

Applying Self-Organizing Competition Artificial Neural Networks to Classify the Soil

Xiangdong Ma*, Qiang Fu**

* School of Water Resources & Environment, Hohai University, Nanjing, Jiangsu 210024, China;

** School of Water Conservancy & Civil Engineering, Northeast Agricultural University, Harbin, Heilongjiang 150030, China, fuqiang100@371.net

Abstract: Through applying the clustering function of Self-Organizing Mapping (SOM) network, the writer uses MATLAB 5.3 software to classify 21 kinds of soil samples in Sanjiang Plain. Through comparing the result with references that uses the method of fuzzy clustering, the paper concludes that the SOM network can reflect the complicated information among each soil samples. The effect of classification is good, and can be applied on soil classification. [Nature and Science 2003;1(1):75-81].

Key words: SOM network; soil classification

1. Introduction

Soil classification is not only the basis of soil science, but also the synthetic symbol of the developed level of soil science (Liu, 1988; Fu, 2003). The soil analyzing system studied out by human in different period can reflect the understanding level and soil science seedtime of this period. At present, the study on soil classification has developed toward the direction of using evaluated index and quantity index (Liu, 1988; Xie, 2000; Liu, 1997; Zhao, 2000). Fuzzy clustering analysis has been applied broadly. But in fuzzy clustering analysis, there are different methods used by different people to demarcating the data, such as correlation coefficient method, distance method and so on. Different methods can obtain different fuzzy similarity matrix. The final result of classification has partial difference. Different demarcating method can use and pick-up different information in the soil samples. Otherwise, when we calculate the best-classified result (the best threshold value λ), we often apply the method of mathematical statistics. That means that the best threshold value will be calculated through building statistical value F . We will analyze and judge further whether the threshold value λ is the best or not. These show that a certain soil classification method can only provide a tool and thought. For a concrete issue, we should combine the practical condition to making a right judgement. Based on this, the author put forward a new method to classify the soil. This is Self-Organizing Feature Map (SOM) network that has the function of clustering.

2. SOC Artificial Neural Networks

2.1 Brief Introduction of SOC Networks

Self-organizing compete (SOC) artificial neural networks is a kind of networks without teaching. It has the function of self-organizing. Through training by

itself, it can classify the input samples automatically. The basic thought of SOC networks is that each nerve cell in completed layer completes the responded chance to input mode. At last, only one nerve cell becomes the winner. At the same time, it can adjust the linked weights relative to the winner nerve cell toward the better direction. So, self-organizing compete artificial neural networks can be applied to classification aspect (Wen, 2000; Wang, 1995).

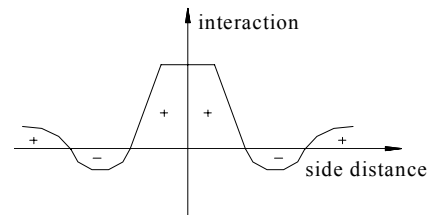


Figure 1. The Relation of Lateral Interaction

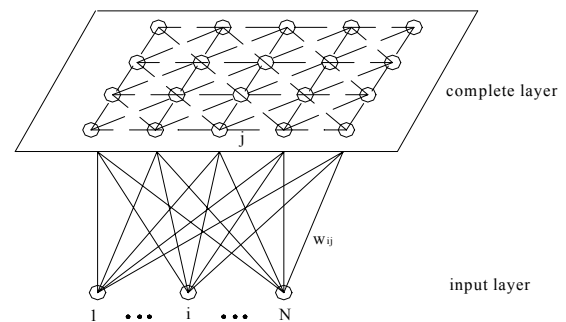


Figure 2. The Structure of Self-organizing Character Mapping Network

2.2 Self-Organizing Feather Map (SOM) Artificial Neural Networks (Wen, 2000; Wang, 1995)

SOM has been put forward in 1981 by professor

Kohonen. The corporate feature of nerve cell is that the neighbouring two nerve cells excite through stimulating each other, and the farside nerve cells restrain each other. The farther nerve cells have a little excitement. This kind of part mutual relation can be described in Figure 1. The network configuration sees to Figure 2. The number of nerve cell in input layer is n . The competed layer composes $M = m^2$ nerve cells. Thus, these nerve cells build up a two-dimension plane array. The nerve cells in input layer are all linked with the nerve cells in competed layer. Sometimes, side controlled links each nerve cell in competed layer. There are two linked weights in the networks. One is to respond to exterior input. The other one is among the nerve cells. The function of the second weight is to control the interaction of every nerve cell.

