

基于极限学习机自编码算法的 近红外光谱模型传递的研究

刘楠^{1,2}, 刘翠玲^{1,2*}, 徐金阳^{1,2}, 张善哲^{1,2}, 孙晓荣^{1,2}, 姜传智^{1,2}

(1. 北京工商大学人工智能学院, 北京 100048; 2. 北京工商大学食品安全
大数据技术北京市重点实验室, 北京 100048)

摘要: **目的** 针对食用油的酸值和过氧化值进行分析, 探究极限学习机自编码算法(transfer via extreme learning machine auto-encoder algorithm, TEAM)在近红外光谱上的模型传递。**方法** 使用 MATRIX-F 和 VERTEX-70 两种红外光谱仪采集食用油近红外光谱数据, 利用多元散射校正方法对光谱数据进行预处理。然后基于 TEAM 建立传递模型, 并与直接标准化、分段直接标准化和斜率偏差校正算法的建模效果进行了对比。**结果** 经 TEAM 算法模型传递后提高了模型的精确度, 食用油酸值模型中, 决定系数(R^2)从-1.3984 升高到 0.8553, 预测集均方根误差从 0.6130 mg/g 降低到 0.2578 mg/g, 食用油过氧化值模型中, R^2 从 0.6170 升高到 0.8987, 预测集均方根误差从 16.1530 mmol/kg 降低到 10.4150 mmol/kg。**结论** 极限学习机自编码算法使从机数据更好适应主机模型, 提高了模型的稳定性和准确性。

关键词: 近红外光谱技术; 极限学习机; 模型传递; 酸值; 过氧化值

Study of near infrared spectral model transfer based on an extreme learning machine auto-encoder algorithm

LIU Nan^{1,2}, LIU Cui-Ling^{1,2*}, XU Jin-Yang^{1,2}, ZHANG Shan-Zhe^{1,2},
SUN Xiao-Rong^{1,2}, JIANG Chuan-Zhi^{1,2}

(1. School of Artificial Intelligence, Beijing Technology and Business University, Beijing 100048, China; 2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China)

ABSTRACT: Objective To investigate the model transfer of the transfer via extreme learning machine auto-encoder algorithm (TEAM) on near infrared spectrum by analyzing the acid value and peroxide value of edible oil. **Methods** The near infrared spectral data of edible oils were collected by MATRIX-F and VERTEX-70 infrared spectrometers, and the spectral data were preprocessed by multiple scattering correction method. Then, a transfer model was established based on TEAM, and the modeling effects were compared with those of direct standardization, piecewise direct standardization and slope deviation correction algorithms. **Results** After TEAM algorithm model transmission, the accuracy of the model was improved. In the edible oil acid value model, coefficient of determination (R^2) increased from -1.3984 to 0.8553, the root mean square error of prediction set decreased from

基金项目: 国家重点研发计划项目(2017YFC1600605-01)、北京市自然科学基金项目(4132008)

Fund: Supported by the National Key Research and Development Program of China (2017YFC1600605-01), and the Beijing Natural Science Foundation of China (4132008)

*通信作者: 刘翠玲, 博士, 教授, 主要研究方向为智能测量技术与数据处理。E-mail: liucl@btbu.edu.cn

*Corresponding author: LIU Cui-Ling, Ph.D, Professor, Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University, Beijing 100048, China. E-mail: liucl@btbu.edu.cn

0.613 mg/g to 0.2578 mg/g, and in the edible oil peroxide value model, R^2 increased from 0.6170 to 0.8987. The root mean square error of prediction set decreased from 16.153 mmol/kg to 10.4150 mmol/kg. **Conclusion** The extreme learning machine self-coding algorithm makes the slave data better adapted to the host model and improves the stability and accuracy of the model.

KEY WORDS: near infrared spectroscopy; extreme learning machine; model transfer; acid value; peroxide value

0 引言

近红外(near infrared, NIR)光谱技术是一种新颖、简单、快速的新型分析技术, 可以实现对复杂事物的无损分析。近年来, 该技术被广泛应用于各个领域, 包括食品^[1-3]、农业^[4-5]、制药^[6]、生物医学^[7]、石化^[8]等。但在实际实验过程中, 会出现在不同仪器之间的校正模型不能完全通用的问题, 使得实验结果可能会产生较大的偏差。有效解决该类问题的方法是模型传递, 其本质为保持样品在不同仪器上的量测信号或光谱一致性^[9-10]。

