文章编号:1006 - 0464(2004)02 - 0140 - 04

主成分分析法用于土壤样品分类

李芳清[1,2],倪永年[1],杨水平[2]

(1. 南昌大学 化学系,江西 南昌　330047;2. 东华理工学院 化学系,江西 抚州　344000)

摘　要:采用火焰原子吸收光谱法对江西省南丰县南丰蜜桔种植区和非种植区土壤中锰、铜、锌、钴、镍 5 种元素的含量进行测定;用电位滴定法测定上述土壤样品的酸度;用灼烧法测定其中有机物质的含量;用干燥法测定自然水的含量。进而采用主成分分析法(PCA)对所测得的土壤样品的 8 个变量进行分类研究,发现蜜桔种植区和非种植区的样品能得到较好的分离识别,证明南丰蜜桔的质量与土壤中的微量元素的含量、有机物质的含量以及土壤的酸度之间存在相关性。

关键词:主成分分析;土壤;分类

中图分类号:O657.32　　**文献标识码**:A

主成分分析是一种常用的多元统计分析技术,是模式识别中的一种降维映射方法,该法将多维空间的信息在低维(二维或三维)空间表现出来,以消除众多化学信息中相互重叠的部分。它将原始变量进行转换,使少数几个新变量为原变量的线性组合,同时又使这些变量尽可能多地表征原变量的数据结构特征而不丢失信息。这种方法已被广泛应用于食品分类[1,2]。Gomez - Plaza 等[3]对不同葡萄产区的葡萄酒中的微量无机元素、有机物含量和组成等进行主成分分析,以区分不同产地的葡萄酒。Yoon 等[4]应用主成分分析将不同地区来源的蜂蜜进行了分类。Dupuy 等[5]用主成分分析法对不同产地的向日葵油、橄榄油、花生油进行了分类。

1　方法原理

1.1　主成分分析法

对于一个有 n 个变量 x_1, x_2, …, x_n 的多维空间,主成分分析法可构造 n 个新变量,称为主成分 t_1, t_2, …, t_n。主成分必须满足以下条件:(1) 每个主成分是各个原始变量的线性组合;(2) 各个主成分之间为正交关系,即相关系数为 0;(3) 经线性变换得到的 t_1 的方差为最大,t_2 次之,依次类推。t_1, t_2, …, t_n 称为第 1,2, …, n 个主成分,与它们对应的特征值为 λ_1, λ_2, …, λ_n,且有 $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$。一般而言,方差大的主成分含较多原始变量的信息。因此,t_1 所含原变量的信息最大,t_2 次之,其余类推。如果前 m 个主成分的方差贡献已足够大,且可基本反映原变量的信息,则原多维空间的大部分信息可由前面 m 个主成分组成的低维(二维或三维)空间表现出来。主成分是原始变量的线性组合,如原始变量之间的量纲不同,则此线性组合的意义就比较模糊,较难给出合理的解释。本工作在进行统计分析前将测试数据作标准化预处理,使各个变量的均值为 0,均方差为 1,以消除由原始变量的量纲不同而对主成分分析的结果产生影响。

2　实验部分

2.1　仪器与试剂

WYX - 9003A 原子吸收分光光度计(沈阳分析仪器厂)锌标准储备溶液(100 μg/mL):称取 0.100 0 g锌粒(99.99 %),用 20 mL HCl(1∶1)溶解,移入 1 L 容量瓶中,用去离子水稀释至刻度,移入聚乙烯塑料瓶中贮存备用;铜标准储备溶液(1 mg/mL):称取 1.000 g 金属铜(99.8 %)溶解于 15 mL HNO_3(1∶1)中,移入 1L 容量瓶,稀释至刻度;钴标准储备溶液(1 mg/mL):称取 1.000 g 钴粉(大于 99.4 %),溶解于 10 mL HNO_3(1∶1)中,移入 1 L 容量瓶,并稀释至刻度;镍标准储备溶液(1 mg/mL):称取 1.000 g 光谱纯镍粉溶解于 10 mL HNO_3(1∶1)中,移入 1 L 容量瓶,并稀释至刻度;锰标准储备溶液(1 mg/mL):称取 2.749 g $MnSO_4$ 溶解于 10 mL HCl(1∶1)中,移入 1 L 容量瓶中,并稀释至刻度。

