

近红外光谱快速检测马铃薯全粉还原糖

孙旭东¹, 董小玲²

(1. 华东交通大学机电工程学院, 南昌 330013; 2. 华东交通大学外国语学院, 南昌 330013)

摘 要: 还原糖含量是评价马铃薯全粉品质的重要指标之一, 该文研究基于近红外光谱技术结合最小二乘支持向量机 (least squares support vector machine, LSSVM) 算法的马铃薯全粉还原糖含量非线性数学模型。采用移动窗口偏最小二乘法 (moving windows partial least square, MWPLS) 和连续投影算法 (successive projections algorithm, SPA) 组合方法筛选出 20 个特征变量, 作为 LSSVM 的输入向量。优化径向基函数 (radial basis function, RBF) 的惩罚因子和核参数, 训练 LSSVM 校正模型。经比较, LSSVM 校正模型预测结果最优, 预测相关系数为 0.984, 预测标准差为 0.223%, 相对分析误差 (standard deviation ratio, SDR) 为 5.62。结果表明: 近红外光谱结合 LSSVM 算法提高了马铃薯全粉还原糖含量的预测精度。

关键词: 近红外光谱, 无损检测, 算法, 变量选择, 马铃薯, 还原糖

doi: 10.3969/j.issn.1002-6819.2013.14.033

中图分类号: TP374⁺.52; TS235.2

文献标志码: A

文章编号: 1002-6819(2013)-14-0262-07

孙旭东, 董小玲. 近红外光谱快速检测马铃薯全粉还原糖[J]. 农业工程学报, 2013, 29(14): 262-268.

Sun Xudong, Dong Xiaoling. Rapid detection of reducing sugar for potato granules by near infrared spectroscopy [J]. Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE), 2013, 29(14): 262-268. (in Chinese with English abstract)

0 引 言

还原糖含量是影响马铃薯全粉加工产品品质的重要因素之一。质量分数超过 0.25% 时, 还原糖与 α -氨基酸发生美拉德, 导致马铃薯加工产品颜色褐变, 影响了马铃薯加工产品的品质^[1]。还原糖传统化学分析法, 例如容量分析法、比色法、电化学法等, 虽然检测精度高, 但耗时、费力、消耗化学药品, 难以满足马铃薯加工产业快速发展的需求。探讨马铃薯全粉还原糖含量快速检测方法, 调控马铃薯加工产品品质, 具有重要的应用前景。

近红外光谱技术结合化学计量学方法实现了物料成分的快速定量检测^[2]。国内外相关研究人员成功应用近红外光谱技术检测薯片的脂肪和水分含量^[3], 马铃薯块茎水分、淀粉、蛋白质、还原糖和粗纤维含量^[4-5], 马铃薯块茎钾含量^[6], 鉴别薯片品牌^[7]。物料成分与近红外光谱间服从比尔定律, 呈线性关系^[8-9]。但光谱采集过程中的杂散光、温度等因素易导致物料成分与光谱间的规律偏离比尔定律, 呈现非线性关系^[10]。基于统计学习理论的机

器学习算法支持向量机 (support vector machine, SVM), 在解决近红外光谱的非线性问题方面已表现出优势^[11-12]。最小二乘支持向量机 (least squares support vector machine, LSSVM) 改进了传统的支持向量机算法, 采用解线性方程组替代二次规划求解问题, 提高了求解的速度和收敛精度。

试验中采用移动窗口偏最小二乘法 (moving windows partial least square, MWPLS) 结合遗传算法 (genetic algorithm, GA)、连续投影算法 (successive projections algorithm, SPA) 分 2 步筛选近红外光谱特征变量, 建立了马铃薯全粉还原糖含量的 LSSVM 校正模型, 解决光谱与还原糖含量间的非线性问题, 提高模型的预测精度。

1 材料与方 法

1.1 试验材料

马铃薯全粉由内蒙古某食品公司提供, 共 110 份样品, 每份 100 g, 均采用聚乙烯 (polyethylene, PE) 真空包装袋密封包装。样品寄至实验室后, 于约 20℃ 室温中放置 24 h, 待达到室温后, 拆开包装采集近红外光谱。试验中共采用 110 份马铃薯全粉样品, 依据校正集样品还原糖含量变化范围大于预测集样品且均匀分布的原则, 大致按照 3:1 的比例, 将光谱差异较大的 83 份样品选入校正集, 其余较相近的样品选入预测集, 建模集和预测集马铃薯样品还原糖含量统计值见表 1。

收稿日期: 2013-03-26 修订日期: 2013-06-27

基金项目: 国家自然科学基金 (61240018); 江西省科技支撑计划 (20121BBF60054); 江西省教育厅青年基金项目 (GJJ12317)