近年来, 国内外研究人员对模型传递问题进行了大量的深入研究: 翁海勇等^[11]应用直接标准化(direct standardized, DS)算法对柑橘溃疡病进行模型传递, 提高了模型稳健性, 对建立稳健可靠的柑橘溃疡病判别模型具有重大意义; 田静等^[12]将分段直接标准化(piecewise direct standardization, PDS)算法应用到小麦粉粗蛋白含量的模型传递上, 降低了光谱之间的差异性, 有效的降低不同光谱仪器间建立校正模型的成本; FAN 等^[13]利用斜率偏差校正(slope/bias, S/B)算法针对可溶性固体含量来研究模型传递, 使生物变异对可溶性固体含量预测的干扰可以忽略不计。这些传统算法都可以实现模型传递, 但容易受到高斯噪声的影响, 且传递非线性光谱差异的能力较差^[14-16]。

极限学习机自编码法(transfer via extreme learning machine auto-encoder method, TEAM)以极限学习机自编码器(extreme learning machine auto-encoder, ELM-AE)为基础, 具有通用逼近的能力^[17], 具有学习速度快^[18]、容易实现和最小的人工干预等显著优势^[19-20]。与传统算法相比, 可以构造从光谱和主光谱之间的非线性依赖关系, 传递后的预测误差明显减小^[21]。所以 TEAM 算法可以用来校正主仪器和从仪器产生的信号之间的系统差异^[22]。但目前基本未见其在近红外光谱检测领域的应用。鉴于此, 本研究基于 TEAM 建立传递模型, 分别研究食用油酸值和过氧化值的校正模型在不同仪器之间传递, 并与传统 DS 算法、PDS 算法和 S/B 算法进行对比研究, 以期将模型传递更好地应用于近红外光谱检测领域。

1 材料与方法

1.1 材料

实验中用到的食用油样本采集于北京古船食品有限

公司, 其中玉米油 19 个, 芝麻香油 25 个, 橄榄油 9 个, 小磨香油 20 个, 大豆油 56 个, 合计 129 个样本。并按照 LS/T 6107—2012《动植物油脂 酸值和酸度测定 自动滴定分析法》与 GB/T 5009.227—2016《食品安全国家标准 食品中过氧化值的测定》测定所有食用油样本的酸值与过氧化值。

1.2 实验仪器

本研究在北京工商大学光谱技术与品质检测实验室完成, 光谱检测仪器是 VERTEX-70 傅里叶红外光谱仪和 MATRIX-F 傅里叶红外光谱仪(美国 Bruker 公司)。在实验中设定光谱仪 MATRIX-F 为主机, 光谱仪 VERTEX-70 为从机进行光谱采集。表 1 是所用仪器的详细参数。

表 1 光谱仪器和参数
Table 1 Spectroscopic instruments and parameters

仪器	主/从机类型	分辨率/cm ⁻¹	波长范围/cm ⁻¹
MATRIX-F	主机	2	12800~4000
VERTEX-70	从机	4	25000~20

1.3 数据划分与处理

1.3.1 样本划分方法

在进行模型传递分析之前, 首先将食用油样本集划分成两个子集, 分别是训练集和测试集。Kennard Stone (KS)^[23]算法是一种高效、应用广泛的转换样本集选取方法, 算法的核心是通过计算样品之间的欧氏距离(Euclidean distance)^[24]来选择训练集的样本, 剩下部分作为测试集, 从而提高校正模型的性能。

1.3.2 光谱数据预处理

为了防止人为或客观因素而导致实验数据含有噪声等影响时, 在模型建立之前需要对主机实验数据进行预处理。本研究采用的数据预处理方法是: 多元散射校正(multiplicative scatter correction, MSC)^[25], 其特点是: 认为每个样本任意波长点下的吸光度值和其平均光谱吸光度之间的关系是近似线性的^[26], 线性的截距和斜率是由光谱集线性回归获得的, 并用于校正每个光谱^[27]。

1.4 基于 TEAM 算法的模型传递

1.4.1 ELM-AE 算法原理

ELM-AE 可以看作是极限学习机的一种特例, 也由输入层、隐含层、输出层组成, 输入层与隐含层、隐含

层与输出层之间的神经元由全连接层连接^[28-29]。对于单隐层神经网络, ELM-AE 模型将输入数据作为输出数据, 并选择随机生成的隐藏节点的权值和偏差正交^[30], 不仅能减少训练时间, 还能获得全局最优解, 可以用于回归拟合问题^[31]。

1.4.2 基于 TEAM 算法的模型传递

基于 TEAM 算法的传递需要对隐节点进行多次随机初始化, 在同一个训练数据集上训练多个 ELM-AE 模型, 得到一系列光谱。通过对这些光谱进行平均, 最终可实现模型传递。其主要步骤为:

①样品分为两个部分: 训练集 C_m 和验证集 P_m 。然后, 在从机上采集相同的子集样本对应为 C_s 和 P_s 。

②隐藏节点从 300~500 中选取 50 个元素, 激活函数选择 $\tanh$, 在主从机的训练集 C_m 和 C_s 上训练多个 ELM-AE 模型, 在每个 ELM-AE 中, C_m 和 C_s 之间关系如公式(1):

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot C_s(j,:) + b_i) = C_m(j,:), j=1, \dots, M \quad (1)$$

