Agent-based modelling for spatial simulation and optimization

Xia Li
School of Geography and Planning
Sun Yat-sen University

What are agents?

- The term “agent” is used (and misused) increasingly to describe a broad range of computational entities. This tends to obscure the differences between radically different approaches.
  - Some agents perform tasks individually... others need to work together.
  - Some are mobile... some static.
  - Agents communicate via messages... some don't communicate at all.
  - Some learn and adapt... others don't.

- Despite this diversity, we can identify some common properties.

Properties of Agents

<table>
<thead>
<tr>
<th>Property</th>
<th>Other Names</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve</td>
<td>Sensing &amp; acting</td>
<td>Responds in a timely fashion to changes in the environment.</td>
</tr>
<tr>
<td>Autonomous</td>
<td>Pro-active</td>
<td>Exercises control over its own actions.</td>
</tr>
<tr>
<td>Goal-oriented</td>
<td>Futuristic</td>
<td>Does not simply act in response to the environment.</td>
</tr>
<tr>
<td>Transparency</td>
<td>Collaborative</td>
<td>Is a combination of connecting processes.</td>
</tr>
<tr>
<td>Communicative</td>
<td>Socially active</td>
<td>Communicate with other agents, perhaps including people.</td>
</tr>
<tr>
<td>Learning</td>
<td>Adaptive</td>
<td>Changes its behavior based on its previous experiences.</td>
</tr>
<tr>
<td>Mobile</td>
<td>Able</td>
<td>Able to transport itself from one machine to another.</td>
</tr>
<tr>
<td>Flexible</td>
<td>Actors</td>
<td>Actors are not assigned.</td>
</tr>
<tr>
<td>Character</td>
<td>Behavioral</td>
<td>Behavioral personality and emotional state.</td>
</tr>
</tbody>
</table>

Agents and environments

- Recently, agent-based models (ABMs) have attracted growing attention because of their effectiveness in representing human and social factors for urban and land use modeling.
- ABMs can address the fact that the accumulated impact of individual decisions causes human-induced environmental changes.
The prerequisite of ABMs is that individual-level information must be available for constructing these models. Census data, GIS data, or investigation data at individual-level can be used to define agents’ behavior (Li and Liu 2007). When these individual data are available, agents with the potential for heterogeneous preferences can be used to represent discrete choices for estimating the likely locations of development (Brown et al. 2004). ABMs can address the fact that the accumulated impact of individual decisions causes human-induced environmental changes. The implementation of ABMs requires the understanding of human behavior and nature’s responses so that the real world can be translated into formal model specifications. For example, economic theories and urban growth theories can be incorporated in the definitions of agents’ behaviors. The implementation of ABMs requires sophisticated techniques, such as sample surveys, participant observation, and model configuration and calibration.

The model structures of ABM is not so simple as CA. It should take considerable effort to understand even the simplest of ABMs (Brown et al. 2004). Generally, the implementation of ABMs for modeling land use dynamics involves elaborated techniques, such as sample surveys, participant observation, field and laboratory experiments, companion modeling, and GIS and remotely sensed data (Robinson et al. 2007). The heterogeneity of households, and specialized household surveys should provide the basic information for constructing ABMs. However, some information may be difficult to collect as people may not know how to express themselves in a quantifiable way or may not wish to reveal certain information.

Spatial simulation and optimization
We can define two major types of agents:
- human-like agents (individual residents, household, institutions, organizations);
- animal-like agents (birds, fish, and ants)

Spatial simulation and optimization
Human-like agents for simulating land use dynamics
- the Time-extended model from urban economic theories; a simple MCE method by using GIS data
Animal-like agents for solving complex spatial optimization problems
- Artificial ant intelligence
Simulation aims to generate realistic scenarios under given conditions, whereas the goal of optimization is to provide optimal solution(s) to a given planning problem.

In terms of Yeh’s framework that classifies the tasks of GIS in planning into three categories: description, prediction, and prescription (Yeh 1999), simulation is a major approach to prediction, whereas optimization belongs to prescription.

When being performed separately, simulation usually adopts an inertial strategy, i.e., it assumes that future development will follow the historical trend (Liu et al. 2010), and optimization in most cases also assumes the problem to be static.

Case study 1

Defining agents’ behaviors to simulate complex residential development using MCE

Introduction

- Fast residential development has been witnessed in many Chinese cities because of growing population.
- The residential development encroaching on valuable agricultural land has caused intensified land-use conflicts.
- Urban sprawl and fragmented use of land resources are a major problem for these fast growing cities (Yeh and Li, 2001).
- There are growing concerns on the side effects of growth such as sprawl, congestion, housing affordability, and loss of open space (Waddell, 2002).
- Urban models can be used to predict how a region will change according to the existing trajectory of development.

Introduction (cont)

- Cellular automata (CA) have attracted increasing attention as a powerful modeling tool in simulating geographical phenomena.
- Studies have demonstrated that very complex behaviors and global patterns can be generated by applying some simple local rules in CA models.
- Simulation of complex urban systems is one of the successful examples of using CA.
- Modeling urban systems can help to understand the mechanisms of urban evolution and examine existing urban theories.
CA models may have apparent limitations on reflecting the decisions and behaviors of individuals, such as governments, residents and investors, in shaping urban dynamics.

The influences of human factors are difficult to implement in traditional CA models.

Agents' spatial behaviors should be considered so that complex spatial interactions can be addressed in urban simulation.

Actually, agent-based models (ABMs) have been widely employed to represent individual decision-makers in social science research.

It is still an initial stage in the application of ABMs in urban modeling. A major problem with ABMs is how to define agents’ properties (attributes) by using empirical data. There is a general lack of methodology for defining agents’ decision behaviors in a more consistent way.

