

投影寻踪等级评价模型在土壤质量变化综合评价中的应用*

赵小勇¹ 付强^{1,2†} 邢贞相³

(1 东北农业大学水利与建筑学院, 哈尔滨 150030)

(2 北大荒农垦集团公司博士后科研工作站, 哈尔滨 150040)

(3 河海大学水资源与环境学院, 南京 210024)

APPLICATION OF PROJECTION PURSUIT GRADE EVALUATION MODEL IN COMPREHENSIVE EVALUATION OF CHANGES IN SOIL QUALITY

Zhao Xiaoyong¹ Fu Qiang^{1,2†} Xing Zhenxiang³

(1 College of Hydraulic and Architecture, Northeast Agricultural University, Harbin 150030, China)

(2 Doctoral Working Station of Beidahuang Agricultural Reclamation Group Company, Harbin 150040, China)

(3 College of Water Resources and Environment, Hehai University, Nanjing 210024, China)

关键词 土壤质量; 评价; 投影寻踪; 遗传算法
中图分类号 S153.6 文献标识码 A

土壤质量动态变化研究是以土壤质量评价为基础,通过土壤质量指数的时空变化来反映。因评价实体、目标、指标体系的不同,评价模式(方法)也存在差异。由于涉及多个评价指标,所以属于高维数据处理问题。针对传统评价方法存在指标权重确定没有统一的理论和计算方法的不足,本文提出用基于实码加速遗传算法(Real coding based accelerating genetic algorithm,简称RAGA)的投影寻踪等级评价(Projection pursuit grade evaluation,简称PPE)模型^[1-7]。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,用倒S曲线建立投影寻踪土壤质量综合评价模型,从而实现土壤样本的评价。

1 土壤质量综合评价的投影寻踪等级评价模型

下面提出一套基于RAGA的投影寻踪等级评价

模型用于土壤质量综合评价新方法,建模具体步骤如下^[6,7]:

步骤1:构造投影指标函数。设根据土壤质量评价标准表产生的土壤样本的经验土壤质量等级及其土壤质量指标分别为 $y(i)$ 及 $\{x^*(i, j) | j=1 \sim p\}$, $i=1 \sim n$ 。其中, n, p 分别为土壤样本个数和土壤质量指标个数。最低等级设为1、最高等级设为 N 。并对 $\{x^*(i, j) | j=1 \sim p\}$ 进行归一化处理,即

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

式中, $x_{\max}(j)$ 为第 j 个指标值的最大值。

为建立土壤质量综合评价模型就是建立 $\{x^*(i, j) | j=1 \sim p\}$ 与 $y(i)$ 之间的数学关系。PPE方法就是把 p 维数据 $\{x^*(i, j) | j=1 \sim p\}$ 综合成以 $a = (a(1), a(2), \dots, a(p))$ 为投影方向的一维投影值 $z(i)$

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

* 国家“863”项目(No. 2002AA2Z4251-09)、国家自然科学基金(No. 30400275)、黑龙江省青年基金(No. QC04C28)和黑龙江省教育厅科技项目(No. 10541033)联合资助

† 通讯作者:付强(1973~),男,辽宁锦州人,东北农业大学水利与建筑学院,教授,博士生导师,主要从事农业水土资源系统分析、节水灌溉及农业系统工程建模与优化技术研究。电话:0451-55191294;E-mail:fuqiang100@371.net

作者简介:赵小勇(1980~),男,江苏南通人,东北农业大学水利与建筑学院,硕士,主要从事农业水土资源优化利用与系统分析方面研究

收稿日期:2005-09-08;收到修改稿日期:2006-03-17

然后根据 $z(i) \sim y(i)$ 的散点图建立适当的数学模型。为消除各预测因子的量纲效应,使建模型具有一般性,式(2)中, a 为单位长度向量。

在综合投影值时,要求投影值 $z(i)$ 应尽可能大地提取 $\{x(i, j)\}$ 中的变异信息,即 $\{x(i, j)\}$ 的标准差 S_z 达到尽可能大;同时要求 $z(i)$ 与 $y(i)$ 的相关系数的绝对值 $|R_{zy}|$ 达到尽可能大。这样得到的投影值就可望能尽可能多地携带预测因子系统 $\{x(i, j) | i=1 \sim n, j=1 \sim p\}$ 的变异信息,并且能够保证投影值对预测对象 $y(i)$ 具有很好的解释性。基于此,投影指标函数可构造为

