

变量优选校正算法的鲜枣可溶性固形物检测模型传递方法研究

孙海霞, 张淑娟*, 薛建新, 赵旭婷, 邢书海, 陈彩虹, 李成吉

山西农业大学工学院, 山西 太谷 030801

摘要 在水果的品质检测和分级分选中, 存在不同仪器所建检测模型难以共享的难题。为此, 以壶瓶枣为研究对象, 利用可见/近红外光谱技术探讨仪器间可溶性固形物含量(SSC)检测模型的传递方法。首先, 采用美国 ASD(Analytical Spectral Device)公司生产的两台仪器采集样本的光谱信息, 采用最小二乘支持向量机(LS-SVM)建立原始光谱、Savitzky-Golay 一阶导数处理、标准正态变量变换后的 SSC 检测模型, 预测不同仪器采集的光谱时 3 种方法的预测能力均较差。预测同一台仪器的光谱时, 基于原始光谱的主仪器所建模型最优, 预测集的决定系数(R_p^2)和均方根误差(RMSEP)分别为 0.73 和 1.36%。在此基础上, 采用 Kennard/Stone 算法选取标样, 利用专利算法(Shenk's)、直接标准化(DS)、斜率/偏差算法(S/B)进行模型传递。然后, 根据回归系数提取主仪器(24 个)和从仪器(28 个)的特征波长, 优选出单一变量(SV)24 个、共性变量(CV)23 个、融合变量(FV)29 个, 均涵盖了 SSC 的主要吸收谱带。利用优选的变量分别建立主仪器的 LS-SVM 检测模型, 采用主仪器的预测结果($R_p^2=0.78\sim 0.80$, RMSEP=1.07%~1.13%)明显好于全波段所建模型, 但预测从仪器时 RMSEP 为 6.62%~7.88%, 模型失效。最后, 基于波长位置偏移和分子振动的吸收特性提出了共性变量优选结合差值校正(CV-MC)、单一变量优选结合差值校正、融合变量优选结合差值校正、共性变量优选结合波长校正算法(CV-WC)进行模型传递, 并与 SV-Shenk's, CV-Shenk's, FV-Shenk's, SV-DS, CV-DS, FV-DS, SV-S/B, CV-S/B 和 FV-S/B 进行对比分析。结果表明, 基于全波段进行模型传递时, 预测结果均较差($R_p^2=0.03\sim 0.34$, RMSEP=2.44%~4.67%); 基于优选变量所建模型经 SV-Shenk's, CV-Shenk's, FV-Shenk's 传递后的结果较差, 经其他算法传递后的结果($R_p^2=0.47\sim 0.73$, RMSEP=1.30%~1.90%)好于全波段; 基于共性变量传递后的结果好于单一变量和融合变量, CV-MC 结果最佳($R_p^2=0.73$, RMSEP=1.30%), CV-WC 传递后的预测结果(RMSEP=1.62%)与 CV-DS 和 CV-S/B 相近。研究表明, CV-MC 和 CV-WC 均是一种有效模型传递算法, 对建立不同仪器间通用的鲜枣品质检测模型具有重要意义。

关键词 可见/近红外光谱; 模型传递; 鲜枣; 无损检测

中图分类号: S123 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)04-1041-06

引言

红枣富含多种营养成分, 为药食同源食物。可溶性固形物含量(soluble solids content, SSC)是水果内部品质的一个重要指标, 与消费者的购买需求息息相关。近红外光谱技术因其快速、无损、实时等优点, 在水果品质检测中被广泛应用。在实际应用中, 由于不同仪器间的光路设计、仪器老化、元器件更换、检测环境等存在差异, 致使一台仪器所建检测

模型应用于另一台仪器时误差较大或直接失效。因此, 光谱检测模型传递是目前研究的一个关键共性基础问题。

专利算法(Shenk's)^[1]、斜率/偏差算法(slope/bias, S/B)^[2]、直接校正(direct standardization, DS)和分段直接校正(piecewise direct standardization, PDS)^[3]等模型传递算法被广泛使用, 并发展出多种方法。刘娇等^[4]提出了一种分段直接校正结合线性插值算法, 实现了猪肉含水率高光谱模型在不同品种间传递。翁海勇等^[5]利用 DS 算法进行了不同型号高光谱成像仪间柑橘溃疡病检测模型传递, 预测集的认识率

