doi:10.6041/j.issn.1000-1298.2020.03.020

基于人工嗅觉系统的土壤有机质检测方法研究

朱龙图^{1,2} 李名伟^{1,2} 夏晓蒙^{1,2} 黄东岩^{1,2} 贾洪雷^{1,2}

(1. 吉林大学生物与农业工程学院, 长春 130022; 2. 吉林大学工程仿生教育部重点实验室, 长春 130022)

摘要:为了实现土壤有机质快速、准确的测量,提出了一种基于人工嗅觉的土壤有机质含量检测方法。首先,由不同温度控制的10个气体传感器所构成的阵列对土壤样品气体进行采集;然后,提取每个传感器响应曲线上的7个特征(包括最大值、最小值、平均值、平均微分系数、响应面积、第30秒的瞬态值和第60秒的瞬态值),构建嗅觉特征空间;对特征空间优化后,采用回归算法建立预测模型。为减小不同测定算法、异常样本以及冗余特征对模型预测性能的影响,在应用蒙特卡罗抽样(Monte Carlo sampling, MCS)法剔除异常样本的基础上,采用主成分分析(Principal component analysis, PCA)法对特征空间进行降维处理,评估了包括偏最小二乘法回归(Partial least square regression, PLSR)、支持向量机回归(Support vector machine regression, SVR)和 BP 神经网络(Back propagation neural network, BPNN)等3种建模方法对土壤有机质含量的预测性能,选用决定系数 *R*²、均方根误差(RMSE)和预测偏差比(RPD)评价各模型的预测性能。测试集验证结果表明, PLSR、SVR 和 BPNN 这3种模型的预测值和样本的观测值之间的 *R*²分别为 0.86、0.91 和 0.85, RMSE 分别为 2.49、2.05、2.68 g/kg, RPD 分别为 2.49、3.02 和 2.32。SVR 模型的预测性能高于 PLSR 模型和 BPNN 模型,可对土壤有机质含量进行准确预测。

关键词:土壤有机质;人工嗅觉系统;蒙特卡罗抽样;特征降维;预测模型

中图分类号: S158.2; TP212.9 文献标识码: A 文章编号: 1000-1298(2020)03-0171-09



Soil Organic Matter Detection Method Based on Artificial Olfactory System

ZHU Longtu^{1,2} LI Mingwei^{1,2} XIA Xiaomeng^{1,2} HUANG Dongyan^{1,2} JIA Honglei^{1,2}

(1. College of Biological and Agricultural Engineering, Jilin University, Changchun 130022, China
2. Key Laboratory of Bionic Engineering, Ministry of Education, Jilin University, Changchun 130022, China)

Abstract: In order to measure soil organic matter content quickly and accurately, a method based on artificial olfactory was proposed. Firstly, the response curves of soil gas were collected by an array composed of 10 gas sensors controlled at different temperatures. And then seven features, including the maximum value, minimum value, mean value, mean differential coefficient value, response area value, transient value at the 30th second and transient value at the 60th second were extracted from each sensor response curves to build an olfactory feature space. Finally, the prediction model was established by using the regression algorithm. To reduce the influence of different regression algorithms, abnormal samples and redundancy characteristics on the prediction performance of the model, the Monte Carlo sampling (MCS) method was used to eliminate abnormal samples, and the principal component analysis (PCA) method was used to reduce the dimension of olfactory feature space. Moreover, three modeling methods, including partial least square regression (PLSR), support vector machine regression (SVR) and back propagation neural network (BPNN), were used to predict soil organic matter content. And the predictive performance of each model were evaluated by coefficient of determination (R^2) , root mean square error (RMSE) and ratio of prediction derivation (RPD). The results showed that the R^2 values of PLSR, SVR and BPNN were 0.86, 0.91 and 0.85, respectively; the RMSE values were 2.49 g/kg, 2.05 g/kg and 2.68 g/kg, respectively; and the RPD values were 2.49, 3.02 and 2.32, respectively.

基金项目:国家重点研发计划项目(2016YFD070030201)和吉林省科技计划项目(20190302116GX)

收稿日期: 2019-12-06 修回日期: 2020-01-07

作者简介:朱龙图(1988一),男,博士生,主要从事保护性耕作技术与人工智能装备研究,E-mail: zhult17@ mails.jlu.edu.cn

通信作者:黄东岩(1976—),男,教授,博士生导师,主要从事农业机械自动化研究,E-mail: cchdy760829@ sina. com

The prediction performance of SVR model was higher than that of PLSR model and BPNN model, which can accurately predict the organic matter content. The results can provide a reference method for the prediction of soil organic matter.