2.3 The Learning and Working Rule of SOM Networks (Wen, 2000; Wang, 1995)

In Figure 2, let the input mode is

$$P_k = (p_1^k, p_2^k, \dots, p_n^k), \quad k = 1, 2, \dots, q.$$

The nerve cell vector in competed layer is

$$A_j = (a_{j1}, a_{j2}, \dots, a_{jm}), \quad j = 1, 2, \dots, m.$$

Where: P_k —— continue value; A_j —— number quantity.

The linked weight between nerve cells j in competed layer and nerve cells in input layer is as follows.

$$W_j = (w_{j1}, w_{j2}, \dots, w_{ji}, \dots, w_{jN}), \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, M.$$

The learning and working rule is as follows.

(1) Initialize. Endow the linked weight ($\{w_{ij}\}$) of networks with random value in the region of $[0, 1]$. $i = 1, 2, \dots, N, j = 1, 2, \dots, M$. Define the original value of learning rate $\eta(0)$ ($0 < \eta(0) < 1$) and neighbouring threshold $N_g(t)$. ($N_g(t) = N_g(0)$). The total learning time is T .

(2) Select any one mode (P_k) in the total mode (q) to input layer. At the same time, we should normalize the data as follows.

$$\bar{P}_k = \frac{P_k}{\|P_k\|} = \frac{(p_1^k, p_2^k, \dots, p_n^k)}{[(p_1^k)^2 + (p_2^k)^2 + \dots + (p_n^k)^2]^{1/2}}$$

(3) Normalize the linked weight $W_j = (w_{j1}, w_{j2}, \dots, w_{ji}, \dots, w_{jN})$. Calculate the distance between \bar{W}_j and \bar{P}_k .

$$\bar{w}_j = \frac{w_j}{\|w_j\|} = \frac{(w_{j1}, w_{j2}, \dots, w_{jn})}{[(w_{j1})^2 + (w_{j2})^2 + \dots + (w_{jn})^2]^{1/2}}$$

$$d_j = \left[\sum_{i=1}^N (\bar{p}_i^k - \bar{w}_{ji})^2 \right]^{1/2}, \quad j = 1, 2, \dots, M$$

(4) Find out the least distance d_g . Define the winner nerve cell g . $d_g = \min[d_j], \quad j = 1, 2, \dots, M$

(5) Adjust the linked weight. Modify all the linked weight between the nerve cells in the neighbouring threshold ($N_g(t)$) in competed layer and in input layer.

$$\overline{w_{ji}(t+1)} = \overline{w_{ji}(t)} + \eta(t) \cdot [\bar{p}_i^k - \overline{w_{ji}(t)}], \quad j \in N_g(t), \quad j = 1, 2, \dots, M, \quad (0 < \eta(0) < 1)$$

(6) Select another learning mode to networks input layer. Return to step (3). The learning process will not end until all the learning modes have been provided to networks.

(7) Renovate the learning rate $\eta(t)$ and $N_g(t)$.

$$\eta(t) = \eta(0) \left(1 - \frac{t}{T}\right)$$

$\eta(0)$ —— original learning rate; t —— learning time; T —— the total learning time.

Now, let the coordinate value of any nerve cell (g) in competed layer is (x_g, y_g) . The scope of neighbouring region is a square. The top right corner of the square is $(x_g + N_g(t), y_g + N_g(t))$. The down left corner of the square is $(x_g - N_g(t), y_g - N_g(t))$. The

revised formula is $N_g(t) = INT \left[N_g(0) \left(1 - \frac{t}{T}\right) \right]$.

$INT[x]$ —— the symbol of getting the integer; $N_g(0)$ is the original of $N_g(t)$.

(8) Let: $t = t + 1$. Return to step (2). The whole process will not end until $t = T$.

3. Soil Classification Based on SOM Networks.

Now, we use the data of Liu (1988). There are 21 samples in Songhua River area. There are 9 indexes of each sample. The original data see to Table 1.

Ensure the input mode of the networks is $P_k = (p_1^k, p_2^k, \dots, p_n^k), k = 1, 2, \dots, q. (q = 21, N = 9)$.

Calculate these data based on software MATLAB 5.3.

(1) Using function newsom (P, [D1, D2, ...Di]) to build up a SOM networks.

P —— input vector. i —— the dimension number.

