POPULATION SYNTHESIS IN ACTIVITY-BASED TRAVEL DEMAND

Ljupko Šimunović, Mario Ćosić, Dino Šojat, Julijan Jurak
University of Zagreb, Faculty of Transport and Traffic Sciences, Croatia

Abstract

A Synthetic Population is first part of creating travel demand model by using activity-based approach. Population synthesis is application of algorithms that expanded representative samples of people or household with characterises (such as gender, car ownership, age or ethnicity etc.) to entire area of researching. Because of complexity of people decisions before or during travel, one attribute is not enough to fully describe what factors have impact on them. Population synthesis iterate a set of attributes for each person in the sample and after expansion and assigning weights create simulated people or household with their characteristic. Basic components are marginal distribution targets of household and person attributes, household and person samples and algorithm for selecting the sample records into a synthetic population such that the attributes of that population match the marginal targets. Goal of this paper is to present population synthesis and her importance for activity-based approach in travel demand modelling. The paper will consist of introduction, literature overview, presenting benefits and complexity of population synthesis, discussion and conclusion.

Keywords: representative samples, set of attributes, marginal distribution, algorithm, marginal targets

1 Introduction

Activity-based concept present new way to handle predication of travel demand. Creating travel demand model is detecting relevant total of trips which occur at some specific area (for example large, middle or small city, various urban areas, rural areas or similar). Prediction of travel demand is determing total future trips with real possibility that will occur at some specific area. Most used method for travel demand modelling is four-step aggregate model. But that model has some limitations regarding to different traffic solutions. Influence of solutions that effect behaviour of people rather that infrastructure measures are hard to predict. That problem is trying to solve by activity-based approach. Activity-based approach is third and newest generation of travel demand model.
Main different between this to types of travel demand models, is level of aggregation and focus on trips or activities. Activity-based approach states that trips and mobility are sequence of some specific activity. Also, activity-based approach is focused on single household or people not zone. Product of activity-based approach is O-D matrix of single modes that are modelled. O-D matrix between zones are created by two separate steps. First steps are population synthesis and second is long-term, middle-term and short-term decisions. Population synthesis is method to create simulated population at some area by expanding relevant sample. Creating simulated population is achieving by various algorithms. In this paper will be presented basic definition of population synthesis and her role in for activity-based approach.

DOI: https://doi.org/10.5592/CO/CETRA.2020.1150

2 Literature overview

Various authors in the past try to detect best and most applicable methodology for creating travel demand at some area. Activity-based approach in the current literature shows various advantages regarding to traditional four-step travel demand approach. But also, that advantages are detected primarily in urban areas where is more and more important how to affect on human behaviour in transport, rather than building new infrastructure or making big and complex interventions in space. First step of conducting activity-based approach is to collect data. Two type of Data is collected: relevant sample (based on number of all population at specific area) and marginal targets.

After collecting data, comes first step of activity-based approach, called population synthesis. Methods for creating baseline synthetic populations of households and persons using 1990 census data are given [1]. By using Public Use Microdata Sample (PUMS), and Iterative Proportional Fitting (IPF) authors are estimating the proportion of households in a block group or census tract with a desired combination of demographics. In paper is estimate the proportion of households in a block group or census tract with a desired combination of demographics [1]. New activity-based models require more detailed information on household demographics and employment characteristics [2]. Using Monte Carlo microsimulation these models aim at reproducing human behaviour at the individual level, i.e. how individuals choose between options following their perceptions, preferences and habits subject to constraints, such as uncertainty, lack of information, and limits in time and money [2].

Because of certain problems regarding to data availability, these kinds of approach must be generated that represents individuals in the form of households and household members [2]. While most existing procedures concentrate on iterative proportional fitting (IFP), this paper show combing different approaches [2]. Until this date, the conventional approach to synthesizing the base-year population has been based on the iterative proportional fitting procedure [3]. Paper show a new population synthesis procedure is presented that addresses the limitations of the conventional approach. Also, validation results indicate that the new procedure can produce a synthetic population that more closely represents the real population size than the conventional approach [3].

In agent-based microsimulation models for land use, the initial step is the definition of agents – usually, persons and households [4]. Authors in this paper are trying to summarize recent efforts to population synthesis for microsimulation. Authors also stated in past papers from various authors that they share two principle: adjustment of an initial population, taken from a past census or other survey data, to current constraints, and selecting households and optionally assigning them to geographic areas [4].

