

文章编号: 0253-2239(2009)12-3556-05

基于非线性径向基核函数支持向量机的堆肥产品 近红外光谱分析研究

黄光群 韩鲁佳

(中国农业大学工学院, 北京 100083)

摘要 探讨了径向基核函数(RBF)支持向量机(RBF-SVR)建立定量分析模型时主要参数的优选方法及其在近红外光谱分析畜禽粪便堆肥产品含水率、挥发性固体含量和碳氮比中的应用,并与偏最小二乘回归法所建近红外定量分析模型的预测能力做了比较。供试样品为我国22省市不同种类的120个畜禽粪便堆肥产品样品,利用傅里叶变换型光谱仪获取样品在4000~10000 cm⁻¹内的光谱数据信息。研究发现,逐步寻优循环优选支持向量机建模参数方法具有较好的可行性,其所建近红外定量分析模型均优于基于偏最小二乘法所建模型,所建立的含水率和挥发性固体近红外模型验证决定系数(r^2)均大于0.90,相对分析误差(RPD)均大于4.0,具有实际应用价值;所建碳氮比近红外模型验证决定系数(r^2)为0.85,RPD值大于2.5,也可用于定量分析,但精度有待于进一步提高。

关键词 光谱学;堆肥品质分析;近红外漫反射光谱;支持向量机

中图分类号 S14;X7 文献标识码 A doi: 10.3788/AOS20092912.3556

Near Infrared Reflectance Spectroscopy Analysis of Compost Products Using Nonlinear Support Vector Machine With RBF Nucleus

Huang Guangqun Han Lujia

(College of Engineering, China Agricultural University, Beijing 100083, China)

Abstract This study explored a new method to choose optimal parameters for support vector machine regression with RBF nucleus (RBF-SVR) and its application on the estimation of moisture content, volatile solid (VS) and the ratio of carbon to nitrogen (C/N) in animal manure compost products using near-infrared reflectance spectroscopy (NIRS). The efficiency of RBF-SVR method was compared with partial least-squares regression (PLSR) mainly using the determination coefficient of prediction (r^2) of the standard error of prediction (SEP) and ratio of performance to standard deviation [RPD (SD/SEP)]. In this study, 120 commercial animal manure compost samples were collected from 22 provinces in China. Spectra of the orient samples were scanned with a SPECTRUM ONE NTS from 4000~10000 cm⁻¹, respectively. Results showed stepwise search for optimal parameters was a feasible method for RBF-SVR. The efficiency of RBF-SVR method for moisture content, VS and C/N were all better than PLSR. Robust models using RBF-SVR were developed for moisture content and VS ($r^2 > 0.90$, RPD > 4.0) and for C/N ($r^2 > 0.85$, RPD > 2.5), respectively. Results showed the potential of NIRS with RBF-SVR to evaluate the products quality of animal manure compost, but further research would be needed for the higher precision.

Key words spectroscopy; compost quality analysis; near-infrared reflectance spectroscopy; support vector machine

收稿日期: 2009-01-16; 收到修改稿日期: 2009-03-24

基金项目: 国家“十一五”科技支撑项目(2006BAD10B05, 2006BAD07A13)资助课题。

作者简介: 黄光群(1979—),男,博士研究生,讲师,主要从事生物质资源开发与利用等方面的研究。

E-mail: huangguangqun@126.com

导师简介: 韩鲁佳(1964—),女,教授,博士生导师,主要从事生物质资源开发与利用等方面的研究。

E-mail: hanlj@cau.edu.cn(通信联系人)

1 引言

随着畜牧养殖业集约化发展,我国畜禽粪便年产量呈现不断增长的趋势。大量的畜禽粪便若处理不当,极易造成生态和环境污染。近年来,畜禽粪便高温好氧堆肥化处理已成为无害化和资源化利用的重要途径^[1]。含水率、挥发性固体含量和碳氮比对于衡量堆肥腐熟度以及堆肥品质评价具有重要意义。传统的分析方法不仅需要专用仪器设备,且操作步骤繁琐、费时费力。国外已有研究表明近红外光谱法测定畜禽粪便堆肥含水率、挥发性固体含量和碳氮比具有较好的可行性^[2~4]。