We should give the scope of each input vector.

net=newsom([0 1; 0 1; 0 11; 0 10; 0 50; 0 70; 0 70; 0 2; 0 50], [6 1] (The number can be adjusted));

Table 1. Soil Sample and Its Character Index

| Sample number | Soil name | Total nitrogen (%) | Total phosphor (%) | Soil character | | | | | Capability $g \cdot cm^{-3}$ | Field moisture capacity (%) |
|---------------|--|--------------------|--------------------|--------------------|--------|-------------------------------|-----------------------|---------------------------|------------------------------|-----------------------------|
| | | | | Organic matter (%) | pH | Substitio n quantity (me/100) | Cultivated depth (cm) | Physical clay content (%) | | |
| 1 | Folium mucosity bottom white slurry black soil | 0.270 | 0.142 | 6.46 | 5.5 | 35.8 | 21 | 45.3 | 1.03 | 29.3 |
| 2 | Thick-level mucosity bottom black soil | 0.171 | 0.115 | 3.46 | 6.3 | 33.0 | 60 | 45.3 | 0.78 | 38.9 |
| 3 | Folium mucosity bottom black soil | 0.114 | 0.101 | 2.43 | 6.4 | 26.5 | 25 | 51.0 | 1.13 | 31.6 |
| 4 | Thick-level mucosity bottom black soil | 0.173 | 0.123 | 3.30 | 5.8 | 28.9 | 65 | 45.6 | 1.09 | 36.1 |
| 5 | Folium mucosity bottom black soil | 0.145 | 0.131 | 3.28 | 6.0 | 28.5 | 25 | 51.0 | 1.03 | 30.5 |
| 6 | Thick-level black meadow | 0.173 | 0.140 | 3.45 | 5.8 | 33.4 | 60 | 49.0 | 0.98 | 35.2 |
| 7 | Middle-level black meadow | 0.250 | 0.177 | 5.51 | 7.2 | 42.5 | 45 | 46.6 | 0.93 | 29.8 |
| 8 | Folium black meadow soil | 0.237 | 0.189 | 5.37 | 6.1 | 32.9 | 27 | 45.0 | 1.00 | 33.0 |
| 9 | Folium gully black meadow soil | 0.319 | 0.227 | 7.04 | 5.8 | 35.9 | 24 | 39.3 | 1.03 | 28.8 |
| 10 | Thick-level flat meadow soil | 0.163 | 0.124 | 3.73 | 6.2 | 30.6 | 61 | 48.1 | 1.28 | 26.0 |
| 11 | Middle-level flat meadow soil | 0.194 | 0.201 | 4.50 | 5.7 | 30.9 | 35 | 47.4 | 1.25 | 43.9 |
| 12 | Thick-level gully meadow soil | 0.142 | 0.185 | 3.79 | 6.4 | 32.5 | 55 | 51.0 | 1.10 | 22.8 |
| 13 | Thick-level gully latent raised meadow soil | 0.240 | 0.217 | 4.92 | 6.5 | 37.3 | 41 | 63.6 | 1.17 | 35.5 |
| 14 | Folium gully latent raised meadow soil | 0.253 | 0.172 | 4.63 | 6.8 | 35.7 | 20 | 44.1 | 1.15 | 33.0 |
| 15 | Thick-level plat carbonate meadow soil | 0.357 | 0.289 | 7.21 | 7.5 | 42.1 | 40 | 48.3 | 0.80 | 33.0 |
| 16 | Middle-level gully carbonate meadow soil | 0.280 | 0.204 | 10.68 | 6.7 | 42.5 | 31 | 41.5 | 1.05 | 41.0 |
| 17 | Middle-level meadow | 0.164 | 0.141 | 3.05 | 4.8 | 19.7 | 30 | 30.9 | 1.22 | 29.0 |
| 18 | Arenaceous meadow soil | 0.095 | 0.099 | 1.51 | 6.0 | 16.0 | 20 | 26.4 | 1.27 | 26.0 |
| 19 | Eroded dark and brown soil | 0.392 | 0.240 | 6.62 | 5.3 | 37.3 | 14 | 34.7 | 1.10 | 26.7 |
| 20 | Folium white slurry soil | 0.267 | 0.208 | 6.25 | 5.8 | 39.5 | 19 | 42.4 | 1.10 | 28.0 |
| 21 | Yellow and white slurry soil | 0.137 | 0.111 | 3.04 | 5.1 | 23.7 | 18 | 43.1 | 1.37 | 34.0 |
| Average value | | 0.2160 | 0.1684 | 4.7729 | 6.0810 | 32.6286 | 35.0476 | 44.7429 | 1.0886 | 32.0048 |

(2) Using function train () and function sim () to train and simulate. Let the training time is 5000. The final classified results see to Table 2. From Table 2 we can see that the character of sample (4) and (6) is very close. They all belong to thick-level mucosity bottom black soil. Sample (9) and (20) belong to the same kind.

