Neighbourhood effects on the demand of e-scooter services - - A comparison between spatial regression models and convolutional neural network (CNN)

Authors

Shiva Habibi[1][1], Kateryna Melnyk[1], Joakim Rosell[1], Ruiting Qin[2]

1 Department of Mobility and Systems, Research Institute of Sweden
2 Department of Statistics, Uppsala University, Sweden

Submission Topic

□ Machine learning and other new methods

Keywords

e-scooter sharing mobility, out-of-sample prediction, spatial distribution, spatial regression, CNN, demand

Extended Abstract

In recent years, a new generation of shared micromobility, i.e., shared electric scooters or shared e-scooters, has rapidly become popular in many cities globally. As a result, the number of publications on the subject has recently increased in an effort to understand this emerging travel mode. However, more research is needed to better understand the role shared electric scooters play in the transport system to provide city planners, operators and decision-makers with information to strengthen the positive impacts and mitigate the negative ones.

In this paper, we develop two different types of models, deep learning-based and econometric-based, to identify the relation between the demand for e-scooter sharing and built environment characteristics. The results specify the spatial distribution of e-scooter sharing trips over different city neighbourhoods. The estimated (trained) models are then used to predict e-scooter sharing demand for a new city. The prediction results will help get insight into defining the geographical boundary of e-scooter sharing before its expansion to a city. Moreover, we compare the performance of a deep learning method with spatial regression in terms of both fit and prediction power.

Different studies have focused on different aspects of e-scooter sharing mobility, such as substitution patterns (see, e.g., Fitt & Curl, 2019; Zuniga-Garcia & Machemehl, 2020), trip characteristics (see, e.g., Liu e. al., Engdahl et al., 2020, Maxwell, 2019; Younes et al., 2020), purpose of trips (McKenzie, 2019), environmental effects of e-scooters (Hollingsworth et al., 2019; Griswold, 2019; Møller & Simlett, 2020), parking issues (Fang et al., 2018; James et al., 2019; Brown et al., 2020), safety (Badeau et al., 2019; Blomberg et al., 2019; Bekhit et

[1] Corresponding author: shiva.habibi@ri.se

al., 2020), and effects of Covid (Li et al., 2021). Recently, Li et al. (2022) have made an extensive comparative study of e-scooter sharing mobility across 30 European cities.

In this study, we use spatial regression and convolutional neural network (CNN) to estimate the spatial distribution of demand for e-scooter sharing over different city neighbourhoods accounting for spatial dependencies of different neighbouring zones and non-linear effects of the spatial characteristics on the demand. Prior works on e-scooter spatial distribution (see, e.g., Jiao & Bai, 2020; Bai & Jiao, 2020) do not account for spatial dependencies. In the context of electric bike-sharing, Caspi et al. (2020) use spatial regression to account for spatial dependencies when estimating the distribution of e-bike sharing demand. They predict the demand using both spatial regression and random forests. However, the classical random forest algorithm, introduced by Breiman (2001), does not account for spatial dependencies while CNN does and, therefore, provides a similar foundation to compare the classical spatial regression models with machine learning/deep learning methods. We test the prediction on several cities and evaluate the predictive power of the models. To the best of our knowledge, our work is the first to conduct this comparison in the e-scooter sharing context and perform out-of-sample prediction.

This paper contributes to the existing literature on e-scooters in three main ways. First, identifying built environment characteristics on e-scooter sharing demand using spatial regression and CNN. Second, predicting the demand for e-scooter sharing and its spatial distribution for a new city. Third, comparing the performance of machine learning/deep learning methods such as CNN with the theoretical models used in spatial econometrics. These empirical comparative studies shed light on the strengths and weaknesses of using each method in different contexts in transport studies.

Methods

We aim to compare the performance of a CNN against a spatial regression model in quantifying the neighbourhoods' characteristics that affect the demand for e-scooter sharing as well as their out-of-sample predictive power when predicting the e-scooter sharing demand for a new city. The models describe a relation between the number of bookings (dependent variable) and sociodemographic and spatial characteristics of the neighbourhoods (explanatory or independent variables). There has been a growing recognition in employing spatial econometrics in the presence of spatial dependencies or spatial autocorrelation in transport studies (see, e.g., Caspi et al., 2020; Becker et al., 2017). Spatial dependencies exist if the value of the variable of interest at a given location is dependent on the values of the same variables at other locations in the system. In this case, the number of e-scooter sharing bookings in a given neighbourhood not only depends on the spatial characteristics of its neighbourhood but also on the number of bookings in the adjacent neighbourhood. Spatial econometrics accounts for spatial dependencies (autocorrelation) and spatial heterogeneity in regression models (Anselin, 2001).