$$Q(a) = S_z |R_{zy}| \quad (3)$$

式中, $| \cdot |$ 为取绝对值, S_z 为投影值 $z(i)$ 的标准差,

$$S_z = \left[\sum_{i=1}^n (z(i) - E_z)^2 / (n - 1) \right]^{0.5} \quad (4)$$

R_{zy} 为 $z(i)$ 与 $y(i)$ 的相关系数,即

$$R_{zy} = \frac{\sum_{i=1}^n (z(i) - E_z)(y(i) - E_y)}{\left[\sum_{i=1}^n (z(i) - E_z)^2 \sum_{i=1}^n (y(i) - E_y)^2 \right]^{0.5}} \quad (5)$$

式(4)、式(5)中, E_z 、 E_y 分别为序列 $\{z(i)\}$ 和 $\{y(i)\}$ 的均值。

步骤 2:优化投影指标函数。当给定经验土壤质量等级及其土壤质量指标的样本数据 $\{y(i) | i=1 \sim n\}$ 和 $\{x^*(i, j) | i=1 \sim n, j=1 \sim p\}$ 时,投影指标函数 Q_a 只随投影方向 a 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。可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

$$\max Q(a) = S_z / R_{zy} \quad (6)$$

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

这是一个以 $\{a(j) | j=1 \sim p\}$ 为变量的复杂优

化问题,常规方法处理很困难。模拟生物进化中优胜劣汰规则与群体内部染色体信息交换机制的加速遗传算法(RAGA),是一种通用的全局性优化方法,用它来求解上述优化问题则十分简便和有效。

步骤 3:建立投影寻踪土壤质量综合评价模型。把由步骤 2 求得的最佳投影方向的估计值 a^* 代入式(1)后即得第 i 个土壤样本投影值的计算值 $z^*(i)$,根据 $z^*(i) \sim y(i)$ 的散点图可建立相应的数学模型。通过对土壤质量评价的研究表明,用倒 S 曲线作为土壤质量综合评价模型是很合适的,即:

$$y^*(i) = \frac{N}{1 + e^{c(1) + c(2)/(z^*(i) + 1.5)}} \quad (8)$$

式中, $y^*(i)$ 为第 i 个土壤样本质量等级的计算值;最大土壤质量等级 N 为该曲线的上限值; $c(1)$ 、 $c(2)$ 为待定参数,它们通过求解如下最小化问题来确定:

$$\min F(c(1), c(2)) = \sum_{i=1}^n (y^*(i) - y(i))^2 \quad (9)$$

同样可用加速遗传算法(RAGA)来处理这一非线性优化问题。

2 实例分析

基于上述建模理论与步骤,采用时间对比法,对三江平原富锦市永富乡耕作土壤进行剖面取样分析,计算土壤表层土(耕作层)的指标投影值,通过不同开垦年限土壤质量综合评价值(最佳投影函数值)的大小来评价开垦后土壤变化趋势。本文主要借用文献[8]的资料,选择与土壤质量相关的物理、化学指标(生物学指标难以测定),指标因子包括:有机质、全氮、全磷、全钾、速效氮、速效磷、速效钾、CEC、pH 值。各个评价指标的分级标准见文献[8]。富锦市永富乡表层土壤的原始实测数据见表 1 所示。