收稿日期: 2018-11-28, 修订日期: 2019-03-01

基金项目: 国家自然科学基金项目(31801632, 31271973), 晋中市科技重点研发计划(Y172007-4), 山西农业大学科技创新基金项目(2016YJ04)资助

作者简介: 孙海霞, 女, 1985 年生, 山西农业大学工学院博士研究生 e-mail: sunhaixia1121@126.com

* 通讯联系人 e-mail: zsujuan1@163.com

从校正前的 38.1% 提高到了 86.2%。Liu 等^[6]提出了一种无标样的线性模型校正算法,使用约束优化方法可将主模型的系数转换为从模型的系数进而实现模型传递。Liang 等^[7]提出了 Rank-Kennard-Stone-PDS 算法,在优化传递样本集选择的同时,有效降低了传递后模型的预测均方根误差(root mean squared error of prediction, RMSEP)。徐惠荣等^[8]采用基于平均光谱差值校正的 DS 算法进行皇冠梨糖度模型转换后, RMSEP 降为 0.453°Brix。

上述研究主要针对全谱区光谱,通过光谱转换或结果校正进行模型传递,少有从变量优选和波长偏移角度出发进行信号修正的模型传递方法研究。因此,本研究基于近红外光谱技术,采用不同仪器采集鲜枣样本的光谱信息以建立 SSC 检测模型;针对全波段光谱信息,采用 Shenk's, DS 和 S/B 算法进行模型传递;在此基础上,利用波长位置偏移和分子振动在近红外区域的吸收特性提出共性变量优选结合波长修正算法和共性变量优选结合差值修正算法进行模型传递,探讨上述方法进行不同仪器间模型传递的可行性。

1 实验部分

1.1 样本采集

壶瓶枣样本采集于山西省太谷县小白乡,采后当天运达实验室并筛选,剔除畸形等缺陷果实,选择尺寸、形状一致的完好样本 177 个。将表面清洁、编号后装入自封袋中,在实验条件下放置 2 h 以恢复室温。随机将样本分为校正集(132 个)和预测集(45 个)。

1.2 光谱数据采集与理化值测定

采用美国 ASD (Analytical Spectral Device) 公司生产的两台光谱仪,分别采集鲜枣样本的漫反射光谱信息,采集范围为 350~2 500 nm。每个样本扫描 3 次取其平均值作为光谱数据,仪器 1(记作 M1)和仪器 2(记作 M2)采集的鲜枣样本平均光谱曲线见图 1。M1 和 M2 在反射率上差异较为显著,但曲线变化趋势较为相似。在 350~450 nm,两台仪器间曲线差异明显,但噪声大;2 400~2 500 nm 信噪比较低,因此选择 450~2 400 nm 进行研究。M1 测定的反射率明显小于 M2,但 970 和 1 400 nm 附近均明显存在与 O-H 键伸缩振动有关的水分吸收峰。

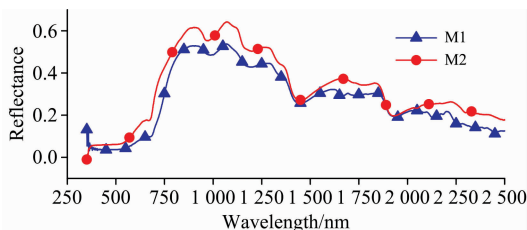


图 1 M1 和 M2 平均光谱曲线图

Fig. 1 The mean spectrum curve of M1 and M2

采用带有温度补偿的手持折光计分别测定每个鲜枣样本的 SSC,最大值为 36.6%,最小值为 21.1%,均值为 29.6%。

1.3 数据处理及模型传递方法

采用最小二乘支持向量机 (least squares support vector machines, LS-SVM)^[9]建模,基于偏最小二乘 (partial least square, PLS) 原理根据回归系数 (regression coefficient, RC) 的局部极值提取特征波长,采用校正决定系数 (determination coefficient of calibration, R_c^2)、预测决定系数 (determination coefficient of prediction, R_p^2)、校正均方根误差 (root mean squared error of calibration, RMSEC) 和 RMSEP 进行模型评价。提出了共性变量优选结合波长修正 (common variable-wavelength correction, CV-WC) 和共性变量优选结合差值修正算法 (common variable-subtraction correction, CV-SC) 2 种共性变量优选修正方法,并结合 DS, Shenk's 和 S/B 算法进行模型传递。