2.2　实验方法

2.2.1　土壤样品的制备　在南丰县的蜜桔种植区

收稿日期:2003 - 09 - 15
基金项目:国家自然科学基金资助项目(20065002);江西省自然科学基金资助项目(0320014)和江西省教育厅科研基金资助项目(03 - 12)
作者简介:李芳清(1971 -),女,讲师.

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

和非种植区(共 20 个点)进行土壤采样。将土壤样品置于阴凉、通风处晾干,平铺于硬质白纸板上,用玻棒压散(勿破坏自然粒径),用玛瑙研钵研磨至样品全部通过 100 目筛。

2.2.2　土壤中含水率,活性酸(pH)和有机质的测定　称取已风干的土壤样品 5.00 g(2 - 3 份)于已经恒重的称量瓶中,放入(105 ±2) ℃烘箱中烘 4h 后取出,放入干燥器中冷却,30 min 后称重,重复烘干 30 min,干燥器放置至恒重。计算土壤含水率(%) = (风干样重 -烘干样重)/风干样重 ×100 %,各样品含水率见表 1。采用电位滴定法测定各样品活性酸。将土壤样品置于高温炉中灼烧(350 - 400 ℃),直接用称重法测出土壤中有机质含量,该法适用于砂质土壤。结果均列于表 1 中。

2.2.3　土壤样品的消解和样品中微量金属元素的测定　称取已风干的土壤样品 0.500 0 g 置于 150 mL 三角烧瓶中,用去离子水润湿样品,慢慢加入王水(浓硝酸 5 mL 和浓盐酸 15 mL 混合而成) 20 mL,并同时做空白实验。将装有样品及试剂空白的三角瓶置于用变压器控制温度的电热板上,加热分解。开始时用低温,逐渐提高温度,但温度不宜过高,以防样液溅出,液面始终保持在微沸状态(140 - 160 ℃),让其充分分解。当反应完毕,大部分有机物分解后,取下三角瓶,冷却后沿着瓶的四周分别加入 10 - 20 mL 高氯酸(视样品中有机质含量而定),继续加热,赶走残留的高氯酸,当样液很少时,可适当降低加热温度,至高氯酸白烟冒尽。样品发白近干,取下三角瓶,冷却至室温,加入约 20 mL 1 %的稀硝酸,继续加热溶解,冷却后用中速定量滤纸过滤于 100 mL 容量瓶中,三角瓶、滤纸和沉淀物用 1 %的稀硝酸洗涤,并将洗涤液并入容量瓶中,稀释至刻度定容。

利用火焰原子吸收光谱方法对经预处理后的样品进行分析,测定最佳工作条件见表 2,测得的各个样品中锰、铜、锌、钴和镍的含量见表 1。

3　结果与讨论

3.1　样品消解方法的选择

分别采用硝酸、硫酸、硝酸 - 盐酸、硝酸 - 氢氟酸、硝酸 - 氢氟酸 - 高氯酸等 5 种方法对土壤样品进行消解,从试验结果可知最后一种方法消解最完全,故本工作采用硝酸 - 氢氟酸 - 高氯酸消解土壤样品。

3.2　共存离子的影响

采用原子吸收光谱法测定锰、铜、锌、钴和镍时,很少有化学干扰,一般的共存元素在浓度不太大时(< 1.0 g/L)也不影响测定。含量大于 4 %的硫酸根对铜和锌有分子吸收而产生正干扰; 大量硅酸根

