CHINESE GEOGRAPHICAL SCIENCE
Volume 12, Number 2, pp. 136 – 141, 2002
Science Press, Beijing, China

APPLYING PPE MODEL BASED ON RAGA TO CLASSIFY AND EVALUATE SOIL GRADE

FU Qiang[1,2], LU Tie-guang[2], FU Hong[2]
(*1. College of Hydraulic Power, Sichuan University, Chengdu 610065, P. R. China;*
2. Northeast Agricultural University, Harbin 150030, P. R. China)

ABSTRACT: The research of soil classification and soil grade evaluation is often based on fuzzy theory. So, the traditional method has an inevitable problem about weight matrix which given by some experts, and the final result can be influenced by artificial factors. The essentials of fuzzy synthetically judge is to handle the data of high dimension. That is to reducing the dimension number. The weight matrix in fuzzy theory is corresponding to low dimension projection value of each index. But we can't define whether the weight matrix given by experts is the best projection value or not. So, the authors apply a new technique of falling dimension named projection pursuit to soil study, through using the improved real coding based accelerating genetic algorithm to optimize the projection direction. Thus, it can transfer multi-dimension data into one dimension data, through searching for the optimum projection direction to realize the soil classification and its grade evaluation. The method can avoid the artificial disturbance, and acquire preferably effect. Thus, the paper provides a new method to the research of soil classification and grade evaluation.

KEY WORDS: RAGA; PPE; soil classification; soil evaluation

CLC number: S155 Document code: A Article ID: 1002-0063(2002)02-0136-06

1 INTRODUCTION

Soil classification is not only the basis of soil science, but also the synthetic symbol of the developed level of soil science(HU and LIU, 1999). The soil analyzing system studied by human in different periods can reflect the understanding level and soil science seedtime of this period. Furthermore, soil quality can reflect synthetically the soil characteristic. It is the most sensitive index to post the dynamic movement of soil condition. And soil quality can incarnate the influence caused by human function(HU and LIU, 1999; LIU, 1988). Soil erosion is very severe currently in China. Human being has destroyed more and more soil resources. Soil quality has degenerated seriously and isn't fit for cultivation in some areas. So, through studying the movement of soil quality and soil classification under the influence of human being, we can not only provide a theoretical method to describe the change trend of soil quality, but also post the influence direction of soil environment caused by human being. Furthermore, it can prevent soil erosion, improve soil, use limited resources, and advance the agriculture and society to develop continuously.

The essential of soil quality evaluation and classification is a synthetically evaluation about different soil sample books according to different indexes. Because there are several evaluation indexes, the evaluation and classification is to deal with high-dimension data. The traditional method adopts weight judgement, which is to give a weight to each index to make fuzzy synthetically judgement. During the course of this method, how to choosing the weight is difficult. Thereby, the different weights can induce different results.

Recently, an effective technique to fall dimension named projection pursuit (PP) has been raised. The method can put the dot in the high-dimension data space into low-dimension sub-space through projection. Because there are many projection directions during the course of mapping from high-dimension space to low-dimension space, the projection direction that can make the projection index function reaching the maximal

Received date: 2001-12-24
Fundation item: Under the auspices of China Postdoctoral Science Fund and the Youth Fund of Sichuan University(432028).
Biography: FU Qiang(1973 –), male, a native of Jinzhou City, Liaoning Province, associate professor, postdoctor in Sichuan University. His research interests include building up mathematical models and its optimization about water saving irrigation, agricultural water resources and agricultural system engineering.

value is the best projection direction. Many indexes should been optimized at the same time. It will be very difficult to optimize these indexes at the same time if we adopt traditional method. Now, another new optimization method named genetic algorithm (GA) has risen. The GA method is fit for the problem about multi-dimension and can convergence overall situation (JIN and DING, 2000; XIANG and SHI, 2000; ZHANG, 2000). The authors put forward a new method named real coding based accelerating genetic algorithm (RAGA). Through combining the RAGA with PPE (Projection Pursuit Evaluation) model, the authors can use RAGA to optimize the parameters in PPE model at the same time. The result can convert several indexes into one synthetical index. Then, according to the projection value, the stylebooks can be identified. Thus, we can realize the evaluation and classification of soil. The authors provide a new method for soil quality evaluation and classification.