Convolutional neural network (CNN) is a class of artificial neural networks that has become dominant in various computer vision tasks. There are several reasons to use CNN instead of simpler machine learning methods such as random forest algorithm. The first reason is that CNN makes an assumption about the locality of pixel dependencies which means that neighboring pixels tend to be correlated while faraway pixels are usually not correlated (Krizhevsky et al., 2012). Therefore, CNN can be used to capture special dependencies as spatial regression does. Moreover, most machine learning methods require an initial feature selection step before building a model whereas CNN has the ability to automatically discover multiple levels of joint representation of the input data. The final reason is computational efficiency. Due to a set of convolution operations, the dimensionality of the data shrinks significantly and the number of parameters to be learned decreases as well, leaving only those which have the biggest impact on the classification or prediction power. Further, to assess the neighborhood's characteristics that might correlate with the demand for e-scooter sharing, we will make use of state-of-the-art methods for the interpretation of deep learning methods, in our case CNN. The interpretation methods such as Layer-wise relevance propagation (Bach et al., 2015) help to better understand "why" the model arrives at a certain prediction and how much each feature in the input data (e.g., a pixel) contributes to the prediction of the model.

Data

There are three primary data sources for this project: e-scooter sharing booking data, geo-coded data on sociodemographic and urban points of interest.

The first database is the e-scooter sharing booking data containing the following variables for each rental: vehicle ID, starting and ending time of the rental, geographic position of vehicles at the start and end time. The data are from different operators, including 16 different cities located in Europe and North America, from different operators, resulting in approximately 10 million vehicle movements. This rich data enable us to do comparative studies that are essential to understanding the underlying drives for e-scooters demand and usage patterns as well as performing out-of-sample prediction to compare the predictive power of the models.

The second database is hectare-based geo-coded sociodemographic such as age and gender distribution, income distribution and the number of workplaces. The third database is Open Street Maps where geo-coded data of points of interest such as the number of residential and commercial buildings as well as transport-related variables such as the number of public transport stops available. The goal is to locate all three data sources in a grid of hectare-sized cells over the operating area of the e-scooter services.

Expected results

We are in the process of preparing, merging, and analysing the data as prerequisites for modelling efforts. We will continue by estimating spatial regression and CNN models on Gothenburg. The estimated model will be used to predict the distribution of bookings in Stockholm, Paris and Berlin.

In this section, we present some results of descriptive analysis of the datasets of Gothenburg. Figure 1 shows the distribution of the average number of rentals per day.

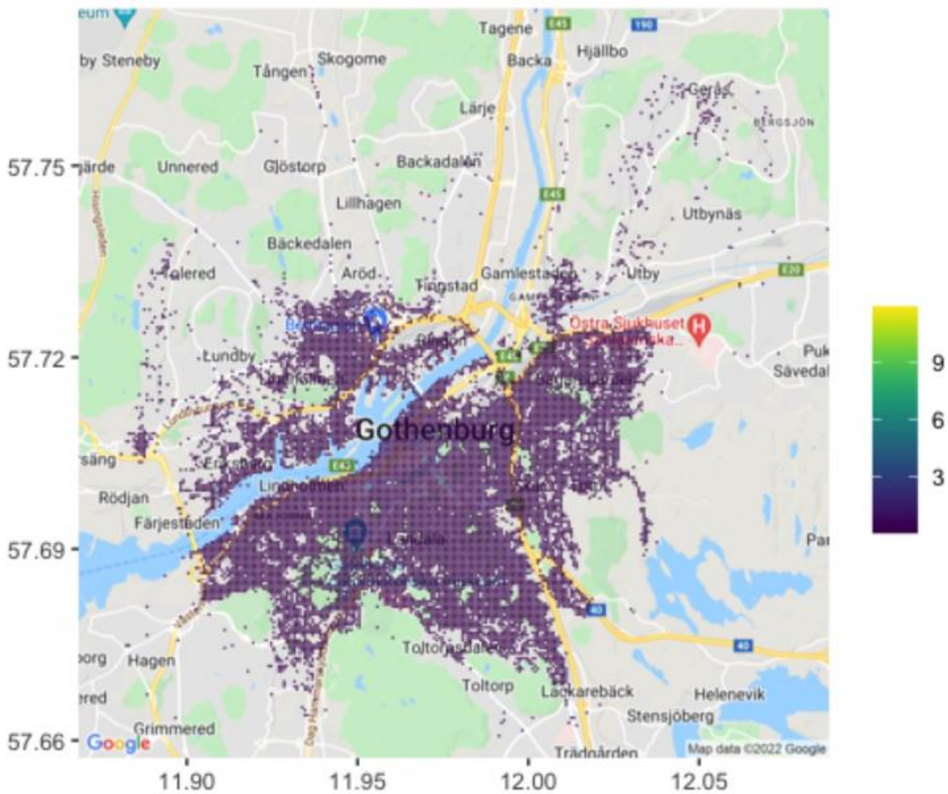

Fig. 1. Average number of rentals per day in Gothenburg

As mentioned above the provided datasets contain individual vehicle IDs for all of the trips within the datasets, which enable us to analyse data of individual e-scooters. Further, the so-called state of charge (SoC) at every trip are also included in the datasets. The SoC is a dimensionless number indicating the level of charge of the scooters' electric battery relative to its capacity. Charging trips. i.e., trips with a SoC at the beginning larger than the SoC at the end of the trip.