表 1　样品测定结果 (平均值,金属含量均折算为 μg/g)

样品号	酸度	有机质	铜	钴	锰	镍	锌	含水率
1	5.65	3.066 6	308.6	17.49	479.09	37.92	337.91	2.985
2	6.01	2.631 0	96.13	19.01	372.48	24.91	147.03	2.188
3	6.08	2.596 9	165.27	15.71	339.73	26.12	236.87	2.031
4	6.92	3.028 1	96.08	20.91	467.61	12.83	173.33	2.850
5	6.59	3.605 5	57.92	12.17	249.49	18.59	158.37	2.159
6	6.14	3.595 2	146.79	21.08	439.20	27.36	216.30	4.627
7	6.83	2.152 7	180.6	17.71	397.76	23.23	101.00	3.420
8	6.09	4.855 1	353.49	41.15	308.13	28.37	205.43	3.704
9	5.90	3.614 2	352.83	17.63	267.47	18.53	80.07	2.755
10	6.25	3.365 0	84.57	41.80	365.57	56.16	186.09	3.140
11	8.20	2.823 0	245.72	35.21	342.35	15.82	242.19	2.503
12	7.80	1.965 0	380.54	18.68	423.52	21.15	310.52	2.224
13	6.50	2.218 0	280.48	44.20	565.25	20.45	150.68	2.021
14	4.20	2.726 6	375.62	19.21	502.15	25.42	265.22	1.325
15	7.50	2.956 0	298.66	16.52	456.58	35.64	316.32	3.126
16	6.92	3.823 0	346.68	15.36	302.14	22.16	191.26	1.085
17	7.20	1.202 2	146.29	38.22	460.81	15.66	56.24	1.229
18	4.60	3.542 0	336.20	25.82	650.26	16.32	92.13	3.215
19	6.70	3.156 0	224.12	41.20	642.10	42.15	120.36	2.786
20	4.30	2.104 8	220.53	22.62	720.60	26.52	350.83	1.365

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

表 2 仪器最佳工作条件

元素	波长 / nm	灯电流 / I/ mA	光谱通带 / nm	乙炔流量 / m^3/ h	空气流量 / m^3/ h	狭缝宽度 / nm
Cu	324.8	3	0.5	0.04	0.35	0.32
Zn	213.9	5	0.5	0.04	0.35	0.32
Mn	279.5	8	0.2	0.04	0.35	0.32
Ni	232.0	8	0.2	0.04	0.35	0.15
Co	240.7	5	0.2	0.04	0.35	0.32

离子对锌会产生负干扰;大量铁、钙、钾、钠盐会使镍的吸光度增加而产生背景吸收。在我们的工作中,这些干扰均可采用氘灯或塞曼效应背景校正器扣除。镍的灵敏线 232.0 nm 附近有别的吸收线产生干扰,可选择最小的光谱通带予以克服。

3.3 主成分分析法处理结果

样品的测试数据(表 1)经主成分分析法处理后可得到各样品的主成分(见表 3),前 3 个主成分对应的特征值、方差贡献率及各原变量在各主成分中的系数列于表 4 中。可以看到,主成分 1 与主成分 2 这两个分量已包含样本的较大部分信息(累计方差贡献率为 52.8 %),含水率、有机质和微量元素锰在主成分 1 中起决定作用;微量元素镍、pH 值、微量元素锰和锌对主成分 2 贡献显著。由主成分 1 与主成分 2 构成的二维判别平面图见图 1,图中每个点都由原 8 维空间的样本点降维映射而来,南丰地区的土壤样品与其他地方土壤有明显的区别而“类聚”在一起。含水率、有机质和微量元素锰这三个变量聚聚集于南丰土壤旁成一“簇”,它们在南丰土壤中有较高的含量,即这些变量是决定南丰土壤与其他地域土壤分类的主要因素,因而适当提高土壤中的有机质和微量元素锰含量对蜜桔的种植是有利的。而微量元素钴、锰和锌对主成分 3 也起着重要的作用,由主成分 1 与主成分 3 构成的二维判别平面如图 2。从表 3 中可见主成分 1、主成分 2 和主成分 3 的累计方差贡献率为 70.8 %,由其构成的三维空间图如图 3,可以看到在三维空间两种土壤样品之间同样有“类聚”现象。

