Authors Michael Zilske, Kai Nagel

Title **Calibration of an Agent-Based Travel Demand Model Constructed from Mobile Phone Data**

Track General Papers

Director Mark Wardman

Abstract

Introduction

The purpose of this work is to investigate replacing travel diaries with sets of call detail records (CDRs) as input data for an agent-oriented traffic simulation. Synthetic CDRs are used in order to study the effect of this substitution in isolation. An experimental design is used where a detailed synthetic transportation scenario with individual simulated travellers is combined with a model of mobile phone usage to collect synthetic CDRs. This set of artificial CDRs, combined with a number of link traffic counts from the same scenario, is then considered as input for another instance of the same traffic model, disregarding all other information. The model output is then compared to the initial scenario.

Simulation and Calibration

In the MATSim transport microsimulation, each traveler is modeled as an agent which can generate and choose from a set of plans. A plan is typically a travel diary with complete route information. All agents are simulated on the network, carrying out their plans. The result of the network simulation is fed back to the agent to re-assign choice probabilities to its plans. This process is iterated until an equilibrium is reached. The initial set of plans is a required input to the simulation, and is typically obtained from travel diaries.

While a travel diary contains start and end times of activities, a CDR data set only witnesses the presence of the participant at a certain point in time in a certain mobile phone cell. Trips during which no call is made are not witnessed. When a traffic demand model is constructed from a set of CDRs, assuming it can be scaled with respect to the sample size, the demand will likely be underestimated because of missing trips which are not covered by calls. Population segments with a lower mobile phone ownership are underrepresented[2]. Furthermore, there is uncertainty, in particular about start and end times of activities. The latter can be modeled by assigning several plans to an agent, each consistent with the set of sightings.

The Cadyts calibration system works by reassigning plan choice probabilities considering known traffic counts. It effectively directs the synthetic population towards choices more consistent with measurements. This can be seen as reducing uncertainty about aspects of behavior such as activity times, but the re-assigned choice probabilities also reflect which segments of the demand are underestimated: These plans will receive a higher probability offset[1].

We use synthetic CDRs to study the possibility of setting up a MATSim model directly from CDRs, with Cadyts reducing uncertainty by incorporating traffic counts, as well as identifying underestimated demand segments.

Synthetic CDRs

We start with a full implementation of an agent-oriented traffic model. The output of this model is a set of complete descriptions of mobility behavior of an agent population with labeled activities and space-time trajectories on the level of network links, annotated with mode of transport. We consider this the ground truth of a hypothetical scenario.

We then apply a model of cell phone usage to the synthetic population in order to obtain call detail records. The setup allows for any behavior to be plugged in, e.g. allowing different population segments with different calling habits, or the situation where all cell handovers are recorded. Similarly, traffic volume counts on an arbitrary number of links can be recorded at any time scale, reflecting varying availability of such data in real-world situations. The output of this step is a synthetic set of CDRs[3] and a synthetic set of link traffic counts. We consider this the available data for traffic modeling in the hypothetical scenario. This framework allows us to study methods for constructing demand models from CDRs, and how much information from CDRs and traffic counts is needed to re-approximate the state of the traffic system in the ground truth scenario to which degree. It isolates these questions from the different question of how good the traffic simulation model itself is at approximating reality.

Travel demand from CDRs

Each CDR trajectory is converted into one MATSim plan in a straightforward way. Every call is converted into an activity. Activities are connected by trips. Several calls in the same zone without a call in a different zone between them are fused, since there is no evidence of travel between them. Similarly, there is no evidence for detours, so the best guess at route choice is the least-cost route. The only degree of freedom is the departure time from each activity location. The simplest initial solution is to set the activity end time to the time of the last sighting at the activity location: the phone call is assumed to have taken place at the time the agent leaves the activity.

The resulting plans are simulated. The output of this step is of the same form as the ground truth scenario. The two scenarios can now be compared to asses the approximation quality

Scenario

We consider a realistic travel demand model generated from a household survey, which contains complete trip diaries from one specific day of 2\% of the Berlin population. We select all individuals who only travel by car and obtain 18377 individuals. We simulate the scenario and obtain a set of trajectories and link traffic counts. This is our base scenario.

Agents homogeneously place calls with uniform probability throughout the day, at a specified daily rate. We carry out parametric runs of our procedure, varying the call rate.

We define the missed total travelled distance in the network as the error measure. As expected, the error drops with increasing call rate, on account of fewer trips to and from activities missed by the sampling, as well as routes being traced more accurately. We find that even with a high average call rate of 20 calls per day throughout the population, the approach still misses about 15% of traffic, demonstrating the necessity of a compensation method[4].

Table 1. Reproduction of total travel distance.

rate	total travel distance [km]	relative to base
base	920928	1.0000
2	180407	0.1959
5	467898	0.5081
10	668082	0.7254
20	798749	0.8673
50	878649	0.9541
100	898857	0.9760
150	902353	0.9798

We now expand the agent population by adding several copies of each agent. These cloned agents are equipped with a second plan, which lets the agent stay at home. The choice probabilities of these two plans are calibrated by Cadyts. The cloned population effectively serves as a buffer population, which the calibrator uses to direct the demand towards matching the known link volume counts by calibrating the probability of each clone to be in the scenario. The final calibration weights can be analyzed to identify demand segments underestimated by the CDRs.

Discussion

Both MATSim and Cadyts have been used for large-scale, real-world scenarios. We expect that applying the process described above to realistic travel demand data will show that it can be used to re-scale demand segments which are underrepresented in CDR data. When using realistic call frequencies, we expect this in particular to be those travellers with activities of short durations such as shopping trips, as these will be more likely to be missed because they are not covered by calls.

References

[1] Flötteröd G, Chen Y, Nagel K. Behavioral Calibration and Analysis of a Large-Scale Travel Microsimulation. Networks and Spatial Economics. 2012; 12: 481-502.

[2] Wesolowski A, Eagle N, Noor AM, Snow RW, Buckee CO. The Impact of Biases in Mobile Phone Ownership on Estimates of Human Mobility. J R Soc Interface. 2013; 10: 20120986.

[3] Isaacman S, Becker R, Cáceres R, Martonosi M, Rowland J, Varshavsky A, Willinger W. Human Mobility Modeling at Metropolitan Scales. Proc. 10th International Conference on Mobile Systems, Applications, and Services. 2012; 239-252.

[4] Zilske M, Nagel K. Studying the Accuracy of Demand Generation from Mobile Phone Trajectories with Synthetic Data. Accepted for presentation at the 4th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTRANS).