

未确知聚类在蔬菜产地环境质量评价中的应用

葛莹莹,刘伟

(河北工程大学 信息与电气工程学院,河北 邯郸 056038)

摘要:文中针对蔬菜生产产地污染程度评价具有不确定性,将未确知聚类算法应用到蔬菜产地环境质量评价中。以江苏省南京市六合区六个乡的土壤环境质量为研究对象,采用我国《土壤环境质量标准》作为评价标准,选取土壤中最常见的六种污染物 Cd、Hg、As、Cu、Pb、Cr 为参评因子,对各采样点土壤重金属污染程度进行综合评价,并求得各类类中心及相应的隶属度。分析实验结果可以发现,与实际情况相符,表明该方法应用到蔬菜产地环境质量评价中是可行的。

关键词:环境质量评价;未确知聚类;类中心;隶属度

中图分类号:X53

文献标识码:A

Application of unascertained cluster on environmental quality evaluation for vegetable bases

GE Ying-ying, LIU Wei

(School of Information and Electrical Engineering, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: For the environmental quality evaluation of vegetable bases standards uncertainty, the paper applied the unascertained cluster method to the environmental quality evaluation. In the paper it selects soil environmental quality of six township of Liuhe district, Nanjing city, Jiangsu province as the research objects. Taking "China's soil environmental quality standard" as evaluation criteria, and selecting six common pollutants as evaluation factor, such as Cd, Hg, As, Cu, Pb, Cr. The environmental qualities comprehensively based on the measured data of Jiangsu province in the experiment were assessed. At last, six samples of soil environmental quality were divided into clean, light pollution and severe pollution. And the class centre, subjection degree and classification results were obtained. Experimental results show that the classification result is consistent with actual situation and the method applied to the vegetable origin in environmental quality assessment is feasible.

Key words: environmental quality evaluation; unascertained cluster; class center; subjection degree

蔬菜在生长过程中,主要从土壤中汲取养分,一旦土壤被污染,那么种植出来的蔬菜必然也不会是健康蔬菜。因此,为了保证蔬菜的质量安全,保护人类的生命健康,需要及时客观地对蔬菜产地污染程度进行评价,以便及时采取措施对土壤污染进行治理。随着我国城乡工业化程度的不断提高,土壤的污染源主要有两个:一是工业中“三废”的排放以及城镇生活垃圾的增加;二是污泥农用和蔬菜种植过程中使用有机肥、含重金属的农药以及化肥的不合理使用等。在各类污染物中,

蔬菜产地污染的主要来源是土壤重金属污染,其中,Hg、Cd、Cr、Pb、Zn、As、Cu、Co 等是最主要的重金属污染物^[1]。

近年来,土壤重金属污染的评价方法主要有模糊数学综合评判法^[2]、改进层次法^[3]、灰色聚类法^[4]、物元分析法^[5]等等,但各有优势及不足。未确知聚类是一种基于未确知集合的无监督分类,是针对解决不确定性问题而提出的,已成功地应用于煤矿矿井工作面瓦斯涌出量的分类评价以及井下空气质量的评价中。因此,本研究将针对以

上方法的不足,且考虑到其评价的不确定性,将未确知聚类应用到蔬菜产地环境质量评价中,对江苏省南京市六合区六个乡的土壤质量实测数据进行实例研究^[6],最后得到新的标准状态以及对应的评价准则。

1 未确知聚类

未确知聚类^[7]是一种基于未确知集合的无监督分类,它不同于一般的聚类分析,而是一种不确定的分类方法。对于任意给定集合 U , 数据样本个数为 N , 指标向量长度为 d , 拟将其分为 C 类, Γ_k ($k=1, 2, \dots, p$) 是第 k 类, 那么 U 中任意对象 x_i 都以测度 $\mu_{\Gamma_k}(x_i)$ 属于 Γ_k 类, $\mu_{\Gamma_k}(x_i)$ 是 $[0, 1]$ 区间中的实数, 且满足

$$\sum_{k=1}^p \mu_{\Gamma_k}(x_i) = 1 \tag{1}$$

1.1 未确知测度的确定

在得到初始分类的基础上, 得到 C 个初始聚类中心 $m_1^{(0)}, m_2^{(0)}, \dots, m_c^{(0)}$, 而任一训练样本 x_i 关于以 m_k 为类中心的 Γ_k 类有一个确定的测度 μ_{ik} 。随着 $\|x_i - m_k\|$ 的增大而减小 (x_i 到 Γ_k 类的类中心的加权距离越大, x_i 隶属于 Γ_k 类的隶属度越小), 因此, 只需要求得其加权距离 D_{ik} , 知道隶属度之间的相对大小即可。

