Generating Artificial Populations

Hongxia Zhao
Chinese Academy of Sciences, China
hongxia.zhao@ia.ac.cn

Shuming Tang
Shandong University of Science and Technology, China

Yisheng Lv
Chinese Academy of Sciences, China

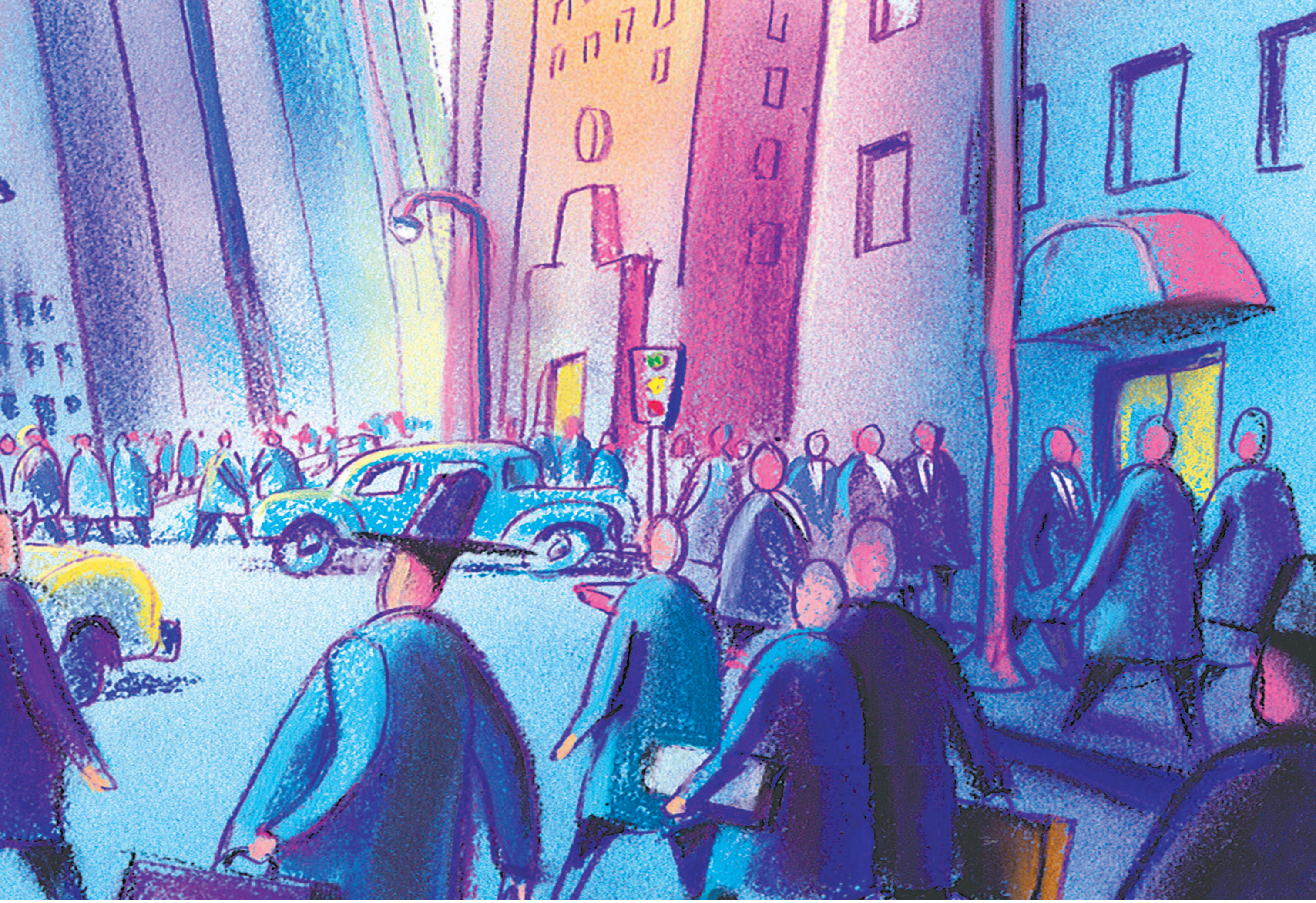
Abstract — Travel demand is derived from people's participation in daily activities scattered in time and space. Traffic microsimulation starts by generating individuals to participate in activities. In this paper, we propose a framework called Artificial Population Systems (APS) in order to automatically generate artificial populations for traffic microsimulation. Different from the conventional approach using empirical data to generate synthetic populations for the base year, our APS framework can generate artificial populations which evolve with time and space. So, with artificial populations, we will get more reasonable traffic demand forecast and analytical results in long term traffic microsimulation.