Key words: soil organic matter; artificial olfactory system; Monte Carlo sampling; feature dimension reduction; prediction model

0 引言

土壤有机质是指土壤中各种含碳有机物,包括 植物和动物残体、土壤生物的细胞和组织,以及这些 生物残体不同阶段的分解物质^[1],它是土壤肥力和 养分的重要指标。土壤有机质对土壤的阳离子交换 能力、土壤结构、水分入渗率、持水能力、土壤的可蚀 性和保持性、农药吸附等理化、生物特性有很强的影 响^[2]。因此,测量土壤有机质含量、掌握其动态变 化对改善土壤结构和指导农业生产具有重要意义。

目前,测量土壤有机质含量的方法有重铬酸钾 容量法、CO₂检测法和灼烧法等^[3]。其中,重铬酸钾 容量法因测量结果准确、适用于批量测量而被广泛 使用,也是我国测量土壤有机质含量的标准方法。 然而,该方法需要在实验室进行分析处理,存在操作 复杂、耗时长、成本高和破坏性大的缺点^[4]。因此, 对快速、经济、无损、准确的土壤有机质含量预测方 法的需求越来越大^[5]。近年来,由于近端土壤遥感 技术的普遍应用,可见光和近红外(Vis – NIR)漫反 射光谱法受到关注,并被认为是一种可行的土壤分 析方法^[6-10]。光谱分析法虽然准确,但其缺点是易 受土壤粒度、土壤湿度和氧化铁的影响^[11-13]。

土壤中气体的产生和消耗主要与土壤中的微生 物活动过程有关^[14],而土壤有机质是土壤微生物生 命活动所需养分和能量的主要底物[15]。在微生物 的降解过程中,养分和能量的供应底物在土壤中产 生许多挥发性有机化合物(Volatile organic compounds, VOCs)和气体^[14]。因此,土壤中的 VOCs 和气体必然与土壤有机质存在某种相关性, 这种相关性为土壤有机质的快速、低成本检测提供 了可能。气体检测成本很低,尤其是基于固态化学 传感器的检测^[16]。然而,土壤气体的组成成分复 杂^[17],采用单一的气体传感器对其识别很困难。由 金属氧化物半导体 (Metal oxide semiconductor, MOS)气体传感器阵列和模式识别组成的人工嗅觉 系统(又称为电子鼻)被认为是实现复杂气体检测 的有效手段。虽然人工嗅觉系统不会给出任何关于 挥发性气体化合物的具体信息,也不会给出它们的 特性^[18],但是借助适当的模式识别算法,人工嗅觉 系统可以识别特定样本的气体模式,进而将不同样 本区分开来^[19-20]。目前,人工嗅觉系统在食品和饮料、医药、环境保护、工业生产和军事等领域有着广泛的应用^[21-24],且在土壤特性检测方面也有研究报道。例如,LAVANYA等^[25]应用电子鼻测量了土壤中腐植酸和黄腐酸的含量,ANDRZEJ等^[26]采用电子鼻评估了土壤湿度状况。但电子鼻(或称人工嗅觉系统)在土壤有机质含量方面的检测却鲜有文献报道。

针对上述问题,本文提出一种基于人工嗅觉系统的土壤有机质检测方法。首先,采用10个由不同 温度控制的 MOS 气体传感器构建检测阵列;然后, 以此阵列获取土壤挥发性有机化合物的响应曲线; 通过提取曲线上的最大值、最小值、平均值、平均微 分系数、响应面积、第30秒的瞬态值以及第60秒的 瞬态值等7个特征来构建嗅觉特征空间;最后,对特 征空间优化后,采用回归算法建立预测模型。

1 土壤样品采集及特征空间构建

1.1 研究区域及土壤样品

本研究126份土壤样品采集于吉林省各个地区,采样点分布如图1所示。



图 1 研究区域及采样点分布图 Fig. 1 Study area and sampling sites

吉林省地貌差异明显,地势由东南向西北倾斜, 呈现东南高、西北低的特点。吉林省的主要土壤类 型为暗棕壤、黑钙土、白浆土、草甸土和黑土,主要种 植作物玉米、大豆、小麦。由于频繁的耕作导致土壤 退化,施用化肥成了农业生产不可或缺的选择。因 此,研究土壤特性有助于优化施肥、改良土壤结构。 采样时间为2018年秋季,采样前除去杂物和浮土, 采样深度为5~20 cm。每个样品以S形布点采样, 一个样点采集11个点位的土壤,均匀混合后,挑出 落叶、秸秆和石块,然后用"四分法"保留1kg 土壤 样品。根据试验需求,将每个土壤样品分成两份,一 份用于化学测量分析,另一份用于人工嗅觉分析。 化学测量分析所用的各个土壤样品经标记后送往实 验室,24℃无风自然条件下风干。人工嗅觉分析所 用的土壤样品通过喷施蒸馏水和风干的方法使各样 品的相对湿度为65%,然后分别称取80g置于 250 mL的密闭集气瓶内,之后将集气瓶存放在通风 良好的黑暗房间里24 h。

1.2 土壤样品化学分析

用于化学分析的各土壤样品,经自然干燥、研磨、过 0. 25 mm 筛网处理后,采用重铬酸钾容量法测量。测量结果采用 SPSS 13 统计描述,结果如图 2 所示。图 2 中,土壤样品的有机质质量比范围为 10. 62~48.79 g/kg,均值为 23. 44 g/kg。土壤有机质含量的变异系数(Variable coefficient, CV)为 32. 17%,说明样本有机质含量分布呈现出较大的空间变异性。K-S检验值是 0. 224(P>0.05),表明 样本数据服从正态分布。