1.3.1 共性变量优选结合差值修正算法

CV-SC 是在优选共性变量的基础上,利用主从仪器共性变量间的光谱差异进行反射率值和预测结果校正,具体步骤为:

(1) 优选共性变量:根据回归系数分别提取主从仪器的特征波长,分析主从仪器间的波长偏移和分子振动在近红外区域的吸收特征,优选出共性变量的矩阵为 X 。

(2) 光谱反射率校正:主仪器标准样品中 X 的光谱矩阵为 S ,平均光谱反射率矩阵为 \bar{S} ;从仪器标准样品中 X 的光谱矩阵为 S_1 ,平均光谱反射率矩阵为 \bar{S}_1 ,计算主从仪器间 X 所对应的平均光谱反射率差值矩阵 ΔS

$$\Delta S = \bar{S} - \bar{S}_1 \quad (1)$$

从仪器未知样品中 X 的光谱矩阵 S_3 根据式(2)转换为与主仪器相匹配的光谱矩阵 S'_3 , S_1 根据式(3)转换为与主仪器相匹配的光谱矩阵 S'_1 。

$$S'_3 = S_3 + \Delta S \quad (2)$$

$$S'_1 = S_1 + \Delta S \quad (3)$$

(3) 预测结果校正:基于主仪器所建检测模型 B,以 S'_1 为模型输入量计算出从仪器标准样品的预测值 Y'_1 ,建立 Y'_1 与实测值 Y_1 间的关系,见式(4)。根据式(5)将 S'_3 为模型输入量计算出的预测值 Y' 转换为校正后的预测值 Y 。

$$Y_1 = AY'_1 + D \quad (4)$$

$$Y = AY' + D \quad (5)$$

1.3.2 共性变量优选结合波长修正算法

CV-WC 需先建立从仪器的共性变量和偏移后的波长所对应光谱矩阵间函数关系,具体步骤为:

(1) 从仪器波长校正:优选出共性变量的矩阵 X ,对应的从仪器提取的特征波长矩阵为 X_2 ,从仪器标准样品的 X 对应光谱矩阵为 S_1 , X_2 对应的光谱矩阵为 S_2 ,建立 X 与 X_2 间的函数关系,见式(6)。对从仪器未知样品中 X 的光谱矩阵 S_3 按式(7)转换为校正后的光谱矩阵 S_4 。

$$S_2 = kS_1 + b \quad (6)$$

$$S_4 = kS_3 + b \quad (7)$$

(2) 光谱反射率校正:(CV-SC 光谱反射率校正),将 S_2 和 S_4 转化为校正后的光谱矩阵 S'_2 和 S'_4 。

(3) 预测结果校正:(CV-SC 预测结果校正),得到转换后的未知样品预测值 Y 。

2 结果与讨论

2.1 基于全波段的模型传递分析

利用两台仪器的全波段光谱信息,采用 Savitzky-Golay 一阶导数(Savitzky-Golay-1st derivative, SG-1st Der)、标准正态变量变换(standard normal variate transformation, SNV)进行预处理,采用LS-SVM建立原始和预处理后的SSC检

测模型,结果见表1。SG-1st Der处理后两台仪器的 R_c^2 和RMSEC好于原始光谱和SNV处理,但 R_p^2 和 R_p^2 ,RMSEC和RMSEP的差值明显大于另两种方法,且对同一仪器预测集的预测结果较原始光谱差。基于三种方法处理后的光谱,M1所建模型的预测能力整体上好于M2所建模型;在预测不同仪器采集的光谱时,SG-1st Der和SNV处理后较原始光谱稍好,但是预测能力仍非常差。因此,选择M1作为主仪器,基于原始光谱进行模型传递修正以提高泛化能力。