I. Introduction

With the accelerated process of urbanization, the rapid growth of urban population and the increment of urban socio-economic activities, urban traffic volume often grows in two to three times the growth rate of population growth. Meanwhile the trend of diversification and individuation in travel demand becomes apparent day-by-day.

Digital Object Identifier 10.1109/MITS.2009.934643

1939-1390/09/$26.00©2009IEEE



© MASTERSERIES

for Traffic Microsimulation

Transportation researchers have recognized that travel demand comes from individual activities, which stimulates an increasing interest in the investigation of individuals' activities [1–3]. Traffic microsimulation has shift from vehicle to people as the behavioral units. Figure 1 shows the most important modules of traffic microsimulation. Population generation module, which can generate the needed number of individuals with demographic properties, is the basis of traffic microsimulation.

Population synthesis [4–7] is the traditional approach to generate the needed number of individuals. It uses aggregate data and disaggregate data to get the synthetic population which is a representative sample of the base year but cannot remain representative over a long period of time. Moreover, activities are scattered in time and space, and the choice of activities is always impacted greatly by the demographic factors such as employment status, gender, age and having a child or not. As a result, there are still challenges to generate consistent and specific individuals with different features according to the requirement of the simulation environment.

To deal with these problems, we propose a population generating framework called Artificial Population Systems (APS), which can generate artificial populations that reflect the demographic changes and travel behavior of the real

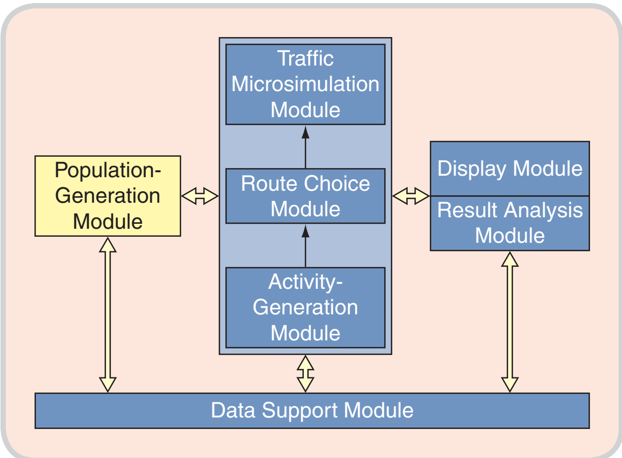


FIG 1 Important modules of traffic microsimulation.

people. The artificial population evolves and communicates with outsides according to certain governing mechanism. It is no longer the static synthetic population of the base year. So we can do experiments with artificial population to quantitatively analyze the relationship between population and traffic demand. The design of the framework is based on the ACP approach which is artificial societies for modeling, computational experiments for analysis, and parallel execution for control and management.

The rest of the paper is organized as follows. Section II is a short review of previous work in the area of population synthesis. Section III discusses the concept of artificial population systems. And the problems of agent-based modeling, experiment, quantitative analysis, and decision-making are expressed in details. Section IV describes the requirements that artificial populations for traffic micro-simulation needs to be fulfilled and then it gives a simple example of the artificial population. Section V finally concludes this paper with a brief summary and discussion on the future direction.

