

基于 TrAdaBoost 算法的近红外光谱模型传递研究

刘翠玲^{1,2} 徐金阳^{1,2} 孙晓荣^{1,2} 张善哲^{1,2} 咎佳睿^{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: 随着近红外光谱检测仪种类的增多,不同仪器间的校正模型存在无法共享问题,可利用模型传递解决。以食用油为研究对象,在主机上建立油酸质量比的极限学习机校正模型,利用迁移学习中的 TrAdaBoost 算法把主机模型传递到从机上,探讨标准化样品数量对模型传递效果的影响,并与直接标准化算法、缺损数据重构算法和极限学习机自编码器的模型传递算法进行对比。结果表明:主机模型经 TrAdaBoost 算法模型传递后,从机预测集决定系数 R^2 从 0.489 上升到 0.892,预测集均方根误差(Root mean square error of prediction, RMSEP)从 4.824 mg/g 下降到 0.267 mg/g,且模型效果几乎不受标准化样品数量的影响。说明 TrAdaBoost 算法可以有效应用于模型传递领域,实现了不同光谱仪器之间的共享。

关键词: 近红外光谱; 模型传递; 迁移学习; TrAdaBoost; 食用油

中图分类号: S123; TS227 文献标识码: A 文章编号: 1000-1298(2022)02-0239-07

OSID:



Near Infrared Spectroscopy Calibration Transfer Based on TrAdaBoost Algorithm

LIU Cuiling^{1,2} XU Jinyang^{1,2} SUN Xiaorong^{1,2} ZHANG Shanzhe^{1,2} ZAN Jiarui^{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: With more and more types of near infrared spectroscopy detectors, the inability to share calibration models between different instruments has become the main problem that limits its application, and calibration transfer has become the key to solve this problem. Taking edible oil as the research object, the extreme learning machine model of its acid value on the master instrument was established. And the TrAdaBoost algorithm in transfer learning was used to transfer the master model to the slave model, and the dependence of calibration transfer on the number of standardization samples was explored. It was also compared with the direct standardized, missing data recovery and transfer via extreme learning machine auto-encoder method. The results showed that the predictive power of the slave samples after the TrAdaBoost calibration transfer algorithm was most effective and very close to the predictive value of the master sample-master model. The R^2 of the validation set was increased from 0.489 to 0.892, the root mean square error of prediction (RMSEP) was reduced from 4.824 mg/g to 0.267 mg/g. Specifically, the model effect was almost independent of the number of standardized samples. The next degree of effect was the transfer via extreme learning machine auto-encoder method algorithm (TEAM), the missing data recovery algorithm (MDR) and direct standardized algorithm (DS) in decreasing order, respectively. It was shown that the TrAdaBoost can be effectively applied to the calibration transfer, and it can realize the communication between different spectroscopic instruments, which provided an idea for the calibration transfer.

Key words: near infrared spectroscopy; calibration transfer; transfer learning; TrAdaBoost; edible oil

0 引言

随着近红外 (Near infrared, NIR) 光谱学的发

展,近红外光谱技术已成功地应用于食品、医药和农业等许多领域^[1-4]。然而,由于仪器的多样性,所建立的校正模型已经不适用于新样本,但是重新建立

校正模型需要大量的工作和时间。模型传递 (Calibration transfer) 为这类问题的解决提供了可行途径,其本质是克服样本在不同仪器间测量信号的 不一致性,使得主机建立的校正模型可以用于其他 仪器^[5-6]。

近年来,国内外学者对模型传递在食品安全领 域的应用进行了大量研究。刘锐等^[7]使用直接标 准化算法 (Direct standardized, DS) 和分段直接标 准化 (Piecewise direct standardization, PDS) 对牛奶的成 分进行模型传递研究。赵政^[8]将斜率截距 (Slope/ bias, S/B) 算法应用到新鲜度猪肉的挥发性和盐基 氮原子含量的模型传递。CHEN 等^[9]提出了一种基 于极限学习机自编码器的模型传递方法 (Transfer via extreme learning machine auto-encoder method, TEAM),通过对玉米、烟草、药品数据集的实验证明 其预测性能较好。这些传统算法都成功将校正模型 进行传递并取得不错效果,但存在参数设置复杂、效 率低等缺点。

随着计算机技术的快速发展,迁移学习 (Transfer learning) 逐渐成为研究热点,其核心思想 是将某个领域上学习到的知识迁移到不同但相关的 领域中^[10-11]。这与模型传递将主机建立的校正模 型迁移到从机上的思路十分相近,所以迁移学习为 模型传递提供了一种新思路。但现阶段迁移学习主 要用于图像^[12-13]、文本^[14]、语义^[15]等方面,戴文 渊^[16]提出 TrAdaBoost 算法并成功将其应用到跨领 域的文本分类。迁移学习具有数据小、效率高和鲁 棒性较好等优点^[17],但其在模型传递领域的应用却 鲜有报道。