In the past, multi-agent models were used mostly in purely social contexts. However, these types of agent-models have few spatial details. It is considered that these ABMs methodologies and existing tools are over-general and underestimate, if not ignore, the importance of space and spatial behavior.

This study proposes a multi-agent system integrated with CA and GIS to simulate residential development in urban systems.

Agents make decisions under the constraints of the environment, which is heterogeneous in a two-dimensional landscape and subject to change in time.
The environmental layers

- Land use
  - Land use is one of the most important factors that should be considered in urban simulation.
  - This factor provides the basic environment for agents to make decisions, but it is also subject to changes with regard to agents' activities.
  - Agents will have different decision behaviors related to land-use types.

Integration of MAS with CA and GIS

- This integrated model consists of two major types of layers—immobile environment layers and mobile agent layers.
- The environment is not uniform in spatio-temporal dimensions when using GIS.
- A GIS database can provide the basic spatial information for agents' location choices.
- The agent layers are used to accommodate the mobile entities that play a key role in shaping the evolution of urban structures.

Land price

- Land price is an important element in determining housing prices which are a major concern for a potential home buyer.
- Residents with high incomes can afford to buy good-quality homes in locations with high land prices.

Surrounding environment

- The attraction of a site for residential development is also related to its surrounding living environment.
- In this study, the amenity of the surrounding environment is mainly based on green space and water.
- Therefore, the amenity is measured by using two indicators, the percentages of green land and water in the neighborhood.

A moving window is used to calculate the percentages of green land and water in the classified TM image which has a resolution of 30 m.
- A window size of 9*9 is adopted because this size seems to be appropriate to capture the spatial patterns of these two variables in the study area according to experiments.
- Finally, the utility (attraction) of a site related to this amenity is obtained by using the following equation:

\[ E_{\text{ams}} = \frac{1}{2} G_{\text{perc}} + \frac{1}{2} W_{\text{perc}} \]
### Accessibility

- Accessibility represents a locational characteristic that permits a place to be reached by the efforts of those at other places using various transport tools.
- It is related to its geographical location (e.g. distance to roads and town centers) and the conditions of road networks.

*The site will be more likely to develop if it is easily accessed.*

The accessibility can be conveniently calculated by using GIS functions. The utility (benefits) related to accessibility can be represented as follows:

\[
B_{access} = \frac{1}{3} e^{-\gamma_1 D_{road}} + \frac{1}{3} e^{-\gamma_2 D_{neighbor}} + \frac{1}{3} e^{-\gamma_3 D_{town}}
\]

### General Public Facilities

- The provision of public facilities is another important factor to affect the decision of a potential buyer.
- The sites will be more attractive if they are closer to general public facilities, such as hospitals, gardens, commercial centers and entertainment centers.

*Distance decay functions can be defined to reflect the utility of a site in terms of facility provision.*

The utility is calculated as follows:

\[
B_{gfac} = \frac{1}{4} e^{-\eta_1 D_{hospital}} + \frac{1}{4} e^{-\eta_2 D_{garden}} + \frac{1}{4} e^{-\eta_3 D_{commercial}} + \frac{1}{4} e^{-\eta_4 D_{entertainment}}
\]

### Education Benefits

- Education is a special type of public facility, and is the most important factor affecting home buying for the Chinese.
- Distance functions can also be used to represent the convenience of accessing education facilities, such as schools and libraries.
- More benefits can be achieved if there is a shorter distance to these facilities.

*Education is a special type of public facility, and is the most important factor affecting home buying for the Chinese.*

\[
B_{edu} = \frac{1}{2} e^{-\delta_1 D_{school}} + \frac{1}{2} e^{-\delta_2 D_{library}}
\]

where \(B_{edu}\) is the utility related to the provision of educational facilities in terms of schools and public libraries; \(D_{school}\) and \(D_{library}\) are the Euclidian distances to these facilities, respectively.
Defining agents’ decision behaviors using MCE techniques

- This model has three major types of agents—residents, property developers, and governments.
- An agent can represent not only a single individual, but also a group of individuals.
- In this model, an agent represents a number of residents according to the total population.
- Each type of agent has unique features.

- The definition of agents’ decision behavior is essential to agent-based models.
- In most situations, agents’ behaviors can be defined heuristically since there are not unique methods.
- However, it is more robust to define agents’ decision behaviors based on GIS data.
- Each major type of agents can be divided into subtypes according to the agents’ properties.
- These agents will influence each other in making decisions.

Resident agents

- There are two kinds of residents—new residents moving in from outside and existing residents relocating to new places to live.
- The behaviors of these mobile residents will affect the investment strategies of property developers.
- The interactions between these agents are responsible for the formation and evolution of urban structures, such as social and ethnic segregation, self-organization and urban expansion.

- A utility function is defined for a resident agent to assess the value of a potential site as a residence.
- His main objective is to maximize the utility function as much as possible in location decisions.
- The utility function of location \( (i,j) \) for resident agent \( k \) can be represented as follows:

\[
U(k, j) = w_{\text{size}}B_{\text{size}} + w_{\text{cost}}B_{\text{cost}} + w_{\text{amen}}B_{\text{amen}} + w_{\text{nat}}B_{\text{nat}} + w_{\text{soc}}B_{\text{soc}} + c_i
\]

- Resident agents may have different preferences in choosing locations for residency, which can be reflected by the weights in the utility function.
- In this study, these weights are decided by using Saaty’s pairwise comparison procedure.
- The comparison is mainly based on experts’ knowledge and preferences.

