

基于 GSA 的厌氧发酵原料碳氮比 NIRS 快速检测

刘金明^{1,2} 程秋爽¹ 甄峰³ 许永花⁴ 李文哲^{1,5} 孙勇^{1,5}

(1. 东北农业大学工程学院, 哈尔滨 150030; 2. 黑龙江八一农垦大学电气与信息学院, 大庆 163319;
3. 中国科学院广州能源研究所中国科学院可再生能源重点实验室, 广州 510640;
4. 东北农业大学电气与信息学院, 哈尔滨 150030;
5. 黑龙江省寒地农业可再生资源利用技术及装备重点实验室, 哈尔滨 150030)

摘要: 在以预处理后玉米秸秆、秸秆粪便混合物为原料进行厌氧发酵生产沼气时, 为了对厌氧发酵原料碳氮比进行快速检测, 将近红外光谱 (NIRS) 与偏最小二乘 (PLS) 回归相结合构建快速检测模型, 并基于遗传模拟退火算法 (GSA) 构建遗传模拟退火区间偏最小二乘算法 (GSA-iPLS) 和双重遗传模拟退火偏最小二乘算法 (DGSA-PLS) 分别用于特征谱区优选和特征波长点优选, 以提高回归模型的检测精度和效率。全谱 1 844 个波长点经 GSA-iPLS 进行谱区优选后, 得到 641 个波长变量, 再经 DGSA-PLS 进行特征波长点优选后, 得到 628 个波长变量。DGSA-PLS 回归模型验证集的决定系数 (R_p^2) 为 0.920, 预测均方根误差为 7.178, 相对分析误差为 3.805。与全谱建模相比, DGSA-PLS 模型的 RMSEP 减小了 15.87%。通过波长优选, 参与建模的波长点数量显著减少, 有效降低了变量维度和模型复杂度, 提升了预测精度和预测能力。本文通过优选碳氮比的敏感波长变量, 有效提高了预测模型的鲁棒性, 为直接、快速、准确测量厌氧发酵原料的碳氮比提供了新途径。

关键词: 厌氧发酵; 碳氮比; 近红外光谱; 偏最小二乘回归; 遗传模拟退火算法

中图分类号: O657.33 **文献标识码:** A **文章编号:** 1000-1298(2019)11-0323-08

Rapid Determination of C/N Ratio for Anaerobic Fermentation Feedstocks Using Near Infrared Spectroscopy Based on GSA

LIU Jinming^{1,2} CHENG Qiushuang¹ ZHEN Feng³ XU Yonghua⁴ LI Wenzhe^{1,5} SUN Yong^{1,5}

(1. College of Engineering, Northeast Agricultural University, Harbin 150030, China
2. College of Electrical and Information, Heilongjiang Bayi Agricultural University, Daqing 163319, China
3. CAS Key Laboratory of Renewable Energy, Guangzhou Institute of Energy Conversion, Chinese Academy of Sciences, Guangzhou 510640, China
4. School of Electrical and Information, Northeast Agricultural University, Harbin 150030, China
5. Heilongjiang Key Laboratory of Technology and Equipment for the Utilization of Agricultural Renewable Resources, Harbin 150030, China)

Abstract: The essence of anaerobic fermentation (AF) is the cultivation process of microorganisms, and the carbon-nitrogen ratio (C/N) is an important factor that affects the production of biogas. To quickly detect the C/N for the AF feedstocks, such as the pretreated corn stover, the mixture of corn stover and feces, a rapid detection model was constructed based on near infrared spectroscopy (NIRS) combined with partial least squares (PLS) regression. To further improve the detection accuracy and efficiency of the model, the genetic simulated annealing interval partial least squares algorithm (GSA-iPLS) and double genetic simulated annealing partial least squares algorithm (DGSA-PLS) based on genetic simulated annealing algorithm (GSA) were proposed for selecting the efficient spectral regions and characteristic wavelength points for NIRS, respectively. Totally 1844 wavelength points of the whole spectrum were selected by GSA-iPLS, and 641 wavelength variables were obtained, and 628 wavelength variables were obtained after the characteristic wavelength points were optimized by DGSA-PLS. The

收稿日期: 2019-06-10 修回日期: 2019-07-06
基金项目: “十二五”国家科技支撑计划项目(2015BAD21B03)和中国科学院可再生能源重点实验室开放基金项目(Y907k81001)
作者简介: 刘金明(1981—),男,博士生,黑龙江八一农垦大学副教授,主要从事光谱分析技术应用研究,E-mail: jinmingliu2008@126.com
通信作者: 孙勇(1974—),男,教授,博士生导师,主要从事农牧废弃资源高值化利用技术与装备研究,E-mail: sunyong740731@163.com

coefficients of multiple determination for prediction (R_p^2), root mean squared error of prediction (RMSEP) and residual predictive deviation (RPD) in DGSA-PLS regressive model were 0.920, 7.178 and 3.805, respectively. Compared with the whole spectrum model, the RMSEP was decreased by 15.87% in the DGSA-PLS model. It was shown that the number of wavelengths was significantly decreased after the optimization, and the performance of regressive model was obviously higher than that of the whole wavelengths. The research improved the adaptability of the prediction model based on optimizing sensitive wavelength variables for C/N, which provided a new way for directly rapid and accurate measurement of the C/N of AF feedstock.