II. Related Work of Population Synthesis

The conventional methods of population synthesis always use aggregate data from one source and disaggregate data from another source. The aggregate data are typically drawn from aggregate census data, such as the Summary Files (SF) of the U.S. and the Small Area Statistics (SAS) file of the U.K. The disaggregate data are usually drawn from 5% representative sample of almost complete census records, such as the Public-Use Microdata Sample (PUMS) of the U.S. and the Sample of Anonymized Records (SAR) of the U.K.

Since, we know at least the marginal distributions of the demographic variables (age, sex, household income, household size, etc.) for the entire population and the complete individual population records for the 5% sample, Iterative Proportional Fitting (IPF) is used to generate individual population records for the entire population that are collectively consistent with the cross tabulations provided by the aggregate data, then the individuals are chosen from that distribution with the appropriate probability in order to get the synthetic population.

The procedure of population synthesis is illustrated in Figure 2 (individual population unit is household).

Beckman et al. [4] are the first to adopt IPF to create baseline synthetic populations of households and persons using the aggregate data of 1990 census data given in Census Standard Type File 3A (STF-3A) and the disaggregate data of the Public Use Microdata Sample (PUMS). The method is used in the population synthesizer module in TRANSIMS.

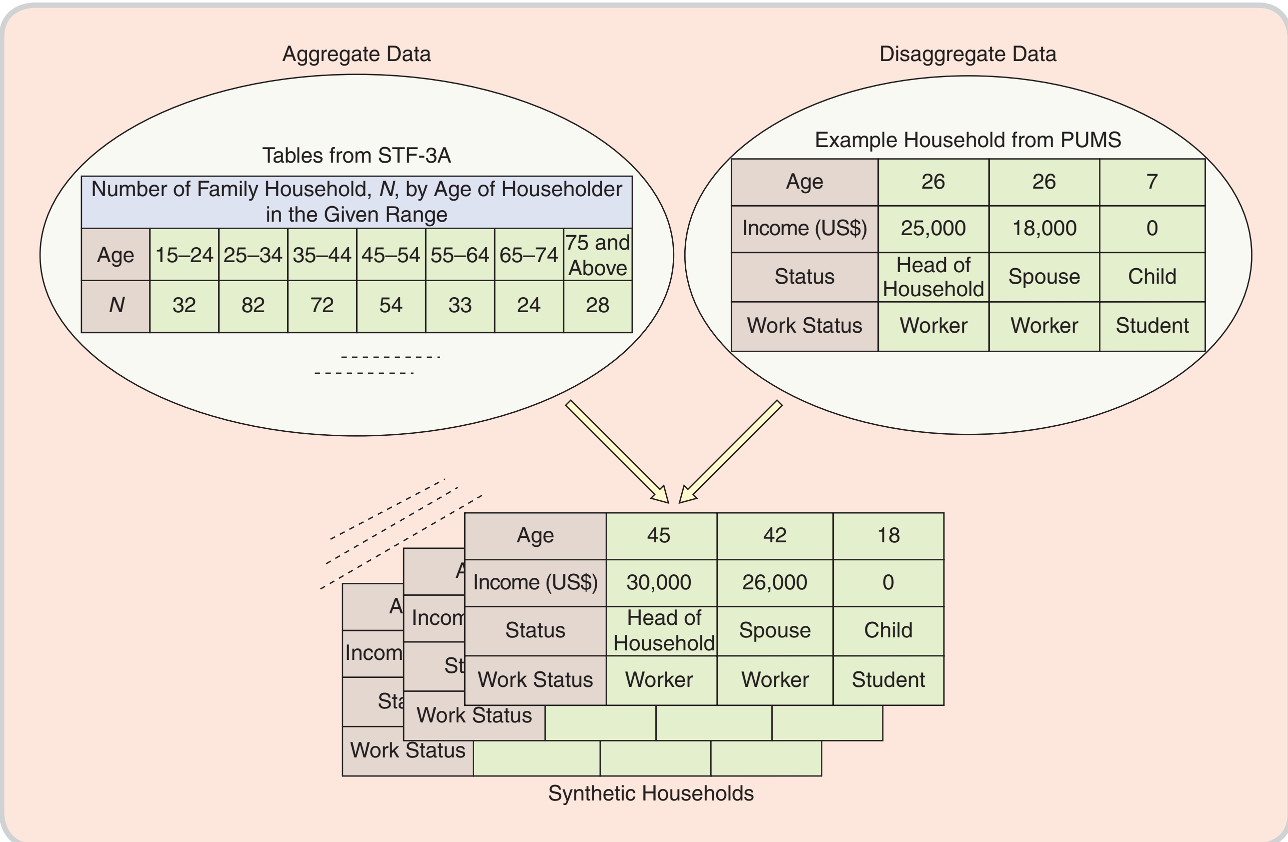


FIG 2 Population synthesis with aggregate and disaggregate data.

Frick [5] uses the Swiss Public Use Sample (PUS), which is a 5% representative sample of complete census records often used in transportation studies, as well as Swiss microcensus data to generate synthetic population with two sociodemographic variables representing the people in the Glattal. In another paper, Frick and Axhousen [6] present further results of their effort to disaggregate the available census data to a hectare based level for employment, age, and sex of the population. Guo and Bhat [7] propose modification and enhancement to Beckman et al's approach. The new procedure is capable of producing synthetic population for a target area closer to the true population compared to the conventional approach. And it is implemented into an operational software system and used to generate synthetic populations for the Dallas/Fort-Worth area in Texas.

The artificial population system can generate artificial populations which evolve with time and space.