基于近红外光谱分析建立定量模型的常用方法有多元线性回归和偏最小二乘法等。支持向量机(SVM)是借助最优化方法解决机器学习问题的新工具,它利用核函数的思想在机器学习领域取得了成功^[5],国内外研究者对其在近红外建模中的应用进行了探讨^[6~8],但核参数优选方法尚不成熟。本文探讨支持向量机在近红外光谱分析畜禽粪便堆肥产品含水率、挥发性固体含量和碳氮比中的应用,并探讨径向基函数(RBF)核参数优选途径,旨在建立和优化畜禽粪便堆肥产品含水率、挥发性固体含量和碳氮比的近红外光谱定量分析方法和模型。

2 实验部分

2.1 实验样品的收集与制备

本研究自2003年5月至2005年10月共收集畜禽粪便堆肥产品样品120个。供试样品主要来源于黑龙江、内蒙古、北京、河北、山东、河南、上海、湖南、安徽、四川、福建、广东、贵州、海南、宁夏和新疆等22省市。堆制原料包括牛粪、猪粪和鸡粪等不同种类畜禽粪便,调理剂以农作物秸秆为主,包括不同堆制方式(条堆或垛堆、强制通风或自然通风、机械翻堆或人工翻堆)、不同腐熟度(以成品为主)和不同形状(粉末状或颗粒状)。样品收集后按四分法取样,并分为两部分,一部分用于近红外光谱的采集,一部分用于实验室化学分析。

2.2 堆肥样品成分含量的实验室化学分析

含水率的测定采用烘箱干燥法(70℃,24h),挥发性固体含量测定采用灼烧法(550℃下至恒重),总有机碳含量的测定采用重铬酸钾外加热法,总氮含量的测定采用凯氏定氮法,碳氮比为总有机碳与总氮含量比。每个试样取两个平行样进行测定,以其算术平均值为最终结果。以上方法参照我

国有机肥料农业行业标准和美国堆肥协会测定标准^[9,10]。

2.3 近红外光谱仪与样品光谱扫描

实验中所用仪器有傅里叶变换型近红外光谱仪(SPECTRUM ONE NTS,美国Perkin Elmer公司),积分球附件,InGaAs检测器。样品近红外漫反射光谱采集时仪器工作参数:光谱范围4000~10000 cm⁻¹,扫描次数32次,分辨率8 cm⁻¹。每个样品重复装样3次进行扫描,取其平均光谱为样品光谱。

2.4 近红外光谱定量分析模型的建立与评价

首先根据样品理化指标成分含量大小排序后隔3取1作为独立验证集,其余样品组成校正集^[11]。回归方法除采用支持向量机方法外,同时选用偏最小二乘法(PLS)作比较,分别结合平滑(Smooth)、导数(Derivative)、变量标准化(SNV)和多元散射校正(MSC)等光谱预处理方法建立近红外光谱校正模型,并通过独立的验证集进行模型的外部验证。校正模型建立过程中根据马氏距离和化学值绝对误差分别对光谱和化学值进行异常值检验与剔除^[12],根据校正决定系数(R^2)、校正标准差(SEE)、验证决定系数(r^2)、验证标准差(SEP)和相对分析误差(RPD,SD/SEP)评价所建模型^[13],如果RPD大于3,说明定标效果良好,所建模型可以用于实际检测;如果RPD介于2.5和3之间,说明利用NIRS进行定量分析是可行,但精度有待于提高;如果RPD小于2.5,则说明该成分难于进行NIRS定量分析^[2]。

偏最小二乘法和支持向量机建立近红外定量分析模型过程分别使用美国Perkin Elmer公司QUANT⁺定量分析软件和Matlab7.0平台实现。

2.5 支持向量机建模流程及参数优选方法

将经过预处理的样品光谱在4000~10000 cm⁻¹范围内对应的吸光度值以数据点格式表示,数据间隔2 cm⁻¹,每个样品光谱得到不同厘米波数处对应的吸光度数据点3001个。将校正集样品光谱数据和样品理化指标含量测定值作为原始空间中的训练集,独立的验证集检验所建模型^[14]。

建模时选用高斯核函数(RBF核): $k(x, x_i) = \exp(-\gamma |u - v|^2)$ 。RBF核支持向量机建模需优选的参数主要有 v 和 γ 和正则化系数 C 。其中 v 和 γ 对模型预测效果影响最大。本研究中,固定 v ,将 (C, γ) 作为一组优化对象, C 赋值为 $10^1 \sim 10^7$, γ 赋值为 $10^{-4} \sim 10^4$,递进调为1.1, v 值默认为0.1,以预测标准差SEP最小为原则进行逐步比较求优循环运

