

蚁群算法在土壤质地高光谱预测建模中的应用

邓浩然, 吕成文*, 陈东来, 徐 伟

(安徽师范大学地理与旅游学院, 安徽 芜湖 241003)

摘 要: 为提高土壤质地高光谱预测模型精度, 以巢湖流域 177 个土样光谱为基础数据源, 运用蚁群算法选择特征波长, 结合 BP 神经网络构建土壤质地光谱预测模型, 并与全光谱构建的光谱预测模型进行比较。结果表明, 运用蚁群算法选择特征波长构建的光谱预测模型精度优于全光谱构建的预测模型精度, 土壤粉粒含量预测模型预测集决定系数 R^2 为 0.76, $RPIQ$ 为 2.23, 土壤砂粒含量预测模型预测集决定系数 R^2 为 0.72, $RPIQ$ 为 1.94; 全光谱土壤粉粒含量预测模型预测集 R^2 为 0.57, $RPIQ$ 为 1.75, 全光谱土壤砂粒含量预测模型预测集 R^2 为 0.48, $RPIQ$ 为 1.82。运用蚁群算法选择光谱特征波长建模, 减少了数据冗余, 提高了预测模型精度。

关 键 词: 土壤质地; 蚁群算法; 特征选择; 高光谱预测模型

中图分类号: S159, TP391 **文献标识码:** A **文章编号:** 0564-3945(2021)05-1063-06

DOI: 10.19336/j.cnki.trtb.2021032901

邓浩然, 吕成文, 陈东来, 徐 伟. 蚁群算法在土壤质地高光谱预测建模中的应用 [J]. 土壤通报, 2021, 52(5): 1063 - 1068

DENG Hao-ran, LÜ Cheng-wen, CHEN Dong-lai, XU Wei. Application of Ant Colony Optimization in Hyperspectral Prediction Modeling of Soil Texture[J]. Chinese Journal of Soil Science, 2021, 52(5): 1063 - 1068

高光谱数据具有光谱分辨率高、光谱信息丰富等优点, 可以使得高光谱数据在探测地物光谱特性的微小差异上有很大的优势。但是由于高光谱数据相邻波长间的又有很强的相关性, 并不是所有的波长对后续光谱的处理都有着同等重要的影响^[1]。应用全光谱进行土壤属性的分析建模, 不仅计算量和复杂度大, 还可能会由于冗余、无效信息的存在, 降低构建模型的预测效果^[2]。因此, 特征波长的选择在高光谱数据的处理与分析中的重要性受到越来越广泛的认同, 目前已经成为了高光谱数据建模分析过程中的关键步骤^[3]。

近年来国内外很多学者在高光谱特征波长的选择方面做了大量研究, 目的是为了选择出包含有用信息的波长或滤除包含噪声、冗余和无效信息的波长。王文才等^[4]运用连续投影算法等方法选择特征波长, 建立土壤总氮含量的光谱预测模型, 结果表明, 选择特征波长建立的光谱预测模型精度均优于全光谱建立的预测模型精度。苗雪雪等^[5]运用遗传算法等四种方法筛选特征波长, 并结合偏最小二乘法建立大米水分含量的光谱预测模型, 也提高了预测模型的精度。杨海清^[6]等运用 9 种方法选择特征波长, 构

建土壤有机质含量预测模型, 均取得了不错的结果。遗传算法、连续投影算法等特征选择的算法, 能够减少建模变量的个数, 减少波长冗余, 降低光谱数据间的相关性, 但遗传算法对样本数量有一定的要求^[7], 连续投影算法因为采用单一指标评估波长重要性, 容易选出过多的特征波长, 导致过拟合现象^[8]。蚁群算法作为一种全局搜索算法, 具有较强的鲁棒性和并行计算的能力, 通过蚂蚁之间分泌信息素的浓度变化来交流, 形成正反馈机制, 并且适用范围较广^[9]。

张娜等^[10-12]人的相关研究表明, 可见光—近红外波长光谱与土壤质地间存在相关性, 可用来预测土壤质地。本文以巢湖流域 177 个土壤样本为基础数据, 运用蚁群算法对土壤光谱样本进行特征波长的选择, 结合 BP 神经网络构建了土壤质地的预测模型, 并与全光谱构建的预测模型的精度进行对比讨论, 以期能为相关研究工作提供参考。