2 PROJECTION PURSUIT EVALUATION MODEL (PPE)

2.1 Brief Introduction of PP Model

The main characteristics of PP model are as follows. Firstly, PP model can handle the difficulty named dimension disaster, which has been brought by high-dimension data. Secondly, PP model can eliminate the jamming, which is irrespective with data structure. Thirdly, PP model provides a new approach to handle high-dimension problem using one dimension statistics method. Fourthly, PP method can deal with non-linearity problem (JIN and DING, 2000; XIANG and SHI, 2000; ZHANG, 2000).

2.2 Step of PPE Modeling

To building up PPE model includes 4 steps as follows (JIN and DING, 2000; XIANG and SHI, 2000; ZHANG, 2000).

Step 1: Normalizing the evaluation indexes set of each stylebook. Now, we suppose the stylebook set is $\{x^*(i, j) \mid i=1-n, j=1-p\}$. Where $x^*(i, j)$ is the index value of j and stylebook of i. n, the number of stylebook. p, the number of index. In order to eliminate the dimension influence and unite the change scope of each index value, we can adopt the following formulas to normalize the data.

$$x(i,j) = \frac{x^*(i,j) - x_{\min}(j)}{x_{\max}(j) - x_{\min}(j)} \quad \text{(1-a)}$$

or

$$x(i,j) = \frac{x_{\max}(j) - x^*(i,j)}{x_{\max}(j) - x_{\min}(j)} \quad \text{(1-b)}$$

where $x_{\max}(j)$ and $x_{\min}(j)$ stand for the maximun and the minimum of j index value. $x(i, j)$ is the index list after moralization.

Step 2: Constructing the projection index function $Q(a)$. PP method is to turn p dimension data ($\{x^*(i, j) \mid j=1-p\}$) into one dimension projection value $z(i)$ based on projection direction a.

$$a = \{a(1), a(2), a(3), \ldots, a(p)\},$$

$$z(i) = \sum_{j=1}^{p} a(j)\, x(i,j) \quad (i=1-n) \quad (2)$$

Then, we can classify the stylebook according to one-dimension scatter figure of $z(i)$. In formula (2), a stands for unit length vector.

Thus, the projection index function can be expressed as follows.

$$Q(a) = S_z D_z \quad (3)$$

where S_z is the standard deviation of $z(i)$, D_z is the partial density of $z(i)$.

$$S_z = \sqrt{\frac{\sum_{i=1}^{n} (z(i) - E(z))^2}{n-1}} \quad (4)$$

$$D_z = \sum_{i=1}^{n} \sum_{j=1}^{n} (R - r(i,j)) \cdot u(R - r(i,j)) \quad (5)$$

In formula (4) and (5), $E(z)$ is the average value of series $\{z(i) \mid i=1-n\}$; R is the window radius of partial density, commonly, $R=0.1S_z$; $r(i, j)$ is the distance of stylebook, $r(i, j)=|z(i)-z(j)|$; $u(t)$ is a unit jump function, if $t \geq 0$, $u(t)=1$, if $t<0$, $u(t)=0$.

Step 3: Optimizing the projection index function. When every indexes valus of each stylebook have been fixed, the projection function $Q(a)$ change only according to projection direction a. Different projection directions reflect different data structure characteristic. The best projection direction is the most likely to discovery some characteristic structure of high-dimension data. So, we can calculate the maximum to estimate the best project direction.

Function: Max: $Q(a) = S_z \cdot D_z$ (6)

Restricted condition: s. t: $\sum_{j=1}^{p} a^2(j) = 1$ (7)

Formula (6) and (7) is a complex non-linearity optimization, which take $\{a(j) \mid j=1-p\}$ as optimized variable. Traditional method is very difficulty to calculate. Now, we adopt RAGA to handle the kind of problem.

