Knowledge discovery for geographical cellular automata

LI Xia[1] & Anthony Gar-On Yeh[2]

1. School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China;
2. Centre of Urban Planning and Environmental Management, The University of Hong Kong, Hong Kong SAR, China

Correspondence should be addressed to Li Xia (email: gplx@zsu.edu.cn or lixia@graduate.hku.hk)

Received July 1, 2003; revised November 11, 2003

Abstract This paper proposes a new method for geographical simulation by applying data mining techniques to cellular automata. CA has strong capabilities in simulating complex systems. The core of CA is how to define transition rules. There are no good methods for defining these transition rules. They are usually defined by using heuristic methods and thus subject to uncertainties. Mathematical equations are used to represent transition rules implicitly and have limitations in capturing complex relationships. This paper demonstrates that the explicit transition rules of CA can be automatically reconstructed through the rule induction procedure of data mining. The proposed method can reduce the influences of individual knowledge and preferences in defining transition rules and generate more reliable simulation results. It can efficiently discover knowledge from a vast volume of spatial data.

Keywords: knowledge discovery, cellular automata, geographical simulation, geographical information systems.

DOI: 10.1360/01yd0368

Recently, cellular automata (CA) has been increasingly applied to the simulation of geographical phenomena, especially urban simulation[1,2]. The researches of using CA have been carried out in China with many publications nationally and internationally[3—7]. CA can be applied to the simulation of many geographical phenomena, such as diffusion of wildfire[8], population fluctuation of animals[9], evolution of urban systems and land use[1,2], the formation of idealized urban forms[3,6], planning for sustainable land use[10], and automatic generation of agricultural protection zones[11].

Many geographical phenomena have manifested the features of complex systems, which cannot be represented and simulated by using mathematical equations. Studies have demonstrated that CA are useful tools for simulating complex systems[12]. The simulation of urban systems is one of these successful examples of CA applications[1,2,6]. Some of the early urban CA studies were done by Couclelis[13]. Her studies indicate that the CA simulation can be used as the analog to realistic urban systems. Batty and his colleagues have also done some interesting CA studies on simulating urban systems[1]. Their early experiments were based on the diffusion limited aggregation (DLA) techniques for simulating the expansion of built-up areas. However, they began to use CA techniques for the simulation of urban development later.

Transition rules are the core of CA, but the determination of these rules is very tedious. Heuristic methods have been used to define transition rules[14], including using matrix[2], multicriteria evaluation[15], and grey state[6]. These methods are subject to a lot of uncertainties and have various forms. Moreover, they are implicit because of using mathematical equations. The definition of parameter values is very difficult. We used to propose the method of using neural networks to retrieve these parameter values automatically[7,16]. However, the neural networks are black-box approaches. Users have problems in comprehending the meanings of these parameter values and the mechanisms. These methods have limitations in applications.

In this study, a new method based on knowledge discovery or machine learning is proposed to reconstruct the transition rules of geographical CA. Data mining has been applied to the classification of remote sensing data for improving the performance. It has also been applied to the knowledge discovery in GIS databases, such as classification of soil types[17]. CA usually involve a large amount of spatial data in simulating complex geographical phenomena. The use of data mining can significantly enhance the simulation capability. The derived transition rules are explicit without using mathematical equations. However, no such studies have been reported so far for CA simulation by using this technique.

1 Data mining and geographical cellular automata

Data mining is to discover knowledge automatically from databases. It has been proposed to solve the problems of the difficulties and uncertainties in knowledge solicitation. This is usually done through machine learning. The most common machine learning algorithms include: ID3, C4.5, CART, IB1, IB2, MPIL1, and MPIL2. C4.5 developed by Quinlan is the most popular algorithm for data mining[18]. See5 for Window and its Unix counterpart C5.0 are the most updated version of C4.5.

The series of C4.5 use the ‘information gain ratio’ to determine the splits at each internal node of the decision tree[18]. First, imagine selecting one case at random from a training data set S and announcing that it belongs to some class C_j. The information from such a message (entropy) is calculated by

$$\mathrm{info}(S) = -\sum_{j=1}^{k} \frac{\mathrm{freq}(C_j, S)}{|S|} \times \log_2 \frac{\mathrm{freq}(C_j, S)}{|S|}, \qquad (1)$$

where $\mathrm{freq}(C_j, S)$ is the number of cases in S belonging to class C_i, and $|S|$ is the total number of observations in S.

