

BPNN 对不同人为活动区域的盐渍土 Na^+ 高光谱估测

田安红¹, 付承彪¹, 熊黑钢², 赵俊三³

(1.曲靖师范学院 信息工程学院, 云南 曲靖 655011; 2.北京联合大学 应用文理学院, 北京 100083; 3.昆明理工大学 国土资源工程学院, 昆明 650093)

摘要:土壤盐分阳离子 Na^+ 在盐渍土的形成过程中起着重要作用, 以新疆无人区为活动(A区)和有人区为活动(B区)区域的土壤为研究对象, 采集野外高光谱和土壤 0—20 cm 样本, 化验 Na^+ 含量, 利用 BP 神经网络(BPNN)、偏最小二乘(PLSR)和逐步多元回归(SMLR)模型对比分析 Na^+ 的高光谱估测, 并力图解释 Na^+ 在不同人为活动区域的估算精度机理。结果表明: Na^+ 在 A 区和 B 区的 4 种阳离子(Ca^{2+} , Mg^{2+} , K^+ , Na^+)中, 所占比例最高分别为 48.4% 和 62.3%, 均值最大分别为 1.590, 2.148。对原始(R)和倒数(1/R)两种光谱变换进行一阶与二阶微分预处理, 提取出相关系数通过 0.05 检验的波段为特征波段, 3 种建模方法在两个研究区域中共有 24 种模型, 且 1/R 在二阶微分处的 BPNN 模型均是 A 区和 B 区的最佳预测模型, 分别迭代 19 次和 9 次时精度满足要求。相对分析误差 RPD、决定系数 R^2 和均方根误差 RMSE 在 A 区分别为 2.461 6, 0.860 9, 0.350 1, 在 B 区分别为 2.169 8, 0.800 6, 0.803 5。BPNN 对 Na^+ 离子的预测能力很好, PLSR 的预测能力一般, SMLR 的预测能力很差。研究成果可为改良干旱区的盐渍化土壤提供科学依据。

关键词:盐渍土; Na^+ 离子; 微分处理; BP 神经网络; 野外高光谱

中图分类号: S151.9

文献标识码: A

文章编号: 1005-3409(2020)02-0364-06

Hyperspectral Estimation of Na^+ Ion in Saline Soils in Areas With Different Human Activities Using BPNN Model

TIAN Anhong¹, FU Chengbiao¹, XIONG Heigang², ZHAO Junsan³

(1.College of Information Engineering, Qujing Normal University, Qujing, Yunnan 655011,

China; 2.College of Applied Arts and Science, Beijing Union University, Beijing 100083, China;

3.Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming 650093, China)

Abstract: Cation Na^+ of soil salinity plays an important role in the formation of saline soil. The soils in no human activity (Zone A) and human activity (Zone B) of Xinjiang were studied. Field hyperspectral data and 0—20 cm soil samples were collected, and Na^+ contents were tested. BP neural network (BPNN), partial least squares (PLSR) and stepwise multiple regression (SMLR) models were used to compare and analyze the hyperspectral estimation of Na^+ , and to explain the mechanism of Na^+ estimation accuracy in areas with different human activities. The results show that the Na^+ has the highest proportion and the largest mean value among the four cations (Ca^{2+} , Mg^{2+} , K^+ , Na^+), the proportion of Na^+ in Zone A and B is 48.429% and 62.274%, respectively, and the mean value is 1.590 and 2.148, respectively. First-order and second-order differential processing were been performed on the Original (R) and reciprocal (1/R) spectral transforms, and the band whose correlation coefficient was checked by 0.05 was extracted as the characteristic band. There are 24 models in Zone A and B for the three modeling methods, and the BPNN models with 1/R at the second-order differential are the best prediction models for Zone A and B, which meet the accuracy requirements when they are iterated by 19 times and 9 times, respectively. The relative analysis error RPD, the decision coefficient R^2 , and the root mean square error RMSE are 2.461 6, 0.860 9, and 0.350 1 in Zone A, and 2.169 8, 0.800 6, and 0.803 5 in Zone B, respectively. The prediction ability of BPNN for Na^+ ions is very good, the prediction ability of PLSR is general, and the prediction ability of SMLR is very poor. These

收稿日期: 2019-09-22

修回日期: 2019-10-12

资助项目: 国家自然科学基金青年项目(41901065); 国家自然科学基金面上项目(41671198); 校级项目(2019JZ001)

第一作者: 田安红(1984—), 女(土家族), 贵州安顺人, 硕士, 副教授, 主要从事干旱区盐渍土的高光谱研究。E-mail: tianfucb@163.com

通信作者: 付承彪(1982—), 男, 云南宣威人, 硕士, 讲师, 主要从事高光谱遥感图像研究。E-mail: fucb305@163.com

research results can provide a scientific basis for improving salinized soil in arid regions.