本文采用 TrAdaBoost 算法,并结合极限学习机 (Extreme learning machine, ELM)^[18]建立传递模型, 实现食用油酸值的校正模型在不同仪器之间传递。 并与直接标准化算法 (DS)、基于极限学习机自编码 器的模型传递算法 (TEAM) 和缺损数据重构算法 (Missing data recovery, MDR)^[19]进行对比研究,以 期将迁移学习更好地应用于模型传递领域。

1 材料与方法

1.1 材料

实验食用油样本来源于北京古船食品有限公 司,选择了 5 种食用油 (玉米油、芝麻香油、大豆油、 橄榄油、小磨香油) 共计 129 个样本。并依据 GB/T 5530—2005《动植物油脂 酸值和酸度测定》^[20]测定 所有食用油样本的酸值。

1.2 实验仪器

实验在北京工商大学光谱技术与品质检测实验

室完成,光谱检测仪器为 Bruker 公司的 VERTEX - 70 型傅里叶红外光谱仪和 MATRIX - F 型傅里叶红 外光谱仪。实验设定 MATRIX - F 型光谱仪为主 机,VERTEX - 70 型光谱仪为从机。仪器参数详见 表 1。

表 1 光谱仪器和参数

Tab. 1 Spectroscopic instruments and parameters

光谱仪型号	主/从机类型	分辨率/cm ⁻¹	波数/cm ⁻¹
MATRIX - F	主	2	12 800 ~ 4 000
VERTEX - 70	从	4	25 000 ~ 20

1.3 光谱采集

使用 2 台近红外光谱仪分别采集食用油样本信 息。采集时仪器参数设置为:分辨率 16 cm⁻¹;样本 扫描次数 32;背景扫描次数为 32;光谱的采集范围 9 000 ~ 5 000 cm⁻¹;光阑 6 mm;扫描频率 10 kHz。

1.4 数据分析与处理

1.4.1 光谱数据预处理

由于客观存在或人为因素,实验采集到的光谱 数据通常会掺入噪声干扰,甚至使得数据不再完整。 因此,光谱数据在建立校正模型之前必须进行预处 理,压缩建模光谱集数据的规模,从而平滑噪声并剔 除奇异数据^[21]。在全光谱范围内比较了有限脉冲 响应 (Finite impulse response, FIR)^[22]、多元散射校 正 (Multiplicative scatter correction, MSC)^[23]、正交信 号校正 (Orthogonal signal correction, OSC)^[24]和标准 正态变量变换 (Standard normalized variate, SNV)^[25] 共 4 种预处理方法对 ELM 模型性能的影响,发现经 SNV 预处理后的光谱能有效提高 ELM 模型的性能, 因此,本实验以 SNV 预处理后的光谱作为后续分析 的基础。

1.4.2 样本划分方法

建立校正模型前,需要将食用油样品集划分为 训练集和测试集。Kennard - Stone^[26]算法是一种有 效的样品集划分方法。其算法是通过计算样品之间 的欧氏距离 (Euclidean distance),选择代表性强的 样品作为训练集样品,其余的作为测试集,从而提高 校正模型性能。

1.5 建立 PCA - ELM - TrAdaBoost 模型传递分 析模型

1.5.1 PCA 降维

由于食用油数据集中的样本数量远小于波数, 因此需要降维处理减少数据的复杂度。PCA 是常 用的化学计量工具,它可以将数据从高维空间投影 到低维空间,并尽可能保留原始数据的有效信息^[27]。

1.5.2 ELM 校正模型

ELM 的拓扑结构是一个经典的前馈神经网络,具有输入层、隐藏层和输出层共 3 层。隐藏神经元参数是随机分配的,输出权重可以通过使用 Moore - Penrose 广义逆矩阵进行分析,ELM 算法具有学习速度快、可调参数少等优点^[28]。本文采用加权 ELM 算法建立校正模型,将降维后的食用油数据作为神经元输入,油酸值预测值作为神经元输出,隐藏层节点设为 20,并选用 Sigmoid 作为激活函数。

1.5.3 TrAdaBoost 模型传递方法原理

现阶段迁移学习大致分为 3 类:基于实例的迁移、基于特征的迁移和基于共享参数的迁移^[29-30]。其中 TrAdaBoost 算法就是基于实例的迁移学习,其核心思想是:对源域 D_s 的标记数据实例进行有效权重分配,使源域实例分布接近目标域 D_t 的实例分布,从而在目标领域中建立一个可靠的校正模型实现模型传递^[31-32]。其主要步骤为:

(1)初始化设置。将源域 S (主机)和目标域 T (从机)的样本数据合并,形成一个新的训练集 $X^i(i=1,2,\cdots,m+n)$ 。并设置最大迭代次数 N ,初始化源域权重 w_s^i 、目标域权重 w_t^i 和学习机权重 β ,即

$$w_s^i = \frac{1}{m} \quad (i=1,2,\cdots,m) \tag{1}$$

$$w_t^i = \frac{1}{n} \quad (i=1,2,\cdots,n) \tag{2}$$

$$\beta = \frac{1}{1 + \sqrt{2\ln \frac{m}{N}}} \tag{3}$$

式中 m,n ——源域、目标域训练集样本数量

(2)计算误差。建立加权 ELM 校正模型。将训练集 X^i 经 PCA 降维后输入到校正模型中,输出预测值 Y^i ,其与真实值 T^i 的预测误差为 ε^i ,计算式为

$$\varepsilon^i = \sum_{i=1}^{n+m} \frac{w(Y^i - T^i)^2}{\max |Y^i - T^i|} \quad (i=1,2,\cdots,m+n) \tag{4}$$

其中 $w \in \{w_s^i; w_t^i\}$ 。且为避免算法停止, ε^i 必须小于 0.5,当超过 0.5 时,设置 $\varepsilon^i = 0.5$ 。

(3)更新权重。更新每一轮迭代的源域权重 w_s^i 和目标域权重 w_t^i 。如果源域中的标注样本被误预测,说明样本与目标域的相关性较小。然后,降低其权重以减少其影响。相反,如果目标域中的标记样本被错误预测,则增加其权重以提高其在下一轮的效果,即

$$w_s^{i+1} = w_s^i \beta_k^{\varepsilon^i} \quad (i=1,2,\cdots,m) \tag{5}$$

$$w_t^{j+1} = w_t^j \beta_k^{-\varepsilon^j} \quad (j=m+1,m+2,\cdots,m+n) \tag{6}$$

其中

$$\beta_k = \frac{\varepsilon^i}{1 - \varepsilon^i}$$

式中 β_k ——每次迭代的学习机权重

(4)达到最大迭代次数 N 后,输出 ELM 校正模型目标域的最终预测值 Y_N 。

1.6 模型传递的评价

模型的评价参数选择决定系数 (Correlation coefficient of cross-validation, R^2) 和预测集均方根误差 (Root mean square error of prediction, RMSEP)。 R^2 越大表明光谱信息与食用油理化值的相关性越好, RMSEP 越小,表明预测性能越好,模型传递的效果越好。

2 结果与分析

2.1 食用油光谱特征分析

主机和从机所采集的食用油样品光谱经 SNV 预处理后波数在 9 000 ~ 5 000 cm^{-1} 范围内,共 2 074 个波数点 (图 1)。观察发现,8 700 cm^{-1} 和 8 200 cm^{-1} 附近有 2 处主要的吸收峰,且其分别可能是由 C—H (CH_3, CH_2) 基团的 2ν 二倍频和 $2\nu + 2\beta$ 组合频作用所引起的^[33]。

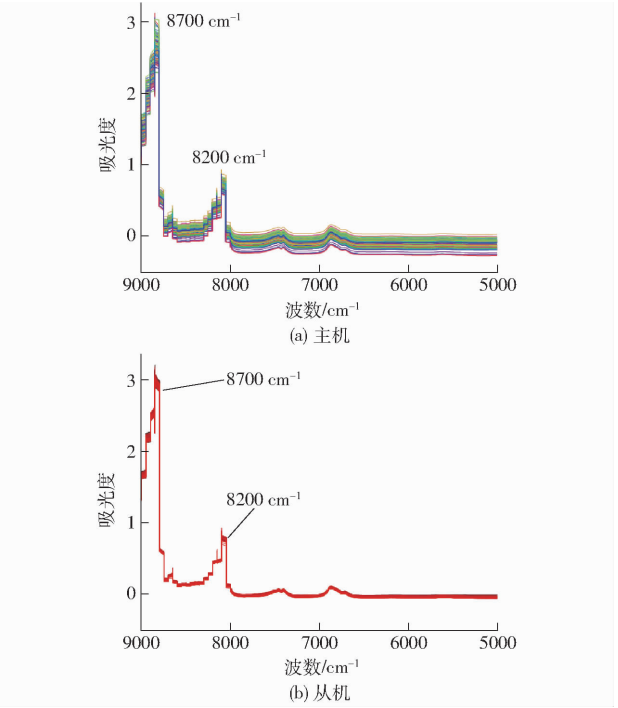


图1 经 SNV 预处理后的主机与从机的食用油吸光度
Fig.1 Spectra of edible oil from master and slave after SNV pretreatment

