

尚梦佳,周忠发,王小宇,等.基于支持向量机的喀斯特山区土壤环境质量评价:以贵州北部一茶叶园区为例[J].中国岩溶,2018,37(4):575-583.

DOI:10.11932/karst20180411

## 基于支持向量机的喀斯特山区土壤环境质量评价 ——以贵州北部一茶叶园区为例

尚梦佳<sup>1,2</sup>,周忠发<sup>1,2</sup>,王小宇<sup>1,2</sup>,黄登红<sup>1,3</sup>,张珊珊<sup>1,3</sup>

(1. 贵州师范大学喀斯特研究院/地理与环境科学学院,贵阳 550001;

2. 贵州省喀斯特山地生态环境国家重点实验室培育基地,贵阳 550001;

3. 国家喀斯特石漠化防治工程技术研究中心,贵阳 550001)

**摘 要:**以贵州北部一茶叶园区 80 个表层土壤样品为研究对象,对其 Hg、As、Cd、Pb、Cr 和 Cu 含量进行测试,在 MATLAB 中应用支持向量机构建土壤环境质量评价模型,并与模糊综合评价法和内梅罗综合污染指数法的评价结果对比分析,探究支持向量机模型在喀斯特山区土壤环境质量评价中的适用性,其结果表明:研究区土壤质量 I 类与 II 类样品比例为 33:7,土壤环境质量大多数为 I 类;支持向量机方法的评价结果与模糊综合评价法和内梅罗综合污染指数法结果的相同率分别达到 82.5% 和 80.0%,并分析结果有差异的样品,发现支持向量机评价结果更符合实际情况,这说明该模型适用于土壤环境质量的评价。

**关键词:**喀斯特土壤;支持向量机;环境评价

**中图分类号:**X825

**文献标识码:**A

**文章编号:**1001-4810(2018)04-0575-09

### 0 引 言

茶叶质量安全问题主要源于农药残留、有害重金属残留和有害微生物的存在<sup>[1]</sup>。重金属在土壤中积累到一定程度会影响茶叶品质安全,进而对人体的健康造成威胁<sup>[2-3]</sup>,故对茶业园区中土壤重金属含量进行等级评价以采取相应的管控措施十分重要。目前,常用的土壤环境质量评价方法按性质可分为 3 大类:定性、定量和定性定量相结合,其中包括单因子污染指数法<sup>[4]</sup>、内梅罗综合污染指数法<sup>[5]</sup>、聚类分析法<sup>[6]</sup>、灰色聚类法<sup>[7]</sup>、模糊数学法<sup>[8]</sup>、层次分析法<sup>[9]</sup>、潜在生态评价法<sup>[10]</sup>等。茶叶园区土壤环境是多种因子构成的,在评价过程中各重金属评价因子与评价等

级之间存在着非线性映射关系,传统方法不能有效处理非线性问题,且计算过程较为复杂,受主观因素的影响较大,需人为给定评价因子权重,因此其评价结果缺乏可靠性<sup>[11]</sup>。

机器学习方法具有超强的容错和容差能力,可有效消除人为和外界干扰,支持向量机(Support Vector Machine,简称 SVM)有严格的数学基础,可通过用内积函数所定义的非线性变换将输入空间变换到一个高维空间,对小样本的泛化能力非常强<sup>[12]</sup>,SVM 模型已广泛应用于洪水灾情评价<sup>[13]</sup>、湖水富营养化评价<sup>[14]</sup>、遥感图像识别<sup>[15]</sup>、CO<sub>2</sub> 排放量预测<sup>[16]</sup>、地下水质量预测<sup>[17]</sup>等,但 SVM 模型在喀斯特地区土壤环境质量评价中的应用十分少见。本文以贵州北部喀

基金项目:国家自然科学基金地区项目“喀斯特石漠化地区生态资产与区域贫困耦合机制研究”(41661088);贵州省科技计划“基于北斗卫星的山地高效农业产业园区智能管理系统开发与应用”(黔科合 GY 字[2015]3001);贵州省高层次创新型人才培养计划——“百”层次人才(黔科合平台人才[2016]5674);国家遥感中心贵州分部平台建设(黔科合计 Z 字[2012]4003)(黔科合计 Z 字[2013]4003)

第一作者简介:尚梦佳(1993—),女,硕士,研究方向:地理信息系统与遥感。E-mail:shangmengjia521@sina.cn.

通信作者:周忠发(1969—),男,教授,博士生导师,研究方向:GIS 与遥感,喀斯特资源与环境。E-mail:fa6897@163.com.

收稿日期:2017-08-22

斯特山区茶叶园区为例,利用 SVM 模型评价土壤环境质量,并与模糊综合评价法和内梅罗综合污染指数法的评价结果进行对比,旨在探究 SVM 在喀斯特山区土壤环境质量评价中的适用性,以期为当地改善土壤环境现状提供帮助。