Step 4: Classification. We can put the best projection direction a^* into formula (2), and then we can obtain the projection value of each stylebook dot. Compare $z^*(i)$ with $z^*(j)$. If $z^*(i)$ is closer to $z^*(j)$, that means stylebook i and j are trend to the same species. If we dispose $z^*(i)$ from big to small, we can obtain the new stylebook list from good to bad.

3 REAL CODING BASED ACCELERATING GENETIC ALGORITHM

3. 1 Brief Introduction of GA

Genetic Algorithm has been put forward by Professor Holland in USA. The main operation includes selection, crossover and mutation (JIN and DING, 2000; ZHOU and SUN, 2000).

3. 2 Real Coding Based Accelerating Genetic Algorithm (RAGA)

The coding mode of traditional GA adopted binary system. But binary system coding mode has many abuses. So, through consulting literature, the authors put forward a new method named RAGA (Real Coding Based Accelerating Genetic Algorithm). RAGA includes 8 steps as follows. For example, we want to calculate the following best optimization problem.

$$\text{Max: } f(X)$$
$$s.t.: a_j \leqslant x_j \leqslant b_j$$

Step1: In the scope of $[a_j, b_j]$, we can create N group uniformity distributing random variable $V_i(0)$ $(x_1, x_2, \ldots, x_j, \ldots, x_p)$. $i = 1 - N$, $j = 1 - p$. N is the group scale. p is the number of optimized parameter.

Step 2: Calculate the target function value. Putting the original chromosome $V^{(0)}$ into target function, we can calculate the corresponding function value $f^{(0)}(V^{(0)}{}_i)$. According to the function value, we dispose the chromosome from big to small. Then, we obtain $V^{(1)}$.

Step 3: Calculate the evaluation function based on order expresses as eval(V). The evaluation function gives a probability for each chromosome. It makes the probability of the chromosome to be selected is fit for the adaptability of other chromosomes. When the adaptability of chromosome is better, it is much easy to be selected. Now, if parameter $a \in (0, 1)$, the evaluation function based order can be expresses as follows.

$$\text{eval}(V_i) = a(1-a)^{i-1}, \quad i = 1, 2, \ldots, N \qquad (8)$$

Step 4: Selecting operation. The course of selecting is based on circumrotating the bet wheel N times. We can select a new chromosome from each rotation. The bet wheel selects the chromosome according to the adaptability. We obtain a new group $V_i^{(1)}$ after selecting.

Step 5: Crossover operation. Firstly, we define the parameter P_c as the crossover probability. In order to ensure the parent generation group to crossover, we can repeat the process from $i = 1$ to N as follows. Create random number from $[0, 1]$. If $r < P_c$, we take V_i as parent generation. We use V'_1 V'_2, ... to stand for male parent to be selected. At the same time, we divide the chromosome into random pair based on arithmetic crossing method. That is as follows.

$$\begin{aligned} X &= c \cdot V'_1 + (1-c) \cdot V'_2 \\ Y &= (1-c) \cdot V'_1 + c \cdot V'_2 \end{aligned} \qquad (9)$$

where c is a random number from $(0, 1)$.

We can obtain a new group $V_i^{(2)}$ after crossover.

Step 6: Mutation operation. Define the P_m as mutation probability. We select the mutation direction d randomly from R^n. If $V + M_d$ isn't feasible, we can make M a random number from 0 to M until the value of $V + M_d$ is feasible. M is a enough big number. Then, we can use $X = V + M_d$ replace V. After mutation operation, we obtain a new group $V_i^{(4)}$.

Step 7: Evolution iteration. We can obtain the filial generation $V_i^{(4)}$ from step 4 to step 6, and dispose them according to adaptability function value from big to small. Then, the arithmetic comes into the next evolution process. Thus, the above steps have been operated repeatedly until the end.

Step 8: The above seven steps make up of Standard Genetic Arithmetic (SGA). But SGA can't assure the whole astringency. The research indicates that the seeking optimization function of selecting and crossover has wear off along with the iteration times increasing. In practical application, SGA will stop to working when it is far away from the best value, and many individuals are conforming or repeated. Enlightened from LIU (1988), we can adopt the interval of excellence individual during the course of the first and the second iteration as the new interval. Then, the arithmetic comes into step 1, and runs SGA over again to form accelerate running. Thus, the interval of excellence individual will gradually reduce, and the distance is closer to the best dot. The arithmetic will not stop until the function value of best individual less than a certain value or exceed the destined accelerating times. At this time, the current group will be destined for the result of RAGA.