- For resident \( k \), the probability of location \( (i,j) \) to be selected is equal to the utility probability that the utility value at that location is greater or equal to those at other locations:

\[
P^f_{\text{resi}}(k,j) = P(U(k,j) \geq U(k,i)) = \frac{\exp(U(k,j))}{\sum_i \exp(U(k,i))}
\]
### Resident agents
- After a satisfactory location has been identified by a resident agent, there are three situations:
  1. The location has been developed and occupied by another resident agent;
  2. The location has been developed and is available for residency;
  3. The location has not been developed.
- The third situation is the essential part of this study.

### Developer agents
- Property developers play an important role in influencing residential development in fast growing cities.
- They need to consider the preferences of residents in home buying and the policies of governments in managing land resources.
- The main criterion is to achieve a certain amount of profit above expectations.
- This criterion can be used to determine the decision behaviors of developer agents.

### Developer agents
- The following equation is used for the assessment of development potential:
  \[ D_{\text{dev}} = W_{\text{dev}} - L_{\text{dev}} - D_{\text{cut}} \]
- The development probability related to developer agents can thus be represented as follows:
  \[ P_{\text{dev},\text{agent}}(k, l) = \frac{W_{\text{dev}} - L_{\text{dev}} - D_{\text{cut}}}{\text{Exp} - D_{\text{cut}}} \]

### Government agents
- Government agents will decide if an application for land development is approved or not according to a number of factors.
- They make decisions not only by considering environmental factors, but also by communicating with resident agents and developer agents.
- Government agents will examine the development suitability of a site according to land use, surrounding environment, transportation, general facilities, and educational benefits.
- The initial development probability of each site can be estimated based on these environmental factors.
- Firstly, existing land use is a major factor in determining land use conversion.
- Secondly, the approval probability is also related to development plans.
The behaviors of government agents are also affected by the behaviors of resident agents and developer agents. When a site has been requested for residency more often by resident agents in the simulation, it has a higher probability of development. This interaction reflects the communication and negotiation processes in urban simulation.

The following equation can be used to represent this type of negotiation between government agents, resident agents and developer agents in affecting the development probability of a cell \((i,j)\):

\[
\begin{align*}
\Delta P_{d(i,j)} &= P_{d(i,j)} - P_{d(i,j)}^{\text{old}} + \alpha \Delta P_1 + \beta \Delta P_2 \\
\text{If } P_{d(i,j)}>1, \text{ then } P_{d(i,j)} = 1
\end{align*}
\]

The behaviors of government agents are also affected by the behaviors of resident agents and developer agents. When a site has been requested for residency more often by resident agents in the simulation, it has a higher probability of development. This interaction reflects the communication and negotiation processes in urban simulation.

The following equation can be used to represent this type of negotiation between government agents, resident agents and developer agents in affecting the development probability of a cell \((i,j)\):

\[
\begin{align*}
\Delta P_{d(i,j)} &= P_{d(i,j)} - P_{d(i,j)}^{\text{old}} + \alpha \Delta P_1 + \beta \Delta P_2 \\
\text{If } P_{d(i,j)}>1, \text{ then } P_{d(i,j)} = 1
\end{align*}
\]

CA mainly focuses on local interactions of physical factors whereas agent-based systems pay much attention to individual’s behaviors. The integration of these two models is essential for the simulation of residential development.

In this study, the development probability related to physical factors is estimated by using a standard-CA. It is based on a logistic-CA model by taking \(s\) number of proximity variables into account (Wu, 2002):

\[
P_{d(i,j)} = \frac{1}{1 + \exp\left(-d + \sum_{k=1}^{s} \alpha_k n_k(i)\right)}
\]

The final decision is based on the interactions between government agents, resident agents, developer agents and the environment. The probability for residential development at a cell is obtained by the product of these probabilities:

\[
P_{d(i,j)} = P_{\text{residential}(i,j)} \cdot P_{\text{developer}(i,j)} \cdot P_{\text{government}(i,j)}
\]

The procedure of modeling between \(t\) and \(t+1\) follows this order:

- Estimating the potential annual sprawl with CA-
- Application for residence by resident agents-
- Estimating the demand and benefit for developer agents before investment-
- Evaluation of the applications by government agents before issuing the permit.
Application

- Study area and data
- Defining resident agents’ properties
- Simulation results
- Model validation

Study area and data

- The Haizhu district of Guangzhou city has been selected for testing the proposed model.
- The model is used to simulate the land-use dynamics from 1995 to 2004 for this fast growing region.
- Landsat TM images dated on 30 December 1995 and 13 June 2004 were used to obtain training data about actual land-use conversion, which can be used to calibrate the logistic-CA model.

GIS data were also used to represent the independent factors that determine land-use changes.
- Social and economic data were also obtained from statistical yearbooks and the Fifth National Census.
- All the original raster data are re-sampled into the resolution of 100*100m to reduce the computation time for the agent-based model.

Study area and data

GIS data were also used to represent the independent factors that determine land-use changes.
- Social and economic data were also obtained from statistical yearbooks and the Fifth National Census.
- All the original raster data are re-sampled into the resolution of 100*100m to reduce the computation time for the agent-based model.

GIS data were also used to represent the independent factors that determine land-use changes.
- Social and economic data were also obtained from statistical yearbooks and the Fifth National Census.
- All the original raster data are re-sampled into the resolution of 100*100m to reduce the computation time for the agent-based model.

Study area and data

GIS data were also used to represent the independent factors that determine land-use changes.
- Social and economic data were also obtained from statistical yearbooks and the Fifth National Census.
- All the original raster data are re-sampled into the resolution of 100*100m to reduce the computation time for the agent-based model.

