Machine Learning based Microsimulation Approach for the Spatial Distributions of Automated Vehicle Preferences

Keywords: Preferences, Shared Automated Vehicles, Private Automated Vehicles, Machine Learning, Population Synthesizing

Introduction

AV technology is positioned to transform the existing urban systems (Fagnant & Kockelman, 2015). The trajectory of urban development may vary significantly depending on users' adoption of such technology, especially their preferences towards two business models of AVs, i.e., SAVs and Privately-owned AVs (PAVs). A wealth of AV simulation studies suggests that SAVs are more sustainable compared with PAVs (Zhang et al., 2018). However, in reality, the impact of AVs still relies heavily on how many users will adopt SAVs (Gkartzonikas & Gkritza, 2019). Although several user preferences surveys and mode choice experiments have been conducted to understand key factors influencing the adoption of AVs, there remains a data gap regarding the spatial distributions of SAV and PAV adopters. Such knowledge is critical to fuel more robust AV simulations and support policymaking to encourage the adoption of SAVs. Motivated by the aforementioned research need, in this study, we developed a generalizable microsimulation framework to synthesize the neighborhood-level preferences towards SAVs and PAVs. The work will serve as an anchor point to merge the existing AV preference and simulation studies. The rest of the paper is organized as follows.

Methods

The proposed three-step microsimulation framework is developed based on data from three sources, including (1) National AV Preference Survey (NAVPS), (2) 2017 National Household Travel Survey (NHTS), and (3) American Community Survey (ACS) data. NAVPS is an attitudinal online survey, which collected a nationally representative sample in September 2018 by Wang, Jiang, Noland, and Mondschein (Wang et al., 2019). The survey asked attitudinal questions about the preferences towards PAVs and SAVs. NAVPS also collected information about respondents' socioeconomic, demographic, attitudinal, and behavioral attributes. The dataset includes 721 completed responses. A series of household and individual-level attributes are shared by the NAVPS and 2017 NHTS datasets. The 2017 NHTS data, however, do not include preferences towards any business model of AVs. Lastly, the ACS data are released by U.S. Census Bureau and provide marginal distributions of various socioeconomic and demographic attributes at different geographic unit levels.