$$D_{ik} = \|x_i - m_k\| = \sqrt{\sum_{j=1}^d \omega_j \cdot (x_{ij} - m_{kj})^2} \tag{2}$$

$$\mu_{ik} = (D_{ik} + \varepsilon)^{-1} \cdot \sum_{l=1}^c (D_{il} + \varepsilon) \tag{3}$$

$$k = 1, 2, \dots, C$$

式中 ε 为控制常数, 作用是调整当 D_{ik} 过小时对隶属度的影响; 而各种特征的分类权重 ω_i 的求解方式如下:

$$\bar{m}_j = \sum_{k=1}^c m_{kj} / C (j = 1, 2, \dots, d) \tag{4}$$

$$\sigma_j^2 = \frac{a_j}{C} \sum_{k=1}^c (m_{kj} - \bar{m}_j)^2 \tag{5}$$

$$\omega_j = \sigma_j^2 / \sum_{i=1}^d \sigma_i^2 \tag{6}$$

1.2 类中心的确定

在 d 维特征空间中, x_i 关于 Γ_k 类的基本隶属度 μ_{ik} 可以被当做点质量赋予样本点 x_i , 这样就能得到 N 个质点构成的质点组 $\{(x_1, \mu_{1k}), (x_2, \mu_{2k}), \dots, (x_N, \mu_{Nk})\}$, 下一步可依据物理方法确定其质心, 新类中心即为

$$m_k^{(1)} = \frac{\sum_{i=1}^N \mu_{ik} \cdot x_i}{\sum_{i=1}^N \mu_{ik}} \quad k = 1, 2, \dots, C \tag{7}$$

初始类中心被利用测度求得的新类中心替代, 即可建立起求未确知聚类中心向量的迭代算法。经过 t 次迭代, 当 $\max \|m_k^{(t)} - m_k^{(t-1)}\| < \sigma$ 满足时, 即停止迭代, 此时即可输出 C 个类中心 m_k^t ($k=1, 2, \dots, C$) 和各样本属于各类的隶属度 μ_{ik} 。经过多次修正后得到的类中心向量能够将该类数据的典型特征很好的体现出来, 即可视为新的标准状态。

2 实例研究

一般来说, 土壤重金属指的是比重大于或等于 5.0 g/cm^3 的重金属 (如 Cr、Cd、Pb、Cu、Zn), 土壤重金属污染是指由于人类活动将重金属污染加入到土壤中, 致使土壤重金属含量明显高于原有含量, 并造成生态环境质量恶化的现象^[1]。考虑到数据获取的难易程度, 文中仅选取 Cd、Hg、As、Cu、Pb、Cr6 种土壤中最常见的重金属污染物为评价因子, 根据我国《土壤环境质量标准》(GB156182 - 1995) 对土壤各采样点进行分类^[8-9], 具体评价标准见表 1。I 级标准为清洁, 即保护区域自然生态、维持自然背景的土壤质量的限制值; II 级标准轻度污染, 即为保障农业生产, 维护人体健康的土壤限制值; III 级标准为重度污染, 即保障农林生产和植物正常生长的土壤临界值^[10]。

表 1 土壤环境质量标准 (mg/Kg)

Tab. 1 Environmental quality standard for soil (unit: mg/Kg)

重金属	I 级	II 级	III 级
Cd	0.3	0.6	1
As	40	30	25
Hg	0.3	0.5	1
Pb	250	300	350
Cu	50	100	100
Cr	150	200	250

笔者借助参考文献^[11]中的数据, 以江苏省南京市六合区 6 个乡的土壤环境质量为研究对象^[12], 利用未确知聚类对其分类。其中, 各采样点土壤利用方式为蔬菜用地, 主要污染类型为农药化肥污染, 具体数据见表 2。

针对表 2 中的实测数据, 对未确知聚类算法运用 Matlab7.0 编程, 将样本分为三类, 分类精度 $\varepsilon = 0.001$, 最终得到的样本分类结果及隶属度如表 3 所示, 各类的类中心见表 4。表 3 显示, 样本 6

属于第 I 类;样本 2,3,5 属于第 II 类;样本 1,4 属于第 III 类。

表 2 土壤各重金属元素含量的分析数据(mg/Kg)

Tab. 2 Content of heavy metal elements in soil samples (unit:mg/Kg)