Consider that S has been partitioned into n outcomes for a test X. The expected information is

$$\mathrm{info}_x(S) = \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times \mathrm{info}(S_i). \qquad (2)$$

The information gained by splitting S using X equals

$$\mathrm{gain}(X) = \mathrm{info}(S) - \mathrm{info}_x(S). \qquad (3)$$

The bias inherent in the gain criterion with a large number of splits should be corrected by normalizing gain(X) using split info(X)

$$\text{split info}(X) = -\sum_{i=1}^{n} \frac{|S_i|}{|S|} \times \log_2 \left(\frac{|S_i|}{|S|} \right). \qquad (4)$$

Then

$$\text{gain ratio}(X) = \text{gain }(X) / \text{split info}(X). \qquad (5)$$

S will be recursively split to ensure that the gain ratio is maximized at each node of the tree. This procedure continues until each leaf node contains only observations from a single class or no gain in information is yielded by further splitting. The above procedure automatically creates decision trees or rule sets based on the criterion of ‘information gain ratio’. The rule induction procedure is convenient and robust.

The above techniques can be used to discover geographical knowledge from GIS databases, such as the spatial distribution patterns. The integration of data mining and CA can automatically generate transition rules from observation data, and calibrate the models simultaneously. Transition rules determine the conversion of state for each cell, such as the conversion from agricultural land to urban land. Many existing urban CA do not provide concrete transition rules, but use mathematical equations to estimate conversion probability. They may use linear or logistic equations to represent the relationship between land use conversion and spatial variables. These equations are not straightforward for decision makers to use. Actually, decision makers are more familiar with the type of explicit rules. For example, it is much easier for them to adopt some actions according to the following explicit rules:

Rule 1:

If Land use types = Forest or Wetland

Then No development is allowed (confidence = 0.85)

Rule 2:

If Land use types = Cropland

Distance to urban centres < 10 km

The number of developed cells in the neighborhood >16

Then Development is allowed (confidence = 0.95)

This paper attempts to discover transition rules of CA from remote sensing and GIS using urban simulation as examples (Fig. 1). Remote sensing images in two different years are used as the observation data for discovering transition rules. It may be ideal if the observation interval (ΔT) is equal to or close to the iteration interval (Δt) so that the mined transition rules can be directly used in urban simulation. However, the observation interval of remote sensing data is usually year-based while the iteration interval of CA is much smaller. It is impractical to collect data within the iteration interval of Δt. Moreover, the observation data cannot comprehend the long-term trend if the observation interval is too short.

Some adjustments are required when the extracted rules from the observation data are applied to each iteration of urban simulation. First, the relationship between the number of iterations (K), the iteration in-