Defining resident agents’ properties

- Detailed information for each agent is unavailable in most situations.
- Agents should be classified into a few categories so that their properties can be defined.
- The attributes for the aggregated agents are obtained by using social and economic data.
- Each group of resident agents has distinct behaviors or preferences in the location choice of residency.
- In this model, their preferences are reflected by the weights in the utility function. These weights are given according to experts’ experiences.

![Utility models of cultural spatial variables prepared by a geospatial GIS](image)

<table>
<thead>
<tr>
<th>Group of Resident Agents</th>
<th>Without Children</th>
<th>With Children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household size</td>
<td>Low income</td>
<td>Middle income</td>
</tr>
<tr>
<td>Income</td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>Low</td>
<td>9%</td>
<td>39%</td>
</tr>
<tr>
<td>Middle</td>
<td>9%</td>
<td>31%</td>
</tr>
<tr>
<td>High</td>
<td>9%</td>
<td>6%</td>
</tr>
</tbody>
</table>
Simulation results

The procedures for simulating residential development are as follows:

1) Determining the total number of resident agents according to the actual amount of residential development in 1995-2004 from the classification of remote sensing data;

2) Using Monto Carlo method to create resident agents in space using the actual proportion of various types of residents in census data;

3) Compute location utility for each type of resident agents. Selecting the locations with the highest utility values and estimating development probability for these places according to the interactions between residents, property developers and governments;

4) Determining whether the location of the highest utility value will be developed by using the Monto Carlo method. If yes, the location will be marked and go to step 1 to start the next round of site selection. If no, the next site will be evaluated by choosing the second highest utility value;

5) This procedure continues until all the residents have been accommodated.

### Table: Weights for different groups of resident agents obtained by using the Sazy’s method

<table>
<thead>
<tr>
<th>Types of residents</th>
<th>Land price</th>
<th>Surrounding environment</th>
<th>Accessibility</th>
<th>Public facilities</th>
<th>Education</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low income without children</td>
<td>0.443</td>
<td>0.009</td>
<td>0.204</td>
<td>0.155</td>
<td>0.105</td>
<td>1</td>
</tr>
<tr>
<td>Low income with children</td>
<td>0.491</td>
<td>0.083</td>
<td>0.154</td>
<td>0.088</td>
<td>0.281</td>
<td>1</td>
</tr>
<tr>
<td>Middle income without children</td>
<td>0.175</td>
<td>0.279</td>
<td>0.165</td>
<td>0.194</td>
<td>0.097</td>
<td>1</td>
</tr>
<tr>
<td>Middle income with children</td>
<td>0.216</td>
<td>0.076</td>
<td>0.143</td>
<td>0.140</td>
<td>0.222</td>
<td>1</td>
</tr>
<tr>
<td>High income without children</td>
<td>0.048</td>
<td>0.526</td>
<td>0.194</td>
<td>0.141</td>
<td>0.091</td>
<td>1</td>
</tr>
<tr>
<td>High income with children</td>
<td>0.084</td>
<td>0.434</td>
<td>0.171</td>
<td>0.076</td>
<td>0.235</td>
<td>1</td>
</tr>
</tbody>
</table>
Model validation

- Model validation is usually required when urban models are applied to the simulation of real cities.
- In this study, the proposed model is assessed in two ways:
  1. comparing the simulated patterns with the actual ones;
  2. comparing the simulation patterns between the integrated CA-agent model and the pure CA model.

---

The visual comparison can provide a rough estimation about the accuracy of this proposed model for simulating urban development.

A further quantitative analysis is to produce a confusion matrix indicating the concordance between the simulated and the actual development patterns.

In this study, the simple indicator Moran’s I is applied to the measurement of land-use patterns.

---

Multi-agent systems usually involve numerous parameters.

These parameters are usually heuristically defined according to the approach of trial and error.

Different sets of parameters can be defined and experiments are required to assess the model’s sensitivity to parameters.
Model validation

- Compare this integrated model with the pure CA models.
- The overall accuracy of a logistic-CA is 0.510, and the Moran’s $I$ is 0.597 for this CA model.
- This indicates that the agent-based model can produce better outcomes for modeling residential development in this study area.
- This is because the behavior of various players in actual urban systems can be well addressed by using this agent-based approach.

Conclusion

- The simulation and prediction of urban growth is important for urban planners to formulate sustainable development strategies.
- Cities are complex systems which are difficult to represent by using mathematical equations.
- One problem with CA models is that human and social factors are difficult to incorporate in the simulation.
- Agent-based modeling techniques can be used to simulate the complex residential development which involves various players in shaping urban morphology.

Case study 1

Ant intelligence for solving spatial optimization problems

1. Spatial optimization problem and ant intelligence
2. Ant intelligence for sites selection
3. Ant intelligence for path optimization
4. Ant intelligence for area object optimization
5. Conclusion and expectation
Spatial optimization problem

An often encountered spatial decision problem is to search for the best site or sites to accommodate one or more facilities to generate the best utility values (e.g. the maximum population coverage and minimum transport cost).

Traditional location-allocation methods before GIS data were available only use relatively small datasets. An effective method is needed for spatial optimization.

Current algorithms and method

- Multi-criteria evaluation
- Linear programming
- Heuristic algorithms: genetic algorithm, GA; simulated annealing, SA; Tabu search

We propose ant colony optimization (ACO)

Ant colony optimization

- Swarm intelligence methods are used to solve such complicated problems.
- Particle swarm optimization, ant colony optimization.
- ACO is a type of computer algorithms for solving combinatorial optimization problems using artificial intelligence.
- ACO is devised by simulating ants’ behaviors of selecting the best route from a food source to their nest. Ants release and deposit pheromone on the ground along their way, to guide others in finding foods efficiently.