算.最后进行 v 值在 $0 \sim 1$ 区间的循环运算,选出建模最优参数组合(v, C, γ)。

3 结果与讨论

3.1 堆肥样品理化指标测定分析结果

120 个畜禽粪便堆肥产品样品的含水率、挥发性固体含量和碳氮比见表 1。由表中结果可知,所收集堆肥产品样品的含水率、挥发性固体含量和碳氮比涵盖了较大的范围,具有较好的代表性。

表 1 堆肥样品含水率、挥发性固体含量和碳氮比范围
Table 1 Variation range of moisture, volatile solid and C/N ratio in compost samples

Parameter	Sample number	Min	Max	Mean	SD
Moisture / (g/kg)	120	31.98	809.33	224.91	156.00
Volatile solid / (g/kg)	120	100.37	642.42	387.00	142.11
C:N	120	0.83	26.73	11.16	4.47

3.2 堆肥样品的近红外漫反射光谱

图 1 为堆肥样品的近红外漫反射光谱图,图中横坐标为波数,纵坐标为吸光度。由图中可以看出,样品来源(原料种类和腐熟度等)虽然具有较大的差

表 2 偏最小二乘法近红外建模信息统计

Table 2 Statistics of NIRS models using PLS method

Parameter	Number of abnormal samples	Pretreatments of spectra	Factor	R^2	SEE	r^2	SEP	RPD	Bias
Moisture / (g/kg)	2	1(9)	6	0.982	21.98	0.983	20.99	7.48	-2.24
Volatile solid / (g/kg)	0	1(5)	6	0.937	37.29	0.938	35.07	4.05	-0.03
C:N	0	1(25)	5	0.741	2.29	0.831	2.01	2.39	-0.08

注:1(9)为 1 阶 9 点导数

3.4 支持向量机近红外建模参数优选

利用支持向量机建立堆肥样品中含水率、挥发

异性,但其光谱图具有一致的相似性,且光谱信息主要集中在长波近红外谱区内。图中在 6940 cm^{-1} 和 5155 cm^{-1} 附近出现的两个较明显的吸收峰表明水分在该区获得强烈吸收。

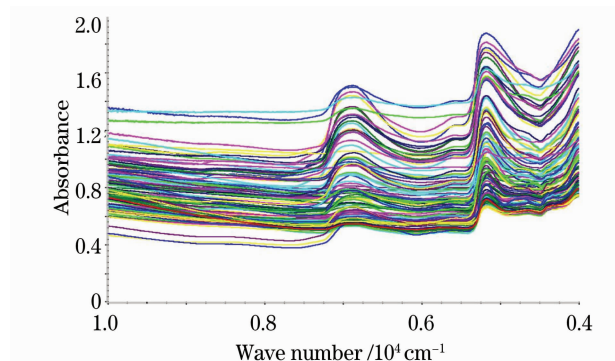


图 1 堆肥样品的近红外漫反射光谱

Fig. 1 NIR spectra of compost samples

3.3 偏最小二乘法近红外光谱定量模型的建立

表 2 基于偏最小二乘法的近红外光谱建模的异常样品剔除、最优光谱预处理方法及模型信息统计。由表中可知,光谱的一阶导数光谱处理有利于光谱特征变量信息的提取。根据 Malley 等^[2]关于近红外定量分析模型评判准则,偏最小二乘法所建立的含水率、挥发性固体含量和碳氮比 NIRS 模型均具有较高的预测精度,可以用于定量分析。