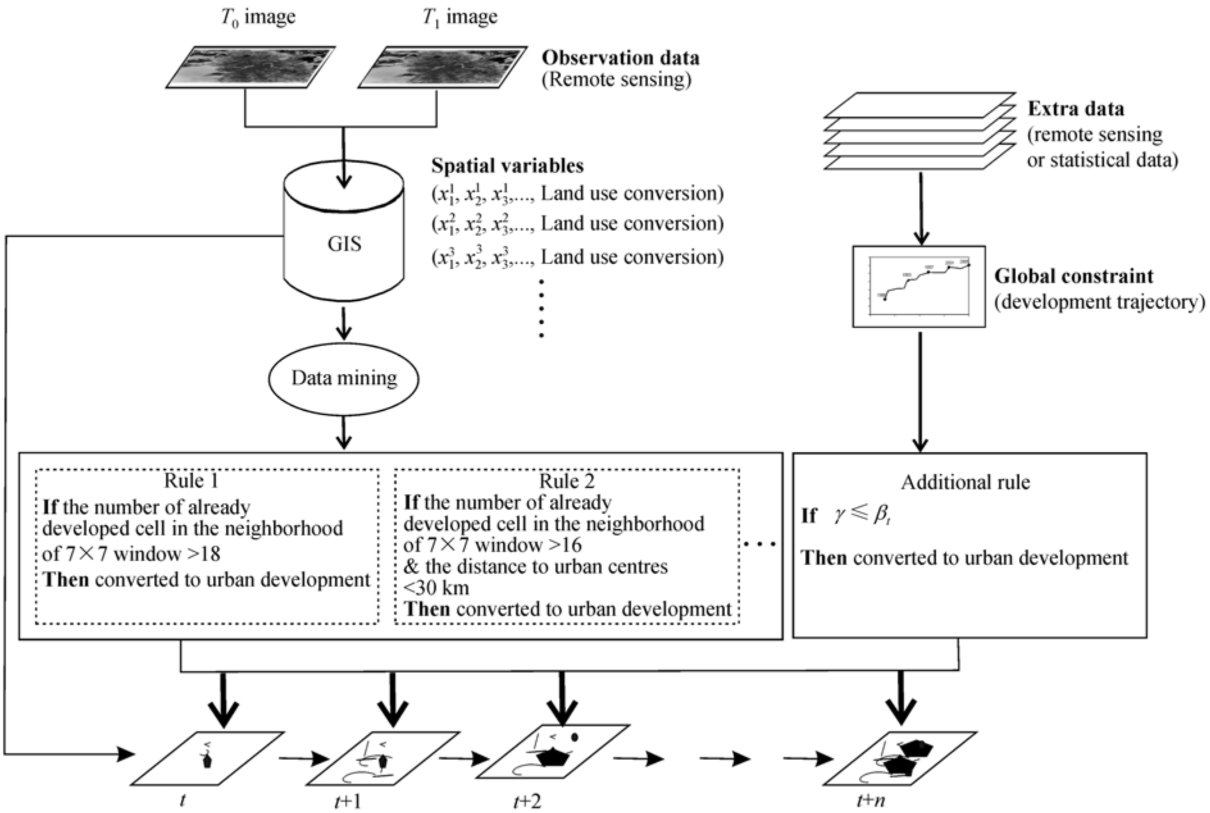
T_0 image
T_1 image
Observation data
(Remote sensing)
Extra data
(remote sensing
or statistical data)
Spatial variables
$(x_1^1, x_2^1, x_3^1$,..., Land use conversion)
$(x_1^2, x_2^2, x_3^2$,..., Land use conversion)
$(x_1^3, x_2^3, x_3^3$,..., Land use conversion)
GIS
Global constraint
(development trajectory)
Data mining
Rule 1
If the number of already developed cell in the neighborhood of 7×7 window >18
Then converted to urban development
Rule 2
If the number of already developed cell in the neighborhood of 7×7 window >16
& the distance to urban centres <30 km
Then converted to urban development
Additional rule
If $\gamma \leqslant \beta_t$
Then converted to urban development
t
t+1
t+2
t+n

Fig. 1. Data mining for automatically reconstructing the transition rules of geographical CA.

terval (Δt) and the observation interval (ΔT) is as follows:

$$K = \Delta T / \Delta t, \tag{6}$$

where ΔT is the observation interval by using two years of remote sensing data, and Δt is the iteration interval between t and t+1, K is the number of iterations.

The amount of land use conversion ΔQ_0 can be determined from remote sensing for the larger interval of ΔT. Only a portion of land use conversion took place in the iteration interval Δt. The proportion of land use conversion between t and $t+1$ can be obtained by using the following equation:

$$\Delta q_0 = \Delta Q_0 / K, \tag{7}$$

where Δq_0 is the amount of land use conversion for the iteration interval.

Therefore, an additional rule is used to determine land use conversion besides the use of the original derived transition rules. This additional rule is as follows:

If $x(i, j)$ should be converted according to the original transition rules

and $x(i, j)$ have not developed at t–1

and $\gamma \leq \beta_0$.

Then $x(i, j)$ will be developed at t,

where $x(i, j)$ is the cell x at location (i, j); and

$$\beta_0 = \frac{\Delta q_0}{\Delta Q_0} = \frac{1}{K}. \tag{8}$$

The above rule assumes that the urban growth rate is constant. It is not true because the growth rate is dynamic. More than two dates of remote sensing data can be used to capture this fluctuation of urban growth. Since the amount of land use conversion (ΔQ_t) is changing with time, the above additional transition rule should have the following generic form:

If $x(i, j)$ should be converted according to the original transition rules

and $x(i, j)$ have not developed at t–1

and $\gamma \leq \beta_t$.

Then $x(i, j)$ will be developed at t,

where $$\beta_t = \beta_0 \times \frac{\Delta Q_t}{\Delta Q_0} = \frac{1}{K} \times \frac{\Delta Q_t}{\Delta Q_0}. \tag{9}$$

2 The implementation and simulation results

2.1 Test area and spatial data

The proposed model has been tested in the same study area where our previous CA models were applied. The selection of the same study area can allow the comparison of the effects of various CA models. We have used neural networks to simplify the procedure of defining transition rules and facilitate the calibration of CA[16]. However, the transition rules of this model are not transparent because of the back-box approach of neural networks. The extraction of explicit transition rules from data mining is very important for understanding the mechanisms of complex urban systems. This model has more advantages than our previous models because of using data mining techniques. There are no studies on deriving transition rules from observation data by using data mining techniques so far.

