

马海蛟,崔晨风. 基于可见-近红外光谱和神经网络的土壤类型鉴别[J]. 江苏农业科学,2014,42(4):284-286.

# 基于可见-近红外光谱和神经网络的土壤类型鉴别

马海蛟, 崔晨风

(西北农林科技大学水利与建筑工程学院,陕西杨凌 712100)

**摘要:**提出一种利用近红外光谱快速、无损鉴别土壤种类的方法。首先利用近红外光谱仪测定不同种类土壤的光谱特征曲线,利用主成分分析法提取主成分,再结合人工神经网络建立模型进行类型鉴别。主成分分析表明,主成分 1、2、3 的累积方差贡献率达到 99.839%,可以很好地代表原始数据特征。以主成分分析得到的前 3 个主成分作为神经网络输入,以土壤类型为输出,通过对 30 个样本的训练学习,分别建立了反向传播人工神经网络(BP)和径向基函数人工神经网络(RBF)。对 10 个样本进行预测,结果表明 2 种模型预测的准确性均达到 100%。RBF 神经网络运行时间明显小于 BP 网络,具有一定优势。

**关键词:**可见-近红外光谱;土壤类型;主成分分析;BP 神经网络;RBF 神经网络;分类鉴别

**中图分类号:** S155;O657.33 **文献标志码:** A **文章编号:** 1002-1302(2014)04-0284-03

光谱技术作为一种简单、快捷、非接触、非破坏的分析方法,已经被广泛地应用到土壤和植物研究、食品行业等领域<sup>[1-4]</sup>。由于土壤的光谱反射率能综合反映土壤理化特征和内在结构,所以土壤光谱测量为土壤光谱特征分析、土壤分类提供了新途径<sup>[5-6]</sup>。目前已有基于土壤光谱特征对土壤各项生理指标参数的研究<sup>[7]</sup>,但是基于土壤近红外光谱和神经网络的土壤分类研究较少<sup>[8]</sup>。刘家雄<sup>[9]</sup>、付强等<sup>[10]</sup>分别利用土壤理化指标,结合主成分分析、聚类分析、神经网络等方法对土壤分类进行研究。由于土壤理化性质测量方法复杂,工程

量大,无法快速、无损地对土壤进行分类。曾庆猛等利用土壤近红外光谱和聚类分析对土壤分类进行了研究<sup>[11]</sup>,宋海燕等利用近红外光谱和正交信号-偏最小二乘法对土壤进行分类,但分类精度最高仅达到 85%<sup>[12]</sup>。

神经网络模型作为一种强有力的学习系统,模拟了人脑思维过程,能够实现输入层、输出层的非线性映射,已经在很多领域得到广泛应用,并取得良好效果<sup>[13-14]</sup>,但目前还没有基于土壤近红外光谱和神经网络[反向传播人工(BP)神经网络和径向基函数人工(RBF)神经网络]的土壤分类研究。本研究采用新疆土、垆土、沙土、阜阳土等 4 种类型土壤的室内可见-近红外光谱反射数据为研究对象,利用主成分分析方法,结合 BP 神经网络和 RBF 神经网络技术进行土壤分类研究,探讨基于土壤反射光谱特性的土壤分类技术,以期提高土壤分类精度和运行时间,为实现基于遥感方法进行土壤快速分类奠定基础。

## 1 材料与方法

### 1.1 仪器设备

使用美国 ASD(analytical spectral device)公司的 Hand-held Field Spec 型光谱仪,其光谱测定范围为 325~1 075 nm,

收稿日期:2013-08-25

基金项目:国家自然科学基金-新疆联合基金(编号:U1203182);国家自然科学基金(编号:51279166);高等学校学科创新引智计划(编号:B12007);水资源与水电工程科学国家重点实验室项目(编号:2011B083);西北农林科技大学大学生创新创业训练计划(编号:1210712077)。

作者简介:马海蛟(1990—),女,甘肃武威人,研究方向为“3S”技术在农业水土工程中的应用。E-mail:18700807193@163.com。

通信作者:崔晨风,博士,讲师,研究方向为“3S”技术在农业水土工程和水文水资源中的应用、大坝安全监测、精密工程测量。E-mail:cuichenfeng@163.com。

[3] 宋开山,刘殿伟,王宗明,等. 三江平原过去 50 年耕地动态变化及其驱动力分析[J]. 水土保持学报,2008,22(4):75-81.

[4] 孙国军,刘普幸,李慧芳,等. 酒泉市耕地变化的驱动力研究及多元线性回归预测[J]. 干旱区资源与环境,2009,23(3):73-76.

[5] 李穗英,孙新庆. 青海省近 10 年耕地面积动态变化及驱动因子分析研究[J]. 中国农业资源与区划,2009,30(5):39-44.

