

文章编号: 1000-7032(2018)10-1458-08

利用改进自动编码器光谱法预测土壤有机质

史 杨^{1,2}, 王儒敬^{1,2*}, 汪玉冰¹

(1. 中国科学院合肥智能机械研究所, 安徽合肥 230031;

2. 中国科学技术大学自动化系, 安徽合肥 230027)

摘要: 提出一种改进自动编码器方法,用于利用近红外光谱预测大尺度下土壤有机质含量等级。首先,提出改进自动编码器算法框架,将传统的用于重建输出的自动编码器与分类器相结合;对改进自动编码器中的损失函数进行定义。然后,将改进自动编码器应用于预测土壤有机质含量等级的近红外光谱分析建模问题中,使用双层前馈神经网络实现了改进自动编码器的编码器、解码器和分类器。最后,使用大尺度土壤光谱数据集对模型进行训练,预测土壤有机质含量等级,并与主成分回归、支持向量机等方法的效果进行对比。实验结果表明,基于改进自动编码器的土壤有机质含量等级分类准确率为63.05%,高于其他方法。利用该模型预测大尺度下土壤有机质含量等级有较好的表现。

关键词: 近红外光谱; 自动编码器; 土壤有机质; 建模方法

中图分类号: TP394.1; TH691.9

文献标识码: A

DOI: 10.3788/fjxb20183910.1458

Prediction of Soil Organic Matter by Improved Auto Encoder Based on Near-infrared Spectroscopy

SHI Yang^{1,2}, WANG Ru-jing^{1,2*}, WANG Yu-bing¹

(1. Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, China;

2. Department of Automation, University of Science and Technology of China, Hefei 230027, China)

* Corresponding Author, E-mail: rjwang@iim.ac.cn

Abstract: This paper presents a calibration model, namely, improved auto encoder, which can be used to predict the grade of soil organic matter content in large scale based on near infrared spectroscopy. First, the framework of improved auto encoder model was proposed, which combined traditional auto encoder and classifier, and loss function of the model is defined. Then, the proposed improved auto encoder was applied to predict the grade of soil organic matter content based on near infrared spectroscopy. The encoder, decoder and classifier were implemented with two-layer feed-forward neural networks. Finally, a large scale soil spectral dataset was used to train the model for predicting the grade of soil organic matter content. The performance was compared with the results of principal component regression and support vector machine. The results show that the classification accuracy of soil organic matter content grades based on the proposed improved auto encoder model is 63.05%, which is better than other methods. This model can be used to predict the grades of soil organic matter content in a large scale.

Key words: NIR spectroscopy; auto encoder; soil organic matter; modeling

收稿日期: 2017-12-20; 修订日期: 2018-03-06

基金项目: 国家自然科学基金(31671586)资助项目

Supported by National Natural Science Foundation of China(31671586)

1 引 言

土壤养分含量分析对农业生产、研究非常重要,是研究土壤肥力分布、精准施肥、农田资源管理等的基礎。尽管传统实验室检测方法精度很高,但局限于时间成本和经济成本,在生产研究中的应用难以大规模开展。近红外光谱分析技术出现后,由于其非接触式信息获取、检测成本低廉的特性,应用前景广阔,因此能否利用该技术对土壤养分进行快速检测吸引了大量研究人员的关注^[1-2]。尽管利用土壤近红外光谱在预测精度上不及实验室方法直接检测,但是当检测样本的数量巨大时(如土壤制图),近红外光谱是一种低成本的、有效的信息来源^[3]。利用近红外光谱技术对土壤各种信息进行间接获取的研究工作正在大量开展,已尝试应用在土壤质地分类^[4]、含水量预测^[5-6]、氮素含量预测^[7-9]、有机质含量预测^[10]等方面。