The above 8 steps make up of RAGA.

3. 3 PPE Model Based on RAGA

Take projection function $Q(a)$ as the most target function in the PPE model and the projection $a(j)$ of each index as optimized variable. Through running the 8 steps of RAGA, we can obtain the best projection direction $a^*(j)$ and projection value $z(i)$. Comparing the $z(i)$ each other, we can obtain the evaluated result. At the same time, if we build PPE model about the soil grade evaluation standard according to the above steps, we obtain the best projection value $Z(i)$. Then, through comparing the distance between $z(i)$ and $Z(i)$, the smallest distance between any two stylebooks, then, the number is the soil stylebook grade.

4 APPLICATION EXAMPLE

Now, we select the grade evaluation standard and take the soil grade as stylebooks(five grades.) The evaluated indexes are total nitrogen, total phosphor, organic matter, pH value, cultivation depth, physical clay grain content and so on. Then, we can build up PPE model of standard stylebook according to synthetic characters.

First, we build up PPE model according to the data of Table 1. There are 5 soil grade standards (stylebook). Each stylebook has 6 evaluated indexes. That means the problem belongs to high-dimension data.

During the course of RAGA, the parent generation scale is 400($n = 400$). The crossover probability is 0.80($p_c = 0.80$). The mutation probability is 0.80($p_m = 0.80$). The number of excellence individual is 20. $\alpha = 0.05$. Through accelerating 8 times, we can obtain the best projection value. That is 0.2769. The best projection direction: $a^* =$ (0.4037, 0.4894, 0.4708, 0.1979, 0.4619, 0.3512). Putting a^* into formula (2), we can obtain the projection value of each soil quality standard stylebook. That are $Z_1^*(j) =$ (2.3749, 1.7717, 1.2415, 0.8195, 0.5171). Then, the PPE model of soil nutrient grade evaluation is as follows (Fig. 1).

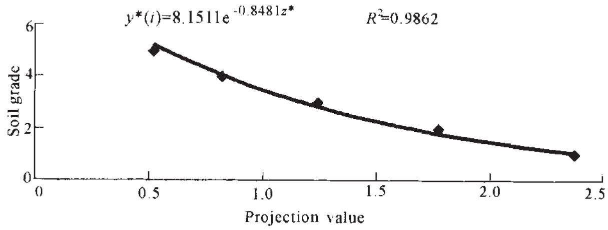

Fig. 1 The relation between projection value of every stylebooks and soil grade number

Table 1 Grade table of soil nutrient

Grade	Total nitrogen (%)	Total phosphor (%)	Organic matter (%)	pH value	Cultivation depth (cm)	Physical clay grain content (%)
Ⅰ	> 0.2	> 0.2	> 4	6.5 – 7.5	> 40	> 60
Ⅱ	0.15 – 0.2	0.15 – 0.2	3 – 4	5.5 – 6.5	25 – 40	50 – 60
Ⅲ	0.1 – 0.15	0.1 – 0.15	2 – 3	4.5 – 5.5	15 – 25	40 – 50
Ⅳ	0.075 – 0.1	0.05 – 0.1	1 – 2	<4.5	10 – 15	30 – 40
Ⅴ	0.05 – 0.075	<0.05	0.6 – 1	<4.5	<10	<30

We know that the precision of RAGA-PPE model is high from Table 2. Through using the above PPE model, we can calculate the projection value of every practical stylebook firstly. Then, putting the projection value into RAGA-PPE model, we can obtain the calculated value of soil grade. Through judging the distance between stylebook and grade standard, we can determine the grade of each soil stylebook. The indexes of every practical stylebook are listed in Table 3.