Sample (3) and (5) are all folium mucosity bottom black soil, and have the similar character. If we classify all the samples to two groups, we can see that most of folium soil belongs to one group. The soil of thick and middle-level belong to another group. SOM networks can classify the soil according to the different networks

dimension. So, the dimension number determines the final results. But the total trend varied a little. From Table 2 we know that the result can reflect the practical complexion.

Table 2. Soil Classify Based on SOM Network (Trained 5000 times)

| Networks dimension | Numbers of classification | Serial number of sample x_i |
|--------------------|---------------------------|---|
| [1 1] | 1 | (1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21) |
| [2 1] | 2 | (2 4 6 7 10 11 12 13 15 16)(1 3 5 8 9 14 17 18 19 20 21) |
| [3 1] | 3 | (2 4 6 10 12)(7 11 13 15 16)(1 3 5 8 9 14 17 18 19 20 21) |
| [4 1] | 4 | (2 4 6 10 12)(7 11 13 15 16)(1 3 5 8 9 14 20)(17 18 19 21) |
| [5 1] | 5 | (2 4 6 10 12)(7 13 15)(11 16)(1 3 5 8 9 14 20)(17 18 19 21) |
| [6 1] | 6 | (2 4 6 10 12)(7 13 15)(11 16)(17 18)(19 21)(1 3 5 8 9 14 20) |
| [7 1] | 7 | (2 4 6 10)(12)(7 13 15)(11 16)(1 3 5 8 9 14 20)(19 21)(17 18) |
| [8 1] | 8 | (2 4 6 10)(12)(7 13 15)(11 16)(3 5 8)(1 9 14 20)(19 21)(17 18) |
| [9 1] | 9 | (2 4 6 10)(12)(7)(13 15)(11 16)(3 5 8)(1 9 14 20)(19 21)(17 18) |
| [10 1] | 10 | (2 4 6 10)(12)(7)(13 15)(11 16)(3 5 8)(1 14)(9 19 20)(21)(17 18) |
| [11 1] | 11 | (2 4 6 10)(12)(7)(13 15)(11 16)(3 5)(8)(1 14)(9 19 20)(21)(17 18) |
| [12 1] | 12 | (2 4 6 10)(12)(7)(13 15)(11 16)(3 5)(8)(1 14)(9 19 20)(21)(17 18) |
| [13 1] | 13 | (2 4 6)(10)(12)(7)(13)(15)(11 16)(3 5)(8)(1 14)(9 19 20)(21)(17 18) |
| [14 1] | 14 | (2)(4 6)(10)(12)(7)(13)(15)(11 16)(3 5)(8)(1 14)(9 19 20)(21)(17 18) |
| [15 1] | 15 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3 5)(8)(1 14)(9 19 20)(21)(17 18) |
| [16 1] | 16 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3 5)(8)(1 14)(9 19 20)(21)(17)(18) |
| [17 1] | 17 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3 5)(8)(1)(14)(9 19 20)(21)(17)(18) |
| [18 1] | 18 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3 5)(8)(1)(14)(9 20)(19)(21)(17)(18) |
| [19 1] | 19 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3)(5)(8)(1)(14)(9 20)(19)(21)(17)(18) |
| [20 1] | 20 | (2)(4 6)(10)(12)(7)(13)(15)(11)(16)(3)(5)(8)(1)(14)(9)(20)(19)(21)(17) (18) |
| [21 1] | 21 | (2)(4)(6)(10)(12)(7)(13)(15)(11)(16)(3)(5)(8)(1)(14)(9)(20)(19)(21)(17) (18) |

4. Fuzzy Clustering Method to Classify the Soil

We can apply fuzzy clustering method to classify the soil samples. The final result of classification is as follows.

If $\lambda = 1.0$, we can divide it into 21 kinds.

If $\lambda = 0.9810$, we can divide it into 20 kinds.