近红外光谱仪器获得土壤样品光谱曲线,目的是获得土壤信息,因此近红外光谱分析是一种间接获取信息的手段,需要通过化学计量学建立光谱与实验室测试土壤信息之间的校正模型,再将模型应用在未知土壤样本信息的预测中,校正模型如何建立直接影响预测的准确性^[11]。在利用近红外光谱对土壤成分进行定量分析时,常采用多元线性回归(MLR)、主成分回归(PCR)、偏最小二乘回归(PLSR)等线性回归方法。光谱数据通常维度较高,不同波长的变量之间多重相关,因此在进行光谱分析时,基于 PCR、PLSR 等的模型使用线性变换对高维度光谱数据进行降维。近年来,研究人员开始尝试将机器学习技术应用到光谱分析建模中,以提升预测效果。Nawar 等^[12-13]使用支持向量回归(SVR)、随机森林、神经网络、梯度提升机等算法对土壤中的有机质、黏土和总碳含量进行预测,比常用的线性回归方法表现更优秀。纪文君等^[14]发现,在使用全谱数据进行挖掘来预测有机质含量时,先利用 PLSR 提取出若干主成分,再将其作为多层神经网络的输入进行建模,可以获得较好的预测精度。在较大面积的土壤成分预测应用中,由于土壤差异性较大,而线性模型表示能力有限,因此采用局部建模方法或者考虑模型容量更大的非线性模型。局部建模方法的思路是从大规模的光谱数据库中根据

定义的光谱距离量度,选出距离相近的样本,仅利用这些相近的样本进行建模预测,是一种基于内存的和基于模型的混合方法。陈颂超等^[15]使用局部加权回归算法成功预测了五个省范围的土壤全氮含量。自动编码器模型是基于神经网络的非线性模型,以重建输入为目标,训练后的网络可以获得高维数据的非线性特征表示,降低了输入数据的维度,可作为后续分类、回归模型的输入^[16]。

本文提出了一种新的改进自动编码器算法,将传统的用于重建输出的自动编码器与分类器相结合,即构建一个多输出的神经网络模型,同时获得输入光谱的非线性特征表示和非线性分类器的分类结果,并将其应用在土壤近红外光谱预测有机质含量等级问题中。实验证明,利用改进自动编码器模型预测土壤有机质含量等级比其他分类方法的准确率更高。

2 基于改进自动编码器的光谱分析

2.1 自动编码器模型

自动编码器模型属于神经网络模型,可以认为是前馈神经网络的一个特殊形式。自动编码器由编码器和解码器两部分组成。输入向量 X , 通过编码器 $f_{\text{encoder}}(\cdot)$ 产生编码表示 R , 编码表示 R 再通过解码器 $f_{\text{decoder}}(\cdot)$ 产生输入向量 X 的重建 X_{rec} , 如下式所示:

$$R = f_{\text{encoder}}(X), \quad (1)$$

$$X_{\text{rec}} = f_{\text{decoder}}(R), \quad (2)$$

其中编码器 $f_{\text{encoder}}(\cdot)$ 和解码器 $f_{\text{decoder}}(\cdot)$ 常采用多层神经网络实现。自动编码器的训练过程就是不断更新编码器和解码器的神经网络模型参数,以最小化重建 X_{rec} 与输入向量 X 之间的差异。如果编码表示 R 的维度小于输入向量 X 的维度,则该编码器为欠完备自动编码器。训练欠完备自动编码器时,自动编码器会捕捉训练数据中的显著特征,这种特性常被用于数据降维或特征提取^[17]。如果编码器和解码器均为非线性函数,训练得到的编码表示 R 即是原始输入信号 X 的非线性特征表示,可用于后续分类或回归模型的输入。

2.2 改进自动编码器算法

在使用多层神经网络作为自动编码器中编码器和解码器的实现对真实土壤样本光谱信号进行特征提取时,由于实现编码器和解码器的神经网

络结构的模型容量较大,得到的非线性特征尽管能很好地重建输入信号,但是在后续的预测土壤有机质等级时,往往效果较差。在对输入光谱进行降维或特征表示时,传统的自动编码器与 PCA 算法类似,仅考虑输入光谱 X 的特征。PLSR 算法对输入光谱 X 进行分解时,考虑了预测输出 Y 的分布,通常预测效果更好。受该思路启发,本文提出一种新的改进自动编码器模型,将重建输入信号的自动编码器训练过程与预测土壤有机质含量的分类器训练过程结合起来。

图 1 为改进自动编码器的结构示意图,其中虚线框内为传统的自动编码器。传统的自动编码器与预测分类器结合起来,形成一个单输入多输出的神经网络模型。训练改进自动编码器时得到的原始输入的特征表示 R ,既能使用解码器很好地重建原始输入,又能准确地预测土壤有机质含量的级别。