Although these research works make good use of data from the existing data, using statistical result of the real world data still meets challenges. On one hand, almost all data, both aggregate data and disaggregate data, come from empirical data obtained by surveys. The data collection costs lots of time and some data are not available because there exist laws to protect the privacy of people. On the other hand, these approaches of generating synthetic population only create static populations, while the demographic dynamic development is not considered.

To overcome these problems, we use artificial population systems to generate artificial populations.

III. Artificial Population Systems

The idea of APS comes from the ACP approach, which consists of artificial societies for modeling, computational experiments for analysis, and parallel execution for control and management of complex systems with social and behavioral dimensions [8–11].

The ACP approach is a paradigm shift for studying complex systems, especially those involving human behaviors and social organizations. Its underlying assumption is that the human capacity for modeling, analyzing, and predicting in complex systems is fundamentally limited. In many aspects, this approach is driven by new developments in agent technology, computing architectures, and networked operational environments. It has been used to build Artificial Transportation Systems which is believed to be an effective tool in the study of the integrated, coordinated and sustainable development of transportation systems [12–15].

Population systems are typical examples of such complex systems involving human behaviors and social organizations. APS is a human population system that allodial evolutes and communicates with outsides according to certain governing mechanism. It is the digital, dynamic and integrated population database and model. It is an idea perfectly suited to complex population problems when testing and validating real population systems is either economically or legally prohibitive. It allows populations to "grow" in a bottom-up fashion, providing an alternative to real systems for experimental investigations and thus "elevating" simulations to experiments. The main problems of APS are agent-based modeling, experiment, quantitative analysis and decision-making. We will express these in details in the following three subsections.

A. Agent-Based APS Modeling

Modeling with artificial population systems has three main parts as shown in Figure 3: agents, environments, and rules for interactions.

