Mining Weights of Land Evaluation Factors Based on Cloud Model and Correlation Analysis

HU Shiyuan  LI Deren  LIU Yaolin  LI Deyi

Abstract  The veracity of land evaluation is tightly related to the reasonable weights of land evaluation factors. By mapping qualitative linguistic words into a fine-changeable cloud drops and translating the uncertain factor conditions into quantitative values with the uncertain illation based on cloud model, and then, integrating correlation analysis, a new way of figuring out the weight of land evaluation factors is proposed. It may solve the limitations of the conventional ways.

Keywords  cloud models; correlation analysis; land evaluation factor; weight data mining

CLC number  P208; P271

Introduction

As one of the critical steps in the processes of urban land evaluation, weight decision for land evaluation factors, whose accuracy has direct impact on land grading, can be recognized as one of the core actions in such work. Generally, the ways currently adopted to measure weight include Delphi method, AHP method, factor paired comparison method, gray assessment method, regression analysis, principal component analysis, etc., all of them are based on exactly collected data, and the final result is calculated through mathematic way. Whereas real world is a nonlinear and dynamic uncertain system, containing multiple parameters, the data used to describe it are also uncertain, thus leads to hard description for real world in precision with conventional mathematical model. Nevertheless, a new cloud model, as a quality-quantity interchangeable model combining traditional fuzzy mathematics and probability and statistics, realized their conversion between qualitative words and quantitative values. Based on the cloud model and correlation analysis, taking deliberate consideration of the fuzziness and randomness existing in objects and human knowledge, this paper successfully builds the conversion model and proposes a new way for weight decision of land evaluation factors.

1  Description of evaluation factors with cloud model

1.1  Concept of cloud model

Cloud model is a uncertainty conversion model between some qualitative concept expressed with linguistic words and its relative quantitative value. Set \( U \) as a domain comprising accurate numbers, whose accorded qualitative concept is \( \tilde{A} \), for any given element \( X \) in domain, there is a random number

Received on May 31, 2007.

HU Shiyuan, School of Resource and Environment Science, Wuhan University, 129 Luoyu Road, Wuhan 430079, China; Key Laboratory of Geographic Information System of Ministry of Education, 129 Luoyu Road, Wuhan 430079, China.

E-mail: hushiyuan@vip.sina.com
having stable tendency \( y = \mu_x(x) \), which is called certainty degree of \( x \) for \( A \), and the distribution of \( y \) in domain is called cloud model, and simple named cloud. The cloud consists of a number of cloud drops, each of which presents one point projected from qualitative concept to numeral domain, that is a concrete realization. For its every realization, cloud model can denote the certainty degree of the point reflecting some qualitative concept.

Cloud is identified by three digital characteristics, \( E_x \) (expected value), \( E_y \) (entropy) and \( H_y \) (hyper entropy), which express the quantity feature of qualitative concept.

The algorithm of generative cloud is called cloud generator, which includes forward cloud generator, condition \( X \) cloud generator, condition \( Y \) cloud generator and backward cloud generator. When cloud drops are produced by the digital characteristics of cloud, the cloud generator is called forward cloud generator; when cloud's three digital characteristics \( E_x, E_y, H_y \) and the special value \( x_0 \) are given, the cloud generator is called condition \( X \) cloud generator; when cloud's three digital characteristics \( E_x, E_y, H_y \) and the special value \( \mu_y \) are given, the cloud generator is called condition \( Y \) cloud generator; giving a group of cloud drops \( \mu_x(x) \) reflects the according qualitative rule’s activation intensity) and pipes into \( CG_y \), this figure \( \mu_y \) controls \( CG_y \) to produce a group of random drops \((y, \mu_y)\). It need a brief account that the cloud drop and output produced by generator are not only one and uncertain, so it can achieve uncertainty illation.

1.2 Uncertainty illation based on cloud model

The formalization description of a qualitative rule is: If \( A \) then \( B \). \( A \) and \( B \) are cloud object denoting language value. The cloud generator is the basis of doing uncertainty illation by cloud model. Connecting a condition \( X \) cloud generator and a condition \( Y \) cloud generator can make a single rule generator. Fig.1 is the schematic diagram about single rule generator, in the figure, \( CG_x \) denotes condition \( X \) cloud of the input language value \( A \). \( CG_y \) denotes condition \( Y \) cloud of the output language value \( B \). When we input some special value \( x_0 \) to activate \( CG_x \), \( CG_x \) produces a group of figures \( \mu_x \) (\( \mu_x \) reflects the according qualitative rule’s activation intensity), and pipes into \( CG_y \), this figure \( \mu_y \) controls \( CG_y \) to produce a group of random drops \((y, \mu_y)\). It need a brief account that the cloud drop and output produced by generator are not only one and uncertain, so it can achieve uncertainty illation.