$\{x_4, x_6\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_9\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{19}\}, \{x_{20}\}, \{x_{21}\};$

If $\lambda = 0.9801$, we can divide it into 19 kinds.

$\{x_4, x_6\}, \{x_9, x_{20}\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_5\}, \{x_7\}, \{x_8\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{13}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{19}\}, \{x_{21}\};$

If $\lambda = 0.9719$, we can divide it into 18 kinds.

$\{x_4, x_6\}, \{x_9, x_{20}\}, \{x_5, x_{13}\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_7\}, \{x_8\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{19}\}, \{x_{21}\};$

If $\lambda = 0.9694$, we can divide it into 17 kinds.

$\{x_4, x_6\}, \{x_9, x_{20}\}, \{x_5, x_8, x_{13}\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{19}\},$

$\{x_{21}\}$;

If $\lambda=0.9693$, we can divide it into 16 kinds.

$\{x_4, x_6\}, \{x_9, x_{19}, x_{20}\}, \{x_5, x_8, x_{13}\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{14}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9674$, we can divide it into 15 kinds.

$\{x_4, x_6\}, \{x_9, x_{19}, x_{20}\}, \{x_5, x_8, x_{13}, x_{14}\}, \{x_1\}, \{x_2\}, \{x_3\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9665$, we can divide it into 14 kinds.

$\{x_4, x_6\}, \{x_9, x_{19}, x_{20}\}, \{x_3, x_5, x_8, x_{13}, x_{14}\}, \{x_1\}, \{x_2\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9659$, we can divide it into 13 kinds.

$\{x_4, x_6\}, \{x_1, x_9, x_{19}, x_{20}\}, \{x_3, x_5, x_8, x_{13}, x_{14}\}, \{x_2\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{16}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9635$, we can divide it into 12 kinds.

$\{x_4, x_6\}, \{x_1, x_9, x_{19}, x_{20}\}, \{x_3, x_5, x_8, x_{13}, x_{14}, x_{16}\}, \{x_2\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9576$, we can divide it into 11 kinds.

$\{x_2, x_4, x_6\}, \{x_1, x_9, x_{19}, x_{20}\}, \{x_3, x_5, x_8, x_{13}, x_{14}, x_{16}\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9550$, we can divide it into 10 kinds.

$\{x_2, x_4, x_6\}, \{x_1, x_3, x_5, x_8, x_9, x_{13}, x_{14}, x_{16}, x_{19}, x_{20}\}, \{x_7\}, \{x_{10}\}, \{x_{11}\}, \{x_{12}\}, \{x_{15}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9504$, we can divide it into 9 kinds.

$\{x_2, x_4, x_6\}, \{x_1, x_3, x_5, x_8, x_9, x_{13}, x_{14}, x_{16}, x_{19}, x_{20}\}, \{x_{10}, x_{12}\}, \{x_7\}, \{x_{11}\}, \{x_{15}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9470$, we can divide it into 8 kinds.

$\{x_2, x_4, x_6\}, \{x_1, x_3, x_5, x_8, x_9, x_{13}, x_{14}, x_{16}, x_{19}, x_{20}\}, \{x_7, x_{15}\}, \{x_{10}, x_{12}\}, \{x_{11}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9414$, we can divide it into 7 kinds.

$\{x_2, x_4, x_6\}, \{x_1, x_3, x_5, x_7, x_8, x_9, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}, \{x_{10}, x_{12}\}, \{x_{11}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9271$, we can divide it into 6 kinds.

$\{x_2, x_4, x_6, x_{10}, x_{12}\}, \{x_1, x_3, x_5, x_7, x_8, x_9, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}, \{x_{11}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9260$, we can divide it into 5 kinds.

$\{x_2, x_4, x_6, x_{10}, x_{12}\}, \{x_1, x_3, x_5, x_7, x_8, x_9, x_{11}, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.9109$, we can divide it into 4 kinds.

$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}, \{x_{17}\}, \{x_{18}\}, \{x_{21}\}$;

If $\lambda=0.8997$, we can divide it into 3 kinds.

$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}, \{x_{17}, x_{21}\}, \{x_{18}\}$;

If $\lambda=0.8712$, we can divide it into 2 kinds.

$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}\}, \{x_{18}\}$;

If $\lambda=0.7881$, we can divide it into 1 kinds.

$\{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}\}$.

The results of dynamic clustering see to Figure 3. The best threshold value λ can be calculated by statistical quantity F (Table 3). In Table 3, r is the number of the classifications.

