doi:10.6041/j.issn.1000-1298.2021.10.011

基于多传感器人工嗅觉系统的土壤有机质含量检测方法

李名伟[1,2]　朱庆辉[1,2]　夏晓蒙[1,2]　刘　鹤[1,2]　黄东岩[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: 为了实现对土壤有机质含量的快速、方便、准确测量,本文提出了一种基于多传感器人工嗅觉系统的土壤有机质含量检测方法。选取 10 个不同型号的氧化物半导体式气体传感器组成传感器阵列,并采用不同浓度的硫化氢、氨气和甲烷等标准气体对传感器阵列进行了响应测试,从响应曲线可以看出,传感器阵列对不同浓度、种类的标准气体皆有响应且响应结果不同,随着标准气体浓度的增大传感器阵列的响应曲线也随之上升,表明传感器阵列具有较高的特异性和一定的交叉敏感性。提取每个传感器土壤气体响应曲线上的响应面积、最大值、平均微分系数、方差、平均值和最大梯度 6 个特征构建人工嗅觉特征空间。采用偏最小二乘法回归(PLSR)、支持向量机回归(SVR)和 BP 神经网络(BPNN)算法建立人工嗅觉特征空间与土壤有机质含量关系的预测模型,使用决定系数(R^2)、均方根误差(RMSE)和绝对平均误差(MAE)评估预测模型的性能。试验结果表明,PLSR、BPNN、SVR 测试集的 R^2 分别为 0.808 78、0.871 79 和 0.919 57,RMSE 分别为 3.678 4、3.161 4、2.425 4 g/kg,MAE 分别为 3.107 9、2.415 4、2.138 9 g/kg。SVR 算法建立的模型 R^2 最高,RMSE、MAE 最小,比 PLSR、BPNN 具有更好的预测性能,可用于土壤有机质含量的测量。

关键词: 土壤有机质; 人工嗅觉系统; 气体传感器阵列; 预测方法; 回归算法

中图分类号: S158.2; TP212.9　　**文献标识码**: A　　**文章编号**: 1000-1298(2021)10-0109-11　　**OSID**:

Detection Method of Soil Organic Matter Based on Multi-sensor Artificial Olfactory System

LI Mingwei[1,2]　ZHU Qinghui[1,2]　XIA Xiaomeng[1,2]　LIU He[1,2]　HUANG Dongyan[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 achieve a rapid, convenient and accurate measurement of the soil organic matter (SOM) content, a multi-sensor array based on artificial olfactory technology was designed to detect SOM content. Totally ten different types of oxide semiconductor gas sensors were selected to form the sensor array, and different concentrations of hydrogen sulfide, ammonia and methane standard gases were used to test the response of the sensor array. It can be seen from the response curve that the sensor array responded to different concentrations and types of standard gases, and the response results were not the same. With the increase of the standard gas concentration, the response curve of the sensor array was also increased. The test results showed that the sensor array had high specificity and certain cross-sensitivity. And from the soil gas response curve of each sensor, six features (response area value, maximum value, average differential coefficient, variance, average value and maximum gradient) were extracted to construct an artificial olfactory feature space (AOFS). Then, partial least square regression (PLSR), back propagation neural network (BPNN) and support vector machine regression (SVR) algorithms were used to establish the prediction model of AOFS and SOM content, and the coefficient of determination (R^2), root mean square error (RMSE) and absolute average error (MAE) were used to evaluate the performance of the prediction model. The test results showed that the R^2 of the PLSR, BPNN, and SVR test sets were 0.808 78, 0.871 79 and 0.919 57, the RMSE were 3.678 4 g/kg, 3.161 4 g/kg and

收稿日期: 2021-07-31　修回日期: 2021-08-22

基金项目: 吉林省科技发展计划项目(20200502007NC、20190302116GX)

作者简介: 李名伟(1992—),男,博士生,主要从事农业机械自动化研究,E-mail: lmw271314@163.com

通信作者: 黄东岩(1976—),男,教授,博士生导师,主要从事保护性耕作技术及其智能农机装备研究,E-mail: cchdy760829@sina.com

2.425 4 g/kg, and the MAE were 3.107 9 g/kg, 2.415 4 g/kg and 2.138 9 g/kg, respectively. The model established by the SVR algorithm had the highest R^2, the smallest RMSE and MAE. It had higher predictive performance than PLSR and BPNN, and can be used for the measurement of SOM content.

Key words: soil organic matter; artificial olfactory system; gas sensor array; prediction method; regression algorithm

0 引言

土壤有机质(Soil organic matter,SOM)是指存在土壤中的所有含碳的有机物质,主要包括动植物残体、土壤微生物及其分解和合成的各种有机物,是土壤固相部分的重要组成成分[1-2]。土壤有机质的含量仅占土壤组成的5%左右,但对改善土壤的物理性质,促进土壤中营养元素的分解,提高土壤肥力都有极为重要的作用[3]。掌控土壤有机质含量的动态变化,根据土壤性状调节作物的施肥量,高效地利用各类农业资源,是实现精准农业和促进农业可持续发展的重要内容之一。

目前,测量土壤有机质含量的方法主要有灼烧法、湿烧法、重铬酸钾容量法等[4-5]。其基本原理都是根据有机碳的含量测定有机质,是一种碳成分直接测定法。其中,重铬酸钾容量法因技术原理简单、测量结果准确且成本较低而被广泛使用[6]。但是该方法需要在化学实验室中进行,且存在耗时耗力、操作复杂和造成污染等问题。近年来,随着近红外光谱技术的发展,光谱测定土壤有机质的方法得到了广泛的关注和研究[7]。近红外光谱有快速、无损和高效等特点,可以实时测量大量土壤样本的参数,适用于精准农业[8-10]。然而,土壤有机质的光谱受土壤水分、氧化铁、质地等因素的影响[11-13]。

土壤气体是土壤的组成部分之一,是由土壤中生物活性和气体转移之间的平衡引起的[14]。土壤中气体和挥发性有机化合物的释放与微生物活动有关,并会不同程度地影响土壤中生物的生活[15-16]。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和气体[17-19]。土壤中的挥发性有机化合物和气体与土壤有机质存在相关性,利用这种相关性可以通过人工嗅觉技术检测土壤气味从而测算出土壤有机质含量。

人工嗅觉技术是一种综合性检测技术,融合了传感器技术、信号处理、计算机科学、模式识别和深度学习等,其工作原理是模拟人体嗅觉功能对被测气体进行感知、分析和识别。由多种气体传感器融合而成的传感器阵列和模式识别方法组成的人工嗅觉系统,被认为是检测复杂气体的有效手段。人工嗅觉系统不是直接得出监测气体的具体成分或特性,而是生成样本气体的"指纹图谱",借助合理的模式识别算法,实现对样本气体的准确分类。目前,人工嗅觉技术被广泛应用在食品安全、医疗分析和环境检测等领域[20-23],且在土壤特性和土壤有机质检测方面也有些研究。例如,ANDRZEJ 等[24]采用电子鼻对不同湿度的土壤进行了分类,LAVANYA 等[25]利用电子鼻检测了土壤中透明质酸和游离脂肪酸的含量;文献[26-27]利用单种传感器阵列的人工嗅觉系统实现了对土壤有机质含量的检测,由于采用单种传感器阵列,不能全面反映土壤中气体的响应,土壤有机质含量的检测精度和响应时间有待提高。