表 1 原始实测数据

采样地点	开垦年限(a)	有机质	全氮	全磷	全钾	速效氮	速效磷	速效钾	CEC	pH
		(g kg ⁻¹)			(mg kg ⁻¹)			(cmol kg ⁻¹)		
富锦市	荒地	96.54	6.23	1.30	5.41	840.4	13.2	217.6	92.7	6.4
永富乡	5	49.62	3.35	0.89	6.90	360.3	10.50	281.6	85.7	6.1
	10	34.00	2.61	1.15	8.05	289.1	7.28	125.3	98.4	5.9
	25	22.15	1.46	0.89	6.63	317.8	8.91	65.7	87.9	6.2

根据土壤质量分级标准、表 1 数据,就可利用 PPE 对富锦市永富乡表层土壤质量变化进行综合评价。计算过程为:

在土壤质量分级标准中各等级取值范围内均匀随机产生各 5 个样本 $x^*(i, j)$, 与对应的等级一起组成样本系列, 并对 $x^*(i, j)$ 进行归一化处理为 $x(i, j)$, ($i =$

$1, 2, \dots, 25; j = 1, 2, \dots, 9$), 用 PPE 中的式(1) - 式(6)得:

最大投影指标函数值 $Q(a^*) = 0.7339$

最佳投影方向 $a^* = (0.3241, 0.3556, 0.3154, 0.4064, 0.3557, 0.3585, 0.3283, 0.3706, 0.0605)$

把 a^* 代入式(2)后即得各土壤样本投影值的计算值 $z^*(i)$, 见表 2。

表 2 土壤质量等级的经验值和 PPE 模型的计算值的对比结果

土样 序号	土壤质量指标									投影值	土壤质量等级	
	有机质	全氮 (g kg^{-1})	全磷	全钾	速效氮 (mg kg^{-1})	速效磷	速效钾	CEC (cmol kg^{-1})	pH		经验值	计算值
1	156.0	8.81	3.23	35.14	387.6	120.3	355.4	283.7	6.42	2.178	1	0.979
2	114.3	6.51	3.72	45.54	538.3	183.9	505.7	304.5	6.13	2.441	1	0.800
3	103.9	9.94	2.87	41.01	649.7	105.0	618.8	299.7	6.36	2.394	1	0.828
4	99.70	8.04	2.09	42.08	360.2	136.8	509.0	339.8	6.23	2.146	1	1.005
5	144.4	5.39	2.34	34.13	648.1	187.3	654.6	378.1	6.08	2.458	1	0.790
6	64.62	4.18	1.90	24.52	298.2	75.96	329.7	192.3	6.84	1.382	2	2.010
7	66.09	4.31	1.93	22.34	297.3	87.04	325.6	194.0	5.94	1.386	2	2.003
8	73.23	4.37	1.61	19.07	315.4	82.46	347.7	160.7	6.87	1.334	2	2.107
9	76.78	4.44	1.51	18.20	271.3	88.94	305.4	186.4	5.86	1.311	2	2.154
10	78.17	4.46	2.01	20.45	339.6	77.14	270.8	186.8	6.71	1.382	2	2.010
11	52.17	2.03	1.46	16.88	261.0	58.11	255.6	126.3	5.19	0.989	3	2.927
12	43.79	2.23	1.30	16.13	240.1	51.11	239.7	108.7	5.07	0.922	3	3.105
13	45.69	2.64	1.29	12.85	259.6	49.15	269.2	132.2	5.13	0.956	3	3.014
14	52.58	2.56	1.16	10.27	214.3	61.53	218.3	123.9	5.34	0.900	3	3.165
15	44.64	2.28	1.22	14.44	270.1	59.38	201.0	134.4	5.43	0.945	3	3.044
16	29.72	1.73	0.87	5.58	125.7	18.17	184.6	60.13	7.41	0.563	4	4.063
17	23.87	1.55	0.75	7.45	179.5	31.08	173.6	98.99	7.01	0.634	4	3.885
18	29.38	1.27	0.88	8.38	129.4	36.23	153.1	66.00	7.22	0.597	4	3.980
19	22.68	1.36	0.51	7.09	169.0	30.78	122.3	77.74	7.13	0.500	4	4.095
20	24.79	1.77	0.67	3.40	144.8	39.01	175.5	67.31	7.25	0.569	4	4.048
21	17.83	0.22	0.09	2.23	17.36	1.99	51.26	33.61	8.01	0.185	5	4.748
22	13.64	0.19	0.45	0.85	38.66	5.47	105.3	13.57	4.48	0.215	5	4.712
23	1.30	0.17	0.26	1.70	75.36	0.15	48.25	48.00	4.67	0.197	5	4.735
24	4.14	0.23	0.19	1.52	7.20	0.84	74.32	6.05	5.34	0.139	5	4.797
25	1.01	0.09	0.16	0.10	20.04	6.65	47.32	8.30	9.18	0.109	5	4.825