Key words: anaerobic fermentation; carbon-nitrogen ratio; near infrared spectroscopy; partial least squares regression; genetic simulated annealing algorithm

0 引言

碳氮比(碳、氮元素质量的比值)对厌氧发酵微生物的生长繁殖和产物合成具有重要影响,测定厌氧发酵底物的碳氮比已成为优化厌氧发酵原料配比的重要环节^[1-2]。在以玉米秸秆为主要原料进行厌氧发酵生产沼气时,通过预处理打破玉米秸秆自身紧密的木质纤维素结构,能够有效提高玉米秸秆的生物转化利用率^[3-4]。同时,由于玉米秸秆碳氮比过高,常与碳氮比较低的畜禽粪便混合共同发酵,以提高厌氧发酵产沼气的效率和能力^[5-6]。为了分析碳氮比对厌氧发酵的影响,进而对预处理后玉米秸秆及秸秆和粪便混合物的厌氧发酵过程进行有效调控,有必要对发酵原料的碳氮比进行快速、准确的测定,采用传统化学方法测定其碳氮比时存在测试速度慢、成本高的问题。

近红外光谱(Near infrared spectroscopy, NIRS)分析技术具有快速、低成本及多组分同步检测等优点^[7-8],已广泛用于有机动植物废弃物碳氮含量的快速测定^[9-10]。当使用 NIRS 对畜禽粪便中的碳、氮含量进行检测时,氮含量的检测精度明显优于碳含量^[11-12]。在使用 NIRS 对植物中所含碳、氮成分进行检测时,能够获得较高的氮含量检测精度^[13-14],但碳含量检测精度较低^[15-16]。针对厌氧发酵过程中原料碳氮比的快速检测需求,以及 NIRS 在秸秆和粪便碳氮含量检测方面的优势与不足,本文提出使用 NIRS 对厌氧发酵原料直接进行碳氮比的快速检测。

针对以全谱建模时冗余波长点严重影响模型检测精度和效率的问题,相关学者提出应用遗传算法(Genetic algorithm, GA)^[17-18]进行 NIRS 波长变量组合优化。GA 因其具有较强的鲁棒性和全局搜索能力,在 NIRS 特征波长优选方面得到了广泛应用,其随机搜索能力能够有效解决光谱波长点之间的共线性问题^[19],还可以与其他光谱谱区优化算法相结合进行特征波长点的优选^[20-21]。但 GA 存在早熟

问题,且进化后期搜索效率较低。遗传模拟退火算法(Genetic simulated annealing algorithm, GSA)^[22]通过将 GA 与模拟退火算法(Simulated annealing algorithm, SA)相结合,引入温度参数对适应度函数进行改进设计,在充分发挥算法强大搜索能力的同时,有效解决了 GA 存在的不足。

本文提出基于 GSA 构建遗传模拟退火区间偏最小二乘算法(Genetic simulated annealing interval partial least squares algorithm, GSA-iPLS)和双重遗传模拟退火偏最小二乘算法(Double genetic simulated annealing partial least squares algorithm, DGSA-PLS)分别用于特征谱区优选和特征波长点优选,进而获取与碳氮比相关性高的有效特征波长变量,建立厌氧发酵原料碳氮比的 NIRS 快速检测模型。

1 材料与方法

1.1 样品采集与制备

实验用玉米秸秆取自东北农业大学校内实验田,猪粪取自哈尔滨市三元畜产实业公司,牛粪取自哈尔滨市宇峰奶牛养殖农民专业合作社,羊粪取自东北农业大学阿城实验实习基地。玉米秸秆、猪粪、牛粪和羊粪各采集 1 个样品,用于后续秸秆预处理和秸秆粪便混合样品制备。采集的玉米秸秆样品自然风干后装袋保存,采集的猪粪、牛粪和羊粪样品于 -18°C 冷冻保存。在厌氧发酵样品制备过程中,先将风干玉米秸秆切成 10 mm 长的秸秆段,再将秸秆段、猪粪、牛粪、羊粪干燥、粉碎,过 40 目筛后装袋备用。依据秸秆厌氧发酵过程中碱性预处理的高效性和生物预处理的环境友好性,分别采用地衣芽孢杆菌(生物方法)、NaOH 溶液(碱性试剂)、猪粪沼液(富含微生物的弱碱性试剂)和 NaOH 沼液溶液(富含微生物的强碱性试剂)对玉米秸秆进行预处理实验。生物方法预处理秸秆样品取自本课题组进行的地衣芽孢杆菌秸秆降解实验^[23]。按最优培养条件活化培养地衣芽孢杆菌后,将其液体菌种接种于秸

秆粉末固体培养基上,进行为期 10 d 的降解实验,每 2 d 采样 1 次,共计采样 5 个。其它方法预处理实验过程中,将 10 mm 长的秸秆段浸泡于处理液中 3 s 后,捞出挤压排水并装入自封袋进行密封处理。定时采样并用蒸馏水充分洗涤样品 5 次后,将样品干燥、粉碎过 40 目筛后装袋密封保存,制备预处理秸秆样品 45 个。NaOH、沼液预处理秸秆实验方案如表 1 所示。

表 1 NaOH 和沼液预处理玉米秸秆实验设计
Tab.1 Experimental design of corn stover pretreatment by NaOH combined with biogas slurry

试剂	溶剂	质量分数/%	采样间隔/d	采样个数
NaOH	水	1	2	5
NaOH	水	2	2	5
NaOH	水	3	2	5
NaOH	水	4	2	5
NaOH	沼液	1	2	5
NaOH	沼液	2	2	5
NaOH	沼液	3	2	5
NaOH	沼液	4	2	5
	沼液		10	5