From Table 3 we can see that there are 11 F to satisfy the inequation $F > F_{0.05}(r-1, n-r)$. Now, let's check $(F - F_\alpha)$. We can select the $(F - F_\alpha)$ that has rather big value as the main result of classification.

$\lambda=0.9470$ (8 kinds), $\lambda=0.9260$ (5 kinds). So, the best

soil classification is 5 kinds or 8 kinds.

4.3 Analyze the Result

When $\lambda=0.8712$, these soil samples can be divided into two kinds as follows.

$\left\{ \begin{array}{l} \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}\}, \\ \{x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}\} \\ \{x_{18}\} \end{array} \right\}$

Table 3. Calculating F to Obtain the Best Threshold Value λ in Soil Classify

| | | | | | | | | | | | |
|----------------------|--------|---------|---------|--------|--------|--------|--------|--------|--------|--------|--------|
| λ | 1.0000 | 0.9810 | 0.9801 | 0.9719 | 0.9694 | 0.9693 | 0.9674 | 0.9665 | 0.9659 | 0.9635 | 0.9576 |
| r | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 | 11 |
| F | 0 | 15.1063 | 17.2141 | 4.4997 | 4.4751 | 5.1556 | 4.9366 | 5.6952 | 6.6561 | 5.8316 | 6.9994 |
| $F_{0.05}(r-1, n-r)$ | | 247.5 | 19.4 | 8.68 | 5.84 | 4.62 | 3.96 | 3.55 | 3.28 | 3.10 | 2.98 |
| $(F - F_{\alpha})$ | --- | --- | --- | --- | --- | 0.5356 | 0.9766 | 2.1452 | 3.3761 | 2.7316 | 4.0194 |
| λ | 0.9550 | 0.9504 | 0.9470 | 0.9414 | 0.9271 | 0.9260 | 0.9109 | 0.8997 | 0.8712 | 0.7881 | |
| r | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 | 2 | 1 | |
| F | 5.4332 | 6.6171 | 8.1403 | 6.5647 | 7.3600 | 8.6611 | 1.6738 | 2.4192 | 2.398 | 0 | |
| $F_{0.05}(r-1, n-r)$ | 2.90 | 2.85 | 2.83 | 2.85 | 2.90 | 3.01 | 3.20 | 3.55 | 4.38 | | |
| $(F - F_{\alpha})$ | 2.5332 | 3.7671 | 5.3103 | 3.7147 | 4.46 | 5.6511 | --- | --- | --- | --- | |