表 1 两台仪器对鲜枣 SSC 的预测结果

Table 1 Prediction results of SSC on fresh jujube by M1 and M2 models

Instrument for modelling set	preprocessing methods	Instrument for validation set	R_c^2	RMSEC	R_p^2	RMSEP
M1	Original	M1	0.76	1.30	0.73	1.36
		M2			0.03	5.39
M2	Original	M1	0.70	1.47	0.02	6.40
		M2			0.70	1.44
M1	SG-1st Der	M1	0.96	0.61	0.72	1.38
		M2			0.40	2.35
M2	SG-1st Der	M1	0.91	0.88	0.05	2.29
		M2			0.51	1.64
M1	SNV	M1	0.72	1.42	0.72	1.41
		M2			0.40	2.19
M2	SNV	M1	0.76	1.32	0.03	4.02
		M2			0.52	1.65

针对校正集样本,采用 Kennard/Stone(KS)算法^[10]选取标样,建立 PLS 模型,交叉验证均方根误差(root mean squared error of cross validation, RMSECV)随样本选择数量的变化曲线见图2。当RMSECV达到最小值1.43%(A点所示)时,样本数量为27。132个样本全部用于建模时,RMSECV为1.72%(B点所示),变化趋势相对平缓。

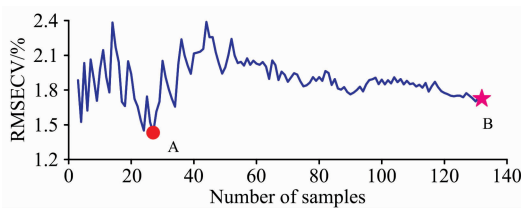


图 2 RMSECV 随样本数量变化图

Fig. 2 Change s of RMSECV with the number of samples

分别选择 27 个和 132 个标样,采用 Shenk's 和 DS 算法进行光谱校正后的曲线见图3。从仪器经校正后和主仪器同样品的平均光谱曲线反射率差异明显减小,变化趋势一致。132 个样本 DS 校正后与主仪器差异性最小;Shenk's 算法中,27 个样本校正后的差异性明显比 132 个样本小。27 个样本 DS 校正后差异最大,可能由于 DS 对全谱区转换时,标样过少,校正矩阵获取信息不足导致校正后光谱曲线拟合精度低。

利用全波段光谱信息,采用 Shenk's, DS 和 S/B 算法分别进行模型传递后的预测结果见表2。132 个标样的 DS 算法

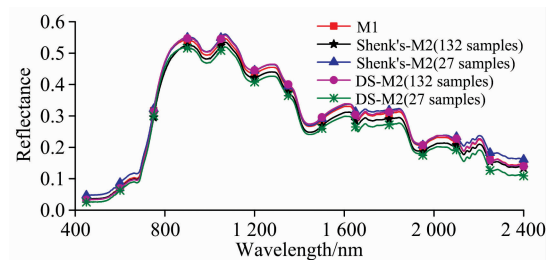


图 3 M1 和校正后 M2 的平均光谱曲线图

Fig. 3 The mean spectrum curve obtained by M1 and M2 after correction

表 2 基于全波段的模型传递结果

Table 2 Prediction results of model transfer using full wavelengths

Model transfer method	Number of standard samples	R_p^2	RMSEP
Shenk's	132	0.03	4.18
	27	0.21	3.47
DS	132	0.34	2.44
	27	0.03	4.67
S/B	132	0.24	2.48
	27	0.24	2.84

预测结果最优, RMSEP 由传递前的 5.39% 降至 2.44%; 27 个标样的 DS 算法传递效果最差, RMSEP 降为 4.67%; Shenk's 算法中窗口大小为 3, 27 个标样的传递结果(RM-