将粉碎玉米秸秆按比例与粉碎后的猪粪、牛粪、羊粪粉末进行混合,制备秸秆和粪便混合发酵原料样品 36 个。混合比例(质量比)为 9:1、8:2、7:3、6:4、5:5、4:6、3:7、2:8、1:9 和 3 个随机比例。连同玉米秸秆、猪粪、羊粪、牛粪样品 4 个,共计采集与制备样品 90 个。

1.2 碳氮比的测定

样品碳、氮含量的测定按照干烧法的原理^[24],采用 EURO EA3000 型元素分析仪测定,测试模式为碳氢氮模式,测试温度为 980℃,运行时间 320 s,样品杯为 5 mm×9 mm 锡囊,装样质量 1~3 mg,氦气为载气,反应管型号为 E13041。标样为琥乙红霉素标准品(C43H75NO16),其碳、氢、氮、氧质量分数分别为 59.911%、8.769%、1.625%、29.695%,与玉米秸秆碳氮比接近,适于测试玉米秸秆碳氢氮含量。每个样品测试 3 次,取 3 次的平均值作为待测样品的碳、氮含量值,然后通过计算可得样品的碳氮比。

1.3 光谱数据采集

对发酵原料样品使用 Bruker TANGO 型近红外光谱仪进行积分球漫反射光谱扫描,光谱采集范围 3 946~11 542 cm⁻¹(波长 866~2 534 nm),分辨率为 8 cm⁻¹,样品扫描 32 次,装样方式为 50 mm 样品杯旋转台扫描,装样质量约 7 g,将所测吸光度进行保存。在光谱扫描时,使用镀金样品杯上盖压实样品粉末,并采用 3 次扫描的平均值作为样品的原始

光谱。每个样品原始光谱的波长点数为 1 845 个。

1.4 波长优选方法

1.4.1 GSA 算法

GSA 算法^[25]融合了 GA^[26]的高效遗传操作和 SA^[27]的退温策略,采用二进制编码方案,以偏最小二乘(Partial least squares, PLS)回归模型的交叉验证均方根误差(Root mean squared error of cross-validation, RMSECV)为目标函数,通过结合温度参数设计适应度函数,基于 Metropolis 判别准则实现扰动解的选择复制,有效解决了 GA 算法存在的早熟收敛和进化后期搜索效率低的两点不足,又克服了 SA 算法进化速度慢的问题。

1.4.2 GSA-iPLS 算法

针对 GSA 以全谱波长点个数为码长进行二进制编码时,码长过长容易导致解空间发散的问题,将 GSA 与区间偏最小二乘法(Interval PLS, iPLS)^[28]相结合构建 GSA-iPLS 算法,通过选取多个有效的光谱子区间参与建模,能够有效提高模型的性能。

GSA-iPLS 基于 iPLS 的思想,将 NIRS 划分为 n 个等宽子区间,然后使用 GSA 优选出有效的特征谱区建模,以提高模型精度。GSA-iPLS 采用二进制编码方式,以子区间个数为码长,进行 GSA 的种群初始化。“1”和“0”分别表示对应子区间所包含波长点对应的数据“是”、“否”选中参与运算。根据种群初始化结果计算各染色体的目标函数值,确定初始温度和降温操作,并计算各染色体的适应度函数值。然后依据适应度函数值对种群中的染色体依次执行带最优保留策略的赌轮选择、离散重组交叉、离散变异和 Metropolis 选择复制操作,完成一轮次的 GSA 种群进化过程。经过多个轮次的种群进化,满足设定的算法终止条件后,即完成 NIRS 特征谱区优选。按如上方法,执行多次特征谱区优选算法,求出不同子区间个数下碳氮比对应的多个备选特征子区间组合,通过综合评测模型性能后,基于 RMSECV 确定发酵原料碳氮比对应的最佳子区间个数和最佳特征谱区。

1.4.3 DGSA-PLS 算法

GSA-iPLS 优选的光谱子区间内部仍然可能存在不相关的波长点和波长点之间的共线性问题。DGSA-PLS 算法利用 GSA 对 GSA-iPLS 优选的光谱子区间进行二次优化,能够有效去除谱区内的不相关波长点,解决波长点之间的共线性问题。

DGSA-PLS 以 GSA-iPLS 优选后特征谱区包含的特征波长点数为码长,进行二进制编码和种群初始化。“1”和“0”分别表示该波长点对应的数据“是”、“否”选中参与运算。在确定初始温度、降温

操作,计算适应度函数值后,执行多个轮次的 GSA 选择、交叉、变异和 Metropolis 选择复制进化操作,完成 NIRS 特征波长点的优选。针对 GSA 优化结果的随机性问题,多次执行特征波长点优选算法,并选择多次重复选中的波长点为特征波长变量建立 PLS 回归模型,能够得到较高的回归模型性能。

1.5 回归模型建立及评价

在进行 NIRS 特征波长优选前,先对光谱数据进行一阶导数预处理,再使用 Kennard - Stone (KS) 法将预处理后的光谱数据划分为 60 个校正集样本和 30 个验证集样本。然后,再使用 GSA - iPLS 和 DGSA - PLS 进行特征波长优选,并建立发酵原料碳氮比 PLS 定量分析模型,并采用校正决定系数 R_c^2 、验证决定系数 R_p^2 、校正均方根误差 (Root mean squared error of calibration, RMSEC)、预测均方根误差 (Root mean squared error of prediction, RMSEP) 和相对分析误差 (Residual predictive deviation, RPD) 来评价回归模型的优劣。决定系数 R^2 (R_c^2 和 R_p^2) 表示自变量对因变量的解释程度,展示了模型在数据变化时的测量能力,越接近 1 越好,但其值也受到待测目标属性分布范围的影响, R^2 只能作为参考指标。均方根误差 (Root mean squared error, RMSE),包括 RMSEC、RMSEP 和 RMSECV,作为评价 NIRS 模型回归性能的基础指标,体现了模型预测的精度和准确性, RMSE 越接近 0 越好。RPD 体现了模型的预测能力,越大越好。当 RPD 值大于 3 时,表明 NIRS 校正模型回归性能良好,能够满足实际应用需求^[29]。