Goal of paper was to describe, analyse and evaluate the characteristics of the all mentioned approaches. Authors in paper [5] state that previous research have already developed a few techniques for generating a synthetic population: iterative proportional fitting and combinatorial optimization. This paper provides a guideline for using the synthetic population techniques by introducing terminologies, related research, and giving an account for the working process to create a synthetic population [5]. A method based on iterative proportional fitting (IPF) is developed for generating synthetic populations for the application of Albatross, a rule-based and activity-based model of travel demand [6].

A method is proposed to generate synthetic households based on data on distributions of individuals [6]. This method uses the concept of relation matrices to convert distributions of individuals to distributions of households in a pre-processing step [6]. Furthermore, a method is proposed to address differences in populations that relate to locational characteristics [6]. A crucial step in developing agent-based models is the definition of agents, e.g., household and persons [7]. This paper lists the most prominent techniques for population synthesis: iterative proportional fitting (IPF), iterative proportional updating (IPU), combinatorial

optimization, Markov-based and fitness-based syntheses, with other emerging approaches [7]. But, until present, authors stated that is no clear godliness for advises how and when use any of the available techniques. Also, authors claim that this paper present a comprehensive synthesis of available examples and literatures [7]. Population synthesis techniques are commonly used as alternative to supplement the lack of availability and completeness of microdata for microsimulation modelling [9]. This paper describes the process of generating a synthetic baseline population for Sydney Greater Metropolitan (GMA) using 2006 Population Census [9]. Those data were used to evaluate their representativeness with aggregated census data [9]. Paper [10] proposed a cross-entropy optimization model in which generalized constraints for different demographic characteristics of the synthetic population could be included. A quasi-Newton algorithm was used to solve the proposed problem [10].
Results from the model show that the proposed method held much promise for generating a more realistic synthetic population with different types of demographic characteristics and could be generally applied in different geographic areas [10]. Paper [11] show a new land use classification method for its explaining travel behaviour and as a new dimension in population synthesis. The method reproduces the four types of land use environments and improves ability to create a finer-grain geographic classification based on land use [11]. Paper also show similar indications about the difference between urban dwellers rural residents [11].
Techniques such as iterative proportional fitting (IPF) have been applied extensively to estimate data for the population, synthetically [12]. Paper proposes a binary linear programming model for tabular rounding in which the integer-converted table totals and marginal sums perfectly fit the input data [12]. The empirical comparison of the proposed method with eight existing methods demonstrates that the proposed model outperforms the tested methods [12]. Paper also show that in this paper, deterministic methods outperform stochastic methods in accuracy and perfect fit to census data [12].
Paper [13] endeavours to develop a synthetic population, based on the Simulated Annealing (SA) algorithm for the activity-based travel demand model. Hill climbing and cooling schedule are essential elements to be considered when applying SA into the synthetic population [13]. Also, Metropolis-Hasting Algorithm was employed to decide whether to select or dismiss the follow-up distribution so that hill climbing phenomenon can be prevented [13]. Based on this result, the current condition of micro sample and census data were utilized to compare the IPF (Iterative Proportional Fitting) of previous methodology with the establishment result of suggested algorithm [13]. Paper proved that the SA algorithm is valid and built with the synthetic population through statistical verification.

3 Population synthesis as basis of activity-based travel demand

Population synthesis is not exclusive to only activity-based models [14]. Synthetic population is used as the basis for forecasting the behaviour of the households and persons in the modelled area [15]. Most of aggregate and disaggregate models have a population that can be used with most of the relevant characteristics available for the base year because of data collected by national household travel survey [14]. That population needs to be synthesised for future years based on the parameters that have been forecast (number of people per zone/district, income, car ownership, etc.) [14].
Attributes, for example: distribution of household sizes, age distribution, school and university attendance, multiple vehicle ownership, etc. need to be estimated, usually at the level of the representative households (for each zone or district) [14]. The first part is to create a synthetic population and then simulate the behaviour of the households and persons in that population [14]. Population synthesis is creating by generating an artificial population by expanding the disaggregate sample data to mirror known aggregate distributions of household and person variables of interest [14].

The first input is marginal "control data," and the second input is "sample data" [15]. The control data represent the attributes that are being explicitly accounted for in the generation of the synthetic population [15]. Control data must be provided at relatively detailed geographic levels. The second data input is sample data. Samples of households and persons to create a list of households and persons that matches distributions from marginal (control data) [15]. The process starts by creating a base year synthetic population from available data while using aggregate demographic and land use forecasts to create a synthetic population of future years.
First step is estimating a demographic distribution of households is for each TAZ or small census area, and then matching sample of households that is chosen from an available data [14]. Second phase is identifying person attributes from within each household [14]. The final output is a synthetic population in which each artificial household and its members have many clearly defined characteristics of interest and together they match the estimated demographic distribution within determined zones.

