

卢宏亮, 赵明松. 基于神经网络模型的安徽省土壤 pH 预测[J]. 江苏农业学报, 2019, 35(5): 1119-1123.
doi: 10.3969/j.issn.1000-4440.2019.05.017

基于神经网络模型的安徽省土壤 pH 预测

卢宏亮¹, 赵明松^{1,2}

(1. 安徽理工大学测绘学院, 安徽 淮南 232001; 2. 中国科学院南京土壤研究所土壤与农业可持续发展国家重点实验室, 江苏南京 210008)

摘要: 以土壤 pH 为研究对象, 利用一般反向传播(Back propagation, BP)神经网络模型、带回溯的弹性反向传播(Resilient back propagation with backtracking, RPROP-WB)神经网络模型、不带回溯的弹性反向传播(Resilient back propagation without backtracking, RPROP-OB)和最小绝对梯度反向传播(Smallest absolute gradient resilient back propagation, SAG-RPROP)神经网络模型进行安徽省土壤 pH 的预测及制图, 选用均方根误差(RMSE)、绝对平均误差(MAE)及决定系数(R^2)为评价标准, 比较 3 种改进的神经网络模型与一般 BP 神经网络模型对于土壤 pH 的预测能力。结果表明: 研究区域内, 4 种神经网络模型的拟合能力高低依次为: SAG-RPROP>RPROP-WB>RPROP-OB>BP。由建模集可以看出, RPROP-WB、RPROP-OB 2 种模型与 BP 神经网络模型的预测精度一致, 4 种模型中预测精度最高的为 SAG-RPROP, R^2 比其他 3 种模型提高 0.07。对于验证集, 预测能力高低依次为: SAG-RPROP>RPROP-WB>RPROP-OB>BP, 预测精度和泛化能力最高的为 SAG-RPROP 模型, RMSE、MAE 和 R^2 分别为 0.67、0.50 及 0.59。空间预测图结果显示, 4 种模型所得安徽省土壤 pH 空间分布基本类似, 均呈“南酸北碱”趋势, 一般 BP 神经网络对于土壤 pH 预测区分度较低, 预测所得安徽省南部地区的土壤 pH 均集中在 5.57 至 6.50 之间, RPROP-WB、RPROP-OB 及 SAG-RPROP 所得预测图则区分更为明显。综上所述, RPROP 及其改进算法可以有效地进行土壤属性的预测, 且精度均高于一般 BP 神经网络模型。

关键词: 土壤 pH; 空间预测; RPROP 算法; GRPROP 算法; 神经网络; 安徽省

中图分类号: S127 **文献标识码:** A **文章编号:** 1000-4440(2019)05-1119-05

Prediction of soil pH in Anhui province based on RPROP and GRPROP algorithms

LU Hong-liang¹, ZHAO Ming-song^{1,2}

(1. School of Geodesy and Geomatics, Anhui University of Science and Technology, Huainan 232001, China; 2. State Key Laboratory of Soil and Sustainable Agriculture, Institute of Soil Science, Chinese Academy of Sciences, Nanjing 210008, China)

Abstract: In this study, soil pH was taken as the research object, using general back propagation (BP) neural network model, resilient back propagation with backtracking (RPROP-WB) neural network model, resilient back propagation without backtracking (RPROP-OB) neural network model and smallest absolute gradient resilient back propagation (SAG-RPROP) neural network model for predicting and mapping soil pH in Anhui province, and the root mean square error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) were used as evaluation criteria to compare the prediction ability of three improved neural network models and the general BP neural network model for soil pH. The results

showed that: the fitting ability of the four neural network models in the study area from high to low was: SAG-RPROP>RPROP-WB>RPROP-OB>BP. It can be seen from the training set that the prediction accuracy of the RPROP-WB and RPROP-OB models was consistent with that of the BP neural network model. The highest prediction accuracy among the four models was SAG-RPROP, and the

收稿日期: 2018-11-13

基金项目: 国家自然科学基金项目(41501226); 安徽省高校自然科学基金项目(KJ2015A034); 土壤与农业可持续发展国家重点实验室开发基金项目(Y412201431)