Character of ACO

- The activity of single ant is very simple, but the behalf of colony are very complicated
- ACO, as a heuristic algorithm, has the merits of searching intelligently, optimizing globally, robust and positive feedback. This method is based on the positive feedback of artificial ants, in which the coordination among ants is achieved by exploiting the stigmergic communication mechanism. An ant colony system is robust, and the integrity of the overall system is not easily affected by the failure of one or several agents
- ACO has been applied to the solution of traveling salesman problems successfully.

Basic of ant colony algorithms

- Ant intelligence for solving the TSP
  The transition probability from city $u$ to city $v$ for the $k$th ant at time $t$ is given as follows:
  $$p_{uv}^k(t) = \begin{cases} \frac{[\tau_{uv}^k(t)][\eta_{uv}^k(t)]^\alpha}{\sum_{w \in \text{allowed}} [\tau_{uw}^k(t)][\eta_{uw}^k(t)]^\alpha} & \text{if } v \text{ is allowed}, \\ 0 & \text{otherwise} \end{cases}$$

  $\tau_{uv}^k(t)$ is the amount of pheromone trail on edge $(u,v)$
  $\eta_{uv}^k(t)$ is a heuristic function
An ant has a higher probability of selecting the shorter route between two cities. The heuristic function is defined as the inverse of the distance between cities \( u \) and \( v \):

\[
\eta_{uv}(t) = \frac{1}{d_{uv}}
\]

At each iteration \( t \), the trail density is updated according to the following formula:

\[
\tau_{uv}(t + 1) = (1 - \rho)\tau_{uv}(t) + \Delta \tau_{uv}(t)
\]

\[
\Delta \tau_{uv}(t) = \sum_{k} \Delta \tau_{uv}^{k}(t)
\]

where \( \rho \) is a coefficient such that \((1-\rho)\) represents the evaporation of trail between \( t \) and \( t+n \). \( \Delta \tau_{uv}^{k}(t) \) is the quantity per unit of length of trail substance laid on path \((u,v)\) by the \( k \)th ant between time \( t \) and \( t+n \).

The trail density \( \Delta \tau_{uv}^{k}(t) \) is calculated by using the following equation:

\[
\Delta \tau_{uv}^{k}(t) = \begin{cases} 
Q \frac{1}{L_k} & \text{if the } k \text{th ant visits } (u,v) \\
0 & \text{otherwise} 
\end{cases}
\]

where \( Q \) is a constant, and \( L_k \) is the tour length of the \( k \)th ant.

In this study, distributed ants’ intelligence is used to solve the hard optimization problems of sitting facilities. The proposed method is devised according to the ACO algorithm for TSP. In TSP, the question is to find a closed tour of minimal length connecting \( N \) given cities. ACO is dependent on two terms, the trail density and the visibility (distance), to choose the optimal route.

Typical problem of site selection: such as p-median problem, coverage problem, p-center problem. P-median problem is investigated in our research. P-median: the objective is to find the best \( N \) locations that can produce the largest value of a utility function, generally minimize the total cost, such as transportation cost, transportation distance or transportation time.
The objective is to find the best $N$ locations (targets) that can produce the largest value of a utility function.  

The optimal site selection is fulfilled by using pheromone updating of ACO.  

In the initial stage, each cell will have an equal amount of pheromone. A certain amount of pheromone will be deposited on the cells visited by an ant.  

The combination of cells for sitting the facility is evaluated according to a utility function.  

It is expected that ants are likely to visit the selected cells of higher utility values according to the greedy criterion. As a result, more amount of pheromone will be deposited on these potential cells. This will in turn attract more ants to visit them.

According to Equation (1), the probability that a cell ($x$) can be selected for a visit by the $k$-th ant at time $t$ is estimated as follows:  

$$p^k(t) = \begin{cases} \frac{[r^t(x)]^{\alpha} [\eta^t(x)]^{\beta}}{\sum_{x \in \text{allowed}} [r^t(x)]^{\alpha} [\eta^t(x)]^{\beta}} & \text{if } x \in \text{allowed} \\ 0 & \text{otherwise} \end{cases}$$

Therefore, the total transportation cost for visiting the $N$ sites (targets) of the facility is given by the following equation:

$$L_{\text{trans}} = \sum_{x} d(x)p_{\text{dens}}(x)A$$

where $L_{\text{trans}}$ is the total transportation cost for the $k$-th ant, $d(x)$ is the Euclidian distance between cell $x$ and the closest target, and $A$ is the area of each cell.

Multi-criteria evaluation could be used to combine all these various spatial factors.  

In this study, the modified total transportation cost for sitting $N$ targets is calculated by the following equation:

$$L'_{\text{trans}} = \sum_{x} d(x)p_{\text{dens}}(x)A \sum_{d \in 2} w_{d} \sum_{x} e^{-\rho_{d,x}}$$
Pheromone updating is then defined as follows:

\[
\Delta \tau^*_i (t) = \begin{cases} 
Q \frac{1}{Q \cdot (x+1)} L_{\text{mean}} & \text{if } x \text{ falls within } 3 \times 5 \text{ window of a closest target at timer } t_i \\
0 & \text{otherwise}
\end{cases}
\]

