

基于多模型加权预测的近红外定量分析方法

刘浩¹,李光尧¹,贾利红¹,王毅¹,张国宏¹

(1.四川长虹电器股份有限公司, 四川 绵阳 621000)

摘要: 为了提升便携式近红外光谱仪对未知样品的预测能力, 本文提出了一种基于多模型加权预测的近红外定量分析方法。首先利用便携式近红外光谱仪对酒醅样本光谱数据进行采集, 将采集后的数据进行划分, 分别构成按训练集、验证集、盲测集, 然后基于不同准则选取多个基础模型, 并基于模型相关系数值进行模型权重计算, 最终结合权重值进行多模型加权预测。对比传统单模型建模预测结果, 多模型加权预测方法对酒醅样本各成分预测的准确率平均提升了约 11%。结果表明, 基于多模型加权预测的近红外定量分析方法有效提升了模型预测准确率, 是一种可以在便携式近红外光谱仪数据建模中推广应用的实用方法。

关键词 便携式近红外光谱仪; 多模型; 权重系数; 酒醅

中图分类号: O657.3 **文献标识码:** A

1 引言

随着微机电技术的发展, 近几年, 便携式近红外光谱仪被广泛应用于化工^[1,2]、食品^[3,4]、饮料^[5]、医药^[6,7]等各个领域, 相较于传统的大型傅里叶变换光谱分析系统, 其具有结构简单、成本低廉、携带方便、结果实时可见等优势, 目前已成为光谱领域的热门产品^[8]。但是便携式近红外光谱仪易受光源、检测器、使用方法、环境条件等影响, 使得采集的光谱数据稳定性差, 精度低, 进而造成预测结果不稳定、预测准确率低等问题。

为了解决上述问题, 提升便携式近红外光谱分析结果的准确性, 提高便携式近红外光谱仪分析模型的预测稳定性, 目前行业内对近红外光谱数据建模优化方案, 主要集中于对单一模型的数据源筛选、预处理算法优化、模型筛选算法优化等工作, 此类建模过程的优化主要适用于高精度的傅里叶近红外光谱仪。对于自身硬件分辨率较低的便携式近红外光谱仪, 优化后模型质量提升不明显。因此本文在基于 PLS 算法建模的基础上, 提出多模型加权预测的方法, 以特定准则选取相对稳定、准确率较高的若干个光谱模型, 结合模型自身系数进行加权预测的方式来提升便携式近红外光谱仪整体性能, 进而提升便携式近红外光谱仪的预测稳定性及准确率。

2 多模型加权预测的近红外定量分析方法

2.1 样本集合划分

使用便携式近红外光谱仪采集多个待测样品光谱数据，采用 Kennard Stone (K-S) 算法对样本进行划分，将样本划分为训练集、验证集、盲测集，其中样本数量训练集包含 X 个、验证集包含 Y 个，盲测级包含 Z 个，X:Y:Z=2:1:1，避免人为划分样本的主观性。

2.2 光谱预处理

便携式近红外光谱仪采用常规的单一预处理方式效果不佳，本文采用双预处理嵌套的方式对样品进行处理，其中第一级预处理方式选择平滑，具体包含高斯平滑或者 SG 平滑，第二级预处理方式选择高斯求导、SG 求导、归一化、标准正态变量 (SNV)、去趋势及 log 中的一种。第一级、第二级预处理均可设置不同的预处理参数，通过不同预处理方式，预处理参数的设置可以获取多种预处理结果。

2.3 光谱建模

近红外光谱定量分析建模方法包括多元线性回归 (MLR)、主成分回归 (PCR)、偏最小二乘回归 (PLS)、人工神经网络 (ANN) 和支持向量机 (SVM) 等。其中，PLS 算法应用最为广泛，选用 PLS 算法进行建模。通过多种预处理结果、不同 PLS 主成分数选择组合建立多个光谱定量分析模型，若设定一级预处理为 M 种，二级预处理为 N 种，PLS 主成分数选择为 T 种，则通过不同排列组合可以建立合计 $M*N*T$ 个光谱定量分析模型。

2.4 光谱模型选择

基础模型一：在上述建立的大量光谱定量分析模型中，选择光谱模型最通用的两大表征系数，即模型相关系数 (R^2 值) 以及均方根误差 (RMSECV 值) 进行基础模型筛选，选择模型相关系数最大的模型为基础模型 A，模型均方根误差最小的模型为基础模型 B

基础模型二：在基础模型一中引入验证集，通过训练集建立的多个光谱模型对验证集进行预测，将预测值与验证集标定值进行计算，获取偏差值，选择偏差值最小的光谱模型为基础模型 C

