



Forecasting travel behavior using Markov Chains-based approaches



Ismail Saadi*, Ahmed Mustafa, Jacques Teller, Mario Cools

University of Liège, ARGenCo, Local Environment Management & Analysis (LEMA), Quartier Polytech 1, Allée de la Découverte 9, BE-4000 Liège, Belgium

ARTICLE INFO

Article history:

Received 23 January 2016
 Received in revised form 22 June 2016
 Accepted 27 June 2016
 Available online 5 July 2016

Keywords:

Travel behavior analysis
 Travel demand modeling
 Activity sequences
 Daily activity-travel patterns
 Simulation-based population synthesis
 Profiling analysis

ABSTRACT

Recent advances in agent-based micro-simulation modeling have further highlighted the importance of a thorough full synthetic population procedure for guaranteeing the correct characterization of real-world populations and underlying travel demands. In this regard, we propose an integrated approach including Markov Chain Monte Carlo (MCMC) simulation and profiling-based methods to capture the behavioral complexity and the great heterogeneity of agents of the true population through representative micro-samples. The population synthesis method is capable of building the joint distribution of a given population with its corresponding marginal distributions using either full or partial conditional probabilities or both of them simultaneously. In particular, the estimation of socio-demographic or transport-related variables and the characterization of daily activity-travel patterns are included within the framework. The fully probabilistic structure based on Markov Chains characterizing this framework makes it innovative compared to standard activity-based models. Moreover, data stemming from the 2010 Belgian Household Daily Travel Survey (BELDAM) are used to calibrate the modeling framework. We illustrate that this framework effectively captures the behavioral heterogeneity of travelers. Furthermore, we demonstrate that the proposed framework is adequately adapted to meeting the demand for large-scale micro-simulation scenarios of transportation and urban systems.

© 2016 Elsevier Ltd. All rights reserved.

1. Background

Activity-based analyses of travel behavior within large-scale micro-simulation models are particularly adapted to understanding the dynamics and the transitional patterns of travel behavior. In this context, most activity-based models require a full (synthetic) population to obtain greater behavioral realism within such complex transport and urban systems. However, the lack of highly disaggregate data suggests the use of more efficient methods for synthesizing individual/household socio-demographic attributes as well as their daily activity information (Beckman et al., 1996). Obtaining accurate individual-level information for a large population is a great challenge especially in the context of restrictive data availability. Typically, such detailed data can be derived from national censuses. However, in practice, only aggregate information is available to researchers and practitioners (Anderson et al., 2014).

Furthermore, micro-samples may also include incomplete observations, making the true population identification process through a virtual process more complex. To overcome this problem, several techniques have been developed to address multiple data sources and incomplete datasets. Generally, detailed individual-level micro-data and aggregate

* Corresponding author.

E-mail address: ismail.saadi@ulg.ac.be (I. Saadi).

large-scale datasets are the two components used as input for population synthesizing purposes. In the literature, the classical techniques identified are Iterative Proportional Fitting (IPF) (Beckman et al., 1996; Mohammadian et al., 2010), Iterative Proportional Updating (IPU) (Ye et al., 2009), Combinatorial Optimization (CO) (Williamson et al., 1998; Voas and Williamson, 2000) and probabilistic models (Farooq et al., 2013; Sun and Erath, 2015; Saadi et al., 2016b).

In this paper, we opt for a simulation-based approach because of its ability to address incomplete datasets using partial conditional distributions (Farooq et al., 2013). More specifically, this operational approach based on a Gibbs sampler can generate agents using partial views of the true joint distribution as if the synthesized agents were drawn from the real population. In this context, the method enables the generation of populations of any given size. A comparative study established by Farooq et al. (2013) between IPF and simulation-based approaches revealed that the latest technique clearly outperforms IPF using, for both methods, almost the same amount of data. We build a representation of the true population by implicitly merging different information provided by multiple micro-data sources using the simulation-based approach. Then, from this representative/synthesized population, we can group the population into homogeneity clusters characterized by similar attributes to understand their related travel behavior and to enable a comparative attribute-based study based on the activity sequences.