1 数据来源与研究方法

1.1 数据来源

土壤样本源于课题组前期在巢湖流域的相关工

收稿日期: 2021-03-29; 修订日期: 2021-06-02

基金项目: 基于土壤高光谱与环境变量信息的巢湖流域数字土壤制图研究(41371229)资助

作者简介: 邓浩然(1996-), 女, 安徽亳州人, 硕士研究生, 主要从事遥感与 GIS 应用研究。E-mail: 1791456400@qq.com

*通讯作者: E-mail: chengwen@ahnu.edu.cn

作, 共选择 177 个土壤样本, 土壤类型主要有水稻土、黄棕壤、黄褐土、粗骨土和紫色土等。去除土壤样本中的杂物后, 研磨过筛备样, 用于土壤质地测定和土壤高光谱测量。

土样质地采用激光粒度仪测定^[13], 按照美国农业部制进行土壤质地的分级: 粉粒 (0.002 ~ 0.05 mm)、砂粒 (0.05 ~ 2 mm) 和黏粒 (< 0.002 mm)。本研究主要对粉粒和砂粒进行分析。

土样光谱采用美国 ASD 公司生产的 FieldSpec 4 型地物光谱仪测量。运用 Savitzky-Golay 卷积平滑法 (多项式阶数为 2) 对测得的原始光谱曲线进行平滑去噪处理。

1.2 蚁群算法

1.2.1 蚁群算法原理

蚁群算法 (Ant Colony Optimization, ACO), 又被称为蚂蚁算法, 是一种用来寻找最优路径的群体智能算法, 最早于 1992 年意大利人 Marco Dorigo 在他的博士论文中提出。蚂蚁在寻找食物源的过程中, 会在其走过的路径中释放一种分泌物, 称为信息素, 后面的蚂蚁根据遗留在路径中的信息素判断觅食方向。当某条路径较短时, 在该路径上行走的蚂蚁会留下较多的信息素, 从而会吸引更多的蚂蚁选择这条路径, 而当路径上行走过的蚂蚁数量较多时, 留下的信息素的量也就更多, 形成了一种正反馈机制。蚁群算法波长选择就是以优化土壤光谱预测模型精度为目标, 模拟蚂蚁觅食行为, 根据路径上信息素变化寻求特征波长变量。

1.2.2 蚁群算法流程

运行蚁群算法首先要进行的是参数和初始化信息素的设置, 参数包括蚂蚁初始种群大小, 最大变量数 (波长数) 和最大迭代次数等; 初始化信息素设置为 1, 即在开始的时候每个波长点都有相同的被选择的概率。其次, 启动蚂蚁, 采用轮盘赌方法选择变量, 直到变量数达到设定的最大值。根据每只蚂蚁选择的波长组合, 运用神经网络计算、比较并获得当前迭代中最优波长组合。最后, 判断是否达到设定的最大的迭代次数, 若没有, 更新信息素表, 重复上述步骤, 直至迭代结束。

1.3 模型的建立与验证

本研究采用 BP 神经网络 (Back Propagation Neural Network, BPNN) 建立土壤质地的预测模型。BP 神经网络由输入层、隐含层和输出层三部分组成, 每一层都包含若干个节点, 通过对训练数据的学习,

计算每个节点的权重, 得到训练结果, 并将结果与预先设定的训练误差比较, 若不能满足误差的期望, 则通过前向反馈, 修改节点的权重。通过不断对训练数据的学习和训练, 来对土壤质地含量进行预测^[14-15]。随机选取全部样本的 80% 作为训练样本, 20% 作为预测样本进行建模, 建模在 matlab 2018a 中完成。建模的评价指标有决定系数 (Determinant Coefficient, R^2), 均方根误差 (Root Mean Square Error, $RMSE$) 和 $RPIQ$ (Ratio of Performance to Inter-Quartile)。 $RPIQ$ 为预测集的四分位数间距与 $RMSE$ 的比值, 将预测集样本的数值从小到大排列, 25% 出现的数值为 $Q1$, 75% 出现的数值为 $Q3$, 四分位间距即 $Q3$ 与 $Q1$ 的差值。决定系数 R^2 越大, 均方根误差 $RMSE$ 越小, 建模精度越高, 反之, 建模精度越低^[16]。Bellon-Maurel 等指出土壤的理化数据一般存在非正态分布的情况, 对于这种非正态土壤理化数据, $RPIQ$ 指数比 RPD 指数的评价更为客观^[17]。通常认为, $RPIQ$ 越大, 模型的预测能力越好, $RPIQ < 1.7$ 时, 模型预测的可信度较低; $1.7 \leq RPIQ < 2.2$ 时, 模型具有较均衡的预测能力; $RPIQ \geq 2.2$ 时, 模型的预测可信度较高^[18]。

