Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 130 (2018) 920–925

Procedia Computer Science

www.elsevier.com/locate/procedia

The 9th International Workshop on Agent-based Mobility, Traffic and Transportation Models, Methodologies and Applications (ABMTrans 2018)

An Activity Based Demand Model for Large Scale Simulations

Mohammad Saleem[a, *], Oskar Blom Västberg[a], Anders Karlström[a]

[a]Department of Urban Planning and Environment, KTH Royal Institute of Technology, Stockholm, Sweden

Abstract

This paper presents the ongoing development of SCAPER, a random utility based travel demand model that consistently incorporates time decisions. The paper focusses on how SCAPER can be used for large scale simulations, and more specifically: 1. How computational speed of SCAPER is improved using sampling of locations, and how it influences the simulation results. 2. Interfacing SCAPER to MATSim simulation framework, and estimating the SCAPER model with travel times and travel costs produced by the simulation of Stockholm demand (simulated) over Stockholm network using MATSim. 3. Preliminary results from 1 and 2.

© 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.

Keywords: SCAPER, dynamic discrete choice model, travel demand model, estimation, sampling, MATSim

1. Introduction

This paper presents the interface of SCAPER with an agent based traffic simulation model (MATSim). SCAPER is a dynamic discrete choice based activity demand model. It generates full day activity schedules for travellers. The activity schedule may have any number of trips, each of which is a combination of six activities, twelve hundred and fourty locations and four modes. In SCAPER, individuals are assumed to make decisions sequentially in time about whether to stay or travel, starting from home in the morning and ending at home in the evening. The choice is made based on a random utility, calculated as the sum of one-stage utility of an action and the expected future utility in the reached state. In previous work, an estimation procedure has been proposed using sampling of alternative travel patterns[1]. The model was estimated based on recorded travel diaries and the subsequent simulation results indicated that the model was able to reproduce departure times, trip lengths, number of trips distribution and mode shares with

* Corresponding author. Tel.: +4687906897.
E-mail address: salee@kth.se

1877-0509 © 2018 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2018.04.090

in the sample very well.

Demand for travel comes from demand for activity participation. With this realization, the travel demand models have evolved recently from trip and tour based models into activity based models. The problem with activity based models is that if one considers a full day activity schedule; there are a huge number of possible ways to plan a day. Most of the activity based models create full day activity schedules, but fail to take the dependency of preference of time over "time of day". E.g. Bowman and Ben-Akiva[2] have employed a nested logit structure that treats tours and activities sequentially based on individual's preference. In this model, primary tours are temporally consistent but secondary tours are not temporally consistent. The model was later extended by combining with a duration and departure time model, which was not integrated into the nested logit structure[3,4]. Habib[5] presented a discrete continuous random utility model for weekend traveling, in which agents are able to choose mode, destination and activity based on the utility of the combination of these three. Future time is time-of-day dependent, parameterized and estimated. This is somewhat relevant to the model presented in this paper. Similarly, Arentze et al.[6] follow a similar approach where choices are made sequentially in time and a decision-rule heuristic system decides the next action at any particular time step. One of the main difficulties with these approaches is treating the value of time consistently.

Balmer et al.[7] and Horni et al.[8] presented models where individuals search for full day patterns that maximize their full day utility. Balmer et al.[7] simulated day paths and took a search algorithm based approach to find paths which maximise total full day utility. The lack of randomness in this model makes it unsuitable for location choice, an issue that was targeted by Horni et al.[8] by using MNL model.

Mixed-integer programming has also been used in literature to solve activity scheduling problem. Recker[9] presented a model to solve the choice of activity schedules including mode, activity duration and participation, and estimated the model using a genetic algorithm[10]. They included time and space constraints into the programming problem. Kang and Recker[11] later extended the work to include destination choice.