图 1 为最佳投影值 $z^*(i)$ 与经验等级 $y(i)$ 的散点图。从 $z^*(i)$ 与 $y(i)$ 的散点图可以看出, $z^*(i)$ 与 $y(i)$ 的图形与倒 S 型曲线较为接近, 因此

可以用倒 S 型曲线所对应的函数, 即式(8)建立土壤质量变化综合评价的数学模型。

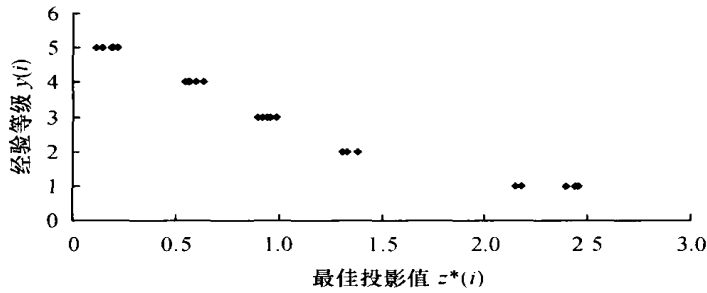


图1 投影值与经验等级的散点图

用 PPE 计算的土壤质量变化综合评价模型为：
$$y^*(i) = 5 / \left(1 + e^{5.0914 - 13.5291 / (z^*(i) + 1.5)} \right) \quad (10)$$
 用 PPE 中的式(10)计算的等级与各个样本的经验值的误差分析见表 3。

表 3 各土壤样本等级的经验值 y(i) 与 PPE 计算值 y*(i) 的误差分析结果

绝对误差值落在下列区间的百分比 (%)			平均绝对误差	平均相对误差 (%)
[0,0.1]	[0,0.2]	[0,0.3]		
48	76	100	0.112 6	4.82

于是可用 PPE 模型和表 1 的实测数据对三江平原地区富锦市永富乡表层土壤质量变化进行综合评价,见表 4。

表 4 用 PPE 模型评价土壤质量等级

采样地点	开垦年限(a)	投影值	土壤质量等级计算值(PPE)
富锦市	荒地	1.320	2.136
永富乡	5	0.852	3.296
	10	0.713	3.677
	25	0.594	3.988

文献[8]对富锦市永富乡表层土壤开垦年限为荒地、5年、10年、15年的土壤质量,分别为352、290、255和220。没有给出富锦市永富乡表层土壤质量综合评价的具体等级,而PPE模型对富锦市永富乡表层土壤质量变化综合评价给出了土壤具体所属的评价等级的数值,PPE模型的结果较常规方法更合理、精确。

从各开垦年限土壤质量等级计算值可知,富锦市永富乡表层土壤质量呈下降趋势,富锦市永富乡开荒前表层土壤质量属于 级;开荒5年后,表层土壤质量 级;开荒10年后,表层土壤质量属于 级;而开荒25年后表层土壤继续退化,表层土壤质量任属于 级。

