

朱淑鑫, 杨 宸, 顾兴健, 等. K 均值算法结合连续投影算法应用于土壤速效钾含量的高光谱分析[J]. 江苏农业学报, 2020, 36(2): 358-365.

doi: 10.3969/j.issn.1000-4440.2020.02.015

K 均值算法结合连续投影算法应用于土壤速效钾含量的高光谱分析

朱淑鑫¹, 杨 宸¹, 顾兴健¹, 张永春², 艾玉春², 徐焕良¹

(1. 南京农业大学信息科技学院, 江苏 南京 210095; 2. 江苏省农业科学院农业资源与环境研究所, 江苏 南京 210014)

摘要: 为解决在土壤速效钾含量的高光谱定量预测分析过程中, 光谱数据维数高、冗余度较大等问题, 提出了一种结合 K 均值算法(K-means)和连续投影算法(SPA)的高光谱特征波段选择方法。该算法首先将全波段数据分别根据不同的距离度量进行 K-means 聚类分析, 之后对聚类后的每个波段簇分别使用 SPA 法提取其中的特征波段。对全波段组合、传统 SPA 法提取的特征波段组合以及结合 K-means 聚类与 SPA 法提取的特征波段组合分别建立土壤速效钾含量的 BP 神经网络预测模型, 通过对比模型预测效果来比较特征波段选择方法的性能。以盐城市 348 份土壤样品进行试验, 结果表明, 结合 K 均值算法与连续投影算法的特征波段选择方法可以有效地解决光谱预测分析过程中的数据冗余问题, 实现对土壤速效钾含量快速精确预测分析。

关键词: 土壤; 高光谱; 连续投影法(SPA); K-means 聚类分析法; BP 神经网络模型

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-4440(2020)02-0358-08

K-means algorithm combined with successive projection algorithm for hyperspectral analysis of soil available potassium content

ZHU Shu-xin¹, YANG Chen¹, GU Xing-jian¹, ZHANG Yong-chun², AI Yu-chun², XU Huan-liang¹

(1. College of Information Technology, Nanjing Agricultural University, Nanjing 210095, China; 2. Institute of Agricultural Resources and Environment, Jiangsu Academy of Agricultural Sciences, Nanjing 210014, China)

Abstract: In order to solve the problems of high dimensionality and redundancy of hyperspectral spectral data in hyperspectral quantitative prediction and analysis of soil available potassium content, a hyperspectral band selection method based on K-means algorithm and successive projection algorithm(SPA) was proposed. Firstly, the full-band data were clustered by K-means based on different distance measures, and then the characteristic bands were extracted by SPA method for each band cluster after clustering. BP neural network prediction models of soil available potassium content were established for full-band combination, combination of characteristic bands extracted by traditional SPA method and combination of characteristic bands extracted by K-means clustering and SPA, respectively. The performance of characteristic band selection methods was evaluated by comparing the prediction effects of the models. The 348 soil samples from Yancheng City were experimented. The characteristic band selection method based on K-means algorithm and successive projection algorithm can

effectively solve the problem of data redundancy in the process of spectral prediction and analysis, and achieve the rapid and accurate prediction and analysis of soil available potassium content.

Key words: soil; hyper-spectrum; successive projection algorithm(SPA); K-means clustering analysis;

收稿日期: 2019-08-19

基金项目: 中央高校基本科研业务费专项资金项目(KYGL201808);
江苏省农业科技自主创新基金项目[CX(17)1001]

作者简介: 朱淑鑫(1978-), 女, 江苏盐城人, 硕士, 副教授, 主要从事
农业信息化与土壤数据研究。(E-mail) zsx@njau.edu.cn

通讯作者: 徐焕良, (E-mail) huanliangxu@njau.edu.cn

BP neural network prediction models

土壤中速效钾含量作为表征土壤中钾元素供应状况的重要指标之一,对农作物的质量和产量具有重要影响^[1]。传统的土壤速效钾含量测定是在实验室中通过化学方法完成的,操作复杂且耗时费力。高光谱遥感技术具有光谱分辨率极高、波段多等优点^[2],能够获得连续的光谱信息,可作为土壤速效钾含量测定的一种有效技术方法,实现更为准确的监测或反演^[3]。在实际运用光谱数据建立模型的过程中,光谱数据往往由于其维度过高、信息冗余度大等特点,而造成模型建立过程复杂度较高、耗时较长等问题。因此,应当选取适合的光谱特征波段提取方法来剔除光谱数据中的无效信息,简化预测模型的建立。