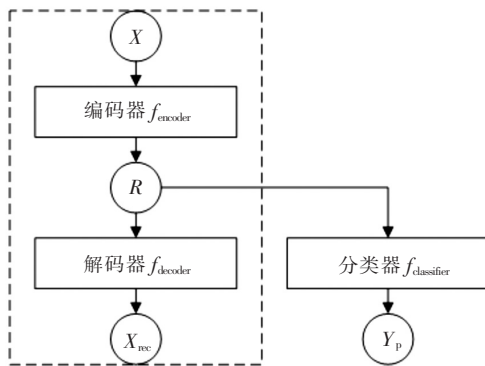


图 1 改进自动编码器结构

Fig. 1 Structure of improved auto encoder

2.3 改进自动编码器模型的训练

反向传播算法建立在梯度下降法的基础上,常用来对多层前馈神经网络进行训练。反向传播算法由正向传播过程和反向传播过程两部分组成。在改进自动编码器模型训练的正向传播过程中,输入样本经过编码器得到特征向量,然后分别通过解码器、分类器分别得到输入样本光谱曲线的重建输出和有机质含量级别分类的预测输出。将输出与监督信息进行比对后,进行神经网络损失函数的计算,计算的损失将作为反向传播过程修改神经网络各层参数的依据。

对于改进自动编码器中解码器的样本重建输出,其监督信号为输入信号,采用下式计算均方损失:

$$J_{\text{rec}}(W, b) = \frac{1}{N} \|X - X_{\text{rec}}\|^2, \quad (3)$$

其中 W 和 b 分别为神经网络的权值和偏置, N 为样本数。

对于改进自动编码器中分类器的土壤有机质含量分类预测输出,其监督信号为有机质含量分类的标注信息,由于采用 softmax 层作为多分类问题的输出层,采用损失函数如下:

$$J_p(W, b) = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^D \mathbf{1}\{y_p^{(i)} = j\} \cdot$$

$$\log p(y_p^{(i)} = j | x^{(i)}; W, b), \quad (4)$$

其中 $\mathbf{1}\{y_p^{(i)} = j\}$ 为示性函数,即大括号中条件满足时函数值取 1; $x^{(i)}$ 和 $y_p^{(i)}$ 分别表示第 i 个样本的输入和预测类别; $p(y_p^{(i)} = j | x^{(i)}; W, b)$ 表示在神经网络参数权值和偏置取值为 W 和 b 时,输入样本 $x^{(i)}$, 预测输出为第 j 类的概率,即 softmax 输出层的对应数值; D 为土壤有机质等级分类的类别总数。

对于改进自动编码器整个神经网络结构来说,由于两种输出损失量纲上的差异,采用的损失函数为两者的加权和,即

$$J(W, b) = J_{\text{rec}}(W, b) + \eta \cdot J_p(W, b), \quad (5)$$

反向传播过程中,由输出层到输入层逐层计算损失函数对各层权值、偏置的偏导数,更新神经网络模型中的对应参数数值。改进自动编码器参数训练时,将建模集中的样本反复循环迭代,神经网络参数不断依此修改,根据神经网络在建模集上的效果以及验证集上的效果综合评估性能变化,确定训练是否完成。

2.4 分类模型效果评估

对于多分类问题的评估,常采用混淆矩阵、准确率、精确率、召回率、F1 分数等指标进行综合评估。混淆矩阵是评估样本的真实分类和模型预测类别的汇总,准确率是所有分类正确的样本占有所有样本的比例。对于具体某一类别,精确率表示正确分为此类的样本数与预测分为此类的样本数之比,召回率表示正确分为此类的样本数占应分为此类的样本数之比, F1 分数为前两者的调和均值;对于多分类问题,精确率、召回率和 F1 分数为各分类对应评价的加权平均数,权值为预测为该类别的样本数,此时准确率与召回率的数值是一致的。

3 实验及分析

3.1 LUCAS 土壤数据集及处理

实验采用的数据集来自 LUCAS 土壤数据集。LUCAS 土壤数据集包含 2008—2012 年欧盟开展欧洲土地利用及覆盖统计调查 (European Land Use/Cover Area frame Statistical Survey, LUCAS) 期间收集的大量土壤样本, 其采样点遍及欧洲 23 个国家^[18-19]。LUCAS 数据集中的所有采样点使用了一致的样本收集方法, 土壤样本的理化特性分析由 ISO 认证的实验室完成。