SEP=3.47%)好于 132 个标样(RMSEP=4.18%),表 2 与图 3 的分析结果相吻合。S/B 算法的传递效果介于 DS 和 Shenk's 之间,且标样数量对传递效果的影响比 DS 和 Shenk's 小。尽管模型传递后对从仪器的预测结果好于无处理前,但精度仍较低,难以满足实际要求。

2.2 变量优选

全波段光谱信息中,存在数据冗余现象,采用 RC 选取特征波长,主仪器提取 24 个特征波长见图 4(a),从仪器提取 28 个特征波长见图 4(b)。两仪器提取的特征波长中均有 917 nm,其余波长均不同。将主仪器提取的 24 个特征波长定义为单一变量(single variable, SV)。由于仪器内部元件、外部环境、被测样本内部存在化学物质的不同等,波长会发生一定的位置偏移^[11-12]。结合分子倍频和组合频吸收的特性,在主仪器提取到的特征波长中选取两台仪器间波长位置偏移在[-30 nm, 30 nm]范围内所对应的波长作为共性变量。以主仪器中所提取的 960 nm 为例,从仪器所提取的特征波长中存在 951 nm,两者存在位置偏移,但偏移范围在[-30 nm, 30 nm]内,且均可能与 O-H 键伸缩振动有关,故将主仪器的 960 nm 和从仪器的 951 nm 作为一个共性变量。结合图 4(a)和(b),表明从仪器提取到了主仪器特征波长中的 569, 621, 654, 683, 705, 785, 864, 917, 960, 1 062, 1 146, 1 201, 1 295, 1 389, 1 581, 1 853, 2 018, 2 104, 2 154, 2 266, 2 307, 2 328, 2 363 nm,将上述 23 个波长作为共性变量(common variable, CV)。由于糖类的吸收谱带在 972, 1 009, 1 450, 2 100 和 2 280~2 330nm 等附近有主要分布,同时在 880~915, 1 150~1 210 和 1 300~1 500 nm 等附近存在 C-H 官能团的振动^[13-14],故所选共性变量已涵盖了 SSC 的主要吸收谱带。除共性变量外,两仪器还存在其他特征波长,如图 4(b)中 526 nm。为避免信息丢失,将两仪器特征波长融合,即在共性变量的基础上增加主

仪器提取到的 1 696 nm 和从仪器提取到的 526, 1 341, 1 423, 1 523, 1 908 nm 共同组成融合变量(fusion variable, FV),共 29 个特征波长。

分别针对主仪器的单一变量、融合变量、共性变量及从仪器的 28 个特征波长(target variable, TV)所对应的光谱信息,建立 LS-SVM 模型,预测结果见表 3。预测主仪器光谱时,单一变量所建模型预测精度最高, R_p^2 为 0.80, RMSEP 为 1.07%。共性变量的建模结果($R_p^2 = 0.78$, RMSEP = 1.11%)稍差于单一变量,表明共性变量比单一变量丢失了极少量的有用信息,但对主仪器预测效果仍较好,对模型影响很小;对从仪器的预测结果(RMSEP=7.49%)差于全波段模型(RMSEP=5.39%)。融合变量所建模型对主仪器的预测结果($R_p^2 = 0.78$, RMSEP=1.13%)较单一变量和共性变量稍差,表明融合变量中可能存在极少量的噪声等冗余信息,但明显好于全波段模型($R_p^2 = 0.73$, RMSEP=1.36%)。整体而言,基于 3 类变量所建主仪器模型预测主仪器光谱时,均比全波段所建模型精度高,且预测能力的差异性不明显;对从仪器预测时,均比全波段所建模型精度低。

表 3 两台仪器对鲜枣 SSC 的预测结果

Table 3 Prediction results of SSC on fresh jujube by M1 and M2 models

Instrument for modelling set	Variable	R_p^2	RMSEC	M1		M2	
				R_p^2	RMSEP	R_p^2	RMSEP
M1	SV	0.78	1.24	0.80	1.07	0.02	6.62
	CV	0.78	1.25	0.78	1.11	0.01	7.49
	FV	0.78	1.23	0.78	1.13	0.01	7.88
M2	TV	0.77	1.27	0.01	11.77	0.77	1.20