Agents are the "people" of artificial population systems. Each agent has its own internal states and behavioral rules. Some states are fixed for the agents' life, while others change through interaction with other agents or the external environment. Environments or space is the place where agents live, which could be the real physical environments, virtual mathematical or computational process, and is generally represented as the grid formed by agents' activities

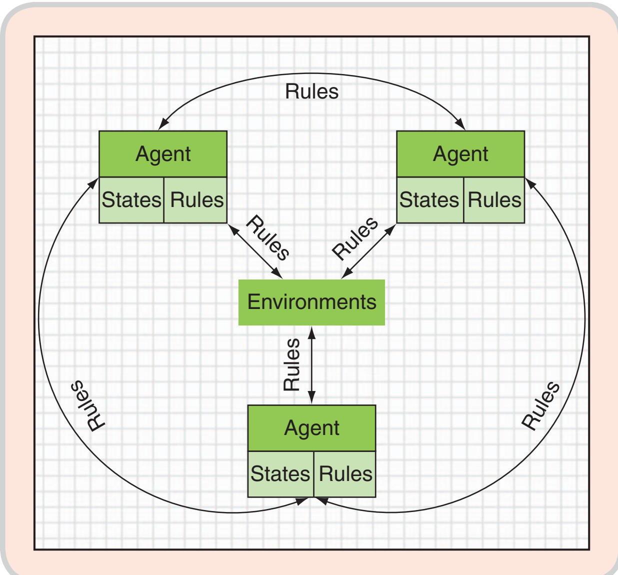


FIG 3 Agent-based modeling of artificial population systems.

With artificial populations, we will get more reasonable traffic demand forecast and analytical results in long term traffic microsimulation.

and their organizations. Rules are interacting norms and steps of agents, organizations, agents and agents, organizations and organizations, agents and organizations, from simple agents' rules to complicated culture, wars and trade rules. Individuals can be viewed as agents whose decision process is structured according to a hierarchy of if-then rules, akin to those performed by classifier systems in complex systems. This approach of agent-based modeling is built on agreeable simple objects or relationships and gives full consideration to the initialization and randomness of simple agents. Then different phenomena emerge through the interaction of simple but autonomous agents according to specified rules in given environments.

In artificial population systems, agents are individual actors that follow rules of behavior and interact with other individual subjects. It is relatively easy to include feedback mechanisms and to integrate micro-based demographic behavioral theories with aggregate-level demographic outcomes.

It is relatively easy to introduce heterogeneous agents that are not fully rational as compared to mathematical modeling. There is no doubt that the heterogeneity of populations is at the core of population models. It can differentiate the population by age, sex, family status, education and various other demographic and social characteristics, as well as those factors that cannot be observed.

Using agent-based approach, it is possible to construct models for which explicit analytical solutions do not exist, taking social interaction and generally non-linear models as examples.

Adopting a bottom-up approach, it is possible to try out the study of hypotheses on decisional processes and behavior at an individual level. The micro individual behaviors can produce complex macro situations, which is an unusual formulation for demography.

Furthermore, the accuracy of approximation to real population systems is no longer the only objective as it is in traditional computer simulations. Instead, the artificial population system is considered real—an alternative possible realization of the target society. Along this line of thinking, the real population system is also one possible realization. As a result, the behaviors of two systems, real and artificial, are different but considered equivalent for evaluation and analysis. This is the important difference between agent-based artificial population systems modeling and traditional population systems modeling.

B. Computational Experiments with APS

It is difficult and infeasible to do experiments with real population systems. Using artificial population systems, we can treat computers as population laboratories. Computational experiments are ideal tools to validate goals and objectives or evaluate strategies and decisions for a coordinated and sustainable development of artificial population systems. The theories and methods of computational experiments used to analyze some basic problems of artificial population systems will be helpful to overcome the difficulty to make experiments in real population systems, to create conditions to further analysis for the behaviors of other complex systems, and to evaluate the feasibility of its strategies. The computational experiments procedure for artificial population systems is presented in Figure 4.

