Exploring on Urban Land Development Intensity based on Artificial Neural Network Methods

Minghao Liu1, 2
1. School of Computer Science, Chongqing University of Posts and Telecoms, P.R. Chongqing, 400065, China
2. Spatial Information System Research Center, Chongqing University of Posts and Telecommunications, Chongqing 400065, P. R. China
Email: liumh@cqupt.edu.cn
Yuan Tao, Donghong Li, Baobao Xia, and Yaoxing Wang
School of Computer Science, Chongqing University of Posts and Telecoms, P.R. Chongqing, 400065, China
Email: {179128381, 823618636, 1036542372}@qq.com

Abstract—A modest development intensity for urban land benefits both ecological protection and living spaces improvement. It is an important indicator for urban land development intensity (ULDI) to measure the city livable and sustainable development. As an example of Chongqing Metropolitan Area, firstly, spatial database about land development intensity and its driving factors was established in sample regions, and BP artificial neural network methods were used to construct the land development intensity simulation model based on data driven in urban area with the help of MATLAB7 software. Secondly, two different scheme and algorithm were adopted to simulate land development intensity. Artificial neural network methods were detected by comparing the difference between real development intensity and the simulation results. Lastly, the land development intensity in Chongqing Metropolitan Area (9 districts) was simulated. Meanwhile, the results were compared by using the methods of neural network forecasting model and the multiple linear regression model with a wider range. The results shows that: (1) BP artificial neural network method is a good way to simulate the ULDI; (2) it is important to choose the reasonable driving factors and training algorithm; (3) the research scale has a certain impact on the results. Although the BP artificial neural network method can not explicitly explain the relationship between land development intensity and its driving factors in urban area, when data is sufficient, it is better to evaluate the ULDI than the method of regression analysis.

Index Terms—GIS, Urban intensity of land development, building density, artificial neural network, MATLAB

I. INTRODUCTION

With the development of geographic information system (GIS), collection of spatial data becomes easier. But there are still many restrictions to reflect spatio-temporal dynamic changes of geographical phenomenon, and to forecast reasoning on spatial data. The spatio-temporal data mining technology promote the development of temporal and spatial information technology. It has been becoming the hot spot of research to integrate the models that combine time with space and to develop related industries based on models[1].

At present, some models about spatio-temporal forecasting methods which have been used include single variable extrapolation method, multivariate regression method, the data adaptive method, the method of system dynamics, cellular automata (CA) method, agent-based modeling (ABM) and the evolutionary tree prediction model[2] [3]. Data adaptive method form the model structure by data-driven, which includes neural network method, genetic programming method, and so on.

Urban land development intensity (ULDI) is an important index to evaluate city livability and sustainability.

The existing models are usually constructed by single index such as economic [4], traffic[5], environment, landscape, culture etc. [6] [7]. These methods are successful for the local environment, but it is failing in the comprehensive prediction in the macro level. At present, comprehensive analysis models of ULDI are just simply weighted analysis from single factor analysis models, which has a strong subjectivity.

At present, whether single factor analysis models or comprehensive analysis models can not avoid the spatial neighborhood phenomenon. Tobler's First Law of Geography gives neighborhood phenomenon (spatial correlation and spatial heterogeneity) a scientific explanation, which states that everything is spatial correlation, and the closer in spatial, the more related each other in attribute values. Recently, neighborhood problems are often expressed by Moran’s and Geary’ C in spatial econometric models [8] or autoregressive statistics models [9]. Sometimes, attribute similarity relationship [10] or experience analogy [11] is also used.

Theoretical assumptions based on experience or similar in attributes is that land development intensity should be consistent when the location is similar. In recent years, the artificial neural network is being used for knowledge extraction and mining of large data under uncertain environment. No matter the issue of spatial
correlation of geography phenomenon or experience analogy, knowledge mining or experience extraction can be also completed with the artificial neural network method.

II. RESEARCH METHODS AND DATA SOURCES

A. Research Methods

The artificial neural networks (artificial neural network, ANN) is an extensive and parallel interconnection function network which is composed of simple unit with data compatibility and data adaptive ability.

