Car fleet synthesis for agent-based mobility models: a comparison of machine learning and discrete choice methods

Marjolaine Lannes[1, 2], Nicolas Coulombel[1], and Yelva Roustan[2]
[1]LVMT, École des Ponts, Université Gustave Eiffel, Marne-la-Vallée, France
[2]CEREA, École des Ponts, EDF R&D, Marne-la-Vallée, France
Correspondence: Marjolaine Lannes (marjolaine.lannes@enpc.fr), Nicolas Coulombel (nicolas.coulombel@enpc)

As various countries and metropolises seek to improve air quality by enforcing vehicle fleet regulation measures, better understanding the determinants of household car ownership and vehicle type choice is increasingly important. The European Union has as a matter of fact adopted pollutant emission standards for car manufacturers and many European countries have also established Low Emission Zones (LEZ) which limit traffic within a restricted zone. LEZs and their impacts have been extensively assessed through the use of agent-based mobility models (e.g. Dias, Tchepel, and Antunes 2016; Fosset et al. 2016; de Bok, Tavasszy, and Thoen 2022), including MATSim (e.g. Adnan et al. 2021). The representation of a disaggregated vehicle fleet at the household level would thus provide better inputs for MATSim, with the prospect of calculating traffic-related daily emission profiles based on a synthetic population and a synthetic vehicle fleet and assessing prospective scenarios for LEZs.

Despite the extensive literature on modeling car ownership, little research has been conducted on the type of fuels used in cars and even less on their age or emissions standards. Moreover, car ownership and vehicle type choice are usually estimated from household characteristics using discrete choice models (Purvis 1994; Jong et al. 2004). But recent studies point out the contribution of machine learning methods for transportation choice modeling (van Cranenburgh et al. 2021) and in particular for the estimation of car ownership (e.g. Paredes et al. 2017; Kaewwichian, Tanwanichkul, and Pitaksringkarn 2019; Dixon et al. 2021).

This work investigates the performance of several classification models in the prediction of vehicle ownership and vehicle type at the household level. We compare a discrete choice model against various machine learning classification methods (e.g. Gradient Boosting, Random Forest) for the estimation of household car ownership, fuel type and car pollutant emissions standards. Explanatory variables include household socio-economic characteristics (household type, income) as well as local and metropolitan accessibility variables (parking availability, public transport accessibility). The methodology is applied to the Paris region, using the “Enquête Globale de Transport” mobility survey. Considering the Matthew Correlation Coefficient, F1 score and Cohen’s kappa as evaluation metrics, we conclude that logistic regression slightly outperforms artificial intelligence models for car ownership whereas Gradient Boosting classifier gets the best results for vehicle type estimation. Our results show a strong relationship for car ownership prediction and a slight agreement for fuel type and emission standard predictions, with a preponderant importance of household composition and accessibility variables.

Bibliography:

Adnan, Muhammad, Fatma Outay, Shiraz Ahmed, Erika Brattich, Silvana di Sabatino, and Davy Janssens. 2021. 'Integrated Agent-Based Microsimulation Framework for Examining Impacts of Mobility-Oriented Policies'. *Personal and Ubiquitous Computing* 25 (1): 205–17. https://doi.org/10.1007/s00779-020-01363-w.

Bok, Michiel de, Lorant Tavasszy, and Sebastiaan Thoen. 2022. 'Application of an Empirical Multi-Agent Model for Urban Goods Transport to Analyze Impacts of Zero Emission Zones in The Netherlands'. *Transport Policy*. https://doi.org/10.1016/j.tranpol.2020.07.010.

Cranenburgh, Sander van, Shenhao Wang, Akshay Vij, Francisco Pereira, and Joan Walker. 2021. 'Choice Modelling in the Age of Machine Learning - Discussion Paper'. *Journal of Choice Modelling*, December, 100340. https://doi.org/10.1016/j.jocm.2021.100340.

Dias, Daniela, Oxana Tchepel, and António Pais Antunes. 2016. 'Integrated Modelling Approach for the Evaluation of Low Emission Zones'. *Journal of Environmental Management* 177 (July): 253–63. https://doi.org/10.1016/j.jenvman.2016.04.031.

Dixon, James, Sofia Koukoura, Christian Brand, Malcolm Morgan, and Keith Bell. 2021. 'Spatially Disaggregated Car Ownership Prediction Using Deep Neural Networks'. *Future Transportation* 1 (1): 113–33. https://doi.org/10.3390/futuretransp1010008.

Fosset, Pierre, Arnaud Banos, Elise Beck, Sonia Chardonnel, Christophe Lang, Nicolas Marilleau, Arnaud Piombini, et al. 2016. 'Exploring Intra-Urban Accessibility and Impacts of Pollution Policies with an Agent-Based Simulation Platform: GaMiroD'. *Systems* 4 (1): 5. https://doi.org/10.3390/systems4010005.

Jong, Gerard De, James Fox, Andrew Daly, Marits Pieters, and Remko Smit. 2004. 'Comparison of Car Ownership Models'. *Transport Reviews* 24 (4): 379–408. https://doi.org/10.1080/0144164032000138733.

Kaewwichian, Patiphan, Ladda Tanwanichkul, and Jumrus Pitaksringkarn. 2019. 'Car Ownership Demand Modeling Using Machine Learning: Decision Trees and Neural Networks'. *International Journal of GEOMATE* 17 (62): 219–30. https://doi.org/10.21660/2019.62.94618.

Paredes, Miguel, Erik Hemberg, Una-May O'Reilly, and Chris Zegras. 2017. 'Machine Learning or Discrete Choice Models for Car Ownership Demand Estimation and Prediction?' In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 780–85. https://doi.org/10.1109/MTITS.2017.8005618.

Purvis, Charles L. 1994. 'Using 1990 Census Public Use Microdata Sample to Estimate Demographic and Automobile Ownership Models'. *Transportation Research Record* 1443: 21.