基础模型三：在基础模型二中引入准确率，在近红外快检的实际应用中，对于预测偏差值在一定阈值范围内的样本定义为准确预测样本，若超出阈值则为预测错误样本，选择准确率最高的光谱模型为基础模型 D

2. 权重系数

光谱模型不同，对盲测样本的预测能力不同，结合光谱模型的模型相关系数 (R^2 值) 或者均方根误差 (RMSECV 值) 计算各个模型的预测权重，以模型相关系数为例：

$$T_i = R_i / (R_1 + R_2 + R_3 + R_4)$$

其中，R1 为基础模型 A 的模型相关系数，R2 为基础模型 B 的模型相关系数，依此类推。Ti 为各个基础模型对应的权重系数。

2.6 多模型加权预测

分别采用光谱基础模型 A、B、C、D 对盲测集 100 个样本进行预测，以盲测集单个样本为例，四个光谱模型对应获取四个预测值 a、b、c、d，结合权重系数计算最终单一预测值 S：

$$S=a*T1+b*T2+c*T3+d*T4$$

3 实验

3.1 实验设备

硬件设备为四川长虹研发的 PV800-III 便携式近红外光谱仪套件，图 1a 为便携式近红外光谱仪及配套移动终端，1b 为制备酒醅样品工装件，光谱仪波段范围为 1350nm-2150nm，采样间隔为 6nm，尺寸为Φ100mm×76.8mm，重量约 750g。



图 1 PV800R-I 便携式近红外光谱仪套件 (a. PV800R-I; b. 工装件)

3.2 实验样本

本次应用在常温条件下，将固液混合态的酒醅样品装入光谱仪的配套工装件中，填满压实备用。具体的采样过程为 PV800-III 通过蓝牙与智能移动终端连接，利用智能移动终端的客户端，控制近红外光谱仪对已经制备好的样品进行采样，然后将采集的数据通过无线网络传送到样品光谱数据库中以备分析使用。为了保证光谱数据采样的准确性及稳定性，每一个样本的原始光谱数据均通过光谱仪在工装件上均匀旋转采样 6 次，取平均所得。采集数量为 400 个酒醅样品，划分为训练集包含 200 个酒醅样品，验证集 100 个酒醅样品，盲测集 100

个酒醅样品，其中酒醅水分、淀粉、酸度等理化指标均由车间经验丰富化验员按常规化学方法测定所得。

3.3 实验结果

结合余梅^[9]等提出的优化光谱预处理算法及李良^[10]等在酒醅分析中提出的基于 PLS 的单模型建模算法对本实验酒醅数据进行建模，通过多种预处理组合优化，有效剔除了光谱数据中的大量无用信息，并结合 PLS 算法，将 300 维的光谱数据进行有效降维，提升了光谱数据的有效性 & 准确性。最后以 RMSECV 作为模型筛选指标，利用筛选的最优模型对盲测集 100 条未知样本进行模型外验证，图 2a~2c 依次给出了传统单模型水分、酸度、淀粉 3 个指标的模型外预测分布情况，图中横坐标为标定值，纵坐标为预测值，黄色区域为模型允许的误差范围（水分、淀粉允许误差为绝对偏差 ± 1 ，酸度允许误差范围为绝对偏差 ± 0.3 ）。