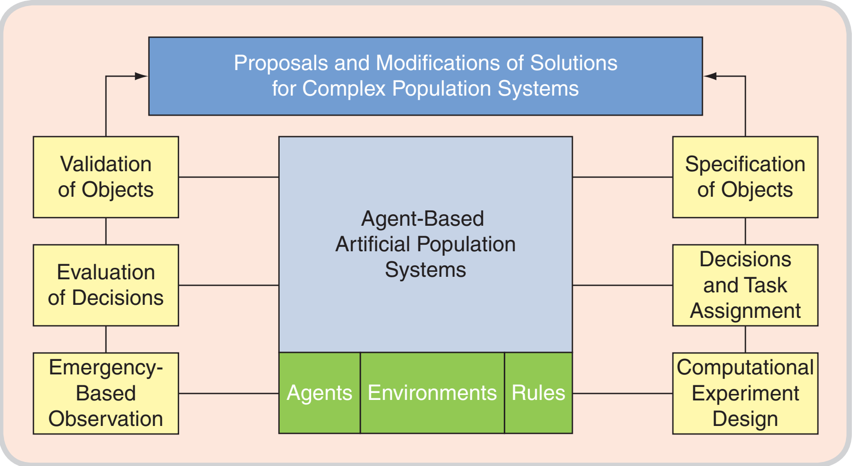


FIG 4 Computational experiments for artificial population systems.

Computational experiments are controllable. We can control the environment, specify the interaction rules, and then evaluate and quantitatively analyze various factors in demographic problems. For example, we can change the rules of decision-making processes in partnership formation to see the population-level activities and outcomes.

Computational experiments can be designed and repeated. So we can design accelerated experiments for limit, disasters or failures to see what will happen. For example, we can speed up the time steps in the artificial population systems to see the effects of the aging of the population, abnormal sex ratio at birth to the national economy, labor market in the future.

C. Parallel Execution

Parallel population systems are mainly composed of real population systems and artificial population systems as illustrated in Figure 5.

Real population systems can dynamically adjust mechanism and strategy according to the predictions of artificial population systems. Artificial population systems are used to do computational experiments, analyze the results and evaluate the effects of different population solutions. So artificial population systems can be learning and training center of management and control complex population systems.

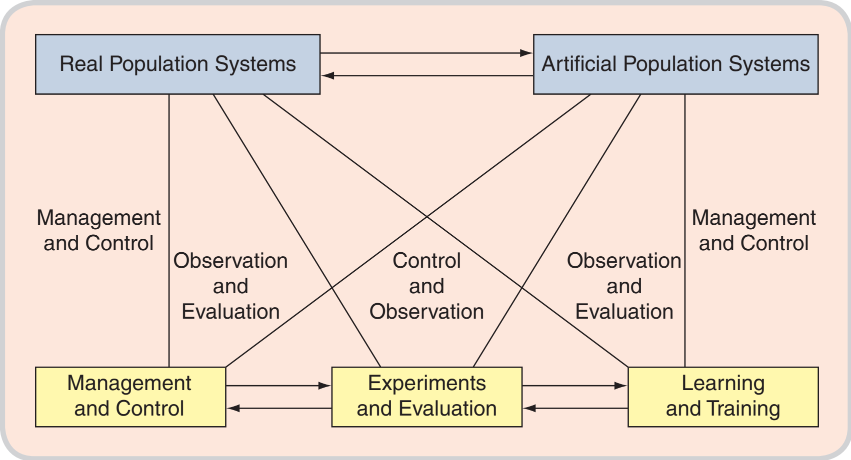


FIG 5 The framework of parallel population systems.

To realize the parallel interaction, reference and adjustment between real population systems and artificial population systems, we must firstly define the processes and protocols between real and artificial population systems. Then we need to build an evaluation system of parallel systems with multi-objective and effective solutions, internal feedback mechanisms for control and adaptation, as well as perturbation analysis and ordinal optimization for effective operation based on parallel systems. At last, we use computational experiments and parallel systems to do active and online identification of behavior patterns.