2.2 样本划分

采用 Kennard - Stone 算法对 129 个食用油样品进行数据划分,训练集与预测集的比例约为 3:1,所以有 97 个样本作为训练集,32 个样本作为预测集。食用油样本的数据集划分的具体情况如表 2 所示,训练集的油酸质量比为 0.08 ~ 2.12 mg/g ,涵盖了较宽的范围,有助于构建稳定的数学模型。且预测集

的油酸质量比为 0.09 ~ 1.82 mg/g,在训练集的油酸质量比范围之内,表明该子集可以对模型的预测性进行验证。

表 2 食用油样品的数据集划分
Tab.2 Data set division of edible oil

数据集	数量	油酸质量比/(mg·g ⁻¹)			
		最大值	最小值	平均值	标准偏差
训练集	97	2.12	0.08	0.58	0.62
预测集	32	1.82	0.09	0.80	0.82

2.3 主成分分析

由于原始光谱数据包含 2 074 个波数点,为了降低模型的复杂度和计算量,故对数据采用 PCA 主成分分析。前 5 个主成分的贡献率分别为 70.52%、16.79%、4.15%、3.29% 和 1.21%。通过计算可知,前 5 个主成分的累计贡献率已经达到了 95.96%,故本文使用 5 个主成分进行建模分析。

2.4 模型预测结果

利用在主机上已建立好的 ELM 模型分别预测主机和从机上的 32 个预测集样品的油酸质量比,并循环 20 次取平均值。利用主机模型预测主机样品集时,预测值与真实值的决定系数 R^2 为 0.922,预测集均方根误差(RMSEP)为 0.198 mg/g,预测效果较好。而当从机样品集未进行模型传递直接代入主机模型进行预测时,决定系数 R^2 下降到 0.489,预测集均方根误差(RMSEP)提高到 4.824 mg/g,预测结果产生较大差异。

图 2 为 2 台光谱仪的预测集与真实值差值图,其中 Δ_1 为真实值与主机预测值的差值, Δ_2 为真实值与从机样品集直接应用于主机模型预测值的差值。可以看出 Δ_1 几乎为 0,主机预测值与真实值相差不大,说明预测较为准确。但当从机样品集直接应用于主机模型时, Δ_2 较大,预测值与真实值产生较大偏差,所以主机建立的校正模型无法直接应用于从机样品,需要进行模型传递改善预测效果。

2.5 基于 TrAdaBoost 算法模型传递及预测结果

首先,按照算法要求将 97 个主机训练集和 97 个从机训练集进行合并,形成一个新的训练集。其次,对新训练集进行 PCA 处理并选取 5 个主成分,建立 PCA-ELM-TrAdaBoost 模型传递分析模型。最后,将从机的 32 个样本代入上述传递后的主机模型中进行预测,并循环 20 次取平均值。

图 3 为循环 20 次的模型传递与未传递的决定系数 R^2 对比图,经过模型传递后的从机预测值与真实值的 R^2 较高且比较稳定,在 0.9 左右浮动,而未经过模型传递的决定系数 R^2 相对较低且不稳定。

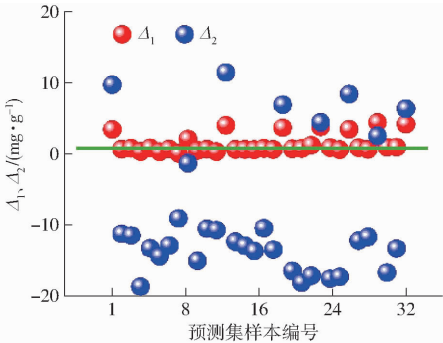


图 2 2 种模型的油酸质量比预测值与真实值差值图

Fig.2 Difference of predicted and true values between two models

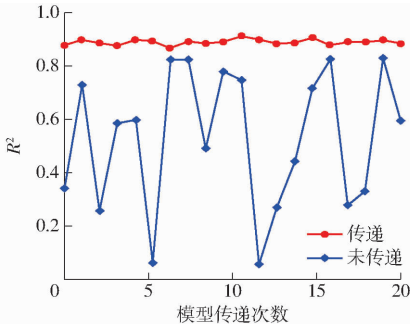


图 3 传递与未传递的决定系数 R^2 对比

Fig.3 R^2 comparison between transfer and un-transfer

图 4 为经过模型传递后从机预测值与真实值差值图,可以看出经模型传递后预测值与真实值差值(Δ)几乎为 0,预测值接近真实值,说明模型预测能力有所改善。决定系数 R^2 从 0.489 提高到 0.892, RMSEP 从原先的 4.824 mg/g 降低到 0.267 mg/g。说明经 TrAdaBoost 算法传递后的主机模型能更有效适用从机样本,从而减少主机与从机间的数据差异。