## 1 评价模型

### 1.1 支持向量机模型

SVM 模型基于统计学理论发展而来,通过 VC 维理论和寻求结构风险最小化原理来得到最优结果,其数学原理严谨,学习能力更强,精度较高,在很大程度上可以解决困扰机器学习方法的模型选择、非线性、维数灾难等问题<sup>[18-19]</sup>。SVM 模型是基于核函数  $K(x_i, x_j)$  的展开运算,通过非线性映射  $\varphi$  将低维空间不可分的变量转换至高维的特征空间,使问题转化为特征空间的线性分类问题,进而得到最优分类超平面<sup>[20]</sup>。而在最优超平面平行的线上或者面上,有一些样点,这些样点即为支持向量(SV)<sup>[21]</sup>。在评价土壤环境质量时,评价因子与土壤质量评价等级存在着复杂性和非线性,需通过核函数来实现维度的变化。常用的核函数为:线性核函数(Linear)、多项式核函数(Polynomial)、径向基核函数(Radial Basis Function)和二层神经网络核函数(Sigmoid)。

#### 1.1.1 评价标准

参考《无公害食品茶叶产地环境条件》(NY5020-2001),本文选用对茶叶健康水平影响较明显的 Hg、As、Cd、Pb、Cr 和 Cu 6 种重金属作为评价因子。依据《土壤环境质量分级标准》(GB15618-1995),将评价结果分为 3 个等级,6 种评价因子的分级标准如表 1 所示。利用商业数学软件 MATLAB 中的 Rand 函数在 3 个评价等级的取值范围内生成随机的样本,若各评价因子的取值均在某一等级范围内,则该样本肯定符合该等级土壤环境质量标准。

#### 1.1.2 模型构建

本文在土壤环境质量小于一级标准,一级、二级标准之间,二级、三级标准之间分别随机生成 200 组样本,分别为 I、II、III 类,每组样本包含随机生成的 Hg、As、Cd、Pb、Cr、Cu 重金属元素的含量,共 600 组样本。随机选取 80% 的样本作为训练样本,剩余的样本作为测试样本。模型构建步骤如下:

表 1 6 种评价因子的分级标准

Table 1 Classification standard with six kinds of the evaluated factors

| 评价因子 | 一级标准                             | 二级标准 | 三级标准 | 土壤背景值 | 权重 |
|------|----------------------------------|------|------|-------|----|
|      | $\text{mg} \cdot \text{kg}^{-1}$ |      |      |       |    |
| Hg   | 0.15                             | 0.3  | 1.5  | 0.11  | 3  |
| As   | 15                               | 40   | 40   | 20    | 3  |
| Cd   | 0.2                              | 0.3  | 1.0  | 0.659 | 3  |
| Pb   | 35                               | 250  | 500  | 35.2  | 3  |
| Cr   | 90                               | 150  | 300  | 95.9  | 2  |
| Cu   | 35                               | 50   | 400  | 32    | 2  |

(1)数据归一化、去噪 将生成的 600 组样本数据归一化至  $[0, 1]$  之间,归一化采用公式(1):

$$\tilde{x} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

(2)确定模型结构 以 Hg、As、Cd、Pb、Cr、Cu 重金属元素的含量作为输入要素,土壤环境质量等级为输出要素,建立 SVM 土壤重金属评价模型。

(3)确定核函数 由于不同地区的土壤含量组成不同,必须建立适合研究区的评价模型,核函数的选取是评价模型建立的关键。此次选用不同的核函数构造不同的支持向量机进行对比,最终确定使用径向基函数<sup>[22]</sup>,其核函数为:

$$K(x, x_i) = \exp \left\{ - \frac{\|x - x_i\|^2}{\sigma^2} \right\} \quad (2)$$

构造的支持向量机判别函数为:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^s \alpha_i \exp \left\{ - \frac{\|x - x_i\|^2}{\sigma^2} \right\} - b \right\} \quad (3)$$

(4)建立模型 由于土壤环境质量分类是非线性分类,故采用 C-支持向量分类(C-SVC)模型,此模型需选择最优惩罚因子( $c$ )和最优核函数系数( $g$ ),在 MATLAB 中构建模型,将  $c$  和  $g$  带入 SVM 模型,训练随机样本,建立评价模型,并对实测的 80 组土壤样品进行评价,得到土壤环境质量评价结果。

## 1.2 模糊综合评价法

模糊综合评价法在模型中引入隶属度和评价因子权重值,能较准确刻画环境质量状况的模糊性和渐变性,把定性评价转化为定量评价<sup>[23-24]</sup>。评价过程中需确定评价因子权重,由于参照污染物的超标情况确定权重的方法应用较为广泛,故选用污染超标赋权法计算取得权重值。计算方法如下:

$$W_i = \frac{C_i/S_i}{\sum_{i=1}^n C_i/S_i} \quad (4)$$

式中:  $W_i$  为第  $i$  个评价因子的权重值;  $C_i$  为第  $i$  种评价因子  $n$  个采样点实测浓度的算术平均值;  $S_i$  为第  $i$  个评价因子各个评价级别标准限定值的算术平均值。

### 1.3 内梅罗综合污染指数法

单因子指数法可以体现单种元素污染情况,而内梅罗综合污染指数法可以突出污染较重的重金属污染物的作用,能够综合评判土壤污染情况。

单因子污染指数法,计算公式为:

$$P_i = \frac{C_i}{S_i} \quad (5)$$

式中:  $P_i$  为第  $i$  个评价因子的单因子污染指数;  $C_i$  为第  $i$  个评价因子的实测值;  $S_i$  为第  $i$  个评价因子的参照值,本文采用国家二级标准作为参照值。