LUCAS 数据集中包含矿质土样共 17 272 个。土壤样本的有机碳含量依据 ISO 10694-1995 干烧方法进行测量, 数据集中有机碳含量基本信息统计见表 1。土壤样本经风干、过筛处理后, 消除了土壤水分、质地、结构及紧实度等因素对光谱的影响^[5], 最后使用 FOSS XDS 近红外光谱分析仪进行光谱测量, 波长范围为 400 ~ 2 500 nm, 光谱数据间隔为 0.5 nm。

表 1 LUCAS 土壤数据集基本信息

Tab. 1 Basic information of LUCAS soil dataset

采集地点	样本数	有机碳含量/(g · kg ⁻¹)		
		最小值	最大值	均值
欧洲 23 国	17 272	0	165.7	25.2

土壤中的有机质含量使用土壤中的有机碳比例乘以系数 1.724 进行换算^[1], 并依据中国第二次土壤普查养分分级标准进行分类, 分类后样本分布见表 2。

表 2 LUCAS 数据集的有机质含量分级

Tab. 2 Classification of SOM in LUCAS dataset

土壤等级	土壤有机质含量/(g · kg ⁻¹)	样本数
一级	>40	6 791
二级	30 ~ 40	2 616
三级	20 ~ 30	3 861
四级	10 ~ 20	3 360
五级	6 ~ 10	469
六级	<6	175

为了对所建立模型的评估具有说服力, 从数据集包含的 17 272 个土壤样本中随机选择 15 000 个样本作为建模集, 再从剩余的样本里选择 1 000 个样本作为验证集, 最后剩余的 1 272 个样本作

为最终评价建模性能的测试集。建模集、验证集、测试集中的土壤样本独立不交叉。

3.2 光谱曲线特征

图 2 是将土壤样本按照有机质含量等级进行分类后, 对每一个类别中所有土壤样本的光谱计算平均, 得到了不同有机质含量级别的光谱平均曲线。从图 2 中可以看出, 不同有机质含量等级的光谱均在 1 400, 1 900, 2 200 nm 左右有明显的峰值, 整体光谱曲线趋势一致; 有机质含量级别越高的土壤样本的平均光谱在整个可见光近红外波段吸光度都高于有机质含量级别较低类别。

由于有机质含量在 20 g/kg 以下区间是以 10 g/kg, 6 g/kg 为分界分成了四级到六级, 这 3 个类别的光谱平均曲线比较接近。

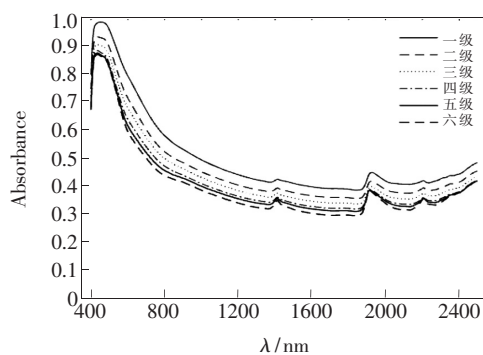


图 2 不同有机质含量等级的土壤平均光谱曲线

Fig. 2 Mean spectra of soil samples with different grades of soil organic matter

3.3 改进自动编码器的实现

改进自动编码器中的编码器、解码器和分类器可以通过不同的模型实现, 在基于土壤近红外光谱的有机质含量等级分类应用中, 编码器、解码器、分类器均通过多层前馈神经网络实现。调节神经网络模型层数及各层的属性 (如全连接层、卷积层等) 可以控制模型容量、特征种类。实验中, 编码器、解码器和分类器均采用两个全连接层实现 (图 3), 3 个部分各层的神经元数目和激活函数, 根据多次尝试后确定。表 3 为最终实现的改进自动编码器中各层神经网络的神经元数目和激活函数的组合。

由于近红外光谱原始数据的维度很高, 而光谱数据中存在较强的共线性, 因此在输入模型前, 将原始的 4 200 维光谱依据波长等间隔采样为 525 维的数据作为模型的输入, 大大减少了神经网络中需要训练参数的数量。