其中, M 为样本个数, L 为隐层节点数, ω_i 为输入层与隐含层节点之间的权重, β_i 为输出层与隐含层节点之间的权重, $g(x)$ 为隐含层激活函数, b_i 为隐含层节点偏差。

③然后对 C_m 和 C_s 传递谱训练的每一个 ELM-AE 模型计算得到 $P_k^{\text{new}} (k=1, 2, \dots, K)$, K 是训练的 ELM-AE 模型数, 则计算可表示为公式(2)。

$$\sum_{i=1}^L \beta_i g(\omega_i \cdot P_s(j,:) + b_i) = P_k^{\text{new}}(j,:), j=1, \dots, M \quad (2)$$

④从一系列传递光谱中得到平均传递光谱 P_{average} , 其是独立测试集从机到主机传递的光谱, 计算如公式(3)。

$$P_{\text{average}} = \frac{\sum_{k=1}^K P_k^{\text{new}}}{m} \quad (3)$$

1.5 模型评价参数

评价模型的参数有相关决定系数 R^2 、交互验证均方根误差(root mean square error of leave-one-out cross validation, RMSECV)、校正集均方根误差(root mean square error of calibration set, RMSEC)和预测集均方根误差(root mean square error of prediction set, RMSEP)。 R^2 越大越接近 1, 且 RMSECV、RMSEC 和 RMSEP 的数值越小, 说明模型的预测性能越好, R^2 、RMSECV、RMSEC 和 RMSEP 计算公式如(4)~(7):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,\text{actual}} - y_{i,\text{predicted}})^2}{\sum_{i=1}^n (y_{i,\text{actual}} - \bar{y}_{i,\text{actual}})^2} \quad (4)$$

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (y_{i,\text{actual}} - y_{i,\text{predicted}})^2}{n}} \quad (5)$$

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (y_{i,\text{actual}} - y_{i,\text{predicted}})^2}{n-1}} \quad (6)$$

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (y_{i,\text{actual}} - y_{i,\text{predicted}})^2}{m-1}} \quad (7)$$

其中, $y_{i,\text{actual}}$ 是第 i 个样本参考方法的测定值, $\bar{y}_{i,\text{actual}}$ 是训练集或验证集所有样本参考方法测定值的平均值, $y_{i,\text{predicted}}$ 是验证集预测过程中第 i 样本的光谱方法预测值, m 、 n 为验证集、训练集的样本数。

2 结果与分析

2.1 食用油光谱特征分析

利用两台光谱检测仪分别采集食用油样本的近红外光谱。两台仪器的分辨率设置为 16 cm^{-1} , 样本重复扫描 32 次, 背景重复扫描 32 次, 采集的光谱范围是 $12000 \sim 4000 \text{ cm}^{-1}$, 光阑设置为 6 mm , 扫描速度设为 10 kHz 。本次研究主要利用的数据是食用油样本在 $9000 \sim 5000 \text{ cm}^{-1}$ 的近红外光谱取平均得到的平均光谱, 共 1039 个波数点。

图 1 是主、从机的食用油平均吸光度及光谱差值图, 可以看出, 在 8700 和 7200 cm^{-1} 附近有两处明显的吸收峰, 可能是由烃类和脂肪烃类物质中 $-\text{HC}=\text{CH}-$ 和 $-\text{CH}_2-$ 基团的 C-H 键二级倍频伸缩震动产生^[32-33]。在主从机的吸光度差值图中可以说明, 主机从机检测的食用油吸光度是存在差异的, 且在 $9000 \sim 8500 \text{ cm}^{-1}$ 范围内吸光度差值最明显。

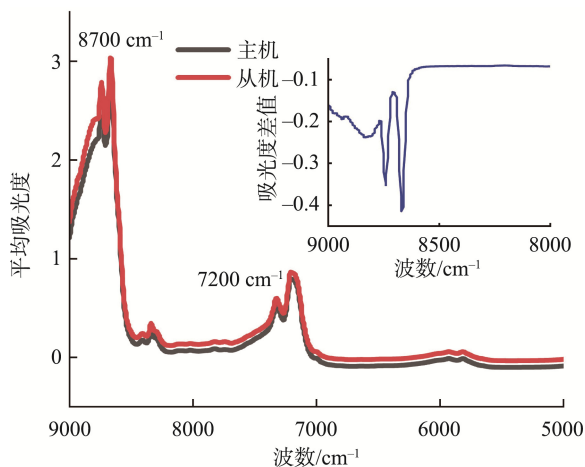


图 1 主、从机的食用油平均吸光度及光谱差值图

Fig.1 Mean absorbance and spectral difference map of edible oil from master and slave

2.2 样本划分

共 129 个食用油样本数据, 采用 K/S 算法按 3:1 把样本分为两个集合训练集与验证集。所以训练集共计 97 个样本, 验证集共计 32 个样本。表 2 是食用油样本酸值和过氧

