# GLOBAL URBAN TYPOLOGY DISCOVERY WITH A LATENT CLASS CHOICE MODEL

**Yafei Han** (Corresponding Author)
MIT, Department of Civil & Environmental Engineering
77 Massachusetts Avenue, Room 10-463, Cambridge MA 02139
Email: yafei@mit.edu; Phone: 1-857-600-8423

**Jimi Oke**
MIT, Department of Civil & Environmental Engineering
77 Massachusetts Avenue, Room 1-249, Cambridge MA 02139
Email: oke@mit.edu; Phone: 1-617-253-9729

**Sean Hua**
MIT, Intelligent Transportation Systems Lab
77 Massachusetts Avenue, Cambridge MA 02139
Email: bosshua@mit.edu

**Jin Zhou**
Department of Statistics, Columbia University, New York, NY, United States
Email: jz2763@columbia.edu; Phone: 1-617-253-9729

**Carlos Lima Azevedo**
MIT, Department of Civil & Environmental Engineering
77 Massachusetts Avenue, Room 1-278, Cambridge MA 02139
Email: cami@mit.edu; Phone: 1-617-452-2482

**Christopher Zegras**
MIT, Department of Urban Studies & Planning
77 Massachusetts Avenue, Room 10-403, Cambridge MA 02139
Email: czegras@mit.edu; Phone: 1-(617) 452-2433

**Joseph Ferreira**
MIT, Department of Urban Studies & Planning
77 Massachusetts Avenue, Room 9-532, Cambridge MA 02139
Email: jf@mit.edu; Phone: 1-(617) 253-7410

**Moshe Ben-Akiva**
MIT, Department of Civil & Environmental Engineering
77 Massachusetts Avenue, Room 1-181, Cambridge, MA, 02139
Email: mba@mit.edu

Submission Date: Nov 15, 2017
Extended abstract word counts: 1,653
**Acknowledgement:** This work was supported through the MIT Energy Initiative's Mobility of the Future study. We also thank Dalia Research for generating and supplying the mobility survey data for 52 countries.

## INTRODUCTION

Cities around the world are experiencing rapid changes driven by advances in technology and major economic and behavioral shifts. Understanding the fundamental dynamics of urban mobility of cities around the world - its differences and similarities – is a key factor in the selection of future behavior-sensitive policies, technological solutions and overall urban scenarios of interest. This requires discovering urban typologies based on the latest global city data and travel behavior data.

The conventional method for city classification is exploratory, such as factor analysis and/or clustering, with typology characterized only by city-level attributes with various focus. Martin et al. (1) classify 300 U.S. cities and find 8 clusters based on 15 socioeconomic indicators. Huang et al. (2) classify 77 metropolitan areas worldwide by urban form using satellite images. Louf and Barthelemy (3) focus on street patterns and obtain 4 clusters out of 131 cities. Mobility-oriented city classification is rare. Priester et al. (4) cluster 41 megacities across the globe based on 59 mobility-related indicators from the 1995 UITP Millennium Cities Database (5). They extract 13 factors and obtain 6 clusters. This study is useful to understand current mobility types, but it covers only megacities and lacks indicators such as economy, climate, built environment, and more importantly, behavior indicators that can affect future mobility.

We proposed a novel supervised approach for city classification: a Latent Class Choice Model (LCCM) that identifies latent classification of cities and its relationship with multiple observed/stated individual travel choice indicators. After applying our proposed framework to a large database of urban and individual-specific behavioral indicators for 331 worldwide cities, we discover 10 latent city classes with distinct city characteristics and travel-related choice preferences. Our model also provides a probabilistic class membership assignment for each city. Comparing LCCM results with traditional exploratory clustering results, we find 2/3 of the cities stay in the same class while the rest move to a different class. The changes in class membership suggest that incorporating individual behavior helps update belief about a city's class membership.

## METHODOLOGY

The LCCM consists of a class membership model and 9 discrete mobility choice models. The class membership model uses city attributes to predict the latent class. It is parameterized as a logit model. The choice models predict individual mobility choices conditioned on the class membership of a person's city and person attributes. They all take the form of logit model.

The LCCM has several advantages. First, it is supervised such that the latent classification is optimized using the observed/stated choice indicators. Second, travel behavior segmentation drives latent city class structure. Preference parameters specific to each class are obtained at the same time. Thirdly, this method generates a probabilistic class membership for each city, which provides richer profiles and continuous spectrum of variation than a single cluster assignment. Also, this predictive model is free from problematic distance metrics involving mixed variable types in clustering analysis.

Lastly, this structure allows for updating posterior belief about class membership over time given more individual data from a city.

## DATA

We collected 66 indicators of 331 cities from open-data sources[1]. City indicators cover population, geography, economy, transport networks, mode shares, traffic, technology, environment quality, among others, for the years of 2015 and 2016. Individual behavior data was obtained from a global mobile-phone survey, including 44,000 individual observations collected worldwide from 52 countries (6). 225 cities with surveyed individuals matched open-data city set. The survey provides individual socio-demographic information, current mobility choices and future vehicle technology preferences.