Ant intelligence search

Optimal site selection according to the pheromone updating of ACO

Model validation and result analysis

The population density for the districts of Guangzhou

Road networks of the study area from the GIS database

Comparison of SS, GA and ACO

<table>
<thead>
<tr>
<th>Number of targets</th>
<th>SS</th>
<th>GA</th>
<th>ACO</th>
<th>(ACO-SS)/ SS</th>
<th>(ACO-GA)/ GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.7</td>
<td>3.9</td>
<td>3.9</td>
<td>5.3%</td>
<td>0.0%</td>
</tr>
<tr>
<td>4</td>
<td>5.0</td>
<td>5.2</td>
<td>5.3</td>
<td>5.5%</td>
<td>0.4%</td>
</tr>
<tr>
<td>6</td>
<td>6.0</td>
<td>6.2</td>
<td>6.3</td>
<td>2.7%</td>
<td>1.3%</td>
</tr>
<tr>
<td>8</td>
<td>6.9</td>
<td>7.1</td>
<td>7.2</td>
<td>3.9%</td>
<td>1.2%</td>
</tr>
<tr>
<td>10</td>
<td>7.7</td>
<td>8.0</td>
<td>8.2</td>
<td>2.8%</td>
<td>2.5%</td>
</tr>
</tbody>
</table>

Comparison of the computation time between the SS, GA and ACO methods

<table>
<thead>
<tr>
<th>Number of targets</th>
<th>SS</th>
<th>GA</th>
<th>ACO</th>
<th>(ACO-SS)/ SS</th>
<th>(ACO-GA)/ GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>32</td>
<td>45</td>
<td>4</td>
<td>12.5%</td>
<td>8.9%</td>
</tr>
<tr>
<td>4</td>
<td>95</td>
<td>138</td>
<td>2</td>
<td>29.5%</td>
<td>20.3%</td>
</tr>
<tr>
<td>6</td>
<td>198</td>
<td>247</td>
<td>84</td>
<td>20.7%</td>
<td>16.6%</td>
</tr>
<tr>
<td>8</td>
<td>333</td>
<td>419</td>
<td>15</td>
<td>16.2%</td>
<td>12.9%</td>
</tr>
<tr>
<td>10</td>
<td>525</td>
<td>832</td>
<td>3</td>
<td>13.9%</td>
<td>8.8%</td>
</tr>
</tbody>
</table>

The ACO method needs much less computation time than both the SS method and the GA method and find much better result.
Site selection using multi-scale method

A multi-scale ACO approach is used to reduce the computation time.

Comparison of the total utility value between the multi-scale ACO method and the single ACO method

<table>
<thead>
<tr>
<th>Number of targets</th>
<th>multi-scale ACO</th>
<th>single ACO</th>
<th>mult-scale/single ACO</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.409</td>
<td>0.409</td>
<td>1.0000</td>
</tr>
<tr>
<td>2</td>
<td>0.786</td>
<td>0.786</td>
<td>1.0000</td>
</tr>
<tr>
<td>3</td>
<td>1.176</td>
<td>1.176</td>
<td>1.0000</td>
</tr>
<tr>
<td>4</td>
<td>1.567</td>
<td>1.567</td>
<td>1.0000</td>
</tr>
<tr>
<td>5</td>
<td>1.958</td>
<td>1.967</td>
<td>1.0000</td>
</tr>
<tr>
<td>6</td>
<td>2.349</td>
<td>2.349</td>
<td>1.0000</td>
</tr>
<tr>
<td>7</td>
<td>2.740</td>
<td>2.740</td>
<td>1.0000</td>
</tr>
<tr>
<td>8</td>
<td>3.131</td>
<td>3.140</td>
<td>1.0000</td>
</tr>
<tr>
<td>9</td>
<td>3.522</td>
<td>3.532</td>
<td>1.0000</td>
</tr>
<tr>
<td>10</td>
<td>3.913</td>
<td>3.923</td>
<td>1.0000</td>
</tr>
<tr>
<td>11</td>
<td>4.304</td>
<td>4.314</td>
<td>1.0000</td>
</tr>
<tr>
<td>12</td>
<td>4.695</td>
<td>4.705</td>
<td>1.0000</td>
</tr>
<tr>
<td>13</td>
<td>5.086</td>
<td>5.096</td>
<td>1.0000</td>
</tr>
<tr>
<td>14</td>
<td>5.477</td>
<td>5.487</td>
<td>1.0000</td>
</tr>
<tr>
<td>15</td>
<td>5.868</td>
<td>5.878</td>
<td>1.0000</td>
</tr>
<tr>
<td>16</td>
<td>6.259</td>
<td>6.269</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

It is found that these two methods can obtain almost the same result in terms of the utility value and the multi-scale ACO is feasible.

Comparison of multi-scale to other methods

Comparison of the computation time for identifying 10 targets using the SS, GA and ACO methods for the study area

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Computing time (hour)</th>
<th>Total utility value (10^-6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SS</td>
<td>82.5</td>
<td>7.7</td>
</tr>
<tr>
<td>GA</td>
<td>5.2</td>
<td>8.0</td>
</tr>
<tr>
<td>single ACO</td>
<td>2.1</td>
<td>8.2</td>
</tr>
<tr>
<td>multi-scale ACO</td>
<td>0.5</td>
<td>8.2</td>
</tr>
</tbody>
</table>

The experiment indicates that the multi-scale ACO has 5.5% improvement of the total utility value over the SS method and has 2.5% improvement over the GA. It has many more advantages of reducing the computation time than the SS, genetic algorithms and the single ACO.

Comparison of overlaying results

Identifying 10 optimal sites using SS, GA, ACO and multi-scale ACO.

