

基于 RAGA 的 PPE 模型在土壤质量等级评价中的应用研究

付强¹, 金菊良², 门宝辉³, 梁川³

(1. 东北农业大学 水利与建筑学院, 哈尔滨 150030; 2. 合肥工业大学 土建学院, 安徽 合肥 230009; 3. 四川大学 水电学院, 四川 成都 610065)

摘要: 以往的土壤分类与等级综合评价多是建立在模糊数学 (Fuzzy) 基础上的模糊聚类分析与模糊综合评判, 因此不可避免的涉及权重矩阵人为干扰, 导致分类与评价结果的人为倾向。本文将高维降维技术——投影寻踪评价模型 (PPE) 应用到土壤学科领域, 利用改进的加速遗传算法 (RAGA) 优化投影方向, 将多维数据指标转换到低维子空间, 通过寻求最优投影方向及投影函数值来实现对土壤的分类与等级评价, 避免了主观赋权的人为干扰, 取得了较好效果, 可为土壤分类与等级评价研究提供一条新思路。

关键词: RAGA; PPE; 土壤分类; 评价

文献标识码: B

文章编号: 1000-288X(2002)05-0051-04

中图分类号: S155.1

Applying RAGA-based PPE Model to Classify and Evaluate Soil Grade

FU Qiang¹, JIN Ju-liang², MEN Bao-hui³, LIANG Chuan³

(1. College of Water Conservancy and Civil Engineering, Northeast Agricultural University, Harbin 150030, China;
2. College of Hydraulic Power, Sichuan University, Chengdu 60065, China;
3. College of Architecture, Hefei Industry University, Hefei 230009, 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. The finally result can be influenced by artificial factor. A new technique of falling dimension named projection pursuit is applied 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 disturb, and acquire preferably effect. Thus, it provides a new method to the research of soil classification and grade evaluation.

Keywords RAGA; PPE; soil classification; evaluation

1 引言

土壤分类既是土壤科学的基础,也是土壤科学发展水平的综合标志^[1]。不同时期所拟定的土壤分析系统反映了该时期人们对土壤的认识水平和土壤科学本身的发展阶段。土壤质量是土壤特性的综合反映,也是揭示土壤条件动态的最敏感的指标,而且能体现人类活动对土壤的影响^[2,3]。对一个地区的土壤作出适宜的分类与评价,对于发展该地区的农业生产、改良土壤、防治水土流失、揭示人类活动对土壤环境的影响均具有重要的意义。

以往的土壤分类与等级评价多是建立在模糊数

学基础上的模糊聚类分析与模糊综合评判,因此不可避免的涉及专家对每个土壤指标赋予相应的权重。不同专家可能赋予不同的权重,因此人为干扰较大。近年来,发展了一项新兴而又有效的降维技术——投影寻踪 (Projection Pursuit, 简称 PP), 它可以将高维数据空间上的点通过投影映射到低维子空间上。由于高维空间上的点映射到低维子空间上可以有不同的投影方向, 其中使投影指标函数达到最大的投影方向即为我们所求的最佳投影方向。由于评价指标较多 (例如本例中涉及 6 个), 因此就具有多个需要优化的参数。传统的优化方法同时优化这么多参数相当困难。近年来得到普遍发展的适于多维、全局收敛的优化方

收稿日期: 2002-01-02

资助项目: 中国博士后科学基金 (2000) 与四川大学校青年基金资助项目 (432028)

作者简介: 付强 (1973-), 男 (汉族), 辽宁锦州, 博士后, 副教授。主要从事农业水资源、节水灌溉及农业系统工程建模与优化研究。

电话 (0451) 5391502, E-mail fuqiang@mail.neau.edu.cn 或 fuqiang100@371.net

法——遗传算法 (Genetic Algorithm 简称 GA) 却能很好的解决这个问题。本文将改进的 GA 方法——基于实数编码的加速遗传算法 (Real coding based Accelerating Genetic Algorithm, 简称 RAGA) 与投影寻踪分类模型 (Projection Pursuit Evaluation Model, 简称 PPE 模型) 有机的结合^[4], 通过 RAGA 优化 PPE 模型中的投影方向参数, 完成高维数据向低维空间的转换, 即将每个样本的多个评价指标综合成一个综合指标, 然后按投影值进行排序与识别, 从而实现土壤样本的分类与评价。为土壤分类与评价研究提供一条新思路。