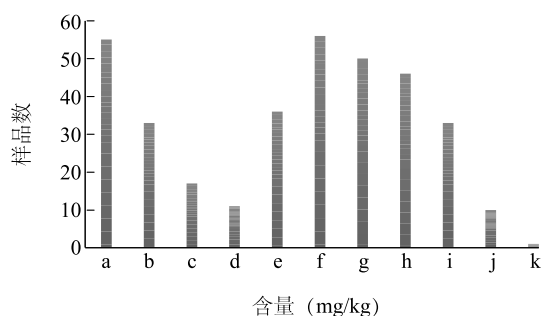
目前国内外学者研究光谱特征波段选取的主要方法有逐步回归法、遗传算法、无信息变量消除法、连续投影法等。连续投影法作为一种使矢量空间共线性最小化的前向变量选取算法,可有效去除众多波段之间共线性影响,极大地减少数据量,有效地提高运算效率和模型精度,具有简便、快速等优点。刘明博等^[3]通过对水稻叶片氮含量光谱检测的研究,讨论了使用连续投影算法选取有效波长的可行性。王武等^[4]采用连续投影法对勾兑梨汁中原汁含量的近红外光谱进行了特征波段提取,并建立定量分析模型。陈定星^[5]采用连续投影法对土壤有机质 NIR 光谱数据进行特征波段选择,建立了土壤有机质快速分析的回归模型。但是传统的连续投影算法往往由于选取初始波段的随机性,造成所提取的特征波段可能为无效信息波段或仍旧携带较多冗余信息,大大降低了预测精度。针对这一问题,本研究采用 K-means 聚类算法和连续投影算法相结合的特征波段选择方法,解决在土壤速效钾含量高光谱定量预测过程中光谱数据冗余信息较多等问题,简化预测模型的复杂度,提升预测精度。

1 材料与方法

1.1 土样采集与制备

采集江苏省盐城市 8 个不同地区的 348 份土壤样品。土壤样品经过风干、碾磨及 2 mm 筛选处理后,分别进行化学分析和光谱数据测定。土样的化学分析在江苏省农业科学院化学实验室中完成,348

份土壤样品速效钾含量测定分析结果为:极大值 504.00 mg/kg,极小值 81.00 mg/kg,均值 293.79 mg/kg,标准差 89.71 mg/kg。对 348 份土壤样品速效钾含量的化学分析结果进行分布统计,绘制对应的分布直方图(图 1)。



a: 81 ~ 120 mg/kg; b: 121 ~ 160 mg/kg; c: 161 ~ 200 mg/kg; d: 201 ~ 240 mg/kg; e: 241 ~ 280 mg/kg; f: 281 ~ 320 mg/kg; g: 321 ~ 360 mg/kg; h: 361 ~ 400 mg/kg; i: 401 ~ 440 mg/kg; j: 441 ~ 480 mg/kg; k: 481 ~ 504 mg/kg。

图1 土壤速效钾含量分布直方图

Fig.1 Histogram of soil available potassium content distribution

采用型号为 InspectorV10E 的高光谱图像采集系统测定土壤样本的光谱数据,该系统包括高感度 EM-CCD 相机(RaptorEMCCD,镜头为 f/1.4,17 mm)、150 W 卤素光源,配有暗箱、散热风扇、穿透及发射光源固定开关、20 cm×20 cm 载物移动台及装有 Spectral-Image 取像软件、HISAnalyzer 分析软件的联想工作站。光谱波段测定范围为 400~1 000 nm,采样间隔 1.1 nm,共 546 个波段。在光谱数据获取过程中,每一份土壤样品选取 5 个 50×50 像素区域作为实际取样点,取 5 个区域的平均值作为每份土样的光谱点数据,共得到 348 份有效土壤样品光谱数据。

1.2 K 均值聚类算法

K 均值聚类算法(K-means)是一种广泛应用于聚类问题的无监督算法^[6],它首先在变量矩阵中随机选取 K 个变量作为初始聚类中心,然后通过迭代计算的过程确定其余每个变量与聚类中心的相似程度,根据相似度的大小将该对象分配给不同的聚类中心,形成不同的聚类簇。更新后的聚类簇以簇内所有变量的平均值作为新的聚类中心,再次计算每个变量与新确定的聚类中心的相似度并进行再次分配,重复此过程直到聚类中心不再发生变化或误差

