ODD Protocol for the Replicate Model

Overview

1. Purpose and patterns

The main purpose of this model is to replicate a published Agent-based Model (ABM) based on its implemented model. The difference between the original conceptual and implemented model is presented in Section "Design Concepts: Objectives". The original model, entitled "SLUCE's original model for exploration (SOME)" (see details in Brown and Robinson 2006), was a residential location ABM that had been used in multiple academic publications. Specifically, by populating the model through the empirical survey data, the SOME model aimed to understand what effects that two types of heterogeneities, e.g., variability and categorization, represented by various value settings of 3 residential location preference factors could have on the simulated urban sprawl patterns. The SOME was not constructed based on one specific study area thereby some virtual landscapes were used to support simulations to obtain general discoveries about the effects of introducing heterogeneity on residential location selections. Since this model aims to replicate the SOME model, the overall structure and major design concepts are identical to the SOME.

2. Entities, state variables, and scales

There are three types of entities in this model: residents, service centers and patches of the landscape. Each resident has 6 different state variables which are used for three kinds of purposes: 1) "rnum" and "cluster" are used to determine which class (or group) the resident belongs to. "rnum" is a randomly generated number ranging from 1 to 100, and based on its "rnum", each resident will be assigned to one of the resident classes, which is referred by "cluster". There are totally 7 classes of residents thereby the "cluster" is an integer value ranging from 1 to 7. 2) The second type of state variables, e.g., "alphaq", "alphas", and "alphac", is used to represent the weights of each resident's 3 preference factors when selecting destination locations. Specifically, "alphaq" is the weight of residents' preferences to the aesthetic quality, "alphas" is the preference weight of the distance between residential location to the service center, and "alphac" refers to the preference weight of residents' social comfort levels in their neighborhoods. All three preference weight variables are numeric values ranging from 0 to 1. 3) The "utility" is the last type of residents' state variable, and it is an evaluation criterion of residents selecting locations. The utility value is calculated based on a function constituted by 3 preference factors and higher utility value of one location denotes a higher preference of residents to select such location.

The patches construct a discrete square grid landscape of 151 × 151 patches, and each patch contains 4 state variables beyond its coordination: The "quality" represents the aesthetic quality of each patch (location), the value of "quality" is static and stored in an external file. The "sddist" refers to the Euclidean distance between the patch and its nearest service center, and the "comfort" denotes the patches' social comfort level. The last state variable "ltype" is a Boolean variable that describes whether a patch has been selected by a resident, e.g., "ltype" = 1 means being selected and "ltype" = 0 means unoccupied. This model does not specify the patch size and the length of each time step because of its generalization. However, the model will not stop until 3420 locations have been selected. The service center as the last kind of entity does not have state variables, one center will enter the landscape at the beginning of the model running and then after every 100 residents entered the landscape, a new service center will be generated to support more residents.

3. Process overview and scheduling

In each time step, 10 residents will enter the landscape and occupy 10 desired locations. And 10 residents do not select locations simultaneously, they follow a sequential order to do the selection. For each resident, there are four sub-steps to do: 1) firstly, if residents need to be classified into different groups, residents will be randomly assigned to one of 7 classes of residents, i.e., different classes also differ in location preferences. 2) then, according to which class they belong to, residents' 3 preference weight values (i.e., "alphaq", "alphas", "alphac") will be determined based on the empirical survey data. 3) next, each resident will evaluate 15 randomly selected potential locations and select the location with the highest utility as the ultimate destination. After 10 residents entering the landscape, the model will examine whether additional service center needs to be added. Each service center can maximally service 100 residents, and if the existing centers exceed their limits, a new center will be added to accept more incoming residents. Finally, the model will check whether the number of residents is over 3420, if the answer is yes, the model will stop running. Moreover, in each time step, the output is updated, such as the number of residents entered, and plot figures of the trend of development cells beyond radius. After the last time step, the model will output a ".csv" file including multiple statistical measurements and a ".txt" file to map the final landscape.

Design Concepts

Since the basic principle of this model is to understand the challenges and opportunities to replicate a published model, the design concepts of this model will rigorously follow the original model (SOME model). Therefore, the following sections will be presented based on the SOME model.

4. Design Concepts

1) Basic principles

The basic principle of the SOME model is that in reality, residents select locations based on heterogeneous preferences, assuming homogeneous residents significantly limit model performances in investigating urban sprawl problems. Two types of heterogeneity (variability and categorization) may have different effects on shaping urban sprawl patterns, which need to be deeply understood. The SOME model examines the heterogeneity's effects by comparing outputs of experiments assuming residents' homogeneous preferences with those introducing different types of heterogeneous preferences.

2) Emergence

The urban sprawl patterns and the distribution of resident utilities emerge from residents' adaptive location choosing behaviors, the optimization-oriented utility function for residents to evaluate potential locations, and the interaction between residents and the landscape where patches have different aesthetic qualities.