Keywords: saline soil; Na^+ ; differential processing; BP neural network; field hyperspectral data

新疆作为典型的干旱半干旱农业区,拥有广袤无垠的土地资源^[1-2],但其盐渍化现象严重,土壤贫瘠,盐分含量高,养分含量低,实际中可用于耕地的面积很少。然而,土壤盐渍化是全球的生态环境问题,会严重限制农业的可持续利用与发展,现已成为土壤退化过程的主要类型^[3]。近年来,针对土壤元素的预测模型多数采用偏最小二乘和逐步多元回归等线性模型来反演^[4-5],但因土壤高光谱数据与土壤元素之间呈现为非线性关系,传统线性模型的估测精度不理想。然而,BP 神经网络(Back propagation neural network, BPNN)理论上可以逼近任意的非线性函数^[6-7],在定量反演非线性问题时具有较大的优势。如刁万英等^[8]以砂土和壤土为研究对象,利用人工神经网络模型结合室内光谱数据预测土壤的含水量,仿真结果表明与指数、对数和线性模型相比较,人工神经网络模型对两种类型土壤含水量的预测精度最高,该模型更具有普适性。卢志宏等^[9]采用 70 个特征波段和 BP 神经网络模型估算草原矿区的氮和磷含量,结果表明当隐藏层的节点数为 10, 3, 6 时, BP 模型对氮和磷的预测精度最好,具有较大的拟合优度和较小的均方根误差。郭云开等^[10]以长沙县的土壤为研究对象,利用 BP 神经网络反演土壤重金属元素 Cu 的含量,其拟合优度为 0.721,远大于线性回归模型的拟合优度值,且预测精度为 0.823。

但已有文献采用 BPNN 模型对土壤元素的估算主要是针对同一研究区域的土壤,且预测对象多为盐分、水分、有机质、氮、磷、钾等土壤元素。然而,目前针对不同人为活动区域的盐渍土元素的高光谱反演鲜有报道,尤其是土壤盐分的主要阳离子研究报道则更少。但是,人为活动对土壤的干扰在自然界中较为常见,人为活动的强度大小、方式类型和持续作用时间,都会改变土壤的理化性质^[11],造成土壤元素在空间上的变异。同时,新疆盐渍土的主要类型之一为易溶盐,易溶盐主要包含硫酸钠和氯化钠,土壤盐分的阳离子 Na^+ 在盐渍土的形成过程中起着重要作用。因此,本研究以新疆没有受到人为活动影响的土壤和受到人为活动影响的土壤为研究对象,利用 BPNN 模型、偏最小二乘模型和逐步多元回归模型对比分析不同人为活动区域的盐渍土 Na^+ 离子的高光谱估算结果,并力图解释 Na^+ 离子在不同研究区域的反演精度的机理,为合理利用新疆的土地资源提供一定的理论依据。

1 试验材料与方法

1.1 研究区域与土壤样品采集

以新疆维吾尔自治区阜康市的盐渍土为研究区,其坐标为 $43^{\circ}29' - 45^{\circ}45' \text{N}$, $87^{\circ}44' - 88^{\circ}46' \text{E}$ 。根据野外考察情况,主要以人类活动对土壤的干扰程度进行划分,即 A 区(无人活动的区域)和 B 区(有人活动的区域)。由于 A 区和 B 区中间有一条水渠,其长度为 15.30 km, A 区因被水渠相隔,基本没有人类活动,土壤表层维持住原有的生态风貌,主要植被为红柳、猪毛菜和梭梭等。而 B 区距离新疆 102 建设兵团的距离很近,该区域经常受到人类日常活动的影响,并且,近 2 a 内有大部分土壤被人工种植开发为育苗林地、榆树林地等。