平方和最小为止^[7]。本研究 K-means 算法的实现流程为:(1)输入 348×546 的光谱反射率矩阵,348 为样品数量,546 为光谱波段数量;(2)从 546 个波段列向量中随机选取 k 个波段作为起始聚类中心;(3)根据相似度衡量指标,依次计算其余波段向量与每个聚类中心的相似度,并根据相似度的大小将其分配给不同的聚类中心;(4)根据分配后的聚类簇计算每个簇内的波段向量平均值作为新一轮的聚类中心;(5)重复第 3、4 步骤直到聚类中心收敛;(6)输出 k 个聚类簇下的波段组合。

由 K 均值聚类算法原理可知,在分配列向量时,不同的相似度度量标准会得到不一样的聚类结果。传统的 K-means 算法通常采用欧式距离作为相似度度量标准,对于光谱数据而言,由于波段之间的相关性影响,简单地采用欧式距离作为相似度指标很难得到最佳的聚类效果。本研究在传统欧式距离标准的基础上,采用加权欧式距离(Squeclidean)、绝对值距离(Cityblock)、余弦相似度(Cosine)和相关距离(Correlation)4 种不同的相似度度量指标分别进行聚类,其计算公式^[8]如下:

$$(1) \text{ 加权欧式距离: } d(x_i, x_j) = \left[\sum_{k=1}^p \frac{(x_{ik} - x_{jk})^2}{s_k^2} \right]^{\frac{1}{2}} \quad (1)$$

$$(2) \text{ 绝对值距离: } d(x_i, x_j) = \sum_{k=1}^p |x_{ik} - x_{jk}| \quad (2)$$

$$(3) \text{ 余弦相似度: } d(x_i, x_j) = \frac{\sum_{k=1}^p x_{ik} x_{jk}}{\sqrt{\sum_{k=1}^p x_{ik}^2} \sqrt{\sum_{k=1}^p x_{jk}^2}} \quad (3)$$

$$(4) \text{ 相关距离: } d(x_i, x_j) = \frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^p (x_{jk} - \bar{x}_j)^2}} \quad (4)$$

1.3 连续投影算法

连续投影算法^[9](Successive projections algorithm, SPA)是一种使矢量空间共线性最小化的前向变量选取算法,可有效去除众多波段之间共线性影响,极大地减少数据量,有效提高运算效率和模型精度,缩减数据建模时间,具有简便、快速等优点。其算法基本原理^[10-11]如下:首先以一个波段为起点,逐次循环,计算它在其他未选入波段上的投影,并将投影向量最大时对应的波段增加到选取的波段组合中,算法循环 N 次结束, N 为需要提取的波长数量。

这样可以确保每个新选入的波段都与前一个具有最小的线性关系。SPA 算法的具体流程如下:(1)初始化, $n=1$ (第 1 次迭代),在光谱矩阵中任选一列向量;(2)集合 S 定义为未被选入波段组合的列向量,分别计算所选的列向量在 S 向量上的投影向量;(3)记录最大投影的序号;(4)将最大投影序号对应的向量作为下轮的投影向量;(5) $n=n+1$,如果 $n < N$,返回步骤(2)继续投影;(6)根据循环次数 N 的不同共得到 N 组波段组合,对不同的波段组合依次建立回归模型,选取均方误差最小的组合作为最优波长组合。

1.4 模型建立

在模型的建立过程中,如果仅使用 1 个样本集,则会导致信息交叉,使得所建立的模型存在说服力不够等问题。本研究采用 Kennard-Stone 法划分建模集与验证集^[12]。Kennard-Stone 法的基本原理是:通过计算样本不同光谱数据间的差异性,将差异较大的划分为建模集,剩余的样本划分为验证集。这样可以保证建模集中所有样本的覆盖范围最广,避免分布不均的问题,提高模型的稳定性。将高光谱测得的 348 份样品数据进行 Kennard-Stone 法划分,共得到 278 份建模集样本,70 份验证集样本。