![Fig.1 Single rule generator](image)

1.3 Description of evaluation standard of land evaluation factors with cloud model

For residential land of a city, we choose prosperity degree, traffic condition, municipal infrastructure, environment and land potential to be land evaluation factors, every factor include some specific sub-factors. Table 1 shows the evaluation standards of the land evaluation factors.

| Factor          | Sub-factor            | Evaluation standard       |
|-----------------|-----------------------|---------------------------|
| Prosperity degree | To emporium           | Nearer | Near | Common | Far | Farther |
|                 | To market             | Nearer | Near | Common | Far | Farther |
| Traffic condition | Road accessibility    | Better | Good | Common | Bad | Worse |
|                 | Bus facility          | Better | Good | Common | Bad | Worse |
| Municipal infrastructure | Drainage | Better | Good | Common | Bad | Worse |
|                 | Heating               | Better | Good | Common | Bad | Worse |
| Environment     | Quality of environment| Clearer | Clear | Light polluted | Middle polluted | Heavy polluted |
| Land potential  | Urban plan            | High grade residence | Agminate residence | Common residence | Scattered residence | Others |
evaluation factors.

In Table 1, the evaluation standard of every evaluation factors are according with the knowledge of experts respectively. They use a qualitative and fuzzy language value to express a certain numerical range. For example, evaluation standard “nearer, near, common, far, farther” in factor “to emporium” are according with the distance degree from a certain land to emporium respectively. The evaluation standard of the factor “heating” are according with the heating rate, so do others.

The five qualitative concepts of the evaluation standard of the factor “to emporium” are expressed with cloud model in distance field as follows:

\[ D_A = \begin{cases} 
1, & x \in [0,300] \\
D(300,200/3,0.05), & \text{others} 
\end{cases} \]

\[ D_A = D(750,300/3,0.05) \]

\[ D_A = D(1050,150/3,0.05) \]

\[ D_A = D(1350,300/3,0.05) \]

\[ D_A = \begin{cases} 
D(1800,300/3,0.05), & \text{others} \\
1, & x \in [1800,\infty] 
\end{cases} \]

Fig.2 shows the unite distribution between the cloud drops and theirs certainty degree of the five qualitative concepts of the evaluation standard “to emporium”.

**2 Mining weights base on cloud model and correlation analysis**

**2.1 Data preparation**

According to the factor system of land style to be evaluated and the spacial distribution of samples, we collected a number of representative samples data about this land style which include the price of samples and their evaluation factors’ concrete status. Table 2 shows the sample data of the residential land.

| Sample price /yuan • m² | 3722 | 2896 | 1389 | 934 | 758 | 300 |
|-------------------------|------|------|------|-----|-----|-----|
| To emporium             | Nearer | Near | Common | Far | Farther | Farther |
| To market               | Nearer | Near | Near | Common | Far | Far |
| Road accessibility      | Better | Better | Good | Common | Bad | Bad |
| Bus facility            | Better | Better | Better | Better | Bad | Worse |
| Drainage                | Better | Better | Better | Better | Good | Common |
| Heating                 | Better | Better | Middle polluted | Light polluted | Light | Light |
| Environment quality     | Clear | Clear | Clearer | Common residence | Common | Scattered |
| Urban plan              | High grade residence | High grade residence | Agminate residence | |

**2.2 Cloud model description of samples and uncertainty illation**

Firstly, we use cloud model to express the factor’s evaluation standard of the appointed land style according to the expert’s knowledge and experience by means of the cloud model description method in Section 1.3.

The quality of land is mapped as land price of some land style, then there is a positive relation between land price and factors which influence land quality. Before analyzing the relation between land price and evaluation factors, we can translate the uncertain factor conditions into quantitative impact values which land evaluation factor impacts on land with the uncertain illation based on cloud model.