对照表1,最佳投影方向各分量绝对值的大小实质上反映了各土壤质量指标对土壤质量等级的影响程度,各分量绝对值越大则对应的土壤质量指标

对土壤质量等级的影响程度就越大,据此可进一步检验原订土壤质量评价标准的合理性。在本例中,最佳投影方向 $a^* = (0.324 1, 0.355 6, 0.315 4, 0.406 4, 0.355 7, 0.358 5, 0.328 3, 0.370 6, 0.060 5)$ 表明,全钾、CEC、速效磷、速效氮、全氮、速效钾、有机质、全磷和 pH 对土壤质量等级的影响程度依次减小,这与土壤质量管理的经验相一致,说明原订土壤质量评价标准是合理的。

3 结 论

1) 用 PPE 模型对三江平原地区富锦市永富乡表层土壤质量变化综合评价,将土壤质量评价指标作为多个投影参数来寻求其最佳投影方向,并由最佳投影值计算出各土壤样本的所属等级,避免了人为赋予权重的干扰,取得了满意效果。

2) PPE 模型具有适应性、通用性强的特点,只要求给出评价标准表和评价对象的各指标值,对式(8)进行很小的修改,即可适应新的评价问题。PPE 模型与所评价问题的性质无关,可广泛应用于各种评价问题中。

参考文献

- [1] 王建国,杨林章,单艳红. 模糊数学在土壤质量评价中的应用研究. 土壤学报, 2001, 38(2): 176~183
- [2] 王辉,董元华,安琼. 南京市郊区蔬菜地土壤环境质量评价. 土壤, 2005, 37(3): 295~298
- [3] Hall P. On polynomial-based projection indices for exploratory projection pursuit. The Annals of Statistics, 1989, 17(2): 589~605
- [4] Fu Q. Study on the PPE model based on RAGA to classify the county energy. Journal Systems Science and Information, 2004, 2(1): 73~82
- [5] Friedman J H, Turkey J W. A projection pursuit algorithm for exploratory data analysis. IEEE Trans. on Computer, 1974, 23(9): 881~890
- [6] Fu Q, Fu H. Applying PPE model based on RAGA in the investment decisionmaking of water saving irrigation project. Nature and Science, 2003, 1(1): 57~61
- [7] Fu Q, Xie Y G, Wei Z M. Application of projection pursuit evaluation model based on real-coded accelerating genetic algorithm in evaluating wetland soil quality variations in the Sanjiang Plain, China. Pedosphere, 2003, 13(3): 249~256
- [8] 卢铁光,杨广林,王立坤. 基于相对土壤质量指数法的土壤质量变化评价与分析. 东北农业大学学报, 2003, 34(1): 56~59

《土壤学报》荣获“百种中国杰出学术期刊”称号

根据中国科技论文与引文数据库(CSTPCD)的最新统计结果,中国科学技术信息研究所发布消息,《土壤学报》荣获 2005 年“百种中国杰出学术期刊”称号。《土壤学报》自 2002 年以来已连续四年获此荣誉称号。

又据 2006 年版《中国科技期刊引证报告(核心版)》,2005 年度《土壤学报》总被引频次和影响因子分别为 2 183 和 1. 825,在 77 种农学类期刊中分别排名第 3 位和第 2 位,在 1 652 种中国科技论文统计源期刊中分别排名第 46 位和第 13 位。

2005 年农学类期刊总被引频次和影响因子前 10 名排序

刊名	总被引频次	学科内排名	刊名	影响因子	学科内排名
中国农业科学	2 835	1	PEDOSPHERE	2. 835	1
作物学报	2 617	2	土壤学报	1. 825	2
土壤学报	2 183	3	土壤	1. 488	3
园艺学报	1 987	4	水土保持学报	1. 169	4
水土保持学报	1 955	5	作物学报	1. 169	4
农业工程学报	1 638	6	中国水稻科学	1. 161	6
土壤	1 108	7	玉米科学	1. 007	7
植物病理学报	1 106	8	中国农业科学	0. 975	8
土壤通报	1 078	9	植物营养与肥料学报	0. 896	9
植物营养与肥料学报	1 041	10	植物病理学报	0. 848	10

数据来源:2006 年版《中国科技期刊引证报告(核心版)》