Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 19 (2016) 81 – 93

Transportation Research Procedia

www.elsevier.com/locate/procedia

International Scientific Conference on Mobility and Transport Transforming Urban Mobility, mobil.TUM 2016, 6-7 June 2016, Munich, Germany

Constructing a synthetic population of establishments for the SimMobility microsimulation platform

Diem-Trinh Le [a]*, Giulia Cernicchiaro [a], Chris Zegras [b], Joseph Ferreira Jr. [b]

[a] *Singapore-MIT Alliance for Research and Technology, Singapore*
[b] *Massachusetts Institute of Technology, USA*

Abstract

This paper presents a method for building a synthetic population of establishments in Singapore for incorporation into SimMobility platform, an integrated microsimulation model that represents households' and firms' short-, medium-, and long-term decisions, ranging from lane changing behavior, to daily activity pattern choices, to location choices. The synthetic population includes data on (1) establishments' locations, (2) establishments' industry type, (3) establishments' employment size, and (4) their occupied floor area.

Peer-review under responsibility of the organizing committee of mobil.TUM 2016.

Keywords: population synthesis; microsimulation; iterative proportional fitting, SimMobility

1. Introduction

Along with households, firms are key drivers of urban dynamics. As such, understanding firmography – or the growth, failure, and migration of new or existing firms – is critical in urban research and modeling. Firms' activities require accessibility to capital and labour, to other firms, and to customers; all of which influence their location

* *Corresponding author. Tel.: +65-6601-1548; fax: +65-6684-2118.*
E-mail address: diem@smart.mit.edu

Peer-review under responsibility of the organizing committee of mobil.TUM 2016.
doi:10.1016/j.trpro.2016.12.070

choices to some degrees. The recent wave of land use-transportation modeling tools reflects a move towards *micro*-simulation e.g. ILUTE, ILUMASS, Ramlas, and UrbanSim, aiming to represent all agents in the model system and their relevant decisions. Such models require a reasonable representation of these agents – the population of interest. However, in most cases, micro-data on firms are considered sensitive and thus not available or not accessible for research. As information for individual firms are needed for micro-simulation models, a synthetic population is often generated for this purpose.

There have been several studies on methods for synthesizing household population (e.g. Zhu & Ferreira 2014; Rich & Mulalic 2012; Beckman et al. 1996). Nevertheless, the literature on firms' synthetic population is comparably sparse. Generating a baseline synthetic population of firms or establishments is necessary to simulate firmography processes and incorporate into land-use transport models. A common method for household population synthesis is Iterative Proportional Fitting technique with Monte Carlo simulation. The same method has also been applied for firms in several studies (e.g. Khan et al. 2002; Maoh & Kanaroglou 2005; Moeckel 2009). In most cases, an establishment is described by its consumption and production, age, required floor space, type of industry, and location. However, how close the synthetic population resembled the reality was not discussed. Cernichiaro and Ferreria (2015) presented a method to build a synthetic population of establishments using directory websites. Firm characteristics simulated include employment size, floor area, and industry type. However, the method was specifically applied to the service sector and manufacturing firms were excluded.

This paper demonstrates a method for creating a synthetic population of establishments in Singapore using limited available data. The population synthesis aims to reflect the reality of firms in Singapore in 2012 for incorporation into SimMobility model (Ben-Akiva 2010; Lu et al. 2015), an integrated microsimulation model that represents households' and firms' short-, medium-, and long-term decisions, ranging from lane changing behavior, to daily activity pattern choices, to location choices. For the synthetic firm population, data to be generated include (1) establishments' locations, (2) establishments' industry type, (3) establishments' employment size, and (4) their occupied floor area.

The paper is organized as follows. First, we will briefly introduce the SimMobility framework. Following an explanation of the simulation methods, we will present the generated synthetic population. The paper concludes with a summary of the methods and some implications for future work.

2. SimMoblity Framework

SimMoblity is a system of mobility sensitive behavioural models integrated in a multi-scale simulation platform that considers land-use, transportation, and communication interactions. It focuses on the impacts on transportation networks, intelligent transportation services, and vehicular emissions, thereby enabling the simulation of a portfolio of technology, policy, and investment options under alternative future scenarios (Adnan et al. 2016). The framework consists of three different sub-models:

- Short-term (ST) simulator is a traffic micro-simulator, extended with a communications simulator as well as pedestrians and public transport. The time step can be a fraction of a second and agent decisions include lane changing, braking, accelerating, gap acceptance, but also route choice.
- Mid-term (MT) simulator is a mesoscopic simulator, designed for activity-based modelling, with explicit pre-day and within-day behaviour including re-routing and re-scheduling, and multiple transport modes. The time step is in the range of seconds to minutes and agent decisions include route choice, mode choice, activity pattern and its (re)scheduling, departure time choice.
- Long-term (LT) simulator is a land-use and transport (LUT) simulator, with a market transaction bidding model. The time step is in the range of days to months to years, and agent decisions include house location choice, job location choice, land development, and car ownership.

The LT simulator is responsible for the generation and updating of a population of agents and their corresponding demographic and locational attributes. In the beginning, a two-stage data synthesis methodology is employed for construction of a synthetic population of households and firm establishments at building scale. The approach is designed to accommodate the need for spatially disaggregated details in a manner that can be readily adjusted and rerun to incorporate new data sources, changed time frames, and updated relationships and hierarchies across overlapping datasets. Long-term behaviours of agents and their effects on urban form, markets and other