Fig. 2 Statistical description results of soil organic matter

1.3 人工嗅觉系统结构及工作原理

人工嗅觉系统主要由传感器阵列、信号处理模 块、数据采集卡和计算机等构成,如图 3 所示。其 中,传感器阵列采用单类传感器阵列,即由多个型号 相同传感器构成。检测阵列由 10 个 SGAS707 型传 感器构成,放置在密闭测试盒内。SGAS707 型传感 器是美国集成设备技术公司(Integrated Device Technologies,Inc.,IDT)生产的一款用于检测 VOCs 的专用 MOS 气体传感器,其内部集成了一个加热电 阻,可进行温度调制,能为传感器提供不同的工作温 度。传感器阵列通过 FFC 软线与信号处理模块相 连接。数据采集卡通过杜邦线连接信号处理模块相 连接。数据采集卡通过杜邦线连接信号处理模块相

图中 V_{heat}为加热电阻的供电电压,通过一个三端电压调节器 LM317 对其进行调制,进而为传感器提供一个恒定工作温度。调节电阻 PR1,可以改变



图 3 人工嗅觉系统装置

 Fig. 3
 Artificial olfactory system device

 1.数据采集卡
 2.信号处理模块
 3.吸气泵
 4.密闭测试盒

 5.土壤气体样品
 6.氦气罐
 7.计算机



 V_{heat} 。在测量单一组分挥发性气体(如辛烷、甲醛和 异丁烯等)时,IDT公司推荐 SGAS707 型传感器的 最优工作温度为 150℃。然而,诸多研究^[27-29]已经 证实,MOS 气体传感器在不同的工作温度下,其对 不同组分气体分子的吸附能力不一样。因此,通过 温度调制可以提高传感器阵列对混合气体的检测灵 敏度。为了提高单类传感器阵列对复杂土壤气体的 选择性以及灵敏性,本文采用不同的 V_{heat} 对 SGAS707 型传感器的工作温度进行差异性调制。设 计中,10 个传感器的 V_{heat} 分别设置为 1.25、1.50、 1.75、2.00、2.25、2.50、2.75、3.00、3.25、3.50 V。

系统工作时,先启动测试按钮,然后采用氦气清 洗密闭测试盒,待传感器阵列输出值稳定后,停止洗 气并关闭测试按钮;接着,再使用一个 20 mL 的注射 器抽取集气瓶顶部的土壤气体,并通过密闭测试盒 上面的注射孔快速转移至密闭测试盒内,同时开启 测试按钮。当采样时间到达 100 s 时,停止采样,并 重新用氦气清洗密闭气室,以便进行下次测量。一 般情况下,较高的采样频率更能反映传感器的响应, 但会增加后期数据处理的难度,而较低的采样频率 会造成关键数据的丢失。测试过程中,采样频率设 置为 10 Hz。

人工嗅觉系统测量土壤 VOCs 的典型响应曲线 如图 5 所示(以土壤有机质质量比为 27.95 g/kg 的 土壤样品为例)。图中 S1~S10 分别为 10 个传感器 编号。从图 5 可以看出,10 Hz 的采样频率能够有效 地获得传感器的响应变化曲线,确保了适宜的数据 量。此外,不同温度控制下的传感器对土壤 VOCs 的响应不同,表现出特定的选择性和灵敏性。然而, 在传感数据测量中,由于外部环境变化、测量系统本 身误差,响应曲线将不可避免产生一些毛刺。



为了消除毛刺的不利影响,采用一维中值滤波 算法对系统输出响应曲线进行平滑处理,并设置平 滑点数为30。图6为滤波后的响应曲线。为减少 后期数据处理难度、加快预测模型测量效率,本研究 选取测量开始后的前60s数据作为分析区域。



1.4 特征提取

从传感器响应曲线上提取恰当的特征构建嗅觉 特征空间是建立预测模型的前提条件。常用的特征 提取方式包括^[30]:最大值(Maximum value) V_{max} 、最 小值(Minimum value) V_{min} 、平均值(Mean value) V_{mean} 、平均微分系数(Mean differential coefficient value) V_{mde} 、响应面积(Response area value) V_{ra} 、时刻 t的瞬态值 V_t 和稳态值 V_s 等。本研究在提取的数据 分析区域采用 V_{max} 、 V_{min} 、 V_{mean} 、 V_{ra} 、第 30 秒的瞬 态值 V_{30} 和第 60 秒的瞬态值 V_{60} 这 7 个特征来构建 特征空间。其中, V_{mde} 和 V_{ra} 的计算式分别为

$$V_{\rm mdc} = \frac{1}{N-1} \sum_{i}^{N-1} \frac{D_{i+1} - D_{i}}{\Delta t}$$
(1)

$$V_{\rm ra} = \sum_{i=1}^{N} D_i \Delta t \tag{2}$$

式中 D_i——第 i 个采样数据, V

 Δt ——采样间隔时间,s

经过特征提取后,每个传感器响应曲线上的7 个特征将被提取,而传感器阵列由10个传感器构成。因此,一个样品将被提取70个特征,所有土壤 样品将形成一个126×70的嗅觉特征空间。为了消 除量纲和数量级对特征的影响,采用 z-score 方法对 选取的特征进行标准化处理^[31]。