传统神经网络算法中常使用双曲正切函数或 Sigmoid 函数作为神经元的激活函数,然而应用在较深层网络时,常发生神经元饱和、梯度扩散的问题。采用修正线性单元 (Rectified linear unit, ReLU) 能有效避免梯度扩散问题,并具有加快网络训练的效果^[20-22],因此在改进自动编码器的实现中,E2、E3、D2、C2 层均采用了 ReLU 作为神经元激活函数。考虑到 D3 层的输出为光谱信号的重建,考虑其取值范围,采用线性单元作为激活函数;C3 层输出为土壤有机质含量等级的类别,采用输出 6 类的 Softmax 函数作为激活函数。

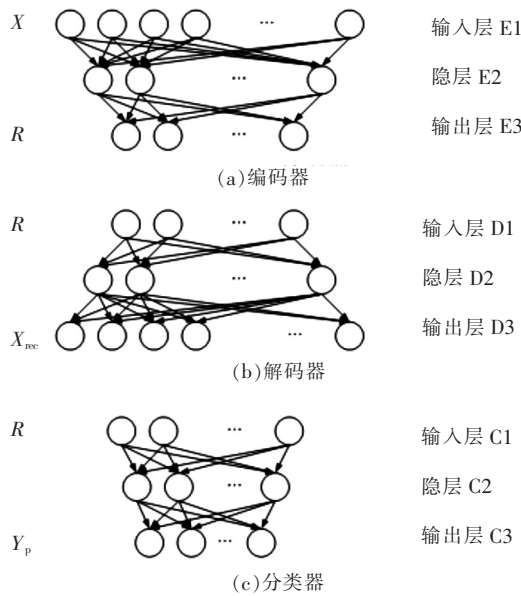


图 3 改进自动编码器的编码器、解码器、分类器的实现。
Fig.3 Implementation of encoder, decoder and classifier in proposed improved auto encoder.

表 3 改进自动编码器各层超参数和激活函数

Tab.3 Hyper parameters and activation function of layers in proposed improved auto encoder

层名	神经元数目	激活函数
输入层 E1	525	-
隐层 E2	200	ReLU
输出层 E3	40	ReLU
输入层 D1	40	-
隐层 D2	200	ReLU
输出层 D3	525	Linear
输入层 C1	40	-
隐层 C2	100	ReLU
输出层 C3	6	Softmax

从图 1、图 3 和表 3 可以看出,525 维的光谱信号通过编码器,首先维数降低到 200 维,然后被表示为 40 维的特征向量;特征向量通过与编码器近似对称的解码器重建为 525 维的光谱信号;另外,特征向量通过两层神经网络分类器分成了六类,即土壤有机质含量等级的分类。

3.4 基于改进自动编码器的土壤分类模型结果

训练改进自动编码器模型前,将 LUCAS 数据集中的所有样本的光谱信息统一进行归一化处理,并根据样本的有机质含量参照表 2 划分为 6 个级别。训练模型时,将建模集中土壤样本光谱作为编码器的输入信号和解码器的监督信号,将土壤样本有机质含量类别作为分类器的监督信号。改进自动编码器中的参数使用随机梯度下降法进行更新。训练完成后,将建模集、验证集、测试集中土壤样本光谱输入模型,预测对应样本的有机质含量等级。基于改进自动编码器的土壤有机质含量分类结果汇总见表 4、表 5。

表 4 基于改进自动编码器的土壤有机质等级分类结果

Tab.4 Classification results of soil organic matter based on improved auto encoder

数据集	准确率/%	精确率/%	召回率/%	F1 分数/%
建模集	85.84	85.88	85.84	85.82
验证集	59.80	60.08	59.80	59.86
测试集	63.05	62.98	63.05	62.99

表 5 基于改进自动编码器的土壤有机质等级分类在测试集上的混淆矩阵

Tab.5 Confusion matrix on test set of soil organic matter based on improved auto encoder

		预测分级					
		一级	二级	三级	四级	五级	六级
实际分级	一级	420	46	26	5	0	0
	二级	46	78	56	11	1	0
	三级	27	59	140	57	4	0
	四级	8	12	70	150	8	3
	五级	0	0	4	17	9	3
	六级	0	1	1	3	2	5