内梅罗综合污染指数法,计算公式:

$$P_{\text{综}} = \sqrt{\frac{(\bar{P})^2 + (P_{i\text{max}})^2}{2}} \quad (6)$$

$$\bar{P} = \frac{\sum_{i=1}^n w_i P_i}{\sum_{i=1}^n P_i} \quad (7)$$

式中:  $P_{\text{综}}$  为综合污染指数;  $P_{i\text{max}}$  为各种金属元素单因子污染指数中的最大值;  $\bar{P}$  为各种金属元素单因子污染指数的加权平均值;  $w_i$  为第  $i$  个评价因子的权重。对于权重  $w_i$  的确立,Swaine<sup>[25]</sup> 按照重金属对环境的影响程度,将环境研究中微量元素分别赋值为 3、2、1 作为权重(表 1),表 2 所示为内梅罗综合污染指数法评价等级与污染水平。

表 2 内梅罗综合污染指数法评价等级

Table 2 Evaluation grade of Nemerow comprehensive pollution index method

| 内梅罗综合指数    | 等级  | 污染水平 |
|------------|-----|------|
| $\leq 0.7$ | I   | 清洁   |
| 0.7~1      | II  | 尚清洁  |
| $\geq 1$   | III | 污染   |

## 2 案例分析

### 2.1 研究区概况

研究区位于云贵高原至湖南丘陵的过渡地带,区

内大部分出露寒武系、奥陶系地层,喀斯特地貌发育,峰丛沟谷交错;境内属亚热带高原湿润季风气候,年降水量为 1 000~1 300 mm,降水充沛,但因季节变化而分布不均,年平均气温在 12.6~13.1℃,无明显热害和冷冻无霜期,年无霜期平均为 284 d,年日照时数为 1 000~1 300 h,是全国太阳辐射低值区之一;土壤类型以黄壤、黄棕壤为主。其良好的气候和土壤条件为茶叶的生长提供了优越条件,但地质特点也为土壤重金属的迁移扩散提供了便利。

此次研究土壤采样布点主要分布在 E107°28'3"~107°39'49.2", N27°53'11"~28°4'19.2", 根据区内土地利用现状及其特征,采用蛇形布点法,采样深度为 0~20 cm,取样 1 kg,共采集 80 组样品,并以北斗手持机记录样点位置信息,土壤采样点分布见图 1。根据《土壤环境监测技术规范》<sup>[26]</sup> 要求,将土壤样品风干,过 100 目尼龙筛,使用电感耦合等离子体质谱仪进行重金属元素测定<sup>[27]</sup>。

### 2.2 SVM 模型评价结果

惩罚因子( $c$ )与核函数系数( $g$ )的选取直接影响土壤环境质量模型的分类质量, $c$  用来调节模型复杂度并且可以对模型的误差进行控制, $g$  用来调节模型的识别能力。传统参数运算方法繁琐,且不易找到最优参数,如表 3 所示,本研究使用网格划分法(Grid)、遗传算法(Ga)和粒子群优化法(Pso)分别选取最优参数,3 种寻优方法的测试集分类准确率均达到了 100%,三者中网格划分法的预测集分类准确率最高,故选用网格划分法进行参数寻优(图 2),最终得到最优参数为  $c=0.003\ 906\ 3$ ,  $g=0.003\ 906\ 3$ 。基于构建的 SVM 模型,对实测的 80 个土壤样品进行评价,等级为 I 类的有 66 个,II 类的有 14 个,没有 III 类的(图 3)。

## 3 分析与评价

### 3.1 描述性统计分析

从表 4 中可看出: Hg、As、Cd、Pb、Cr 和 Cu 含量的平均值为 0.24 mg·kg<sup>-1</sup>、21.77 mg·kg<sup>-1</sup>、0.03 mg·kg<sup>-1</sup>、41.91 mg·kg<sup>-1</sup>、118.89 mg·kg<sup>-1</sup> 和 28.61 mg·kg<sup>-1</sup>,其中 Cd 含量最低,Cr 含量最高,但均未超过国家二级标准。与贵州省土壤背景值相比,Hg、As、Pb、Cr 平均值分别为土壤背景值的 2.14、