作者简介: 孙旭东 (1978-), 男, 吉林辽源人, 主要从事农产品无损检测研究。南昌 华东交通大学机电工程学院, 330013。

Email: sunxudong_18@163.com

表 1 马铃薯全粉样品还原糖含量
Table 1 Reducing sugar in potato granules

	样本数 Sample number	最大值 Maximum/%	最小值 Minimum/%	平均值 Mean/%	标准偏差 Standard deviation
校正集 Calibration set	83	6.06	0.27	2.59	1.33
预测集 Prediction set	27	6.06	0.38	2.60	1.25

1.2 试验仪器

试验仪器采用 NIRFlex N-500 傅立叶近红外光谱仪 (BUCHI Labortechnik AG, Switzerland), 配备 Peltier 冷却铟镓砷检测器。光谱范围为 $10\,000\sim 4\,000\text{ cm}^{-1}$, 分辨率为 4 cm^{-1} , 波长精度为 $\pm 0.2\text{ cm}^{-1}$ 。光谱仪配有培养皿、反射板测样附件。

1.3 数据获取

马铃薯全粉样品拆开包装后, 置于培养皿中压实, 盖上反射板, 防止光从样品中溢出。采用漫反射方式, 光斑直径约 2 cm 。培养皿测样附件旋转一周, 扫描 32 次, 记录近红外光谱, 每份样品重复采谱 3 次, 取平均吸光度光谱作为试验数据, 共 110 条光谱。

马铃薯全粉还原糖含量真实值由供样公司提供, 采用 3, 5-二硝基水杨酸比色法测定。称取马铃薯全粉 10 g 加入三角瓶, 加入 80 mL 蒸馏水混匀。 50°C 恒温水浴保温 20 min , 不断搅拌浸出还原糖, 以 $12\,000\text{ r/min}$ 离心 10 min 。取上层清液 1 mL 加入 1% 的 3, 5-二硝基水杨酸溶液 1 mL , 混合均匀, 在分光光度计下测定 540 nm 处吸光度值, 查标准曲线计算还原糖含量^[13]。

1.4 算法和模型评价

MWPLS 是一种波长区间优选方法, 该方法以一定宽度的窗口移过全谱区间, 在每个窗口区间下分别建立偏最小二乘 (partial least square, PLS) 模型, 区间的交叉验证均方根误差 (root mean square error of cross validation, RMSECV) 值小于全谱 RMSECV 的区间入选^[14-15]。GA 应用生物物种进化机制, 以适应度函数为依据, 对全谱中某个光谱变量施加选择、交叉、变异操作, 实现光谱变量优化选择^[16-17]。SPA 计算光谱矩阵中某一变量对其它变量的投影, 根据投影信息量的多少决定入选变量^[18-20]。PLS 是基于线性回归方式的多元校正方法, 通过光谱矩阵 X 的得分矩阵 T 和浓度矩阵 Y 的得分矩阵 U 间的线性回归, 建立 PLS 模型。LSSVM 通过定义适当的核函数实现非线性变化, 然后寻找最优拟合曲线。合理选择径向基函数 (radial basis function, RBF) 的惩罚因子 γ 和核参数 σ^2 参数, 可提高 LSSVM 的机器泛化能力^[21-23]。

MWPLS、GA、SPA 和 LSSVM 均在 Matlab R2010a 软件中完成。PLS 建模及模型预测均在 Uscrmblr 8.0 软件中完成。采用预测相关系数 R_p 、校正标准差 (standard error of calibration, SEC) 和预测标准差 (standard error of prediction, SEP) 和相对分析误差 (standard deviation ratio, SDR) 评价模型性能, 其中 SDR 为标准差与预测标准差 SEP 的比值^[24], R_p 和 SDR 值越高模型预测能力越强, SDR 一般要求在 3 以上。

2 结果与讨论

2.1 近红外光谱特性

有机物在近红外谱区的吸收峰主要由含氢基团各级倍频与合频振动引起, 绝大多数含氢基团振动信息分布在 $2\,000\text{ cm}^{-1}$ 以上。图 1 为马铃薯全粉近红外吸光度光谱, 在 $4\,000\sim 8\,000\text{ cm}^{-1}$ 谱区范围内, 3 条光谱分别是校正集中还原糖含量最小 (0.27%)、最大 (6.06%) 和中间值 (2.58%) 马铃薯全粉样品的光谱, 马铃薯还原糖含量由小到大, 对应的光谱吸收峰由弱到强; 在 $8\,000\text{ cm}^{-1}$ 以后不完全服从这一规律, 在特征变量选择中, $8\,000\text{ cm}^{-1}$ 以后谱区中的变量也很少被选中。马铃薯全粉近红外光谱吸收峰主要是由近红外光引起各种含氢基团振动引起的, 具体的谱峰归属及来源见表 2。