Artificial neural network was developed with reference to the biological neural network. Its organization is able to simulate the reaction of the biological nervous system to the real world. They do not require the users have a clear understanding of the mechanism of evolution and development of things, the output of the system depends on the input and connection values between the inputs to output, which got by training the events happened in the history [12]. Compared with conventional statistical model, the artificial neural network is more suitable for non-linear problems whose prior distribution is not clear [13].

BP neural network is the most representative one of the widely used artificial neural network, usually consists of input layer, having a plurality of nodes of the hidden layer and one or a plurality of output nodes in the output layer. It is shown as below:

![Diagram of Neural Network](image)

Figure1. The structure of neural network

The learning process of the BP neural network algorithm consists of forward propagation and back propagation. In the forward propagation process, the input information from the input layer is treated through the hidden layer to the output layer step by step. Each layer of neurons’ state affects only the states of the neurons in the next layer. If the output layer does not get the desired output, go back and propagate reversely, retracing the error information between the output value and the expected value. By modifying the weights of each neuron, reduce errors, to continue the cycle, until the whole network error converges to a prescribed range [16-20].

Partial derivative of error to all weight coefficient is calculated to reduce the error of each by using the gradient descent method to modify the connection weights.

Input signal is firstly treated by a transmission function, secondly treated signal spreaded to the nodes of the hidden layer, finally to the output node.

B. Data Sources and Data Processing

Data include ULDI and its influence factors. Orthographic projection data of building got from GOOGLE remote sensing data. Street block data, green land use data were extracted from city planning data of Chongqing. Data of road, including trunk roads (railway and expressway) and other road, water body were extracted from second national land survey data and open street map. DEM image data got from the global land cover Research Institute: University of Maryland Earth Science Data Download Interface. Demographic data got from the Chongqing Statistical Yearbook-2009. Driving factors were resampled to unified spatial resolution of 90 meters.

1) Boundary extraction from building information

There were many indexes to indicate ULDI such as building density, building height and floor area ratio (FAR), building capacity, etc. Building density was the concentration (amount) of buildings in a given geographic area. Floor area ratio (FAR) is term for the ratio of a building's total floor area to the size of the parcel of land upon which it is built.

Building density is one of the quota of city planning. It is the main index to measure the efficiency of land use and the ecological environment livable or not in the city.

In this paper, building density is chosen to measure development intensity which is the ratio of buildings projection area to the size of the parcel of block land. The formula of building density is determined as below:

$$I = \frac{A_h}{A_z} \tag{1}$$

Where $I$ is the building density. $A_h$ is the projected area of buildings and $A_z$ is the size of the parcel of block land.

The planning data is a format of CAD, so there involves a data conversion process. Firstly, convert the dwg data format of CAD into shp data format of ARCGIS. Secondly, convert the line elements into polygon elements in ARCGIS. Lastly, merge the block data and spatial distribution data of buildings to form an integrated map.

The selection of sample should be distributed as uniformly as possible in space and different land use conditions should also be taken into account.

Sample region cover Yuzhong District and its surrounding area, Green Belt along the mountainous region, Beibei, Tongjiqiao, Lijiatuo and Dadukou, etc.

Orthographic projection of the building information in the sample area is extracted. The area of square and road are not included in the calculation of building density in this study.

2) Effect of factor selection and data processing
Land use is the interaction results by socio-economic factors and natural environment factors. Driving factors of ULDI include the natural environment factors, socio-economic factors and Accessibility factors. ULDI driving factors selected in the study were as follows:

① Topographical factors include elevation, slope, waviness and roughness.

② Accessibility factors include the distance to roads, railways, water bodies, rivers, and the center of markets.

③ Socio-economic factors include population density, cultural sites, and the distance to the green space.

In order to facilitate the simulation in MATLAB, we need to do some corresponding processing to each factor in ARCGIS.

C. Simulation

In this paper, three-layer artificial neural network model is introduced. The first layer is data input layer. There are 13/14 neurons in the data input layer, which represents the variables that affect the dynamic evolution of urban land use correspondingly. The second layer is the hidden layer. According to the Kolmogorov theorem, the hidden layer neurons number must be at least 2n/3 (n is the number of input layer neurons) for the three layer nonlinear neurons. In this paper, the activation function of tansig and logsig is used. The number of neurons in the hidden layer is 12. The third layer is the output layer, composed of only one neuron, corresponding to the intensity of urban land development.