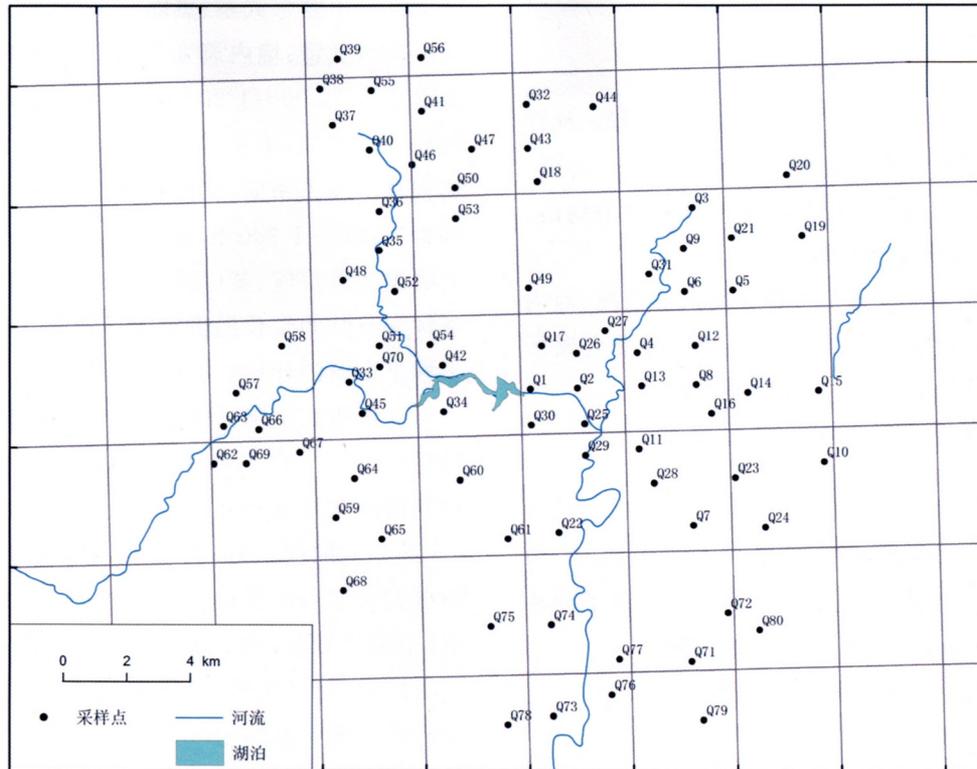


图 1 土壤采样点分布图

Fig. 1 Distribution map of soil sampling points

表 3 支持向量机模型参数寻优法精度对比

Table 3 Accuracy comparison of parameter optimization method of SVM

| 参数寻优法  | 惩罚因子( $c$ ) | 核函数系数( $g$ ) | 测试集分类准确率      | 预测集分类准确率      |
|--------|-------------|--------------|---------------|---------------|
| 网格划分法  | 0.003 906 3 | 0.003 906 3  | 100%(120/120) | 82.5%(66/80)  |
| 遗传算法   | 1.589 10    | 0.128 27     | 100%(120/120) | 43.75%(35/80) |
| 粒子群优化法 | 0.01        | 0.01         | 100%(120/120) | 80.00%(64/80) |

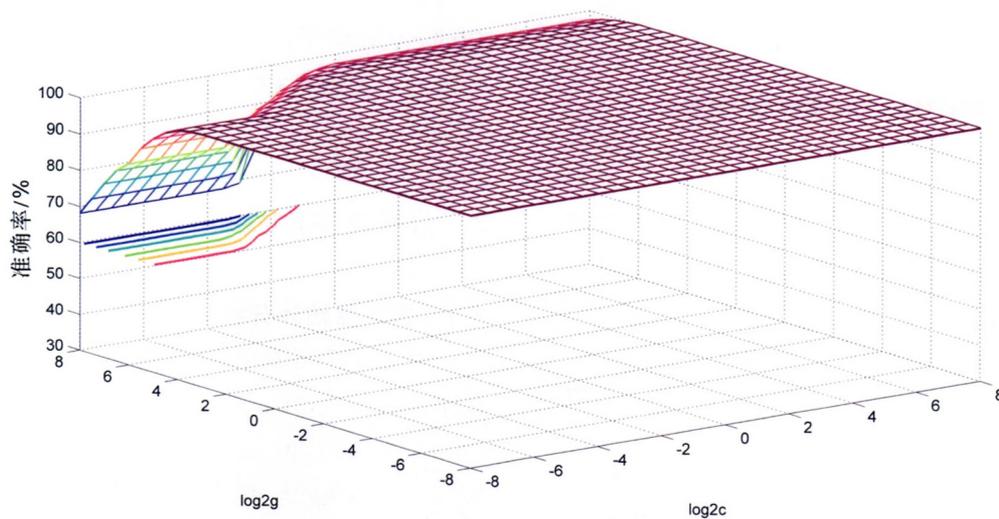


图 2 最优参数结果

Fig. 2 Optimal parameter result

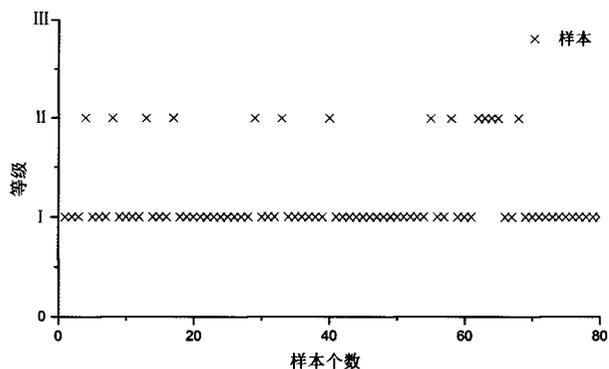


图 3 SVM 模型评价结果图

Fig. 3 Evaluation results diagram of SVM model

1.09、1.19、1.24 倍, Hg 含量与土壤背景值相差最多, 存在一定富集现象。与《无公害食品茶叶产地环境条件》(NY 5020-2001)要求的最低浓度对比, 评价因子均值都未超过限值, 但并不能代表所有样品均不存在污染, 进一步参照土壤重金属含量(图 4)可知, 所有样品中的 Hg、As、Cd、Pb、Cu 均在安全范围内, 有 7 个样品中存在 Cr 含量超过浓度限值, 占整个样品总量的 8.75%, 这表明绝大多数的土壤样品属于安全样品, 此茶园符合无公害、优质茶园建设。