2 投影寻踪评价模型建立^[3, 4]

(1) 样本评价指标集的归一化处理。设各指标值的样本集为 $\{x^*(i, j) | i = 1 \sim n, j = 1 \sim p\}$, 其中 $x^*(i, j)$ 为第 i 个样本第 j 个指标值, n, p 分别为样本的个数和指标的数目。为消除各指标值的量纲和统一各指标值的变化范围, 可采用下式进行极值归一化处理: 例如对于越大越优的指标:

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

式中: $x_{\max}(j)$ ——第 j 个指标值的最大值; $x(i, j)$ ——指标特征值归一化的序列。

(2) 构造投影指标函数 $Q(a)$ 。PP 方法就是把 p 维数据 $\{x^*(i, j) | j = 1 \sim p\}$ 综合成以 $a = \{a(1), a(2), a(3), \dots, a(p)\}$ 为投影方向的一维投影值

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

然后根据 $\{z(i) | i = 1 \sim n\}$ 的一维散布图进行分类。式 (2) 中 a 为单位长度向量。综合投影指标值时, 要求投影值 $z(i)$ 的散布特征应为: 局部投影点尽可能密集, 最好凝聚成若干个点团; 而在整体上投影点团之间尽可能散开。因此, 投影指标函数可以表达成:

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

式中: S_z ——投影值 $z(i)$ 的标准差; D_z ——投影值 $z(i)$ 的局部密度, 即:

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

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

式中: $E(z)$ ——序列 $\{z(i) | i = 1 \sim n\}$ 的平均值; R ——局部密度的窗口半径, R 可以根据试验来确定, 一般可取值为 $0.1S_z$; $r(i, j)$ 表示样本之间的距离, $r(i, j) = |z(i) - z(j)|$; $u(t)$ 为一单位阶跃函数,

当 $t \geq 0$ 时, 其值为 1, 当 $t < 0$ 时其函数值为 0

(3) 优化投影指标函数。当各指标值的样本集定时, 投影指标函数 $Q(a)$ 只随着投影方向的变化而变化。不同的投影方向反映不同的数据结构特征, 最佳投影方向就是最大可能暴露高维数据某类特征结构的投影方向, 因此可以通过求解投影指标函数最大化问题来估计最佳投影方向, 即

最大化目标函数:

$$\text{Max } Q(a) = S_z \cdot D_z \quad (6)$$

约束条件:

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

这是一个以 $\{a(j) | j = 1 \sim p\}$ 为优化变量的复杂非线性优化问题, 用传统的优化方法处理较难。因此, 本文应用模拟生物优胜劣汰与群体内部染色体信息交换机制的基于实数编码的加速遗传算法 (RAGA) 来解决其高维全局寻优问题。

3 基于实数编码的加速遗传算法^[3-7]

遗传算法由美国密执安大学的 Holland 教授提出的, 是模拟生物在自然环境中的遗传和进化过程而形成的一种自适应全局优化概率搜索算法。主要包括选择 (selection)、交叉 (crossover) 和变异 (mutation) 等操作。例如求解如下最优化问题:

$$\text{Max: } f(X)$$

$$s. t. a_j \leq x_j \leq b_j$$

(1) 在各个决策变量的取值变化区间随机生成组均匀分布的随机变量; (2) 计算目标函数值, 从大到小排列; (3) 计算基于序的评价函数 (用 $eval(V)$ 表示); (4) 进行选择操作, 产生新的种群; (5) 对步骤 4 产生的新种群进行交叉操作; (6) 对步骤 5 产生的新种群进行变异操作; (7) 进化迭代; (8) 上述 7 个步骤构成标准遗传算法 (SGA)