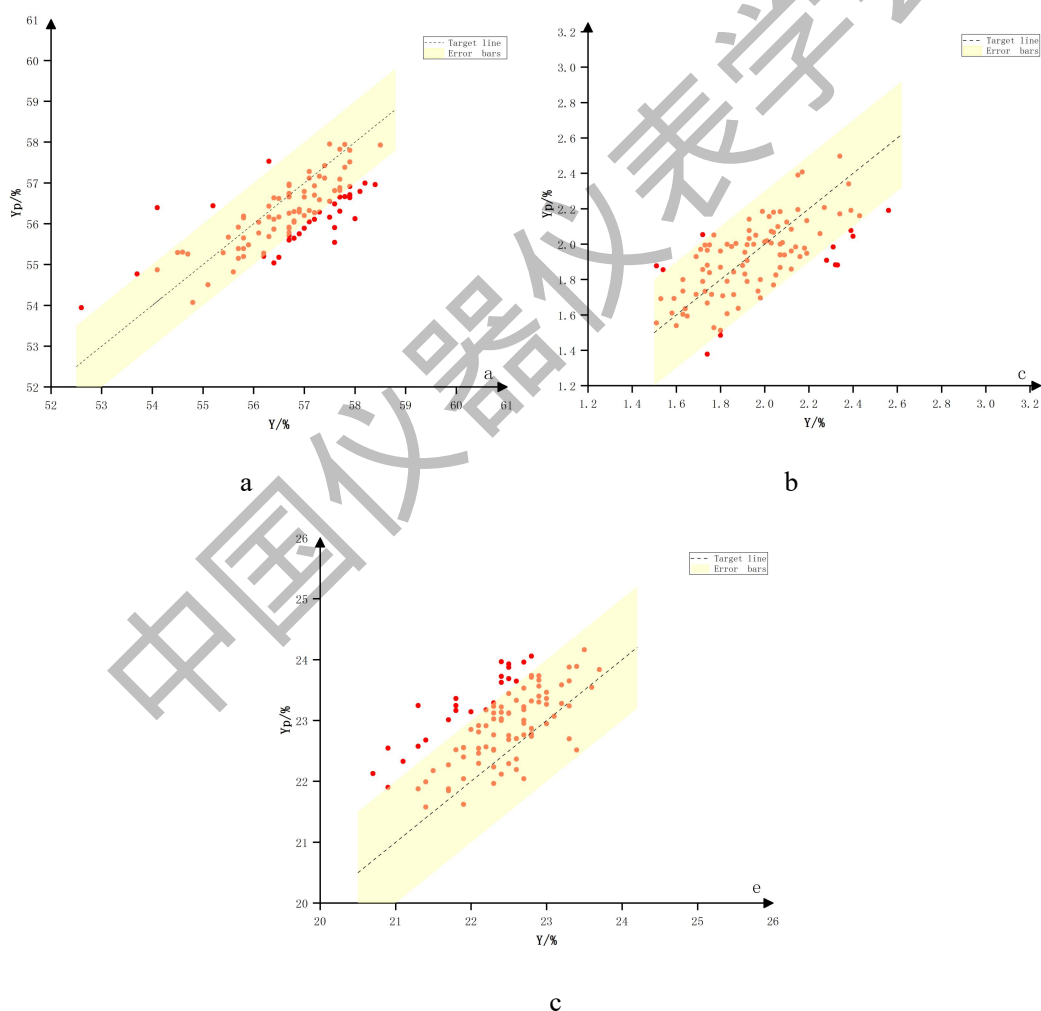


图 2 酒醅样品各指标传统单模型预测结果 (a.水分; b.酸度; c.淀粉)

采用基于多模型加权预测的近红外定量分析方法对上述酒醅光谱数据进行建模, 利用筛选的多个光谱模型对盲测集 100 条未知样本进行加权预测, 图 3a~3c 依次给出了多模型加权预测方法水分、酸度、淀粉 3 个指标的模型外预测分布情况。