The first step of the method is to prepare the training data for the rule discovery. A series of spatial data, which are from remote sensing and GIS, are used for the data mining. The data include the layers of urban development, proximity variables, neighborhood conditions, and physical attributes (Table 1). Satellite TM images in four years, namely 10 December 1988, 24 December 1993, 29 August 1997, and 20 November 2001, are used to calibrate the CA model. The observation data mainly include the 1988 and 1993 satellite TM images for deriving the transition rules. The 1997 and 2001 images are used for capturing the urban development trend for simulating future urban development. The development trend can even be obtained from statistical yearbooks when there are no enough satellite images.

Table 1 Spatial variables used for data mining

Spatial variables	Acquisition methods	Value ranges
1. Target variable	Classification of satellite TM images	1-converted to urban areas;
Urban development in 1988—1993		0-non-converted
2. Proximity variables		
Distance to the city proper (PropD)	*Eucdistance* of ARC/INFO GRID	0—60 km
Distance to town centres (TownD)		0—30 km
Distance to roads (RoadD)		0—20 km
Distance to expressways (ExprD)		0—60 km
Distance to railways (RailD)		0—60 km
3. Neighborhood function		
Number of developed cells in the 7×7 neighborhood (Nsum)	*Focalsum* of ARC/INFO GRID	0—49
4. Physical attributes of a site		1-crop;
Land use types (Land)	Classification of satellite TM images	2-bared soil;
		3-construction sites;
		4-orchard;
		5-built-up areas;
		6-forest;
		7-water
Agricultural suitability (Agsu)	Land evaluation of GIS	0—1
Slope (Slope)	DEM of GIS	1—90º

2.2 Data mining for deriving transition rules

These spatial variables have huge sets of data volume. It is inefficient to process the entire set of spatial data for data mining. Even though See5 is relatively fast, a much longer time is needed for building decision trees, especially when options such as boosting are employed. Secondly, it is undesirable to use a whole set of data for mining because of spatial autocorrelation. Bias will be introduced to analysis results if the training data have severe correlation. Using a smaller set of training cases may be at the cost of possible reduction in the classifier's predictive performance. The training data were divided into two groups: one for deriving rules, and another for examining predictive accuracy. Fig. 2 clearly shows the relationships between the increase of sampling points and the prediction error. The prediction error is 35.2% by using 1% of the training data, and becomes 25.0% by using 10% of the training data. The improvement rates are insignificant after the first 10% of the data. Therefore, this study only used the sample of 20% to derive the transition rules.

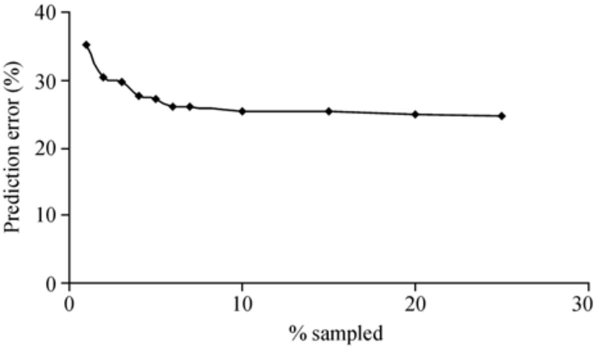

Fig. 2. Sampling rate and the prediction error.

A very large and complex decision tree is often produced because the tree may overfit training data. If the training data contain errors, overfitting the tree to the data can result in poor performance. The original tree must be pruned to minimize such a problem. The default value of pruning rate (25%) from See5.0 was used to simplify the decision tree.

The 1988 and 1993 satellite TM images were used to derive the transition rules. These derived transition rules were then used to simulate the urban dynamics in 1988—2005. Fig. 3 shows the development trajectory of the study area in 1988, 1993, 1997 and 2001 according to the classification of satellite images. Urban expansion was astonishing in 1988—1993. The rate of urban expansion became less in the later periods as a result of government intervention. If the projection of urban development is only based on the 1988 and 1993 observation data without using information on development trend, the simulated urban areas will be much larger than the actual ones for the later periods.