In order to use SCAPER for forecasting in combination with a traffic simulation model, there were two main problems that had to be solved: Firstly, the model is computationally very demanding to evaluate. Simulating a daily travel pattern for one individual takes close to 10 seconds when 1240 locations are considered for each trip. Secondly, the level-of-service attributes used for estimation are not the same as the level-of-service attributes obtained from the traffic simulator. It would be desirable if the modelling system would reproduce observed behavior from travel diaries in the fixed point with estimated parameters.

To speed up the simulation process, this paper proposes using "sampling of locations". A simple MNL model is estimated from which locations are sampled for each individual. The resulting travel patterns are compared with observed behavior to determine the number of locations that has to be sampled in order to reach an acceptable level of bias. To avoid the discrepancy in travel times reported in travel diaries, the proposed model uses travel times and travel costs generated through a dynamic traffic flow simulation model for the city of Stockholm, using MATSim. Although sampling of locations is proposed for computational efficiency, this model has also been solved for the entire number of 1240 locations (zones) in Stockholm. This enables us to know and compare the systematic bias sampling different number of locations may generate in the system. To our knowledge, this systematic comparison of bias in the selection of a benchmark number of locations for sampling has not been focused before.

2. Model

This section provides the detailed specification of SCAPER. All aspects of travel pattern are interconnected, and so a correct representation of time is crucial in activity based models as some aspects of time are fixed and cannot be violated. The choice of travel decision can be explained as trade-offs to spend a limited amount of available time, and it makes sense to assume that travellers consider full day travel pattern, while deciding what activities to engage in. If a state s_t represents a location and time of day, and an action a_t defines activity type, duration and the mode of transport that moves the traveller to a new state s_{t+1}; then a full day path actually is a sequence of actions starting in the morning and ending in the evening. The travel times vary, and hence add some kind of randomness to the model. States are also only partially visible, as unexpected needs or opportunities may also arise, adding further stochasticity to the model. A rational agent following a policy π in a stochastic environment, starting in a state s_t, will take action a_t in state s_t, that maximizes the expected future utility:

$$V(s_0) = \max_{\pi} E_s \left\{ \sum_{t=0}^{T} \beta^t u(s_t, a_t) \middle| s_0 = s \right\} \tag{1}$$

Where $u(s_t, a_t)$ is a one-stage utility function and β is the discount factor. We assume β as 1, and assume $u(s_t, a_t)$ to be additively separable into $u(x_t, a_t) + \varepsilon_t$. $u(x_t, a_t)$ represents the known part (known to both econometricians and the individual), while ε_t is the unknown part (not known to econometrician, only known to individual over current time period t), capturing all uncertainty in the system. ε_t is independent and identically distributed over alternatives and time. We assume that the state s_t is divided into (x_t, ε_t), such that ε_t is Gumbel distributed and $\varepsilon_t \sim G(-\gamma, 1)$, meaning the mean of ε is 1 rather than γ . With these assumptions, the expected value function becomes:

$$EV(x_t) = \int V(s_{t+1})\, p(ds_{t+1} \mid s_t, a_t) = log\left(\sum_{a \in A(x_t)} e^{u(x_t,\, a) + EV(x_{t+1})} \right) \tag{2}$$

The probability that an alternative a is chosen when in state x , can be given by:

$$P(a_t \mid x_t) = e^{u(a_t, x_t) + EV(x_{t+1}) - EV(x_t)} \tag{3}$$

Where $EV(x_t)$ represents the expected value of the value function in state x at time t . For an individual, the one stage utility can be obtained by adding the utility of performing the activity and (dis)utility of travelling to the activity location. The travel disutility is in turn dependent on the mode, time of day, origin and destination. The utility of performing an activity is calculated from a time-of-day dependent constant utility for starting the activity, and an activity-duration and time-of-day dependent utility. The duration utility is calculated by integrating the marginal utility of activity participation at time t (from time t to time $t + \Delta t$), where Δt is the duration of the activity episode.