The hidden layer produces a certain response value of these signals and exports them to the next layer. The response function of the neurons in the hidden layer can be expressed as follow:

$$f(x) = \tan \text{sig}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{2}$$

After receiving the input signal, the output layer will produce an output value which has the value from 0 to 1. The activation function of neurons in the output layer can be expressed as follows:

$$g(x) = \log \text{sig}(x) = \frac{1}{1 + e^{-x}} \tag{3}$$

In order to introduce the BP network to solving urban land development intensity, the problem is divided into the following five steps for processing in this study:

1) Data Importing. Data with format both of raster and vector are used to simulate ULDI in the study area. In the Grid-based simulation process, first of all, convert the data of ULDI and its driving factors from raster to point, and export the attribute table of point data to dbf format. Secondly, save factors data as txt file. Lastly, import the test data and training data into MATLAB for further analysis.

In the vector-based simulation process, to get land development intensity and its impact factors, we need to join the raster format data of driving factors to sampled vector data in the block. In this study, we use 1686 sample data as training data.

2) Network training. MATLAB software is used and BP neural network function was called. Different schemes are developed to test the influence of drive factors selection to the simulation results. One Scheme include the planning factors, the other does not. The number of nodes in the input layer is 13 (scheme 1), or 14 (scheme 2), while the number of nodes in the output layer is only one.

LM (Levenberg-Marquardt) algorithm and DX algorithms were used to train the neurons. LM algorithm combined gradient descent method with Gauss -Newton's law, so it not only has partial convergence of the Gauss - Newton's law, but also has global superiority of the gradient descent. DX combined the algorithm of adaptive and modify learning rate with algorithm of Momentum batch gradient descent, with quick speed of training [14].

Tansig transfer function was used between the layers of input and hidden, while logsig/purelin function were used between the layer of hidden and output. If sigmoid function was used in the last layer of BP network, the output of the network is a continuous quantity; limited to a small range from 0 to 1, as to purelin function, the output can be any value [15].

3) Simulation. We simulated with the data by using trained BP neural network.

4) Compared the results of simulation with real samples data. Calculate the error, the average error and standard deviation between the predicted value and the actual value. Obtain the average error and the standard deviation by training. If the average error and the standard deviation within the desired range, global simulation was reasonable.

5) Simulation of new data (global data). On the basis of above analysis, simulation of ULDI of the whole Metropolitan Area of Chongqing is completed.

III RESULTS ANALYSIS

A. Comparison of Different Schemes

In this study, two schemes and two different algorithms is designed to test results. Scheme 1 does not contain planning factor. There are 13 driving factors, Scheme 2 contains planning factors (commercial, industrial, residential land etc.), and there are 14 driving factors. LM algorithm and DX algorithm are chosen in this paper.

The test data will be imported into the trained network. The average error of simulation results is 0.0666. The standard deviation is 0.1039. The error is small between predicted value and actual value. The simulation results are satisfactory. The model can be used to simulate development intensity in urban in Chongqing Metropolitan Area. As can be seen from the chart, the red line represents the actual value and the blue line represents the predicted value (Figure 2).
Results show that, when increasing the driving factor, the average error, standard deviation and error accuracy are significantly improved in the scheme 2 (Table 1). Compared with the actual value, predictions obtained using different algorithms have different error value. LM algorithm is better than DX algorithm. 0.12031, 0.14787, 0.0266532 is the average error, standard deviation, error accuracy by using DX algorithm, while the average error, standard deviation and error accuracy is 0.10572, 0.13482, 0.0199608 by using DX algorithm (Table 1).

<table>
<thead>
<tr>
<th>Scheme 1</th>
<th>Scheme 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>average error</td>
<td>0.1325</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.15962</td>
</tr>
<tr>
<td>error accuracy</td>
<td>0.0539307</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme 1</th>
<th>Scheme 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>average error</td>
<td>0.13615</td>
</tr>
<tr>
<td>standard deviation</td>
<td>0.15489</td>
</tr>
<tr>
<td>error accuracy</td>
<td>0.027931</td>
</tr>
</tbody>
</table>

Simulation results based on DX and LM algorithm reflect the actual situation of land development intensity and are in accordance with the city space distribution law.