本文算法 (包括光谱预处理、样品集划分、GSA - iPLS 算法、DGSA - PLS 算法及回归模型构建等) 全部在 Matlab R2012b 软件平台中实现。

2 结果与分析

2.1 采集数据分析

样品的光谱数据如图 1 所示。图 1a 中 NIRS 低波数频段吸收峰较强,波形比较尖锐,分辨能力较好。其中 $4\,000 \sim 5\,000\text{ cm}^{-1}$ (波长 $2\,000 \sim 2\,500\text{ nm}$) 区域对应着 C—C、C = O、—CH、—CH₂、—CH₃、—NH₂ 和 —COOH 基组合频, $5\,000 \sim 5\,500\text{ cm}^{-1}$ (波长 $1\,818 \sim 2\,000\text{ nm}$) 区域对应着 —NH₂ 和 —COOH 基团的一级倍频, $5\,500 \sim 6\,000\text{ cm}^{-1}$ (波长 $1\,667 \sim 1\,818\text{ nm}$) 区域对应着 —CH、—CH₂ 和 —CH₃ 基团的一级倍频, $6\,500 \sim 7\,500\text{ cm}^{-1}$ (波长 $1\,333 \sim 1\,538\text{ nm}$) 区域对应着 —CH、—CH₂、—CH₃、—NH 和 —NH₂ 基团的二级倍频, $8\,250 \sim 9\,000\text{ cm}^{-1}$ (波长 $1\,111 \sim 1\,212\text{ nm}$) 区域对应着 —CH、—CH₂ 和 —CH₃ 基团的二级倍频高频区域。由图 1b 可知,采用一阶

导数预处理能够有效消除基线漂移和背景噪声对光谱的干扰,增强了光谱数据的分辨率和信噪比,起到了提高建模效果的作用。

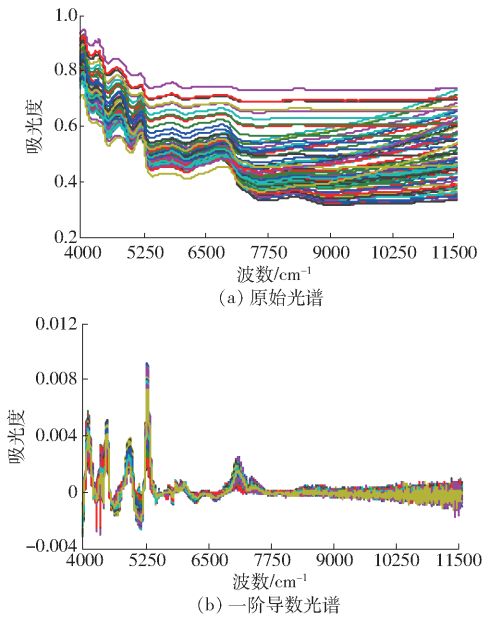


图 1 样品光谱数据

Fig. 1 Spectroscopic data of samples

对 90 个样品的原始光谱经一阶导数预处理后,使用 KS 法按 2:1 的比例进行样本划分,得到校正集样本 60 个、验证集样本 30 个,对应的碳氮比如表 2 所示。

表 2 厌氧发酵原料碳氮比

Tab. 2 C/N ratio of anaerobic fermentation materials				
样本	平均值	最大值	最小值	标准偏差
校正集	54.455	114.613	10.891	29.034
验证集	62.834	108.882	15.113	27.784

对预处理后的 NIRS 进行主成分 (Principal components, PCs) 分析,第 1、第 2 和第 3 主成分的贡献率分别为 71.136%、8.596% 和 5.327%,前 3 个 PCs 的累积贡献率达 85.059%。校正集和验证集的三维主成分空间分布情况如图 2 所示。在样本主成分空间分布图中,左侧为秸秆粪便混合物样本对应数据点,右侧为预处理秸秆样本对应数据点,产生如此清晰分类的结果与样品性状差异、原始光谱数据分布吻合。