针对上述问题,本文提出一种基于人工嗅觉技术的多传感器阵列检测土壤有机质含量的检测方法。土壤有机质在厌氧环境中产生气态烃、硫化氢、氨和醛类等挥发性气体,土壤中的挥发性气体与土壤有机质存在相关性。将土壤样本放在集气瓶中密封48 h,采用10种不同类型的氧化物半导体式气体传感器构建传感器检测阵列,搭建基于人工嗅觉的土壤有机质检测系统,以此系统获取土壤样本中气体的响应曲线。提取响应曲线的响应面积、最大值、平均微分系数、方差、平均值和最大梯度等6个特征构建人工嗅觉特征空间,采用BP神经网络算法(Back propagation neural network,BPNN)、支持向量机回归算法(Support vector machine regression,SVR)和偏最小二乘回归算法(Partial least squares regression,PLSR)依据人工嗅觉特征空间建立土壤有机质含量的预测模型,从而利用人工嗅觉技术实现对土壤有机质含量的检测。

1 材料与方法

1.1 研究区域与土壤取样

土壤采样区域为吉林省,位于中国东北地区中部,如图1所示。

吉林省位于中纬度欧亚大陆的东侧,属于温带大陆性季风气候,年平均气温在-5~8.6℃,年平均降雨量为400~600 mm。土壤类型以暗棕壤、黑钙土、白浆土、草甸土、黑土、风沙土、新积土和水稻土为主,以水稻、玉米、大豆为主要粮食作物。频繁的耕作导致土壤退化,施用化肥可有效提高作物产量,

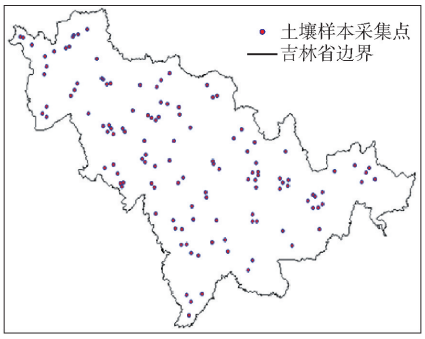


图 1　研究区域和采样地点

Fig. 1　Study area and sampling site

成为农业生产不可或缺的选择。

2019 年秋季在图 1 所示 130 个采样点共采集土壤样本 130 份。由于土壤的不均一性，使各个体都存在着一定程度的变异，为了尽可能还原土壤样本在田间的实际状态，采用对角线布点法进行土样采集，每个采样点采集 9 份土样，采样深度为 0 ~ 20 cm。将得到的 9 份土样混合，按四分法每个采样点保留 1 份土壤样品(1 kg)。根据试验的需求，将 130 份土壤样品在温度 23℃下自然风干后，将土壤碾碎并过 0.25 mm 的筛网处理。将样品分为 2 份，分别用于化学氧化法和人工嗅觉系统检测其土壤有机质含量。化学氧化法采用重铬酸钾法[28]，是标准的土壤有机质测定方法(GB 9834—88)[29]，此方法测得土壤有机质含量为实际值。人工嗅觉系统所用土壤样品通过喷施蒸馏水使各样品的相对湿度维持在 55%，每份土壤样本称取 80 g 置于 250 mL 的密闭集气瓶内，最后将集气瓶在黑暗房间里存放 48 h，部分密封后的土壤样品如图 2 所示。

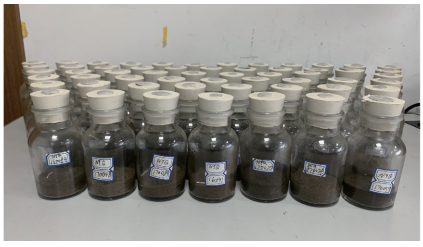


图 2　部分密封后的土壤样品

Fig. 2　Partly sealed soil samples

1.2　人工嗅觉试验方法

试验所采用装置为基于人工嗅觉的土壤有机质检测系统，如图 3 所示。该系统由传感器阵列(安装在封闭的反应室)、信号处理电路、NI 数据采集卡和便携式计算机组成。传感器阵列是整个系统的核心部件，由 10 个不同型号的氧化物半导体式气体传感器组成，如图 4 所示。每个传感器的型号和可测量的气体类型，如表 1 所示。信号处理电路主要负责传感器的供电、对传感器数据的处理，传感器阵列与信号处理电路之间通过 FFC 软线连接。数据采集卡通过杜邦线与信号处理电路连接，用于对传感器数据的采集，并将采集的数据通过 USB 数据线输入计算机来显示、存储。

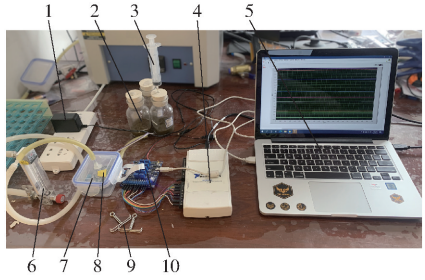


图 3　基于人工嗅觉的土壤有机质检测系统

Fig. 3　Soil organic matter detection system based on artificial olfaction

1. 12 V 电源　2. 土壤样本　3. 顶空采样注射器　4. NI 数据采集卡　5. 便携式计算机　6. 流量计　7. 传感器阵列　8. 注射口　9. 止气夹　10. 信号处理电路

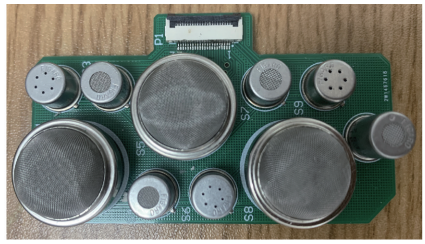


图 4　气体传感器阵列

Fig. 4　Gas sensor array

表 1　气体传感器型号及参数

Tab. 1　Gas sensor model and parameters

传感器编号	型号	监测气体类型	测量范围/(mg·m^{-3})
S1	TGS826	氨气	30 ~ 300
S2	TGS2602	空气污染物(氨气、VOC、硫化氢等)	1 ~ 30
S3	TGS2610	丁烷、LP 气体	500 ~ 10 000
S4	TGS2620	乙醇、有机溶剂	50 ~ 5 000
S5	TGS821	氢气	100 ~ 1 000
S6	TGS2603	空气污染物(三甲胺、甲基硫醇等)	1 ~ 10
S7	TGS2611	甲烷、天然气	500 ~ 10 000
S8	TGS823	甲烷、乙醇	50 ~ 300
S9	TGS2600	空气污染物(氢气、酒精等)	1 ~ 30
S10	TGS2612	甲烷、丙烷、异丁烷	3 000 ~ 9 000