Overlay of the 10 simulation results of the GA method.
Comparison the overlaying result

Overlay of the 10 simulation results of the ACO method

<table>
<thead>
<tr>
<th>No.</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACO</td>
<td>8.20</td>
<td>8.09</td>
<td>8.19</td>
<td>8.11</td>
<td>8.18</td>
<td>8.19</td>
<td>8.18</td>
<td>8.19</td>
<td>8.20</td>
<td>8.20</td>
<td>8.17</td>
<td>0.039</td>
</tr>
<tr>
<td>GA</td>
<td>8.07</td>
<td>8.10</td>
<td>7.84</td>
<td>8.04</td>
<td>7.93</td>
<td>7.96</td>
<td>7.93</td>
<td>8.08</td>
<td>7.96</td>
<td>8.05</td>
<td>8.00</td>
<td>0.084</td>
</tr>
</tbody>
</table>

ACO can produce higher utility values and smaller standard deviation (SD) over GA.

Experiment result

Experiments in Guangzhou indicate that this multi-scale ACO method can produce similar results but use much less computation time, compared with the single ACO method. Good position accuracies can be maintained although this approach is based on an approximation method. This method has better performance than conventional methods, such as the SS method and the GA method, for solving site search problems.

ACO can produce higher utility values and smaller standard deviation (SD) over GA.

Spatial optimization problem and ant intelligence

1. Spatial optimization problem and ant intelligence
2. Ant intelligence for sites selection
3. Ant intelligence for path optimization
4. Ant intelligence for area object optimization
5. Conclusion and expectation

Path optimization problem

- Path-finding problems have attracted widespread research interests in many disciplines, such as robot path planning, emergency evacuation etc. The objective is to choose the best travel path according to the costs in terms of time, distance or safety. The path-covering optimization is usually much more complex than the point-covering optimization because it involves a huge solution space.
- Functions have been provided in most GIS to find the optimal paths between origins and destinations based on the least-cost. The Dijkstra’s algorithm, which has been widely used to solve the shortest-path problem. But it pays no attention to coverage problem. A virtual network has to be constructed for raster data.
- The methods of path-covering optimization in a raster space have not been well explored because of the complexities. Now we use ACO to solve such problem.

Path optimization model

Path optimization usually concerns two common objectives in planning practice:
1. The maximum service coverage to the population (benefit).
2. The minimum total travel distance (cost).
3. Incorporate to one object

| Improvement | a direction function | A utility function represent the multi-objects | Modification of tabu and Pheromone update |
How to construct a path

1. Just start from the original cell. Set it as the current cell.
2. Select one of the eight neighbor cell of the current cell. Set it as the next current cell. Repeat step 2 until the destination cell is found.
3. The destination cell is thought to be found if ant reach the certain area near it.

Define utility function

Utility function:

\[
F = \frac{\text{path}_{\text{area}}}{\text{path}_{\text{length}}} = \frac{\sum_{i} \text{area}_i}{\sum_{i} \text{length}_i}
\]

Where:
- \(\text{path}_{\text{area}}\): area whose background color is gray
- \(\text{path}_{\text{length}}\): the length of the line connecting the original cell to the destination cell

Select a neighbor cell

Eight candidate neighbor cell, its transition probability is calculated as follows:

\[
p^c_i(0) = \frac{[c_i(0)]^{[g(0,0)]^k}}{\sum_{i} [c_i(0)]^{[g(0,0)]^k}} \text{ if } v_i \in S^c\]

otherwise

Pheromone model

- The difference of pheromone is very small between these eight direct neighbors because of spatial autocorrelation.
- The pheromone at the cells of a further distance \(r\) should be used to determine the probability of walking direction correctly to enlarge the difference of pheromone in its eight neighbor cell.
- For example, the pheromone at \(\mathbf{v}\) is replaced by that at \(\mathbf{v}^*\) which has the largest amount of pheromone at that direction.

Direction function

- Two objectives: service coverage and travel distance.
- For satisfying the first objective, The ant will be attracted to go through cells of large local population density.
- For satisfying the second objective, the ant is encouraged to move towards its final destination.
- A direction function is devised to balance the trade-off between the two objectives as the right figure shown.

Tabu list update strategy

- A modified strategy for updating the tabu list
- More cells should be included in the tabu list besides the path itself for two reasons.
  - Firstly, if an ant just explores a neighborhood, the increase in the service coverage can't compensate the increase in the travel distance.
  - Secondly, unconstrained walking will cause the overlapping of service coverage and thus reduce the total served population.
Pheromone update strategy

- Additional strategies for pheromone updating
  - The pheromone at the cells near the destination is set to high values.
  - The best utility value will be recorded for each cell.
  - The pheromone will be updated according to the best utility for each cell.
  - A moving window of 3*3 is used to rank the best utility values in the neighborhood. Only those cells with the first top three utility values will update the pheromone.

Suppose \( F_5 > F_4 > F_1 > F_8 > F_7 > F_9 > F_2 > F_6 > F_3 \). We will update pheromone of the 5th, 4th, and the first cell.

Posterization procedures

A constructed path may consist of some redundant cells or unnecessary curves. Some posterization procedures are required to improve the path shape such as thinning and straightening procedure.

Model validation and result analysis

- Parameters for the model

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Ants</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \rho )</th>
<th>( R_{cov} )</th>
<th>( r )</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>20</td>
<td>2</td>
<td>1</td>
<td>0.9</td>
<td>10</td>
<td>3</td>
</tr>
</tbody>
</table>

Parameters used in this ACO path-covering model

Optimization result

- ant algorithm has a good convergence rate for finding the optimal path or near optimal one.