2.3 基于变量优选修正算法的模型传递分析

针对 SV, CV, FV 所建立的模型,采用多种算法进行模型传递,预测结果见表 4。

表 4 基于变量优选的模型传递结果

Table 4 Prediction results of model transfer using selected variables

Model transfer method	Number of wavelengths	R_p^2	RMSEP
SV-Shenk's	24	0.21	3.55
CV-Shenk's	23	0.22	2.56
FV-Shenk's	29	0.01	7.94
SV-DS	24	0.60	1.62
CV-DS	23	0.60	1.62
FV-DS	29	0.50	1.83
SV-S/B	24	0.60	1.62
CV-S/B	23	0.61	1.60
FV-S/B	29	0.47	1.90
SV-MC	24	0.70	1.43
CV-MC	23	0.73	1.30
FV-MC	29	0.51	1.80
CV-WC	23	0.60	1.62

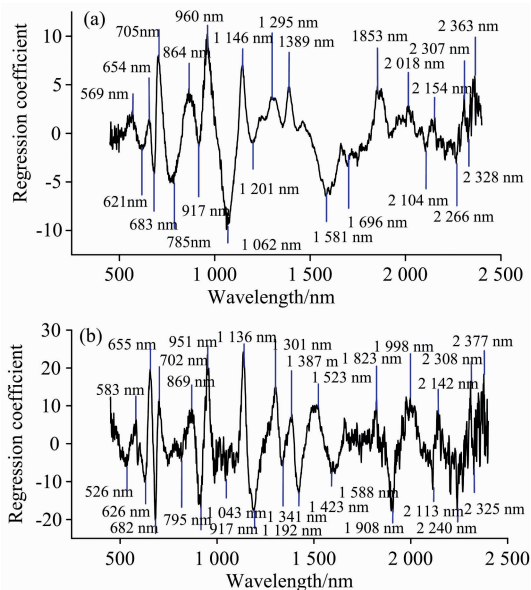


图 4 M1(a)和 M2(b) PLS 模型的回归系数图

Fig. 4 Regression coefficient of PLS model on M1 (a) and M2 (b)

变量优选的 Shenk's 算法传递中, 选取全波段传递时所对应的 27 个样本为标样, 窗口大小为 3, CV-Shenk's 传递后的 RMSEP 由全波段传递后的 3.47% 降为 2.56%, 但 SV-Shenk's 和 FV-Shenk's 均比全波段传递后的结果差, 可能由于单一变量和融合变量中各波长点在窗口光谱段内样本间、仪器间的信号不一致性明显, 致使 Shenk's 中的波长点重构和插值得到的光谱矩阵误差较大, 进而导致预测结果不精确; 经 DS 和 SB 算法传递后, 基于变量优选所建模型 ($R_p^2 = 0.47 \sim 0.61$, RMSEP = 1.60% ~ 1.90%) 的预测精度明显好于全波段所建模型 ($R_p^2 = 0.03 \sim 0.34$, RMSEP = 2.44% ~ 4.67%), 表明变量优选后有效降低了全波段光谱信息中无关信息的干扰, 并适度降低了光谱信息间的差异。

SV-MC, FV-MC 与 CV-MC 算法的步骤相同, 区别在于输入量由共性变量分别变为单一变量和融合变量。SV-MC, FV-MC 和 CV-MC 算法进行模型传递后的预测结果分别优于相同变量的其他方法, RMSEP 降至 1.30% ~ 1.80%, 明显好于全波段所建模型传递后的结果, 且 CV-MC 算法的传递结果 ($R_p^2 = 0.73$, RMSEP = 1.30%) 优于 SV-MC 和 FV-MC。