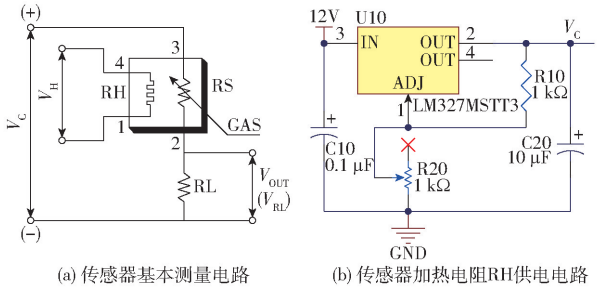
在高耗氧条件下，土壤有机物在微生物作用下分解的最终产物为二氧化碳、水、硝酸盐、硫酸盐和磷酸盐，增加了土壤的耗氧量和二氧化碳含量。在厌氧条件下，土壤有机质的分解促进了气态烃(CH_4、C_2H_4、C_2H_6、C_3H_8等)、硫化氢、氨和醛类物质的形成[30]。本文将土壤样本在集气瓶中密封 48 h，使其在厌氧环境中产生气态烃、硫化氢、氨和醛类等挥发性气体。利用气体传感器的特异性和交叉灵敏

性等特点,选择 10 个对土壤在厌氧环境下产生气体有响应的不同型号气体传感器组成传感器阵列,对土壤样本产生的挥发性气体进行检测。

气体传感器的基本测量电路如图 5a 所示,且每个气体传感器均由电阻器 RS 和 RH 组成。其中,RS 为传感器电阻器,其电阻 R_S 随测量气体的浓度变大而变小;RH 是加热电阻器。传感器正常工作需要加热器电压 V_H 和回路电压 V_C,V_H 作用于 RH,可以为 RS 提供适合的工作温度;V_C 则是用于测定与传感器串联的负载电阻器 RL 上的回路输出电压 V_{RL}。V_H 和 V_C 的电压均设置为(5 ±0.2) V。加热电阻器 RH 为固定值且电阻 R_H 较小,为确保传感器工作温度,每个传感器的 RH 都采用单独供电,其供电电路如图 5b 所示。负载电阻器 RL 与传感器电阻器 RS 串联,可通过回路输出电压 V_{RL} 反映出传感器的响应(传感器电阻 R_S 的变化),R_S 的计算公式为

$$R_S = \left(\frac{V_C}{V_{RL}} - 1\right) R_L \tag{1}$$

式中 R_L——电阻器 RL 的电阻



(a) 传感器基本测量电路 (b) 传感器加热电阻RH供电电路

图 5 传感器工作电路

Fig. 5 Sensor operating circuit

数据采集卡为 NI(National Instruments)公司生产的多功能 I/O 设备 USB-6210,该采集卡集成了 16 路 16 位模拟输入(AI)端口,最大采样速率为 250 kS/s;分别集成了 4 路数字输入(DI)和数字输出(DO)端口,并具有 2 个 32 位计数器。采用 NI 公司开发的 LabVIEW 软件和 DAQ 驱动,开发了一套数据采集系统,其软件界面如图 6 所示。图中,DAQ 助手模块用于设置数据采集卡的参数,本文选用 Ai1 ~ Ai10 共 10 个 AI 采集端口,采集方式为连续测量,采集频率为 10 Hz。

工作时,先打开系统电源,使传感器预热 15 min;然后用空气泵以 2 L/min 的流速抽取洁净空气清洗装有传感器阵列的反应室,清洗时间为 1 min;使用 20 mL 的注射器抽取集气瓶顶部的土壤气体,注射到反应室中;启动 LabVIEW 检测程序,开始采集数据,采集时间为 2 min。完成一次采样后,使用空气泵以 2 L/min 的流速清洗反应室,其余样本数据的采集重复上述过程。

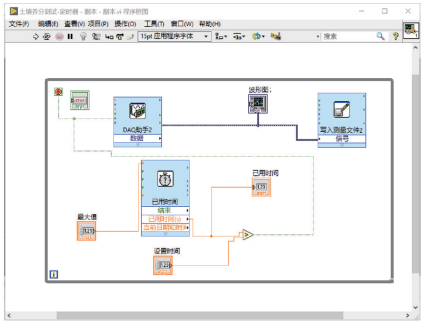


图 6 LabVIEW 程序框图

Fig. 6 LabVIEW program block diagram

1.3 传感器阵列响应测试

基于人工嗅觉的土壤有机质检测系统搭建完成后,还需测试传感器阵列的响应效果。测试所用仪器主要包括土壤有机质检测系统、多种标准气体、YQAr-01L 型标准气体减压阀,图 7 为标准气体罐。标准气体类型及参数如表 2 所示。



图 7 标准气体罐

Fig. 7 Standard gas

表 2 标准气体类型及参数

Tab. 2 Standard gas types and parameters

名称	质量浓度/(mg·m^{-3})	钢瓶填充压力/MPa	钢瓶体积/L
硫化氢	15.2 80.3	9	4
氨气	79.8 808.5	9	4
甲烷	507 18 180	9	4

试验时,先打开土壤有机质检测系统电源,使传感器预热 15 min;然后启动 LabVIEW 检测程序,利用 NI 数据采集卡采集传感器阵列在空气中的响应数据 30 s,接着打开标准气体气阀,以 0.8 L/min 的流速将标准气体通入反应室,通气时间为 30 s,然后用止气夹封闭反应室,采集时间为 3 min;最后用空气泵以 2 L/min 的流速抽取洁净空气清洗反应室,清洗时间为 1 min,完成一次标准气体响应测试。每种标准气体测试 3 次,取平均值。

1.4 特征选择

为消除传感器在采集时产生的噪声对数据处理

的不利影响，采用一维中值滤波算法对响应曲线进行平滑处理，并设置平滑点数为 25。为提高后期数据处理效率，选取测量开始后的前 40 s 数据用作处理分析。本研究提取响应面积 V_{rav}、最大值 V_{max}、平均微分系数 V_{mdc}、方差 V_{vv}、平均值 V_{mean} 和最大梯度 V_{mgv} 这 6 个特征构建特征空间。其中，V_{rav}、V_{mdc}、V_{vv} 和 V_{mgv} 计算公式为

$$V_{rav} = \sum_{i=1}^{N} X_i \Delta t \tag{2}$$

$$V_{mdc} = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{X_{i+1} - X_i}{\Delta t} \tag{3}$$