1.5 训练集与测试集划分

为建立一个适当的预测模型,需要将嗅觉特征 空间划分为训练集和测试集两部分。训练集可用于 训练模型,测试集可以对模型预测性能进行测试。 合理的划分训练集和预测集有利于模型性能提升, 采用 Kennard – Stone 算法给出的较为合理的划分比 例^[32],即训练集和测试集之比为 7:3。

2 模型构建方法

2.1 PLSR 模型

主成分因子(Principal component factor, PCF)数 量是造成 PLSR 模型过拟合或欠拟合的主要因素。 本研究采用留一交叉验证法结合赤池信息量 (Akaike information criterion, AIC)准则来判断最优 PCF 数量。其中,赤池信息量(AIC 值)计算公式为

$$AIC = M \lg P_{rss} + 2p \tag{3}$$

式中 M——训练集样本数目

P_{rss}——训练集样本的预测残差平方和

p----PCF 数量

2.2 SVR 模型

SVR 是一种基于支持向量机(Support vector machine, SVM)的回归技术^[33]。LIBSVM 工具箱提供了两类回归方法: ε – SVR 和 ν – SVR。本研究采用 ε – SVR 建立回归模型,应用径向基函数(Radial basis function, RBF)作为核函数。惩罚因子c(c>0)和内核参数 g 是影响 SVR 建模的两个主要参数。本文为了优化 SVR 模型,采用网格搜索法和 5-折交叉验证法,并结合交叉验证均方误差(Mean square error of cross-validation, MSECV)来确定参数组合(c,g)的值。MSECV 越小,参数 c 和 g 的组合越佳。

2.3 BPNN 模型

BPNN 是一种典型的多层前向型神经网络。本 文采用 3 层网络构建 BPNN 模型,隐含层最优神经 个数计算公式为

$$h = \sqrt{n + m} + \alpha$$
 (4)
式中 h ——隐含层神经元节点数
 n ——输入节点数

m——输出节点数

α----1~10的正整数

其中,n等于用于建模的特征向量数量,m为预测因变量数量,本文只对有机质含量作预测,所以m为1。

为了确定 BPNN 模型隐含层神经元节点数 h, 首先根据式(4)确定 h 的范围,然后基于 h 的每一 次不同值在训练集上分别训练 10 个不同的 BPNN 模型,并计算这 10 个 BPNN 模型对训练集样本的预 测均方根误差(Root mean square error, RMSE)的平 均值,记为 MRMSE,最后根据 MRMSE 来确定 h 值。 MRMSE 越小,对应 h 值用于建模的效果越好。

在 BPNN 建模及 h 的优选中,隐含层神经元的激活函数选用 S 形传递函数 tansig,输出层神经元的激活函数采用线性传递函数 purelin,并且设置训练的迭代1000次,学习率为0.01,目标误差为0.001。

2.4 模型评价指标

土壤养分预测模型的常用评价指标有决定系数 (Coefficient of determination) R^2 、均方根误差 (RMSE)和预测偏差比(Ratio of prediction derivation, RPD)。 R^2 越接近1表明模型的拟合效果 越好; RMSE用于表征模型预测值和测量值之间误 差, RMSE越小,表明模型的预测精度越高。RPD是 样品标准差与预测均方误差之间的比值,用于对模 型性能进行进一步评价。一般 RPD 越大,模型性能 越好。RPD 在土壤检测方面,可分为3个等级^[34]: A级(RPD 大于等于 2.0)表示模型性能非常好,可 以进行准确的定量检测;B级(RPD大于1.4小于2.0)表示模型一般,可进行较为粗略检测;C级(RPD小于等于1.4)表示模型性能很差,不可用于定量检测。

3 结果与分析

3.1 初步建模评估结果

根据 Kennard - Stone 分配方法,可以将 126 个 土壤样品的嗅觉特征空间分成两部分,即前88个样 本数据作为训练集,剩余的38个样本数据用作测试 集。分别构建 PLSR、SVR 和 BPNN 预测模型,并采 用训练集对各模型的建模参数进行优化。PLSR 模 型的参数 PCF 可通过 AIC 随 PCF 数量变化曲线来 确定,如图 7a 所示。最优的 PCF 数量可以根据较 小的 AIC 值来确定。但是当 PCF 数量选择过大时, 会使模型的复杂度增加。因此,从图 7a 可以看出, 将 PFC 数量设为 3 用于构建 PLSR 模型较为适宜。 SVR 模型的参数组合(c,g)经网格搜索法和 5-折交 叉验证法优化后,可设为 c = 2,g = 0.015 6, 如图 7b 所示。建立 BPNN 模型时,较小的 MRMSE 更有利 于优选出合适的隐含层神经元节点数。因此,根据 图 7c, 可将 BPNN 建模的隐含层神经元节点数设定 为10。在确定各预测模型的最佳建模参数后,分别 构建 PLSR、SVR 和 BPNN 模型,并用训练集对各模 型进行训练,同时采用测试集对模型进行预测。预 测结果如图8所示。