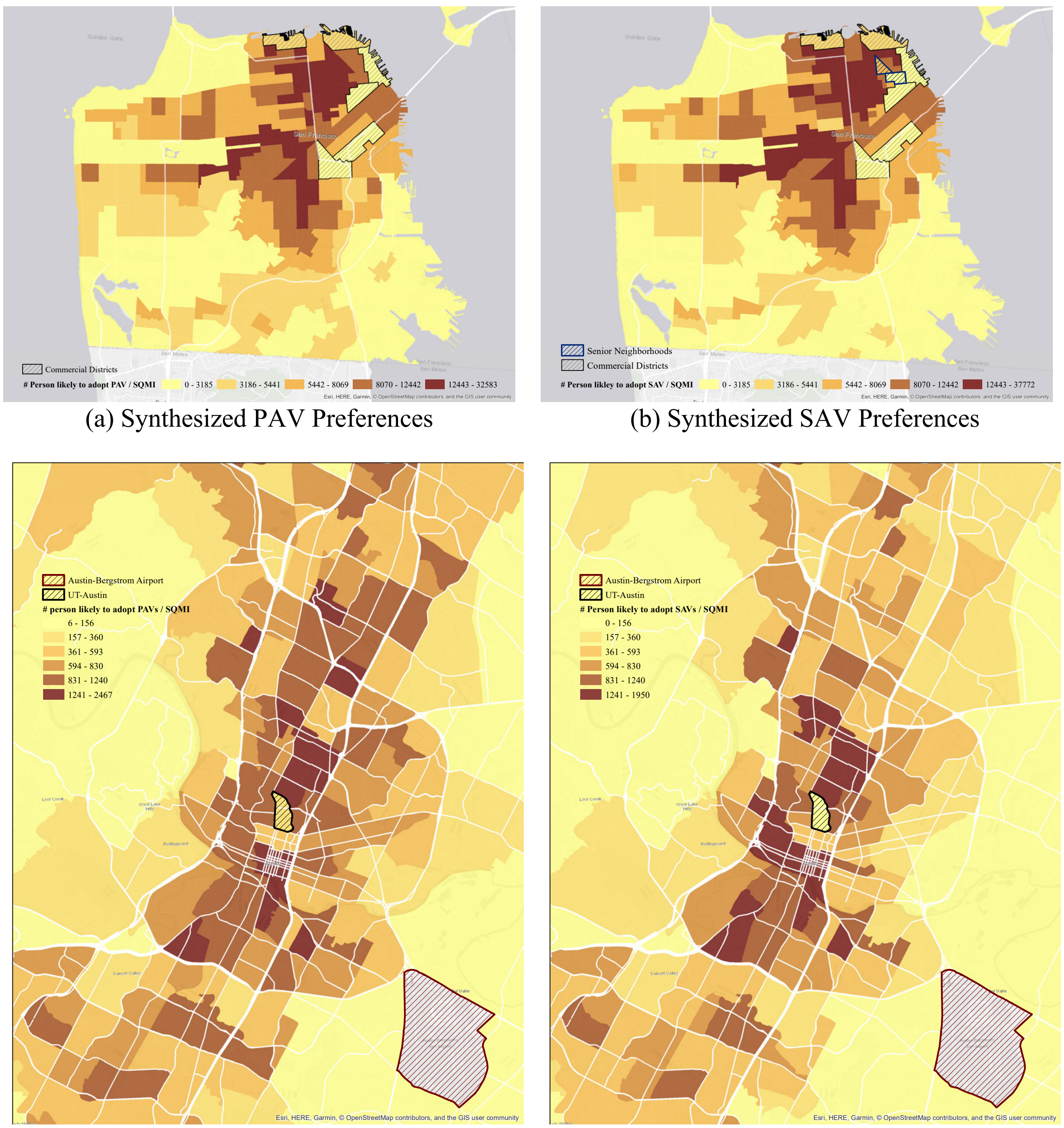
In the proposed three-step microsimulation, we will first train machine learning models to predict PAV and SAV preferences using the NAVPS data, with variables shared by the 2017 NHTS data. Second, we will apply the best-trained model to impute PAV and SAV preferences for each adult (i.e., above 18) in the NHTS person dataset. Finally, we synthesize populations using the imputed NHTS data as the seed matrix and ACS data as the marginal controls. We can eventually obtain the spatial distribution of PAV and SAV adoption preferences at the geographic units that are used to generate the synthetic population.

Machine Learning PAV and SAV Preferences. We used various machine learning classifiers to impute the willingness to adopt PAVs and SAVs using common variables shared by NAVPS and NHTS data. Classifiers, including Support Vector Classifier, Logistic Regression, Ridge Classifier, Random Forest Classifier, and Gradient Boosting Classifier, are trained using the 10-fold cross-validation method. The most predicative models are then applied to NHTS person data to impute the likelihood to adopt PAVs and SAVs. The NHTS households and person data (with imputed PAV and SAV preferences) are subsequently used as seed matrices for population synthesizing.

Population Synthesizing. The final NHTS data outputs from the previous step, together with the marginal distribution of socioeconomic variables obtained from ACS data, are then used to generate a synthetic population of households. This step is implemented in software PopGen 1.1 (Konduri et al., 2016), controlling socioeconomic and demographic variables at both household and individual levels. The final synthetic households and population include both the controlled variables, the existing ride-hailing behavior, and preferences for PAVs and SAVs. The neighborhood level AV preferences and ride-hailing users are then obtained by aggregating the synthesized results.

Results

As proof of concept, we applied the proposed machine learning-based microsimulation approach to two study areas (i.e., San Francisco, CA and Austin, TX) to validate the model. This is because the selected areas are decently sampled in the 2017 NHTS, providing robust seed matrices for population synthesis. Second, the spatial distribution of ride-hailing trips is available, which will be used to validate the synthesized ride-hailing users.