对于划分后的土壤高光谱反射率数据,采用经典的三层神经网络结构训练最优 BP 神经网络模型对土壤速效钾含量进行预测^[13]。在 MATLAB 2018b 软件中实现,神经网络输入层和隐藏层采用的激活函数为 tansig 函数,设置 10 个隐藏层神经元,输出层的激活函数采用 purelin 函数,学习率设定为 0.01,误差设定为 0.001。

2 结果与分析

2.1 原始土壤光谱数据预处理

测得的 348 份原始土壤光谱数据如图 2 所示。由于测量过程中,高光谱仪受到自身构造、环境因素、人为因素、电子干扰及样品颗粒大小不同等因素的影响,导致获得的光谱数据存在大量噪声、基线偏移等不稳定现象。因此,对光谱数据进行预处理,消除光谱中的噪声影响。采用 9 点 Savitzky-Golay (SG)卷积平滑法^[14]对光谱数据进行平滑预处理,平滑处理后的光谱曲线如图 3 所示。

2.2 预处理后的光谱数据 K-means 算法聚类

对预处理后的光谱数据进行 K-means 聚类。

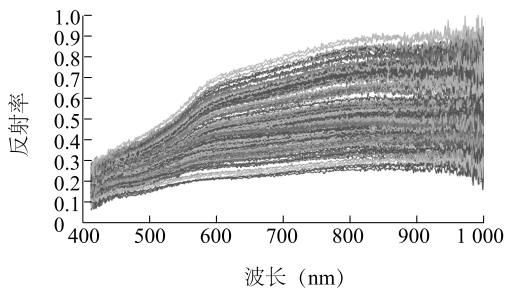


图2 原始光谱数据
Fig.2 Raw spectral data

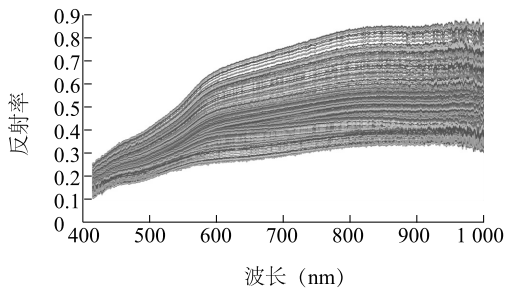


图3 SG 平滑处理后光谱数据
Fig.3 Spectral data after SG smoothing

为研究不同聚类相似度对聚类效果的影响,根据波段向量数量,结合文献[15]中的手肘法确定聚类 k 值为 5,将原始复杂度较高的光谱数据分为 5 个聚类簇,每个簇内的波段列向量间有较高的相似度。分别采用加权欧式距离、绝对值距离、余弦相似度和相关距离作为相似度度量指标,对预处理后的光谱数据进行聚类分析,结果如表 1 所示。其中基于加权欧式距离、绝对值距离、余弦相似度和相关距离这 4 种度量指标下的 K-means 聚类分别简称为 Sqe-kmeans、Cit-kmeans、Cos-kmeans 和 Cor-kmeans。

2.3 基于连续投影算法的波段选择

聚类后的光谱数据每个簇内都由若干个相似度较高的波段组成,此时,对每个簇分别进行连续投影算法选取簇内最优波段组合,对波段进行进一步优选。在使用连续投影算法选择特征波长时,一般认为特征波长选取应为 5~10 个为佳,这样既能使特征波段反映整体信息,又能避免过拟合,故本研究设置特征波段选取数量 (N) 为 10。虽然连续投影算法的初始波段仍具有随机性,但通过与聚类算法的结合,可将随机初始波段的范围限制在相应的聚类簇内,大大降低了最终 SPA

选取结果为无效信息波段的可能。对全波段以及 4 种不同相似度度量下的波段组合应用连续投影算法优选结果如表 2 所示。

表 1 不同相似度度量下的 K-means 聚类结果
Table 1 K-means clustering results under different similarity measures

聚类方式	聚类类别	波段组合	波段数量
Sqe-kmeans	第 1 类	1~75	75
	第 2 类	76~133	58
	第 3 类	134~182	49
	第 4 类	183~297	115
	第 5 类	298~546	249
Cit-kmeans	第 1 类	1~85	85
	第 2 类	86~144	59
	第 3 类	145~221	77
	第 4 类	222~313	92
	第 5 类	314~546	233
Cos-kmeans	第 1 类	1~78、80、82	80
	第 2 类	79、81、83~150、153	71
	第 3 类	151、152、154~252、254~257、259、260	107
	第 4 类	253、258、261~422、424	165
	第 5 类	423、425~546	123
Cor-kmeans	第 1 类	1~78、80	79
	第 2 类	79、81~143	64
	第 3 类	144~249	106
	第 4 类	250~387、389~396、400、401、403、407	150
	第 5 类	388、397~399、402、404~406、408~546	147