图 8 显示 3 种模型的 R²分别为 0.69、0.86 和 0.65; RMSE 分别为 5.71、2.81、3.63 g/kg; RPD 分 别为 1.07、2.16 和 1.68。这表明,土壤嗅觉特征空 间与土壤有机质含量之间存在一定的相关性。但 是,土壤嗅觉特征空间并没有得到充分的优化,因此 需要进一步的分析来确定是否存在其他干扰。嗅觉 特征空间的优化主要包括异常样本剔除和特征降维 这两种处理。它可以消除异常样本和冗余特征对模 型的干扰,达到准确建模、预测的目的。

3.2 异常样本剔除结果

异常样本产生的主要原因可能是操作不当、人 工嗅觉系统本身的误差或温度、湿度等外部因素。 异常样本对模型的预测精度具有重大影响。因此, 有必要对异常样品进行识别和去除。MCS 方法是 基于预测误差(或者预测残差)对异常样本的敏感 性而提出的,被证明是一种有效的异常样本剔除方



法[35]。

本文采用 MCS 对嗅觉特征空间的异常样本进 行检测,具体步骤如下:首先,在嗅觉特征空间中随 机选择 70% 的样本作为训练集,剩余 30% 的样本用 作验证集;其次,基于训练集数据计算 PLSR 的最佳 主成分因子数,并构建 PLSR 预测模型;之后,应用 构建的 PLSR 预测模型对验证集进行预测,并计算 验证集样本的预测残差;然后,重复上述过程进行多 次循环采样,可获得所有样品的预测残差分布;最 后,计算各样本预测残差的平均值(Average value, AVG)和标准偏差(Standard deviation,STD),并根据 AVG – STD 分布图检测异常样本。

图 9 为采用 MCS 方法得到的 AVG - STD 分布 图。在运行 MCS 方法时,将循环采样次数设定为 5 000 次。图中结果表明,1、2、22、64、32 号样本属 于离群点,可视为异常样本。这是因为这几个样本 不稳定,也不适用于基于其余样本构建的模型。

去除这 5 个异常样本后,首先按照前文所述方 法重新划定训练集和测试集,此时训练集包含 85 个 样本数据,测试集包含 36 个样本数据。之后,再基 于新的训练集重新优化各模型的建模参数,得到 PLSR 的 PFC 为 4,SVR 的参数为 c = 4.12 和 g =0.011 8,BPNN 的隐含层神经元节点数为 12。然 后,再次建立新的 PLSR、SVR 和 BPNN 预测模型,测 试集的预测结果如图 10 所示。图 10 显示 3 种模型



图 9 预测残差平均值与标准差的关系 Fig. 9 Prediction residual average values versus STD

的 R² 分 别 为 0.75、0.89 和 0.84; RMSE 分 别 为 5.68、2.74、3.15 g/kg; RPD 分 别 为 1.09、2.29 和 1.97, 优于去除异常样本前预测结果。

3.3 PCA 降维结果

特征向量是影响模型性能的另一个重要因素, 这是因为原始特征空间中包含了大量的与建模无关 的冗余信息。采用未降维的特征空间直接建立模型 将导致较大的计算量,并且会干扰模型的预测精度。 主成分分析(PCA)是一种较通用的特征降维方 法^[36],其通过计算原始特征空间的协方差矩阵的特 征向量,将高维空间向量线性变换为分量不相关的 低维空间向量。为了优化土壤嗅觉特征空间,本研 究采用 PCA 方法作为特征降维手段,步骤如下: ①计算嗅觉特征空间的协方差矩阵。②求出协方差 矩阵的特征值及其对应的特征向量,并根据特征值



图 10 去除异常样本后各模型的预测结果



177

的大小对特征向量进行排序,得到特征向量矩阵。 ③选择特征向量矩阵的前 $k(1 \le k < 70)$ 个向量,则 可将原始的嗅觉特征空间降为k维,其中k即为主 成分数,可通过方差信息累计贡献率G(k)来确定, 计算公式为

$$G(k) = \sum_{i=1}^{k} \lambda_{i} / \sum_{j=1}^{70} \lambda_{j} \times 100\%$$

(k = 1, 2, ..., 69) (5)

$$\alpha_i = \lambda_i / \sum_{j=1}^{\infty} \lambda_j \times 100\% \quad (i = 1, 2, \cdots, k) \quad (6)$$

式中 λ_i ——协方差矩阵第 i(i < k)个排序后特征值

70

- λ_i ——协方差矩阵第 $j(j \leq 70)$ 个排序后特征值
- α_i——第*i*个主成分的方差信息贡献率

当 G(k)大于一个设定的值时,可得到一个降至 k 维 矩阵。

为了获取 PCA 对特征空间的优化结果,对去除 异常样本的土壤嗅觉特征空间(121(样本)×70(维 特征))采用 PCA 方法降维,并设置 *G*(*k*)为 95%, 得到如图 11 所示结果。从图 11 可以看出,当主成 分数 *k* 为 2 时,*G*(*k*)大于 95%,表明采用 2 个主成 分基本能够反映原来特征空间的大部分信息。采用 经 PCA 降维后的特征空间数据分别对新建立的 PLSR、SVR 和 BPNN 模型进行训练与测试,得到各 模型的预测结果如图 12 所示。图中,PLSR 的最优 建模参数 PFC 为2,SVR 的最优建模参数分别为c =104.2 和g = 0.0002,BPNN 的最优隐含层神经元节 点数为14。图 12显示,3种模型的 R^2 分别为0.86、 0.91 和 0.85;RMSE 分别为2.49、2.05、2.68 g/kg; RPD 分别为2.49、3.02 和 2.32。