作者简介: 卢宏亮(1993-), 男, 安徽铜陵人, 硕士研究生, 从事数字土壤制图研究。(E-mail) 17355481665@163.com

通讯作者: 赵明松, (E-mail) zhaomingsonggis@163.com

R^2 was 0.07 higher than the other three models. For the verification set, the prediction ability was as follows: SAG-RPROP>RPROP-WB>RPROP-OB>BP. The SAG-RPROP model had the highest prediction accuracy and generalization ability, and RMSE, MAE and R^2 were 0.67, 0.50 and 0.59. The results of spatial prediction showed that the spatial distribution of soil pH in Anhui province was similar in the four models, all of which showed the trend of "Southern acid North Base". The general BP neural network had lower discrimination for soil pH prediction. The soil pH in the southern part of Anhui province was concentrated between 5.57 and 6.50. The prediction maps obtained by RPROP-WB, RPROP-OB and SAG-RPROP models were more obvious. In summary, RPROP and its improved algorithms can effectively predict soil properties, and the accuracy is higher than the general BP neural network model.

Key words: soil pH; spatial prediction; RPROP algorithm; GRPROP algorithm; neural network; Anhui province

土壤 pH 显著影响土壤的其他物理、化学性质和植物病害^[1-3],如土壤肥力水平、微生物和动物群活动、C/N 比及腐殖质的形成与土壤 pH 密切相关^[4]。研究土壤 pH 的空间分布对于土壤质量的监测与管理有重要意义。传统的土壤 pH 预测方式多采用野外采样和室内化学分析相结合的方式进行,缺点是费时费力,不易于大面积测定。目前随着机器学习技术的不断发展,采用机器学习技术进行土壤属性的预测成为研究热点。

近年来,国内外使用神经网络进行土壤属性的预测研究取得了一定的成果。国外研究中, Kim 等^[5]利用 BP 神经网络估算了来自美国内布拉斯加州林肯附近的土壤侵蚀及径流中的养分含量,结果表明土壤侵蚀量与降雨量和径流量呈正相关,且神经网络模型可以得到径流中养分的可靠估计。Were 等^[6]对比了支持向量回归(Support vector regression, SVR)、人工神经网络(Artificial neural network, ANN)和随机森林(Random forest, RF)3 种模型对于土壤有机碳的预测及制图精度,结果表明 SVR 模型的预测精度最高。国内研究中,董敏等^[7]以土壤有效锌为研究对象,构建了一种遗传径向基(GARBF)神经网络,并比较其与径向基(RBF)神经网络及普通克里格模型对于研究对象的预测精度,结果表明 GARBF 神经网络具有更高的预测精度。郑立华等^[8]利用 BP 神经网络预测方法,建立基于近红外光谱技术的土壤有机质含量和土壤全氮含量的分析模型,结果证明 BP 神经网络模型具有较高的鲁棒性和容错性。江中枫等^[9]应用集成 BP 神经网络模型预测了土壤有机质空间分布,克服了单一 BP 神经网络局部搜索能力差和易陷入全局最优的缺点。

国内外研究大多使用简单的 BP 神经网络进行土壤属性预测制图,经过改进的神经网络模型往往有更高的预测精度,具有一定的研究价值。而目前使用改进的神经网络模型进行土壤属性预测的研究

较少,因此本研究基于 2 种已有的 RPROP^[10]和 GRPROP^[11]算法,分别建立带回溯的弹性反向传播(Resilient back propagation with backtracking, RPROP-WB)神经网络模型、不带回溯的弹性反向传播(Resilient back propagation without backtracking, RPROP-OB)和最小绝对梯度(Smallest absolute gradient, SAG)RPROP 神经网络模型,探讨不同优化方法对 RPROP 神经网络模型的影响和预测精度,并与传统的 BP 神经网络模型作对比,最终建立最优的土壤 pH 预测神经网络模型。