土壤样本的采集时间为 2017 年 5 月 9 日到 23 日,在 A 区布设采样线 5 条, B 区布设采样线 6 条,且在每条采样上间隔 400 m 左右选择 5 个代表性的采样点, A 区共 25 个采样点, B 区共 30 个采样点,一共为 55 个采样点,见图 1。采集土壤样本 0—20 cm,并密封带回实验室经过风干、剔除碎石、研磨、孔筛等处理后,送到中科院新疆生态与地理研究所^[12],由化验员测量土壤 Na^+ 离子含量。

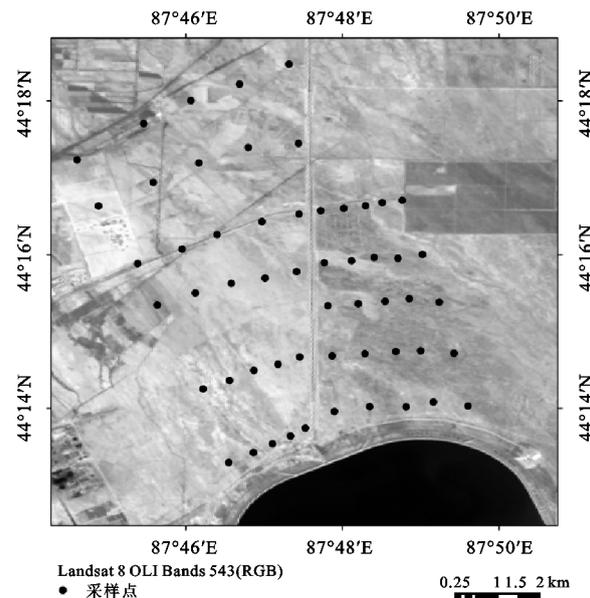


图 1 土壤采样点

1.2 实地高光谱数据采集

采用 ASD FieldSpec3 光谱仪测量研究区的土壤室外高光谱,光谱仪的波长范围为 350~2 500 nm。在室外测量时,选择天空晴朗且无风的天气,测量时

间为新疆 13:00—15:00 之间。测量前进行白板校正,且探头垂直向下测量高光谱,每个采样点重复测量 10 次,按梅花桩采样方法,共计每个样点测量得到 50 条高光谱数据,取其平均值即为该样本点的野外高光谱反射率曲线。同时,采集的高光谱需要删除边缘波段(350~390 nm 和 2 401~2 500 nm)和水分吸收带(1 355~1 410 nm 和 1 820~1 942 nm)附近的波段。

1.3 BP 神经网络的基本原理

BPNN 是一种处理非线性问题的模型,一般由输入层、隐藏层和输出层构成,其基本原理是将神经网络学习的结果情况反馈给隐藏层,并调整权值和阈值,使得总误差达到最小,从而满足预期的学习要求。

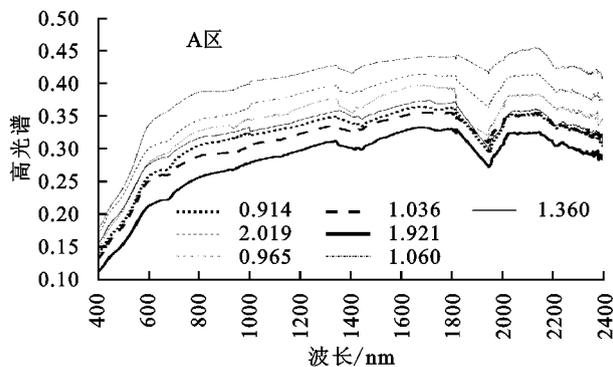
利用 Matlab 2019 年编程实现 BPNN 模型的训练过程,首先,选择 60% 的样本作为训练,40% 的样本作为测试,则 A 区的训练和测试样本分别为 15 个和 10 个, B 区的训练和测试样本分别为 18 个和 12 个。其次,对输入参数和输出参数进行归一化处理,易于网络的收敛。最后,设置训练参数,最大训练迭代次数为 1 000,学习速率为 0.001,隐藏层节点数为 100,训练目标的最小误差为 $1e^{-6}$,训练函数为 trainscg。

表 1 研究区阳离子含量统计特征

阳离子	A 区				B 区			
	最小值	最大值	均值	比例/%	最小值	最大值	均值	比例/%
Ca ²⁺	0.047	2.811	1.476	44.9	0.023	2.904	1.052	30.5
Mg ²⁺	0.014	0.157	0.066	2.0	0.014	1.200	0.157	4.6
K ⁺	0.047	0.313	0.152	4.6	0.029	0.224	0.092	2.7
Na ⁺	0.640	4.890	1.590	48.4	0.622	8.299	2.148	62.3