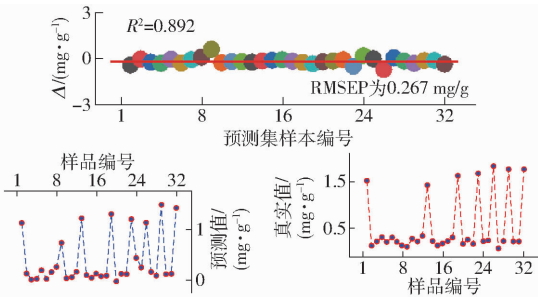


图 4 模型传递后的预测值与真实值差值图

Fig.4 Difference of predicted and true values between transfer and un-transfer

2.6 标准样品数量对模型传递的影响

为了取得较好的模型传递效果,采用 Kennard-Stone 算法依次从主机训练集中选取 0、10、...、90 个样品作为标准化样品集,对 TrAdaBoost 算法进行测试。图 5 为不同标准化样品数量对 TrAdaBoost 模型的影响情况。当主机样品数为 0 时,此时仍是从机

样品直接在主机模型预测的结果。随着主机样品数的增加, RMSEP 下降到稳定值后无明显变化, 说明标准样品集的数量对 TrAdaBoost 算法的影响很小。

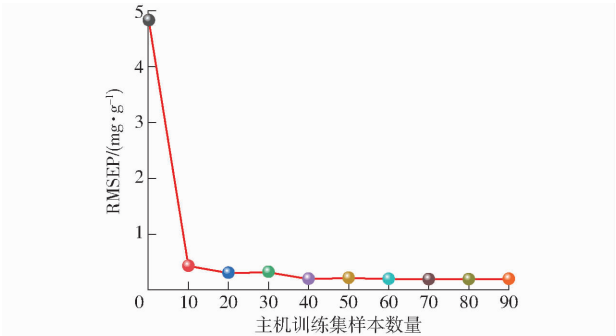


图 5 不同标准化样品数对模型的影响
Fig.5 Influence of number of samples on model

2.7 TrAdaBoost 算法与其他算法的比较

为了评估 TrAdaBoost 算法的模型传递的性能, 采用 DS 算法、TEAM 算法和 MDR 算法进行对比测试。DS 算法的基本思想是先建立主、从机光谱数据的数学函数关系, 再用函数关系转换从机光谱数据, 从而减少不同仪器间所测同一样本光谱数据的差异, 实现模型在不同仪器间传递。TEAM 算法将主机光谱作为学习目标, 选择隐藏层中权重与偏差正交的节点, 利用极限学习机的快速逼近能力与泛化性能建立主机和从机光谱之间的关系, 以减少传递后的预测误差。MDR 算法通过构建光谱转换矩阵, 将待转换光谱视作缺失数据, 通过多次迭代计算, 可逐步实现从机光谱向主机光谱方向的转换, 再预测时能够得到从机与主机光谱数据较小偏差的结果。

为了对比 4 种模型传递算法, 样本的数据集划分、主成分个数以及 ELM 隐藏节点数均相同。经 4 种算法传递后的主机模型分别预测从机样品的油酸质量比, 并循环 20 次计算平均 R^2 和 RMSEP, 以评估传递模型的性能, 如图 6 所示。

对比 4 种算法, 将从机样本代入经 TrAdaBoost 算法转换后的主机模型, 其预测能力效果最好, R^2 为 0.892, RMSEP 为 0.267 mg/g, 十分接近主机本身的样本代入主机模型的预测值, R^2 为 0.922 和 RMSEP 为 0.198 mg/g。预测效果由优到差依次为 TEAM 算

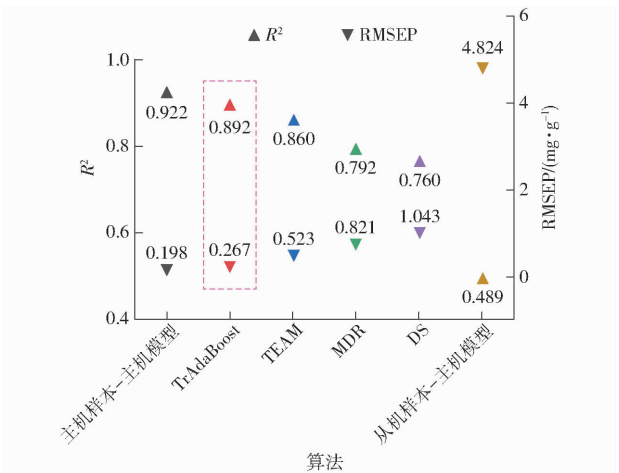


图 6 不同算法下的模型 R^2 和 RMSEP
Fig.6 R^2 and RMSEP results by different calibration transfer methods