2 结果分析与讨论

2.1 土壤质地统计特征分析

根据土壤质地分类系统^[19], 土壤样本的质地类型多为粉壤土。对巢湖流域 177 个土壤样本的质地含量进行统计分析, 如表 1。全部样本的土壤质地粉粒和砂粒含量分别在 20.76% ~ 84.50%、0.45% ~ 69.65% 之间, 平均值分别为 64.81%、14.99%, 标准差分别为 14.02%、16.05%, 偏度和峰度介于 -1.67 ~ 3.28 之间。训练样本和预测样本的分布较为均匀, 并且训练样本和预测样本基本包含了全部样本的数据范围, 具有代表性。

2.2 蚁群算法特征波长选择

以巢湖流域 177 个土壤样本为基础数据, 运用蚁群算法对土壤光谱样本进行特征波长的选择。通过多次实验, 蚁群算法参数设置如下: 初始群体大小为 100, 最大迭代次数为 100 次, 最大变量数 100, 信息素衰减系数设置为 0.65^[20]。根据各次迭代中最优波长组合数据集, 统计每个波长点被选择的频率, 即波长贡献率。根据波长贡献率由低到高, 依次去除贡献率最小的波长点, 进行 BP 神经网络的建模,

表 1 土壤粉粒和砂粒含量特征统计
Table 1 Statistical characteristics of soil silt and sand contents

粒级 Particle grade	样本类型 Sample type	样本数量 (个) Sample number	最小值 (%) Minimum	最大值 (%) Maximum	平均值 (%) Mean	标准差 (%) Standard deviation	偏度 Skewness	峰度 Kurtosis
粉粒	全部样本	177	20.76	84.50	64.81	14.02	-1.10	0.62
	训练样本	142	21.77	84.50	64.74	14.07	-0.98	0.22
	预测样本	35	20.76	79.55	65.09	13.99	-1.67	2.91
砂粒	全部样本	177	0.45	69.65	14.99	16.05	1.62	2.12
	训练样本	142	0.45	69.28	14.79	15.81	1.55	1.86
	预测样本	35	0.69	69.65	15.80	17.22	1.92	3.28

比较各次建模结果并最终得到预测能力最好的模型对应的波长组合为特征波长。图 1、图 2 分别为土壤粉粒和砂粒的特征波长中被选择波长点的贡献率，被选择的频率越高，贡献率越大。土壤粉粒的特征波长包括 19 个波长点，土壤砂粒的特征波长包括 53 个波长点。