2.2 土壤的高光谱曲线特征

图 2A 和图 2B 为在 A 区和 B 区中选择前 7 个土壤样本的高光谱反射率曲线。由于土壤的高光谱是土壤中多种组分信息(盐分、有机质、氮磷钾等)和周围环境(土壤类型、植被、气象等)的综合反映,Field-



1.4 模型精度验证指标

本研究采用 BPNN 模型来建模预测 Na⁺ 离子含量,并对分析偏最小二乘(PLSR)和逐步多元回归(SMLR)模型的预测效果。本研究采用决定系数(R^2),均方根误差(RMSE)和相对分析误差(RPD) 3 个参数指标来评估模型的精度^[13]。 R^2 代表模型的拟合程度,其值越接近于 1,说明拟合程度越好。RMSE 代表样本的真实值与模型的预测值的偏差。RPD 代表模型的预测性能,一般 $RPD \geq 2.0$ 代表该模型的预测性能很好, $1.8 \leq RPD < 2.0$ 代表模型的预测性能较好, $1.4 \leq RPD < 1.8$ 代表模型的预测性能一般, $1.0 \leq RPD < 1.4$ 代表模型的预测性能很差, $RPD < 1.0$ 代表模型不具有预测能力。

2 结果与分析

2.1 研究区的 4 种阳离子含量统计

易溶盐(硫酸钠、氯化钠)是新疆盐渍土的主要类型,Na⁺ 在盐渍土的形成过程中起着重要的作用。研究区化验的 4 种阳离子(Ca²⁺, Mg²⁺, K⁺, Na⁺)统计特征见表 1。Na⁺ 在 A 区和 B 区中所占比例最高分别为 48.4% 和 62.3%,其均值在 4 种阳离子中也最大分别为 1.590, 2.148。

Spec[®]3 Hi-Res 光谱仪测量得到的高光谱反射率曲线与某一具体元素之间呈现为非线性关系,图 2 说明高光谱与 Na⁺ 离子含量为非线性关系,而不是单调的递增或递减的线性关系。

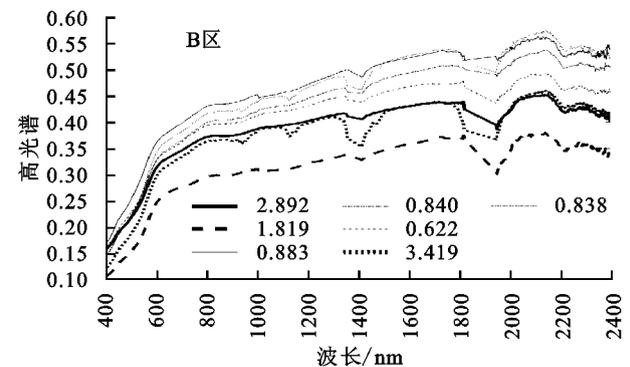


图 2 土壤的高光谱反射率曲线

2.3 选择特征波段

原始高光谱与 Na⁺ 离子的相关系数如图 3 所

示。倒数变换后的高光谱与 Na⁺ 离子的相关系数如图 4 所示。图 3 和图 4 均表明 A 区和 B 区中的 0 阶

微分都没有相关系数通过 0.05 显著性检验,一阶微分和二阶微分有大量波段的相关系数通过 0.05 显著性检验,统计结果为原始高光谱的一阶微分和二阶微分通过 0.05 检验的波段数量,在 A 区分别 92 个和 107 个,在 B 区分别为 119 个和 57 个;倒数变换后高光谱的一阶微分和二阶微分通过 0.05 检验的波段数