法、MDR 算法和 DS 算法, 但这 3 种模型传递算法的预测能力均有所改善, 均大于从机样本直接应用到主机模型上的预测值。TrAdaBoost 模型预测效果比其他 3 种算法好的原因可能是该算法将主机样本和从机样本进行合并, 新产生的训练集因为包含主机样本, 使得拟合效果更接近主机模型, 以致在预测时可以更好地适应主机模型。

3 结论

(1) 将迁移学习的方法应用到近红外光谱中模型传递领域, 采用 TrAdaBoost 算法结合 ELM 模型, 实现了食用油酸质量比校正模型在不同仪器之间的传递。并与 DS 算法、TEAM 算法和 MDR 算法进行对比研究。

(2) 经模型传递后的主机模型的从机样本油酸质量比预测模型 R^2 从 0.489 提高到 0.892, RMSEP 从 4.824 mg/g 降低到 0.267 mg/g。对比其余 3 种算法, TrAdaBoost 算法的预测结果最好, 且预测值十分接近主机模型的预测结果, 此外模型的建立几乎不受标准样品数量的影响。

(3) TrAdaBoost 算法可以有效使不同仪器之间的光谱数据进行转换, 提高了从机样本在主机模型的适应度, 这对迁移学习应用于模型传递领域提供了研究思路, 使近红外光谱技术应用于食用油检测具有实际意义。

参 考 文 献

[1] 褚小立, 史云颖, 陈瀑, 等. 近五年我国近红外光谱分析技术研究与应用进展[J]. 分析测试学报, 2019, 38(5): 603-611. CHU Xiaoli, SHI Yunying, CHEN Pu, et al. Research and application progress of near infrared spectroscopy analytical technique for China in past five years[J]. Chinese Journal of Analysis Laboratory, 2019, 38(5): 603-611. (in Chinese)

[2] 汪六三, 黄子良, 王儒敬. 基于近红外光谱和机器学习的大豆种皮裂纹识别研究[J]. 农业机械学报, 2021, 52(6): 361-368. WANG Liusan, HUANG Ziliang, WANG Rujing. Research on soybean seed coat crack recognition based on near-infrared spectroscopy and machine learning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 361-368.

(in Chinese)