Utility convergence

Utility improvement with iterations by the proposed ACO:
- (a) average utility; (b) maximum utility.
Comparison to traditional ACO

Pheromone updating and optimal path finding

Application of ACO
- Pheromone updating and optimal path finding

Comparison of forward and backward path
- Forward path
- Backward path
- Indicate the model is very robust

Comparison to Dijkstra algorithm
- Dijkstra algorithm
- Comparison of the total utility values between the modified Dijkstra Methods and the proposed ACO method

Spatial optimization problem and ant intelligence

1. Spatial optimization problem and ant intelligence
2. Ant intelligence for sites selection
3. Ant intelligence for line optimization
4. Ant intelligence for area object optimization
5. Conclusion and expectation

Range of application
- Ecological protected area
- Natural protected area
- Farmland protected area
- Park, school
- Other area liked entity

Ant intelligence for sites selection

Ant intelligence for line optimization

Ant intelligence for area object optimization

Conclusion and expectation
Area of single block
- Park, school, etc. which occupy a certain area
  Define its boundary

Area of multi-block
- Farmland protected area
  Define the number of block
  Define the boundary of every block
  Not concern the number of block and pay more attention to the total area of all blocks
  Our concern: how to define the shape, size and location for every block under a pre-defined total area in raster data.

Area of single block
- Park, school, etc. which occupy a certain area
  Define its boundary

Area of multi-block
- Farmland protected area
  Define the number of block
  Define the boundary of every block
  Not concern the number of block and pay more attention to the total area of all blocks
  Our concern: how to define the shape, size and location for every block under a pre-defined total area in raster data.

Area object optimization

- Define the total area through defining the number of cells
- Assemble cells to define the boundary of every area block indirectly
- ACO are introduce for area optimization

Utility function
- The average total ecological suitability of all the selected cells
- The compactness of the pattern
- It is expected that the optimal protected pattern should yield the highest values for the average total ecological suitability and the compactness.
  The utility is defined as follows:

\[
U_{goal} = w_c \times S_c + w_e \times P_e
\]

transition probability

\[
p(t) = \frac{(r(t))}{\sum (r(t))} \quad \text{if } i \text{ is allowed}, \quad \eta_i = \frac{s_i}{\sum s_i} \times 10
\]

- \( s_i \) is the ecological suitability at cell i
- \( \sum s_i \) is the sum of the suitability for all the cells in the study area

Modify the pheromone updating manner
In order to make block more compact, a cell should have a higher probability to be selected if its neighbors have already been included in the protected area. A revision is made as follows:

\[
\Delta \tau_{i}^{c}(t) = \begin{cases} 
\eta_i \frac{U_{goal}^{c}(t)}{\Delta \tau_{i}^{c}(t-1)} & \text{if cell } i \text{ falls within a window } d \text{ of cells at time } t \\
0 & \text{otherwise}
\end{cases}
\]

\( d \) is the distance from the central cell i.

\( U_{goal} \) is the utility of the protected pattern

Accelerate the computing speed
Many cells generally over several thousand have to be selected when construct one solution, if we select one by one it is very time consuming. A multi-gamble wheel method have been adopted:

1. All the cells will be judged during first cycle, but not all of them will be selected.
2. The remaining will be judged during the second cycle
3. Repeat until there is no remaining
4. A solution is retained
Area object optimization model

- Accelerate the computing speed

Factor relating to land use suitability

- \( \text{NDVI} \)
- \( \text{MNDWI} \)
- \( \text{DEM}_{\text{amp}} \)
- \( \text{DEM}_{\text{disturb}} \)
- \( \text{DEM}_{\text{compact}} \)

Land use suitability and compactness

- Land use suitability:
  \[
  S_h = w_1 \text{NDVI}_p + w_2 \text{MNDWI}_p + w_3 (1 - \text{DEM}_{\text{amp}}) + w_4 (1 - \text{DEM}_{\text{disturb}}) + w_5 \text{DEM}_{\text{compact}} + w_6 \text{NDVI}_{\text{std}}
  \]
- Land use compactness:
  \[
  P_c = \frac{\sqrt{A}}{L}
  \]

where \( A \) is the total area of the protection, \( L \) is the perimeter of a protected scenario.

Utility function:

\[
U_{\text{result}} = w_x \times S_h + w_y \times P_c
\]

Model validation and experiment result

- Guangzhou-2008

<table>
<thead>
<tr>
<th>Iteration</th>
<th>( \alpha )</th>
<th>( \beta )</th>
<th>( \rho )</th>
<th>( Q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1000</td>
<td>5</td>
<td>1</td>
<td>0.01</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Sensitivity analysis of weight parameter

- \( W_x = 0.7, W_y = 0.3 \) the experiment result is better
Model validation and experiment result

Validation - Hypothetical Data

Contrast Experiment

Model validation and experiment result

ACO, SA, IR, DS

ACO has the improvement of the compactness over SA, IR, and DS by 4.81%, 7.59%, and 21.23%, respectively. It also has the improvement of the total utility over SA, IR, and DS by 1.33%, 3.21%, and 6.71%, respectively. Therefore, the proposed modified ACO seems to be very attractive for area optimization.

Conclusion and expectation

Spatial decision for environment and resource relates to complex nonlinear relationship and high dimension, large dataset. A optimal solution is difficult to obtain using traditional brute-force method in a limited time. ACO which just depends on the utility function can be used to solve non-linear and high-dimensional problem. ACO as a heuristic algorithm does not need to search all the solution space so it is fast and suitable for optimal problem of large dataset.

The integration of ACO with geographic information systems is proposed to solve site selection problems. A number of modifications have also been introduced so that ACO can be applied for optimal problem of large dataset. The proposed model has better performance than the single search and GA.

This article has developed a new method for solving complicated zoning problems by using ACO techniques. Significant modifications have been made, so that traditional ACO can be extended to the solution of area optimization problems.

The result of urban expansion simulation would be used as an important factor for defining ecological protected area, transportation line planning, land use planning etc.

coupling urban expansion simulation and urban planning

www.geosimulation.cn
Thanks
Any question?