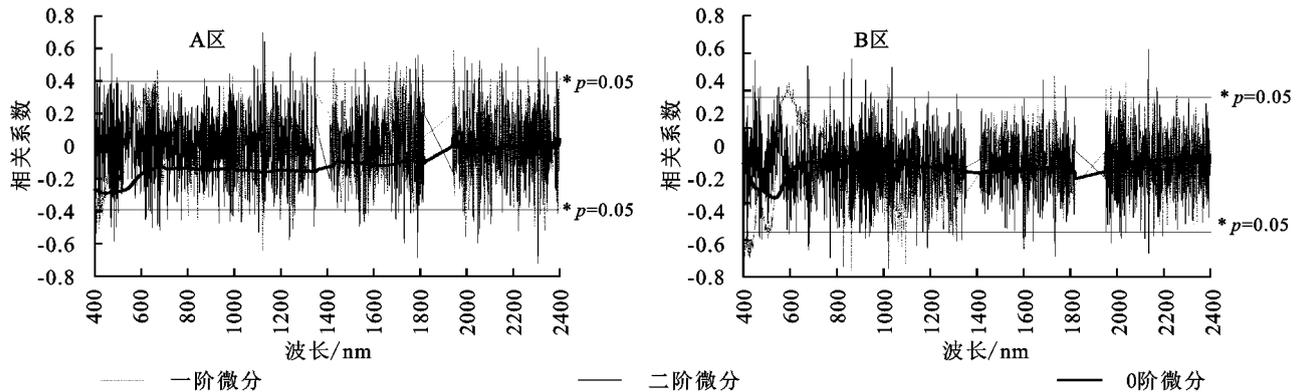


图3 原始高光谱与 Na^+ 离子的相关系数

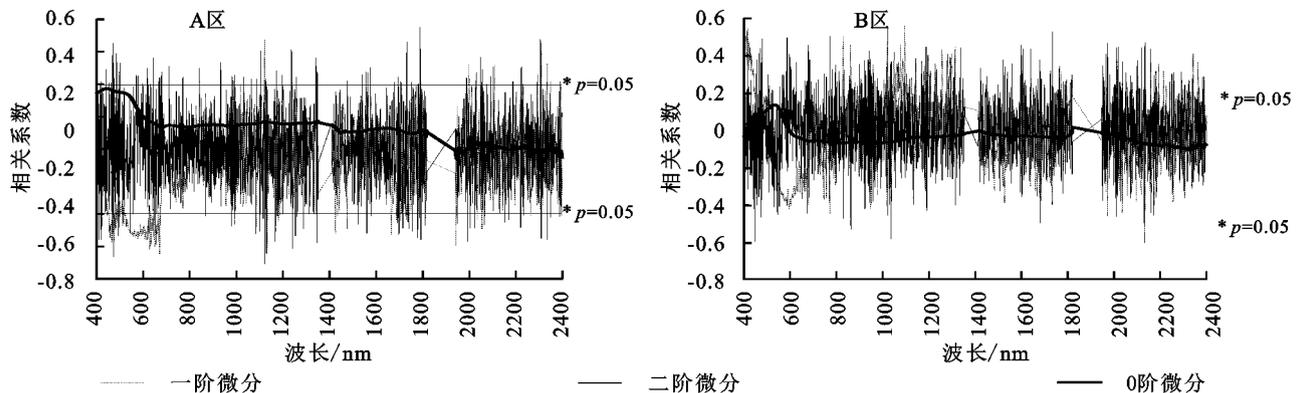


图4 倒数变换后的高光谱与 Na^+ 离子的相关系数

2.4 钠离子高光谱估算模型精度比较

采用 BPNN, PLSR 和 SMLR 这 3 种模型估算 Na^+ 离子含量,两种光谱变换下的模型结果见表 2。在 A 区, BPNN 模型在两种光谱变换中的 RPD 值均大于 2.0,其预测能力很好,且最好的预测效果属于 1/R 的二阶微分,其 RPD 值为 2.461 6, R^2 为 0.860 9, RMSE 为 0.350 1。在 B 区, BPNN 模型在原始高光谱变换中, RPD 值在 1.8 与 2.0 之间,表明其预测能力好; BPNN 模型在倒数变换中, RPD 值大于 2.0,表明其预测能力很好,且最好的预测效果属于 1/R 的二阶微分,其 RPD 值为 2.169 8, R^2 为 0.800 6, RMSE 为 0.803 5。而在 A 区和 B 区中, PLSR 模型在两种光谱变换中的 RPD 值均在 1.4~1.8 之间,其预测能力一般; SMLR 模型在两种光谱变换中的 RPD 值均在 1.0~1.4 之间,其预测能力很差。表 2 显示 BPNN 模型在 A 区对 Na^+ 离子的预测精度高于 B 区,这是因为 A 区基本没有受到人类日常活动的影响,其土壤基本维持着原有的生态风貌;而 B 区