Sqe-kmeans、Cit-kmeans、Cos-kmeans 和 Cor-kmeans 分别为基于加权欧式距离、绝对值距离、余弦相似度和相关距离的 K-means 聚类。

2.4 土壤速效钾含量 BP 神经网络预测模型

分别将全波段光谱数据、全波段下直接使用连续投影算法的优选波段和本研究提出的 4 种相似度度量聚类结合连续投影算法的优选波段作为 BP 神经网络的输入,将土壤速效钾含量的预测值作为模型的输出。以划分后的 278 份建模样本作为建模数据,70 份验证样本数据作为模型精度的验证数据,建立 BP 神经网络预测模型,将得到的预测值和实际值进行比较(图 4)。图 4 中每个圆点代表一份土壤样品,直线为预测值和实测值 1:1 直线,圆点与直线距离越近,则表明模型的预测值越接近实测值,预测精度越高。

表 2 不同相似度度量下 K-means 聚类后波段的连续投影算法 (SPA) 优选

Table 2 Successive projection algorithm (SPA) optimization for bands after K-means clustering under different similarity measures

聚类方式	聚类类别	SPA 筛选波段组合	波段数量
Sqe-kmeans	第 1 类	5、6、8、10、24、40	6
	第 2 类	90、101、102、105、128、133	6
	第 3 类	136、137、148、151、153、170、182	7
	第 4 类	191、227、235、254、263、264、297	7
	第 5 类	459、534、546	3
Cit-kmeans	第 1 类	5、6、8、10、24、40	6
	第 2 类	90、92、96、105、110、128、142	7
	第 3 类	146、149、150、151、153、175、183、191、213、222	10
	第 4 类	228、254、257、298、308、313	6
	第 5 类	459、534、546	3
Cos-kmeans	第 1 类	5、6、8、10、14、24、40	7
	第 2 类	84、90、96、101、105、110、122、128	8
	第 3 类	156、158、161、170、183、192、235、247、254	9
	第 4 类	263、377、412、414、415	5
	第 5 类	459、534、546	3
Cor-kmeans	第 1 类	5、6、8、10、14、24、40	7
	第 2 类	81、84、90、96、105、110、128、143	8
	第 3 类	146、149、150、151、153、175、182、191	8
	第 4 类	253、377、383、398、407	5
	第 5 类	459、534、546	3
全波段	第 1 类	3、4、5、7、23、104、539、545、546	9