表 4 80 个样品重金属含量的描述性统计结果

Table 4 Descriptive statistics of heavy metals in 80 soil samples

| 指标 | 样本 | 均值/<br>mg · kg <sup>-1</sup> | 标准差/<br>mg · kg <sup>-1</sup> | 变异<br>系数 | 二级标准/<br>mg · kg <sup>-1</sup> | 土壤背景值/<br>mg · kg <sup>-1</sup> | 浓度限值/<br>mg · kg <sup>-1</sup> |
|----|----|------------------------------|-------------------------------|----------|--------------------------------|---------------------------------|--------------------------------|
| Hg | 80 | 0.24                         | 0.02                          | 0.09     | 0.3                            | 0.11                            | 0.3                            |
| As | 80 | 21.77                        | 1.68                          | 0.08     | 40                             | 20                              | 40                             |
| Cd | 80 | 0.03                         | 0.02                          | 0.67     | 0.3                            | 0.659                           | 0.3                            |
| Pb | 80 | 41.91                        | 18.39                         | 0.44     | 250                            | 35.2                            | 250                            |
| Cr | 80 | 118.89                       | 20.00                         | 0.17     | 150                            | 95.9                            | 150                            |
| Cu | 80 | 28.61                        | 5.95                          | 0.21     | 50                             | 32                              | 150                            |

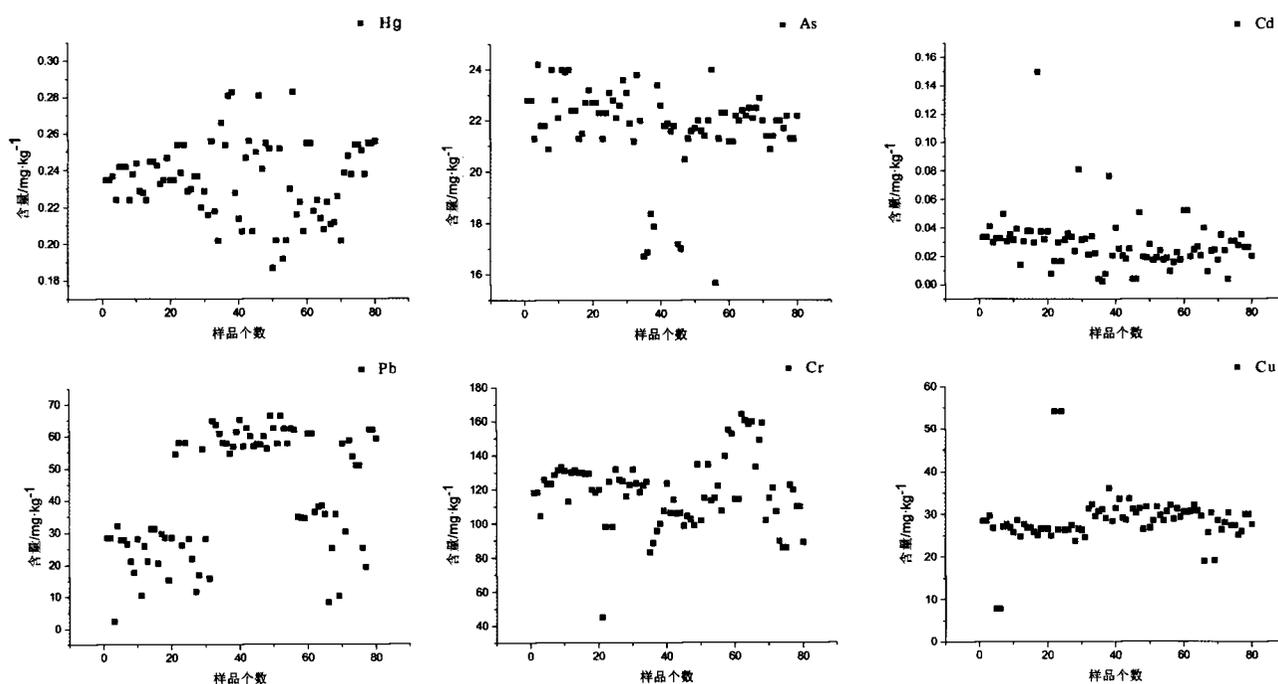


图 4 80 个样品土壤重金属含量图

Fig. 4 Soil heavy metal content of 80 soil samples

变异系数反映了土壤样品评价因子在各采样点的平均变异程度,研究区样品中重金属变异系数由高到低依次为: Cd>Pb>Cu>Cr>Hg>As。Wilding<sup>[28]</sup>将变异系数分为高度变异(>0.36)、中度变异(0.16~0.36)和低度变异(<0.16),其中 Hg、As 变异系数为 0.09 和 0.08,属于低度变异;Cu、Cr 的变异系数为 0.21 和 0.17,属于中度变异;Cd、Pb 变异较大,分别为 0.67 和 0.44,属于高度变异,空间分异相对显著。变异系数最大值是最小值的 8.5 倍,说明

此茶园土壤变异程度差距较大。

### 3.2 SVM 与其他方法结果对比分析

单因子污染指数法的结果表明,各采样点 Hg、As、Cd、Pb 和 Cu 污染程度均为未污染;Cr 元素中有 7 个采样点重金属超标,污染程度属未—轻度。6 种重金属元素污染程度由强至弱依次为: Cr>Hg>Cu>As>Pb>Cd。单因子污染指数仅能反映单一评价指标对土壤环境的污染程度,不能对土壤质量进行等级的判定,故其无法进行综合质量评价。