$$V_{vv} = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N-1} \tag{4}$$

$$V_{mgv} = \frac{X_{jmax} - X_0}{j} \tag{5}$$

式中　X_i——传感器采集的第 i 个数据，V

Δt——相邻 2 个采样点的间隔时间，取 0.1 s

N——采样数据的总数量

$\overline{X}$——采集数据的平均值，V

X_{jmax}——采集数据中最大值，V

X_0——采集数据的初始响应值，V

j——采集数据中最大值所对应的时间，s

气体传感器阵列由 10 个传感器构成，每个传感器的响应曲线提取 6 个特征值，因此，共提取 60 个特征参数，形成一个 130 ×60 的人工嗅觉特征空间。由于 6 种特征值的量纲和数量级不同，需对各特征数据进行无量纲化处理，本研究采用 z - score 方法实现数据的标准化，以消除量纲的影响，其公式为

$$z = (x - u)/s \tag{6}$$

式中　z——标准化后的数据

x——原始数据

u——样本平均值

s——样本标准差

1.5　训练集与测试集划分

预测模型的建立需要划分训练集和测试集，训练集用来估计模型中的参数，使模型能够反映现实，进而预测未来或其他未知的信息；测试集用来评估模型的预测性能。训练集和测试集的合理划分有利于提高所建立模型的预测能力，本研究采用 Kennard - Stone 方法设置训练集和测试集比为 3∶1。

2　模式识别算法

为提高本文检测方法的检测精度，采用 PLSR、BPNN 和 SVR 3 种算法建立人工嗅觉特征空间与土壤有机质含量之间的预测模型，以找出最优的关系模型。

2.1　PLSR 模型

偏最小二乘回归是一种统计学方法，与主成分回归有关系，但不是寻找响应变量和自变量之间最大方差的超平面，而是通过投影分别将预测变量和观测变量投影到一个新空间，来寻找一个线性回归模型。此算法集主成分分析、典型相关分析和多元线性回归分析 3 种分析方法的优点于一身，当变量之间存在高度线性相关时，PLSR 可以建立一个较高精度的预测模型。本研究使用留一交叉验证法寻找主成分因子，通过交叉验证的均方根误差验证模型的稳健性[31]。

2.2　BPNN 模型

BP 神经网络（BPNN）是一种按误差反向传播训练的多层前馈网络，在结构上包括输入层、隐含层和输出层 3 部分，每个部分之间依靠权值和阈值连接[32]。基本的 BPNN 算法是通过信号的前向传播和误差的反向传播 2 个过程来反复修正各单元权值，以此无限地逼近目标值。隐含层神经元数对 BP 神经网络的性能影响较大。若隐含层神经元数较少，则网络不能充分描述输出和输入变量之间的关系；相反，若隐含层神经元数较多，则会导致网络的学习时间变长，甚至会出现过拟合的问题。一般地，确定隐含层神经元数的方法是在经验公式的基础上，对比隐含层不同神经元数对模型性能的影响，从而进行选择。本研究采用 3 层网络构建 BPNN 模型，隐含层最优神经元数计算公式为

$$h = \sqrt{n+m} + \alpha \tag{7}$$

式中　h——隐含层神经元数

n——输入节点数

m——输出节点数

α——1 ~ 10 的正整数

n 为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。此外，神经元激活函数、训练迭代次数和目标误差等的设置也会影响模型效果，但通常都是被设置为固定值。

2.3　SVR 模型

SVR 是一种基于支持向量机（Support vector machine，SVM）的回归技术，是为了利用 SVM 解决回归拟合方面的问题在 SVM 分类的基础上引入 ε 不敏感损失函数提出的[33]。该模型使用径向基函数（Radial basis function，RBF）作为核函数将数据转移到更高维度，然后找到最佳函数（$f(x)$ 或超平面），该函数与目标的偏离不超过 ε，其原理如图 8 所示。SVR 的泛化能力主要受惩罚因子 C 和核参

数 g(RBF 核函数中的方差)的影响,并且 C 大于零。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 越适用于该模型。

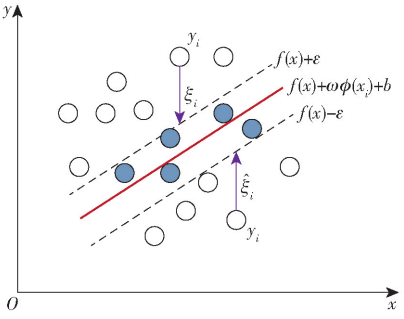


图 8 支持向量回归原理图

Fig. 8 Schematic of support vector regression

2.4 模型评价指标

为了评估上述模块的质量和可靠性,采用决定系数(Coefficient of determination,R^2)、均方根误差(Root mean square error,RMSE)和绝对平均误差(Absolute mean error,MAE)作为性能指标。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,MAE 为其平均绝对误差,RMSE 和 MAE 越小,表明模型的预测精度越高。

3 结果与分析

3.1 传感器阵列响应测试结果

氧化物半导体式气体传感器在长时间未通电时,不管有没有对象气体存在,传感器电阻 R_S 在通电数秒钟后都会突然急剧下降,然后逐渐上升并达到一个稳定的水平。在预热过程中发生的这种反应被称之为“初始动作”。因此,在测试试验开始前,需要对传感器阵列进行预热。

3.1.1 硫化氢响应测试结果

传感器阵列响应测试按照 1.3 节的试验方法进行。图 9 为传感器阵列对不同浓度硫化氢的响应曲线。从图 9 中可以看出,除了传感器 S5,其余传感器对硫化氢均有响应。在低浓度硫化氢中,传感器 S1、S2 和 S6 响应非常强烈,S8 响应也很明显,其余传感器响应不太明显;在高浓度硫化氢中,传感器 S1、S2、S6、S8 响应十分强烈,除了传感器 S5,其余传感器较低浓度时响应明显。

3.1.2 氨气响应测试结果

图 10 为传感器阵列对不同浓度氨气的响应曲线。从图 10 中可以看出,传感器阵列对氨气均有响应,且随着氨气浓度的增大,传感器响应曲线呈上升趋势;传感器 S5 对低浓度氨气响应不明显,但随着氨气浓度的增加,其响应曲线呈明显上升趋势,说明传感器阵列对氨气浓度的变化响应明显。

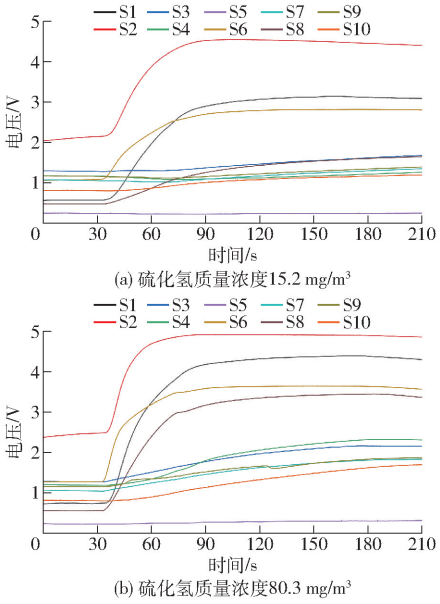


图 9 传感器阵列对不同浓度硫化氢的响应曲线

Fig. 9 Response curves of sensor array to different concentrations of hydrogen sulfide