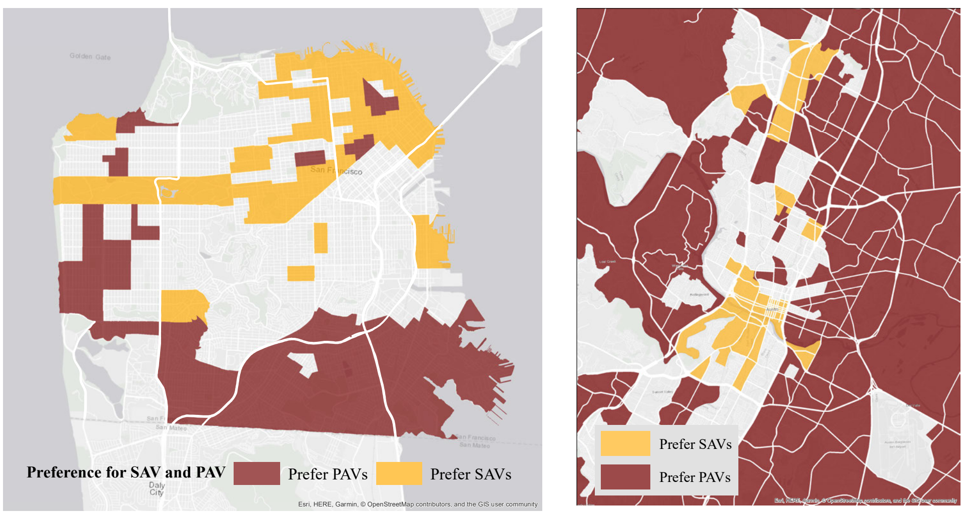
(a) Synthesized PAV Preferences (b) Synthesized SAV Preferences

(c) Synthesized PAV Preferences (d) Synthesized SAV Preferences

FIGURE 1: Austin Synthetic PAV and SAV Results

Both SAV and PAV machine learning models achieved classification accuracies of 0.79, using balanced datasets. Figure 1.a and b show the synthesized PAV and SAV adopters per square miles in San Francisco, CA. Results suggest that people who are willing to use AVs may show interests to both PAVs and SAVs. Senior neighborhoods are less likely to adopt SAVs, which aligns with prior SAV preference survey results (Haboucha et al., 2017; Krueger et al., 2016). The density of PAV and SAV adopters for Austin, TX are displayed in Figure 1.c and d. The density of early adopters in Austin-Bergstrom International Airport (red dashed area) is zero, as there is no residential population in the census tract. Meanwhile, the density of synthesized adopters is also low in the University of Texas, Austin area (black dashed area), because group quarter is not synthesized in this study.

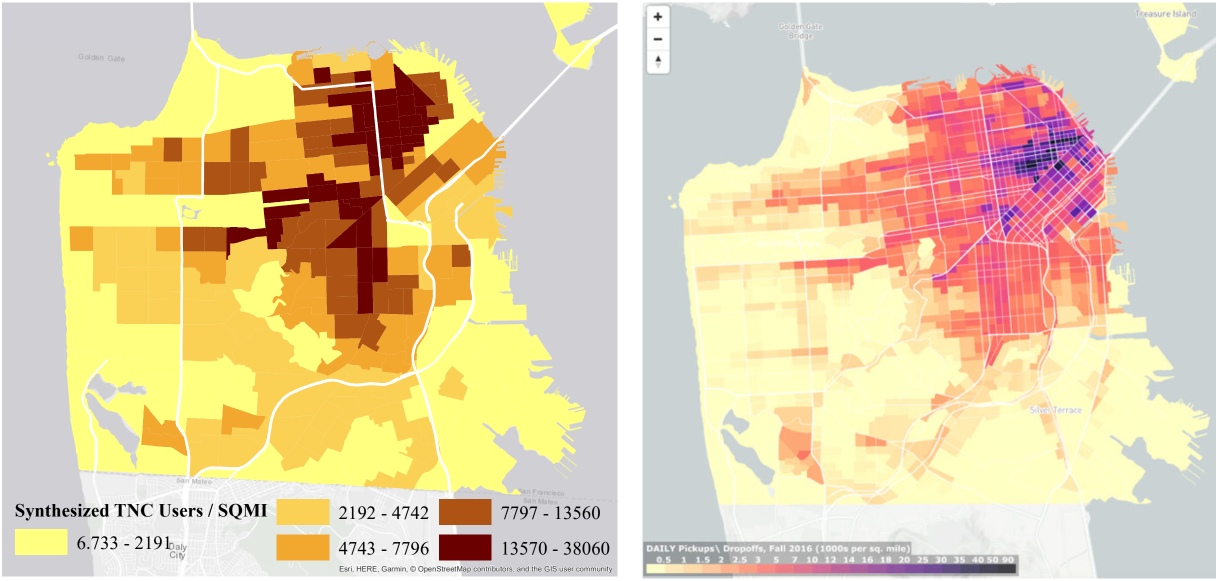
The synthesized results also suggest residents from suburban neighborhoods are more likely to adopt PAVs than SAVs, as shown in Figure 2. The red areas are Census Tracts, where the number of PAV adopters is 10% more than the SAV adopters, and the yellow areas indicate places where there are 10% more SAV adopters. Such results are consistent with the existing SAV preference surveys (Gkartzonikas & Gkritza, 2019).



