Towards a framework for mobile phone data in MATSim

Michael Zilske

Technische Universität Berlin, Verkehrssystemplanung und Verkehrstelematik
michael.zilske@tu-berlin.de

July 18, 2017

1 Mobility models from phone data

Mobile phone data sets are seen as a candidate to replace formal travel surveys as an input to traffic models. Travel surveys are expensive to the point of only being done in large time intervals and small sample sizes, while the main cost component in acquiring mobile phone data sets seems to be privacy compliance. Mobile phone data sets are biased, noisy, and lack information that surveys contain, yet recent research (Jiang et al., 2016) indicates that by fixing a home location for each user, a daily commute for each worker, and person-specifically fitting a three-parameter model to the frequency and timing of the remaining trips, combining that with a destination model based on land use, and then simulating that model, one can reproduce many important features of the movement of individuals as well as the population.

This modeling framework does not require imputing any secondary activity types ("other" suffices), but it also does not generate modes of transport. It does not explicitly consider the transportation system. It does explicitly distinguish stays from trips, and it does impute trip properties for which evidence from the mobile phone trace is not used, notably start time and duration of work.

2 A direct simulation-based approach

In our exploration of how to integrate mobile phone data into MATSim, we started minimalistic and rooted in the traffic system. Instead of having pre-processing steps amounting to behavioral interpretation, such as detecting stays and trips with mode of transport, we create an initial demand as close to the raw input data as possible. The guiding hypothesis is that alternative behavioral interpretations of observations are structurally not very different from alternative choices. Algorithms which propose behavior consistent with observations should be replanning modules. The likelihood of behavior given an observation should be part of the scoring function. Replanning modules and scoring functions are parameterized by simulated network conditions.

More specifically, we put all sightings on the network, connect them by legs, and simulate the traffic flow. No explicit trip-stay-distinction is done, except for fusing successive sightings in the same cell. Intermediary sightings are initially treated just like sightings at activity locations. Since sparse chains of sightings leave uncertainty about the time the change of location takes place, agents randomly realize their traces temporally by picking departure times from each sighting location randomly so that their next location can be reached by the time it is witnessed there (Zilske and Nagel, 2015).

Because agents visit only locations for which there is direct evidence in form of a sighting, trips are missing (Zilske and Nagel, 2014). Using this construction, sparsely using the phone network causes sparsely using the traffic system. While in reality there may or may not be a correlation, it should not enter the model by using a biased construction. To account for this bias, we enrich traces of sparse users with additional locations. We implemented a simple, constructive, geometric approach for adding additional locations in patterns consistent with what is witnessed for more active users. That way, we do not discard the data generated by sparse users: Their simulated plans are still rooted at the times and locations for which there is direct evidence. We prototyped a simple trip/stay distinction, based on the heuristic that the less slack an agent has in choosing a departure time from a location, subject to the constraint of having to reach its next location in time, the more likely it is to be not at an activity location, but on a trip. Note that in this way, the trip/stay distinction is directly informed by simulated network travel times.

The random re-realization of traces is implemented as a MATSim replanning strategy [1]. Since at this point we do not label activities, and also the number of activities is variable, we do not use a behavioral scoring. We do use Cadyts to account for link volume counts as an additional input. This reduces temporal uncertainty from the sparseness of traces: The hourly link counts impose their time structure onto the random realization of the departure times.

3 MATSim providing traffic system constraints to algorithms

MATSim can be a framework for determining what imputations from a mobile phone trace leads to a feasible trip chain. For example, picking a departure time from work which is so late that a visit to another location for which there is direct evidence cannot be reached in time is infeasible. Consider any mode detection algorithm, which could well be added to the framework of (Jiang et al., 2016). If performed within the MATSim simulation loop, a trip made by car on a congested road may be discovered to be infeasible, while a high-speed train trip may be feasible. Model parameters can be optimized to produce trip chains which are both feasible and consistent with counts. We consider the implementations of trip/stay distinction and time choice which were sketched here to be placeholders for any such methods which could benefit from being in the loop with a traffic simulation, and from delaying the imputation of behavior until the constraints imposed by the traffic system are considered.

References

Jiang, S., Y. Yang, S. Gupta, D. Veneziano, S. Athavale, and M. C. González (2016). "The TimeGeo modeling framework for urban mobility without travel surveys". In: *Proceedings of the National Academy of Sciences* 113.37, E5370–E5378. DOI: 10.1073/pnas.1524261113.

Zilske, M. and K. Nagel (2014). "Studying the accuracy of demand generation from mobile phone trajectories with synthetic data". In: *Procedia Computer Science* 32, pp. 802–807. ISSN: 1877-0509. DOI: 10.1016/j.procs.2014.05.494.

— (2015). "A Simulation-based Approach for Constructing All-day Travel Chains from Mobile Phone Data". In: *Procedia Computer Science* 52, pp. 468–475. ISSN: 1877-0509. DOI: 10.1016/j.procs.2015.05.017.

[1] https://github.com/michaz/java8-matsim-playground