如表 4 所示,利用改进自动编码器模型对土壤有机质含量进行预测分类在建模集上的准确率为 85.84%,而在测试集上的准确率为 63.05%,

与验证集上的准确率 59.80% 比较接近。尽管模型在建模集上存在一定的过拟合现象,但模型总体泛化能力良好。实验结果表明,在利用土壤样本的近红外光谱预测有机质含量分级问题中,使用覆盖欧洲 23 国的、包含多种土壤的大尺度土壤数据集对提出的改进自动编码器进行训练,训练好的模型预测土壤有机质含量级别的准确率达到 63.05%;利用近红外光谱间接获取大尺度范围的土壤有机质含量信息具有可行性。

表 5 为利用改进自动编码器模型对土壤有机质含量进行预测分类在测试集上的混淆矩阵,其中,对角线上的数值为正确分类的样本个数。测试集一共包含 1 272 个土壤样本,其中一级到六级正确分类的样本数和级别总样本数占比分别为 420/497, 78/192, 140/287, 150/251, 9/33, 5/12。由混淆矩阵可以计算得出,在测试集中,一级土壤的分类精确率和召回率最高,分别为 83.83% 和 84.51%;四级土壤的分类精确率最低,仅为 37.50%;五级土壤的分类召回率最低,仅为 27.27%。

利用改进自动编码器模型对土壤光谱曲线进行重建的结果见图 4。其中,图片第一列为随机从 LUCAS 数据集中选取的两个样本的光谱曲线;图片第二列为对应样本经过编码器、解码器后重建的光谱曲线。由图 4 中样本光谱曲线的对比观察可以得出,重建得到的光谱曲线与原始曲线基本一致,保留了原始曲线的峰谷特征及数值特征。使用改进自动编码器可以有效地将 525 维原始光谱信息仅使用 40 维的特征向量进行表示,并能很好地保留原始光谱中的信息。

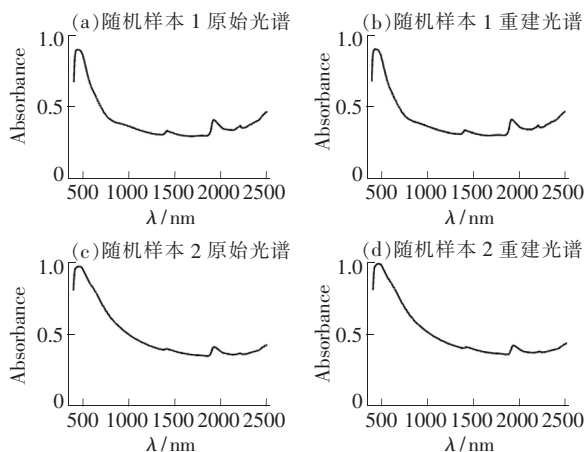


图 4 基于改进自动编码器的土壤光谱曲线重建结果

Fig. 4 Reconstruction of soil spectra based on improved auto-encoder

3.5 不同建模方法性能对比

为了更客观地了解改进自动编码器在土壤有机质含量等级预测的效果,本文还实现了常用于土壤成分预测建模的支持向量机模型、主成分回归模型,并在 LUCAS 土壤数据集上进行有机质含量等级的预测实验,模型的训练采用完全一致的建模集划分,结果评价在完全一致的测试集上进行。其中,支持向量机模型包括分类模型和回归模型两种,实验结果中用 SVM、SVR-C 表示;主成分回归模型实验结果中用 PCR-C 表示。需要说明的是,SVR-C、PCR-C 在建模时使用原始有机质含量数值作为监督输入,分别训练基于 SVR、PCR 的回归模型,再将回归模型预测的数值使用表 2 的有机质含量分级方法判定为各类别。SVM、SVR-C、PCR-C 模型与提出的改进自动编码器算法性能对比见表 6。

表 6 不同建模方法分类结果对比

Tab. 6 Comparison of classification results with different calibration models

数据集	准确率/%	精确率/%	F1 分数/%
改进自动编码器	63.05	62.98	62.99
SVM	56.37	50.16	50.61
SVR-C	55.82	56.58	55.25
PCR-C	51.65	50.82	50.24