3. Sampling of Locations

There can be two motivations for decreasing the size of the choice set. The first is to try to better mimic the way in which individuals actually make decisions; and the second is to reduce the computational complexity of the problem. Either sampling or aggregation of locations are always used (directly[2, 12, 13], or indirectly[14]) in discrete choice travel demand models of larger cities and regions, most often motivated by computational requirements. The number of locations is especially problematic in models allowing tours with multiple non-home location choices, as all locations can then serve as both origins and destinations of trips.

Below we discuss in detail how locations are sampled in SCAPER. Let N represents the total number of alternatives, $h(i) = e^{V(i)}$ represents the function for which the expected utility is approximated, X_i be a random variable, $X_i \in \{1, \ldots, N\}$, C is set of random variables $X_i : C \subset \{1, \ldots, N\}$, J is number of alternatives in C , $q(i)$ is the probability density function for the sample distribution to generate samples X_i , k_i is the number of times random variable X occurs in $C : \sum_1^N K_i = J$, and μ_{IC} represents importance sampling estimates. There are N discrete samples for which we sample, and the purpose of sampling is to approximate

$$\mu = \sum_{l=1}^{N} e^{V_i} = \sum_{l=1}^{N} h_i \tag{4}$$

by sampling a subset $C \subset \{1, \ldots, N\}$ of the alternatives. Approximating this sum is straight forward using importance sampling. If $p(i)$ is a probability density function, h can be written as an expectation under q using:

$$\mu = N \cdot \sum_{j=1}^{N} \frac{h_j}{q_i} \cdot q_i = N \cdot E_q \left[\frac{h_j}{q_i} \right] \tag{5}$$

It can thus be approximated by drawing a set of J samples $X_1, \ldots, X_J$ according to q, to obtain the approximate value. Let k_i be the times alternative i is sampled, the importance sampling approximation of μ is then given by:

$$\mu_{IC} = \frac{N}{J} \cdot \sum_{i=1}^{J} \frac{k_j}{q_j} \cdot h_j \tag{6}$$

The optimal choice of q is dependent on h, and in this specific case it is clear that the optimal choice would be:

$$q_i^* = \frac{e^{V_i}}{\sum_{j=1}^{N} e^{V_j}} \tag{7}$$

in which case:

$$\mu_{IC}^* = \sum_{i=1}^{J} \frac{k_j}{q_j^*} \cdot h_j = \sum_{j=1}^{N} h_j = \mu \tag{8}$$

independently of the number of samples J. However, if it was possible to sample directly from this distribution there would be no need to approximate the sum in the first place. Still, this seems to indicate that a good sampling distribution q should be similar to the optimal distribution q^*. Our idea is therefore to seek an approximation of q^* which is easy to evaluate and from which samples can easily be obtained. This can however only partly solve our problem, as the purpose is to draw a sample and reuse it in all future states. This means that values of $h(i)$ will change, but that the same sample C should be reused to approximate μ. Although importance sampling gives unbiased estimates independently of the value of h, the variance of the estimate will be highly dependent on the sampled values and the weights assigned to each sampled value.

4. Interfacing SCAPER with MATSim

As the details about MATSim can be found elsewhere[15], only a brief overview is provided. MATSim is an activity-based multi-agent traffic simulation framework. It uses a microscopic description of travel demand and performs fast mesoscopic simulation of traffic flows and the congestion resulting from those traffic flows. For this work, the demand used by MATSim is synthetic in nature, but based on real census data, hence fairly representative of the actual Stockholm demand. Each traveler of the synthetic demand has one or more travel plans for the day, representing its intentions for the day. In order to do a performance-based comparison of plans, MATSim allocates a real-valued utility value to each plan. The default utility functions of MATSim have been re-implemented to make them consistent with SCAPER.