3) Adaptation

There are two major resident adaptation behaviors in the SOME model. Firstly, residents do not select locations that have already been occupied, which means that earlier entrants may have advantages on more high-quality choices. Secondly, since social comfort (neighbourhood similarity) is also a factor of residents' location preferences, new residents may choose to live far away from or nearer to certain types of early entrants. In other words, early entrants may alter the land value of not only their selected locations, but their neighboring lands. And the later entrants will make adaptive choices based on updated land values made by former residents. These behaviors are based on both the rule-of-thumb and the survey data.

4) Objectives

Residents in the SOME model used a utility function (equation (1)), as the only criterion to choose locations.

$$u = (1 - |1 - \frac{1}{\mathrm{sddist}}|)^{alphas} \times (1 - |1 - quality|)^{alphaq} \times (1 - |0 - comfort|)^{alphac} \quad (1)$$

In this utility function, 'sddist' is the distance to the nearest service center, 'quality' is the aesthetic quality of the location, and 'comfort' is the neighborhood similarity value of the location, while 'alphas', 'alphaq', and 'alphac' are the weight for three factors respectively. Each resident will randomly select 15 locations to evaluate their utility values and the objective is to select the one with the highest utility. Three factors ('sddist', 'quality', and 'comfort') are chosen based on existing literature related to residential location selection models. Residents' values of 'sddist' and 'quality' are obtained through the interaction between residents and landscape patches, and 'comfort' values are calculated through the interaction between new entrant and its neighboring existing early entrants.

To be clarified, the conceptual SOME model required that three preference factors (alphas, alphaq, alphac) should be constrained to sum to one (normalization); however, its implemented model (the source code) does not include this constraint. Since this replicate model aims to replicate based on the original implemented model, no normalization would be done in this replicate model.

5) Sensing

In the SOME model, sensing plays a critical rule to guide agent behaviours. Each resident has 15 potential locations (patches) to evaluate, the 'quality' value (aesthetic quality level), the x- and y- coordinates of these patches can be identified by the corresponding resident. Additionally, the 'alphas', 'alphaq', and 'alphac' values of surrounding residents of the potential patch can be sensed by a new entrant to calculate the neighborhood similarity. The coordinates of the nearest service center to the potential patch are also accessible by the new entrant to calculate the distance to the service center. Moreover, all new entrants can know all patches' status of whether having been occupied before they are selecting. Finally, the service center can know the number of residents who view it as the nearest service center and make sure the maximal number of residents it services is 100.

6) Interaction

Residents who have selected locations (patches) on the landscape can directly update the destination patches' status from unoccupied to occupied and the occupied patch is not an available choice to the subsequent entrants. Residents can also interact with its neighboring residents through the mediated interaction, e.g., by knowing their neighbors' preference weights to calculate neighborhood similarity, but cannot alter the values of their neighbors' state variables.

7) Stochasticity

Stochasticity is used to represent variabilities in allocating classes (groups) of residents, determining preference weight values, and potential location selections. There are 7 groups of residents and each time before a new resident entering the landscape, it will be randomly allocated to one of 7 groups. This stochastic random allocation may cause that 7 groups of residents do not have an identical number of people, however, in reality, it is also impossible that the number of different groups of people is the same. Additionally, in certain experiments (some experiments assume homogeneous preferences), residents within the same group may have similar preference weights rather than the absolutely same weights. This is because even people in the same household and from the same group may still have slight differences in some preferences. The stochastic preference weight values ensure more heterogeneity within the same group. To achieve this, residents in the same group will have an acceptable preference weight value range, and the ultimate value is randomly generated from the acceptable range. Finally, each resident has randomly selected 15 potential locations to evaluate rather than predefined potential choices. This helps to simulate that residents have a wide range of choices and have a very high level of uncertainties when selecting their next housing locations.

10) Collectives

Each of 7 groups of residents is collective of the SOME model. Residents within the same group have similar behaviors, such as similar preferences when selecting residential locations. The collectives are represented by value settings of preference weights, they are predefined rather than emerged from the model running. Agent behaviors might be affected by defining collectives because residents from the same group may tend to select certain types of lands that make residents from other groups have no opportunities to choose those lands.

11) Observation

Multiple observations are available in the SOME model to help to explore the research purposes. To examine the spatial patterns of the development areas (development area is the location that has been selected by residents), four landscape metrics, which are the largest patch index (LPI), mean patch size (MPS), edge density (ED), and the mean nearest neighbor (MNN). Additionally, to measure the residential location sprawl, a circular area that has a radius of 31 patches from the first service center is determined, and the number of cells occupied by residents outside this circular area (DOR) will be another observation of the model. Moreover, the proportion of residents outside the circular area of each of 7 resident groups (POR) will be recorded to investigate results by the resident group. Then, each group of residents will output a mean resident utility value (MRU) and a variance of utility value (VRU) to measure the distribution of agent utilities, and another MRU and VRU of the entire residents will also be calculated. Furthermore, the GINI coefficient (GINI) of each resident group and the entire residents will also be part of the observations to monitor the disparity of the population. Next, an evenness value (EVENNESS) to measure the equality of 3 preference weight values within the same resident group will be calculated by each group. To be clarified, 4 landscape metrics (LPI, MPS, ED, and MNN) should be calculated by an external software named "FRAGSTATS", to achieve this, one ".txt" file which stores the landscape after the model running should be exported from the model and imported into the "FRAGSTATS". And the other measurements (DOR, POR, MRU, VRU, GINI, AND EVENNESS) will be exported altogether from the model through a ".csv" file. Finally, a graphic of the total residents outside the predefined circular area will not be exported but is also a source of observations to monitor the trend of urban sprawl.