## FINDINGS

We first conduct a traditional exploratory analysis of 331 cities. We obtained 6 factors from a mixed variable factor analysis: metro propensity, bus rapid transit (BRT) propensity, efficiency and equity, auto-dependence and industrialization, rapid growth and congestion. From these, we obtain 10 preliminary clusters using Ward's hierarchical algorithm. The clustering results help initialize parameters for the LCCM.

We then use E-M (Expectation-Maximization) algorithm to estimate the LCCM and tested various model specifications (e.g. number of latent classes, explanatory variables etc). The choice indicators include car availability, commuting mode choices, weekday travel time, weekday miles of driving, propensity to buy electronic vehicle and use self-driving car. Our final model discovers 10 latent classes. The latent class centroids indicate that the 10 classes are well distinguished by population size, density, GDP, population growth, mode shares, subway, Bus Rapid Transit (BRT) and bike-share program availability. The 10 classes are summarized as below:

- Class 1: Medium to small population, low density, rich, car dependent (many U.S. cities, such as Birmingham, Louisville, Edmonton)
- Class 2: Large population, rich, high car share, well developed subway system (e.g. Chicago, New York, Philadelphia, Boston)
- Class 3: Medium to small population, medium density, advanced economy, mixed mode (e.g. Nagoya, Sapporo, Sendai)
- Class 4: Mega city, high density, advanced economy, extensive subway system and high subway ridership (e.g. Seoul, Singapore, Hong Kong)
- Class 5: Medium to small population, medium density, rich, mixed mode with higher car mode share and more extensive bike-share program than class 3 (e.g. Zurich, Nice, Frankfurt)
- Class 6: Large, high density, developing economy, high transit mode share, extensive BRT infrastructure and high BRT ridership (e.g. Quito, Belo Horizonte, Sao Paulo, Bogota)

---

[1] http://web.mit.edu/its-lab/www/dashboard/cities.html

- Class 7: Medium to small population, medium to low density, developing economy, relatively high car mode share and few transit (e.g. Pretoria, Durban, Johannesburg)
- Class 8: Mega city, high density, developing economy, high population growth, mixed mode with high transit and bike mode shares, expanding subway system, largest bike-share program (e.g. Shanghai, Beijing, Hangzhou, Shenzhen)
- Class 9: Medium population, high density, less developed economy, low population growth, few transit systems (e.g. Medellin, Alexandria, Casablanca)
- Class 10: Largest population, high density, high population growth, least developed economy, few transit systems (e.g. Mumbai, Ahmedabad, Surat)

Compared to exploratory clustering results, we find about 2/3 of the cities stay in the same class (according to highest probability prediction by LCCM), which confirms the exploratory analysis. Including individual travel behavior information however, changed some cities' class membership. For example, the LCCM moves Houston, Seattle and San Diego from class 2 (mostly U.S. cities with developed metro systems) to class 1 (car-dependent U.S. cities with few metros). Interestingly, the model also moves Chinese cities from the original class 6 and 7 to class 8 – a Chinese city dominant class. This result suggests that the model classifies cities with similar travel preferences into the same group. Our model also suggests interesting differences in people's tendency to buy electric vehicle and use autonomous vehicle. For example, class 8 (a Chinese city dominant class) and class 2 (mostly U.S. cities with developed metro) exhibits the most interest in purchasing electronic vehicle, while class 1 (U.S. auto-dependent cities) show lower interest than other classes. In terms of how soon a person consider autonomous vehicle will be safe enough to use, class 6 and 9 are most unsure. Class 4 and 8 are the most optimistic.

## CONCLUSIONS

We build a LCCM to learn a probabilistic city classification for 225 global cities using both global city data and individual travel choice data. We find the model has the potential to classify cities based on travel behavior segmentation. This city classification also reveals variations in preferences for future vehicle technologies. Some limitations of this study include a lack of data for certain city indicators (e.g. road traffic and mode share), which is filled by average values; and a measurement of individual choices from the global survey not validated for representativeness by the authors. Yet, the probabilistic city class membership and the class-specific behavior parameters obtained are useful for predicting future mobility behaviors for various types of cities around the world. This model can also be used to generate city prototypes based on a mixture of existing cities for urban simulation and scenario discovery.

## References

[1] Martin, M. K., K. M. Carley, and N. Altman, 300 Cities - *An Exploration in Characterizing US Cities*. CASOS technical report CMU-ISR-08-122, Institute for Software Research, Carnegie Mellon University, Pittsburgh, PA 15213, 2008.

[2] Huang, J., X. Lu, and J. M. Sellers, *A global comparative analysis of urban form:*

*Applying spatial metrics and remote sensing*. Landscape and Urban Planning, Vol. 82, No. 4, 2007, pp. 184–197.

[3] Louf, R. and M. Barthelemy, *A typology of street patterns*. Journal of The Royal Society Interface, Vol. 11, No. 101, 2014.

[4] Priester, R., J. Kenworthy, and G. Wulfhorst, *The Diversity of Megacities Worldwide: Challenges for the Future of Mobility*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 23–54, 2013.

[5] Kenworthy, J. and F. Laube, *The millennium cities database for sustainable transport: (CDROM Database)*. International Union of Public Transport (UITP), Brussels, 2001.

[6] Dalia Research, *How the world moves: the future of mobility & global transport*, 2017.