[6] 郭洪海,宋 民. 山东省耕地动态变化趋势及驱动力分析[J]. 中国农业资源与区划,2009,30(1):51-57.

[7] 曾永年,吴孔江,靳文甦,等. 青海高原东北部耕地变化及驱动力分析[J]. 干旱区资源与环境,2012,26(8):89-92.

[8] 杨 朔,李世平. 关中地区城市化过程中土地利用问题研究[J]. 中国土地科学,2009,23(7):79-80,封三.

[9] 杨 勇. 关中地区土地利用动态模拟与优化配置研究[D]. 西安:陕西师范大学,2010.

[10] 杨忠泰. 陕西培育发展战略性新兴产业的路径选择[J]. 科技管理研究,2013,33(6):72-77.

[11] 曾 波,刘思峰. 基于灰色关联度的小样本预测模型[J]. 统计与信息论坛,2009,24(12):22-26.

[12] 刘思峰,蔡 华,杨英杰,等. 灰色关联分析模型研究进展[J]. 系统工程理论与实践,2013,33(8):2041-2046.

[13] 东亚斌,段志善. 灰色关联度分辨系数的一种新的确定方法[J]. 西安建筑科技大学学报:自然科学版,2008,40(4):589-592.

[14] 熊和金,徐华中. 灰色控制[M]. 北京:国防工业出版社,2005:44-47.

[15] 焦李成. 神经网络系统理论[M]. 西安:西安电子科技大学出版社,1990:8-11.

[16] 沈花玉,王兆霞,高成耀,等. BP 神经网络隐含层单元数的确定[J]. 天津理工大学学报,2008,24(5):13-15.

采样间隔为 10 nm,光谱分辨率为 3.5 nm,探头视场角为 15°。光源采用与光谱仪配套的 14.5 V 卤素灯。采用 ASD View Spec Pro 和 Matlab R2009b 软件分析数据。

## 1.2 样本采集与制备

供试土壤分别为新疆土、埭土、沙土、阜阳土。每种土壤各取 10 个样本,每个样本扫描 20 次,取平均值作为最终光谱反射率数据。

## 1.3 主成分分析(PCA)原理

PCA 是一种降维映射方法,可以有效地压缩数据特征,把原有多维空间信息通过一系列线性变换在低维空间表现出来,从而消除众多信息中的重叠部分。提取的每个特征都是原来特征的函数,使新得到的特征数少于原有特征数,同时保存了原有特征的主要信息。

## 1.4 神经网络原理

**1.4.1 BP 神经网络** BP 神经网络是目前应用最广泛的神经网络方法之一,具有较强的信息处理能力,能够实现输入和输出之间的非线性映射。典型的神经网络具有 1 个输入层、数个隐含层、1 个输出层,层与层之间采用全连接的方法,同层神经元之间不存在相互连接。理论上已证明,1 个具有隐含层的 3 层网络可以逼近任何连续的非线性函数。隐含层中神经元多采用“S”型函数(tansig),输出层的神经元多采用线性传递函数(purelin)。

传统的 BP 算法存在很多未解决问题,如收敛速度较慢,易于陷入局部极小等。在实际应用中,该算法存在网络结构参数和学习训练参数难以确定的问题,这在一定程度上影响了神经网络的推广应用。

本研究采用 3 层 BP 神经网络,该网络由 1 个输入层、1 个隐含层、1 个输出层组成,隐含层采用“S”型函数,输出层采用线性传递函数,目标误差 0.01,设定训练迭代次数 1 000 次。

**1.4.2 RBF 神经网络** 多变量插值的 RBF 神经网络具有优秀的离散数据内插特性,可以提供最优逼近功能,其网络结构域与多层前向型网络类似,是一种 3 层前向型网络,由输入层、隐含层、输出层组成,隐含层神经元传递函数是对中心点径向对称且衰减的非负非线性函数(radbas),从输入层空间到隐含层的空间变换是线性的,从隐含层空间到输出层空间变换也是线性的。RBF 具有结构简单,训练速度快,函数逼近能力和分类能力强,不存在局部最优问题等特点。

本研究采用 3 层 RBF 神经网络,径向基函数扩展速度(spread)为 5.5。

## 2 结果与分析

### 2.1 可见-近红外光谱分析

图 1 为光谱波段为 325 ~ 1 075 nm 的土壤光谱反射率。从图 1 可以看出,不同类型土壤因为内部各种属性不同,表现出来的光谱反射率有明显差异,这为不同类别土壤的鉴别提供了数学基础,可以用土壤光谱特性对其进行分类识别。