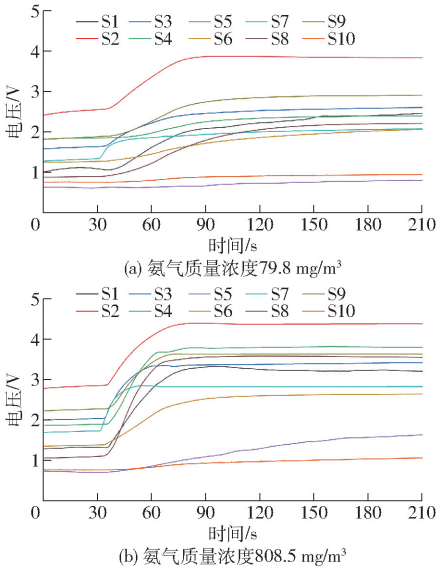


图 10 传感器阵列对不同浓度氨气的响应曲线

Fig. 10 Response curves of sensor array to ammonia at different concentrations

3.1.3 甲烷响应测试结果

图 11 为传感器阵列对不同浓度甲烷的响应曲线。从图 11 中可以看出,传感器 S1、S2、S5 对甲烷几乎没有响应,传感器 S1、S2 在高浓度甲烷中响应曲线有些下降,是受通气时高浓度甲烷气流的影响,导致传感器电阻产生变化;传感器 S6 响应曲线呈下降趋势,且对高低浓度的甲烷响应不明显;其余传感器皆对不同浓度的甲烷气体有明显响应,其中传感器 S7、S8、S10 响应最为明显。

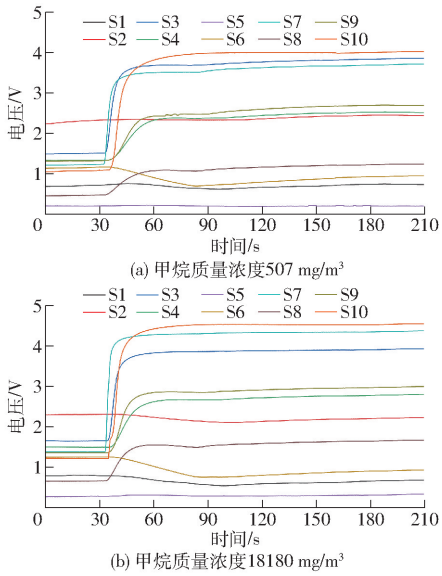


图 11　传感器阵列对不同浓度甲烷的响应曲线

Fig. 11　Response curves of sensor arrays to methane at different concentrations

从传感器阵列对不同浓度、种类标准气体的响应曲线可以看出，传感器 S5 只对高浓度氨气响应，传感器 S1、S2 对甲烷几乎没有响应，其余传感器对不同浓度、种类的标准气体皆有响应且响应结果不同，且随着标准气体浓度的增大传感器阵列的响应曲线也随之上升，测试表明传感器阵列具有较高的特异性和一定的交叉敏感性，为人工嗅觉系统的组建奠定了基础。

3.2　土壤有机质含量化学分析结果及分组

采用重铬酸钾法得出了 130 个土样土壤有机质的含量，结果如表 3 所示，采用 Kennard－Stone 方法设置训练集和测试集之比为 3∶1，即训练集和测试集各有 97、33 个样本。在训练集中，土壤有机质质量比为 13.26～50.55 g/kg，标准差为 7.63 g/kg，变异系数为 24.44%；而测试集中，土壤有机质质量比为 12.19～53.53 g/kg，标准差为 8.48 g/kg，变异系数为 27.41%。研究区域中采集的土壤有机质含量和样本的变异系数变化趋势较大，有利于提高模型的预测能力[34]。测试集中土壤有机质质量比的范围完全覆盖了训练集中土壤有机质质量比的范围，有利于提高模型的泛化能力。

表 3　土壤样本的有机质含量

Tab. 3　Organic matter content of soil samples

参数	训练集	测试集
土壤有机质质量比/(g·kg^{-1})	25.43；30.52；14.19；45.23；40.69；33.46；44.97；31.72；24.23；39.08；28.38；23.16；26.50；21.82；29.99；22.62；21.02；36.54；32.39；23.03；30.39；44.30；27.71；21.42；32.39；39.08；36.27；26.20；18.99；43.77；28.78；40.42；50.55；40.55；29.45；22.89；43.9；31.19；28.51；23.16；31.19；28.78；34.80；35.55；27.57；44.30；15.33；26.37；21.28；36.01；29.99；36.14；39.48；26.50；29.18；33.86；24.77；31.32；26.37；27.71；29.05；25.83；25.43；30.52；43.55；28.66；36.27；25.97；41.36；28.99；33.46；32.66；24.50；47.65；33.59；26.90；33.11；24.77；34.93；27.84；36.01；24.77；36.81；23.29；38.95；13.26；38.01；24.77；31.32；22.09；34.13；23.29；34.8；40.82；36.14；44.17；28.78	34.40；36.81；34.30；45.50；23.70；24.77；37.08；33.19；47.65；25.83；12.19；32.92；53.53；28.51；25.97；21.42；35.34；24.63；38.54；13.52；31.46；35.66；28.11；30.52；28.91；20.21；27.44；26.77；31.59；36.01；27.04；37.08；30.66
最小值/(g·kg^{-1})	13.26	12.19
最大值/(g·kg^{-1})	50.55	53.53
平均值/(g·kg^{-1})	31.24	30.98
标准差/(g·kg^{-1})	7.63	8.48
变异系数/%	24.44	27.41

3.3　PLSR 模型校准与预测

为了充分拟合数据，本研究采用留一交叉验证法寻找主成分因子(Principal component factor，PCF)，通过交叉验证的均方根误差(Root mean square error of cross validation，RMSECV)验证模型的稳健性，得到图 12 所示 RMSECV 和 PCF 的关系图。从图 12 中可以看出，当 PCF 数量为 3、4 个时，均方根误差最小；且较少的 PCF 数量可以降低模型的复杂性，因此本研究采用 4 个 PCF 建立了 PLSR 模型。

利用已确定的 PCF 和人工嗅觉特征空间划分的训练集建立 PLSR 模型，并采用测试集对模型进行验证，如图 13 所示。利用训练集建立的模型，其 R^2 = 0.8102 6，RMSE 为 3.499 2 g/kg，MAE 为 2.688 g/kg(图 13a)，测试集的 R^2 = 0.808 78，RMSE 为 3.678 4 g/kg，MAE 为 3.107 9 g/kg(图 13b)。试验结果表明，PLSR 算法建立的模型，其训练集和测试集的 R^2 大于 0.80，RMSE 和 MAE 均小于 3.7 g/kg，模型具有一定的泛化能力。