Namely, on the first level, as the center of Yuzhong district, development intensity or building density spread to outer layer. On the second level, as the center of Shapingba, Shiqiao-Yangjiaping-Daping. Namping, building density gradually decline.

Compared with the actual value, the simulation results are smoother. The maximum value of simulation is 0.485-0.619, while the actual value is between 0.631-0.980. From the view of residual distribution, most of the area error is smaller, and DX and LM algorithms are similar in spatial distribution characteristic.

In order to test simulated accuracy of block as a unit (vector based) and grid as a unit, we compared the simulation results with the actual development intensity. 0.717006 is the maximum observe value on the Nan’ an, while 0.665935 is the maximum simulated value in block unit and 0.563301 is the maximum simulation value in grid unit (90*90).

From this point of view, better effect of simulation can be obtained with block unit. One possible reason is that land development intensity is theoretically measured by block building density. Generally speaking, development intensity simulation can capture the city development law as block a unit. They can lead to fragmentation and excessive computation as grid a unit.

B. Global Simulation Results

Factors data is imported to the trained BP network and global simulation results of Chongqing Metropolitan Area can be obtained. From figure 4 we can see, the prediction results are in the range of 0-1[0, 1]. On the whole, it is in line with the actual situation, and can be used as the evaluation results of Chongqing Metropolitan Area.

The matrix of simulation results (building density value) was saved as DBF format data after a series of transformation, and loaded into the ARCMAP software, join in with block ID in ARCGIS, so we can get building density distribution map of each block.

Detail steps to visualize the simulation results are:
1) The land development intensity prediction value is derived from the MATLAB to ASCII, and coisnvert the result as EXCEL format file.
2) The building density after conversion is copied to the SPSS, ID code is added, and then DBF format data was saved.
3) The DBF format data is loaded into the ARCMAP.
4) The simulation result data was joined in ID code of the Chongqing City District block data.

From Figure 4, the result of the simulation, the high value of development intensity distribution lie in Jiefang Bei of Yuzhong district, Jielong Town of Dadukou district, lijiatuo of BaNan. Cuntan of Jiangbei district, and Xipeng, Chayuan, Tong Jing, Beibei. The prediction results show a large high strength distribution area is in lijiatuo area, even higher than Jie fang Bei of Yuzhong district. In a certain extent, it suggested that these places are the major development area in the future.

C. Comparision of Neural Network Model and Multiple Linear Regression Model Results

In this research, the simulation results of BP neural network were compared with the results of multiple linear
regression model to analyze the pros and cons of neural network.

Based on the construction the regression equation between land development intensity and influence factors, the global spatial interpolation (trend) method was used to estimate land development intensity of Chongqing Municipality (9 main districts) by using ARCGIS and SPSS software.

Multiple linear regression model is actually established the relationship between the various factors and land development intensity. The multiple linear regression formula is as follows:

\[
\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n
\]

\(\hat{y}\) is the estimated result according to \(x\) variables;

\(b_0\) is a constant;

\(b_1, b_2, \ldots\) is regression coefficient;

\(x\) is the driving factor, such as the elevation, slope, relief, roughness, density of population, cultural sites, parks, green space, the distance away from the main road (highway, railway), the distance away from other waters, the distance away from rivers, and distance from the market.

\[
\begin{array}{|c|c|c|c|}
\hline
\text{Distance from} & \text{Non standardized coefficient} & \text{standardized coefficient} \\
\text{variable} & \text{B coefficient} & \text{standard error} & \text{trial version} \\
\hline
\text{constant} & 0.097 & 0.001 & - \\
\text{Distance from green space} & 7.938E-5 & 0.000 & 0.419 \\
\text{Distance from market} & -5.270E-6 & 0.000 & -0.299 \\
\text{Topographic relief} & 0.000 & 0.000 & -0.062 \\
\text{Distance from traffic branch} & 4.830E-6 & 0.000 & 0.254 \\
\text{Distance from Yangtze River and Jialing River} & -2.463E-6 & 0.000 & -0.126 \\
\text{Elevation} & -7.750E-5 & 0.000 & -0.097 \\
\text{Distance from main road} & -2.274E-6 & 0.000 & -0.040 \\
\text{Slope} & -0.001 & 0.000 & -0.053 \\
\text{Density of population} & -3.943E-5 & 0.000 & -0.023 \\
\text{Distance from other waters} & 1.624E-6 & 0.000 & 0.025 \\
\hline
\end{array}
\]