化值数据集的具体划分情况,可以看出预测集的数据范围都在训练集范围内,说明预测集具备对模型性能进行验证的能力。

2.3 模型主因子数分析

在建立校正模型之前需要确定校正集的主因子数(principal factor number, PCs)。过低的校正因子数,可能会导致建模校正集的信息损失过大,使模型的预测能力降低,但是校正因子数过多,又会引入过多的冗余信息,造成模型的过拟合现象,同时使数据处理的时间加长。提取有效的主成分,能显著降低数据维数,剔除冗余信息,而又不丢失有效信息。所以合理选取主成分因子数对建立模型来说是非常重要的。

本研究按 K 折交叉验证的方法,利用样本的 RMSECV 来评价并确定主因子数,采用 5 折校验后,图 2 是食用油酸值与过氧化值 RMSECV 与主因子数的关系, RMSECV 先随着主因子数的增大呈下降趋势,图 2a 中酸值的主因子数为 8 时 RMSECV 达到最低点,图 2b 中过氧化值的主因子数为 5 时 RMSECV 达到最低点,然后随着主因子数的增大,曲线呈上升趋势,说明最低点以后,增加的主因子数是与被测油品无关的噪声,模型出现过拟合的现象。因此确认酸值的最佳主因子数为 8,过氧化值的最

佳主因子数为 5。

2.4 模型预测结果

2.4.1 主机模型预测结果

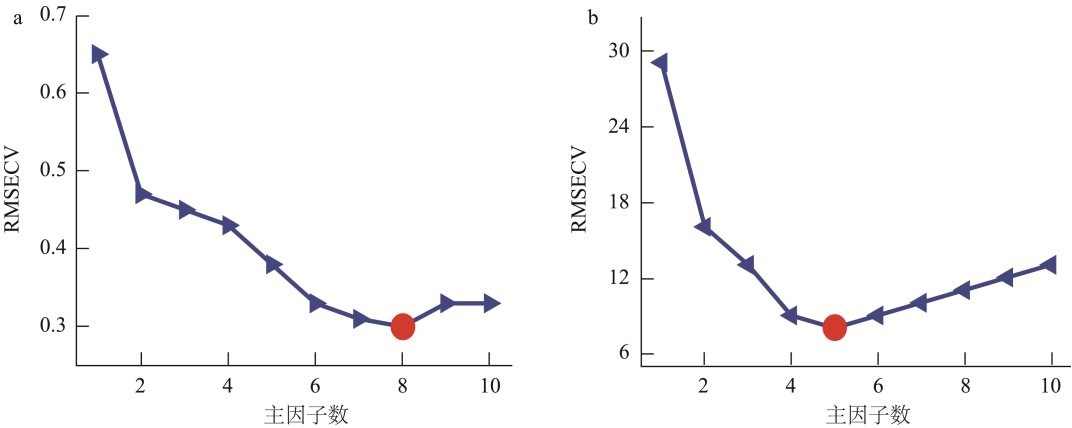
经过 K 折交叉验证的方法确定主因子数后,建立 ELM-AE 校正模型分别预测主从机的 32 个预测集样本的酸值和过氧化值。利用主机模型预测主机样本集时,酸值的预测值与真实值的决定系数 R^2 为 0.7843, RMSEP 为 0.2490 mg/g, 过氧化值的 R^2 为 0.7528, RMSEP 为 7.7490 mmol/kg, 最终的预测模型较好。但是当从机的样本集没有进行模型传递而直接代入主机模型进行预测的时候,酸值的 R^2 下降到 -1.3984, RMSEP 上升到 0.6130 mg/g, 过氧化值的 R^2 下降到 0.6170, RMSEP 上升到 16.1530 mmol/kg, 预测结果产生巨大偏差,说明主机上建立的模型在从机上无法使用,需要对从机样本进行模型传递来改善结果。

2.4.2 TEAM 算法模型预测结果

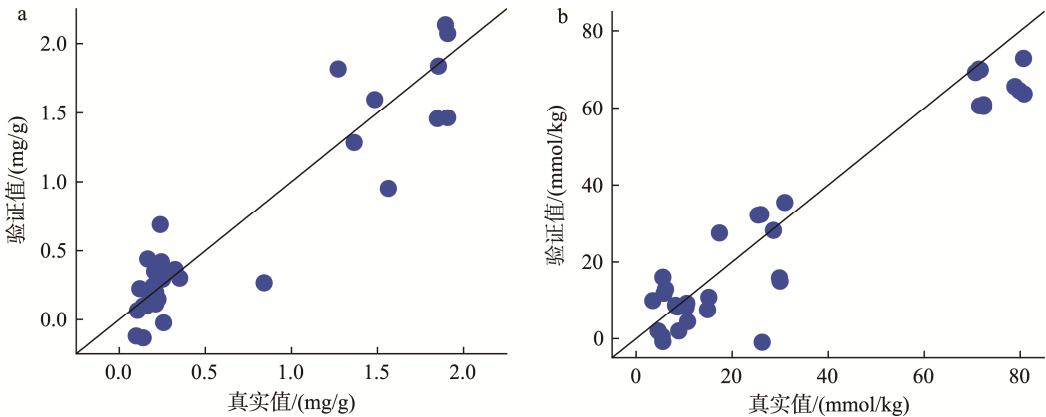
由于主机模型不能直接应用于从机的样本,所以将从机的 32 个食用油样本带入 TEAM 算法模型进行预测,图 3 是酸值和过氧化值经传递后的模型真实值与预测值散点图,可以看出真实值与验证值均匀的分布在 diagonal 两侧,预测值与真实值相差不大,说明预测相对准确。