From the initial generated synthetic demand, MATSim generates initial level-of-service matrices; based on which, SCAPER generates travel patterns consisting of number of trips as well as mode, destination, departure time and purpose for all trips. MATSim then simulates the demand with SCAPER generated travel patterns, and iterate route choice until an approximate stochastic equilibrium is achieved, such that adjusting routes does not yield further improvements on average. After achieving such equilibrium, new level-of-service matrices are calculated, which are then fed to SCAPER to generate new travel patterns. The interactive iteration between SCAPER and MATSim continues until convergence is achieved. We assume convergence in SCAPER-MATSim interaction when the difference between different attributes of alternatives in observed data and simulated data do not change systematically anymore and only stochastic variations exist. Further details of the interface of SCAPER and MATSim will be published in a separate publication.

5. Results and Discussion

While it is intuitive that sampling of locations achieves computational benefits, it is also important to check that sampling of locations does not introduce systematic bias in the simulation results. To check this, we performed a number of simulation experiments with sampling 40, 80, 160, 320 and 640 locations. The computational time and simulations results were then compared to simulation results with-out any sampling of locations.

Fig. 1 shows the computational time versus the number of locations sampled. As we can see in Fig. 1 that the computation time increases quadratically with number of locations sampled.

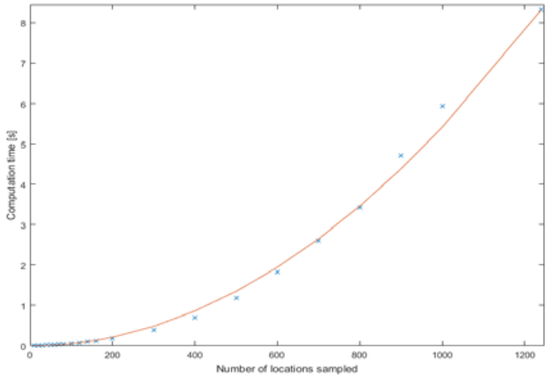

Fig. 1. Computational results of sampling of locations

Table 1 shows simulation values for car, PT, walk and bike mode shares and corresponding average trip travel times with different number of locations sampled. The mode share values represent the average number of trips made per individual per mode, for one day. As one can see, the difference between simulated values without sampling of locations and simulated values with sampling of locations is pretty low, indicating that the simulated results with location sampling are able to reproduce the simulated values without sampling. One can also notice that as we increase the number of locations to sample, the mode shares values get closer to the mode share simulation values without sampling. We think that 80 is a good number of locations to sample, as 80 is a reasonable number of location choices available to consider for any individual who wants to perform an activity. For sampling 80 locations, the computational benefit is also high, and the simulated results are close to the simulated results without sampling. A similar trend to mode shares values can also be observed in average trip times in the table.

Table 1. Average mode share and trip times simulation results with and without location sampling.

Number of Locations Sampled	Car Mode Share	PT Mode Share	Walk Mode Share	Bike Mode Share	Car Time	PT Time	Walk Time	Bike Time
40	1.0855	1.0608	0.4019	0.2502	19.8913	28.9010	10.1477	5.1999
80	1.0792	1.0622	0.3896	0.2490	19.8256	29.0386	10.1357	5.2271
160	1.0772	1.0598	0.3837	0.2474	19.8537	29.0191	10.1600	5.2058
320	1.0765	1.0587	0.3788	0.2475	19.8649	29.0059	10.1285	5.2235
640	1.0750	1.0587	0.3767	0.2479	19.8482	28.9979	10.1183	5.2172
1240 (All Locations)	1.0761	1.0596	0.3769	0.2462	19.8597	29.0305	10.1254	5.1885

We estimated the model based on Stockholm travel survey from 2004, in which the users have reported full day travel diaries. In order to avoid any discrepancy in the travel times and travel costs in the travel survey, we used travel times and travel costs produced by the simulation of Stockholm demand (simulated) over Stockholm network using MATSim, for estimation. Jointly with estimation, we also simulated the Stockholm car owning demand (61,000 sampled individuals). For simulation we have sampled 80 locations out of 1240 locations. It creates a bias to the system, but the bias is within an acceptable threshold. For all important attributes, the bias is under 2 percent. Table 2 shows the values for mode shares and average mode trip times (minutes), both for the observed data and simulated data. As one can see, in the converged setting, the difference is very small, indicating the simulated results are able to reproduce the observed behavior.