1 材料与方法

1.1 区域概况

安徽省(114°54'~119°37'E, 29°41'~34°38'N)地处中国东部,跨长江、淮河中下游,东临以海为中心的长江三角洲经济区,西接中原腹地,总面积 1.396×10⁵ km²。全省处于亚热带向温带过渡带。年均气温 14~16℃,年均降雨量 800~1 800 mm。除安徽西南和南部丘陵地区外,海拔一般不超 100 m。安徽省从北到南分为淮河中游平原、江淮丘陵地区、沿江平原、皖西大别山、皖南丘陵地区等 5 个地理区域。农田主要分布在淮河中游平原、江淮丘陵岗地区和沿江平原,林地和草地主要分布在皖西、皖南丘陵地区的大别山。主要土壤类型有:砂姜黑土、潮土、水稻土、黄壤、黄棕壤、棕壤、黄褐土、红壤、紫色土、石质土、粗骨土、山地草甸土等。

1.2 数据来源

本研究数据来源于野外土壤样本采集和遥感数据。最终数据库由 15 个环境变量和 1 个目标变量(土壤 pH)组成。

1.2.1 土壤数据 采样深度为 0~20 cm。来源于《中国土志·安徽卷》^[12],按照随机性、均匀性和代表性的原则在研究区共采集样点 140 个,采样时

间为 2010–2011 年。以土壤 pH 为研究目标,土壤样本自然风干后,过 2.00 mm 筛以供 pH 测定,实验室测量土壤 pH 采用水浸提法。

1.2.2 地形数据 来源于 SRTM 的数字高程模型 (DEM),空间分辨率为 90 m。利用 ArcGIS 10.2 计算坡度 (Slope),利用 SAGA GIS 6.3.0 提取多尺度谷底平坦度 (MrVBF)、多尺度脊顶平坦度 (MrRTF)、地形湿度指数 (TWI) 及地形位置指数 (TPI)。

1.2.3 归一化植被指数 (NDVI) 和增强植被指数 (EVI) 来源于 MODIS 陆地产品 16 d 合成植被指数 (MOD13Q1),空间分辨率为 250 m,时间为 2010 年 6 月。

1.2.4 气候数据 主要包括年均温度 (MAT)、年均降水量 (MAP)。数据来自中国农业科学院农业资源与农业区划研究所中国生态环境背景层面建设项目完成的栅格数据 (1 km 分辨率),为 1980–1999 年的逐月平均值计算生成。所有环境变量及土壤属性空间预测结果,使用最近邻内插法将分辨率统一为 500 m。

1.3 研究方法

反向传播 (Back propagation, BP) 是使用多层前馈网络进行监督学习的最广泛使用的算法。反向传播学习算法^[13]的基本思想是重复应用链式规则来计算网络中每个权重对任意误差函数的影响。误差函数计算公式为:

$$\frac{\partial E}{\partial \omega_{ij}} = \frac{\partial E}{\partial s_i} \frac{\partial s_i}{\partial net_i} \frac{\partial net_i}{\partial \omega_{ij}} \quad (1)$$

式中 $\partial \omega_{ij}$ 是从神经元 j 到神经元 i 的权重, s_i 是输出, net_i 是神经元 i 输入的加权和。一旦知道每个权重的偏导数,通过执行简单的梯度下降来实现最小化误差函数的目的:

$$\omega_{ij}(t+1) = \omega_{ij}(t) - \varepsilon \frac{\partial E}{\partial \omega_{ij}}(t) \quad (2)$$

式中 ε 是学习速率 (Learning rate)。如何设置 ε 是一个重要的问题,为了解决这个问题,已经有了很多的理论,如 SGD 算法、Nesterov 算法、AdaGrad 算法及 Adam 算法^[14-16]等。RPROP 表示“弹性传播”,是一种有效的神经网络学习方案,可根据局部梯度信息直接调整权重步长。与其他优化技术的重要区别在于,适应的影响不会因梯度行为而模糊^[10]。其基本原理为:首先给定各权重变化一个初始值,设定权重变化的加速与减速因子,在网络前馈迭代中当连续误差梯度符号不变时,采用加速策略,