3.4　BPNN 模型校准与预测

采用 Matlab 9.7.0.1190202(R2019b)中的神经

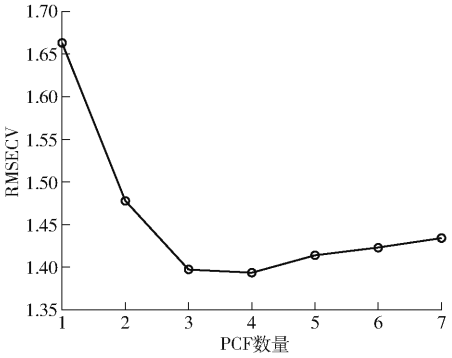


图 12 PLSR 中 PCF 数量与 RMSECV 关系

Fig. 12 Relationship between PCF number and RMSECV in PLSR

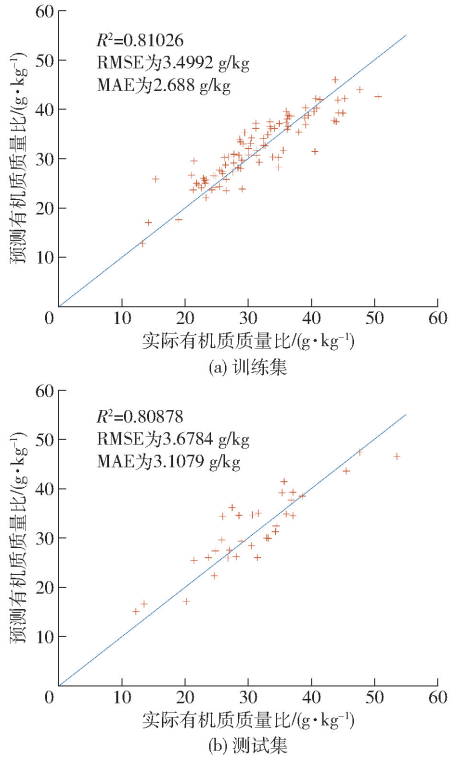


图 13 PLSR 模型训练集和测试集的预测结果

Fig. 13 Predicted results of training set and test set of PLSR model

网络工具箱建立 BPNN 模型。在 BPNN 建模中,隐含层采用 S 型传递函数 tansig,输出层采用线性传递函数 purelin,并采用 feedforwardnet 函数创建前向传播神经网络。将训练迭代次数设置为 1 000,学习率设置为 0.01,目标误差设置为 0.001,用 train 函数将训练集代入神经网络。

为了选择合适的隐含层神经元数 h,在以上确定的 h 取值范围内,选取程序运行 10 次对应决定系数的平均值作为评价指标,结果如表 4 所示。R^2 越大,表明模型的准确性和泛化能力越好。当 h 为 8 时,R^2 平均值为 0.849 31,最大值为 0.893 48,均为最大。因此,模型参数确定为 60 - 8 - 1(60 个输入量,8 个隐含层神经元,1 个输出量),其训练集和测试集的预测结果如图 14 所示。

表 4 隐含层神经元数对 BPNN 模型性能的影响

Tab. 4 Effect of number of hidden layer neurons on performance of BPNN model

神经元数	R^2		
	最小值	最大值	平均值
5	0.512 56	0.792 15	0.695 52
6	0.495 27	0.821 69	0.742 56
7	0.651 68	0.884 25	0.835 48
8	0.786 81	0.893 48	0.849 31
9	0.728 44	0.863 54	0.800 21
10	0.476 29	0.829 95	0.765 15
11	0.584 84	0.775 51	0.725 69

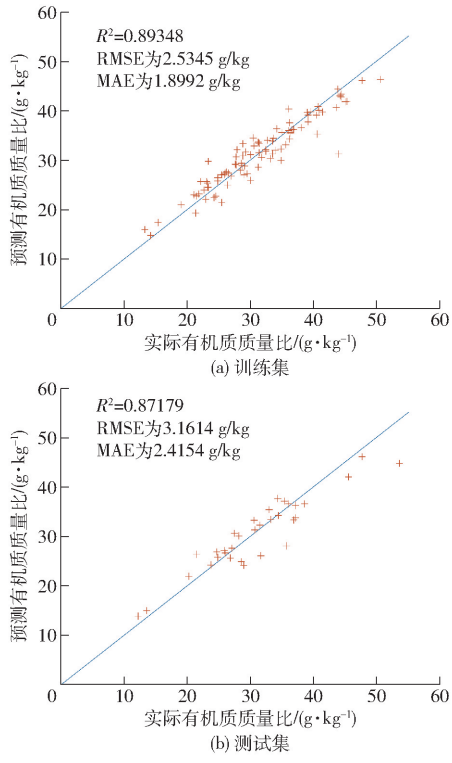


图 14 BPNN 模型训练集和测试集的预测结果

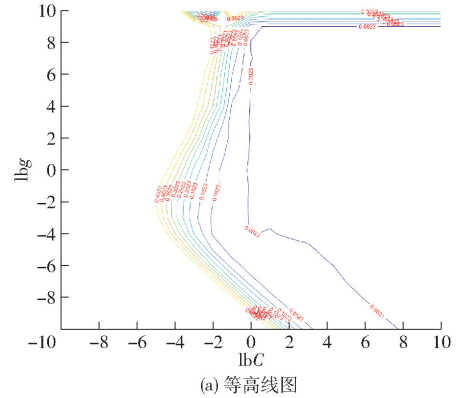
Fig. 14 Predicted results of training set and test set of BPNN model

试验结果表明,BPNN 算法建立的模型,其训练集的 R^2 = 0.893 48,RMSE 为 2.534 5 g/kg,MAE 为 1.899 2 g/kg(图 14a),测试集的 R^2 = 0.871 79,RMSE 为 3.161 4 g/kg,MAE 为 2.415 4 g/kg(图 14b),模型具有较好的拟合效果。

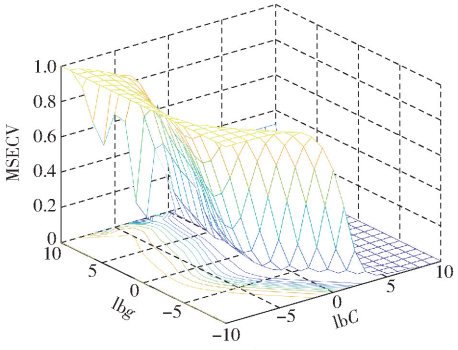
3.5 SVR 模型校准与预测

SVR 模型的泛化能力主要受惩罚因子 C 和核参数 g 的影响,为了提高模型的预测精度,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选择参数组合(C,g)的值。图 15 分别用等高线和 3D 视图(MSECV 经 mapminmax 函数处理为 0 ~ 1 间的数)显示选定的支持向量回归参数的结果,C 和

g 的最佳值分别为 128 和 0.001 953 1，MSECV 为 26.479 7 g/kg。



(a) 等高线图



(b) 3D视图

图 15　支持向量回归(SVR)参数选择

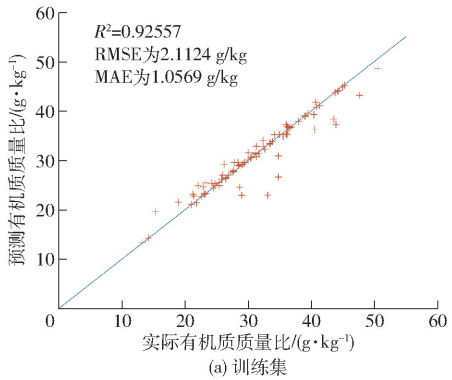
Fig. 15　Support vector regression parameter selection