表 2 食用油样品的数据集划分
Table 2 Data sets partitioning of edible oil samples

数据集	数量	理化值	数值范围	平均值	标准偏差
训练集	97	酸值/(mg/g)	0.07~2.13	0.59	0.63
		过氧化值/(mmol/kg)	0.34~93.01	27.78	26.89
验证集	32	酸值/(mg/g)	0.08~1.91	0.81	0.79
		过氧化值/(mmol/kg)	0.56~91.83	33.62	31.92



注: a 为食用油酸值 RMSECV 与主因子数关系; b 为食用油过氧化值 RMSECV 与主因子数关系。
图 2 食用油酸值与过氧化值 RMSECV 与 PCs 关系
Fig.2 Relationship between RMSECV and PCs by the edible oleic acid values and peroxide values



注: a 为 TEAM 算法预测的食用油酸值散点图; b 为 TEAM 算法预测的食用油过氧化值散点图。

图 3 酸值和过氧化值模型传递后的模型真实值与预测值散点图

Fig.3 Scatter plots of the true values and predicted values of the model after the transmission of the acid value and peroxide value model

表 3 是食用油酸值与过氧化值建模及预测结果,从表 3 中可以看出,通过 TEAM 模型传递算法进行传递后,酸值的 R^2 上升到 0.8553, RMSEP 下降到 0.2578 mg/g, 过氧化值的 R^2 上升到 0.8987, RMSEP 下降到 10.4150 mmol/kg, 模型预测效果有明显提升,说明 TEAM 算法使得从机样本能更好的适用于主机模型。

表 3 食用油酸值与过氧化值建模及预测结果

Table 3 Modeling and prediction results of edible oleic acid value and peroxide value

模型类型	酸值		过氧化值	
	RMSEP/(mg/g)	R^2	RMSEP/(mmol/kg)	R^2
未传递模型	0.6130	-1.3984	16.1530	0.6170
TEAM	0.2578	0.8553	10.4150	0.8987
主机模型	0.2490	0.7843	7.7490	0.7528

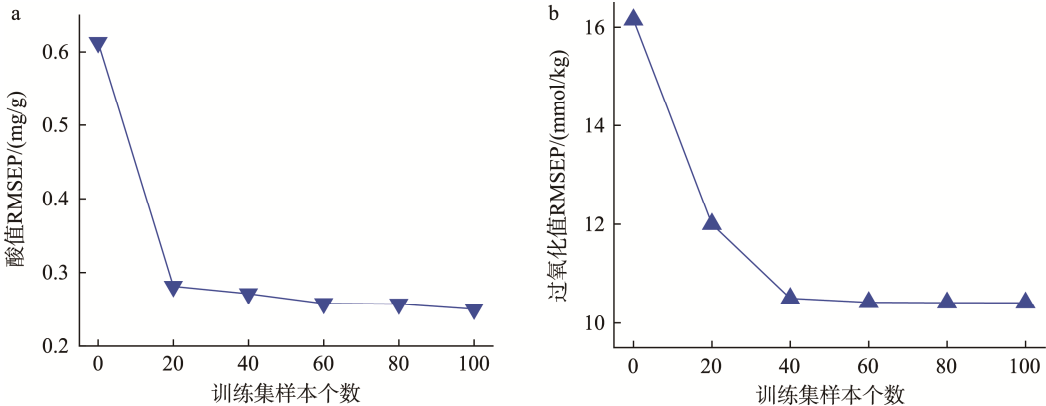
2.5 标准化样品数量对模型传递的影响

为了深入探索如何能够获得更好的传递效果,采用 K/S 算法从同批次的食用油样本训练集中选取 0、20、40、...、100 个样品作为标准化样本集,图 4 所示为不同标准化样本数