如表 6 所示,利用改进自动编码器模型对土壤有机质含量进行预测分类在测试集上的效果最好,准确率、精确率、F1 分数分别为 63.05%、62.98% 和 62.99%,3 项指标均优于其他 3 种模型;SVM 和 SVR-C 方法性能比较接近,准确率分别为 56.37% 和 55.82%,而 SVM 模型的精确率稍低于 SVR-C 模型;PCR-C 模型准确率最低,为 51.65%。实验表明,在大尺度土壤数据集中,通过近红外光谱预测有机质含量级别,使用提出的改进自动编码器模型可以获得比常用的主成分回归模型、支持向量机模型更高的准确率。

利用近红外光谱预测有机质含量等级是一种间接方法,本文中的预测模型训练采用欧洲 23 国范围的土壤样本,因此,该模型可用于同样范围的真实土壤样本预测。模型使用时,应当尽量保证待测样本与训练样本采用同样的土壤采集方法、土壤预处理方法及光谱测量方法,并按同样的波长间隔进行采样后输入模型进行有机质含量等级

的预测。

4 结 论

本文研究了利用近红外光谱预测大尺度下土壤有机质含量等级的分类问题,提出一种改进自动编码器模型,将传统的用于重建输出的自动编码器与分类器相结合,并对改进自动编码器中的损失函数进行定义。然后利用样本覆盖欧洲 23 国、土壤差异性较大的 LUCAS 数据集对改进自动

编码器模型进行训练。最后,将改进自动编码器模型的预测性能与主成分回归、支持向量机等方法的效果进行对比。实验结果证明:利用近红外光谱间接获取土壤有机质含量信息具有可行性;在大尺度的土壤数据集中,基于本文提出的改进自动编码器模型分类准确率达到 63.05%,比常用的主成分回归、支持向量机等模型预测性能更好,基本满足了间接获取土壤有机质含量等级分类的要求。

参 考 文 献:

- [1] 宋海燕. 土壤近红外光谱检测 [M]. 北京:化学工业出版社, 2013.
SONG H Y. *Soil Near Infrared Spectroscopy Test* [M]. Beijing: Chemical Industry Press, 2013. (in Chinese)
- [2] 褚小立, 陆婉珍. 近五年我国近红外光谱分析技术研究与应用进展 [J]. 光谱学与光谱分析, 2014, 34(10):2595-2605.
CHU X L, LU W Z. Research and application progress of near infrared spectroscopy analytical technology in China in the past five years [J]. *Spectrosc. Spect. Anal.*, 2014, 34(10):2595-2605. (in Chinese)
- [3] RAV R, WEBSTER R. Predicting soil properties from the Australian soil visible-near infrared spectroscopic database [J]. *Eur. J. Soil Sci.*, 2012, 63(6):848-860.
- [4] 胡晓艳, 宋海燕. 基于支持向量机和近红外光谱特性的土壤质地分类 [J]. 山西农业科学, 2017, 45(10):1643-1645.
HU X, SONG H Y. Soil texture classification based on support vector machine and near infrared spectral characteristics [J]. *J. Shanxi Agricult. Sci.*, 2017, 45(10):1643-1645. (in Chinese)
- [5] 刘广霖, 郭焱, 劳彩莲, 等. 基于田间原位土壤含水量估测的可见/近红外光谱建模方法 [J]. 中国农业大学学报, 2016, 21(8):125-131.
LIU G L, GUO Y, LAO C L, *et al.*. Estimation of soil water content in situ by using visible/near infrared spectrum modeling [J]. *J. China Agricult. Univ.*, 2016, 21(8):125-131. (in Chinese)
- [6] 吴龙国, 王松磊, 何建国, 等. 基于高光谱成像技术的土壤水分机理研究及模型建立 [J]. 发光学报, 2017, 38(10):1366-1376.
WU L G, WANG S L, HE J G, *et al.*. Soil moisture mechanism and establishment of model based on hyperspectral imaging technique [J]. *Chin. J. Lumin.*, 2017, 38(10):1366-1376. (in Chinese)
- [7] 张瑶, 李民赞, 郑立华, 等. 基于近红外光谱分析的土壤分层氮素含量预测 [J]. 农业工程学报, 2015, 31(9):121-126.
ZHANG Y, LI M Z, ZHENG L H, *et al.*. Prediction of soil total nitrogen content in different layers based on near infrared spectral analysis [J]. *Trans. Chin. Soc. Agricult. Eng.*, 2015, 31(9):121-126. (in Chinese)
- [8] ZHANG Y, LI M Z, ZHENG L H, *et al.*. Soil nitrogen content forecasting based on real-time NIR spectroscopy [J]. *Comput. Electron. Agricult.*, 2016, 124(C):29-36.
- [9] NAWAR S, MOUAZEN A M. Predictive performance of mobile Vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques [J]. *Catena*, 2017, 151:118-129.
- [10] 王儒敬, 陈天娇, 汪玉冰, 等. 基于深度稀疏学习的土壤近红外光谱分析预测模型 [J]. 发光学报, 2017, 38(1):109-116.
WANG R J, CHEN T J, WANG Y B, *et al.*. Soil near-infrared spectroscopy prediction model based on deep sparse learning [J]. *Chin. J. Lumin.*, 2017, 38(1):109-116. (in Chinese)
- [11] 褚小立. 近红外光谱分析技术实用手册 [M]. 北京:机械工业出版社, 2016.
CHU X L. *Practical Handbook for Near Infrared Spectroscopy* [M]. Beijing: China Machine Press, 2016. (in Chinese)