性固体和碳氮比的近红外定量分析模型过程中 RBF 核参数 v, γ 和正则化系数 C 的优选过程如图 2 中所示(以含水率为例)。人工试选方法通常需要操

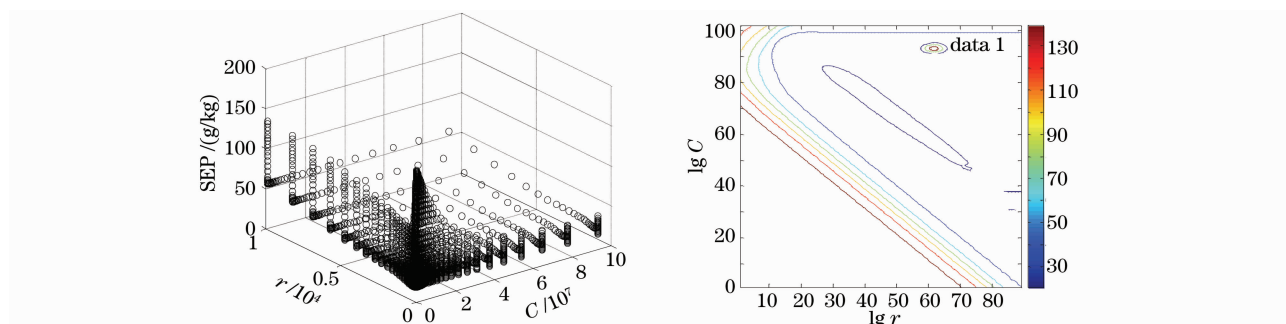


图 2 含水率支持向量机建模参数优选

Fig. 2 Optimization of parameters C and γ for moisture model using SVM

作人员掌握专业的技术知识,且效率较低,而本研究中提出的逐步比较寻优的循环方法与人工试选方法相比更具有更好的操作性,依靠计算机更好的运行效率和高可靠性可以最大限度的确保所选取的参数组合最优,确保模型预测误差最小化。

表 3 堆肥样品支持向量机-近红外建模最优参数

Table 3 Optimal parameters for SVM-NIRS models of compost samples

Parameter	Calibration samples	Validation samples	Mathematical treatment	ν	γ	C
Moisture	88	30	1(9)	0.5	24.23	5.27×10^4
Volatile solid	90	30	1(5)	0.7	4.70	9.10×10^4
C:N	90	30	1(25)	0.7	1337.6	154.07

3.5 支持向量机所建近红外光谱定量分析模型

表 4 为利用支持向量机建立的堆肥样品中含水率、挥发性固体含量和碳氮比近红外校正模型及外部验证结果统计。根据 Malley 等关于近红外定量分析模型评判准则,利用非线性 RBF 核支持向量机所建立的含水率、挥发性固体含量和碳氮比 NIRS 模型均具有较高的预测精度,可以用于定量分析。尤其对于含水率和挥发性固体含量而言,所建模型的验证决定系数(r^2)和验证相对分析误差(RPD)分别大于 0.9 和 4.0,完全可以用于实际定量检测。

比较非线性 RBF 核支持向量机和偏最小二乘法所建三种组分的近红外预测模型,由于非线性 RBF 核的支持向量机回归方法具有良好的机器学习能力和模型泛化能力,其所建模型的预测能力均优于偏最小二乘法所建模型。

基于 RBF 核函数支持向量机的近红外光谱校正模型的分集情况、最优光谱预处理方法及最优建模参数如表 3 所列。其中,各指标建模时采用的校正集和验证集样品与基于最小二乘法建模时采用的校正集和验证集样品相同。

表 4 支持向量机所建近红外模型与验证信息统计

Table 4 Statistics of NIRS calibrations and validations using SVM

Parameter	R^2	SEE	r^2	SEP	RPD	Bias
Moisture / (g/kg)	0.991	15.44	0.988	17.93	8.75	2.89
Volatile solid / (g/kg)	0.948	35.02	0.945	34.86	4.07	2.23
C:N	0.954	1.63	0.853	1.84	2.61	0.03

如图 3 所示,两种回归方法所建立的碳氮比定量分析模型预测精度略低,验证决定系数(r^2)和验证相对分析误差(RPD)分别为 0.85 和 2.5 左右,预测精度有待于进一步提高,可能是由于碳氮比为碳氮两种组分质量的间接比值增加了累计误差。

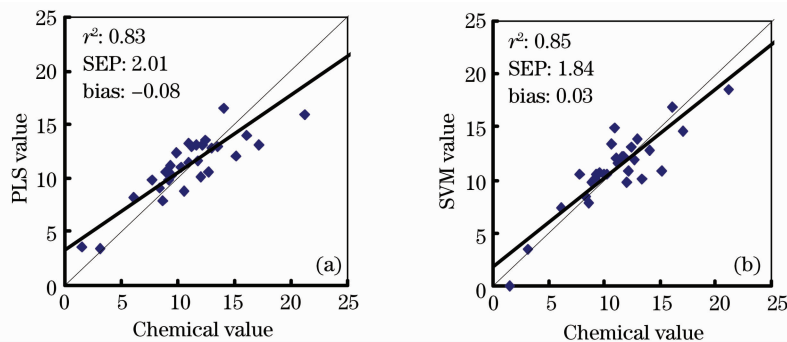


图 3 SVM 与 PLS 所建碳氮比 NIRS 模型预测值与化学值相关关系(验证集)

Fig. 3 Chemical value versus NIRS values of C/N using SVM and PLS(validation)