利用优化后的 C 和 g 构建 SVR 模型，得到如图 16 所示结果。为观察该模型的预测效果，本研究使用测试集对其进行检验。结果显示，训练集 R^2 = 0.925 57，RMSE 为 2.112 4 g/kg，MAE 为 1.056 9 g/kg(图 16a)；测试集 R^2 = 0.919 57，RMSE 为 2.425 4 g/kg，MAE 为 2.138 9 g/kg(图 16b)。这表明，网格搜索法和 5-折交叉验证法选择的组合(C = 128，g = 0.001 953 1)具有良好的预测性能。

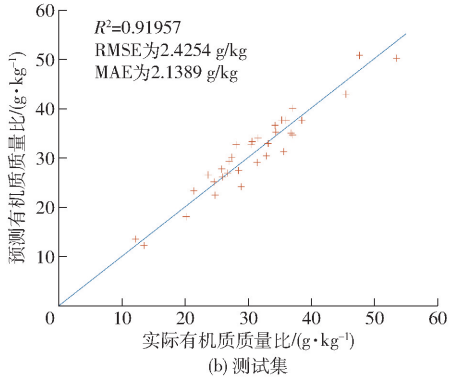
3.6　模型对比

许多回归模型均有较高的预测精度，如 MLR、SVR 和 RBF 等。为了选择合适的预测模型，以提高土壤有机质嗅觉检测模型的泛化能力，采用 BPNN、SVR 和 PLSR 这 3 种算法分别对人工嗅觉特征空间的训练集和测试集进行测试。从图 13a、14a、16a 可以看出，在 PLSR、BPNN、SVR 3 个训练集中，SVR 的预测效果最好(R^2最大，RMSE 和 MAE 最小)，BPNN 次之，PLSR 最差。但 R^2均大于 0.81，表明 3 个模型具有较好的预测能力。

利用测试集对已经训练完成的 BPNN、SVR 和 PLSR 模型进行测试，得到图 13b、14b、16b 中的预测结果。为了进行更直接的比较，在图 17 中绘制了不同模型的预测结果，并在表 5 中列出了相关预测性能。从图 17 中可以看出样本 12、13、22 在 3 种模型



(a) 训练集



(b) 测试集

图 16　SVR 模型训练集和测试集的预测结果

Fig. 16　Predicted results of training set and test set of SVR model

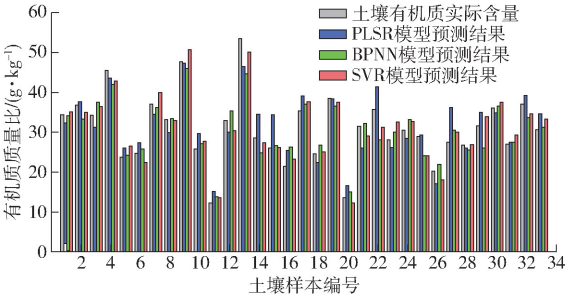


图 17　PLSR、BPNN 和 SVR 预测结果比较

Fig. 17　Comparison of prediction results of PLSR, BPNN and SVR

表 5　PLSR、BPNN 和 SVR 模型的 SOM 预测性能

Tab. 5　SOM prediction performance of PLSR, BPNN and SVR models

预测模型	R^2	RMSE/(g·kg⁻¹)	MAE/(g·kg⁻¹)
PLSR	0.808 78	3.678 4	3.107 9
BPNN	0.871 79	3.161 4	2.415 4
SVR	0.919 57	2.425 4	2.138 9

的预测值与实际土壤有机质含量均有较大差异。原因可能是训练集中一个或几个样本的人工嗅觉测量过程中产生了误差，这可能是由于操作不当、人工嗅觉装置本身的误差或温度、湿度等外部因素造成的。

从测试集的预测结果中可以看出(表 5)，3 个模型的 R^2均大于 0.80，表明所有模型都有较好的预测能力。因此，在 3 种算法中，SVR 的预测精度最

好,其次是 BPNN,PLSR 也获得了较好的预测性能。SVR 和 BPNN 比 PLSR 具有更高的预测性能,可能是因为土壤有机质含量与人工嗅觉特征空间存在一定程度的非线性关联,而支持向量回归机和 BPNN 比 PLSR 更适合非线性回归。SVR 的表现比 BPNN 好,主要原因可能是支持向量回归机的模型学习能力更强,因为它使用了一组最佳参数组合(C,g)。

4 结论

(1)采用 10 种不同类型的氧化物半导体式气体传感器构建传感器阵列,实现了利用人工嗅觉系统对土壤有机质含量的检测。采用不同浓度的硫化氢、氨气和甲烷等标准气体对传感器阵列进行了响应测试,从传感器阵列对不同浓度、种类标准气体的响应曲线可以看出,传感器 S5 只对高浓度氨气响应,传感器 S1、S2 对甲烷几乎没有响应,其余传感器对不同浓度、种类的标准气体皆有响应且响应结果不同,这体现了气体传感器阵列较高的特异性和一定的交叉敏感性,且随着标准气体浓度的增大传感器阵列的响应曲线也随之增大。

(2)对比分析了 PLSR、BPNN 和 SVR 3 种算法在基于多传感器人工嗅觉系统的土壤有机质含量检测方法的检测精度。采用此 3 种算法建立人工嗅觉特征空间与土壤有机质含量之间的关系模型,以 R^2、RMSE 和 MAE 作为性能指标,将 3 种模型的预测能力进行比较。结果表明,3 种模型对土壤有机质含量均有较好的预测能力,其中 SVR 显示出最大的精度和最小的 RMSE 和 MAE。这种表现在一定程度上可能是由于土壤有机质含量和人工嗅觉特征空间之间的非线性相关性,以及支持向量回归机较强的学习能力。因此,支持向量回归模型可以作为评估土壤有机质含量的有效工具。

(3)试验结果表明,人工嗅觉系统给测定土壤有机质含量提供了支持,该方法稳健有效,且为预测和简化土壤有机质含量的测定提供了依据。

参 考 文 献

[1] 于雷, 洪永胜, 耿雷, 等. 基于偏最小二乘回归的土壤有机质含量高光谱估算[J]. 农业工程学报, 2015, 31(14): 103 - 109.
YU Lei, HONG Yongsheng, GENG Lei, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression[J]. Transactions of the CSAE, 2015, 31(14): 103 - 109. (in Chinese)

[2] HUANG D, LIU H, ZHU L, et al. Soil organic matter determination based on artificial olfactory system and PLSR - BPNN[J]. Measurement Science and Technology, 2021, 32(3): 035801.