量,在 A 区分别为 248 个和 140 个,在 B 区分别为 125 个和 75 个。并且,一阶微分和二阶微分对应的相关系数曲线中包含更多的吸收峰和吸收谷。本研究将 A 区和 B 区中通过 0.05 检验的一阶与二阶微分对应的波段选择为特征波段,其对应的高光谱反射率为 BPNN 模型的输入变量。

受到人类日常活动的影响较大,且土壤被人类犁地、翻耕和种植育苗林地,破坏土壤表层的物理结皮和生物结皮,又因研究区域地处干旱区,土壤的水分蒸发非常强烈,土壤中的盐分随着水分的蒸发被带到土壤表层,呈现盐分表聚的现象, B 区土壤的变化更为复杂,其预测难度高于 A 区,因此 A 区的预测精度高于 B 区。

1/R 光谱变换在二阶微分处的 BPNN 模型均是 A 区和 B 区的最佳预测模型,其真实值与预测值的散点图如图 5 所示,其 BPNN 模型的训练过程如图 6 所示。图 5 显示,验证集中 A 区的 R^2 为 0.860 87,拟合方程为 $y = 1.2453x + 0.45153$ 。验证集中 B 区的 R^2 为 0.800 6,拟合方程为 $y = 1.0287x + 0.3654$ 。图 6 显示, A 区 BPNN 模型的输入变量为 140 个高光谱,隐藏层的节点个数为 100,输出层有 1 个变量,共迭代 19 次满足精度要求。 B 区 BPNN 模型的输入变量为 75 个高光谱,隐藏层的节点个数为 100,输出层有 1 个变量,共迭代 9 次满足精度要求。

表 2 验证集中不同建模方法的精度比较

光谱变换	模型	A 区			B 区			
		R^2	RMSE	RPD	R^2	RMSE	RPD	
原始 高光谱(R)	BPNN	一阶微分	0.8074	0.5686	2.1815	0.7140	0.8197	1.8846
		二阶微分	0.8854	0.6418	2.0593	0.7761	0.9886	1.8114
	PLSR	一阶微分	0.6806	0.7312	1.4127	0.5782	1.3900	1.5378
		二阶微分	0.7223	0.7941	1.7581	0.5943	1.1620	1.6393
	SMLR	一阶微分	0.5132	0.7544	1.2287	0.3640	1.1778	1.2540
		二阶微分	0.5470	0.9999	1.1845	0.2107	1.4085	1.0523
原始高光谱 倒数变换(1/R)	BPNN	一阶微分	0.9052	0.2875	2.2765	0.7913	0.9181	2.0036
		二阶微分	0.8609	0.3501	2.4616	0.8006	0.8035	2.1698
	PLSR	一阶微分	0.6025	0.5501	1.4055	0.4924	0.9056	1.4277
		二阶微分	0.7435	0.3219	1.7782	0.5950	1.4690	1.5914
	SMLR	一阶微分	0.2795	0.5296	1.0638	0.2878	2.9772	1.0011
		二阶微分	0.3952	0.6217	1.1648	0.2514	1.2113	1.0906