- [3] 郭文川,朱德宽,张乾,等.基于近红外光谱的掺伪油茶籽油检测[J].农业机械学报,2020,51(9):350-357.
GUO Wenchuan, ZHU Dekuan, ZHANG Qian, et al. Detection of adulteration of camellia seed oil based on near-infrared spectroscopy[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 350-357. (in Chinese)
- [4] 王伟,赵昕,褚璇,等.基于可见/近红外高光谱的八角茴香与莽草无损鉴别[J].农业机械学报,2019,50(11):373-379.
WANG Wei, ZHAO Xin, CHU Xuan, et al. Nondestructive identification of star anise and manggrass based on visible/near infrared hyperspectrum[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(11): 373-379. (in Chinese)
- [5] 刘翠玲,刘浩言,孙晓荣,等.食用油酸值与过氧化值近红外光谱模型转移研究[J].农业机械学报,2020,51(9):344-349.
LIU Cuiling, LIU Haoyan, SUN Xiaorong, et al. Transfer of near-infrared spectroscopy model of acid oil value and peroxidation value[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(9): 344-349. (in Chinese)
- [6] SHI Y Y, LI J Y, CHU X L. Progress and applications of multivariate calibration transfer methods[J]. Chinese Journal of Analytical Chemistry, 2019, 47(4): 479-487.
- [7] 刘锐,梁秋曼,南良康,等.用于牛奶分析的中红外光谱标准化及其在模型传递中的作用[J].中国奶牛,2019,1(4):1-5.
LIU Rui, LIANG Qiuman, NAN Liangkang, et al. Mid-infrared spectroscopy standardization for milk analysis and its role in model transfer[J]. China Dairy Cow, 2019, 1(4): 1-5. (in Chinese)
- [8] 赵政.猪肉新鲜度光谱模型的建立及传递方法研究[D].武汉:华中农业大学,2013:25-42.
ZHAO Zheng. Study on the establishment of pork freshness spectrum model and transfer method[D]. Wuhan: Huazhong Agricultural University, 2013:25-42. (in Chinese)
- [9] CHEN W, BIN J, LU H, et al. Calibration transfer via extreme learning machine auto-encoder[J]. Analyst, 2016, 141(6): 1973-1980.
- [10] 刘鑫鹏,栾悉道,谢毓湘,等.迁移学习研究和算法综述[J].长沙大学学报,2018,32(5):28-31.
LIU Xinpeng, LUAN Xidao, XIE Yuxiang, et al. Overview of transfer learning research and algorithms[J]. Journal of Changsha University, 2018, 32(5): 28-31. (in Chinese)
- [11] 王惠.迁移学习研究综述[J].电脑知识与技术,2017,13(32):203-205.
- [12] 张雪松,庄严,闫飞,等.基于迁移学习的类别级物体识别与检测研究与进展[J].自动化学报,2019,45(7):1224-1243.
ZHANG Xuesong, ZHUANG Yan, YAN Fei, et al. Research and development of category-level object recognition and detection based on transfer learning[J]. Acta Automatica Sinica, 2019, 45(7): 1224-1243. (in Chinese)
- [13] MARGARET L, MELOSSA S, ROBERT B, et al. Amplitude frequency analysis of emotional speech using transfer learning and classification of spectrogram images[J]. Advances in Science, Technology and Engineering Systems, 2018, 3(4): 363-371.
- [14] 吴斌,吉佳,孟琳,等.基于迁移学习的唐诗宋词情感分析[J].电子学报,2016,44(11):2780-2787.
WU Bin, JI Jia, MENG Lin, et al. Sentiment analysis of Tang poetry and Song poetry based on transfer learning[J]. Chinese Journal of Electronics, 2016, 44(11): 2780-2787. (in Chinese)
- [15] 鲁强,刘兴昱.基于迁移学习的知识图谱问答语义匹配模型[J].计算机应用,2018,38(7):1846-1852.
LU Qiang, LIU Xingyu. Knowledge graph question answering semantic matching model based on transfer learning[J]. Computer Applications, 2018, 38(7): 1846-1852. (in Chinese)
- [16] 戴文渊.基于实例和特征的迁移学习算法研究[D].上海:上海交通大学,2009:25-34.
DAI Wenyuan. Research on transfer learning algorithm based on examples and features[D]. Shanghai: Shanghai Jiaotong University, 2009:25-34. (in Chinese)
- [17] QIAN Z, LI H, ZHANG Y, et al. Instance transfer learning with multisource dynamic TrAdaBoost[J]. The Scientific World Journal, 2014, 40(6): 1176-1183.
- [18] 喻黎明,严为光,龚道枝,等.基于ELM模型的浅层地下水位埋深时空分布预测[J].农业机械学报,2017,48(2):215-223.
YU Liming, YAN Weiguang, GONG Daozhi, et al. Prediction of temporal and spatial distribution of shallow groundwater depth based on ELM model[J]. Transactions of the Chinese Society for Agricultural Machinery, 2017, 48(2): 215-223. (in Chinese)
- [19] 杜文.多元校正及模型转移中的缺损数据重构和交替残差多线性方法研究[D].长沙:湖南大学,2016.
DU Wen. Research on defect data reconstruction and alternating residual multilinear method in multivariate calibration and model transfer[D]. Changsha: Hunan University, 2016. (in Chinese)
- [20] 国家粮食局.动植物油脂 酸值和酸度测定:GB/T 5530—2005[S].北京:中国标准出版社,2005.
- [21] 张进,胡芸,周罗雄,等.近红外光谱分析中的化学计量学算法研究新进展[J].分析测试学报,2020,39(10):1196-1203.
ZHANG Jin, HU Yun, ZHOU Luoxiong, et al. New developments in chemometric algorithms in near-infrared spectroscopy analysis[J]. Chinese Journal of Analysis Laboratory, 2020, 39(10): 1196-1203. (in Chinese)
- [22] 张静,李钰,任舜文.基于有限脉冲响应滤波器的实时小波算法及其在色谱信号解析中的应用[J].色谱,2017,35(4):368-374.
ZHANG Jing, LI Yu, REN Shunwen. Real-time wavelet algorithm based on finite impulse response filter and its application in chromatographic signal analysis[J]. Chromatography, 2017, 35(4): 368-374. (in Chinese)
- [23] WU Y F, PENG S L, XIE Q, et al. An improved weighted multiplicative scatter correction algorithm with the use of variable

selection: application to near-infrared spectra[J]. Chemometrics and Intelligent Laboratory Systems,2019,185(1): 114 – 121.

[24] 王其滨,杨辉华,潘细朋,等. 随机森林结合直接正交信号校正的模型传递方法[J]. 激光与红外,2020,50(9):1081 – 1087.
WANG Qibin, YANG Huihua, PAN Xipeng, et al. Model transfer method combining random forest with direct orthogonal signal correction[J]. Laser and Infrared, 2020, 50(9): 1081 – 1087. (in Chinese)

[25] 李杰,李尚科,蒋立文,等. 基于近红外光谱技术与化学计量学的绿茶无损鉴别方法研究[J]. 分析测试学报,2020,39(11):1344 – 1350.
LI Jie, LI Shangke, JIANG Liwen, et al. Research on the non-destructive identification method of green tea based on near-infrared spectroscopy and chemometrics[J]. Chinese Journal of Analysis Laboratory, 2020, 39(11): 1344 – 1350. (in Chinese)

[26] MORAIS C, SANTOS M, LIMA K, et al. Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation Kennard – Stone algorithm approach[J]. Bioinformatics, 2019, 35(24):5257 – 5263.