序号	Cd	Hg	As	Cu	Pb	Cr
1	0.202	0.304	10.701	63.340	33.420	86.260
2	0.453	0.586	8.840	24.198	21.613	64.024
3	0.767	0.303	8.090	22.686	24.131	73.160
4	0.728	0.980	11.550	25.277	21.969	76.921
5	0.754	0.304	8.720	15.776	19.221	41.918
6	0.388	0.357	9.410	24.108	23.927	64.550

表 3 样本的分类结果

Tab. 3 The classification results of the sample

序号	I 级	II 级	III 级	结果
1	0.199 5	0.199 4	0.601 1	III
2	0.480 0	0.480 6	0.039 4	II
3	0.447 5	0.447 8	0.104 6	II
4	0.246 4	0.246 3	0.507 4	III
5	0.442 8	0.443 1	0.114 1	II
6	0.480 3	0.479 7	0.039 9	I

表 4 各类的类中心

Tab. 4 The centers of the clusters

	Cd	Hg	As	Cu	Pb	Cr
I	0.511 3	0.211 7	0.714 9	0.210 2	0.279 5	0.519 7
II	0.644 8	0.211 8	0.333 0	0.210 2	0.279 5	0.519 8
III	0.645 0	0.375 4	0.332 9	0.520 2	0.537 0	0.793 0

3 实验结果分析

针对得到的实验结果,主要从类中心、隶属度以及分类结果三方面进行分析。

1)从类中心(表 4)来看,很容易发现,在 6 种重金属污染物中,只有 As 含量是越高越好,其余五种均是越低越好,这与土壤环境质量标准(表 1)中是一致的,说明未确知聚类能够应用到土壤环境质量评价中。

2)从各样本点以及各类的隶属度(表 3)来看,当某样本点处于某两类之间,其隶属度相差并不大,或者说很难区分,比如样本点 2,对 I 类和 II 类隶属度相差只有 0.000 6,是很微小的,这是因为在求 $y_{ii} - m_{ii}$ 时总体相差并不大,唯一不同的是各指标权重 ω_i 的取值不同,才导致其属于各类的隶属度有微小偏差。

3)从分类结果上来看,样本 6 属于第 I 类;样本 2,3,5 属于第 II 类;样本 1,4 属于第 III 类,与当地实际情况一致,证明未确知聚类能够对土壤

环境质量进行准确评价。同时,未确知聚类不仅给出了分类结果,也给出了各样本相对于各类的隶属度,这可以为下一步改善蔬菜产地环境质量提供一个指导方向。

4 结语

分析实验结果,与实际情况相符,证明此方法能够在蔬菜产地环境质量评价中应用,且该算法同时能够保证蔬菜产地环境质量评价更加精准化和定量化。将未确知聚类应用到土壤环境质量评价中,不仅拓宽了未确知聚类算法的应用领域,同时也为土壤环境质量评价提供了一种新的思路。

参考文献:

- [1] 徐鸿志.安徽省主要土壤重金属污染评价及其评价方法研究[D].合肥:安徽农业大学,2007.
- [2] 朱雷,陈威.模糊指数法在水质评价中的应用[J].武汉理工大学学报,2001,23(8):61-65.
- [3] 吴传星,伍钧,杨刚,等.基于改进 AHP 法重金属污染因子权重的确定及其在农作物重金属安全性评价中的应用[J].四川农业大学学报,2010,28(3):346-347.
- [4] 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(103):34-36.
- [5] 余立斌,张江山,王菲凤.熵权物元分析模型在土壤环境质量评价中的应用[J].黑龙江环境通报,2008,32(1):70-74.
- [6] 李朝奎,王利东.土壤重金属污染评价方法研究进展[J].矿产与地质,2011,25(2):172-176.
- [7] 曹庆奎,任向阳,刘开弟.矿井工作面瓦斯涌出量的未确知聚类研究[J].煤炭学报,2006,31(3):337-341.
- [8] 王国庆,骆永明,宋静,等.土壤环境质量指导值与标准研究 I—国际动态及中国的修订考虑[J].土壤学报,2005,42(4):666-673.
- [9] 夏家淇,骆永明.我国土壤环境质量研究几个值得探讨的问题[J].生态与农村环境学报,2007,23(1):16.
- [10] 刘蕾,姜灵彦.基于可拓物元法的土壤重金属污染程度评价[J].土壤,2013,45(3):517-521.
- [11] 朱青,周生路,孙兆金,等.两种模糊数学模型在土壤重金属综合污染评价中的应用与比较[J].环境保护科学,2004,30(123):53-56.
- [12] 陈建勇,张江山,郑育毅.基于熵权的物元分析法在土壤重金属污染评价中的应用[J].安全与环境工程,2011,18(5):57-60.

(责任编辑 刘存英)