加快训练速度;当连续误差梯度符号变化时,采用减速策略,以期稳定收敛^[17]。GRPROP 算法是基于 RPROP 算法在全局收敛上的改进模型,研究结果表明 GRPROP 比 RPROP 和 IRPROP 表现出更好的收敛速度和稳定性^[11]。

1.4 模型构建与评价

基于 R 语言中的 caret 和 neuralnet 软件包,建立 4 种神经网络模型。由于环境变量数值差异较大,为保证神经网络模型更好的收敛,首先对环境变量进行标准化处理,且由于样本数量较少,为增加模型的稳定性,故在建模中选用 5 次十折交叉验证和独立验证 2 种方式验证建模精度。模型建立之后随机选取 43 个样本点作为独立验证集进行预测,以比较不同模型的预测精度和泛化能力。其中 caret 软件包用于进行十折交叉验证,neuralnet 软件包用于建模。

模型精度评定选用均方误差 (RMSE)、平均绝对误差 (MAE)、以及决定系数 (R^2) 3 个标准。计算方法如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (p_i - o_i)^2} \quad (3)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |o_i - p_i| \quad (4)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (p_i - o_i)^2}{\sum_{i=1}^n (p_i - \hat{o}_i)^2} \quad (5)$$

式中, p_i 和 o_i 分别为预测值和观测值, \hat{o}_i 是平均观测值。

2 结果与分析

2.1 安徽省土壤 pH 统计特征

表 1 为研究区内采样点土壤 pH 基本统计特征。样本土壤 pH 范围为 4.58~8.67,平均值为 6.37,变异系数为 18.21%,属于中等变异。土壤样点中,强酸性样本数为 36,占总体比例 26%;酸性样本数为 45,占总体比例 32%;中性样本数为 33,占总比例 24%;微碱性样本数为 20,占总体比例 14%;强碱性样本数为 6,占总体比例 4%。峰度和偏度分别为 -0.90 和 0.51,属于偏态分布,经对数变换后符合正态分布。整体来说,安徽省土壤酸性比例较大,约占总比例 58%。土壤样点 pH 空间分布 (图 1) 表明,安徽省土壤酸碱度分布大致呈南酸北碱的格局。

建模集均值(6.41)大于验证集(6.25),建模集标准差(1.16)低于验证集(1.17),所以建模集变异系数(18.10%)低于验证集(18.72%)。样本整体分布相似,可以用于模型建立。

表 1 安徽省土壤 pH 基本统计特征

Table 1 Statistical characteristics of soil pH in Anhui province

样本	样本数	土壤 pH 最大值	土壤 pH 最小值	土壤 pH 均值	标准差	变异系数 (%)
总样本	140	8.67	4.58	6.37	1.16	18.21
建模集	108	8.67	4.58	6.41	1.16	18.10
验证集	32	8.51	4.68	6.25	1.17	18.72

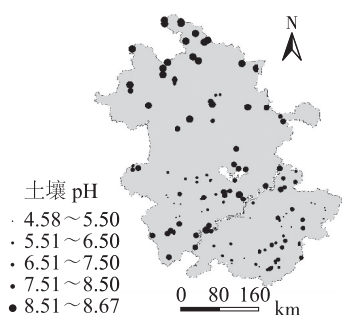


图 1 采样点土壤 pH 空间分布

Fig.1 Soil pH spatial distribution of sampling points

2.2 不同模型精度分析

由表 2 可知,对于建模集,相对于一般 BP 模型,RPROP-WB 及 RPROP-OB 模型的拟合度略有提高,RMSE 和 MAE 也略有降低,优化效果不高。相对而言,SAG-RPROP 模型拟合度最高,RMSE 降低了 0.07,MAE 降低了 0.05, R^2 提高了 0.07,整体拟合度较高。由验证集可知,4 种模型预测精度高低依次为:SAG-RPROP>RPROP-WB>RPROP-OB>BP,预测精度最高的为 SAG-RPROP 模型,RMSE 降低了 0.17~0.12,MAE 降低了 0.10~0.14, R^2 提高了 0.15~0.22,精度提高较大。对比建模集和验证集可知,BP、RPROP-WB、RPROP-OB 及 SAG-RPROP 模型建模集和验证集 R^2 之差分别为 0.24、0.18、0.23 和 0.10,可以看出改进的 BP 神经网络模型稳定性均有所提高,SAG-RPROP 模型稳定性最高。综上所述,SAG-RPROP 模型解释度(建模精度)、泛化能力(验证集精度)和稳定性(建模集和验证集精度对比)比一般 BP 神经网络均有所提高,其他 2 种模型则精度提高较为有限。故 SAG-RPROP 模型更适合用于安徽省土壤 pH 空间预测。

