Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 83 (2016) 34 – 41

Procedia Computer Science

The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016)

A generic data-driven sequential clustering algorithm determining activity skeletons

Wim Ectors[a,*], Bruno Kochan[a], Luk Knapen[a], Davy Janssens[a], Tom Bellemans[a]

[a]*Transportation Research Institute, Hasselt University, Wetenschapspark 5 Box 6, B-3590 Diepenbeek, Belgium*

Abstract

Many activity-based models start by scheduling inflexible or mandatory activities (if present), before more flexible activities. Often work and educational activities are assumed as most stringent and recognized as the only mandatory activities. According to this definition, only 45% of all schedules contains a mandatory activity (OVG single-day travel survey in Flanders, Belgium). This means 55% of schedules does not have a traditional mandatory-flexible activity structure. This research proposes a completely data-driven approach to reveal the real basic structure of individuals' schedules, i.e. the skeleton schedule sequence. To this end, a sequential clustering algorithm was developed. Furthermore, an in-depth analysis of the parameter settings was performed. The proposed method reveals a set of skeleton activity schedules and confirms the importance of work and education.

Peer-review under responsibility of the Conference Program Chairs

Keywords: skeleton schedules; activity patterns; mandatory activities; activity-based modeling

1. Introduction

Despite having some limitations, the traditional four-step model is still abundantly used by small- and medium-sized cities and consultancy agencies across the world. It is relatively easy to implement and has a low complexity. Activity-based models (ABMs) have been under development for many years and remedy many of the traditional four-step model's limitations. They are characterized by the fact that the base unit is an activity rather than a trip or tour.

In 1986 the STARCHILD system of models was introduced. It used a utility-maximizing approach to determine the expected activity pattern (syn. schedule). It distinguished between planned and unplanned activities. Planned activities were predicted and inserted into the schedule first and subsequently used as boundary conditions in the unplanned activity planning problem[1,2]. Also in most recent ABMs mandatory, inflexible activities are determined around which more flexible activities can be scheduled[3,4]. Often only work and school (university) activities are considered for a mandatory activity. This approach and assumption is used in i.a. SWDAS (TASHA), ALBATROSS, CEMDAP, FAMOS, the CT-RAMP family of ABMs in general and recently also in the daily-activity pattern model. However, e.g. the HAPP framework introduced in 1995 does *not* differentiate between flexible or inflexible activities.

* Corresponding author. Tel.: +32-11-269114.
E-mail address: wim.ectors@uhasselt.be

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2016.04.096

However, only 33% of all activity schedules in Flanders, Belgium, contain a work or business activity. Including educational activities, still only 45% of all schedules contains a mandatory activity. Admittedly, one does not necessarily have a mandatory activity scheduled every day. However considering modeling purposes, there could still be meaningful activity patterns in those remaining 55% of schedules which do not obey the popular mandatory- vs flexible activity-structure. Extracting these is the purpose of the current research: revealing *skeleton* activity patterns which can be used as nuclei for subsequent scheduling decisions, in a generic data-driven fashion.

Some previous researchers found indications that the concept of skeleton patterns is plausible. Doherty[5] revealed that approximately 34% of activity time is considered routine or planned weeks to years in advance. This substantial fraction of preplanned and routinized (recurring) activity time can be viewed as a base for a skeleton pattern in one's schedule. Arentze et al.[6] postulate that activity patterns are the outcome of a learning process, of which the result is a set of rules. They reject the notion that individuals systematically compare all possible activity patterns. A skeleton activity pattern implicitly captures some of these rules and would therefore limit the activity pattern choice set. Additionally, Arentze et al.[7] found that predicting tour skeletons instead of assuming them given within the model did not significantly reduce the goodness of fit of their ABM: ALBATROSS 2. They however did not detail the process by which these skeletons are generated.

A k-means-based clustering technique was developed by Allahviranloo et al.[8], ultimately in order to accurately forecast activity patterns of an individual. Individuals with their socio-demographic information are clustered in two stages according to their observed activity patterns. A surprising result was the discovery of a core set of sequences often serving as a centroid for other patterns. This discovery strengthens the belief that a core set of skeleton schedules exists and that these could be used for activity pattern modeling purposes.

Roorda and Ruiz[9] found evidence against the common assumption that a skeleton schedule only consists of mandatory activities (such as work or education). They used a CHASE-based dataset from the TAPS longitudinal survey.

The current research forms the first step towards a multi-agent, rule-based traffic demand model. Similar to ALBATROSS[6] and FEATHERS[10] it might be based on a system of sequential decisions using decision trees. Note that within this research, skeleton activity patterns are defined as the fundamental activity patterns that are the backbone for many complete sequences of activities. Some authors used *skeleton activity patterns* in a slightly different meaning, i.e. that of a set of strictly inflexible or mandatory activities. In the current research activity skeletons are identified based on frequency rather than flexibility. This criterion reveals important patterns that may be used as fundamental patterns for modeling purposes.

2. Methodology

2.1. Data Description

The data used in this research resulted from the *Transportation Behavior Research Flanders* (*Onderzoek Verplaatsingsgedrag* or OVG) travel survey conducted in the Flanders region, Belgium. This large-scale survey is funded by the Ministry of Mobility and Infrastructure. Single-day travel diaries (including weekends) enriched with individual and household socio-demographical information were collected for approximately 17,300 individuals. Participants were asked to keep a detailed log of their trip purpose, departure and arrival time, and trip origin and destination on a randomly assigned day. Individual weights were calculated in order that the OVG sample may correctly represent the true population of Flanders. These weights were used in all reported figures of this paper. The survey was conducted in multiple phases from 2007 until 2013. A new phase started in 2015.

Of the 17,300 participants, approximately 13,200 conducted at least one trip. Using the reported trip purposes of Table 1, single-day activity schedules were created. All of them are assumed to start at home, but do not necessarily end at home. A limitation is that only activities for which a trip was performed are included (i.e. no at-home activities). Multiple identical consecutive activities were merged into a single occurrence, a practice which can be defended when purely studying activity patterns. It is beneficial as it removes some of the variance present in the data.

In total, approx. 2,600 unique single-day schedules could be constructed. The 14 schedules having the highest frequency account for 45% of the observations. None of the other schedules has a share larger than one percent. This shows that quite a large fraction of the observed travel patterns can be grasped in only 14 distinct schedules, but that