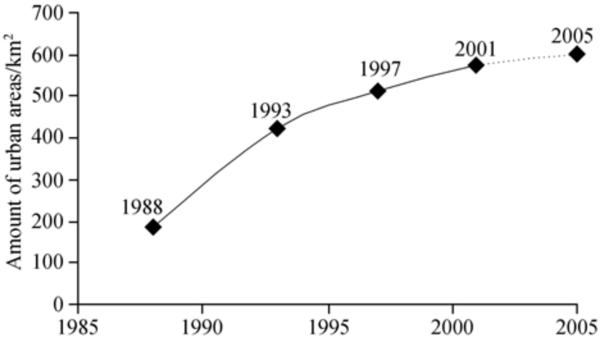

Fig. 3. Monitoring of the development trajectory of Dongguan using the 1988, 1993, 1997 and 2001 TM images.

There are many iterations of simulation before the final outcome is obtained. A shorter interval between *t* and *t*+1 means that a larger number of iterations are required. Although there is no consensus on what exact number of iterations should be used, 100—200 of iterations are quite normal for producing realistic simulation. The subtle patterns cannot be produced if there are too few iterations. It is because local interactions only take place at each iteration of urban simulation.

The transition rules from data mining were used to simulate the urban dynamics of Dongguan in 1988—1993, 1993—1997, 1997—2001 and 2001—2005. A total of 200 iterations were used for the simulation of urban growth in each period. The amount of urban growth (ΔQ_t) for each period was obtained based on the change detection of remote sensing. The global constraint factor (β_t) was calculated according to eq. (9). Table 2 lists the parameter values that were used in the simulation.

Table 2 Iterations, intervals, and the amount of urban growth for each period

	1988—1993	1993—1997	1997—2001	2001—2005
K (iterations)	200	200	200	200
ΔT (year)	5	4	4	4
Δt (year)	1/40	1/50	1/50	1/50
ΔQ_t (km^2)	233.3	90.6	62.9	25.0
Δq_t (km^2)	1.167	0.453	0.315	0.125
β_t	0.0050	0.0019	0.0013	0.0005

The See5 system was used to discover knowledge from GIS and remote sensing data. The rule sets were obtained from the data mining procedure. The following examples are some of the derived rule set:

```
Rule 1
    If    PropD < 30
          RoadD <= 5
          Nsum > 18
          Agsu < 0.8
          Land = 1
    Then  Converted to urban development
[0.92]
```

```
Rule 2
    If    PropD <= 25
          TownD > 7
          Nsum >= 12
          Agsu <= 0.5
          Land = 4
          Slope <= 6º
    Then  Converted to urban development
[0.86]
```

These transition rules are much clearer and simpler than mathematical equations. They can better reflect the mechanism of urban development. Each applicable rule votes for its predicted class with a voting weight equal to its confidence value. The confidence value is also automatically obtained by See5 during the data mining process. The votes are totted up, and the class with the highest total vote is chosen as the final prediction.

Because of the discrepancy between the observation

Rule 3

If PropD <= 48
TownD > 13
RoadD > 1
RoadD <= 5
Nsum >= 9
Agsu > 0.2
Agsu <= 0.4
Then Converted to urban development [0.90]

⋮

interval and the iteration interval, β_t is calculated and the following additional rule is also jointly used to decide the final land use conversion at each iteration from *t* to *t*+1:

Additional rule

$$\textbf{If}\quad \gamma \le \begin{cases} 0.0050 & (\text{in } 1988-1993) \\ 0.0019 & (\text{in } 1993-1997) \\ 0.0013 & (\text{in } 1997-2001) \\ 0.0005 & (\text{in } 2001-2005) \end{cases}$$

Then Converted to urban development

2.3 Simulation results and verification

The model simulates the urban growth of the study area in the period of 1988—1993, 1993—1997, and 1997—2001. The initial stage is based on the 1988 actual urban areas detected from remote sensing. Fig. 4(a) is the simulated urban development in 1988—1993, 1993—1997, and 1997—2001. Fig. 4(b) is the actual urban development in the same periods from the classification of remote sensing. Fig. 5 is the prediction of urban development in 2001—2005 based on the transition rules. The region witnessed fast urban expansion in the early 90s, but the rate of urban expansion was much less in the later periods because of development control. No compact development patterns were formulated because land development usually took place along roads. These dispersed development patterns can significantly increase energy consumption and cause wasteful use of land resources. The simulation of urban development can help to analyze and forecast the impacts of land use policy on land use changes. The CA simulation can be an effective tool for urban planners.

It is unrealistic to reproduce the exact patterns of a natural phenomenon because of the complexity of nature and the limitations of modelling. However, the assessment of goodness-of-fit is usually required to give a general indication on how the simulation is similar to the observation. The assessment can be based on cell-by-cell comparison or aggregated comparison. The first method is very simple just by overlaying the simulated and actual patterns. The second method emphases the aggregated patterns for the assessment instead of just using the cell-by-cell comparison. The spatial patterns of a geographical phenomenon usually involve many features, such as connectivity and fractal dimensions.