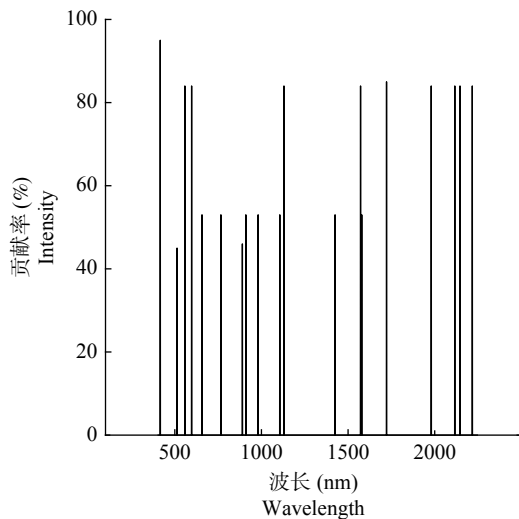


图 1 蚁群算法选择的土壤粉粒光谱特征波长
Fig.1 Spectral feature bands of soil silt selected by ACO

2.3 模型的构建与分析

运用蚁群算法选择特征波长，结合 BP 神经网络建立光谱预测模型，并与全光谱建立的光谱预测模型进行比较，结果如表 2。

通过表 2 对比分析可知，基于蚁群算法选择特征波长建立的 BP 神经网络模型预测精度较全光谱建

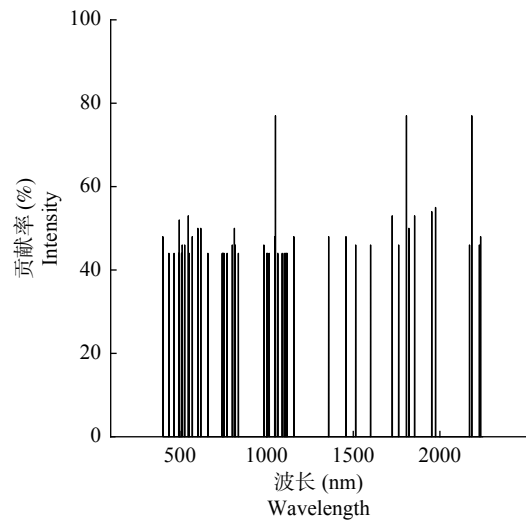


图 2 蚁群算法选择的土壤砂粒光谱特征波长
Fig.2 Spectral feature bands of soil sand selected by ACO

立的预测模型精度明显提高。基于全光谱建立的土壤质地粉粒含量的 BP 神经网络预测模型预测集的决定系数 R^2 为 0.57，均方根误差 $RMSE$ 为 14.94%， $RPIQ$ 为 1.75；经蚁群算法选择特征波长建立的土壤质地粉粒含量的 BP 神经网络模型预测集 R^2 达到 0.76， $RMSE$ 为 8.42%， $RPIQ$ 为 2.23。基于全光谱建立的土壤质地砂粒含量的 BP 神经网络预测模型预测集的决定系数 R^2 为 0.48，均方根误差 $RMSE$ 为 16.85%， $RPIQ$ 为 1.82；经蚁群算法选择特征波长建立的土壤质地砂粒含量的 BP 神经网络模型预测集 R^2 达到 0.72， $RMSE$ 为 9.66%， $RPIQ$ 为 1.94。图 3、图 4 分别为基于全光谱和经蚁群算法选择特征波长构建模

表 2 土壤粉粒和砂粒含量预测模型
Table 2 Prediction model of soil slit and sand contents

粒级 Particle grade	模型 Model	波长数 Number of wavelength	决定系数 Coefficient of determination	均方根误差 (%) Root mean square error	$RPIQ$ Ratio of performance to inter-quartile
粉粒	ACO-BP	19	0.76	8.42	2.23
	全谱-BP	1851	0.57	14.94	1.75
砂粒	ACO-BP	53	0.72	9.66	1.94
	全谱-BP	1851	0.48	16.85	1.82