Take the factor “to emporium” as example, we can establish the qualitative illation rule by the knowl-
edge of expert as follows:

1) If to emporium "nearer", then impact value "higher";
2) If to emporium "near", then impact value "high";
3) If to emporium "common", then impact value "common";
4) If to emporium "far", then impact value "low";
5) If to emporium "farther", then impact value "lower".

The description methods of the five qualitative concepts of the rule’s front with cloud model in distance field is described as foregoing. The five qualitative concepts of the rule’s rear are expressed with cloud model in impact value field as follows:

\[ D_{h_1} = D(90,15 \pm 3,0.05), \quad D_{h_2} = D(70,15 \pm 3,0.05) \]
\[ D_{h_3} = D(50,15 \pm 3,0.05), \quad D_{h_4} = D(30,15 \pm 3,0.05) \]
\[ D_{h_5} = D(10,15 \pm 3,0.05) \]

Using the cloud of the rule’s front to drive forward cloud generator, we can get a group of distance figures, then get theirs average to be the input of condition \( X \) cloud generator, and count the certainty degrees of the five qualitative concepts of the rule’s front respectively, we can find the maximum \( \mu_i \). It is said that the \( i \)-th rule in rule-base is activated, the activation intensity is \( \mu_i \). Using the single rule generator to generate output, and take the output to be this factor’s impact value. When calculating the output by using the algorithm of condition \( Y \) cloud generator, we choose the plus or the minus is according that the input is of up or down. When \( x < E_{a_f} \), we choose the minus, when \( x > E_{a_f} \), we choose the plus. We can get the impact value of the other factors on plot by using uncertainty illation. Factor impact value and sample price of plot are make up of the reference data and compare data. It is shown in Table 3.

### Table 3  Reference data and comparing data

| Sample price/ yuan • m² | 3 722 | 2 896 | 1 389 | 934 | 758 | 300 |
|-------------------------|-------|-------|-------|-----|-----|-----|
| To emporium             | 100.00| 78.86 | 65.38 | 35.86| 18.39| 8.17 |
| To market               | 89.65 | 82.47 | 72.93 | 63.59| 30.56| 21.52|
| Road accessibility       | 98.26 | 89.51 | 78.28 | 52.67| 33.56| 24.42|
| Bus facility            | 87.21 | 86.83 | 82.56 | 81.23| 45.12| 10.85|
| Drainage                | 98.25 | 97.39 | 99.03 | 70.23| 50.78| 10.08|
| Heating                 | 98.37 | 96.12 | 98.10 | 12.56| 11.43| 10.08|
| Environment quality     | 75.91 | 72.49 | 86.28 | 30.28| 51.52| 52.03|
| Urban plan              | 96.67 | 93.41 | 93.68 | 55.31| 45.46| 43.06|

### 2.3 Calculating weights by correlation analysis

First, we must normalize the reference data and compare data, then calculate weights of land evaluation factors according to the following formula:

\[ \zeta_i(k) = \min \min \{ x_i(k) - x_i (k) \} + \zeta \max \max \{ x_i(k) - x_i (k) \} \]

\[ r_i = \frac{1}{N} \sum_{k=1}^{N} \zeta_i(k), \quad w_i = \frac{r_i}{\sum_{i=1}^{N} r_i} \times 100 \]

where \( x_i \) is impact value; \( \zeta_i(k) \) is the correlation coefficient \( x_i \) to \( x_u \) in moment \( K \); \( \zeta \) is distinguish coefficient, it is usually chosen between 0 and 1; \( r_i \) is the correlation degree of each land evaluation factors; \( w_i \) is the weight of each land evaluation factors.

After calculating, the weight of residential land evaluation factors are shown in Table 4.

### Table 4  Weight of residential land evaluation factors

| Factor Sub-factor | Weight |
|-------------------|--------|
| Prosperity degree | 0.29   |
| To emporium       | 0.16   |
| To mart           | 0.13   |
| Traffic condition | 0.25   |
| Road accessibility| 0.13   |
| Bus facility      | 0.12   |
| Municipal infrastr| Drainage | 0.10 |
| (0.24)            | Heating  | 0.14 |
| Environment quality| 0.11 |
| Land potential    | 0.11   |
| Urban plan        | 0.11   |