In this study, the actual urban areas in 1993, 1997 and 2001 were obtained from the classification of satellite TM images. The simulated urban areas were compared with the actual urban areas using overlay analysis. Table 3 lists the overall accuracy obtained from the cross-tabulation of the overlay analysis. The overall accuracy is 82.0% for simulating the urban growth in 1988—1993. It becomes 74.8% and 72.4% for simulating the urban growth in 1993—1997 and 1997—2001 respectively. It can be seen that although the transition rules were derived from the 1988—1993 time period, the urban development process captured by the transition rules has not changed much and quite a high simulation accuracy is obtained in the simulation of urban development in 1993—1997 and 1997—2001. These figures are quite acceptable for urban simulation.

Table 3 The overall accuracy of simulation compared with the actual urban development from satellite images in 1993, 1997 and 2001

Year	1993	1997	2001
Correct (%)	82.0	74.8	72.4

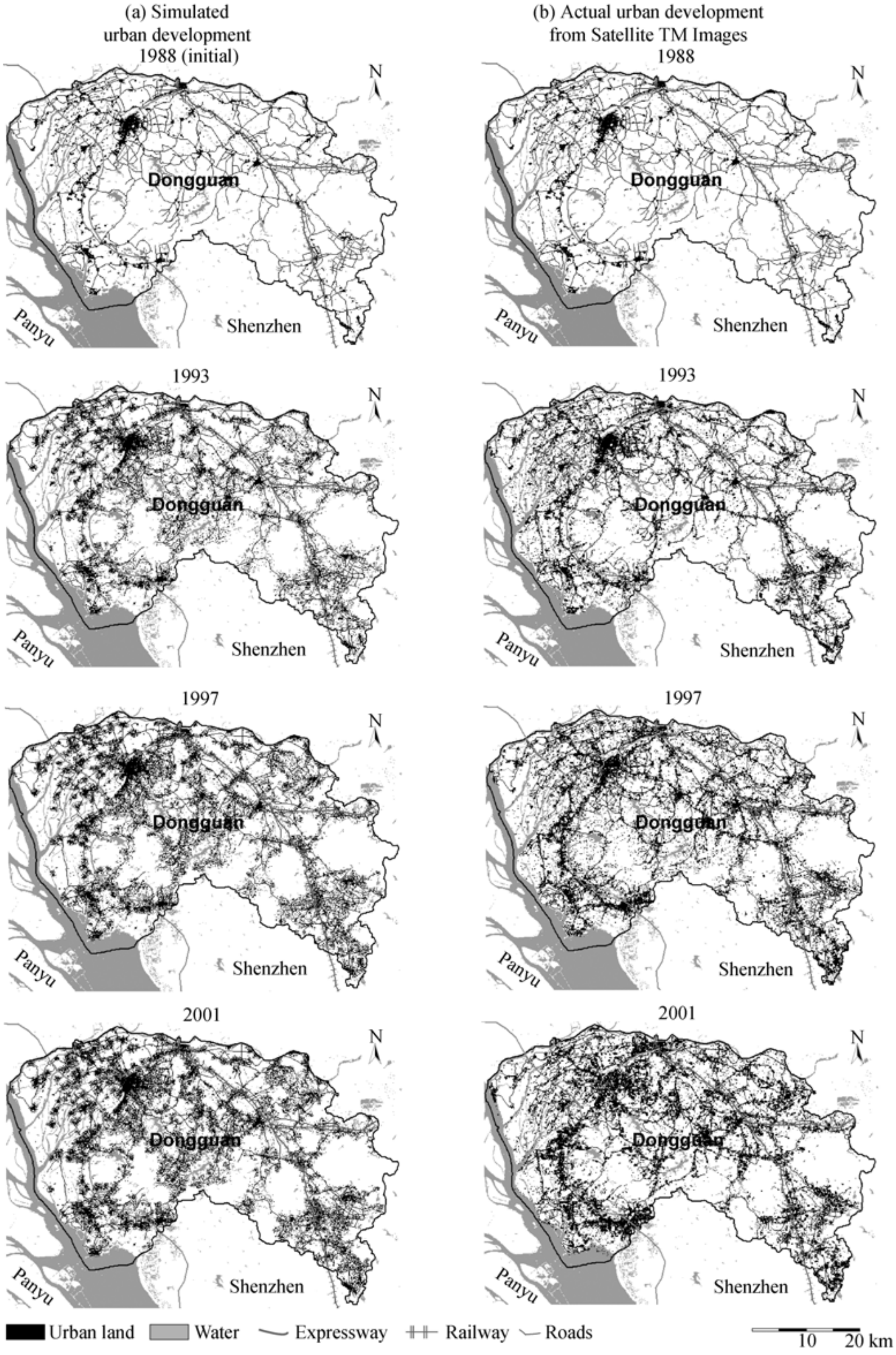

Fig. 4. The simulated and actual urban development of Dongguan in 1988, 1993, 1997 and 2001.