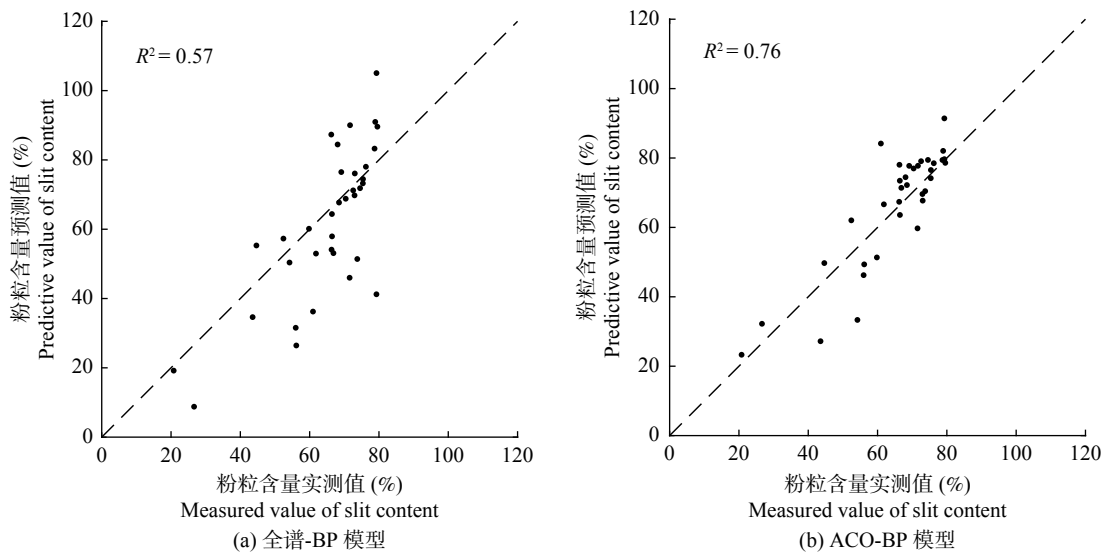


图 3 不同模型下粉粒含量预测集的实测值和预测值散点图

Fig.3 Scatter plots between the measured and predicted values of the soil silt content prediction set under different models

型的土壤粉粒和砂粒实测值和预测值的散点图，可以明显看出经蚁群算法选择特征波长构建的模型预测效果更好，拟合度更高。

3 讨论

运用蚁群算法选取特征波长建立土壤质地光谱预测模型，相较于全光谱建立预测模型有更大的优势。特征波长的选择不仅仅是减化了模型，更重要的是剔除了冗余、无效的变量信息，减少变量间的相关性，进而提高了模型的稳定性和预测力。王风云等^[21]基于蚁群算法选择特征波长以及全光谱两种方式分别建立苹果糖度光谱预测模型。结果显示，

基于蚁群算法选择特征波长建立的苹果糖度光谱预测模型较全光谱建立的预测模型预测精度更高，与本文研究结果相一致。

特征波段选择方法有很多，除蚁群算法外，还包括遗传算法、连续投影算法等，每一种算法各有特点，在不同案例中其优劣可能不尽相同。蚁群算法作为一种全局搜索算法，具有较强的鲁棒性和并行计算的能力，可以通过蚂蚁之间分泌信息素的浓度变化来交流，形成正反馈机制，能够更有效的选取特征波长^[9]。郭志明等^[22]开发了短波近红外光谱苹果品质在线检测系统，分别采用连续投影算法、遗传算法、蚁群算法等提取特征波长并结合偏最小二

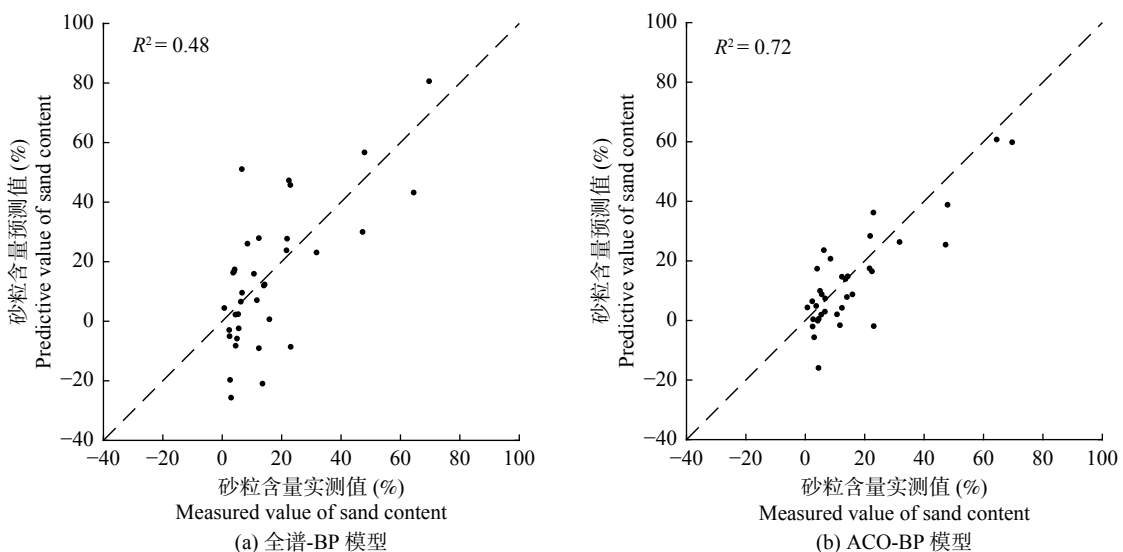


图 4 不同模型下砂粒含量预测集的实测值和预测值散点图

Fig.4 Scatter plots between the measured and predicted values of the soil sand content prediction set under different models