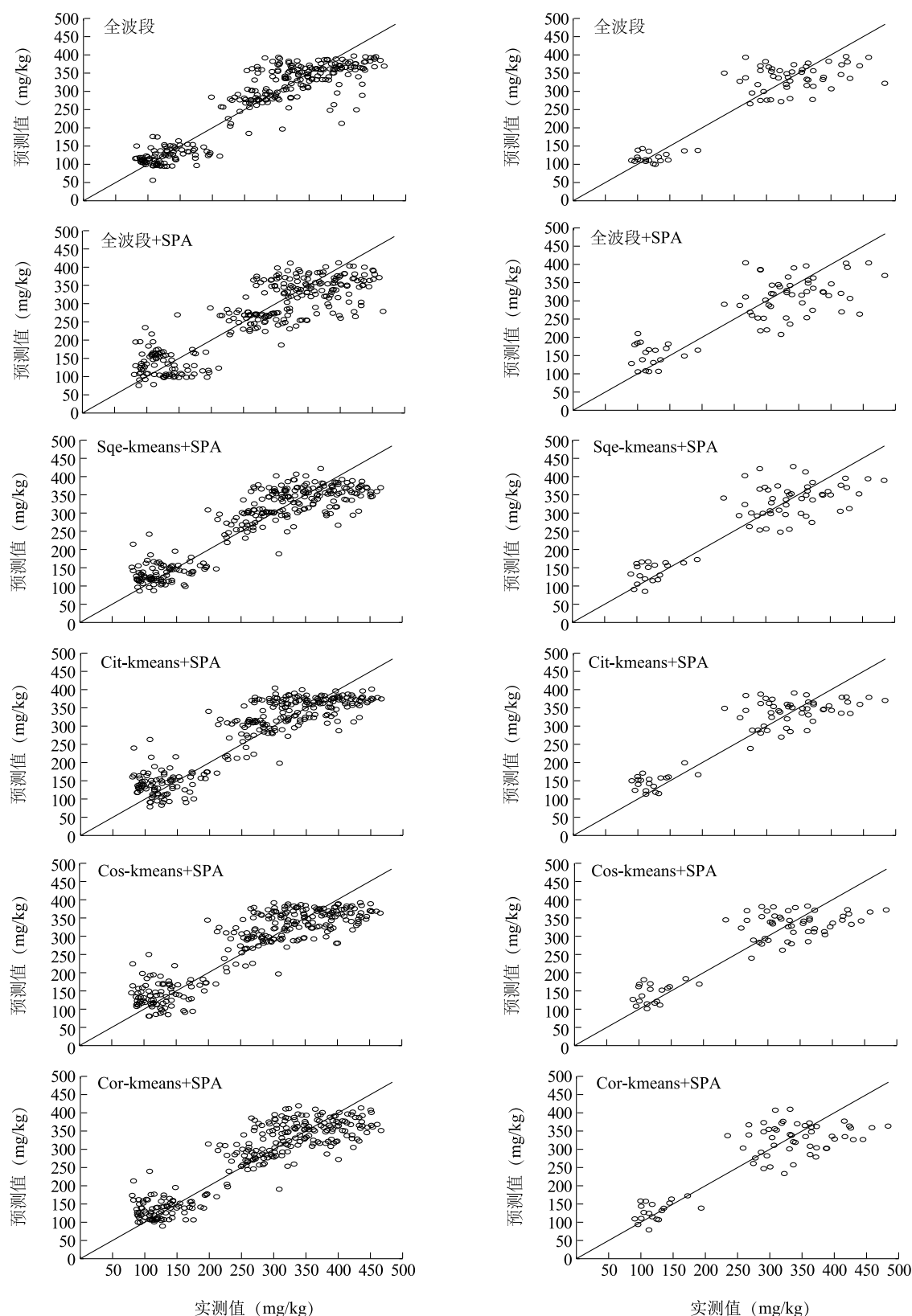
聚类方式见图 2 注。

分别采用决定系数 (R^2)、预测均方根误差 ($RMSEP$) 以及相对分析误差 (RPD) 作为模型评价标准,其中 R^2 反映预测值对实际值变化的解释程度, R^2 值越接近 1,模型预测精度越高。 $RMSEP$ 验证预测值和实际值之间的吻合度, $RMSEP$ 越小,表明模型越稳定,预测越准确。 RPD 为验证集的标准差和均方根误差的比值, RPD 越大,模型的预测能力越强^[16-21]。不同波段组合下土壤速效钾含量 BP 神经网络模型的检验结果如表 3 所示。

由表 3 可以看出,使用高光谱全部波段数据建立土壤速效钾含量光谱预测模型时,验证集的 R^2 、 $RMSEP$ 以及 RPD 分别达到 0.910 3、33.57 和 3.23,模型预测效果在所有建模波段组合中最优,能够对土壤速效钾含量进行定量预测,但由于全波段光谱数据量庞大,模型建立过程较为复杂,建模所需时间达到了 14.42 s,耗时较长。使用对高光谱全波段数据直接应用 SPA 法筛选出的特征波

段组合建立预测模型时,预测模型建立所需时间在所有波段组合中最短,仅需要 0.44 s,但模型预测效果不佳,验证集的 R^2 、 $RMSEP$ 以及 RPD 分别仅为 0.645 7、65.11 以及 1.67,只可对土壤速效钾含量进行粗略估测。使用对高光谱全波段数据先进行 K-means 聚类分析再结合 SPA 法筛选特征波段后的波段组合建立预测模型时,4 种不同聚类方式对应 4 种不同的波段组合,所建立的预测模型在预测精准度上都达到了近似于全波段组合下的模型预测精度,相对于直接运用 SPA 法有了大幅提升,并且在模型建立所需时间上,相对于全波段模型建立时有了显著改善。