Details

5. Initialization

There are 5 different versions of the model to process 5 different experiment settings. The differences among 5 experiments only reflect on setting different groups of resident preference weight values (‘alphas’, ‘alphaq’, and ‘alphac’). Different versions of the model have been stored as separated “.nlogo” files, which means that the initial settings of the preference weight values have been predefined in each “.nlogo”, therefore, the modeler only need to open the specific version and the initialization could be automatically done.

In all model versions, the landscape of the model (the aesthetic quality value of each patch) should be initialized when the model starts. The landscape is stored in a ‘.txt’ file (“151aesth.txt”) and is imported in the model before the first time step, and the “151aesth.txt” file should be stored in the same folder as the model files (‘.nlogo’ files). Beyond the landscape file, there are alternative two values of global variables should be initialized in all model versions: 1) ‘numtests’ refers to how many potential locations should be evaluated by each resident before selecting the final destination. In this model, it should be set as 15; 2) ‘radius’ refers to the radius of the circular area around the first service center, and it should be set as 31 in the SOME model.

6. Input data

There is only 1 input data in this model, which is the landscape file (“151aesth.txt”) to store each patch’s quality.

7. Submodels

The sub-model to set up residents 3 preference weights differs in 5 experiments: 1) “Uniform” experiment: all three preference weights of all residents are set to be a random floating number ranging from 0 to 1; 2) “Homogeneous” experiment: for all residents, ‘alphas’ is set to 0.536, ‘alphaq’ is set to 0.532 and ‘alphac’ is set to 0.494. 3) “Normal” experiment: for all residents, ‘alphas’ is a random floating number generated based on a normal distribution of 0.536 as the mean value and 0.244 as the standard deviation (Std), ‘alphaq’ is a random floating number from a normal distribution which mean value is 0.532 and the Std is 0.261, and ‘alphac’ is also a random floating number from a normal distribution (mean is 0.494 and Std is 0.248). 4) “Group Means” experiment: seven groups’ 3 preference weight values are set according to the following Table:

	alphac	*alphas*	*Alphaq*
Group 1	0.296	0.326	0.373
Group 2	0.776	0.378	0.411
Group 3	0.217	0.773	0.903
Group 4	0.602	0.337	0.687
Group 5	0.450	0.769	0.335
Group 6	0.529	0.787	0.330
Group 7	0.527	0.466	0.756

5) “Group Normals” experiment: 3 preference weight values of 7 resident groups are all generated from the normal distributions, the mean and std values of each normal distribution of each weight will be presented as following:

	alphac		*alphas*		*alphaq*	
	mean	std	mean	std	mean	std
Group 1	0.296	0.264	0.326	0.224	0.373	0.246
Group 2	0.776	0.264	0.378	0.264	0.441	0.251

Group 3	0.217	0.309	0.773	0.430	0.903	0.565
Group 4	0.602	0.354	0.337	0.300	0.687	0.305
Group 5	0.450	0.444	0.769	0.341	0.335	0.323
Group 6	0.529	0.381	0.787	0.327	0.330	0.287
Group 7	0.527	0.300	0.466	0.255	0.756	0.224

The sub-model to calculate the "comfort" value uses the same equation in all experiments (equation (2)).

$$\gamma_{r(x,y)} = \frac{\sum_{j=1}^{n}\sqrt{((alphas_r - alphas_j)^2 + (alphaq_r - alphaq_j)^2 + (alphac_r - alphac_j)^2)}}{\sqrt{3}*n} \quad (2)$$

In equation (2), "r" is the target resident per se, "j" means the jth neighboring residents of "r" (in this model, each resident has 8 neighbours). "n" is the number of neighboring residents that have occupied locations.

The last sub-model, e.g., allocating groups, is only applicable in "Group Normals" and "Group Means" experiments. Each time before a resident selecting locations, its state variable "rnum" will be set to a random floating number ranging from 0 to 100. Then, based on the "rnum" value, the state variable "cluster" of residents will be set to a specific number. The criteria for setting the "cluster" number according to the "rnum" number is in the following table:

		"rnum" value
" cluster" value	1	0 <= rnum <= 24
	2	24 < rnum <= 42
	3	42 < rnum <= 48
	4	48 < rnum <= 59
	5	59 < rnum <= 71
	6	71 < rnum <= 84
	7	84< rnum <= 100

To be clarified, the "cluster" value indicates which group the resident belongs to.

Reference

Brown, D.G. and Robinson, D.T., 2006. Effects of heterogeneity in residential preferences on an agent-based model of urban sprawl. *Ecology and society*, *11*(1).