The average number of trips per hour for each city analysed can be seen in Figure 2. It can be observed that there seem to be a trend of e-scooters being mostly used around 15:00 for most of the cities, with a local minimum at the early hours of 00:04 to 06:00.

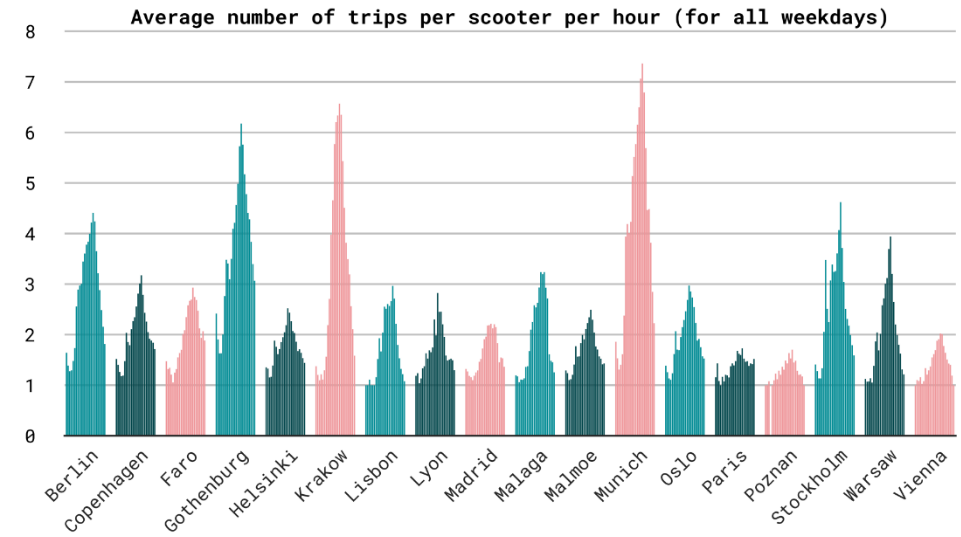

Figure 2. Average number of trips per e-Scooter per hour for all weekdays.

The mean rent duration for the different cities is compared in Figure 2, where Paris seems to have the longest mean trips of the cities analyzed. Paris is the city with the largest Built-Up Land area[1]. In Figure 3, the cities are ordered (descending) from left to right of Built-Up Land area. As can be seen, there seems to be no relation between built-up land area and e-scooters rental duration.

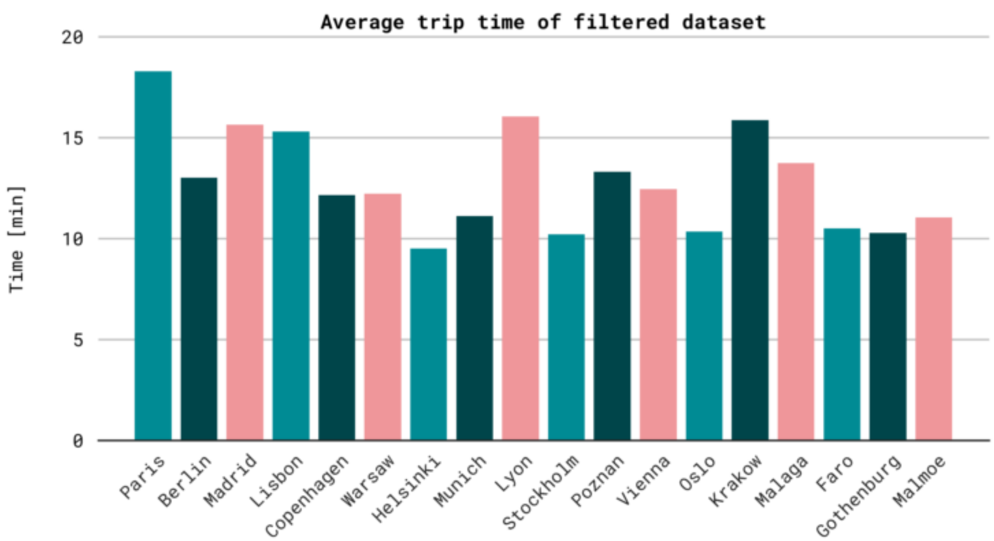

Figure 3. Average trip time for the different cities ordered (descending) from left to right of Built-Up Land area.