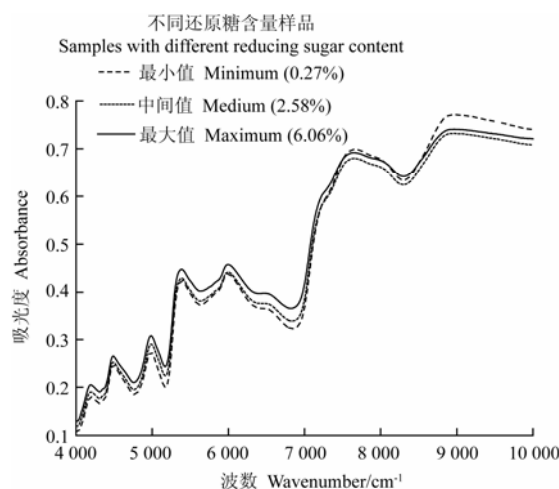


图 1 马铃薯全粉近红外吸光度光谱
Fig.1 Absorbance spectra of potato granules

表 2 马铃薯全粉近红外光谱对应的谱带归属及来源

Table 2 Bands attribution corresponding to near infrared spectra of potato granules and source

波数 Wave number/cm ⁻¹	振动方式 Vibration mode	基团 Group	波数 Wave number/cm ⁻¹	振动方式 Vibration mode	基团 Group
4200	3×O-H def.	纤维素	6876	2×C-O str.	糖类
4472	O-H str. + O-H def.	淀粉	7652	3×C-H str.	CH
4976	N-H sym. str.	CONH ₂	8816	3×C-H str.	CH ₃
5372	O-H str. + 2×C-O str.	淀粉			

注：2×，一级倍频；3×，二级倍频；str.，伸缩振动；def.，变形振动；sym.，对称振动。
Note: 2×, the first overtone; 3×, the second overtone; str., stretch vibration; def., deformation vibration; sym., symmetric vibration.

2.2 特征光谱区间定位

马铃薯全粉近红外光谱既包含还原糖目标信息，同时也受到纤维素、淀粉、蛋白质等非目标或仪器噪声的干扰，而且还还原糖对应的近红外光谱信息分布复杂、区间宽度不等。移动窗口偏最小二乘法，在一定交叉验证均方根误差（RMSECV）水平下，可获得还原糖含量对应的近红外光谱若干个信息区间^[25]。本文采用与文献 25 相同的步骤运行移动窗口偏最小二乘法，运行结果如图 2 所示。将全谱分成若干个等间隔区间，不断增加间隔数量，直到对选择结果无改善为止，确定将全谱分成 31 个子区间。在每个子区间内，采用 PLS 分别建立还原糖校正模型，如某子区间的 RMSECV 小于平均值，则该子区间入选。合并相邻子区间，即为特征谱区，共 470 个变量，具体选择结果见表 3。

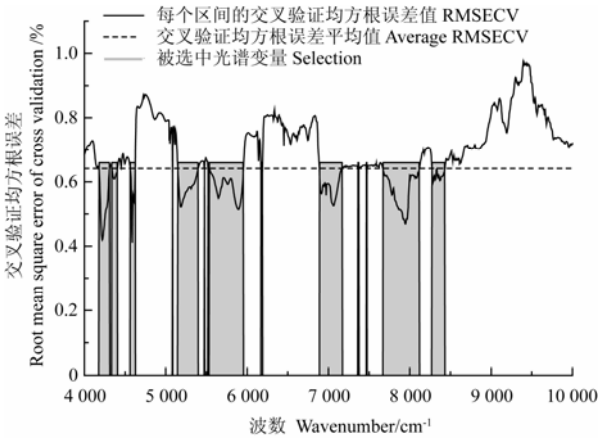


图 2 马铃薯全粉近红外特征谱区移动窗口偏最小二乘法选择结果

Fig.2 Characteristic bands of potato granules selected by moving windows partial least square

表 3 不同输入变量的 PLS 建模及模型预测结果

Table 3 Results of different PLS calibration and prediction models with different input variables