综合 3 类变量的模型传递结果, 共性变量所建模型不仅简化了模型, 提高了主仪器的预测精度, 而且传递结果好于单一变量、融合变量和全谱区数据。CV-WC 算法传递后, RMSEP 由传递前的 7.49% 降为 1.62%, 传递结果与 CV-DS (RMSEP = 1.62%) 和 CV-S/B (RMSEP = 1.60%) 相近, 得到较好的预测结果, 表明 CV-WC 算法也是一种有效的模型传递方法。因此, 本研究采用的 CV-MC 和 CV-WC 算法所需输入变量少, 将随机选取的校正集样本作为转换样品, 采用原始光谱建模时得到好的传递结果, 为鲜枣品质检测和检测

模型在不同仪器间共享的实现提供了基础。在接下来的研究中, 可进一步优化标样选择算法和光谱预处理方法以提高模型的预测能力。

3 结 论

以壶瓶枣为研究对象, 分别采用 M1 和 M2 两台光谱仪采集样本的可见/近红外光谱信息, 采用多种算法进行 SSC 检测模型传递的研究, 得到以下结论:

(1) 利用全波段光谱信息建立 LS-SVM 模型, 预测同一台仪器采集的光谱时, 原始光谱好于 SG-1st Der 和 SNV 处理后所建模型, 但直接预测不同仪器所采集光谱时结果均较差; 基于原始光谱采用 Shenk's, DS 和 S/B 算法进行模型传递的结果均不理想。

(2) 根据 RC 提取主仪器和从仪器的特征波长, 优选出单一变量 24 个、共性变量 23 个和融合变量 29 个, 分别建立 SSC 的 LS-SVM 模型; 主仪器所建模型预测主仪器所采集光谱时 R_p^2 为 0.78 ~ 0.80, RMSEP 为 1.07% ~ 1.13%, 建模精度明显较全波段好, 均可较好的解释鲜枣样本的变量信息, 但对从仪器预测均失效 (RMSEP = 6.62% ~ 7.88%)。

(3) 利用提出的 SV-MC, CV-MC, FV-MC 和 CV-WC 算法进行 SSC 检测模型传递, 共性变量所建模型的传递结果好于单一变量、融合变量和全谱区数据, CV-WC 传递后的预测结果 (RMSEP = 1.62%) 与 CV-DS 和 CV-S/B 相近; CV-MC 得到最优预测结果, $R_p^2 = 0.73$, RMSEP = 1.30%, 为不同仪器间的模型传递及鲜枣品质检测模型的开发应用提供了基础。

References

- [1] LI Xin, BIN Jun, FAN Wei, et al (李鑫, 宾俊, 范伟, 等). Chinese Journal of Analytical Chemistry (分析化学), 2017, 45(7): 958.
- [2] Qin Yuhua, Gong Huili. Infrared Physics & Technology, 2016, 77: 239.
- [3] Mou Yi, Zhou Long, Yu Shujian, et al. Chemometrics and Intelligent Laboratory Systems, 2016, 156: 62.
- [4] LIU Jiao, LI Xiao-yu, GUO Xiao-xu, et al (刘娇, 李小昱, 郭小许, 等). Transactions of the Chinese Society of Agricultural Engineering (农业工程学报), 2014, 30(17): 276.
- [5] WENG Hai-yong, CEN Hai-yan, HE Yong (翁海勇, 岑海燕, 何勇). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2018, 38(1): 235.
- [6] Liu Yan, Cai Wensheng, Shao Xueguang. Spectrochimica Acta Part A Molecular and Biomolecular Spectroscopy, 2016, 169: 197.
- [7] Liang Chen, Yuan Hongfu, Zhao Zhong, et al. Chemometrics and Intelligent Laboratory Systems, 2016, 153: 51.
- [8] XU Hui-rong, LI Qing-qing (徐惠荣, 李青青). Transactions of the Chinese Society for Agricultural Machinery (农业机械学报), 2017, 48(9): 312.
- [9] Zhang Shujuan, Zhang Haihong, Zhao Yanru, et al. Mathematical and Computer Modelling, 2013, 58(3-4): 545.
- [10] Kennard R W, Stone L A. Technometrics, 1969, 11(1): 137.
- [11] Yao Yang, Chen Huarui, Xie Lijuan, et al. Journal of Food Engineering, 2013, 119(1): 22.
- [12] NI Li-jun, HAN Ming-yue, ZHANG Li-guo, et al (倪力军, 韩明月, 张立国). Chinese Journal of Analytical Chemistry (分析化学), 2018, 46(10): 1660.
- [13] Travers S, Bertelsen M G, Petersen K K, et al. LWT-Food Science and Technology, 2014, 59(2): 1107.
- [14] Jerry Workman Jr, Lois Weyer. Practical Guide to Interpretive Near-Infrared Spectroscopy (近红外光谱解析实用指南). Translated by CHU Xiao-li, XU Yu-peng, TIAN Gao-you (褚小立, 许育鹏, 田高友, 译). Beijing: Chemical Industry Press (北京: 化学工业出版社), 2009. 19.