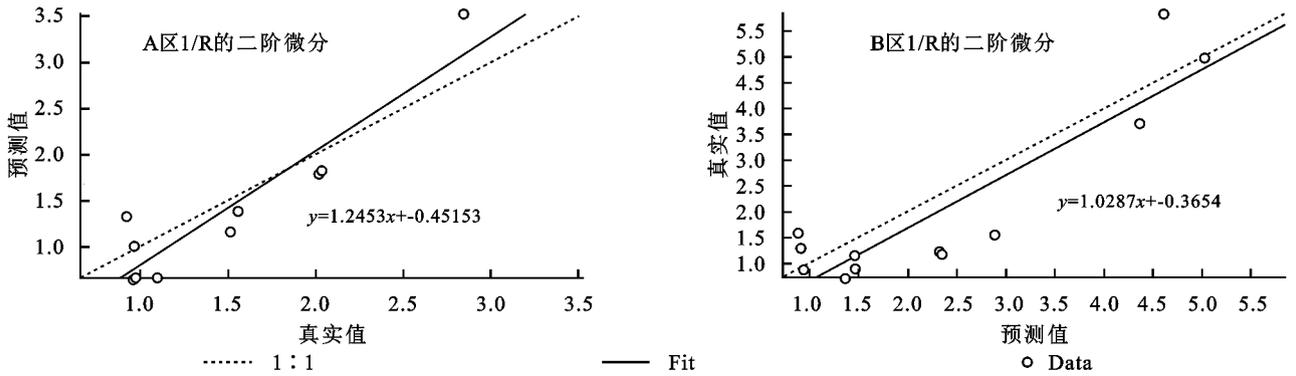


图 5 BPNN 模型的真实值与预测值的散点图

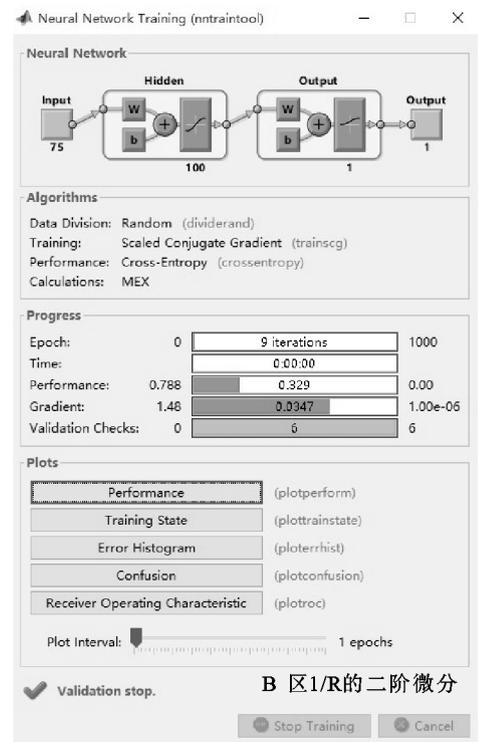
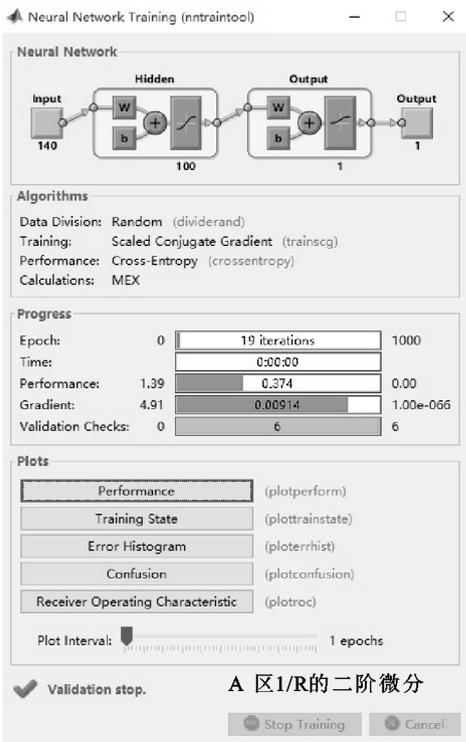


图 6 BPNN 模型的训练过程

3 结论

(1) BPNN 模型的预测精度高, 训练时间短, 迭

代次数少, 且 1/R 二阶微分处理后的 BPNN 均是 A 区和 B 区的最佳反演模型。在验证集中, BPNN 模型在 A 区 $RPD=2.4616, R^2=0.8609$, 在 B 区 $RPD=$

2.1698, $R^2 = 0.8006$ 。

(2) 两种光谱变换在 A 区和 B 区中,PLSR 模型的 RPD 值均在 1.4~1.8 之间,其预测能力一般;SMLR 模型的 RPD 值均在 1.0~1.4 之间,其预测能力很差。说明在不同研究区域中,BPNN 模型均表现出最优的预测结果。

参考文献:

- [1] 阿布都热合曼·哈力克.新疆岳普湖县土壤盐分空间变异及其分布特征研究[J].水土保持研究,2011,18(1):97-100.
- [2] 王飞,杨胜天,丁建丽,等.环境敏感变量优选及机器学习算法预测绿洲土壤盐分[J].农业工程学报,2018,34(22):102-110.
- [3] Wichelns D, Qadir M. Achieving sustainable irrigation requires effective management of salts, soil salinity, and shallow groundwater[J]. *Agricultural Water Management*, 2015,157:31-38.
- [4] 林鹏达,佟志军,张继权,等.基于 CWT 的黑土有机质含量野外高光谱反演模型[J].水土保持研究,2018,25(2):46-52,57.
- [5] 王明宽,莫宏伟,陈红艳.基于近地高光谱的土壤氯离子含量估测[J].水土保持通报,2017,37(6):214-219.
- [6] 郑雯,明金,杨孟克,等.基于波段深度分析和 BP 神经网络的水稻色素含量高光谱估算[J].中国生态农业学报,2017,25(8):1224-1235.
- [7] 龚巧灵,官冬杰.基于 BP 神经网络的三峡库区重庆段水资源安全评价[J].水土保持研究,2017,24(6):292-299.
- [8] 刁万英,刘刚,胡克林.基于高光谱特征与人工神经网络模型对土壤含水量估算[J].光谱学与光谱分析,2017,37(3):841-846.
- [9] 卢志宏,刘辛瑶,常书娟,等.基于 BP 神经网络的草原矿区表层土壤 N/P 高光谱反演模型[J].草业科学,2018,35(9):2127-2136.
- [10] 郭云开,刘宁,刘磊,等.土壤 Cu 含量高光谱反演的 BP 神经网络模型[J].测绘科学,2018,43(1):135-139,152.
- [11] 段鹏程,熊黑钢,李荣荣,等.不同干扰程度的盐渍土与其光谱反射特征定量分析[J].光谱学与光谱分析,2017,37(2):571-576.
- [12] 乔娟峰,熊黑钢,王小平,等.基于最优模型的荒地土壤有机质含量空间反演[J].江苏农业学报,2018,34(1):68-75.
- [13] Kuang B, Mouazen A M. Effect of spiking strategy and ratio on calibration of on-line visible and near infrared soil sensor for measurement in European farms[J]. *Soil & Tillage Research*, 2013,128:125-136.
- [13] 申建双,叶远俊,潘会堂,等.12 份连翘种质资源的核型参数分析[J].植物遗传资源学报,2015,16(1):178-184.
- [14] 聂小飞,郑海金,左继超,等.枯落物覆盖对红壤坡地土壤水分特性的影响[J].水土保持学报,2016,30(6):85-95.
- [15] 廖凯涛,宋月君,张金生,等.无人机遥测技术在水土保持生态果园改造监测中的应用[J].中国水土保持科学,2017,15(5):135-141.
- [16] 姜玉印,楚翠娇.中国长序列资料测站月降水、月均温相关联系的稳定性分析[J].贵州气象,2011,35(3):7-12.
- [17] 包洪福,孙志禹,陈凯麟.南水北调中线工程对丹江口库区生物多样性的影响[J].水生态学杂志,2015,36(4):14-19.
- [18] 全国农业技术推广服务中心.土壤分析技术规范[M].2 版.北京:中国农业技术出版社,2006.
- [19] 郑纯辉,康跃虎,姚素梅,等.基于地理信息系统的植物根系分析方法[J].农业工程学报,2004,20(1):181-183.
- [20] 王彩华,宋连天.模糊论方法学[M].北京:中国建筑工业出版社,1988.
- [21] 李禄军,蒋志荣,李正平,等.3 树种抗旱性的综合评价及其抗旱指标的选取[J].水土保持研究,2006,13(6):253-254.
- [22] 朱慧芬,张长芹,龚洵.植物引种驯化研究概述[J].广西植物,2003,23(1):52-60.
- [23] Mayr H. Die Naturgesetlicher Grundlage des Waldbaues[M]. Berlin: Parey, 1909.
- [24] 郎莹,汪明.春、夏季土壤水分对连翘光合作用的影响[J].生态学报,2015,35(9):3043-3051.
- [25] 李许文,刘文,陈红锋,等.广州适宜的植物引种来源地与气候区选择研究[J].中国园林,2016(4):96-100.
- [26] 杨洋,卫海燕,王丹,等.连翘潜在地理分布预测模型比较[J].生态学杂志,2016,35(9):2562-2568.
- [27] 杜钦,杨淑慧,任文玲,等.植物根系固岸抗蚀作用研究进展[J].生态学杂志,2010,29(5):1014-1020.
- [28] 朱显谟.黄土地区植被因素对于水土流失的影响[J].土壤学报,1960,8(2):110-121.
- [29] 单建平,陶大立.国外对树木细根的研究动态[J].生态学杂志,1992,11(4):46-49.
- [30] 张祖荣.植物根系提高土壤抗侵蚀能力的初步研究[J].渝西学院学报:自然科学版,2002,15(1):31-35.
- [31] 王同顺,孙保平,冯磊,等.不同水分处理对甘蒙柽柳幼苗根系生长特性的影响[J].生态学杂志,2013,32(3):591-596.

(上接第 363 页)