5 结 语

对于土壤分析中所获的大量数据,采用数理统计方法进行处理,能充分利用所得的数据中的信息。由以上分析结果可见,经过主成分分析,原来多个指标(变量)可转化为少数几个互不相关的综合指标,揭示了所考察指标之间的关系,因此采用主成分分析法对土壤样品进行分类识别是可行。根据本文结果可知,南丰蜜桔土壤和其他地方的土壤一般都能采用主成分分析法区分,但也有部分区域相互重叠,这说明南丰县郊区有些蜜桔种植土壤已经有品质下降的趋势,有些已接近并非种植区的土壤,这对南丰蜜桔的栽培是不利的,应引起充分注意。

表 3 各样品的得分主成分

样品	主成分 1	主成分 2	主成分 3
1	- 0.132 4	1.909 9	- 0.949 8
2	- 0.042 1	- 1.001	0.129 9
3	- 0.345 6	- 0.552 6	- 0.735 4
4	0.087 9	- 1.038 3	0.077 9
5	0.936 6	- 1.749 1	- 1.040 1
6	1.762 9	0.699 4	- 0.108 9
7	0.309 5	- 0.995 3	0.276 2
8	2.447 2	1.227 2	- 0.324 8
9	0.874 0	- 0.716 1	- 1.271 9
10	2.094 3	1.159 1	1.962 0
11	0.394 2	- 1.313 6	0.068 5
12	- 1.134 6	- 0.439 5	1.072 0
13	- 1.068 8	- 0.216 2	1.742 0
14	- 2.013 5	1.357 7	- 1.052 2
15	0.359 0	0.796 4	- 0.845 7
16	- 0.064 8	- 0.777 7	- 1.788 4
17	- 1.512 0	- 2.474 9	2.116 8
18	- 0.428 2	1.143 8	0.297 9
19	0.478 7	1.120 5	2.319 2
20	- 3.002	1.860 1	0.198 9

表 4 各变量在各主成分中的载荷系数

	主成分 1	主成分 2	主成分 3
pH 值	0.232 8	- 0.454 7	0.065 2
有机质	0.547 6	0.259 3	- 0.282 9
铜	- 0.164 9	0.303 0	- 0.344 7
钴	0.125 9	0.097 8	0.676 0
锰	- 0.436 4	0.413 0	0.382 2
镍	0.261 4	0.483 7	0.240 9
锌	- 0.190 6	0.403 8	- 0.363 8
含水率	0.554 8	0.238 5	0.058 2
累计方差贡献率/ %	32.3	52.8	70. 8