表 2 基于不同神经网络模型的精度比较

Table 2 Accuracy comparison based on different neural network models

模型	建模集			验证集	
	MAE	RMSE	R^2	MAE	RMSE
BP	0.57	0.72	0.62	0.64	0.83
RPROP-WB	0.57	0.71	0.62	0.60	0.78
RPROP-OB	0.57	0.72	0.62	0.63	0.82
SAG-RPROP	0.52	0.65	0.69	0.50	0.67

RMSE:均方误差;MAE:绝对平均误差; R^2 :决定系数。

2.3 安徽省土壤 pH 空间分布制图

空间制图使用 ArcGIS10.2 软件以 DEM 栅格为基础,进行栅格转点,以 500 m 分辨率为标准,提取建模所使用的环境变量数据,使用 4 种模型分别进行土壤 pH 的空间制图。由图 2 可知,4 种不同神经网络模型对于安徽省土壤 pH 的预测空间分布基本一致,淮河中游平原、沿江平原地区及江淮丘陵地区东部土壤 pH 值较高,皖西丘陵地区、皖南丘陵地区和江淮丘陵地区东部土壤 pH 值较低,安徽省土壤 pH 的空间变化趋势大致为由北至南逐渐降低,北部大致为微碱和偏碱性土壤,南部大多为酸性土壤,整体基本符合“南酸北碱”的特征,表明 4 种神经网络模型用于安徽省土壤 pH 的数字制图精度均较高。相比其他 3 种模型,一般 BP 神经网络模型的所得空间预测分布区分度较低,安徽省南部地区土壤 pH 预测值主要集中在 5.57 至 6.50 之间,并没有很好地区分。

3 结 论

本研究以地形数据、遥感植被指数、气候数据等环境变量为预测因子,利用不同神经网络模型,实现了安徽省区域内的土壤 pH 空间分布预测。主要结论如下:(1)4 种不同神经网络模型分别解释了安徽省土壤 pH 的 62%、62%、62%和 69%,模型均有较高精度,SAG-RPROP 模型拟合度最高。(2)在预测方面,相比一般的 BP 神经网络模型,RPROP-WB 神经网络模型、RPROP-OB 神经网络模型和 SAG-RPROP 神经网络模型的精度更高,其中基于 RPROP 算法优化的 SAG-RPROP 模型具有最高的预测精度,且 SAG-RPROP 模型的泛化能力和模型的稳定性最好。(3)4 种模型的制图结果均呈相同趋势,对于安徽省土壤 pH 的空间分布预测均呈“南酸北碱”趋势。相比一般 BP 神经网络模型而言,其他 3 种模型的制图区分度更高,划分更为细致。综上所述,SAG-RPROP 算法比