分别对 Sqe-kmeans+SPA、Cit-kmeans+SPA、Cos-kmeans+SPA、Cor-kmeans+SPA 波段组合建立预测模型后所得到的土壤速效钾含量预测值进行 $\alpha=0.05$ 下的单因素方差检验,统计描述如表 4 所示,方差检验结果如表 5 所示。



左图和右图分别为建模集和验证集。

图4 土壤速效钾含量 BP 神经网络预测值和实际值比较

Fig.4 Comparison between predicted values and measured values of soil available potassium content by BP neural network

表 3 不同波段组合下土壤速效钾含量的 BP 神经网络模型检验结果

Table 3 Test results of BP neural network model of soil available potassium content under different band combinations

波段组合	建模集		验证集			建模时间 (s)
	R^2	RMSEP	R^2	RMSEP	RPD	
全波段	0.954 1	24.62	0.910 3	33.57	3.23	14.42
全波段+SPA	0.803 6	56.63	0.645 7	65.11	1.67	0.44
Sqe-kmeans+SPA	0.944 9	28.92	0.892 8	35.71	3.04	0.93
Cit-kmeans+SPA	0.948 1	35.05	0.898 9	35.06	3.10	1.01
Cos-kmeans+SPA	0.949 2	27.41	0.901 1	34.64	3.13	1.25
Cor-kmeans+SPA	0.943 1	27.51	0.889 5	36.35	2.99	1.03

R^2 : 决定系数; RMSEP: 预测均方根误差; RPD: 相对分析误差。

表 4 不同波段组合建立模型所得预测值的统计描述

Table 4 Statistical description of predicted values from models established by different band combinations

波段组合	波段数	求和	平均	方差
Sqe-kmeans+SPA	70	19 763.2	282.3	9 554.1
Cit-kmeans+SPA	70	20 102.1	287.2	8 478.4
Cos-kmeans+SPA	70	19 640.9	280.6	8 236.9
Cor-kmeans+SPA	70	19 269.7	275.2	9 362.1

表 5 不同波段组合建立模型所得预测值方差检验结果

Table 5 Variance test results of predicted values from models established by different band combinations

差异源	平方和	自由度	均方差 平方和	检验量	显著性 水平	临界值
组间	5 060.6	3	1 686.8	0.189	0.903	2.637
组内	2 458 472	276	8 907.9			
总计	2 463 632	279				

由表 6 可知, 方差检验显著性水平大于 α , 即 4 种不同聚类方式对应的不同波段组合预测模型预测的土壤速效钾含量差异并不显著。结合文献[22]的研究结果, 说明在 K-means 聚类过程中, 不同的相似度度量得到的聚类结果往往差异度较小, 在不同的聚类对象中, 通常很难确定最优的相似度度量指标。综合考虑本研究预测模型的评价指标和模型建立的时间代价, Cos-kmeans+SPA 法波段组合建立的模型验证集的 R^2 、RMSEP 以及 RPD 分别达到了 0.901 1、34.64 和 3.13, 且建模所需时间仅为 1.25 s, 可在较短的时间内实现对土壤速效钾含量的定量预测, 更适合作为预测模型的特征波段。

3 讨论

本研究分析了在对江苏省盐城市 348 份土壤样品速效钾含量高光谱预测模型建立过程中, 不同特征波段选择方法所得到的不同建模波段组合下 BP 神经网络预测模型的效果。结果表明: (1) 使用高光谱全部波段数据建立的土壤速效钾预测模型有着最优的模型预测效果, 但耗时较长; 使用连续投影法对全波段优选后的波段组合所建立的预测模型耗时最短, 但模型预测效果一般。(2) 使用本研究提出的结合 K-means 聚类法和 SPA 法所得到的波段组合建立的土壤速效钾预测模型分别改善了全波段下模型建立耗时较长和直接应用 SPA 法优选波段组合时模型精度较差的缺点, 在较短的时间内取得了较优的预测效果。结合预测评价指标和方差分析结果, Cos-kmeans+SPA 法所选波段组合更适合作为土壤速效钾含量预测模型的输入波段。

在利用高光谱数据建立土壤速效钾含量的预测模型过程中, 主要分析了不同特征波段组合对预测模型的影响, 实际应用过程中, 光谱数据的去噪效果、预测模型种类等因素都会对模型预测精度产生一定影响。在 K-means 聚类过程中, 重点分析了不同相似度度量对聚类效果的影响, 但对于聚类过程中聚类簇数量对特征波段选取的影响还有待进一步研究。

参考文献:

- [1] 祁亚琴, 吕新, 邵玉林, 等. 基于高光谱数据提取土壤养分信息的研究进展[J]. 中国农学通报, 2014, 30(12): 28-31.
- [2] 王跃明, 贾建鑫, 何志平, 等. 若干高光谱成像新技术及其应用研究[J]. 遥感学报, 2016, 20(5): 850-857.