Fig. 12 Prediction results of each model after PCA dimensionality reduction

3.4 比较分析

图 8、10、12 分别显示了嗅觉特征空间优化前后,不同测定算法所建模型的预测效果。为了更直观地对比分析,将各模型的性能评价指标列于表 1。从表 1 可得出,与未优化前模型性能相比,经 MCS方法剔除异常样本后,PLSR、SVR 和 BPNN 模型的 R²分别提升了 8.7%、3.5%和 29.2%,RMSE 分别 降低了 0.5%、2.5%和 13.2%,RPD 分别提升了 1.9%、6.0%和 17.3%。由此可见,所有模型的性 能指标均得到了提升,这说明 MCS 方法能有效地检 测出异常样本。

在剔除异常样本的基础上,进一步应用 PCA 方 法对特征空间降维,结果表明 PLSR、SVR 和 BPNN 这 3 种模型预测性能得到了更进一步的提升。其中 R² 再次提升了 14.7%、2.2% 和 1.2%; RMSE 再 次降低了 56.2%、25.2% 和 14.9%; RPD 再次提高

表1 模型性能对比结果

Tab.1 Comparison	results	of n	nodel	performance
------------------	---------	------	-------	-------------

模型	特征空间的优	R^2	RMSE/	DDD	等级
	化处理		$(\mathbf{g} \cdot \mathbf{kg}^{-1})$	KPD	
PLSR	未优化	0.69	5.71	1.07	С
	剔除异常样本	0.75	5.68	1.09	С
	特征降维	0.86	2.49	2.49	А
SVR	未优化	0.86	2.81	2.16	А
	剔除异常样本	0.89	2.74	2.29	А
	特征降维	0.91	2.05	3.02	А
BPNN	未优化	0.65	3.63	1.68	В
	剔除异常样本	0.84	3.15	1.97	В
	特征降维	0.85	2.68	2.32	Α

128.4%、31.9% 和 17.8% $_{\circ}$

如果在未优化的特征空间上直接建模、训练和 预测,根据土壤性质 RPD 的分类方法,只有 SVR 模 型具有较好的预测性能,属于 A 级(RPD 为 2.16); BPNN 模型属于 B 级(RPD 为 1.68), 预测性能一 般,但其拟合效果($R^2 = 0.65$)较差;而 PLSR 的预测 等级为C级(RPD小于1.4),表现出较差的预测性 能。经异常样本剔除、特征降维后, PLSR、SVR 和 BPNN 等3种模型预测等级均达到了A级(RPD大 于2.0),这表明异常样本和冗余特征信息对模型预 测性能有很大的影响。在优化后的特征空间上建 模,3种模型的预测指标 R²均不小于 0.85,表现出 较强预测性能。然而,SVR 的预测能力($R^2 = 0.91$ 、 RMSE 为 2.05 g/kg、RPD 为 3.02) 明显高于 PLSR 和 BPNN, 而 PLSR 的预测性能略优于 BPNN。这可 能因为 SVR 是基于结构风险最小化的模型,其对数 据规模和数据分布的要求比较低,具有优异的泛化 能力^[37]。此外,结合网格搜索法和5-折交叉验证法 来优化洗取 SVR 参数在一定程度上也提升了 SVR 的学习能力。

4 结论

(1)提出了一种基于人工嗅觉技术的土壤有机 质含量检测方法,结合蒙特卡罗抽样(MCS)和 PCA 特征降维两种手段,实现了土壤嗅觉特征空间的优 化。采用 PLSR、SVR 和 BPNN 等 3 种回归算法,构 建了土壤嗅觉特征空间与土壤有机质含量之间的关 系模型。模型预测结果表明,这 3 种预测模型对土 壤有机质含量均有较高的预测能力,其中 SVR 模型 的预测性能最好,可对土壤有机质含量进行准确预 测。

(2)采用人工嗅觉技术测量土壤有机质含量是 可行的。考虑到研究区域的土壤样本具有较大的空 间变异性,因此可以认为该测量方法是稳健的。研 究结果可作为土壤有机质含量检测的一种参考方 法。

参考文献

- SALEHI M H, BENI O H, HARCHEGANI H B, et al. Refining soil organic matter determination by loss-on-ignition [J].
 Pedosphere, 2011, 21(4): 473 482.
- [2] DING G, NOVAK J M, AMARASIRIWARDENA D, et al. Soil organic matter characteristics as affected by tillage management
 [J]. Soil Science Society of America Journal, 2002, 66(2): 421 429.
- [3] 吴才武,夏建新,段峥嵘. 土壤有机质测定方法述评与展望[J]. 土壤,2015,47(3):453-460.
 WU Caiwu, XIA Jianxin, DUAN Zhengrong. Review on detection methods of soil organic matter[J]. Soils, 2015, 47(3):453-460. (in Chinese)
- [4] 尼加提·卡斯木,茹克亚·萨吾提,师庆东,等. 基于优化光谱指数的土壤有机质含量估算[J/OL]. 农业机械学报,2018, 49(11):155-163.