Table 2. Simulation results for SCAPER/MATSim integration .

Quantity of Interest	Observed Values	Simulated Values	Difference
Car Mode Share	1.03681	1.01677	1.93257
PT Mode Share	1.05715	1.07582	-1.76564
Bike Mode Share	0.242816	0.240421	0.986139
Walk Mode Share	0.35712	0.35848	-0.380805
Car Time	14.2581	14.0202	1.66883
PT Time	29.3087	29.8187	-1.74013
Bike Time	5.23743	5.22748	0.190012
Walk Time	9.85045	9.89375	-0.439575

References

1. Västberg OB, Karlström A, Jonsson D, Sundberg M. *Including time in a travel demand model using dynamic discrete choice*. Germany:University Library of Munich; 2016.
2. Bowman JL, Ben-Akiva ME. Activity-based disaggregate travel demand model system with activity schedules. *Transportation Research Part A: Policy and Practice* 2001; 35(1): 1-28.
3. Bradley M, Bowman JL, Griesenbeck B. SACSIM: An applied activity-based model system with fine-level spatial and temporal resolution. *Journal of Choice Modelling* 2010; 3(1): 5-31.
4. Vovsha P, Bradley M. Hybrid discrete choice departure-time and duration model for scheduling travel tours. *Transportation Research Record* 2004; 1894(1): 46-56.
5. Habib KMN. A random utility maximization (RUM) based dynamic activity scheduling model: Application in weekend activity scheduling. *Transportation* 2011; 38(1): 123-151.
6. Arentze T, Hofman F, van Mourik H, Timmermans H. ALBATROSS: Multiagent, rule-based model of activity pattern decisions. *Transportation Research Record* 2000; 1706(1): 136-144.
7. Balmer M, Raney B, Nagel K. (2005) Adjustment of activity timing and duration in an agent-based traffic flow simulation. In: Timmermans H, editors. *Progress in Activity-Based Analysis*, Elsevier; 91-114.
8. Horni A, Nagel K, Axhausen KW. *High-resolution destination choice in agent-based demand models.* In: Zürich: Eidgenössische Technische Hochschule, Institut für Verkehrsplanung und Transportsysteme; 2011.
9. Recker W. A bridge between travel demand modeling and activitybased travel analysis. *Transportation Research Part B: Methodological* 2001; 35(5): 481-506.
10. Recker W, Duan J, Wang H. Development of an estimation procedure for an activity-based travel demand model. *Computer-Aided Civil and Infrastructure Engineering* 2008; 23(7): 483-501.
11. Kang JE, Recker W. The location selection problem for the household activity pattern problem. *Transportation Research Part B: Methodological* 2013; 55(0):75- 97.
12. Bowman JL, Bradley M. Disaggregate treatment of purpose, time of day and location in an activity-based regional travel forecasting model. In: *European Transport Conference* 2005; Strasbourg, France.
13. Liao F, Arentze T, Timmermans H. Constructing personalized transportation networks in multi-state supernetworks: a heuristic approach. *International Journal of Geographical Information Science* 2011; 25(11): 1885-1903.
14. Beser M, Algers S. Sampers — the new swedish national travel demand forecasting tool. In: Lundqvist L, Mattsson LG, editors. *National Transport Models: Recent Developments and Prospects*. Springer Berlin Heidelberg, Berlin, Heidelberg; 101-118.
15. Horni A, Nagel K, Axhausen KW. *The multi-agent transport simulation MATSim*; 2016.