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

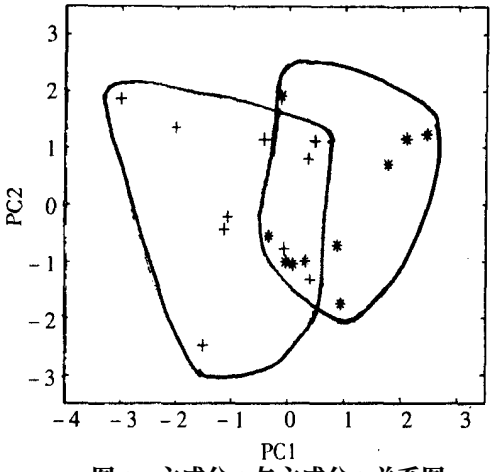

图 1　主成分 1 与主成分 2 关系图

+ :南丰土壤; * :非种植区土壤

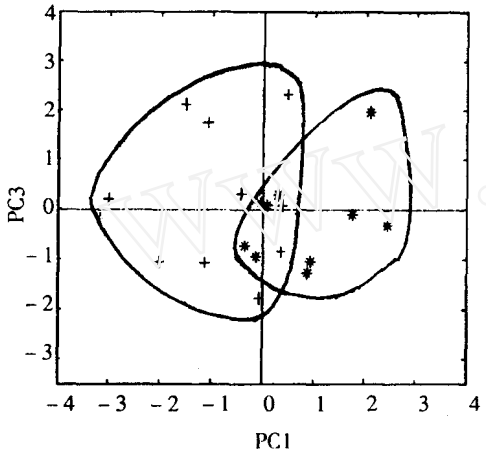

图 2　主成分 1 与主成分 3 关系图

+ :南丰土壤; * :非种植区土壤

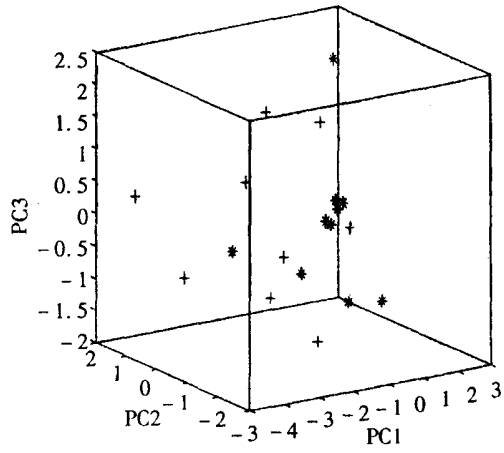

图 3　主成分分析三维图

+ :南丰土壤; * :非种植区土壤

参考文献:

[1] Show P E, Moshonas M G, Buslig B S, et al. Discriminant and PCA to Classify Commercial Orange Juice Based on Relative Amount of Volatile Juice Constituents [J]. J Sci Food Agric, 1999, 79(13): 1 947 - 1 953.

[2] Latorre M J, Pena R, Pita C, et al. Chemometric Classification of Honeys According to Their Type: Metal Content Data [J]. Food Chem, 1999, 60(2): 263 - 268.

[3] Gomez - Plaza E, Gil - Munoz R, Martinez Cutillas A. Multivariate Classification of Wines from Seven Clones of Monastrell Grapes[J]. J Agric Food Chem, 2000, 80(4): 497 - 501.

[4] Yoon J H, Bae S Y, Kim K, et al. Chemometric Aspects and Determination of Sugar Composition of Honey by HPLC [J]. Anal Sci, 1997, 10(5): 362 - 369.

[5] Dupuy N, Duponchel L, Huvene J P, et al. Classification of Edible Fats and Oils by Principal Component Analysis of Fourier Transform Infrared Spectra [J]. Food Chem, 1996, 57(2): 245 - 251.

CLASSIFICATION OF THE SOIL SAMPLES BY APPLICATION OF PRINCIPAL COMPONENT ANALYSIS

LI Fang-qing[1,2], NI Yong-nian[1], YANG Shui-ping[2]

(1. Department of Chemistry, Nanchang University, Nanchang 330047, China;
2. Department of Chemistry, East China College of Teachnology, Fuzhou 344000, China)

Abstract: The classification of soil samples obtained from orange farm Nanfeng countryside in Jiangxi Province, as well as samples obtained from other area (total 20 samples) has been studied. Five metal elements (copper, zinc, cobalt, nickel and manganese) were determined with flame atomic absorption spectrometry (FAAS), the acidity of soil was determined by direct potential method, the moisture of the soil was determined by the drying method, and the organic contents was determined by burn method. The pattern recognition based on the principal component analysis (PCA) was applied to the measured data of the soil samples. The result indicated that abundance statistics characters can be obtained and the samples obtained from different area can also be distinguished effectively when the variables were selected appropriately.

Key words: principal component analysis; soil; classification

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net