For several decades, characterizing peoples activity-travel behavior has been an important issue to researchers (Pendyala and Goulias, 2002; Auld et al., 2016; Saadi et al., 2016a). To include the sequential dependencies of daily activities, some studies have suggested the use of Sequence Alignment Methods (SAMs) (Joh and Timmermans, 2011; Joh et al., 2002; Wilson, 1998), whose inputs representing the daily activity behavior in the form of activity sequences are extracted from activity-travel diaries (Bhat and Singh, 2000; Spissu et al., 2009). Typically, the application of SAMs includes pairwise sequence alignment algorithms for scoring and comparing activity sequences in-between them. In this way, the main activity patterns can be identified quantitatively and qualitatively within their respective subset of activity chains (Joh et al., 2006).

When a relationship can be established between the derived patterns and the variables of interest, general activity-travel behavior trends can be analyzed to investigate more deeply the determinants influencing those travel patterns (Wilson, 1998). Although SAMs have been extensively used for characterizing activity-travel patterns, such methods clearly suffer from serious limitations. Indeed, SAMs are unable to identify the complete nature of a cluster. Even if the information regarding the frequent activity patterns is extracted, irregular activities are not considered; thus, only parts of the clusters are characterized (Liu et al., 2015). In this context, a number of questions have been raised regarding the consistency of such sequence alignment techniques.

In an attempt to obtain improvements, Liu et al. (2015) proposed a profiling method called profile Hidden Markov Models (pHMMs) for characterizing the complete set of activity sequences including irregular activities. pHMMs belong to the family of sequence profile methods that are essentially used to characterize protein sequences. Furthermore, pHMMs may be described as position-specific scoring parameters emanating from a multiple alignment of a group of protein sequences (Durbin et al., 1998).

From a methodological point of view, the positions of alignment are categorized according to three possible states: match, insertion and deletion. Fundamentally, the configuration of the model is a successive layout of match states representing the conserved activities that have been identified within their respective cluster. This successive layout forms the reference activity sequences or a base template characterized by the most frequent activities. In parallel, insertion states model the introduction of new conserved residues, which can be interpreted as additional activities incorporated into the previously defined base template. Conversely, deletion states represent the omission of particular conserved activities from the base template. In this regard, random differences between the activity sequences within the overall characterization process are implicitly included throughout these two states. Subsequently, estimated pHMMs can generate new activity sequences so that the relationships with their corresponding cluster are preserved (Liu et al., 2015).

Furthermore, pHMMs can be considered as quantitative descriptors that assign weights – probabilities of occurrence – to each activity at each corresponding match state characterizing both regular and irregular activities. For instance, considering daily activity-travel sequences, some travelers might see their behavior differing from the general behavior of a cluster (e.g., work cluster). This difference is captured by pHMMs by introducing new activities or, inversely, canceling or ignoring others. Subsequently, the scored activity sequences can be assessed to measure how similar or dissimilar these are from the cluster (Liu et al., 2015).

In this paper, we propose an integrated framework including both a population synthesis approach (Farooq et al., 2013) and a profiling method (Liu et al., 2015) capable of describing and assigning the activity sequences for each individual of the synthetic population. The model is capable of considering, in an efficient manner, the behavioral effects of different influencing factors, which might vary between clusters. In this regard, the main trends in terms of travel behavior can emerge from the characterization achieved using the pHMMs.

The remainder of this paper is organized as follows. First, the data, obtained from the Belgian National Household Travel Survey, are described. Subsequently, in Section 3, the details of the integrated framework are provided to enable implementation in different contexts. Then, the main results and the validation of the integrated framework are discussed (Section 4). Finally, the concluding remarks are formulated in Section 5.

Download English Version:

<https://daneshyari.com/en/article/524739>

Download Persian Version:

<https://daneshyari.com/article/524739>

[Daneshyari.com](https://daneshyari.com)