量对食用油酸值与过氧化值建模效果的影响。样本数量为 0 时,是从机样品直接应用在主机模型的预测结果。从图 4a 可知,食用油酸值经 TEAM 算法传递后的 RMSEP 随着主机样品数的增加在逐渐下降,直到稳定后无明显变化。图 4b 中,食用油过氧化值经 TEAM 算法传递后的 RMSEP 与酸值呈相同规律,随着主机样品数量的增加, RMSEP 逐渐减小直到稳定。说明 TEAM 算法受标准样品集数量的影响较小,且在光谱数量较少的情况下能够找到更精确的从机与主机之间的传递模型,减少两种仪器之间系统差异的影响。

2.6 TEAM 算法与其他算法的比较

为了评估 TEAM 算法的模型传递效果,研究对 DS 算法、PDS 算法和 S/B 算法进行比较。其中 DS、PDS 算法是利用标准样本集对从机光谱进行校正,使得校正后的从机光谱与主机光谱相似度达到最大,进而利用主机模型对其进行预测。而 S/B 算法是一种目标值校正的方法,通过构造主机与从机目标值之间的一元线性关系对未知的样本从机目标值进行校正。



注: a 为酸值在不同标准化样本数目的预测均方根误差; b 为过氧化值在不同标准化样本数目的预测均方根误差。

图 4 不同标准化样本数目的 RMSEP

Fig.4 RMSEP with different number of standardization samples

如表 4 所示, 对比 4 种算法, 把从机的 32 个样本集分别带入各算法转换后的主机模型, 均有效地降低了酸值和过氧化值 RMSEP, 提高了从机样本在主机模型中的适应能力。其中 TEAM 算法的预测能力效果最好。

表 4 不同算法下的模型 RMSEP
Table 4 RMSEP results by different calibration transfer methods

RMSEP	DS 算法	PDS 算法	S/B 算法	TEAM 算法
酸值 RMSEP /(mg/g)	0.3150	0.5590	0.4020	0.2578
过氧化值 RMSEP /(mmol/kg)	11.9150	12.1060	14.6650	10.4150

3 结 论

本研究在 ELM 的自动编码器的基础上开发了 TEAM, 并将其应用于近红外光谱的模型传递。通过对食用油光谱数据集进行模型传递的结果表明, 经 TEAM 算法传递后的从机样本酸值预测模型的 R^2 从-1.3984 升高到 0.8553, RMSEP 从 0.6130 mg/g 降低到 0.2578 mg/g, 过氧化值的预测模型的 R^2 从 0.6170 升高到 0.8987, RMSEP 从 16.1530 mmol/kg 降低到 10.4150 mmol/kg。对比其他 3 种传统算法, TEAM 算法可以有效地校正不同仪器间测量光谱的系统误差, 提高食用油酸值与过氧化值模型的预测结果。在大多数情况下, TEAM 算法也可以通过少量的校准集来建立传递关系, 从而获得最佳的预测集均方根误差, 这对于近红外光谱技术应用用于食用油检测具有重要意义。

但是, TEAM 算法还存在一些问题, 使得模型的验证结果与理想结果还存在差距。一方面, 可能由于样品数量以及种类仍然不够丰富, 使得建模结果缺乏足够的可靠性。另一方面, TEAM 算法可能适用于不同的样品种类, 不同的实验样本可能导致实验的适应度不同。综上, 下一步将从增加食用油样本的种类和数量, 以及尝试不同实验样品种类进行进一步的实验研究。

参考文献

[1] 马雪亭, 罗华平, 高峰, 等. 近红外光谱技术在苹果检测方面的研究与应用[J]. 食品安全质量检测学报, 2022, 13(13): 4219–4227.
MA XT, LUO HP, GAO F, *et al.* Research and application of near infrared spectroscopy in apple detection [J]. J Food Saf Qual, 2022, 13(13): 4219–4227.

[2] 苏丹, 王志霞, 周佳, 等. 基于知识图谱分析近红外光谱技术在茶叶分析中的研究进展[J]. 食品安全质量检测学报, 2022, 13(4): 1193–1200.
SU D, WANG ZX, ZHOU J, *et al.* Research progress of near infrared spectroscopy in tea analysis based on knowledge mapping analysis [J]. J Food Saf Qual, 2022, 13(4): 1193–1200.

[3] SHU XF, JIANG BL, YU X, *et al.* Long-term evaluation of soluble solids content of apples with biological variability by using near-infrared

spectroscopy and calibration transfer method [J]. Postharvest Biol Technol, 2019, 151(5): 79–87.

[4] 王懂, 杨玮, 曹永研, 等. 基于近红外光谱信息的土壤电导率预测模型研究[J]. 农业机械学报, 2022, 53(S1): 218–223.
WANG D, YANG W, CAO YY, *et al.* Research on soil conductivity prediction model based on near infrared spectral information [J]. Agric Mach, 2022, 53(S1): 218–223.