NIJAT Kasim, RUKEYA Sawut, SHI Qingdong, et al. Estimation of soil organic matter content based on optimized spectral index[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(11): 155 - 163. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20181118&flag = 1. DOI:10.6041/j.issn.1000-1298.2018.11.018. (in Chinese)

- [5] WANG J, HE T, LV C, et al. Mapping soil organic matter based on land degradation spectral response units using Hyperion images[J]. International Journal of Applied Earth Observations and Geoinformation, 2010, 12(Supp.2): 171-180.
- [6] NAWAR S, MOUAZEN A M. On-line Vis NIR spectroscopy prediction of soil organic carbon using machine learning [J]. Soil and Tillage Research, 2019, 190:120 - 127.
- [7] SHEPHERD K D, WALSH M G. Development of reflectance spectral libraries for characterization of soil properties [J]. Science Society of America Journal, 2002, 66(3): 988 - 998.
- [8] CONFORTI M, CASTRIGNANO A, ROBUSTELLI G, et al. Laboratory-based Vis NIR spectroscopy and partial least square regression with spatially correlated errors for predicting spatial variation of soil organic matter content [J]. CATENA, 2015, 124: 60-67.
- [9] 王海峰,张智韬,KARNIELIA,等.基于灰度关联-岭回归的荒漠土壤有机质含量高光谱估算[J].农业工程学报,2018, 34(14):124-131.
 WANG Haifeng, ZHANG Zhitao, KARNIELIA, et al. Hyperspectral estimation of desert soil organic matter content based on

gray correlation-ridge regression model[J]. Transactions of the CSAE, 2018, 34(14): 124 - 131. (in Chinese) [10] 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 and Tillage Research, 2016, 155: 510 522.
 [11] 宋海燕,秦刚,韩小平,等. 基于可见光谱的不同质地土壤有机质快速测定[J/OL]. 农业机械学报,2012,43(7):69 72.
- [11] 木母無,条柄,种小干,寺. 至了可见几盲时不问质地上录有优质厌迷例足[J'OL]. 农业优减学报,2012,43(7):69-72. SONG Haiyan, QIN Gang, HAN Xiaoping, et al. Rapid prediction of soil organic matter by using visible infrared spectral technology[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2012, 43(7): 69-72. http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx? file_no = 20120712&flag = 1. DOI:10.6041/j.issn.1000-1298.2012.07. 012. (in Chinese)
- [12] 袁静, 王鑫, 颜昌翔. 不同含水量黑土土壤光谱反射率半经验模型构建[J]. 光谱学与光谱分析, 2019, 39(11): 3514-3518.
 YUAN Jing, WANG Xin, YAN Changxiang. A semi-empirical model for reflectance spectral of black soil with different moisture contents[J]. Spectroscopy and Spectral Analysis, 2019, 39(11): 3514-3518. (in Chinese)
- [13] 彭杰,向红英,周清,等. 土壤氧化铁的高光谱响应研究[J]. 光谱学与光谱分析, 2013, 33(2):502-506.