Model Transfer Method of Fresh Jujube Soluble Solids Detection Using Variables Optimization and Correction Algorithms

SUN Hai-xia, ZHANG Shu-juan*, XUE Jian-xin, ZHAO Xu-ting, XING Shu-hai, CHEN Cai-hong, LI Cheng-ji
College of Engineering, Shanxi Agricultural University, Taigu 030801, China

Abstract The difficulty of sharing detection models built by different instruments is common in the quality inspection and classification of fruits. In the study, the “Huping” jujube was used as the research object, and the transfer method of the soluble solids content (SSC) detection model between instruments was explored using visible/near infrared spectroscopy. First, the spectral information of samples was collected using two instruments produced by American Analytical Spectral Device. Based on the original, Savitzky-Golay first derivative processed and standard normal variable transformed spectrum, SSC detection models were established by least squares-support vector machines (LS-SVM), respectively. The prediction ability of the three methods for the spectra acquired by different instruments was poor. The built model by the original spectrum of the master instrument was optimal in predicting spectra from the same instrument. The determination coefficient (R_p^2) and the root mean squared error of prediction (RMSEP) were 0.73 and 1.36%, respectively. Next, the Kennard/Stone algorithm was used to select standard samples. The Shenk’s, direct standardization (DS) and slope/bias (S/B) algorithm were used for model transfer, respectively. Then, according to the regression coefficient, the sensitive wavelengths of the master instrument (24) and the slave instrument (28) were extracted. 24 single variables (SV), 23 common variables (CV) and 29 fusion variables (FV) were selected, all of which contained the main absorption bands of SSC. LS-SVM detection models of the master instrument were respectively established by the preferred variables, which ($R_p^2=0.78\sim0.80$, RMSEP=1.07%~1.13%) was better than the model built by the full wavelength for the prediction result of the master instrument. However, the model failed in predicting spectra from different instruments (RMSEP=6.62%~7.88%). Finally, based on the wavelength position shift and the absorbed property of molecular vibration, these algorithms named as common variable-subtraction correction (CV-MC), single variable-subtraction correction, fusion variable-subtraction correction and common variable-wavelength correction (CV-WC) were respectively proposed for model transfer. These methods were compared with SV-Shenk’s, CV-Shenk’s, FV-Shenk’s, SV-DS, CV-DS, FV-DS, SV-S/B, CV-S/B and FV-S/B algorithms. The results showed that the prediction results ($R_p^2=0.03\sim0.34$, RMSEP=2.44%~4.67%) were poor when the model was transferred by the full-band. Using the model built by the preferred variables, the results transferred by SV-Shenk’s, CV-Shenk’s and FV-Shenk’s were poor, and the results transferred by other algorithms ($R_p^2=0.47\sim0.73$, RMSEP=1.30%~1.90%) were better than the full wavelength. The CV got better transfer results than the SV and the FV, and the CV-MC result was the best ($R_p^2=0.73$, RMSEP=1.30%). The predicted result after CV-WC transfer (RMSEP=1.62%) was similar to CV-DS and CV-S/B. The research indicates that both CV-MC and CV-WC are effective model transfer algorithms, which are of great significance to establishing a common jujube quality detection model between different instruments.

Keywords Near/infrared spectroscopy; Calibration transfer; Jujube; Non-destructive detection

(Received Nov. 28, 2018; accepted Mar. 1, 2019)

* Corresponding author