[5] 钟翔君, 杨丽, 张东兴, 等. 不同粒径对土壤有机质含量可见—近红外光谱预测的影响[J]. 光谱学与光谱分析, 2022, 42(8): 2542–2550.
ZHONG XJ, YANG L, ZHANG DX, *et al.* Effect of different particle sizes on the prediction of soil organic matter content by visible-near infrared spectroscopy [J]. Spectrosc Spectr Anal, 2022, 42(8): 2542–2550.

[6] ZHENG AB, YANG HH, PAN XP, *et al.* Identifying multi-class drugs by using near-infrared spectroscopy and variational auto-encoding modeling [J]. Spectrosc Spectr Anal, 2020, 40(12): 3946–3952.

[7] 余代鑫, 郭盛, 张霞, 等. 近红外光谱技术结合化学计量学和机器学习算法的干姜产地溯源研究[J]. 中国中药杂志, 2022, 47(17): 4583–4592.
YU DX, GUO S, ZHANG X, *et al.* Origin traceability of dried ginger by near infrared spectroscopy combined with chemometrics and machine learning algorithms [J]. China Chin Mater Med, 2022, 47(17): 4583–4592.

[8] 李泉伦, 陈争光, 孙先达. 近红外光谱的油页岩总有机碳快速检测[J]. 光谱学与光谱分析, 2022, 42(6): 1691–1697.
LI QL, CHEN ZG, SUN XD. Rapid detection of total organic carbon in oil shale by near infrared spectroscopy [J]. Spectrosc Spectr Anal, 2022, 42(6): 1691–1697.

[9] KENRY, DUAN YK, LIU B. Recent advances of optical imaging in the second near-infrared window [J]. Adv Mater, 2018, 30(47): 1802394.

[10] 褚小立, 陈瀑, 李敬岩, 等. 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.
CHU XL, CHEN P, LI JY, *et al.* The latest progress and prospect of near-infrared spectroscopy [J]. J Instr Anal, 2020, 39(10): 1181–1188.

[11] 翁海勇, 何勇. 直接校正算法的柑橘溃疡病高光谱模型传递[J]. 光谱学与光谱分析, 2018, 38(1): 235–239.
WENG HY, HE Y. Direct correction algorithm for hyperspectral model transmission of citrus canker [J]. Spectrosc Spectr Anal, 2018, 38(1): 235–239.

[12] 田静, 陈斌, 陆道礼, 等. 不同分光原理近红外光谱仪光谱标准化方法在小麦粉品质检测中的应用[J]. 中国食品学报, 2022, 22(10): 286–294.
TIAN J, CHEN B, LU DL, *et al.* Application of spectral standardization methods of different spectroscopic principles of near infrared spectroscopy in wheat flour quality testing [J]. Chin Food Sci Technol, 2022, 22(10): 286–294.

[13] FAN S, LI J, XIA Y, *et al.* Long-term evaluation of soluble solids content of apples with biological variability by using near-infrared spectroscopy and calibration transfer method [J]. Postharvest Biol Technol, 2019, 151: 79–87.

[14] WORKMAN JJ. A review of calibration transfer practices and instrument differences in spectroscopy [J]. Appl Spectrosc, 2018, 72(3): 340–365.

[15] ZHANG J, LI BY, HU Y, *et al.* A parameter-free framework for calibration enhancement of near-infrared spectroscopy based on correlation constraint [J].