由表 2 和图 2 可知,校正集样本碳氮比基本涵盖了验证集,且校正集和验证集样本在主成分空间上分布比较均匀,可以使用该样本划分方法进行 NIRS 分析。

2.2 特征波长优选

2.2.1 GSA - iPLS 特征谱区优选

GSA - iPLS 先按照 iPLS 将全谱划分成多个均匀的子区间,再以子区间个数为码长、以 RMSECV

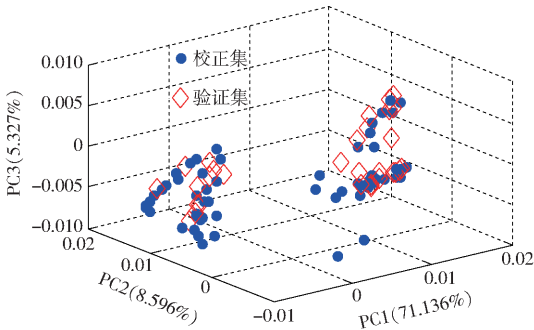


图 2 样本主成分空间分布

Fig. 2 Distribution of samples in PCs space

为目标函数运行 GSA 算法,优选特定子区间数下的特征谱区。为考察分割波长点个数对波长选择及模型预测性能的影响,分别按约 30、40、50、60、80、100、120 个波长点划分子区间,依次将预处理后的一阶导数光谱划分为 61、46、37、31、23、18、15 个子区间,依据 RMSECV 优选有效的子区间组合作为 GSA - iPLS 优选的特征谱区。为解决 GSA 优选结果的随机性问题,在每个子区间划分个数下,执行 10 次 GSA - iPLS 算法,并选定回归模型性能最佳的子区间组合作为该子区间数下的碳氮比特征谱区。在进行 GSA - iPLS 特征谱区优选时,种群规模设为 100,初温确定系数取 200,降温系数取 0.950,进化代数取 200,交叉概率取 0.700,变异概率取 0.010,邻域解扰动位数取码长的 1/10 向上取整。不同子区间数下优选的碳氮比特征谱区信息如表 3 所示。

表 3 GSA - iPLS 优选结果

Tab. 3 Results optimized by GSA - iPLS

子区间数	选中子区间数	波长点数	主成分数	RMSECV
61	25	754	14	15.300
46	14	560	6	15.293
37	13	650	10	15.119
31	12	714	6	15.653
23	8	641	8	14.963
18	13	1 332	10	17.670
15	7	860	10	17.641

由表 3 可知,采用子区间划分个数为 23,优选谱区的选中子区间数为 8,波长点数为 641 时,回归模型的性能最佳。GSA - iPLS 优选谱区如图 3 阴影部分所示。

基于各含氢基团在近红外谱区中的分布特性可知,在选中的 8 个子区间中,3 950 ~ 4 935 cm^{-1} (波长 2 026 ~ 2 534 nm) 对应着 C—C、—CH、—CH₂、—CH₃ 和 —NH₂ 基组合频,7 242 ~ 7 567 cm^{-1} (波长 1 322 ~ 1 381 nm)、7 901 ~ 8 226 cm^{-1} (波长 1 215 ~ 1 266 nm) 和 8 560 ~ 8 885 cm^{-1} (波长 1 125 ~ 1 168 nm) 对应着 —CH、—CH₂ 和 —CH₃ 基团的二级

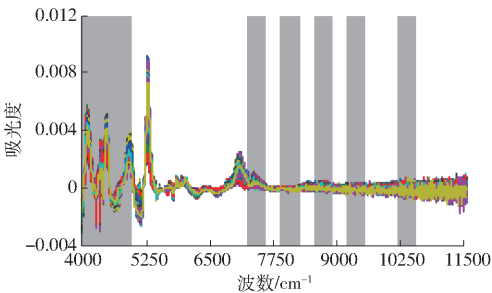


图 3 GSA - iPLS 优选谱区

Fig. 3 Spectral intervals selected by GSA - iPLS

倍频,9 219 ~ 9 544 cm^{-1} (波长 1 048 ~ 1 084 nm) 对应着 —CH、—NH₂ 基团的三级倍频。

当特征波长点在整个谱区中分布比较集中时,GSA - iPLS 谱区优选算法的性能优越,去除冗余波长点的效果较好。当特征波长点的分布比较分散时,GSA - iPLS 以子区间为单位进行特征谱区优选,在去除冗余波长点时会连带去除部分有效波长点,进而影响回归模型的性能。此时,需要增大子区间个数,减小子区间内波长点的数量,防止 GSA - iPLS 算法去除过多的有效波长点。但子区间数过多时,编码码长过长,影响 GSA 算法搜索效率的同时还可能导致解空间的发散问题。因此,在进行问题求解时,需要结合实际情况,设置合理的算法参数,实现算法运行效率和求解精度的统一。

2.2.2 DGSA - PLS 特征波长优选

DGSA - PLS 在进行特征波长点优选时,以 GSA - iPLS 优选的特征谱区波长点数为码长,随机生成 160 个码长为 641 的染色体构建初始种群,邻域解扰动位数取 20,其它算法初始参数与 GSA - iPLS 一致。为消除 GSA 算法的随机性,执行算法 50 次对碳氮比特征波长点进行优选。多次执行时,每次都选中的波长点代表了染色体的优良基因,以这些特征波长点作为特征波长变量建立回归模型时,可以有效消除 GSA 算法的随机性,且能够得到较高的回归模型性能。DGSA - PLS 波长优选结果与预处理后光谱的平均值对比如图 4 所示。

图 4 中,重复选中次数为 1 时,选中 628 个波长

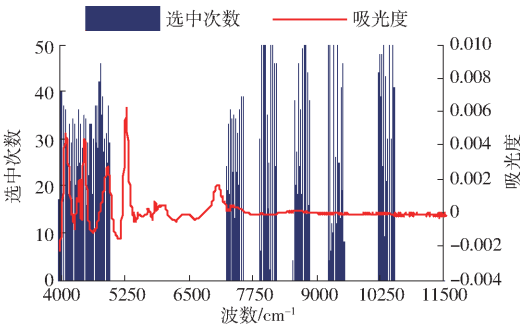


图 4 DGSA - PLS 优选波长变量

Fig. 4 Wavelength variables selected by DGSA - PLS

点;重复选中次数为 50 时,选中 19 个波长点。测试发现,校正集 RMSECV 和验证集 RMSEP 都随选中波长点数的增加呈先减小后增大的趋势,但两者的趋势存在较大差别。为了分析特征波长变量数目与模型性能的关系,绘制 RMSECV、RMSEP 与选中波长点数的关系图,如图 5 所示。