- [3] 刘明博,唐延林,李晓利,等. 水稻叶片氮含量光谱监测中使用连续投影算法的可行性[J]. 红外与激光工程, 2014, 43(4): 1265-1271.
- [4] 王 武,王建明,李 颖,等. 近红外特征波长筛选在勾兑梨汁中原汁含量的快速检测中的应用[J]. 光谱学与光谱分析, 2017, 37(10): 3058-3062.
- [5] 陈定星. 连续投影法应用于土壤有机质 NIR 光谱分析的波长选择[D]. 广州:暨南大学, 2013.
- [6] 林 滨. K-means 聚类的多种距离计算方法的文本实验比较[J]. 福建工程学院学报, 2016, 14(1): 80-85.
- [7] 周本金,陶以政,纪 斌,等. 最小化误差平方和 K-means 初始聚类中心优化方法[J]. 计算机工程与应用, 2018, 54(15): 48-52.
- [8] 陈磊磊. 不同距离测度的 K-means 文本聚类研究[J]. 软件, 2015, 36(1): 56-61.
- [9] 王瑛瑛,宋良图. 土壤有机质近红外光谱分析的波段优选[J]. 仪表技术, 2014(5): 4-6.
- [10] 郝 勇,孙旭东,王 豪. 基于改进连续投影算法的光谱定量模型优化[J]. 江苏大学学报, 2013, 34(1): 49-53.
- [11] LIU K, CHEN X J, LI L M, et al. A consensus successive projections algorithm-multiple linear regression method for analyzing near infrared spectra[J]. Analytica Chimica Acta, 2015, 858: 16-23.
- [12] 陈思明,毛艳玲,邹小兴,等. 基于不同建模方法的湿地土壤有机质含量多光谱反演[J]. 土壤通报, 2018, 49(1): 16-22.
- [13] 杨红飞,郑黎明,郜中要,等. 砂姜黑土土壤有机碳高光谱特征与定量估算模型的研究[J]. 安徽农业大学学报, 2018, 45(1): 101-109.
- [14] 乔星星,冯美臣,杨武德,等. SG 平滑处理对冬小麦地上干生物量光谱监测的影响[J]. 山西农业科学, 2016, 44(10): 1450-1454.
- [15] 王建仁,马 鑫,段刚龙. 改进的 K-means 聚类 k 值选择算法[J]. 计算机工程与应用, 2019, 55(8): 27-33.
- [16] 曹文涛,康日斐,王集宁,等. 基于高光谱遥感的土壤氯化钠含量监测[J]. 江苏农业学报, 2016, 32(4): 817-823.
- [17] 葛 亮,王 斌,张立明. 基于波段聚类的高光谱图像波段选择[J]. 计算机辅助设计与图形学学报, 2012, 24(11): 1447-1454.
- [18] 张 悦,官云兰. 聚类与自适应波段选择结合的高光谱图像降维[J]. 遥感信息, 2018, 33(2): 66-70.
- [19] 纪文君,李 曦,李成学,等. 基于全谱数据挖掘技术的土壤有机质高光谱预测建模研究[J]. 光谱学与光谱分析, 2012, 32(9): 2393-2398.
- [20] 李冠稳,高小红,肖能文,等. 特征变量选择和回归方法相结合的土壤有机质含量估算[J]. 光学学报, 2019, 39(9): 361-371.
- [21] GRIGORIOS T, ARISTIDIS L. The MinMax K-means clustering algorithm[J]. Pattern Recognition, 2014, 47(7): 2505-2516.
- [22] 乔 天,吕成文,肖文凭,等. 基于遗传算法的土壤质地高光谱预测模型研究[J]. 土壤通报, 2018, 49(4): 773-778.

(责任编辑:张震林)