4 Conclusion

Implementation of population synthesis in travel demand modelling is visible in all of models created by using activity-based approach. Regarding to which methodology was used or creating during the phase of decision-making (which come after population synthesis), synthesis of population was base of that phase.
Population synthesis is created by two set of data: marginal "control data," and the "sample data". Purpose of marginal data is determing real distributions of data attributes on fine geographic level. Sample data is representing number of households and persons in population on specific geographic area. Combing two set of data, it is possible to create artificial population (which match real number of people in area) of some area by matching various set of attributes (age distribution, incomes, gender, etc.).
Population synthesis is first step of creating activity-based travel demand model after collecting data. All mentioned papers with their authors try to develop or detected best methodology or procedure to create population synthetic. Docent of methodology were detected or created. Most used principle is IPF (Iterative Proportional Fitting), while others are in phase of researching. Proposed methodologies for population synthesis have their advantages and disadvantages, but in total all of them need to be more researched to detect which is most suitable for what situation.

References

[1] Beckman, J.R., Baggerly, K.A., McKay, M.D.: Creating synthetic baseline populations, Transportation Research Part A: Policy and Practice, 30, pp. 415-429, 1996.

[2] Moeckel, R., Spiekermann, K., Wegener, M.: Creating a Synthetic Population, 8th International Conference on Computers in Urban Planning and Urban Management (CUPUM), pp. 1-18, Sendai, Japan, May 2003.

[3] Guo, J.Y., Bhat, C.R.: Population Synthesis for Microstimulating Travel Behaviour, Transportation Research Record Journal of the Transportation Research Board, pp. 92-101, 2014.

[4] Muller, K., Axhausen, K.W.: Population synthesis for microsimulation: State of the art, STRC 2010, Swiss Federal Institute of Technology Zurich, Zurich, Switzerland, 2010.

[5] Cho, S., Bellemans, T., Creemers, L., Knapen, L., Janssens, D., Wets, G.: Synthetic Population Techniques in Activity-Based Research, Data Science and Simulation in Transportation Research, pp. 1-23, 2014.

[6] Arentze, T., Timmermans, H., Horman, F.: Creating Synthetic Household Populations: Problems and Approach, Transportation Research Record Journal of the Transportation Research Board, pp. 85-91, 2014.

[7] Ramadan, E.O., Sisiopiku, V.P.: A Critical Review on Population Synthesis for Activity- and Agent-Based Transportation Models, Transportation Systems Analysis and Assessment, InTech, Ljubljana, 2019.

[8] Gargeet, D.: Population synthesis for travel demand forecasting, 2013.

[9] Siripirote, T., Sumalle, A., Ho, W.H.: A Statistical Synthetic Population Calibration for Activity-Based Model with Incomplete Census Data, Journal of the Eastern Asia Society for Transportation Studies, 10 (2013), pp. 742-761

[10] Lee, D.H., Fu, Y.: Cross-Entropy Optimization Model for Population Synthesis in Activity-Based Microsimulation Models, Transportation Research Record: Journal of the Transportation Research Board, pp. 20-27, 2011.

[11] McBridem E.C., Davis, A., Goulias, K.: A Spatial Latent Profile Analysis to Classify Land Uses for Population Synthesis Methods in Travel Demand Forecasting, Transportation Research Record Journal of the Transportation Research Board, 49 (2018), pp. 158-170

[12] Choupani, A.A., Mamdooh, A.R.: Population Synthesis in Activity-Based Models: Tabular Rounding in Iterative Proportional Fitting, Transportation Research Record: Journal of the Transportation Research Board, 17 (2015), pp. 223-233

[13] Kim, J., Lee, S.: A simulated annealing algorithm for the creation of synthetic population in activity-based travel demand model, KSCE Journal of Civil Engineering, 20 (2016), pp. 2513-2523

[14] Ortuzar, J.D., Willumsen, L.G.: Modelling transport, 4th edition, Wiley, 2011.

[15] Castiglione, J., Bradley, M., Gliebe, J.: Activity-Based Travel Demand Models: A Primer, The second strategic highway research program, Transportation research board, 2015.

[16] Rich, J., Flotterod, G., Garrido, S., Pereira, F.: Review of population synthesis methodologies, Swedish National Road and Transport Research Institute, 2019.