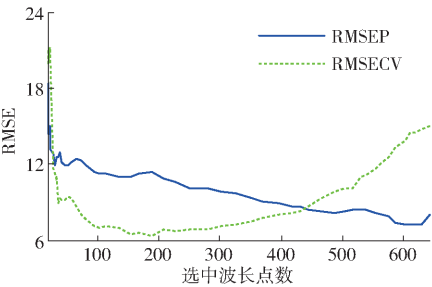


图 5 RMSE 与选中波长点数间的关系

Fig. 5 Relationship between RMSE and number of variables

由图 5 可知, RMSECV 最小值早于 RMSEP 出现,当选中波长点数为 189、重复选中次数为 27 次时, RMSECV 最小。 RMSEP 最小时,对应的波长点数为 628,重复选中次数为 1 次。 RMSECV 与 RMSEP 随选中波长点数变化趋势差异较大的主要原因在于 GSA 以校正集的 RMSECV 为依据进行特征波长优选,验证集性能拐点出现时表明校正集发生了过拟合。因此,选取 RMSEP 最小时对应的 628 个波长点作为 DGSA-PLS 优选的碳氮比特征波长变量。

2.3 回归模型评价与分析

为评测本文构建的两种波长优选算法的建模性能,以 GSA-iPLS 和 DGSA-PLS 优选后的特征波长变量作为 PLS 回归模型的输入,建立厌氧发酵原料碳氮比定量回归模型,并与全谱(Full-PLS)、协同区间偏最小二乘(Synergy iPLS, SiPLS)^[30]和反向区间偏最小二乘(Backward iPLS, BiPLS)^[31]优选特征波长的建模性能进行对比,结果如表 4 所示。

表 4 不同回归模型评价指标

参数	Full-PLS	SiPLS	BiPLS	GSA-iPLS	DGSA-PLS
波长点数	1 844	320	1 364	641	628
主成分数	8	6	10	8	8
R_c^2	0. 906	0. 810	0. 935	0. 906	0. 911
R_p^2	0. 876	0. 841	0. 904	0. 913	0. 920
RMSEC	8. 452	11. 499	7. 118	8. 447	8. 253
RMSEP	8. 532	9. 565	7. 701	7. 566	7. 178
RPD	3. 202	2. 856	3. 547	3. 611	3. 805

由表 4 可知, SiPLS 和 BiPLS 作为两种最典型 iPLS 算法, SiPLS 的建模性能弱于全谱建模,而

BiPLS 的建模性能优于全谱建模。主要原因在于: SiPLS 选取 2~4 个固定个数的子区间作为备选谱区,再通过比较 RMSECV 确定最佳谱区;而 BiPLS 通过剔除相关性较差的子区间,搜索 RMSECV 最小的子区间组合作为特征谱区; BiPLS 比 SiPLS 更适于特征波长点分布比较分散问题的求解。而碳氮比对应着谱区中所有含碳和含氮基团的吸收峰,这些吸收峰在整个谱区中分布较广,适合采用 BiPLS 进行特征谱区优选。 GSA-iPLS 作为一种新型近红外光谱特征谱区优选算法,具有良好的随机搜索能力。在使用 GSA-iPLS 算法进行特征谱区优选时,与 BiPLS 相比扩展了搜索结果的随机性。通过多次搜索并选取建模性能最佳的搜索结果作为 GSA-iPLS 优选特征谱区,该方式能够有效提高算法的特征波长优选性能。

基于 GSA-iPLS 和 DGSA-PLS 两种波长优选算法建立的碳氮比回归模型的 R_c^2 和 R_p^2 都大于 0. 900, RPD 都大于 3. 610,说明建模成功,且优于 Full-PLS 的建模性能和 SiPLS、BiPLS 优选特征波长的建模性能。 GSA-iPLS 优选的建模波长点个数为 641, DGSA-PLS 优选的建模波长点个数为 628, GSA 波长点二次优选只去掉了 GSA-iPLS 优选谱区中 0. 028% 的冗余波长点,原因在于碳氮两种元素相关基团在整个光谱频段内吸收峰分布过于分散。这与 DGSA-PLS 优选的波长点分布情况一致。在执行 GSA-iPLS 和 DGSA-PLS 算法时,在码长固定的情况下,可以通过适当增大种群规模、初温确定系数和遗传代数的方式提高 GSA 算法的寻优精度。采用 DGSA-PLS 作为碳氮比的波长优选方案,以优选后的特征波长建立 PLS 回归模型,并进行性能评测,其结果如图 6 所示。

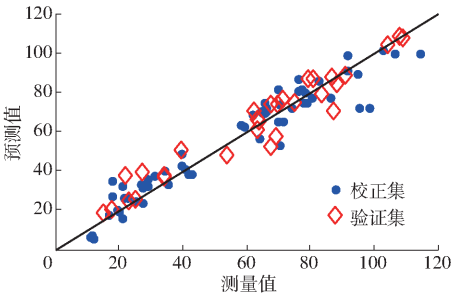


图 6 碳氮比实测值与预测值分布

Fig. 6 Distribution of measured and predicted values for C/N ratio

由图 6 可知,碳氮比的实测值与预测值点基本呈对角线分布。经 t 检验计算,得到样本碳氮比实测值与预测值的 t 值为 0. 329,小于相应的临界值 $t_{(0. 05, 89)} = 1. 987$,且 Matlab 内置的 t 检验函数的返回值检验结果为零、显著性水平概率为 0. 743,说