表 5 SVM 与模糊综合评价法、内梅罗综合污染指数法结果对比

Table 5 Comparison of SVM and fuzzy comprehensive evaluation method and Nemerow comprehensive pollution index method

| 样品编号 | SVM | 模糊综合评价法 | 内梅罗综合污染指数法 | 样品编号 | SVM | 模糊综合评价法 | 内梅罗综合污染指数法 |
|------|-----|---------|------------|------|-----|---------|------------|
| Q1   | I   | I       | I          | Q41  | I   | I       | I          |
| Q2   | I   | I       | I          | Q42  | I   | I       | I          |
| Q3   | I   | I       | I          | Q43  | I   | I       | I          |
| Q4   | II  | I       | I          | Q44  | I   | I       | I          |
| Q5   | I   | I       | I          | Q45  | I   | I       | I          |
| Q6   | I   | I       | I          | Q46  | I   | I       | II         |
| Q7   | I   | I       | I          | Q47  | I   | I       | I          |
| Q8   | II  | I       | I          | Q48  | I   | I       | I          |
| Q9   | I   | I       | I          | Q49  | I   | I       | II         |
| Q10  | I   | I       | I          | Q50  | I   | I       | I          |
| Q11  | I   | I       | I          | Q51  | I   | I       | I          |
| Q12  | I   | I       | I          | Q52  | I   | I       | II         |
| Q13  | II  | I       | I          | Q53  | I   | I       | I          |
| Q14  | I   | I       | I          | Q54  | I   | I       | I          |
| Q15  | I   | I       | I          | Q55  | II  | I       | I          |
| Q16  | I   | I       | I          | Q56  | I   | I       | II         |
| Q17  | II  | I       | II         | Q57  | I   | I       | II         |
| Q18  | I   | I       | I          | Q58  | II  | I       | II         |
| Q19  | I   | I       | I          | Q59  | I   | I       | II         |
| Q20  | I   | I       | I          | Q60  | I   | I       | I          |
| Q21  | I   | I       | I          | Q61  | I   | I       | I          |
| Q22  | I   | I       | I          | Q62  | II  | I       | II         |
| Q23  | I   | I       | I          | Q63  | II  | I       | II         |
| Q24  | I   | I       | I          | Q64  | II  | I       | II         |

续表 5

| 样品编号 | SVM | 模糊综合评价法 | 内梅罗综合污染指数法 | 样品编号 | SVM | 模糊综合评价法 | 内梅罗综合污染指数法 |
|------|-----|---------|------------|------|-----|---------|------------|
| Q25  | I   | I       | I          | Q65  | II  | I       | II         |
| Q26  | I   | I       | I          | Q66  | I   | I       | I          |
| Q27  | I   | I       | I          | Q67  | I   | I       | II         |
| Q28  | I   | I       | I          | Q68  | II  | I       | II         |
| Q29  | II  | I       | I          | Q69  | I   | I       | I          |
| Q30  | I   | I       | I          | Q70  | I   | I       | I          |
| Q31  | I   | I       | I          | Q71  | I   | I       | I          |
| Q32  | I   | I       | I          | Q72  | I   | I       | I          |
| Q33  | II  | I       | I          | Q73  | I   | I       | I          |
| Q34  | I   | I       | I          | Q74  | I   | I       | I          |
| Q35  | I   | I       | I          | Q75  | I   | I       | I          |
| Q36  | I   | I       | I          | Q76  | I   | I       | I          |
| Q37  | I   | I       | II         | Q77  | I   | I       | I          |
| Q38  | I   | I       | II         | Q78  | I   | I       | I          |
| Q39  | I   | I       | I          | Q79  | I   | I       | I          |
| Q40  | II  | I       | I          | Q80  | I   | I       | I          |

将 SVM 模型评价结果分别与模糊综合评价法和内梅罗综合污染指数法的评价结果进行对比验证分析(表 5)。SVM 模型评价研究区土壤质量结果 I 类和 II 类样品比例为 33 : 7, 表明茶园可进行茶叶的安全种植。内梅罗综合污染指数评价结果显示, 80% 的样品属于 I 类(清洁状态), 而 20% 的样品属于 II 类(处于警戒线水平), SVM 模型与内梅罗综合污染指数法评价结果相同率达 80%; 模糊综合评价法的评价结果显示, 80 个样品均属于 I 类, SVM 与模糊综合评价法评价结果相同率达 82.5%。SVM 模型与以上两种方法的评价结果相同率均处于较高水平, 故将 SVM 模型用于评价喀斯特山区茶园土壤环境质量评价是可行的。

由于研究区内土壤污染水平较轻, 内梅罗综合污染指数法易受含量最大的重金属污染物对土壤的影响, 单个评价因子较高会使内梅罗综合污染指数偏高; 模糊综合评价法考虑到不同重金属的隶属度和权重值, 解决了边界问题, 但无法解决复杂的非线性关系, 计算繁琐, 且确定权重及隶属度函数时受人为主观因素影响较大, 从而影响评价结果。以 Q62 样品为例: SVM 和内梅罗综合污染指数法的评价结果均