乘法建立预测模型。结果显示, 三模型均具有较好的预测能力。其中, 基于蚁群算法建立的模型预测能力最佳, 说明蚁群算法在光谱特征波段选择及建模中具有较好应用潜力。

在蚁群算法选择特征波长的过程中, 参数的设置对波长的选择有一定的影响, 包括蚂蚁初始种群大小、最大变量数和最大迭代次数等相关参数的设置, 需通过多次控制参数实验, 并结合前人经验确定。后续研究应关注最优参数的确定方法, 使其能更加快速、准确的寻找到最优结果。

从建模结果来看, 基于蚁群算法建立的土壤粉粒、砂粒含量的光谱预测模型均有较好的预测效果, 其中, 土壤粉粒含量光谱预测模型效果优于砂粒含量光谱预测模型效果。张娜等^[11]以河套灌区土壤样本为研究对象, 研究结果显示, 土壤粉粒含量的光谱预测模型效果优于土壤砂粒含量的光谱预测模型效果, 与本文研究情况一致。王德彩等^[23]以江苏北部黄淮平原地区的土壤样本为研究对象, 土壤砂粒含量的光谱预测模型效果优于土壤粉粒含量的光谱预测模型效果。导致这种情况的原因可能是多方面的, 首先不同预测对象的光谱敏感波段不同, 其次与研究样本土壤类型、数量、甚至建模方法等差异有关。

4 结论

以巢湖流域的 177 个土壤样本为基础数据, 运用蚁群算法选择特征波长, 通过 BP 神经网络建立了土壤质地光谱预测模型, 并与全光谱建立的光谱预测模型进行比较。运用蚁群算法选择特征波长建立的土壤质地光谱预测模型精度较全光谱建立的预测模型精度更高。经蚁群算法选择特征波长建立的土壤粉粒含量光谱预测模型所用波长数为 19 个, 模型预测集决定系数 R^2 为 0.76, $RPIQ$ 为 2.23; 土壤砂粒含量光谱预测模型所用波长数为 53 个, 模型预测集决定系数 R^2 为 0.72, $RPIQ$ 为 1.94。与全光谱建立的土壤质地光谱预测模型相比, 经蚁群算法选择特征波长建立的光谱预测模型, 去除了冗余、无效的光谱信息, 降低了建模复杂度, 进而提高了建模精度。

参考文献:

[1] 赵春晖, 陈万海, 杨 雷. 高光谱遥感图像最优波段选择方法的

研究进展与分析[J]. 黑龙江大学自然科学学报, 2007, 24(5): 592 – 602.