- Anal Chim Acta, 2020, 1142(1): 169–178.
- [16] 许亮, 张紫叶, 陈曦, 等. 基于改进麻雀搜索算法优化 BP 神经网络的气动光学成像偏移预测[J]. 光电子与激光, 2021, 32(6): 653–658.
- XU L, ZHANG ZY, CHEN X, *et al.* Optimized BP neural network based on improved sparrow search algorithm for prediction of pneumatic optical imaging offset [J]. Optoelectron Laser, 2021, 32(6): 653–658.
- [17] CHEN YY, WANG ZB. Cross components calibration transfer of NIR spectroscopy model through PCA and weighted ELM based TrAdaBoost algorithm [J]. Chemometr Intell Lab, 2019, 192(15): 103824–103843.
- [18] LIU T, XU TY, YU FH. A method combining ELM and PLSR (ELM-P) for estimating chlorophyll content in rice with feature bands extracted by an improved ant colony optimization algorithm [J]. Comput Electron Agric, 2021, 186(2): 106–110.
- [19] YE A, ZHOU X, MIAO F. Innovative hyperspectral image classification approach using optimized CNN and ELM [J]. Electronics-Switz, 2022, 11(5): 775–778.
- [20] 崔芳, 李雪, 王督, 等. 近红外光谱快速检测食用植物油中酸价[J]. 食品安全质量检测学报, 2022, 13(8): 2589–2593.
- CUI F, LI X, WANG D, *et al.* Rapid detection of acid value in edible vegetable oil by near infrared spectroscopy [J]. J Food Saf Qual, 2022, 13(8): 2589–2593.
- [21] 周孟然, 王锦国, 宋红萍, 等. 核极限学习机和激光诱导荧光技术在食用油识别中的应用[J]. 激光与光电子学进展, 2020, 57(20): 203001–203005.
- ZHOU MR, WANG JG, SONG HP, *et al.* Application of nuclear limit learning machine and laser induced fluorescence technique in edible oil identification [J]. Laser Optoelectron Prog, 2020, 57(20): 203001–203005.
- [22] CHEN W, BIN J, LU H, *et al.* Calibration transfer via extreme learning machine auto-encoder [J]. Analys, 2016, 141(6): 1973–1980.
- [23] 李华, 王菊香, 邢志娜, 等. 改进的 K/S 算法对近红外光谱模型传递影响的研究[J]. 光谱学与光谱分析, 2014, 31(2): 362–365.
- LI H, WANG JX, XING ZN, *et al.* Study of the effect of improved K/S algorithm on the model transfer of NIR spectra [J]. Spectrosc Spectr Anal, 2014, 31(2): 362–365.
- [24] 吴静珠, 张乐, 李江波, 等. 基于高光谱与集成学习的单粒玉米种子水分检测模型[J]. 农业机械学报, 2022, 53(5): 302–308.
- WU JZ, ZHANG L, LI JB, *et al.* Hyperspectral and integrated learning based moisture detection model for single seed maize [J]. Agric Mach, 2022, 53(5): 302–308.
- [25] 温晓燕, 徐艳艳, 李茂刚, 等. 分段直接校正模型转移结合近红外光谱技术的甲醇汽油中甲醇定量分析方法研究[J]. 分析化学, 2021, 49(10): 1758–1765.
- WEN XY, XU YY, LI MG, *et al.* Quantitative analysis of methanol in methanol gasoline by fractional direct correction model transfer combined with near infrared spectroscopy [J]. Anal Chem, 2021, 49(10): 1758–1765.
- [26] 孙代青, 谢丽蓉, 周延, 等. 基于近红外光谱的 SG-MSC-MC-UVE-PLS 算法在全血血红蛋白浓度检测中的应用[J]. 光谱学与光谱分析, 2021, 41(9): 2754–2758.
- SUN DQ, XIE LR, ZHOU Y, *et al.* Application of SG-MSC-MC-UVE-PLS algorithm based on near-infrared spectroscopy in the detection of whole blood hemoglobin concentration [J]. Spectrosc Spectr Anal, 2021, 41(9): 2754–2758.
- [27] WU YF, PENG SL, XIE Q, *et al.* An improved weighted multiplicative scatter correction algorithm with the use of variable selection: Application to near-infrared spectra [J]. Chemometr Intell Lab, 2019, 185(1): 114–121.
- [28] 徐睿, 梁循, 齐金山, 等. 极限学习机前沿进展与趋势[J]. 计算机学报, 2019, 42(7): 1640–1670.
- XU R, LIANG X, QI JS, *et al.* Cutting-edge advances and trends in extreme learning machines [J]. Chin Computers, 2019, 42(7): 1640–670.
- [29] 饶利波, 庞涛, 纪然仕, 等. 基于高光谱成像技术结合堆栈自动编码器-极限学习机方法的苹果硬度检测[J]. 激光与光电子学进展, 2019, 56(11): 113001–113005.
- RAO LB, PANG T, JI RS, *et al.* Apple hardness detection based on hyperspectral imaging combined with stacked autoencoder-extreme learning machine method [J]. Laser Optoelectron Prog, 2019, 56(11): 113001–113005.
- [30] WANG LJ, DING SF. A spectral clustering algorithm based on ELM-AE feature representation [J]. Caa T Intel Syst, 2021, 16(3): 560–566.
- [31] WANG HP, CHU XL, CHEN P. Partial least squares regression residual extreme learning machine (PLSRR-ELM) calibration algorithm applied in fast determination of gasoline octane number with near-infrared spectroscopy [J]. Fuel, 2021, 309(12): 22–24.
- [32] 刘剑, 刘元法, 李进伟. 基于气相色谱的植物油特征脂肪酸高温热氧化特性研究[J]. 中国油脂, 2019, 44(1): 35–40.
- LIU J, LIU YF, LI JW. Characterization of high-temperature thermal oxidation of fatty acids in vegetable oils based on gas chromatography [J]. China Oils Fats, 2019, 44(1): 35–40.
- [33] LI X, LI J, WANG Y, *et al.* Effects of frying oils' fatty acids profile on the formation of polar lipids components and their retention in French fries over deep-frying process [J]. Food Chem, 2017, 237(15): 98–102.

(责任编辑: 韩晓红 张晓寒)

作者简介



刘楠, 硕士研究生, 主要研究方向为近红外光谱的食品安全检测。
E-mail: ln1966491009@163.com



刘翠玲, 博士, 教授, 主要研究方向为智能测量技术与数据处理。
E-mail: liucl@btbu.edu.cn