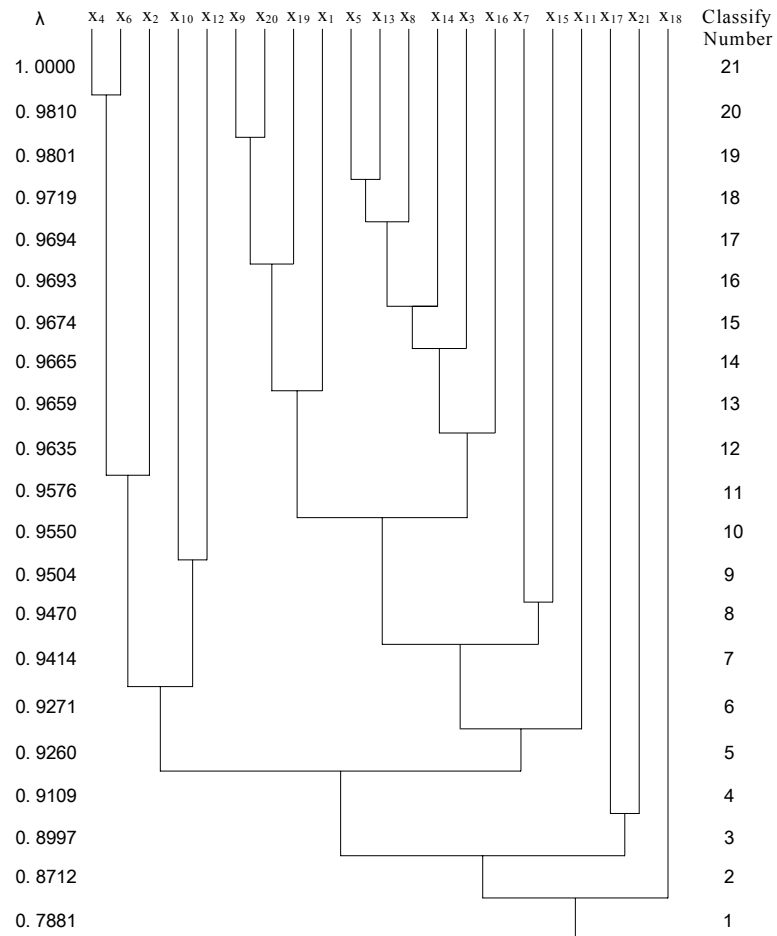


Figure 3. Fuzzy Dynamic Clustering of Soil

The sample 18 belongs to one kind itself. It shows that the indexes of total nitrogen, total phosphor, substitution quantity, organic matter and so on are all lower than other soil samples. So that, it is reasonable for that it has been divided into a single kind.

If $\lambda = 0.9260$, we can divide it into 5 kinds. One is $\{x_2, x_4, x_6, x_{10}, x_{12}\}$.

It shows that the character of most of thick soil is close.

Another is

$\{x_1, x_3, x_5, x_7, x_8, x_9, x_{11}, x_{13}, x_{14}, x_{15}, x_{16}, x_{19}, x_{20}\}$.

It shows that the character of most of folium soil is close.

The others are $\{x_{17}\}$, $\{x_{18}\}$ and $\{x_{21}\}$.

If $\lambda = 0.9470$, we can divide it into 8 kinds.

$\{x_2, x_4, x_6\}$, $\{x_1, x_3, x_5, x_8, x_9, x_{13}, x_{14}, x_{16}, x_{19}, x_{20}\}$, $\{x_7, x_{15}\}$, $\{x_{10}, x_{12}\}$, $\{x_{11}\}$, $\{x_{17}\}$, $\{x_{18}\}$, $\{x_{21}\}$.

It can reflect the same information as other classification. This information is accord with the practical complexion.

5. Discussions

According to Liu (1988), the fuzzy similarity matrix is built by distance method. The method used the distance information among every sample. The final result of the best classification is 12 kinds or 17 kinds. In this paper, the authors apply the method of correlation coefficient. The best classification is 5 kinds or 8 kinds. In SOM network, the training function Linkdist (pos) adopts the distance method also, and the final result is closer to Liu (1988). The different result shows that different methods to build up demarcated matrix can pickup different information in the original data. Thus, the result will have part difference. But the above three methods can reflect the practical complexion. They have the similar classified configuration.

On the other hand, it is very easy to use SOM

networks to classify the soil samples. SOM network can train, learn and compete by itself. It can give us the best-classified result according to the people's wish. Especially when we use the SOM-ANN toolbox in MATLAB 5.3 software, we can make the problem of soil classification much simply. At last, we should give the right result according to combine with practical complexion.

Fund Item: Chinese National "863" High-Technique Programme (No. 2002AA2Z4251-210041).

Correspondence to:

Qiang Fu

School of Water Conservancy & Civil Engineering

Northeast Agricultural University

Harbin, Heilongjiang 150030, China

Telephone: 01186-451-5519-0298

Cellular phone: 01186-13936246215

E-mail: fuqiang100@371.net

References

- Fu Q, Xie YG, Wei Z.M. 2003. Application of Projection Pursuit Evaluation Model Based on Real-Coded Accelerating Genetic Algorithm in Evaluating Wetland Soil Quality Variation in the Sanjiang Plain, China [J]. *Pedosphere* 2003;13(3):249-56.
- Liu X. Applying fuzzy clustering to soil classification. *Journal of Northeast Agricultural College* 1988(2):112-26.
- Liu ZG, Ma XH. Effect of reclamation on soil environment in Sanjiang Plain. *Pedosphere* 1997;7(1):73-8.
- Wang W. The principle of ANN—Introduction and Application. Publishing Company of Beijing Aviation Spaceflight University, Beijing, China. 1995:157-65.
- Wen X, Zhou L, Wang D. Design the ANN with MATLAB. Science Publishing Company, Beijing, China. 2000:271-85.
- Xie J, Liu C. Fuzzy Mathematics Method and Its Application. Middle China Technique Publishing Company, Wuhan, Hubei, China. 2000:81-106.
- Zhao QG. Resource and environmental quality changes and adjustment principles for sustainable development in rapidly developing coastal region of Southeastern China. *Pedosphere*. 2001;11(4):289-99.