- [2] Zhan Y, Li M Z, Zheng L H, et al. Spectral features extraction for estimation of soil total nitrogen content based on modified ant colony optimization algorithm[J]. Geoderma, 2019, 333(1): 23 – 34.
- [3] 宾 俊, 范 伟, 周冀衡, 等. 智能优化算法应用于近红外光谱波长选择的比较研究[J]. 光谱学与光谱分析, 2017, 37(1): 95 – 102.
- [4] 王文才, 赵 刘, 李绍稳, 等. 基于特征波长选择和建模的高光谱土壤总氮含量估测方法研究[J]. 浙江农业学报, 2018, 30(9): 1576 – 1584.
- [5] 苗雪雪, 苗 莹, 龚浩如, 等. 特征波长优选结合近红外技术检测大米中的含水量[J]. 食品科技, 2019, 44(10): 335 – 341.
- [6] 杨海清, 祝 旻. 基于可见-近红外光谱特征波长选择的土壤有机质快速检测研究[J]. 红外, 2015, 36(2): 42 – 48.
- [7] 褚小立, 袁洪福, 陆婉珍. 近红外分析中光谱预处理及波长选择方法进展与应用[J]. 化学进展, 2004, (4): 528 – 542.
- [8] 王轩慧, 郑西来, 韩仲志, 等. 混合式随机森林的土壤钾含量高光谱反演[J]. 光谱学与光谱分析, 2018, 38(12): 3883 – 3889.
- [9] 殷玲玲, 苏剑锋. 浅析蚁群算法及其应用[J]. 价值工程, 2019, 38(8): 156 – 159.
- [10] 张 娜, 张栋良, 李立新, 等. 基于高光谱的区域土壤质地预测模型建立与评价——以河套灌区解放灌域为例[J]. 干旱区资源与环境, 2014, 28(5): 67 – 72.
- [11] 沈掌泉, 单英杰, 王 珂. 田间行走式测定的红外光谱数据与土壤质地之间的相关性研究[J]. 光谱学与光谱分析, 2009, 29(6): 1526 – 1530.
- [12] 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.
- [13] 张甘霖, 龚子同. 土壤调查实验室分析方法[M]. 北京: 科学出版社, 2012: 8-19.
- [14] 贺军亮, 崔军丽, 张淑媛, 等. 基于偏最小二乘的土壤重金属铜含量高光谱估算[J]. 遥感技术与应用, 2019, 34(5): 998 – 1004.
- [15] 汪子豪, 秦其明, 孙元亨, 等. 基于BP神经网络的地表温度空间降尺度方法[J]. 遥感技术与应用, 2018, 33(5): 793 – 802.
- [16] 肖云飞, 高小红, 李冠稳. 土壤有机质可见光-近红外光谱预测样本优化选择[J]. 土壤, 2020, 52(2): 404 – 413.
- [17] Bellon-Maurel V, Fernandez-Ahumada E, Palagos B, et al. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy[J]. TrAC Trends in Analytical Chemistry, 2010, 29(9): 1073 – 1081.
- [18] 曹肖奕, 丁建丽, 葛翔宇, 等. 基于光谱指数与机器学习算法的土壤电导率估算研究[J]. 土壤学报, 2020, 57(4): 867 – 877.
- [19] 李建波, 房宗启, 纪全菊, 等. 基于WPF的三角图自动识别系统构建方法——以土壤质地分类系统为例[J]. 排灌机械工程学报, 2016, 34(4): 339 – 345.
- [20] 郭志明, 黄文倩, 彭彦昆, 等. 适应蚁群优化算法的近红外光谱特征波长选择方法[J]. 分析化学, 2014, 42(4): 513 – 518.
- [21] 王风云, 郑纪业, 阮怀军, 等. 基于高光谱的套袋和不套袋苹果

- 糖度无损预测模型研究[J]. 山东农业科学, 2020, 52(6): 129 – 136.
- [22] 郭志明, 黄文倩, 陈全胜, 等. 近红外光谱的苹果内部品质在线检测模型优化[J]. 现代食品科技, 2016, 32(9): 147 – 153.
- [23] 王德彩, 郇登巍, 赵明松, 等. 平原区土壤质地的反射光谱预测与地统计制图[J]. 土壤通报, 2012, 43(2): 257 – 262.

Application of Ant Colony Optimization in Hyperspectral Prediction Modeling of Soil Texture

DENG Hao-ran, LÜ Cheng-wen*, CHEN Dong-lai, XU Wei
(School of Geography and Tourism, Anhui Normal University, Anhui 241003, China)

Abstract: In order to improve the accuracy of hyperspectral prediction model for soil texture, the soil texture spectral prediction model was conducted based on the 177 soil samples taken from Chaohu Lake Basin using the methods of the ant colony optimization to select the characteristic wavelengths, and combined with BP neural network. The results showed that, compared with the spectral prediction model constructed by full spectrum, the accuracy of the spectral prediction model constructed in this study was better. The determination coefficient R^2 of the prediction set of the soil silt content prediction model was 0.76 and $RPIQ$ was 2.23, that of the soil sand content prediction model was 0.72 and $RPIQ$ was 1.94. The prediction set R^2 and $RPIQ$ of the full spectra soil silt content prediction model were 0.57 and 1.75, respectively, and that of the full spectra soil sand content prediction model were 0.48 and 1.82, respectively. Compared with the full spectrum prediction model of soil texture, the ant colony optimization selected the spectral characteristic wavelengths for modeling could reduce the redundancy of spectral information and improve the accuracy of the prediction model.

Key words: Soil texture; Ant colony optimization; Feature selection; Hyperspectral prediction model

[责任编辑: 韩春兰]