### 2.2 主成分分析

光谱波段从 325 ~ 1 075 nm 共有 751 个点,去除噪声较大的波段,选择可见-近红外波段 400 ~ 899 nm,共有 500 个点。如果采用 500 个点进行计算,则计算量大,会导致神经网络训练时间过长等问题。所以先通过主成分分析,提取新的

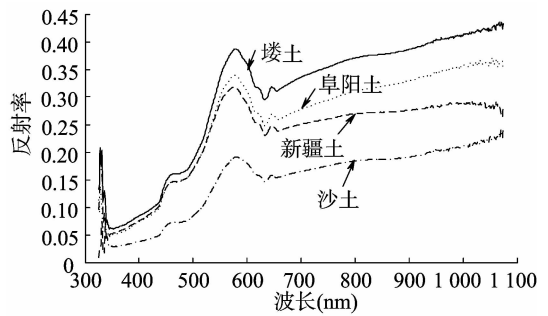


图1 不同类型土壤的光谱反射率

变量作为建立神经网络鉴别模型的输入集,这样可以简化模型,提高运算速度和分类准确度。

通过计算提取出 3 个主成分,由表 1 可知,前 3 个主成分可以代表原始数据信息的 99.839%。

表 1 主成分累积方差贡献率

主成分	累积可信度(%)
PC1	95.980
PC2	99.455
PC3	99.839

由于前 2 个主成分就可以代表原始数据信息的 99.455% (表 1),故采用前 2 个主成分绘制 PC1、PC2 得分图。从图 2 可以看出,所有样本都聚集在第四象限,新疆土、埭土、沙土、阜阳土等 4 种类型土壤大体上可分为 4 类,只有新疆土中的 1 个样本没有很好地聚类,而且每类土壤样本的聚合度也有一定差异,其中阜阳土、埭土聚类呈现明显条状,而新疆土、沙土聚类呈现点状。以上分析说明,通过主成分分析进行聚类还不够精准,须要通过其他模型来进一步鉴别土壤种类。

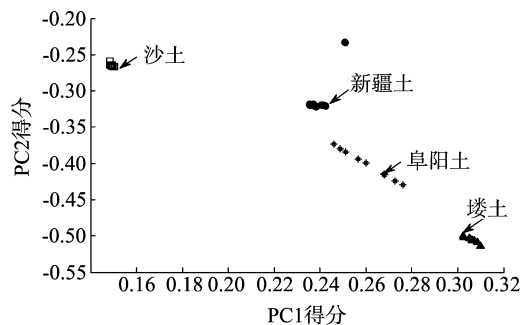


图2 40个样本的PCA得分(PC1×PC2)

### 2.3 BP 神经网络、RBF 神经网络模型分析

将主成分分析提取的前 3 个主成分作为神经网络输入变量建立模型。通过调整网络隐含层节点数和“spread”值来优化网络。通过相对误差、决定系数、运行时间对网络性能进行评价。

相对误差( $E_i$ )计算公式为:

$$E_i = \frac{|\hat{y}_i - y_i|}{y_i}, i = 1, 2, \dots, n \quad (1)$$

式中: $\hat{y}_i$  为第  $i$  个土壤样本的预测值; $y_i$  为第  $i$  个土壤样本的真实值; $n$  为土壤样本数目;1、2、3、4 分别代表新疆土、埭土、沙土、阜阳土。

决定系数( $r^2$ )计算公式为:

$$r^2 = \frac{(\sum_{i=1}^l \hat{y}_i y_i - \sum_{i=1}^l \hat{y}_i \sum_{i=1}^l y_i)^2}{[\sum_{i=1}^l \hat{y}_i^2 - (\sum_{i=1}^l \hat{y}_i)^2][\sum_{i=1}^l y_i^2 - (\sum_{i=1}^l y_i)^2]} \quad (2)$$

相对误差越小,表明模型性能越好。决定系数越接近 1,说明模型性能越好,预测精度越高。

运行时间调用“matlab”函数“t=cputime”来计算。

由于利用不同数据进行模型训练和测试会导致结果不

同,为了避免存在特殊性,在所有样本数据中随机选取 30 个数据作为输入集,剩余 10 个数据作为测试集对网络进行试验。

通过反复试验,得到的最佳网络结构为 3 个输入层、1 个隐含单元、1 个输出层的 3 层 BP 神经网络,“spread”值为 5.5 的 RBF 神经网络,对随机选取的 30 个样本建立模型进行学习,对剩余 10 个样本进行预测,某次运行预测结果见表 2。