IV. Generate Artificial Populations

In traffic microsimulation, a representative sample of population may often exist which can define the set of agents whose behavior is to be simulated. However, we need a population which can be "updated" each time step within the simulation. Because the population dynamics within any area is a key component of the changing demand on urban traffic. For example, different population size, different population structure, different population mobility, and different population distribution in city reflect realistic population dynamics, which can result in travel demand of real population markedly different from travel demand of static population.

The artificial population must be extended to include methods for updating the attributes of the agents so that they continue to be representative at each point of time within the simulation. The nature of this updating obviously depends on the attributes involved, the processes being simulated, the size of the simulation time step, etc.

Let's look at a simple example. We are simulating travel demand of the city over twenty years as an assumption, which is in one-year time steps. We initially realize an artificial population of 10000. The "people" are different in age, sex, family status, education, occupation as well as other demographic and socio-economic characteristics. Demographic and socio-economic processes which are simulated as part of the updating process include births,

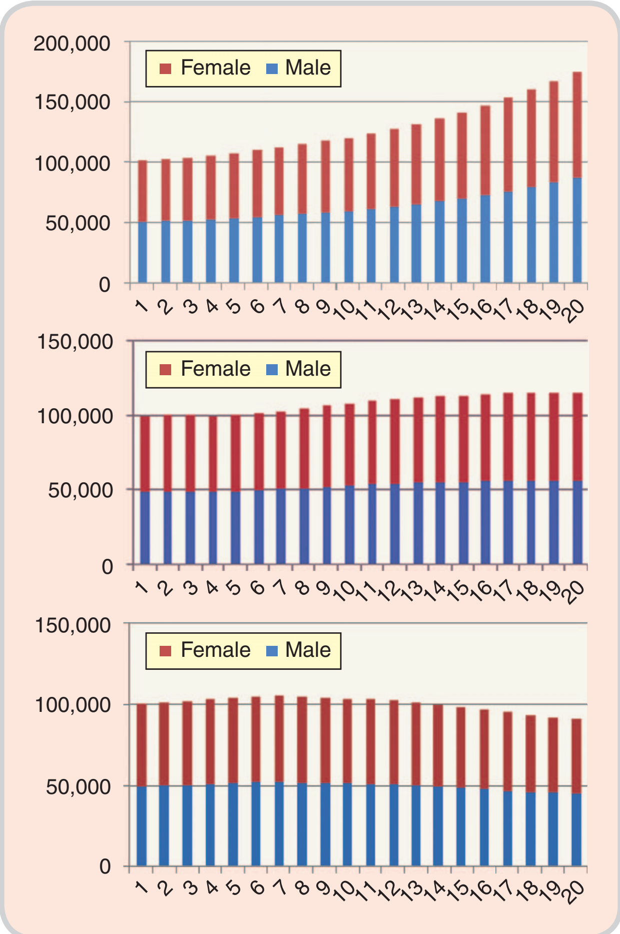


FIG 6 Examples of population dynamics.

FIG 7 Dynamics of working-age population rate.

deaths, aging, changes in labor structure, and so on. Their daily life and movements in the city are various according to their own states and behavioral rules. Travel demand of the city may change quite a lot or remain relatively the same during the twenty years because of different population dynamics.

Given different combinations of birth rate, death rate, and birth gender ratio, we can see different demographic processes in Figure 6. X axis represents time and y axis represents population size. From top to bottom, they are population I, population II and population III which respectively are growing population, asymptotically stable population and decreasing population.

According to demography, we divide the population into the children population aged 0–14, the working-age population aged 15–64, and the elderly population aged 65 and over. In Figure 7, x axis represents time and y axis represents the rate of working-age population. It shows different changing trend of population structures.

From the example, we can see that the artificial population can reflect various dynamic changes of real populations. So we will get more reasonable travel demand forecasting and analysis results in long-term traffic microsimulation.

Conclusion

In this present work, the basic concept, framework, and architecture of artificial population systems are introduced. The ACP approach in complex system study is used to generate population for traffic microsimulation. The problems of artificial population systems: agent-based modeling, experiment, quantitative analysis and decision-making, are discussed in details.