4 结 论

利用非线性 RBF 核支持向量机作为回归方法建立畜禽粪便堆肥产品中含水率、挥发性固体含量和碳氮比近红外定量分析模型具有良好的可行性。逐步寻优循环参数优选方法选取 RBF 核参数具有高效、准确和易操作的特点,与偏最小二乘回归法所

建近红外定量分析模型相比较,非线性 RBF 核支持向量机所建近红外模型优于基于偏最小二乘法所建模型,所建立的含水率和挥发性固体模型验证决定系数(r^2)均大于 0.90,RPD 值均大于 4.0,具有实际应用价值;所建碳氮比模型验证决定系数(r^2)为 0.85,RPD 值大于 2.5,也可用于定量分析。

参 考 文 献

- 1 R. T. Haug. The Practical Handbook of Compost Engineering [M]. London: Lewis Publishers, 1993. 1~10
- 2 D. F. Malley, C. McClure, P. D. Martin *et al.*. Compositional analysis of cattle manure during composting using a field-portable near-infrared spectrometer [J]. *Communication in Soil Science and Plant Analysis*, 2005, **36**(4): 455~475
- 3 G. Q. Huang, L. J. Han, Liu X. Rapid estimation of the composition of animal manure compost by near infrared reflectance spectroscopy [J]. *J. Near Infrared Spectroscopy*, 2007, **2**(1): 79~84
- 4 R. Albrechta, R. Joffreb, R. Grosa *et al.*. Efficiency of near-infrared reflectance spectroscopy to assess and predict the stage of transformation of organic matter in the composting process [J]. *Bioresource Technology*, 2008, **99**(2): 448~455
- 5 Deng Naiyang, Tian Yingjie. New Method in Data Mining: Support Vector Machine [M]. Beijing: Science Press, 2004. 1
邓乃扬, 田英杰. 数据挖掘中的新方法-支持向量机 [M]. 北京: 2004. 1
- 6 A. Taloy, R. W. McCabe. Removal of hydrogen bonding interferences in the near infrared spectroscopic determination of dicarboxylic acids in diols [J]. *J. Near Infrared Spectroscopy*, 2004, **12**(2): 101~104
- 7 Zhang Rongxiang, Zeng Shijie, Xia Yanjun *et al.*. Application of support vector regression for reconstruction of non-uniform strain profile along the fiber grating [J]. *Acta Optica Sinica*, 2008, **28**(8): 1513~1517
张荣祥, 郑世杰, 夏彦君 等. 支持向量回归算法在光纤光栅非均匀应变重构中的应用 [J]. *光学学报*, 2008, **28**(8): 1513~1517
- 8 Zhang Luda, Su Shiguang, Wang Laisheng *et al.*. Study on application of Fourier transformation near-infrared spectroscopy analysis with support vector machine [J]. *Spectroscopy and Spectral Analysis*, 2005, **25**(1): 33~35
张录达, 苏时光, 王来生 等. 支持向量机在傅里叶变换近红外光谱分析中的应用研究 [J]. *光谱学与光谱分析*, 2005, **25**(1): 33~35
- 9 Agricultural Criterion of People's Republic of China-Microbial Organic Fertilizer (中华人民共和国农业行业标准-生物有机肥) [S]. (标准号: NY884-2004)
- 10 W. H. Thompson, P. B. Leege, P. D. Millner *et al.*. Test methods for the examination of composting and compost [S]. Washington: US Composting Council, 2002
- 11 X. Liu, L. J. Han. Evaluation of near-infrared reflectance spectroscopy (NIRS) for predicting chemical composition of straw silage [J]. *J. Animal and Feed Science*, 2006, **15**(2): 329~336
- 12 Min Shungeng, Li Ning, Zhang Mingxiang. Outlier diagnosis and calibration model optimization for near infrared spectroscopy analysis [J]. *Spectroscopy and Spectral Analysis*, 2004, **24**(10): 1205~1209
闵顺耕, 李 宁, 张明祥. 近红外光谱分析中异常值的判别与定量模型优化 [J]. *光谱学与光谱分析*, 2004, **24**(10): 1205~1209
- 13 Lu Wanzhen, Yuan Hongfu, Xu Guangtong *et al.*. The Modern Analysis Technique for Near-Infrared Spectra [M]. Beijing: Chinese Oil and Chemical Press, 2001: 1~13
陆婉珍, 袁洪福, 徐广通 等. 现代近红外光谱分析技术 [M]. 北京: 中国石油化工出版社, 2001. 1~13
- 14 W. Saeys, A. M. Mouazen, H. Ramon. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy [J]. *Biosystem Engineering*, 2005, **91**(4): 393~402