表 2 BP 神经网络、RBF 神经网络对 10 个测试集的预测结果

编号	土壤类型	反射率 真实值	反射率			
			BP 神经网络预测值	RBF 神经网络预测值	BP 神经网络预测值误差	RBF 神经网络预测值误差
1	沙土	3	2.988 8	3.008 7	0.003 7	0.002 9
2	新疆土	1	0.970 4	1.002 0	0.029 6	0.002 0
3	阜阳土	4	3.984 9	4.172 0	0.003 8	0.043 0
4	阜阳土	4	3.979 5	4.014 4	0.005 1	0.003 6
5	埭土	2	2.003 5	2.027 6	0.001 7	0.013 8
6	埭土	2	1.961 5	2.025 5	0.019 2	0.012 7
7	沙土	3	2.999 2	3.001 5	0.000 3	0.000 5
8	埭土	2	1.717 1	2.174 3	0.141 4	0.087 1
9	沙土	3	3.013 2	3.023 1	0.004 4	0.007 7
10	新疆土	1	0.970 0	1.006 3	0.030 0	0.006 3

该次运行时,2 种网络模型对土壤类型的预测都有很好的效果,BP 神经网络模型、RBF 神经网络模型预测值与实际值的相关系数分别达到 0.999、0.996,2 种模型可以准确地对土壤类型进行鉴别。

由表 3 可见,RBF 神经网络运行时间明显少于 BP 神经网络,最多时只占到 BP 神经网络所用时间的 13.17%。相对于 BP 神经网络,RBF 神经网络在处理大量数据时有明显优势,能显著减少网络运行训练时间。

表 3 BP 神经网络、RBF 神经网络运行时间

序号	运行时间(s)		RBF 神经网络运行时间/ BP 神经网络运行时间(%)
	BP 神经网络	RBF 神经网络	
1	1.62	0.14	8.64
2	1.42	0.11	7.75
3	1.93	0.11	5.7
4	1.56	0.11	7.05
5	1.47	0.17	11.56
6	1.4	0.12	8.57
7	2.7	0.12	4.44
8	1.72	0.09	5.23
9	1.67	0.22	13.17
10	1.53	0.09	5.88

3 结论

通过试验获得 4 种类型土壤的可见-近红外光谱特征曲线,结合主成分分析和有导师学习型的 2 种神经网络,分别建立了鉴别土壤类型的模型,2 种模型预测效果很好,识别率均达到 100%。其中 RBF 神经网络克服了 BP 神经网络运行时间长的缺点,在应用时更加快捷方便。利用可见-近红外光谱技术可以快速、准确、无损地对土壤类型进行鉴别,为土壤类型鉴别提供了一种可靠的新途径。

参考文献:

[1]何 勇,李晓丽,邵咏妮.基于主成分分析和神经网络的近红外光谱苹果品种鉴别方法研究[J].光谱学与光谱分析,2006,26(5):850-853.

[2]何 勇,李晓丽.用近红外光谱鉴别杨梅品种的研究[J].红外与毫米波学报,2006,25(3):192-194+212.

[3]李晓丽,何 勇,袁正军.一种基于可见-近红外光谱快速鉴别茶叶品种的新方法[J].光谱学与光谱分析,2007,27(2):279-282.

[4]邵咏妮,何 勇,潘家志,等.基于光谱技术的桔子汁品种鉴别方法的研究[J].光谱学与光谱分析,2007,27(9):1739-1742.

[5]黄应丰,刘腾辉.华南主要土壤类型的光谱特性与土壤分类[J].土壤学报,1995,32(1):58-68.

[6]刘焕军,张 柏,张渊智,等.基于反射光谱特性的土壤分类研究[J].光谱学与光谱分析,2008,28(3):624-628.

[7]张娟娟,田永超,姚 霞,等.基于高光谱的土壤全氮含量估测[J].自然资源学报,2011,26(5):881-890.

[8]王遵义,金春华,刘 飞,等.基于光谱技术的土壤快速分类方法研究[J].浙江大学学报:农业与生命科学版,2010,36(3):282-286.

[9]刘家雄.主成分分析与聚类分析在土壤分类中的应用[J].上海农业学报,2011,27(3):110-113.

[10]付 强,王志良,梁 川.自组织竞争人工神经网络在土壤分类中的应用[J].水土保持通报,2002,22(1):39-43.

[11]曾庆猛,孙宇瑞,严红兵.土壤质地分类的近红外光谱分析方法研究[J].光谱学与光谱分析,2009,29(7):1759-1763.

[12]宋海燕,秦 刚,韩小平,等.基于近红外光谱和正交信号-偏最小二乘法对土壤的分类[J].农业工程学报,2012,28(7):168-171.

[13]吴桂芳,蒋益虹,王艳艳,等.基于独立主成分和 BP 神经网络的干红葡萄酒品种的鉴别[J].光谱学与光谱分析,2009,29(5):1268-1271.

[14]杨 芳,李红睿,田学东.基于 RBF 神经网络的汉字粗分类方法[J].计算机工程与应用,2009,45(6):170-172.