为 II 类, 模糊综合评价法评价结果为 I 类, 而实际情况中, Hg、As、Pb 均属于 II 类, Cr 属于 III 类, 仅有 Cd 和 Cu 属于 I 类, 由此可见, 此土壤环境质量的最终评价结果定为 II 类更符合实际情况。再以 Q33 样品为例: SVM 评价结果为 II 类, 内梅罗综合污染指数法和模糊综合评价法的评价结果为 I 类, 而实际情况中, Hg、As、Pb、Cr 均属于 II 类, 仅 Cd 和 Cu 属于 I 类, 由此可见, 此土壤环境质量的最终评价结果定为 II 类更符合实际情况。将评价结果有差异的样品根据实际情况进行分析, 可知 SVM 模型的评价结果较切合实际。SVM 利用核函数计算将低维空间非线性映射至高维特征空间, 能对非线性模式进行识别, 使复杂的非线性多分类问题得以解决, 且与传统评价方法相比操作简单, 评价客观, 具有很强的实用性。

#### 4 结论与讨论

(1) 研究区样品各重金属变异系数从高到低依次为: Cd > Pb > Cu > Cr > Hg > As, 其中 Hg、As 属于低度变异, Cu、Cr 属于中度变异, Cd、Pb 属于高度变异, 土壤重金属含量空间差异较大; 对比各评价因子

的平均值与贵州省土壤背景值,发现 Cd 和 Cu 小于土壤背景值,其他均介于背景值和国家土壤质量标准的二级标准值之间;对比样品含量值与茶叶产地条件的浓度限值,91.25%的土壤样品属于安全样品,此茶园符合无公害、优质茶园建设。

(2)选用大量样本进行训练和验证,SVM 评价结果显示研究区土壤环境质量全部为 I、II 类,说明区内土壤环境质量处于警戒线以内,土壤环境质量较好。SVM 模型评价方法与模糊综合评价法、内梅罗综合污染指数法的结果相同率分别达到 82.5%和 80.0%,对存在差异的结果进行分析,发现 SVM 更符合实际情况,说明 SVM 模型应用于喀斯特山区土壤环境质量的评价是可行的。

(3)相较于传统方法,SVM 分类模型采用结构风险最小化原则,能够解决复杂的非线性问题,且操作简单,评价不受人为干预的影响,具有很强的实用性,将其运用到土壤重金属分类中,为喀斯特山区土壤环境质量评价提供了新的思路和方法,在土壤重金属污染评价中有广泛的应用前景。但本文训练的 SVM 模型在进行等级分类标准设定时,缺乏对特殊样品的考虑,使得评价结果尚未实现高精度的定量化分析,这需在今后的研究中对模型进行完善。

## 参考文献

- [1] 汪庆华,刘新. 浅谈我国茶叶质量安全现状及应对措施[J]. 茶叶,2006,32(2):66-69.
- [2] 何璐君. 开阳县茶园土壤重金属元素含量及污染评价[J]. 贵州农业科学,2009,37(4):184-186.
- [3] 李云,张进忠,童华荣. 茶园土壤和茶叶中重金属的监测与污染评价[J]. 环境科学与技术,2008,31(5):77-81.
- [4] 李名升,张建辉,梁念,等. 常用水环境质量评价方法分析与比较[J]. 地理科学进展,2012,31(5):617-624.
- [5] 邓琴,吴迪,秦樊鑫,等. 岩溶铅锌矿区土壤重金属污染特征[J]. 中国岩溶,2017,36(2):248-254.
- [6] 柴世伟,温琰茂,张亚雷,等. 广州市郊区农业土壤重金属污染评价分析[J]. 环境科学研究,2006,19(4):138-142.
- [7] 丁立,田林亚,范雷刚,等. 基于灰色聚类法和灰色关联分析的大坝安全评价[J]. 测绘工程,2015,24(4):66-69,73.
- [8] 孙小涛,周忠发,黄智灵,等. 改进模糊综合评价模型对岩溶山区茶叶产地土壤重金属污染评判[J]. 中国岩溶,2016,35(3):282-290.
- [9] 孟宪林,郭威. 改进层次分析法在土壤重金属污染评价中的应用[J]. 环境保护科学,2001,27(1):34-36.
- [10] 任坤,陈志兵,潘晓东,等. 重庆南山老龙洞地下河表层沉积物重金属环境地球化学特征及生态风险评价[J]. 中国岩溶,2016,35(2):144-152.
- [11] 杨忠平,卢文喜,龙玉桥. 改进 BP 算法在城市土壤环境质量评价模型的应用[J]. 重庆大学学报(自然科学版),2010,32(2):99-103.
- [12] 李晓婷,刘勇,王平. 基于支持向量机的城市土壤重金属污染评价[J]. 生态环境学报,2014,23(8):1359-1365.
- [13] 刘德地,陈晓宏. 基于支持向量机的洪水灾情综合评价模型[J]. 长江流域资源与环境,2008,17(3):490-494.
- [14] 张成成,沈爱春,张晓晴,等. 应用支持向量机评价太湖富营养化状态[J]. 生态学报,2013,33(23):7563-7569.
- [15] 张锦水,何春阳,潘耀忠,等. 基于 SVM 的多源信息复合的高空间分辨率遥感数据分类研究[J]. 遥感学报,2006,10(1):49-57.
- [16] 周建国,张希刚. 基于粗糙集与灰色 SVM 的中国 CO<sub>2</sub> 排放量预测[J]. 中国环境科学,2013,33(12):2157-2163.
- [17] Yoon H, Jun S C, Hyun Y, et al. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer[J]. Journal of Hydrology, 2011, 396(1):128-138.
- [18] 梁雪春,龚艳冰,肖迪. 一种多核加权支持向量机的水质预测方法[J]. 东南大学学报:自然科学版,2011,41(S1):14-17.
- [19] 姜雪,卢文喜,杨青春,等. 应用支持向量机评价土壤环境质量[J]. 中国环境科学,2014,34(5):1229-1235.
- [20] 陈其松. 智能优化支持向量机预测算法及应用研究[D]. 贵阳:贵州大学,2009.
- [21] Vapnik V N. The nature of statistical learning theory[M]. Berlin: Springer-Verlag,1995.
- [22] 张强,黄生志,陈晓宏. 基于支持向量机的土壤湿度模拟及预测研究[J]. 土壤学报,2013,50(1):59-67.
- [23] LI Wei-Xin, ZHANG Xu-Xiang, WU Bing, et al. A Comparative Analysis of Environmental Quality Assessment Methods for Heavy Metal-Contaminated Soils[J]. PEDOSPHERE, 2008, 18(3):344-352.
- [24] 梁伟,张慧颖,朱孔来. 基于模糊数学和灰色理论的城市生态环境竞争力评价[J]. 中国环境科学,2013,33(5):945-951.
- [25] Swaine D J. Why trace elements are important[J]. Fuel Processing Technology, 2000, 65:21-33.
- [26] 国家环境保护总局. 土壤环境监测技术规范: HJ/T166-2004 [S]. 北京:中国环境科学出版社,2004.
- [27] 刘红艳,周燕,王铁夫,等. ICP-MS 法测土壤样品中铜铅的不确定度评定[J]. 光谱学与光谱分析,2011,31(4):1115-1118.
- [28] Wilding L P. Spatial variability: Its documentation, accommodation, and implication to soil surveys[M]//NIELSEN D R, BOUMA J. Soil Spatial Variability. Wageningen: Pudoc, 1985: 166-194.