明碳氮比的预测值与实测值无显著性差异,一致性较好。但图中部分样品点距离对角线较远,说明部分样品的预测性能较差。存在部分预测性能较差样本的主要原因在于样品中含有 36 个秸秆粪便混合物样品,这 36 个样品在进行碳氮含量测定时,由于元素分析仪锡囊装料较少(1~3 mg),混合样品的均匀性将严重影响碳氮含量的测试准确性。而秸秆粉末密度小、粪便粉末密度大,很难保证混合时的均匀性,因此带来了较大的测量误差,进而影响所建 NIRS 回归模型的预测精度。DGSA-PLS 回归模型 R_p^2 为 0.920, RMSEP 为 7.178, RPD 为 3.805,与全谱建模相比有效波长点个数减少了 65.94%, RMSEP 减小了 15.87%,说明基于 DGSA-PLS 特征波长优选算法建立的碳氮比 PLS 回归模型性能得到显著提高,基本满足厌氧发酵过程中对原料碳氮比的直接快速检测需求。

3 结束语

探讨了采用 NIRS 技术结合化学计量学方法进行厌氧发酵原料碳氮比快速检测的可行性。为提高 NIRS 回归模型的检测精度和效率,基于 GSA 算法构建了 GSA-iPLS 和 DGSA-PLS 两种算法进行碳氮比特征波长的优选。GSA-iPLS 将光谱数据划分成多个子区间后,以子区间个数为码长,搜索有效的特征波长子区间组合作为特征谱区,有效减少了建模变量个数,提高了碳氮比检测模型的精度和效率。DGSA-PLS 在 GSA-iPLS 优选谱区的基础上,以波长点个数为码长进行特征波长变量优选,有效去除不相关的冗余波长点,得到 628 个特征波长点,建立的碳氮比检测模型 RMSEP 为 7.178, RPD 为 3.805。与全谱建模相比,基于 DGSA-PLS 建立的回归模型有效波长点个数减少了 65.94%, RMSEP 减小了 15.87%,有效地提高了模型的检测精度。

参 考 文 献

- [1] 葛一洪,邱凌, HASSANEIN A A M, 等. 马铃薯茎叶与玉米秸秆混合厌氧消化工艺参数优化[J/OL]. 农业机械学报, 2016, 47(4): 173-179.
GE Yihong, QIU Ling, HASSANEIN A A M, et al. Optimization of anaerobic digestion parameters with mixed material of potato stem leaf and corn straw[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2016, 47(4): 173-179. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20160423&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2016.04.023. (in Chinese)
- [2] LI H, TAN F, KE L T, et al. Mass balances and distributions of C, N, and P in the anaerobic digestion of different substrates and relationships between products and substrates[J]. Chemical Engineering Journal, 2016, 287: 329-336.
- [3] 罗立娜,丁清华,刘鑫,等. 基于 TRIZ 理论的秸秆好氧-厌氧联合发酵特性研究[J/OL]. 农业机械学报, 2018, 49(6): 332-339.
LUO Li'na, DING Qinghua, LIU Xin, et al. Characteristics of aerobic-anaerobic combined digestion process based on TRIZ theory[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(6): 332-339. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20180639&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2018.06.039. (in Chinese)
- [4] 艾平,田启欢,席江,等. 产酸沼渣再利用稻秸两级联合产酸工艺研究[J/OL]. 农业机械学报, 2019, 50(1): 292-298.
AI Ping, TIAN Qihuan, XI Jiang, et al. VFAs production technology by two-stage fermentation from rice straw based on digestate reuse[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(1): 292-298. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190132&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.01.032. (in Chinese)
- [5] JIN W Y, XU X C, YANG F L, et al. Performance enhancement by rumen cultures in anaerobic co-digestion of corn straw with pig manure[J]. Biomass & Bioenergy, 2018, 115: 120-129.
- [6] NING J, ZHOU M, PAN X, et al. Simultaneous biogas and biogas slurry production from co-digestion of pig manure and corn straw: performance optimization and microbial community shift[J]. Bioresource Technology, 2019, 282: 37-47.
- [7] 田潇瑜,黄星奕,白竣文,等. 基于近红外光谱技术的紫薯贮藏期间花青素含量检测[J/OL]. 农业机械学报, 2019, 50(2): 350-355.
TIAN Xiaoyu, HUANG Xingyi, BAI Junwen, et al. Detection of anthocyanin content of purple sweet potato during storage period based on near infrared spectroscopy[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(2): 350-355. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190240&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2019.02.040. (in Chinese)
- [8] ZHU Y, CHEN X, WANG S, et al. Simultaneous measurement of contents of liquiritin and glycyrrhizic acid in liquorice based on near infrared spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 196: 209-214.
- [9] 杨增玲,黄圆萍,沈广辉,等. 基于在线近红外光谱的堆肥全过程关键参数快速检测[J/OL]. 农业机械学报, 2019, 50(5): 356-361, 384.
YANG Zengling, HUANG Yuanping, SHEN Guanghui, et al. Rapidly detection of key parameters in whole composting process based on online near infrared spectroscopy[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(5): 356-361, 384. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20190540&journal_id=jcsam