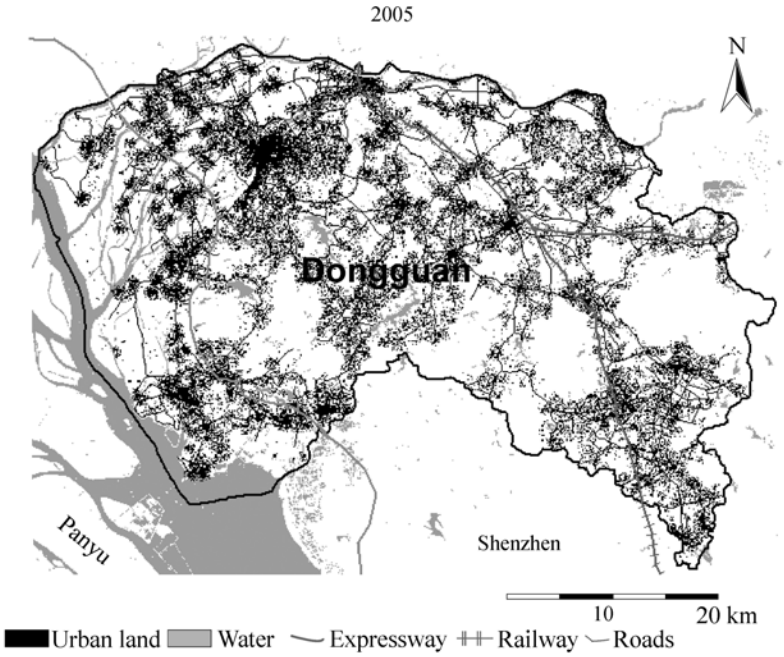

Fig. 5. Prediction of future urban development in 2005 based on the development trend.

This study also compares the simulated patterns with the actual patterns. Numerous indicators have been proposed for describing spatial patterns. However, there are no agreements on which indicator is most suitable for representing spatial patterns. A visual comparison may sometimes provide more meaningful results for calibrating CA models[19]. The visual comparison of the actual and simulated urban development indicates that the model is able to generate plausible simulation results (Fig. 4). In this study, the indicator of Moron I was chosen for the assessment of the aggregated patterns for reducing the uncertainties. It is quite easy to calculate the Moron I values in the GIS package, ARC/INFO GRID. Moran I is a useful spatial indicator that can reveal the degree of spatial autocorrelation. The indicator is able to estimate how close the simulated land use pattern is to the observed one[14].

Table 4 Comparison of Moran I between the actual and simulated urban development in 1993, 1997 and 2001

	1993	1997	2001
Actual urban development	0.44	0.66	0.76
Simulated urban development	0.42	0.58	0.71

The maximum value is one which indicates the absolute concentration of land use. A smaller value, which can be below zero, indicates a more even distribution of land use. Table 4 shows the comparison of the values of Moran I between the actual and simulated urban development in 1993, 1997 and 2001 respectively. The values of Moran I indicate the good conformity between the actual and simulated urban development. The analysis is consistent with the visual comparison. Urban development sites in the earlier stage (1993) are relatively isolated because of the prevailing urban sprawls. Urban developments tend to be more connected in the later years as they continue to grow.

The overall accuracy and Moran I were also calculated for the neural-network-based CA model for the comparison. The overall accuracy is 0.79 and Moran I is 0.40 for the previous model. This indicates that the proposed model has improvements in terms of accuracy. It is because the explicit transition rules are more easily adapted to complex relationships than mathematical equations. The most advantage of this method is that the explicit transition rules can be discovered from observation data automatically without using mathematical equations. This can provide flexibility in the modeling process.

3 Conclusion

Data mining has been applied to geographical researches. It can help to discover the rules about spatial distribution in geography. CA is a useful tool in simulating geographical phenomena. It is essential to define the transition rule in CA simulation. Reliable simulation results cannot be achieved if the transition rules are not defined in a systematic and consistent way. Howsoever, the definition of transition rules is subject to a lot of uncertainties. Transition rules are usually implicitly represented by mathematical equations in general CA models. There are limitations by using these mathematical equations to represent complex natural phenomena.