(a) San Francisco, CA (b) Austin, TX

FIGURE 2: Preferences for PAVs vs. SAVs

We validated our model results by comparing the density of synthesized ride-hailing users with the density of ride-hailing trips in the two study areas. Figure 3 shows a comparison for the San Francisco study area. The observed ride-hailing distribution is developed by the San Francisco County Transportation Authority (SFCTA) and published online (Castiglione et al., 2016). The distributions of the synthesized users and trip generation follow the same trend, except for the commercial zones. The results for Austin are similar. This to some extent validates our modeling approach.



(a) Synthesized Ride-hailing Users Density

(b) Observed Ride-hailing Trip Density (Castiglione et al., 2016)

FIGURE 3: Synthesized Ride-hailing Users and Trip Density in San Francisco

Conclusion

This study develops a machine learning-based microsimulation model to synthesize neighborhood-level preference towards PAVs and SAVs. The model can be potentially applied to any major U.S. cities, with sufficient NHTS sampled households or local travel survey data with shared variables. The model incorporates machine learning and population synthetizing to estimate the number of adults (above 18) who may be interested in the adoption of PAVs and SAVs. The synthetic population produced from this model closes the existing data gaps for neighborhood-level PAV and SAV preferences and can fuel the AV simulation studies for more robust AV impact analysis. The model is implemented and validated using data from San Francisco, CA, and Austin, TX areas.

Acknowledge

Wenwen Zhang and Kaidi Wang are partially funded by 4VA collaborative research grant to enable this study.

References

Castiglione, J., Chang, T., Cooper, D., Hobson, J., Logan, W., Young, E., Charlton, B., Wilson, C., Mislove, A., Chen, L., & others. (2016). TNCs today: A profile of San Francisco transportation network company activity. *San Francisco County Transportation Authority (June 2016)*.

Fagnant, D. J., & Kockelman, K. (2015). Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, *77*, 167–181. https://doi.org/10.1016/j.tra.2015.04.003

Gkartzonikas, C., & Gkritza, K. (2019). What have we learned? A review of stated preference and choice studies on autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *98*, 323–337.

Haboucha, C. J., Ishaq, R., & Shiftan, Y. (2017). User preferences regarding autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *78*, 37–49.

Konduri, K. C., You, D., Garikapati, V. M., & Pendyala, R. M. (2016). Application of an Enhanced Population Synthesis Model that Accommodates Controls at Multiple Geographic Resolutions. *Transportation Research Record, Journal of the Transportation Research Board*, *2563*, 40–50.

Krueger, R., Rashidi, T. H., & Rose, J. M. (2016). Preferences for shared autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *69*, 343–355.

Wang, S., Jiang, Z., Noland, R. B., & Mondschein, A. S. (under review). Attitudes towards privately-owned and shared autonomous vehicles.

Zhang, W., Guhathakurta, S., & Khalil, E. B. (2018). The impact of private autonomous vehicles on vehicle ownership and unoccupied VMT generation. *Transportation Research Part C: Emerging Technologies*, *90*, 156-165.