- _id = jcsam. DOI:10.6041/j.issn.1000-1298.2019.05.040. (in Chinese)
- [10] 杨福芹, 冯海宽, 李振海, 等. 基于可见光-近红外光谱特征参数的苹果叶片氮含量预测[J/OL]. 农业机械学报, 2017, 48(9): 143–151.
YANG Fuqin, FENG Haikuan, LI Zhenhai, et al. Prediction for nitrogen content of apple leaves using spectral features parameters from visible and near infrared lights[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(9): 143–151. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20170918&journal_id=jcsam. DOI:10.6041/j.issn.1000-1298.2017.09.018. (in Chinese)
- [11] LUCE M S, ZIADI N, GAGNON B, et al. Prediction of total carbon, total nitrogen, and pH of organic materials using visible near-infrared reflectance spectroscopy[J]. Canadian Journal of Soil Science, 2018, 98(1): 175–179.
- [12] TAMBURINI E, CASTALDELLI G, FERRARI G, et al. Onsite and online FT-NIR spectroscopy for the estimation of total nitrogen and moisture content in poultry manure[J]. Environmental Technology, 2015, 36(18): 2285–2294.
- [13] SHETTY N, RINNAN Å, GISLUM R. Selection of representative calibration sample sets for near-infrared reflectance spectroscopy to predict nitrogen concentration in grasses[J]. Chemometrics and Intelligent Laboratory Systems, 2012, 111(1): 59–65.
- [14] YARCE C J, ROJAS G. Near infrared spectroscopy for the analysis of macro and micro nutrients in sugarcane leaves[J]. Zuckerindustrie Sugar Industry, 2012, 137(11): 707–710.
- [15] ROSSA U B, ANGELO A C, NISGOSKI S, et al. Application of the NIR method to determine nutrients in Yerba mate (*Ilex paraguariensis* A. St.-Hill) leaves[J]. Communications in Soil Science and Plant Analysis, 2015, 46(18): 2323–2331.
- [16] 李晓金, 朱凯, 牛智有, 等. 基于 PLS 算法的生物质秸秆元素分析 NIRS 快速检测[J]. 华中农业大学学报, 2015, 34(2): 131–135.
LI Xiaojin, ZHU Kai, NIU Zhiyou, et al. NIRS rapid detection of elemental analysis for straw biomass based on PLS algorithm[J]. Journal of Huazhong Agricultural University, 2015, 34(2): 131–135. (in Chinese)
- [17] YANG M, CHEN Q, KUTSANEDZIE F Y H, et al. Portable spectroscopy system determination of acid value in peanut oil based on variables selection algorithms[J]. Measurement, 2017, 103: 179–185.
- [18] YANG Y, LIU X, LI W, et al. Rapid measurement of epimedin A, epimedin B, epimedin C, icariin, and moisture in Herba Epimedii using near infrared spectroscopy[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017, 171: 351–360.
- [19] ZAREEF M, CHEN Q S, OUYANG Q, et al. Prediction of amino acids, caffeine, theaflavins and water extract in black tea using FT-NIR spectroscopy coupled chemometrics algorithms[J]. Analytical Methods, 2018, 10(25): 3023–3031.
- [20] LIANG H, CAO J, TU W J, et al. Nondestructive determination of the compressive strength of wood using near-infrared spectroscopy[J]. BioResources, 2016, 11(3): 7205–7213.
- [21] YANG Y, WANG L, WU Y J, et al. On-line monitoring of extraction process of Flos Lonicerae Japonicae using near infrared spectroscopy combined with synergy interval PLS and genetic algorithm[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2017, 182: 73–80.
- [22] LIU J M, LI N, ZHEN F, et al. Rapid detection of carbon-nitrogen ratio for anaerobic fermentation feedstocks using near-infrared spectroscopy combined with BiPLS and GSA[J]. Applied Optics, 2019, 58(18): 5090–5097.
- [23] SUN Y, QU J B, LI R L, et al. Optimization of the enzyme production conditions of bacillus licheniformis and its effect on the degradation of corn straw[J]. Journal of Biobased Materials and Bioenergy, 2018, 12(5): 432–440.
- [24] LUDWIG B, MURUGAN R, PARAMA V R R, et al. Use of different chemometric approaches for an estimation of soil properties at field scale with near infrared spectroscopy[J]. Journal of Plant Nutrition and Soil Science, 2018, 181(5): 704–713.
- [25] 刘金明, 初晓冬, 王智, 等. 玉米秸秆纤维素和半纤维素 NIRS 特征波长优选[J]. 光谱学与光谱分析, 2019, 39(3): 743–750.
LIU Jinming, CHU Xiaodong, WANG Zhi, et al. Optimization of characteristic wavelength variables of near infrared spectroscopy for detecting contents of cellulose and hemicellulose in corn stover[J]. Spectroscopy and Spectral Analysis, 2019, 39(3): 743–750. (in Chinese)
- [26] SONG J, LI G, YANG X. Optimizing GA-PLS model of soluble solids content in Fukumoto navel orange based on Vis-NIR transmittance spectroscopy using discrete wavelet transform[J]. Journal of the Science of Food and Agriculture, 2019, 99(11): 4898–4903.
- [27] YE D, SUN L, ZOU B, et al. Non-destructive prediction of protein content in wheat using NIRS[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 189: 463–472.
- [28] NØRGAARD L, SAUDLAND A, WAGNER J, et al. Interval partial least-squares regression (iPLS): a comparative chemometric study with an example from near-infrared spectroscopy[J]. Applied Spectroscopy, 2000, 54(3): 413–419.
- [29] STOCKL A, LICHTI F. Near-infrared spectroscopy (NIRS) for a real time monitoring of the biogas process[J]. Bioresource Technology, 2018, 247: 1249–1252.
- [30] JIANG H, LIU G, MEI C, et al. Measurement of process variables in solid-state fermentation of wheat straw using FT-NIR spectroscopy and synergy interval PLS algorithm[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2012, 97: 277–283.
- [31] LEARDI R, NØRGAARD L. Sequential application of backward interval partial least squares and genetic of relevant spectral regions[J]. Journal of Chemometrics, 2004, 18(11): 486–497.