It is the first time that the explicit transition rules of CA are directly deduced from machine learning. Much improvement has been made by using this method. The explicit transition rules can be instantly derived from a vast volume of geographical data by using data mining techniques. No mathematical equations are required for representing transition rules. Calibration is automatically carried out during rule induction from training data. General CA involves a lot of variables and the determination of the parameters for these variables is very difficult through model calibration.

The model is applied to the Pearl River Delta region by using various years of satellite images as the observation data. The transition rules are discovered based on the proposed method. The transition rules are used to simulate the urban growth of the study area in 1988—2001, and also forecast the urban development in 2005. The validity of the model has been assessed based on the cell-by-cell comparison and quantitative indicator of Moran I. The assessment indicates good conformity between the actual and simulated urban development. The study shows that this method has more advantages than previous methods.

Acknowledgements This study was supported by the National Natural Science Foundation of China (Grant No. 40471105), and also by the "985 Project" of GIS and Remote Sensing for Geosciences from the Ministry of Education of China.

References

1. Batty, M., Xie, Y., From cells to cities, Environment and Planning B: Planning and Design, 1994, 21: 531—548.
2. White, R., Engelen, G., Cellular automata and fractal urban form: a cellular modelling approach to the evolution of urban land-use patterns, Environment and Planning A, 1993, 25: 1175—1199.
3. Li, Xia, Yeh, A. G. O., Constrained cellular automata for modelling sustainable urban forms, Acta Geographica Sinica (in Chinese), 1999, 54(4): 289—298.
4. Zhou Chenghu, Sun Zhanli, Xie Yichun, Geo-cellular Automata (in Chinese), Beijing: Science Press, 1999, 1—163.
5. Li Xia, Yeh, A. G. O., Integration of principal components analysis and cellular automata for spatial decision making and urban simulation, Science in China, Ser. D, 2002, 45(6): 521—529.[Abstract] [PDF]
6. Li Xia, Yeh, A. G. O., Modelling sustainable urban development by the integration of constrained cellular automata and GIS, International Journal of Geographical Information Science, 2000, 14(2): 131—152.[DOI]
7. Li Xia, Yeh, A. G. O., Neural-network-based cellular automata for simulating multiple land use changes using GIS, International Journal of Geographical Information Science, 2002, 16(4): 323—343. [DOI]
8. Clarke, K. C., Brass, J. A., Riggan, P. J., A cellular automata model of wildfire propagation and extinction, Photogrammetric Engineering & Remote Sensing, 1994, 60: 1355—1367.
9. Couclelis, H., Of mice and men: what rodent populations can teach us about complex spatial dynamics, Environment and Planning A, 1988, 20: 99—109.
10. Yeh, A. G. O., Li Xia, A constrained CA model for the simulation and planning of sustainable urban forms by using GIS, Environment and Planning B: Planning and Design, 2001, 28: 733—753. [DOI]
11. Li Xia, Yeh, A. G. O., Zoning for agricultural land protection by the integration of remote sensing, GIS and cellular automata, Photogrammetric Engineering & Remote Sensing, 2001, 67(4): 471—477.
12. Wolfram, S., Cellular automata: a model of complexity, Nature, 1984, 31: 419—424. [DOI]
13. Couclelis, H., From cellular automata to urban models: new principles for model development and implementation, Environment and Planning B: Planning and Design, 1997, 24: 165—174.
14. Wu, F., Calibration of stochastic cellular automata: the application to rural-urban land conversions, International Journal of Geographical Information Science, 2002, 16(8): 795—818. [DOI]
15. Wu, F., Webster, C. J., Simulation of land development through the integration of cellular automata and multicriteria evaluation, Environment and Planning B, 1998, 25: 103—126.
16. Li Xia, Yeh, A. G. O., Neural-network-based cellular automata for realistic and idealized urban simulation, Acta Geographica Sinica (in Chinese), 2002, 57(2): 159—166.
17. Moran, C. J., Bui, E. N., Spatial data mining for enhanced soil map modeling, International Journal of Geographical Information Science, 2002, 16(6): 533—549. [DOI]
18. Quinlan, J. R., C4.5: Programs for Machine Learning, San Mateo: Morgan Kaufmann, 1993, 302.
19. Clarke, K. C., Hoppen, S., Gaydos, L., A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area, Environment and Planning B: Planning and Design, 1997, 24: 247—261.