References

Anselin, L. (2001). Spatial econometrics. *A companion to theoretical econometrics*, *310330*.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, *10*(7), e0130140

Badeau, A., Carman, C., Newman, M., Steenblik, J., Carlson, M., & Madsen, T. (2019). Emergency department visits for electric scooter-related injuries after introduction of an urban rental program. *The American Journal of Emergency Medicine, 37*(8), 1531-1533.

Bai, S., & Jiao, J. (2020). Dockless E-scooter usage patterns and urban built Environments: A comparison study of Austin, TX, and Minneapolis, MN. *Travel behaviour and society*, *20*, 264-272.

Becker, H., Ciari, F., & Axhausen, K. W. (2017). Modeling free-floating car-sharing use in Switzerland: A spatial regression and conditional logit approach. *Transportation Research Part C: Emerging Technologies*, *81*, 286-299.

Bekhit, M. N. Z., Le Fevre, J., & Bergin, C. J. (2020). Regional healthcare costs and burden of injury associated with electric scooters. *Injury, 51*(2), 271-277.

Blomberg, S. N. F., Rosenkrantz, O. C. M., Lippert, F., & Christensen, H. C. (2019). Injury from electric scooters in Copenhagen: a retrospective cohort study. *BMJ open, 9*(12): e033988.

Brown, A., Klein, N. J., Thigpen, C., & Williams, N. (2020). Impeding access: The frequency and characteristics of improper scooter, bike, and car parking. *Transportation research interdisciplinary perspectives, 100099.*

Caspi, O., Smart, M. J., & Noland, R. B. (2020). Spatial associations of dockless shared e-scooter usage. *Transportation Research Part D: Transport and Environment, 86, 102396.*

Engdahl H, Englund C, Faxer A, Habibi S, Petterson S, Sprei F, Voronov A, Wedlin J, (2020). Electric Scooters' Trip Data Collection and Analysis. *33rd Electric Vehicle Symposium (EVS33).*

Fang, K., Agrawal, A. W., Steele, J., Hunter, J. J., & Hooper, A. M. (2018). Where Do Riders Park Dockless, Shared Electric Scooters? Findings from San Jose, California. *Mineta Transportation Institute Publications.*

Fitt, H., & Curl, A. (2019). E-scooter use in New Zealand: Insights around some frequently asked questions. *Retrieved August*, *21*, 2020.

James, O., Swiderski, J. I., Hicks, J., Teoman, D., & Buehler, R. (2019). Pedestrians and e-scooters: An initial look at e-scooter parking and perceptions by riders and non-riders. *Sustainability, 11(20), 5591*.

Hollingsworth, J., et al. (2019). "Are e-scooters polluters? The environmental impacts of shared dockless electric scooters." *Environmental Research Letters 14(8): 084031*.

Jiao, J., & Bai, S. (2020). Understanding the shared E-scooter travels in Austin, TX. *ISPRS International Journal of Geo-Information*, *9*(2), 135.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Li, A., Zhao, P., Haitao, H., Mansourian, A., & Axhausen, K. W. (2021). How did micro-mobility change in response to COVID-19 pandemic? A case study based on spatial-temporal-semantic analytics. *Computers, Environment and Urban Systems*, *90*, 101703.

Li, A., Zhao, P., Liu, X., Mansourian, A., Axhausen, K. W., & Qu, X. (2022). Comprehensive comparison of e-scooter sharing mobility: Evidence from 30 European cities. *Transportation Research Part D: Transport and Environment*.

Liu, M., Seeder, S., & Li, H. (2019). Analysis of E-Scooter Trips and Their Temporal Usage Patterns. Institute of Transportation Engineers. *ITE Journal, 89(6), 44-49.*

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Maxwell, B. (2019) Shared Electric Scooters and Transportation Equity: A Multivariate Spatial Regression Analysis of Environmental Factors on Revealed Travel Behavior and Mode Shift Potential.

McKenzie, G. (2019). Spatiotemporal comparative analysis of scooter-share and bike-share usage patterns in Washington, DC. *Journal of transport geography, 78, 19-28.*

Møller, T. H., & Simlett, J. (2020). Micromobility: moving cities into a sustainable future. EY.

Younes, H., Zou, Z., Wu, J., & Baiocchi, G. (2020). Comparing the Temporal Determinants of Dockless Scooter-share and Station-based Bike-share in Washington, DC. *Transportation Research Part A: Policy and Practice, 134, 308-320.*

Zuniga-Garcia, N., & Machemehl, R. (2020). Dockless electric scooters and transit use in an urban/university environment. In *99th Annual Meeting of the Transportation Research Board, Washington, DC*.