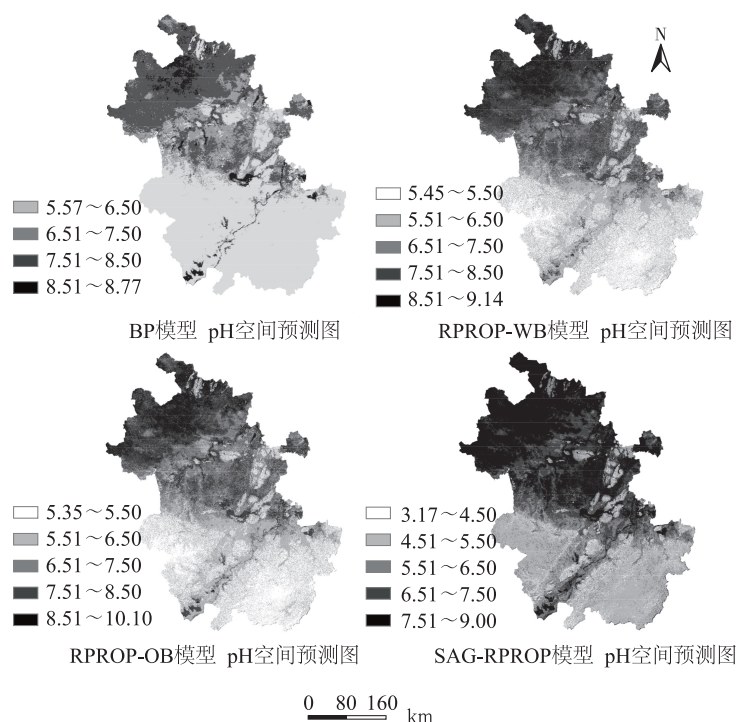


图2 安徽省土壤 pH 属性实测值及预测值空间分布

Fig.2 Spatial distribution of the measurements and predictions for soil pH properties in Anhui province

传统 BP 神经网络更适合作为土壤属性预测的模型。

参考文献:

- [1] 吕贻忠,李保国,李跃进,等.土壤学[M].北京:中国农业出版社,2005:115.
- [2] LAUBER C L, HAMADY M, KNIGHT R, et al. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale[J].Applied and Environmental Microbiology, 2009, 75(15): 5111-5120.
- [3] 蒋欢,彭玉梅,董代文,等.土壤酸碱度与榨菜根肿病的关系分析[J].南方农业学报, 2017, 48(9): 1617-1623.
- [4] HOSSNER L R. Field pH[M].Netherlands: Springer, 2008: 271-272.
- [5] KIM M, GILLEY J E. Artificialneural network estimation of soil erosion and nutrient concentrations in runoff from land application areas[J].Computers & Electronics in Agriculture, 2008, 64(2): 268-275.
- [6] WERE K, BUI D T, OYSTEIN B D, et al. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape[J].Ecological Indicators, 2015, 52: 394-403.
- [7] 董敏,王昌全,李冰,等.基于 GARBF 神经网络的土壤有效锌空间插值方法研究[J].土壤学报, 2010, 47(1): 42-50.
- [8] 郑立华,李民赞,潘变,等.基于近红外光谱技术的土壤参数 BP

神经网络预测[J].光谱学与光谱分析, 2008, 28(5): 1160-1164.

- [9] 江叶枫,郭熙,叶英聪,等.应用集成 BP 神经网络模型预测土壤有机质空间分布[J].江苏农业学报, 2017, 33(5): 1044-1050.
- [10] RIEDMILLER M, BRAUN H. A direct adaptive method for faster backpropagation learning: the RPROP algorithm[J]. Chemical Product & Process Modeling, 1993, 6(A8): 586-591.
- [11] ANASTASIADIS A D, MAGOULAS G D, VRAHATIS M N, et al. New globally convergent training scheme based on the resilient propagation algorithm[J].Neurocomputing, 2005, 64: 253-270.
- [12] 李德成,张甘霖,王华等.中国土系志·安徽卷[M].北京:科学出版社, 2017: 3-24.
- [13] RUMELHART D E, MCCLELLAND J L. Parallel distributed processing[M].Cambridge, Massachusetts: The MIT Press, 1986: 45-76.
- [14] MEULEAU N, DORIGO M. Ant colony optimization and stochastic gradient descent[J].Artificial Life, 2002, 8(2): 103.
- [15] NESTEROV Y. Gradient methods for minimizing composite functions[J].Mathematical Programming, 2013, 140(1): 125-161.
- [16] WEI C H, LI Y, HUANG P J, et al. Estimation of breast density: an adaptive moment preserving method for segmentation of fibroglandular tissue in breast magnetic resonance images.[J].European Journal of Radiology, 2012, 81(4): 618-624.
- [17] 杨克俭,梁佳成,王艳秋.基于 RPROP 神经网络算法的旋转机械故障诊断模型[J].硅谷, 2008(23): 116-116.

(责任编辑:张震林)