[3] 关连珠. 普通土壤学[M]. 北京:中国农业大学出版社, 2016: 42 - 45.

[4] 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 and prospect of soil organic matter determination methods[J]. Soil, 2015, 47(3): 453 - 460. (in Chinese)

[5] 刘军根, 汪文俊, 罗娇, 等. 土壤有机质含量测定方法对比分析研究[J]. 中国化工贸易, 2017, 9(21): 212 - 215.
LIU Jungen, WANG Wenjun, LUO Jiao, et al. Comparative analysis of soil organic matter content determination methods[J]. China Chemical Trade, 2017, 9(21): 212 - 215. (in Chinese)

[6] 赵龙华, 刘小粉, 王雅婧, 等. 基于热分析技术的土壤有机质含量和稳定性分析[J]. 农业工程学报, 2016, 32(10): 105 - 114.
ZHAO Longhua, LIU Xiaofen, WANG Yajing, et al. Thermal analysis determining soil organic matter content and thermal stability[J]. Transactions of the CSAE, 2016, 32(10): 105 - 114. (in Chinese)

[7] CONDIT H R. The spectral reflectance of American soils[J]. Photogrammetric Engineering, 1970, 36: 955 - 966.

[8] 黄应丰, 刘腾辉. 华南主要土壤类型的光谱特性与土壤分[J]. 土壤学报, 1995, 32(1): 58 - 68.
HUANG Yingfeng, LIU Tenghui. Spectral characteristics and soil distribution of main soil types in South China[J]. Acta Pedologica Sinica, 1995, 32(1): 58 - 68. (in Chinese)

[9] STONE E R, BAUNGARDNER M F. Characteristic variations in reflectance of surface soils[J]. Soil Science Society of America Journal, 1981, 45: 1161 - 1165.

[10] VAN W C, MESTDAGH I, LOOTENS P, et al. Possibilities of near infrared reflectance spectroscopy for the prediction of organic carbon concentrations in grassland soil[J]. Journal of Agricultural Science, 2005, 143(6): 487 - 492.

[11] LIU W, BARET F, GU X, et al. Relating soil surface moisture to reflectance[J]. Remote Sensing of Environment, 2002, 81(2 - 3):238 - 246.

[12] BA Y, LIU J, HAN J, et al. Application of vis - NIR spectroscopy for determination the content of organic matter in saline-alkali soils[J]. Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy, 2019, 229: 117863.

[13] BOWERS S A, HANKS R J. Reflection of radiant energy from soil[J]. Soil Science, 1965, 100(2): 130 - 138.

[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 & Biochemistry, 2011, 43: 2094 - 2107.

[15] CONRAD R. Soil microorganisms as controllers of atmospheric trace gases (H_2, CO, CH_4, OCS, N_2O, and NO)[J]. Microbiological Reviews, 1996, 60(4): 609 - 640.

[16] JONATHAN W L, NOAH F. Volatile organic compound (VOC) emissions from soil and litter samples[J]. Soil Biology and Biochemistry, 2008, 40(7): 1629 - 1636.

[17] SHEPPARD S K, LLOYD D. Direct mass spectrometric measurement of gases in soil monoliths[J]. Journal of Microbiological Methods, 2002, 50(2): 175 - 188.

[18] MCNEAL K S, HERBERT B E. Volatile organic metabolites as indicators of soil microbial activity and community composition shifts[J]. Soil Science Society of America Journal, 2009, 73(2): 579 - 588.

[19] FERNADO W G D, RAMARATHNAM R, KRISHNAMOORTHY A S, et al. Identification and use of potential bacterial organic antifungal volatiles in biocontrol[J]. Soil Biology & Biochemistry, 2005, 37(5): 955 - 964.

[20] TOZLU B H, OKUMUS H İ. A new approach to automation of black tea fermentation process with electronic nose[J]. Automatika, 2018, 59(3 - 4): 373 - 381.

[21] SCHNABEL R M, BOUMANS M L L, SMOLINSKA A, et al. Electronic nose analysis of exhaled breath to diagnose ventilator-associated pneumonia[J]. Respiratory Medicine, 2015, 109(11): 1454.

[22] SCHMEKEL B, WINQUIST F, VIKSTROM A. Analysis of breath samples for lung cancer survival[J]. Analytica Chimica Acta, 2014, 840: 82 - 86.

[23] GIL-SANCHEZ L, SOTO J, MARTINEZ-MANEZ R, et al. A novel humid electronic nose combined with an electronic tongue for assessing deterioration of wine[J]. Sensors and Actuators A: Physical, 2011, 171(2): 152 - 158.

[24] ANDRZEJ B, KATARZYNA J G, LUKASZ G, et al. Evaluating soil moisture status using an e-nose[J]. Sensors, 2016, 16(6): 886.

[25] LAVANYA S, NARAYANAN B D S, MURTHY V K, 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] 朱龙图,李名伟,夏晓蒙,等. 基于人工嗅觉系统的土壤有机质检测方法研究[J/OL]. 农业机械学报, 2020, 51(3): 171 - 179.
ZHU Longtu, LI Mingwei, XIA Xiaomeng, et al. Soil organic matter detection method based on artificial olfactory system[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(3): 171 - 179. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx? flag = 1&file_no = 20200320&journal_id = jcsam. DOI: 10.6041/j.issn.1000-1298.2020.03.020. (in Chinese)

[27] 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.

[28] LU P, WANG L, NIU Z, et al. Prediction of soil properties using laboratory vis - NIR spectroscopy and hyperion imagery[J]. Journal of Geochemical Exploration, 2013, 132: 26 - 33.

[29] LI G, FENG X, QIU G, et al. Environmental mercury contamination of an artisanal zinc smelting area in Weining County, Guizhou, China[J]. Environmental Pollution, 2008, 154(1): 21 - 31.

[30] GOODLASS G, SMITH K A. Effect of pH, organic matter content and nitrate on the evolution of ethylene from soils[J]. Soil Biol. Biochem., 1978, 10(3): 193 - 199.

[31] ROSSEL R, WALVOORT D, MCBRATNEY A B, et al. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties[J]. Geoderma, 2006, 131(1 - 2): 59 - 75.

[32] LI X, XIANG S, ZHU P, et al. Establishing a dynamic self-adaptation learning algorithm of the BP neural network and its applications[J]. International Journal of Bifurcation and Chaos, 2015, 25(14): 1540030.

[33] VAPNIK V N, CHERVONENKIS A Y. On the uniform convergence of relative frequencies of events to their probabilities[J]. Theory of Probability & Its Applications, 1971, 16(2): 264 - 280.

[34] KUANG B, MOUAZEN A M. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms[J]. European Journal of Soil Science, 2011, 62(4): 629 - 636.