Now, we build up PPE model according to Table 3. The best projection value of each stylebook is: $z^*(j) =$ (1.3478, 1.3661, 1.0237, 1.3987, 1.1334, 1.4173, 1.5672, 1.4022, 1.5639, 1.4050, 1.3994, 1.4679, 1.6456, 1.3210, 1.9237, 1.7309, 1.0526, 0.7783, 1.5329, 1.4249, 0.9636). At last, through comparing the distance between $z^*(j)$ and $Z^*(j)$, we can determine the soil grade(Table 4).

In Table 4, we know that the number 15 is the best

Table 2 The result and error-analyzing table of RAGA-PPE model

Grade experiential value	Stylebook projection value	Grade calculated value(y^*)	Absolute error	Relatively error (%)
1	2.3749	1.0876	0.0876	8.76
2	1.7717	1.8141	−0.1859	−9.29
3	1.2415	2.8441	−0.1559	−5.19
4	0.8195	4.0679	0.0679	1.70
5	0.5171	5.2572	0.2572	5.14
Average			0.1509 (absolute value)	6.01 (absolute value)

Table 3 Soil stylebooks and its character indexes

Stylebook number	Soil name	Soil characters: Total nitrogen (%)	Total phosphor (%)	Organic matter (%)	pH	Cultivated depth (cm)	Physical clay content (%)
1	Folium mucosity bottom white slurry black soil	0.270	0.142	6.46	5.5	21	45.3
2	Thick-level mucosity bottom black soil	0.171	0.115	3.46	6.3	60	45.3
3	Folium mucosity bottom black soil	0.114	0.101	2.43	6.4	25	51.0
4	Thick-level mucosity bottom black soil	0.173	0.123	3.30	5.8	65	45.6
5	Folium mucosity bottom black soil	0.145	0.131	3.28	6.0	25	51.0
6	Thick-level black meadow	0.173	0.140	3.45	5.8	60	49.0
7	Middle-level black meadow	0.250	0.177	5.51	7.2	45	46.6
8	Folium black meadow soil	0.237	0.189	5.37	6.1	27	45.0
9	Folium gully black meadow soil	0.319	0.227	7.04	5.8	24	39.3
10	Thick-level flat meadow soil	0.163	0.124	3.73	6.2	61	48.1
11	Middle-level flat meadow soil	0.194	0.201	4.50	5.7	35	47.4
12	Thick-level gully meadow soil	0.142	0.185	3.79	6.4	55	51.0
13	Thick-level gully latent raised meadow soil	0.240	0.217	4.92	6.5	41	63.6
14	Folium gully latent raised meadow soil	0.253	0.172	4.63	6.8	20	44.1
15	Thick-level plat carbonate meadow soil	0.357	0.289	7.21	7.5	40	48.3
16	Middle-level gully carbonate meadow soil	0.280	0.204	10.68	6.7	31	41.5
17	Middle-level meadow	0.164	0.141	3.05	4.8	30	30.9
18	Arenaceous meadow soil	0.095	0.099	1.51	6.0	20	26.4
19	Eroded dark and brown soil	0.392	0.240	6.62	5.3	14	34.7
20	Folium white slurry soil	0.267	0.208	6.25	5.8	19	42.4
21	Yellow and white slurry soil	0.137	0.111	3.04	5.1	18	43.1

Table 4 The result of RAGA-PPE model and the grade of soil stylebook

Serial number	Projection value	Arrange from big to small	Stylebook number after arrangement	Calculated value (y^*)	Grade
1	1.3478	1.9237	15	1.5947	Ⅱ
2	1.3661	1.7309	16	1.8779	Ⅱ
3	1.0237	1.6456	13	2.0188	Ⅱ
4	1.3987	1.5672	7	2.1576	Ⅱ
5	1.1334	1.5639	9	2.1637	Ⅱ
6	1.4173	1.5329	19	2.2213	Ⅱ
7	1.5672	1.4679	12	2.3472	Ⅱ
8	1.4022	1.4249	20	2.4344	Ⅱ
9	1.5639	1.4173	6	2.4501	Ⅱ
10	1.4050	1.4050	10	2.4758	Ⅱ
11	1.3994	1.4022	8	2.4817	Ⅱ
12	1.4679	1.3994	11	2.4876	Ⅱ
13	1.6456	1.3987	4	2.4891	Ⅱ
14	1.3210	1.3664	2	2.5589	Ⅲ
15	1.9237	1.3478	1	2.5989	Ⅲ
16	1.7309	1.3210	14	2.6586	Ⅲ
17	1.0526	1.1334	5	3.1171	Ⅲ
18	0.7783	1.0526	17	3.3633	Ⅲ
19	1.5329	1.0237	3	3.4211	Ⅲ
20	1.4249	0.9636	21	3.6000	Ⅳ
21	0.9636	0.7783	18	4.2126	Ⅳ