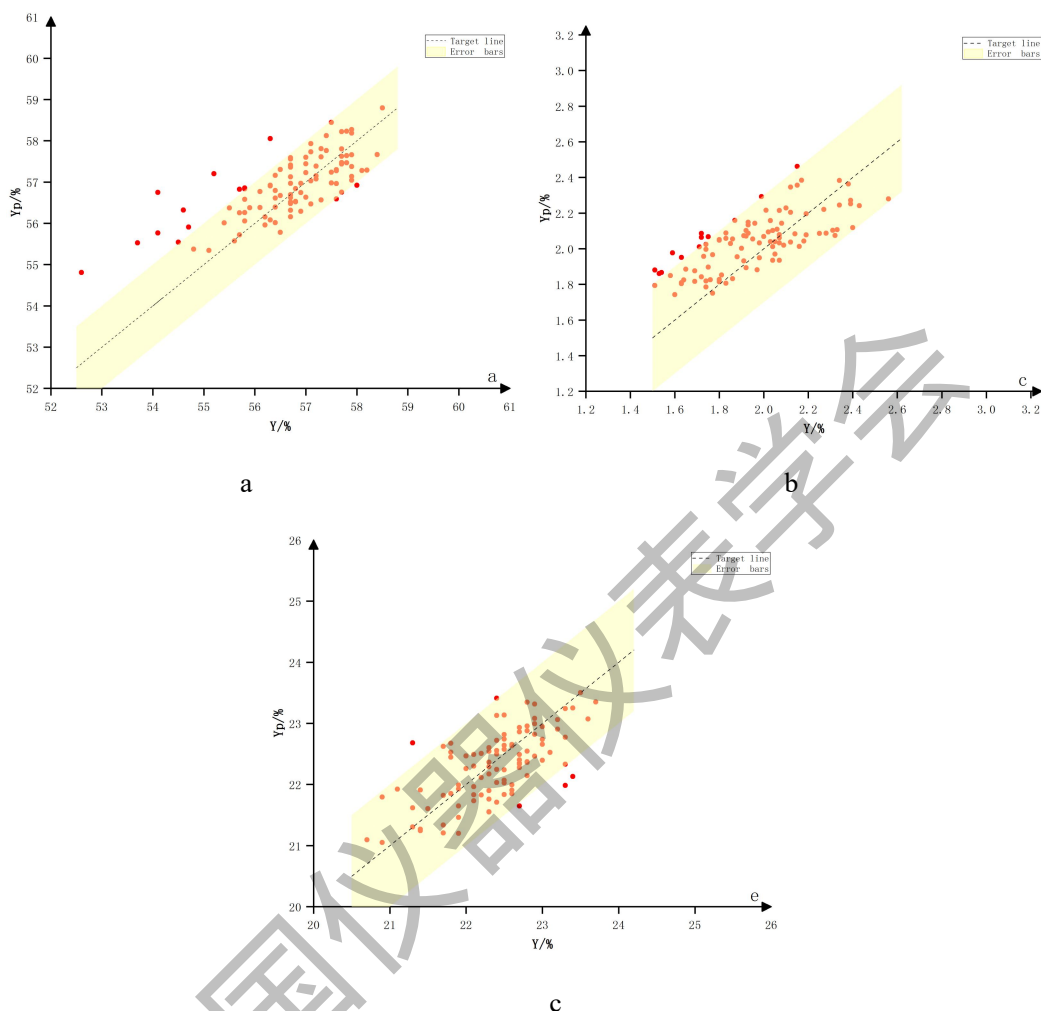


图 3 酒醅样品各指标多模型加权预测结果 (a.水分; b.酸度; c.淀粉)

统计上述单模型与多模型加权预测方法预测准确率如表 1 所示。

表 1 单模型与多模型预测准确率对比

	水分	酸度	淀粉
单模型准确率(%)	72	88	80
多模型准确率(%)	88	89	96

由上表知, 多模型加权预测准确率在酒醅水分、酸度、淀粉的预测准确率上均优于单模型预测准确率, 相较于单模型建模预测结果, 多模型加权预测方法将酒醅样本的水分预测准确率提升了 16%, 酸度预测准确率提升了 1%, 淀粉预测准确率提升了 16%, 各指标准准确率均达到了酿酒车间的应用要求。

4 结论

利用便携式近红外光谱仪,分别以传统单模型建模方法、多模型加权预测方法进行酿酒车间酒醅各成分光谱建模,并对100个未知样本进行模型外预测分析。结果表明,基于多模型加权预测的近红外定量分析方法,可以有效弥补便携式近红外光谱仪在采样精度、稳定性等方面的不足。相较于单模型建模预测结果,多模型加权预测方法将酒醅样本各成分预测准确率平均提升了约11%,甚至可有效逼近大型傅里叶光谱仪设备预测效果,是一种可以在便携式近红外光谱快检领域推广应用的实用方法。

参考文献

- [1] Francine D S, Layla P S, Pedro H P C et al. Discrimination of oils and fuels using a portable NIR spectrometer[J]. Fuel, 2021, 283
- [2] 吴琰. 基于 NIR 的制浆材材性快速检测方法研究[D]. 中国林业科学研究院, 2018.
- [3] Lorena C R S, Gabrielly S F, Layla P S et al. Quantification of beef, pork, and chicken in ground meat using a portable NIR spectrometer[J]. Vibrational Spectroscopy, 2020, 111 : 103158-.
- [4] 王允虎, 孙蕾, 王成忠, 孙锐. 便携式近红外光谱仪在鉴定无花果品质中建模效果研究[J]. 齐鲁工业大学学报, 2019, 33(04): 20-25.
- [5] 马伟, 董大伟, 杨建梅, 王晓慧, 陈文, 叶红, 毕小艳. 近红外光谱仪在酒曲检测中的应用[J]. 酿酒科技, 2017(02): 82-85.
- [6] Fan L H, Liu S Q, Fan W X, He L, Zhang C L, Wu C J, Huang Y L. Quality assessment of *Fritillariae cirrhosae* using portable NIR spectrometer[J]. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2022, 265 : 120325-120325.
- [7] 李庆, 闫晓剑, 黄永亮, 彭善贵, 赵小勤, 周晓英, 罗霄, 文永盛. 西红花吸光度测定方法改进及其品质快速评价[J]. 中国现代中药, 2021, 23(04): 655-659.
- [8] 褚小立, 陈瀑, 李敬岩, 刘丹, 许育鹏. 近红外光谱分析技术的最新进展与展望[J]. 分析测试学报, 2020, 39(10): 1181-1188.
- [9] 余梅, 李尚科, 杨菲, 郑郁, 李跑, 蒋立文, 刘霞. 基于近红外光谱技术与优化光谱预处理的陈皮产地鉴别研究[J]. 分析测试学报, 2021, 40(01): 65-71.
- [10] 李良, 董孝元, 吴昊, 张超, 周玉, 陈茂彬, 方尚玲, 谢逾群, 景艳艳, 张玉. 基于偏最小二乘法建立白酒酒醅近红外分析模型[J]. 酿酒, 2018, 45(03): 57-60.