## Evaluation of soil environmental quality in karst mountain area based on support vector machine: A case study of a tea plantation in northern Guizhou

SHANG Mengjia<sup>1,2</sup>, ZHOU Zhongfa<sup>1,2</sup>, WANG Xiaoyu<sup>1,2</sup>, HUANG Denghong<sup>1,3</sup>, ZHANG Shanshan<sup>1,3</sup>

(1. School of Karst Science/ School of Geography & Environmental Science, Guizhou Normal University, Guiyang, Guizhou 550001, China;  
2. The State Key Laboratory Incubation Base for Karst Mountain Ecology Environment of Guizhou Province, Guiyang, Guizhou 550001, China;  
3. State Engineering Technology Institute for Karst Desertification Control, Guiyang, Guizhou 550001, China)

**Abstract** The content of heavy metals in soil directly affects the quality and safety of tea, and even has a potential threat to human health. It is hence important to monitor, evaluate and control the content of heavy metals in the tea plantation soil. In this paper, we select a tea plantation in karst mountain area of northern Guizhou as a study area. The area is located in the transitional zone from Yun-Gui Plateau to Hunan hilly area, which belongs to the humid monsoon region of tropical plateau, with the annual precipitation of 1,000 – 1,300 mm and the annual average temperature of 12.6–13.1 °C. In this study area, because the Cambrian and Ordovician carbonate strata are widely exposed, karst landform is well developed and is characterised by interlacing occurrence of peak clusters and karst valleys. According to present situation and the characteristics of land use, in the area, 80 surface soil samples were collected for the analyses of heavy metal (such as mercury (Hg), arsenic (As), cadmium (Cd), lead (Pb), chromium (Cr) and copper (Cu)) contents and the environmental quality of the tea plantation soil. To classify and evaluate the sample analytical results, the Support Vector Machine (SVM) model coded in MATLAB was employed. Meanwhile, by comparing the result from Nemerow comprehensive pollution index method with that of fuzzy comprehensive evaluation method, the applicability of SVM in soil heavy metal pollution evaluation was discussed. These results show that, (1) There are significant spatial differences in soil heavy metal contents, with the variation coefficients in the order from high to low of Cr>Hg>Cu>As>Pb>Cd. By comparing the average value of evaluation factors with the soil background value of Guizhou Province, it is found that the values of Cd and Cu are lower than the soil background values, and the others fall in between the background values and the secondary standard values of the national soil quality standard. In fact, the chemical concentrations of 91.25% of the soil samples are below the standard limits for tea producing areas, which represents a soil environment of non-pollution and high-quality for tea plantation. (2) The quality of soil in the study area is good, with its soil environmental quality ranging between category I and II. The evaluation results of SVM method are quite similar to those of fuzzy comprehensive evaluation and Nemerow comprehensive pollution index methods, with a similarity of 82.5% and 80.0%, respectively. During the application of these methods it was found that the results of SVM were more accurate, which showed that the model is suitable for the evaluation of soil environmental quality in karst mountain area. (3) In addition, the SVM can solve complex nonlinear problems, with much easier manual operation and less artificial intervention, comparing with the application of traditional assessment models. It provides a new idea and method for the evaluation of soil environmental quality in karst mountain area.

**Key words** karst mountain area, Support Vector Machine, soil environmental quality, evaluation model, tea plantation

(编辑 黄晨晖)