- [12] NAWAR S, BUDDENBAUM H, HILL J, *et al.*. Estimating the soil clay content and organic matter by means of different calibration methods of Vis-NIR diffuse reflectance spectroscopy [J]. *Soil Tillage Res.*, 2016, 155:510-522.
- [13] NAWAR S, MOUAZEN A M. Comparison between random forests, artificial neural networks and gradient boosted machines methods of on-line Vis-NIR spectroscopy measurements of soil total nitrogen and total carbon [J]. *Sensors*, 2017, 17(11):2428.
- [14] 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.
- JI W J, LI X, LI C X, *et al.*. Using different data mining algorithms to predict soil organic matter based on visible-near infrared spectroscopy [J]. *Spectros. Spect. Anal.*, 2012, 32(9):2393-2398. (in Chinese)
- [15] 陈颂超, 冯来磊, 李硕, 等. 基于局部加权回归的土壤全氮含量可见-近红外光谱反演 [J]. *土壤学报*, 2015, 52(2):312-320.
- CHEN S C, FENG L L, LI S, *et al.*. Vis-NIR spectral inversion for prediction of soil total nitrogen content in laboratory based on locally weighted regression [J]. *Acta Pedolog. Sinica*, 2015, 52(2):312-320. (in Chinese)
- [16] HINTON G E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks [J]. *Science*, 2006, 313(5786):504-507.
- [17] IAN G, YOSHUA B, AARON C. *Deep Learning* [M]. Cambridge: MIT Press, 2016.
- [18] STEVENS A, NOCITA M, TOTH G, *et al.*. Prediction of soil organic carbon at the European scale by visible and near infrared reflectance spectroscopy [J]. *Plos One*, 2013, 8(6):e66409.
- [19] ORGIAZZI A, BALLABIO C, PANAGOS P, *et al.*. LUCAS soil, the largest expandable soil dataset for Europe: a review [J]. *Eur. J. Soil Sci.*, 2017, doi:10.1111/ejss.12499.
- [20] ALEX K, ILYA S, GEOFFREY E H. Image net classification with deep convolutional neural networks [J]. *Adv. Neur. Inform. Proc. Syst.* 25, 2012:1097-1105.
- [21] HE K M, ZHANG X Y, REN S Q, *et al.*. Deep residual learning for image recognition [C]. *2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas*, 2016:770-778.
- [22] SZEGEDY C, IOFFE S, VANHOUCKE V, *et al.*. Inception-v4, Inception-ResNet and the impact of residual connections on learning [C]. *Thirty-First AAAI Conference on Artificial Intelligence, San Francisco*, 2017:4278-4284.



史杨(1987 -),男,安徽肥东人,博士研究生,2011年于中国科学技术大学获得硕士学位,主要从事模式识别、农业信息化等方面的研究。
E-mail: yshi@mail.ustc.edu.cn



王儒敬(1964 -),男,安徽亳州人,博士,研究员,2004年于中国科学技术大学获得博士学位,主要从事农业智能系统的理论、方法与技术的研究。
E-mail: rjwang@iim.ac.cn