由于 SGA 不能保证全局收敛性, 在实际应用中常出现在远离全局最优点的地方 SGA 即停滞寻优工作。为此, 可以采用第一次、第二次进化迭代所产生的优秀个体的变量变化区间作为变量新的初始变化区间, 算法进入步骤 1, 重新运行 SGA, 形成加速运行, 则优秀个体区间将逐渐缩小, 与最优点的距离越来越近。直到最优个体的优化准则函数值小于某一设定值或算法运行达到预定加速次数, 结束整个算法运行。此时, 将当前群体中最佳个体指定为 RAGA 的结果。上述 8 个步骤构成基于实码的加速遗传算法 (RAGA)^[3, 4]。

4 基于 RAGA 的 PPE 模型分类评价

将 PPE 模型中投影指标函数 $Q(a)$ 求最大作为目标函数, 各个指标的投影 $a(j)$ 作为优化变量, 运行 RAGA 上述 8 个步骤, 即可求得最佳投影方向 $a^*(j)$ 及相应的投影值 $z(i)$, 从而求得分类结果。同时将土壤等级评价标准按照上述步骤建立 PPE 模型, 得出最佳投影方向下的投影值 $Z(i)$, 比较 $z(i)$ 与 $Z(i)$ 之间的距离, 距离最近的即为土壤样本的归属级。

5 在土壤质量等级评价中的应用

先选定土壤等级评价标准, 将等级作为样本 (5 级)。本文选定全氮、全磷、有机质、pH 值、耕层厚、物理黏粒含量 6 项作为评价指标, 则可以将其按照综合性状建立标准样本的 PPE 模型, 见表 1^[1,10]。

表 1 土壤养分评级

| 等级 | 全氮 % | 全磷 % | 有机质 % | pH 值 | 耕层厚 /cm | 物理黏粒含量 % |
|-----|------------|------------|------------|----------|---------|----------|
| I | > 0.20 | > 0.20 | > 4.00 | 6.5~ 7.5 | > 40 | > 60 |
| II | 0.15~ 0.20 | 0.15~ 0.20 | 3.00~ 4.00 | 5.5~ 6.5 | 25~ 40 | 50~ 60 |
| III | 0.10~ 0.15 | 0.10~ 0.15 | 2.00~ 3.00 | 4.5~ 5.5 | 15~ 25 | 40~ 50 |
| IV | 0.08~ 0.10 | 0.05~ 0.10 | 1.00~ 2.00 | < 4.5 | 10~ 15 | 30~ 40 |
| V | 0.05~ 0.08 | < 0.05 | 0.60~ 1.00 | < 4.5 | < 10 | < 30 |

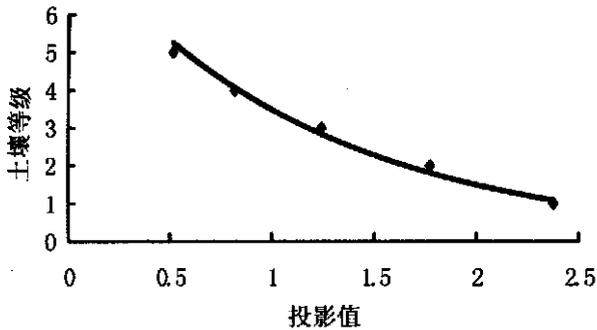


图 1 各样本综合投影值与土壤等级序号关系

表 2 土壤养分等级标准 RAGA-PPE 模型计算及误差

| 等级 经验值 | 样本 投影值 | 等级计算 值 y^* | 绝对 误差 | 相对误 差 % |
|-----------|-----------|-----------------|-----------------|---------------|
| 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 |
| 平均 | | | 0.1509 (绝对值) | 6.01 (绝对值) |

利用上述评价模型, 可以先计算各个实际样本的投影值, 然后将投影值代入 RAGA-PPE 模型, 求得土壤等级计算值, 再判断样本与标准等级之间的距离, 从而判定该土壤样本的归属级。各实测样本各项

