Proceedings of the 2020 Winter Simulation Conference
K.-H. Bae, B. Feng, S. Kim, S. Lazarova-Molnar, Z. Zheng, T. Roeder, and R. Thiesing, eds.

SYNTHETIC TRIP LIST GENERATION FOR LARGE SIMULATIONS

Ben Frederick Martir Intoy
George Panteras
Trey LaNasa

Deloitte Consulting LLP
1919 North Lynn Street
Arlington, VA 22209, USA

Kevin Liberman

Deloitte Consulting LLP
200 Berkeley Street
Boston, MA 02116, USA

ABSTRACT

Decision-makers use large simulations to plan the future of complex systems. Here we use the example of transportation networks where the accuracy of the simulation model is highly dependent upon an accurate representation of the users' behavior in the network. Such user data may be sparse, private, or difficult to obtain and would have to be generated synthetically using available data. We present a method to synthetically generate user travel schedules which are then used in a massive scale agent-based model to make informed decisions and their future impacts.

1 INTRODUCTION

Through innovations in data granularity and computing capabilities, simulations of large systems have become increasingly complex in nature. Novel methods for creating synthetic populations are essential to pushing the boundaries of digital twins and the accuracy of such simulations. With an accurate way to generate synthetic populations, modelers can dynamically test the effects of "what-if?" scenarios with high fidelity input data that drives their results. As an example, we present a method that uses raw spatiotemporal data to generate user schedules. These schedules are then used by agents in a massive scale agent-based model (ABM) for traffic modeling. A full scale agent-based model driven by a realistic population model can then be used to forecast possible different scenarios and their effects in the real world.

2 FROM SPATIOTEMPORAL TO INDIVIDUAL TRIP SCHEDULES

For the purpose of simulating realistic human behavior, synthetic travel patterns are constructed for individuals (users) in a metropolitan area based on cellular geolocation data. These patterns take the form of individual schedules of stationary locations with associated timestamps. The generated schedules, collectively known as trip lists, are then used in a cloud-based simulation environment to model traffic behavior. To create trip lists for each individual we take inspiration from a process that uses spatiotemporal data derived from cellular call detail records (CDR) to train an input-output hidden Markov model (IOHMM) (Yin et al. 2017). Nevertheless, our methodology (Baeder et al. 2019) differs from previous implementations in two important ways. First, we use higher geospatial resolution of individual positions from anonymized mobile location service data. This data is more frequently collected and precise compared to CDR. Second is that the generated schedules are fed into an agent-based model. This way we create a simulated urban environment that captures behavior effects stemming from environmental changes (e.g., adding or removing roads on which cars travel).

In order to derive a trip list, we built a pipeline that consists of four steps: (1) identifying user stay locations based on a density-based clustering algorithm; (2) grouping users' stay locations by temporal characteristics using a Gaussian mixture model to infer location meaning; (3) assigning user stay locations to Traffic Analysis Zones; (4) and lastly, generating individual activity sequences using an IOHMM.

Once a synthetic schedule is generated, the schedule's activities are mapped to primary and subsequent locations via a sampling procedure of residential and commercial areas. Lastly, we run sensitivity analyses to understand how slight variations in the trip list (e.g. 3% volume increase) will affect output metrics. We have added an ad-hoc process to create more synthetic activities by leveraging probability distribution functions fitted to origin-destination arcs segmented by trip purpose. We then sampled from these distributions to obtain new destination points. After an alternative trip list is finalized and selected, it is ready to be ingested into the simulation.

3 SIMULATION ENVIRONMENT AND LOGIC

Agents are assigned these generated schedules and placed into a simulation environment. To create the simulation environment for the agents to interact with, open- and crowd- sourced data sets are used. Roads, road lanes, speed limits, traffic signals, and stop signs are imported and built within the simulation. The agents/cars observe and comply with these structures as they travel between destinations. The agents are also able to interact with each other with car physics and movement based on others around them. Within the simulation, perturbations can be performed such as adding/removing a new road(s) and/or changing the behavior of a traffic light. These perturbations can be interpreted as experiments a planner can perform to assess the effectiveness of traffic infrastructure changes.

4 ANALYSIS OF THE SIMULATION RESULTS

The results of the simulations can be utilized to answer a plethora of questions depending on the scope of the problem. A leading application of this approach is first-order analysis of future transportation networks, policies, and modalities. Specifically, we perform "what-if" analyses to predict how future changes in the transportation network can reduce congestion and greenhouse gas emissions. Changes are expressed and visualized via a baseline of appropriate metrics, such as vehicle miles traveled and transit ridership. When rapid innovation in mobility and exogenous factors like COVID-19 disrupt travel supply and demand, decision-makers must plan and act quicker than allowed by traditional survey-based methods of synthetic population modeling. Our methodology of synthetic trip list generation coupled with agent-based simulation combines long-term sketch planning with iterative and flexible model design and output.

ACKNOWLEDGMENTS

As used in this document, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting. This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor. Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

REFERENCES

Baeder, D., E. Christensen, A. Doanvo, A. Han, B. F. M. Intoy, S. Hardy, Z. Humayun, M. Kain, K. Liberman, A. Myers, M. Patel, W. J. Porter, L. Ramos, M. Shen, L. Sparks, A. Toriel, and B. Wu. 2019. "Massive-Scale Models of Urban Infrastructure and Populations". In *Social, Cultural, and Behavioral Modeling*, edited by R. Thomson, H. Bisgin, C. Dancy, and A. Hyder, 113–122. Cham: Springer International Publishing.

Yin, M., M. Sheehan, S. Feygin, J.-F. Paiement, and A. Pozdnoukhov. 2017. "A Generative Model of Urban Activities from Cellular Data". *IEEE Transactions on Intelligent Transportation Systems* 19(6):1682–1696.