stylebook. That means the total nitrogen and total phosphor account for primacy. At the same time, organic matter, pH value and any other indexes are all rather high. So, the fifth stylebook is the best soil style. On the contrary, every indexes of the eighteenth stylebook are rather low, so it is the worst soil style. The other stylebooks can be analogized in turn. In addition, the projection values of number 4, 6, 8, 10, 11 and 20 are quite approximate. That means the synthetic characters of these six stylebooks are quite approximate through evaluating by the above six indexes. Number 7 and number 9 have the same synthetic characters also. The stylebook will be listed in the former if the projection value is high. That means the grade value is low, the grade is high, and the synthetic quality of soil nutrient is good. Thus, we can divide these 21 stylebooks into different grades according to PPE model. It is obvious that these soil stylebooks that have the closer synthetic characters will be the same kind. At the same time, through evaluating and classifying for soil, we know that the soil synthetic quality of this area belongs to grade Ⅱ, grade Ⅲ and grade Ⅳ basically.

5 CONCLUSION

(1) Through applying PPE model, the authors build up the PPE model of soil quality evaluation and classification. Several evaluation indexes of soil quality have been taken as multi-dimension projection parameters to seeking the best projection direction. The best projection index function value can reflect the quality of each soil stylebook good or bad. Thus, we can avoid the disturbance by artificial factor to endow weight. The result is good.

(2) The author improves on SGA, and put forward a new method named RAGA. Through reducing the interval of excellence individual to accomplish the accelerate process. Thus, the method of RAGA can realize quick convergence and seeking the best result in the whole scope.

(3) Combing RAGA with PPE model, through using RAGA to optimizing the many parameters in the PPE model, we can obtain the best projection direction of evaluation index of each soil quality stylebook. Thus, the process of PPE modeling has been predigested. And the PPE model can be used in many other fields.

(4) Through applying RAGA-PPE model to classify and evaluate the soil stylebooks, we know that most of the soil stylebooks belong to grade Ⅱ, grade Ⅲ and grade Ⅳ. The fifth stylebook is the bestsoil, and the total nitrogen and total phosphor account for primacy. Any other indexes, such as organic matter, pH value and so on, are ratter high. The eighteenth stylebook is the worst soil kind.

REFERENCES

HU Jin-ming, LIU Xing-tu, 1999. Evaluate and analyze the soil quality movement in Sanjiang Plain [J]. *Scientia Geographica Sinica*, 19(5): 417 – 421. (in Chinese)

JIN Ju-liang, DING Jing, 2000. *Genetic Arithmetic and Its Application to Water Science*[M]. Chengdu: Sichuan University Publishing Company: 42 – 47. (in Chinese)

LIU Xing-jiu, 1988. Applying fuzzy clustering to soil classification [J]. *Journal of Northeast Agricultural College*, (2): 122 – 126. (in Chinese)

XIANG Jing-tian, SHI Jiu-en, 2000. *The Statistics Method of Data Processing in the Non-linearity System* [M]. Beijing: Science Press, 172, 190 – 192. (in Chinese)

ZHANG Xin-li, 2000. *Projection Pursuit ant Its Application to Water Resources* [D]. Chengdu, 67 – 73. (in Chinese)

ZHOU Ming, SUN Shu-dong, 2000. *The Theory of Genetic Arithmetic and Its Application*[M]. Beijing: National Defense Industry Publishing Company, 4 – 7, 37 – 38. (in Chinese)