首先, 对表 3 中数据建立 PPE 模型, 表 3 中土壤样本等级标准为 5 个 (5 个样本), 每个样本有 6 个评价指标, 即属于 6 维数据。RAGA 优化过程中选定父代初始种群规模为 $n = 400$, 交叉概率 $p_c = 0.80$, 变异概率 $p_m = 0.80$, 优秀个体数目选定为 20 个, $T = 0.05$, 加速次数为 11, 得出最大投影指标值为: 0.2769, 最佳投影方向 $\hat{T} = (0.4037, 0.4894, 0.4708, 0.1979, 0.4619, 0.3512)$, 将 \hat{T} 代入式 (2) 后即得各个土壤标准的投影值 $Z^*(j) = (2.3749, 1.7717, 1.2415, 0.8195, 0.5171)$, 建立土壤养分等级评价的投影寻踪模型 (PPE): $y^*(i) = 8.1511 e^{-0.8481z^*}$, $R^2 = 0.9682$ (图 1)。从表 2 土壤养分等级标准 RAGA-PPE 模型计算及误差可看出, 土壤质量等级标准 RAGA-PPE 模型的精度较高。

指标详见表 3

对表 3 中相应的数据建立 PPE 模型, 得出各个土壤样本投影值 $z^*(j) = (1.3478, 1.3661, 1.0237, 1.3987, 1.1334, 1.4173, 1.5672, 1.4022, 1.5639, 1.4050, 1.3994, 1.4679, 1.6456, 1.3210, 1.9237, 1.7309, 1.0526, 0.7783, 1.5329, 1.4249, 0.9636)$, 最后, 比较 $z^*(j)$ 与 $Z^*(j)$ 之间的距离, 从而判定样本的归属等级 (见表 4)。

从表中可看出样本 15 排到第一位, 说明其全氮、全磷含量均占首位, 有机质含量、土壤 pH 值、代换量及其它指标均较高, 故排在首位; 而样本 18 的各项指标均较低, 因此排在最后。其它样本依次类推。同时, 样本 (4, 6, 8, 10, 11, 20) 的投影值十分接近, 说明这 6 个样本通过上述 6 个指标来评价的综合性状十分接近, 同样样本 (7, 9) 等也具有同样接近的综合性状。投影值越高的样本其排名越靠前, 等级值越小, 等级就越高, 土壤养分综合质量越好。

这样, 根据 PPE 模型计算出等级评价, 就可将 21 个土壤样本划分为不同等级。可见, 综合性状相近的土壤均划分为同一类。同时, 本次土壤采样分类及等级评价说明, 该地区土壤综合性能指标平等基本属于 II 级, III 级与 IV 级。