The simple example includes two demographic variables: age and gender, briefly describe the basic population updating process. The next step is to generate artificial populations with more demographic and socio-economic variables and updating process for traffic microsimulation.

Artificial population systems aim at building large-scale data sets on the attributes of individuals or households and on the attributes of individual firms or organizations and at analyzing policy impacts on these micro-units through the computational experiments of traffic, economic, demographic and social processes. Due to the fact that computers' calculation speed is increasing and computers' memory space is enlarging, modeling of single units in large networks performed in real time or faster is no more a problem.

Acknowledgment

This work was supported in part by the MOST Grants 2006CB705500 and 2007AA11Z217, NNSFC Grant 60621001, CAS Grant 2F07N04, and Shandong Province Taishan Chair Professor Fund.

References

[1] C. Bhat and F. S. Koppelman, "Activity-based modeling of travel demand," in *Handbook of Transportation Research*. New York: Kluwer, 1999.

[2] J. L. Bowman and M. E. Ben-Akiva, "Activity-based disaggregate travel demand model system with activity schedules," *Transport. Res. A*, vol. 35, no. 1, pp. 1–28, 2001.

[3] M. Ben-Akiva, J. Bottom, S. Gao, H. N. Koutsopoulos, and Y. Wen, "Towards disaggregate dynamic travel forecasting models," *Tsinghua Sci. Technol.*, vol. 12, no. 2, pp. 115–130, 2007.

[4] R. J. Beckman, K. A. Baggerly, and M. D. McKay, "Creating synthetic baseline populations," *Transport. Res. A*, vol. 30, no. 6, pp. 415–429, 1996.

[5] M. Frick, "Generating synthetic populations using Monte-Carlo techniques," in *Proc. 3rd Swiss Transport Research Conf.*, Monte Verità/Ascona, 2003.

[6] M. Frick and K. Axhousen, "Generating synthetic populations using IPF and Monte Carlo techniques: Some new results," in *Proc. 4th Swiss Transport Research Conf.*, Monte Verità/Ascona, 2004.

[7] J. Y. Guo and C. R. Bhat, "Population synthesis for microsimulating travel behavior," *Transport. Res. Record*, pp. 92–101, 2007.

[8] F.-Y. Wang, "Computational experiments for behavior analysis and decision evaluation in complex systems," *J. Syst. Simul.*, vol. 16, no. 5, pp. 893–897, 2004.

[9] F.-Y. Wang, "Parallel systems for control and management of complex systems," *J. Control Decis.*, vol. 19, no. 5, pp. 485–489, 2004.

[10] F.-Y. Wang, "On modeling, analysis, control and management of complex systems," *Complex Syst. Complexity Sci.*, vol. 3, no. 2, pp. 26–34, 2006.

[11] F.-Y. Wang, "Toward a paradigm shift in social computing: The ACP approach," *IEEE Intell. Syst.*, vol. 22, no. 5, pp. 65–67, 2007.

[12] S. M. Tang and F.-Y. Wang, "A preliminary study for basic approaches in artificial transportation systems," *J. Grad. Sch. Chin. Acad. Sci.*, vol. 23, no. 23, pp. 569–575, 2006.

[13] F.-Y. Wang and S. M. Tang, "Artificial societies for integrated and sustainable development of metropolitan systems," *IEEE Intell. Syst.*, vol. 19, no. 4, pp. 82–87, 2004.

[14] J. Y. Li, S. M. Tang, X. Q. Wang, and F.-Y. Wang, "A software architecture for artificial transportation systems—Principles and framework," in *Proc. 10th IEEE Intelligent Transportation Systems Conf.*, 2007, pp. 229–234.

[15] J. Y. Li, S. M. Tang, and F.-Y. Wang, "An investigation on ATS from the perspective of complex systems," in *Proc. 11th IEEE Intelligent Transportation Systems Conf.*, 2008, pp. 20–24.

ITS