- [14] CESARE F D, MATTIA E D, PANTALEI S, et al. Use of electronic nose technology to measure soil microbial activity through biogenic volatile organic compounds and gases release [J]. Soil Biology and Biochemistry, 2011, 43(10): 2094 - 2107.
- [15] GOUGOULIAS C, CLARK J M, SHAW L J. The role of soil microbes in the global carbon cycle: tracking the below-ground microbial processing of plant-derived carbon for manipulating carbon dynamics in agricultural systems [J]. Journal of the Science of Food and Agriculture, 2014, 94(12):2362 - 2371.
- [16] CARON A, REDON N, CODDEVILLE P, et al. Identification of indoor air quality events using a K-means clustering analysis of gas sensors data[J]. Sensors and Actuators B (Chemical), 2019, 297: 126709.
- [17] SECHMAN H, DZIENIEWICZ M, LISZKA B. Soil gas composition above gas deposits and perspective structures of the Carpathian Foredeep, SE Poland[J]. Applied Geochemistry, 2012, 27(1):197 - 210.
- [18] BALASUBRAMANIAN S, PANIRAHI S, KOTTAPALLI B, et al. Evaluation of an artificial olfactory system for grain quality discrimination[J]. LWT-Food Science and Technology, 2007, 40(10):1815-1825.
- [19] SRIVASTAVA A K. Detection of volatile organic compounds (VOCs) using SnO₂ gas-sensor array and artificial neural network
 [J]. Sensors and Actuators B (Chemical), 2003, 96(1-2): 24-37.
- [20] ZHANG X, CHENG J, WU L, et al. An overview of an artificial nose system [J]. Talanta, 2018, 184: 93 102.
- [21] SANAEIFAR A, ZAKIDIZAJI H, JAFARI A, et al. Early detection of contamination and defect in foodstuffs by electronic nose: a review[J]. TrAC Trends in Analytical Chemistry, 2017, 97: 257 - 271.
- [22] 郭森仁,梁丽丽,林雪娟,等. 基于电子鼻的慢性胃炎常见病性证素间的气味图谱特征研究[J]. 中华中医药杂志, 2016, 31(6): 2263 2266.
 GUO Senren, LIANG Lili, LIN Xuejuan, et al. Study on odor map characteristics of common syndrome elements of the chronic gastritis based on electronic nose[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2016, 31(6): 2263 -
- gastritis based on electronic nose[J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2016, 31(6): 2263 2266. (in Chinese)
- [23] LU X, WANG J, LU G, et al. Quality level identification of West Lake Longjing green tea using electronic nose[J]. Sensors and Actuators B (Chemical), 2019, 301: 127056.
- [24] 毕丽君,高宏岩.电子鼻(EN)及其在多领域中的应用[J]. 医学信息,2006,19(7):1283-1286.
 BI Lijun, GAO Hongyan. The electronic nose (EN) and its application in many field [J]. Medical Information, 2006, 19(7):1283-1286. (in Chinese)
- [25] LAVANYA S, DEEPIKA B, NARAYANAN S, et al. Indicative extent of humic and fulvic acids in soils determined by electronic nose[J]. Computers and Electronics in Agriculture, 2017, 139:198-203.
- [26] ANDRZEJ B, KATARZYAN J G, ŁUKASZ G, et al. Evaluating soil moisture status using an e-nose[J]. Sensors, 2016, 16(6): 886.
- [27] 宋凯. 金属氧化物半导体气体传感器气体检测关键问题研究[D]. 哈尔滨:哈尔滨工业大学,2011. SONG Kai. Research on key issues of metal oxide semiconductor gas sensor in gas detection[J]. Harbin: Harbin Institute of Technology, 2011. (in Chinese)
- [28] MARTINELLI E, POLESE D, CATINI A, et al. Self-adapted temperature modulation in metal-oxide semiconductor gas sensors [J]. Sensors and Actuators B (Chemical), 2012, 161(1): 534 - 541.
- [29] DENG Q, GAO S, LING Y, et al. Temperature & light modulation to enhance the selectivity of Pt-modified zinc oxide gas sensor[J]. Sensors and Actuators B (Chemical), 2017, 247: 903-915.
- [30] 殷勇,赵玉珍,于慧春.基于多种变量分析方法鉴别食醋种类电子鼻信号特征筛选[J].农业工程学报,2018, 34(15):290-297.

YIN Yong, ZHAO Yuzhen, YU Huichun. Feature selection of electronic nose signal for vinegar discrimination based on multivariable analysis [J]. Transactions of the CSAE, 2018, 34(15): 290 - 297. (in Chinese)

- [31] 杨鹏圆,李海芳,陈东伟,等. 脑电信号特征的归一化方式与选择方法研究[J]. 计算机应用与软件, 2015, 32(2):48-52.
 YANG Pengyuan, LI Haifang, CHEN Dongwei, et al. Research on normalization manner and selection method of EEG features
 [J]. Computer Applications and Software, 2015, 32(2):48-52. (in Chinese)
- [32] ZHU L, JIA H, CHEN Y, et al. A novel method for soil organic matter determination by using an artificial olfactory system [J]. Sensors, 2019, 19(15): 3417.
- [33] VAPNIK V N. Statistical Learning Theory[M]. New York: Wiley, 1998.
- [34] QI H, PAZKAGAN T, KARNIELI A, et al. Evaluating calibration methods for predicting soil available nutrients using hyperspectral VNIR data[J]. Soil and Tillage Research, 2018, 175: 267 - 275.
- [35] LI X, WEI Y, XU J, et al. Quantitative visualization of lignocellulose components in transverse sections of moso bamboo based on FTIR macro- and micro-spectroscopy coupled with chemometrics [J]. Biotechnology for Biofuels, 2018, 11(1): 1-16.
- [36] 李鸿强,魏小清,王有玺,等. 主成分分析和线性判别分析应用于心电信号特征提取和诊断算法研究[J]. 生物医学工程研究, 2019, 38(2): 145-150.
 LI Hongqiang, WEI Xiaoqing, WANG Youxi, et al. Research of feature extraction and diagnosis algorithm using principal

component analysis and linear discriminant analysis of electrocardiogram signal [J]. Journal of Biomedical Engineering Research, 2019, 38(2): 145 - 150. (in Chinese)

[37] 魏宏业,王建华,何歲. 销售量预测的支持向量机建模及参数选择研究[J]. 系统仿真学报,2005,17(1):33-36.
 WEI Hongye, WANG Jianhua, HE Wei. Study on the support vector machines model for sales volume prediction and parameters selection[J]. Acta Simulata Systematica Sinica, 2005, 17(1):33-36. (in Chinese)