表 3 土壤样本及其性状指标

| 样本号 | 土壤名称 | 全氮 % | 全磷 % | 有机质 % | pH | 耕层厚 /cm | 物理黏粒含量 % |
|-----|------------|-------|-------|-------|-----|---------|----------|
| 1 | 薄层黏底白浆化黑土 | 0.270 | 0.142 | 6.46 | 5.5 | 21 | 45.3 |
| 2 | 厚层黏底黑土 | 0.171 | 0.115 | 3.46 | 6.3 | 60 | 45.3 |
| 3 | 薄层黏底黑土 | 0.114 | 0.101 | 2.43 | 6.4 | 25 | 51.0 |
| 4 | 厚层黏底黑土 | 0.173 | 0.123 | 3.30 | 5.8 | 65 | 45.6 |
| 5 | 薄层黏底黑土 | 0.145 | 0.131 | 3.28 | 6.0 | 25 | 51.0 |
| 6 | 厚层草甸黑土 | 0.173 | 0.140 | 3.45 | 5.8 | 60 | 49.0 |
| 7 | 中层草甸黑土 | 0.250 | 0.177 | 5.51 | 7.2 | 45 | 46.6 |
| 8 | 薄层草甸黑土 | 0.237 | 0.189 | 5.37 | 6.1 | 27 | 45.0 |
| 9 | 薄层沟谷地草甸黑土 | 0.319 | 0.227 | 7.04 | 5.8 | 24 | 39.3 |
| 10 | 厚层平地草甸土 | 0.163 | 0.124 | 3.73 | 6.2 | 61 | 48.1 |
| 11 | 中层平地草甸土 | 0.194 | 0.201 | 4.50 | 5.7 | 35 | 47.4 |
| 12 | 厚层沟谷地草甸土 | 0.142 | 0.185 | 3.79 | 6.4 | 55 | 51.0 |
| 13 | 厚层沟谷地潜育草甸土 | 0.240 | 0.217 | 4.92 | 6.5 | 41 | 63.6 |
| 14 | 薄层沟谷地潜育草甸土 | 0.253 | 0.172 | 4.63 | 6.8 | 20 | 44.1 |
| 15 | 厚层平地碳酸盐草甸土 | 0.357 | 0.289 | 7.21 | 7.5 | 40 | 48.3 |
| 16 | 中层沟谷碳酸盐草甸土 | 0.280 | 0.204 | 10.68 | 6.7 | 31 | 41.5 |
| 17 | 中层泛滥地草甸土 | 0.164 | 0.141 | 3.05 | 4.8 | 30 | 30.9 |
| 18 | 砂质泛滥地草甸土 | 0.095 | 0.099 | 1.51 | 6.0 | 20 | 26.4 |
| 19 | 侵蚀性暗棕壤 | 0.392 | 0.240 | 6.62 | 5.3 | 14 | 34.7 |
| 20 | 薄层白浆土 | 0.267 | 0.208 | 6.25 | 5.8 | 19 | 42.4 |
| 21 | 破皮黄白浆土 | 0.137 | 0.111 | 3.04 | 5.1 | 18 | 43.1 |

表 4 土壤样本 RAGA—PPE 模型计算结果及其等级划分

| 序号 | 样本投影值 | 从大到小排列 | 排序后样本号 | 计算值 y^* | 样本归属级 |
|----|--------|--------|--------|-----------|-------|
| 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 | 1.3661 | 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 |

6 结 论

本文应用 PPE模型建立了土壤标准样本等级评价的 PPE模型,并将函数值相近的划分为同一等级,与实际情况基本吻合,结果令人满意;将标准遗传算法进行改进,建立了完整的建模步骤,并具体应用到

土壤质量分类与等级评价中,避免了人为赋予权阵的干扰,取得良好效果;可望将 RAGA与 PPE结合模型推广到其它相关研究,为该领域研究提供一种新的思路与方法。

[参 考 文 献]

- [1] 胡金明,刘兴土.三江平原土壤质量变化评价与分析[J].地理科学,1999,5(19): 417-421.
- [2] 刘兴久,等.模糊聚类分析在土壤分类中的应用[J].哈尔滨:东北农学院学报,1988(2): 12-126.
- [3] 谢季坚,刘承平.模糊数学方法应用[M].(第二版).武汉:华中理工大学出版社,2000.81-106.
- [4] 金菊良,丁晶.遗传算法及其在水科学中的应用[M].成都:四川大学出版社,2000.42-47.
- [5] 现代数学手册(随机数学卷)[M].武汉:华中科技大学出版社,2001.512-520.
- [6] 项静恬,史久恩.非线性系统中数据处理的统计方法[M].北京:科学出版社,2000.172,190-192.
- [7] 张欣莉.博士学位论文.投影寻踪及其在水文水资源中的应用[M].成都市,2000.67-73.
- [8] 周明,孙树栋.遗传算法原理及其应用[M].北京:国防工业出版社,2000.4-7,37-38.
- [9] 付强,梁川.节水灌溉系统建模与优化技术[M].成都:四川科学技术出版社,2002.140-142.
- [10] 水利部农村水利司,中国灌排技术开发培训中心.水土资源评价与节水灌溉规划[M].北京:中国水利水电出版社,2000.59-65.