[27] 王计平,支晓蓉,黄继红,等. 基于 GIS 和 PCA 降维的森林景观格局等级特征分析[J]. 农业机械学报,2019,50(1):195 – 201.
WANG Jiping, ZHI Xiaorong, HUANG Jihong, et al. Analysis of hierarchical characteristics of forest landscape pattern based on GIS and PCA dimensionality reduction[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(1): 195 – 201. (in Chinese)

[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 Rui, LIANG Xun, QI Jinshan, et al. Frontier development and trend of extreme learning machines[J]. Chinese Journal of Computers, 2019, 42(7): 1640 – 1670. (in Chinese)

[29] 张景祥. 迁移学习技术及其应用研究[D]. 无锡:江南大学,2015:4 – 5.
ZHANG Jingxiang. Research on transfer learning technology and its application[D]. Wuxi: Jiangnan University, 2015:4 – 5. (in Chinese)

[30] 李子彬. 基于聚类动态 TrAdaBoost 迁移学习分类方法[D]. 广州:广东工业大学,2018.
LI Zibin. Dynamic TrAdaBoost migration learning classification method based on clustering[D]. Guangzhou: Guangdong University of Technology, 2018. (in Chinese)

[31] CHEN Y Y, WANG Z B. Cross components calibration transfer of NIR spectroscopy model through PCA and weighted ELM-based TrAdaBoost algorithm[J]. Chemometrics and Intelligent Laboratory Systems,2019,192(15):103824 – 103843.

[32] YU Y, HUANG J P, LIU S S, et al. Cross target attributes and sample types quantitative analysis modeling of near-infrared spectroscopy based on instance transfer learning[J]. Measurement,2021,177(1): 109340 – 109359.

[33] 吴新生,谢益民. 基于相关分析法的木素 C—H 键的近红外光谱特征吸收峰的解析[J]. 造纸科学与技术,2010,29(3): 69 – 72.
WU Xinsheng, XIE Yimin. Analysis of the characteristic absorption peaks of lignin C—H bonds in the near-infrared spectrum based on correlation analysis[J]. Paper Science and Technology, 2010, 29(3): 69 – 72. (in Chinese)

(上接第 238 页)

[23] 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·激光,2021,32(6):653 – 658.
XU Liang, ZHANG Ziyue, CHEN Xi, et al. Improved sparrow search algorithm based BP neural networks for aero-optical imaging deviation prediction[J]. Journal of Optoelectronics · Laser, 2021,32(6):653 – 658. (in Chinese)

[24] 汤安迪,韩统,徐登武,等. 基于混沌麻雀搜索算法的无人机航迹规划方法[J]. 计算机应用,2021,41(7):2128 – 2136.
TANG Andi, HAN Tong, XU Dengwu, et al. Path planning method of unmanned aerial vehicle based on chaos sparrow search algorithm[J]. Journal of Computer Applications, 2021,41(7):2128 – 2136. (in Chinese)

[25] 杨薇,尹青剑,张付杰,等. 基于介电特性的三七粉含水率检测与建模[J]. 昆明理工大学学报(自然科学版),2016,41(4):81 – 87.
YANG Wei, YIN Qingjian, ZHANG Fujie, et al. Predicating and modeling of moisture content of panax notoginseng powder based on dielectric properties[J]. Journal of Kunming University of Science and Technology(Natural Science Edition), 2016, 41(4):81 – 87. (in Chinese)

[26] 牛智有,刘芳宏,刘鸣,等. 平行极板电容传感器介电式颗粒饲料水分检测仪设计与试验[J]. 农业工程学报,2019,35(18):36 – 43.
NIU Zhiyou, LIU Fanghong, LIU Ming, et al. Design of dielectric pellet feed moisture detector based on parallel plate capacitance sensor[J]. Transactions of the CSAE, 2019,35(18):36 – 43. (in Chinese)

[27] 郭交,段凯文,郭文川,等. 基于微波自由空间测量的小麦含水率检测方法[J]. 农业机械学报,2019,50(6):338 – 343.
GUO Jiao, DUAN Kaiwen, GUO Wenchuan, et al. Detection method of moisture content of wheat with microwave free-space measurement[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(6):338 – 343. (in Chinese)