### 2.4 Comparison analysis and conclusions

Calculating weights of residential land evaluation factors by using Delphi, correlation analysis and the combination method of cloud model and correlation analysis respectively, the results are showed in Table 5.
Table 5 Comparison between results of cloud model and other ways

| Method                      | To emporium | To market | Road accessibility | Bus facility | Drainage | Heating | Environment quality | Urban plan |
|-----------------------------|-------------|-----------|--------------------|--------------|----------|---------|---------------------|------------|
| Cloud model and correlation analysis | 0.16        | 0.13      | 0.13               | 0.12         | 0.10     | 0.14    | 0.11                | 0.11        |
| (Prosperity degree)         | 0.29        | 0.13      | 0.13               | 0.12         | 0.24     | 0.14    | 0.11                | 0.11        |
| (Traffic condition)         | 0.17        | 0.13      | 0.13               | 0.12         | 0.25     | 0.14    | 0.12                | 0.10        |
| Delphi                      | 0.16        | 0.12      | 0.13               | 0.13         | 0.10     | 0.14    | 0.10                | 0.12        |
| (Prosperity degree)         | 0.28        | 0.12      | 0.13               | 0.13         | 0.25     | 0.14    | 0.10                | 0.12        |
| (Traffic condition)         | 0.28        | 0.12      | 0.13               | 0.13         | 0.24     | 0.14    | 0.10                | 0.12        |
| Correlation analysis        | 0.28        | 0.12      | 0.13               | 0.13         | 0.25     | 0.14    | 0.10                | 0.12        |
| (Prosperity degree)         | 0.28        | 0.12      | 0.13               | 0.13         | 0.25     | 0.14    | 0.10                | 0.12        |
| (Traffic condition)         | 0.28        | 0.12      | 0.13               | 0.13         | 0.25     | 0.14    | 0.10                | 0.12        |

From Table 5, we can see that distribution of the factors’ weights by using three methods are uniform basically, it reflects that the realization of experts which is about the intensity that land evaluation factors impact on land quality are according with the results of objective data analysis. Firstly, in land evaluation factors, the weight of the factor “emporium” is the largest, which shows that the prosperity degree of land position has the most important influence on land quality. The second is the factor “heating”, this is because of the cold climate of northern city. It reflects that people pay a lot attention to heating equipments. Another important factor in land evaluation is traffic condition. The traffic condition has a direct influence on cost of the moving of people and materials and convenience degree. Besides, from the distribution of weights, we could find that with the development of people’s living standard, they are not only satisfied with just a place to live, but pay more attention to the quality of environment around their residence.

The experiment shows that it is feasible to describe the land evaluation factors by using cloud model. The method based on cloud model’s uncertain illation and correlation analysis can quantitatively reflect the intensity of evaluation factors on land, and make the results of weights closer to fact. Compared with some traditional methods like Delphi and correlation analysis, it has some traits as follows: ① coming from objective data, the results of weights eliminate the influence of subjective factors on evaluation results. ② The sample data collected are no longer sclerous accurate figure, but being described by natural language. It makes full consideration of the fuzziness and randomness about the real world and describes the real world more exactly and perfectly. It coheres with human’s realization rules about objective world. ③ On data collecting and processing, it is easier than other ways.

This method is applicable to confirm the index weight of multi-attribute evaluation, especially to these conditions which the evaluation factors are described by qualitative linguistic words. The result’s accuracy from this method is determined by the accuracy of the factor evaluation standard model.

**References**

[1] Wang Shuliang (2000) Data field and cloud model based spatial data mining and knowledge discovery[D]. Wuhan: Wuhan University (in Chinese)

[2] Li Deyi (2005) Artificial intelligence with uncertainty[M]. Beijing: National Defence Industry Press (in Chinese)

[3] Li Deren, Wang Shuliang, Li Deyi, et al. (2002) Theories and technologies of spatial data mining and knowledge discovery[J]. Geomatics and Information Science of Wuhan University, 27(3):221-233 (in Chinese)

[4] Di Kaichang (2003) Spatial data mining and knowledge discovery[M]. Wuhan: Wuhan University Press (in Chinese)

[5] Li Xingsheng (2003) Study on classification and clustering mining based on cloud model and data field[D]. Nanjing: PLA University of Science and Technology (in Chinese)

[6] Fan Dingguo, He Shuo, Duan Fu, et al. (2003) A comprehensive evaluation model based on cloud model[J]. Sci/Tech Information Development & Economy, 13(12): 157-159 (in Chinese)

[7] Liu Yaolin (2003) Land information system[M]. Beijing: China Agriculture Press (in Chinese)