The results (Table II) show that: (1) the main impact factor is slope, followed by the ecological green space and elevation factors; (2) there are positively correlation between ecological green land, waters, branch traffic and land development intensity; otherwise there are negatively correlation between slope, elevation, main roads and land development intensity; (3) the method can overall reveal the relationships between land development intensity and its driving factors.

The multiple linear regression model is used in simulation results of ULDI in the sample region and actual development intensity was compared. Compared the results from multiple linear regression model with the actual development of strength, the simulation results have a certain bias, especially in Jiefangbei vicinity, however the result from the neural network model is more closer to the real development of strength, reflecting the change trend of ULDI from each group center respectively to the peripheral extension. The results show that advantages for BP neural network model are much more than linear regression model.

The simulation error of the neural network is much smaller than multiple linear regression model. The residual value of the prediction results of neural network model is about from -0.005 to 0.2 while the results of multiple linear regression model is about from -0.2 to 0.6. It is easy to see that method of neural network simulation is much better than multiple linear regression model.

By contrast, the neural network model is much better than multiple linear regression model in simulating ULDI, but the multivariate linear regression model can clearly show the quantitative relation between ULDI and its driving factors, but the neural network model unable to do so.

IV RESULTS AND DISCUSSION

BP neural network method was used to simulate land development intensity by selecting the data of driving factors and land development intensity based on the correlation of data, availability and other conditions. Results show that:

(1) The artificial neural networks method can reveal the relationship between land development intensity and its influence factor in a contiguous urban area. It is an innovation method for quantitative simulation of land development intensity. The simulation results basically reflect the general distribution trend of land development intensity in Chongqing Metropolitan Area.

(2) Through different scheme comparison, we found that the accuracy of simulation will be improved when increasing the driving factors such as planning factors; it suggests that it is very important to simulation results for select reasonable factors.
This research was financially supported by Natural Science, Chongqing University of Posts and Telecoms.

The further work is: (1) To add more driving factors, especially for more human and social economic factors. (2) To capture spatial data using the volume or building capacity to represent the intensity of land development. (3) To improve simulation accuracy by considering quantitative relations between driving factors changes with time.

ACKNOWLEDGMENT

The authors wish to thank the helpful comments and suggestions from my colleagues in School of Computer Science, Chongqing University of Posts and Telecoms. This research was financially supported by Natural Science Foundation Project of CQ CSTC (Grant NO.2011jjA30014).

REFERENCES


© 2013 ACADEMY PUBLISHER
His research addresses questions in urban, land use planning through the application of GIS methods. He is also interested in the development of new analytical methods for geographical research-especially GIS-based 3D geovisualization and geocomputation. Part of his current research attempts to integrate GIS and qualitative methods to simulate urban land use dynamic change.

Dr. Liu has been visiting senior research during the period of time correspondent to May 2011 to June 2012, supported by the Chongqing University of Posts and Telecommunication. He has worked with the area of Geographical Information Systems during his visit to Sweden at the department of informatics at the International Business School at Jönköping. Dr. Liu is now the director of “3S” technology application centre, Chongqing University of Posts and Telecommunications. Publications: (1) Introduction to Geographic Information Systems. Chongqing university press. ISBN 978-7-5624-5638-4. Liu Minghao, Xia Ying, Yuan Zhengwu2010. (2) Economic Analysis of Land banking. Chongqing university press. Liu Minghao2010.


Donghong Li, Female, born in shaanxi, March 1, 1988. From 2007 to 2011, studied at the school of communication and information engineering of Chongqing University of Posts and Telecommunications, get the bachelor's degree; So far in 2011, studied at the school of computer science of Chongqing University of Posts and Telecommunications, the main research direction is the geographic information system.