, tidyr, tidyverse, timechange, twilio and viridis.

References

Allen, M. et al. (2018) *Global Warming of 1.5°C*, IPCC. Geneva, Switzerland: World Meteorological Organization.

Arana, P., Cabezudo, S. and Peñalba, M. (2014) ‘Influence of weather conditions on transit ridership:

A statistical study using data from Smartcards', *Transportation Research Part A: Policy and Practice*, 59, pp. 1–12. doi: 10.1016/j.tra.2013.10.019.

Böcker, L., Dijst, M. and Prillwitz, J. (2013) 'Impact of Everyday Weather on Individual Daily Travel Behaviours in Perspective: A Literature Review', *Transport Reviews*, 33(1), pp. 71–91. doi: 10.1080/01441647.2012.747114.

Friman, M., Larhult, L. and Gärling, T. (2013) 'An analysis of soft transport policy measures implemented in Sweden to reduce private car use', *Transportation*, 40(1), pp. 109–129. doi: 10.1007/s11116-012-9412-y.

Hofmann, M. and O'Mahony, M. (2005) 'The impact of adverse weather conditions on urban bus performance measures', in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*. Institute of Electrical and Electronics Engineers Inc., pp. 84–89. doi: 10.1109/ITSC.2005.1520087.

IPCC (2014) *Climate Change 2014: Synthesis Report*. Geneva, Switzerland. doi: 10.1017/CBO9781139177245.003.

James, G. et al. (2013) *An Introduction to Statistical Learning*. New York, NY: Springer New York (Springer Texts in Statistics). doi: 10.1007/978-1-4614-7138-7.

Kitamura, R. (1988) 'An evaluation of activity-based travel analysis', *Transportation*, 15(1), pp. 9–34.

Loo, B. P. Y. and Li, L. (2012) 'Carbon dioxide emissions from passenger transport in China since 1949: Implications for developing sustainable transport', *Energy Policy*, 50, pp. 464–476. doi: 10.1016/j.enpol.2012.07.044.

Ma, X. et al. (2015) 'Long short-term memory neural network for traffic speed prediction using remote microwave sensor data', *Transportation Research Part C: Emerging Technologies*, 54, pp. 187–197. doi: 10.1016/j.trc.2015.03.014.

Metz, D. (2015) 'Peak Car in the Big City: Reducing London's transport greenhouse gas emissions', *Case Studies on Transport Policy*, 3(4), pp. 367–371. doi: 10.1016/j.cstp.2015.05.001.

Miao, Q., Welch, E. W. and Sriraj, P. S. (2019) 'Extreme weather, public transport ridership and moderating effect of bus stop shelters', *Journal of Transport Geography*, 74, pp. 125–133. doi: 10.1016/j.jtrangeo.2018.11.007.

Moser, S. C. (2010) 'Communicating climate change: History, challenges, process and future directions', *Wiley Interdisciplinary Reviews: Climate Change*, 1(1), pp. 31–53. doi: 10.1002/wcc.11.

Noland, R. B. and Polak, J. W. (2002) 'Travel time variability: A review of theoretical and empirical issues', *Transport Reviews*, 22(1), pp. 39–54. doi: 10.1080/01441640010022456.

Nordhaus, W. D. (2007) 'To Tax or Not to Tax: Alternative Approaches to Slowing Global Warming', *Review of Environmental Economics and Policy*, 1(1), pp. 26–44. doi: 10.1093/reep/rem008.

Polson, N. G. and Sokolov, V. O. (2017) 'Deep learning for short-term traffic flow prediction', *Transportation Research Part C: Emerging Technologies*, 79, pp. 1–17. doi: 10.1016/j.trc.2017.02.024.

Schlich, R. and Axhausen, K. W. (2003) 'Habitual travel behaviour: Evidence from a six-week travel diary', *Transportation*, 30(1), pp. 13–36.

Singhal, A., Kamga, C. and Yazici, A. (2014) 'Impact of weather on urban transit ridership', *Transportation Research Part A: Policy and Practice*, 69, pp. 379–391. doi: 10.1016/j.tra.2014.09.008.

Stott, P. (2016) ‘How climate change affects extreme weather events’, *Science*, 352(6293), pp. 1517–1518. doi: 10.1126/science.aaf7271.

Tao, S., Corcoran, J., Rowe, F., Hickman, M. (2018) ‘To travel or not to travel: “Weather” is the question. Modelling the effect of local weather conditions on bus ridership’, *Transportation Research Part C: Emerging Technologies*, 86, pp. 147–167. doi: 10.1016/j.trc.2017.11.005.

Urry, J. (2015) ‘Climate Change and Society’, in Michie, J. and Cooper, C. L. (eds) *Why the Social Sciences Matter*. London: Palgrave Macmillan UK, pp. 45–59. doi: 10.1057/9781137269928_4.

Wang, Z., He, S. Y. and Leung, Y. (2018) ‘Applying mobile phone data to travel behaviour research: A literature review’, *Travel Behaviour and Society*, 11, pp. 141–155. doi: 10.1016/j.tbs.2017.02.005.

Wei, M. et al. (2019) ‘The influence of weather conditions on adult transit ridership in the sub-tropics’, *Transportation Research Part A: Policy and Practice*, 125, pp. 106–118. doi: 10.1016/j.tra.2019.05.003.

Zhang, X. P. and Cheng, X. M. (2009) ‘Energy consumption, carbon emissions, and economic growth in China’, *Ecological Economics*, 68(10), pp. 2706–2712. doi: 10.1016/j.ecolecon.2009.05.011.

Zhou, F. et al. (2021) ‘Urban flow prediction with spatial–temporal neural ODEs’, *Transportation Research Part C: Emerging Technologies*, 124(March 2020), p. 102912. doi: 10.1016/j.trc.2020.102912.

Biographies

Francisco Rowe is a Senior Lecturer in Human Quantitative Geography at the University of Liverpool, member of the Geographic Data Science Lab and a Project Associate of project IMAGE: Comparing Internal Migration Around the Globe. His areas of expertise are: internal & international migration and computational approaches to population research.

Michael Mahony is a PhD student in the Geographic Data Science Lab at the University of Liverpool. His research primarily focuses on integration of migrants into the UK labour market. He has also published work with the International Organisation for Migration (IOM) measuring public sentiment towards migrants using Twitter data.

Sui Tao is a postdoctoral research fellow at the Institute of Future Cities (IOFC), The Chinese University of Hong Kong (CUHK). He is an urban planner and geographer interested in activity-travel patterns across urban transport systems, public transport dynamics and socio-spatial equity associated with transport infrastructure and urban development.