变量选择方法 Method of variable selection			回归方法 Regression method		建模集 Calibration set		预测集 Prediction set		
方法 Method	变量 Variables	变量数 Variable number	算法 Algorithm	因子数 Factors	相关系数 R _c Correlation coefficient	校正标准差 Standard error of calibration /%	相关系数 R _p Correlation coefficient	预测标准差 Standard error of prediction /%	相对分析误差 Standard deviation ratio
无 None	4000~10000	1501	PLS	9	0.961	0.368	0.961	0.361	3.474
MWPLS	4188~4308, 4336~4404, 4564~4624, 5084, 5148~5396, 5476, 5520, 5536~5952, 6172~6188, 6892~7168, 7364~7372, 7468~7472, 7672~8116, 8272~8428	470	PLS	9	0.976	0.318	0.963	0.346	3.624
MWPLS-GA	4284, 4352, 4596, 5324, 5604, 5716, 5832-2836, 6176-6180, 6944~6952, 7084~7088, 7696, 7856~7860, 8112, 8360, 8400~8408, 8424	24	PLS	9	0.976	0.290	0.970	0.306	4.098
MWPLS-SPA	4224, 4336, 4404, 5084, 5176, 5236, 5268, 5316, 5776, 5876, 6188, 6900, 6968, 7116, 7972, 7996, 8288, 8304, 8336, 8360	20	PLS	9	0.980	0.262	0.976	0.273	4.593

注：MWPLS（moving windows partial least square）为移动窗口偏最小二乘法，MWPLS-GA（genetic algorithm）为移动窗口偏最小二乘法-遗传算法，MWPLS-SPA（successive projections algorithm）为移动窗口偏最小二乘法-连续投影算法，PLS（partial least square）为偏最小二乘。下同。

2.3 特征光谱变量选择

分别采用 GA 和 SPA 在初步定位的特征谱区内，筛选特征变量。GA 采用 0/1 二进制编码，群体数目为 30，杂交概率为 50%，变异概率为 1%，

进化代数 and 运行轮数均为 100，根据 RMSECV 最小的原则，在运行轮数内，出现频率高的波长点被选中，遗传算法运行结果如图 3 所示。将相邻光谱变量合并，共包含 24 个光谱变量，具体变量如表 3

所示。利用遗传算法筛选出的 24 个光谱变量，建立 PLS 校正模型，并进行预测，预测标准差为 0.306%，略优于 MWPLS 的预测结果。遗传算法根据运行轮数内变量被选频率的高低筛选变量，由表 3 可知，6 组相邻的变量被选中。相邻变量之间存在较大的共线性，会影响模型的预测能力。

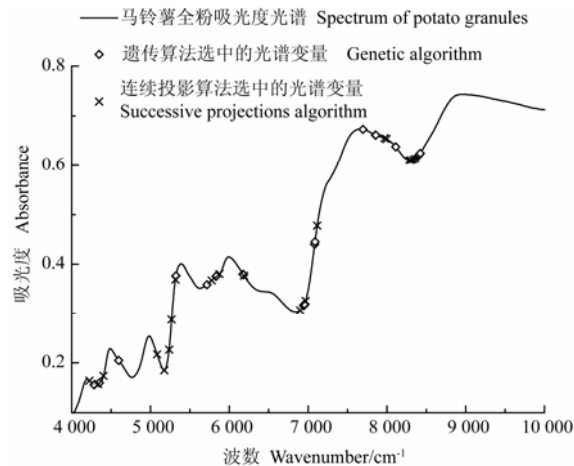


图 3 特征变量 GA 和 SPA 选择结果

Fig.3 Characteristic variables selected by GA and SPA

SPA 设定筛选最小变量数为 20，最大变量数目

为 50，以 RMSECV 最小为原则，选出最优的变量。SPA 运行结果如图 3 所示，共 20 个变量被选中，具体结果见表 3。利用 SPA 筛选的 20 个变量，建立 PLS 校正模型，并进行预测，与 GA 相比，结果略优，预测标准差为 0.273%。SPA 克服了 GA 易选中相邻变量的缺点，较好地避免了相邻变量的共线性问题。

采用不同方法筛选的变量建立 PLS 校正模型，并采用 27 个未参与建模的样品预测模型性能，结果见表 3。变量筛选提高了 PLS 模型的预测能力，且采用 MWPLS 和 SPA 组合筛选的 20 个变量建立的 PLS 校正模型预测精度最高，预测相关系数 R_p 为 0.976，SDR 为 4.593，SDR 大于 3，模型具有实用价值和较强的预测能力^[26]。

2.4 非线性模型建立及预测

RBF 核函数的惩罚因子 γ 控制样本偏差与机器泛化能力间的平衡，核参数 σ^2 控制其欠学习或过学习^[27]。采用二步格点搜索法和留一法交叉验证相结合，进行惩罚因子 γ 和核参数 σ^2 的全局寻优^[28-29]。第 1 步在 10×10 格点中以较大步长初选；第 2 步在粗选基础上，在 10×10 格点中以较小步长搜索，确定最优参数。试验中 γ 和 σ^2 的搜索范围分别为 $1 \sim 10\,000$ 和 $0.01 \sim 100$ 。不同的 LSSVM 模型对应的最优 γ 和 σ^2 见表 4。

表 4 不同输入变量的 LSSVM 模型预测结果

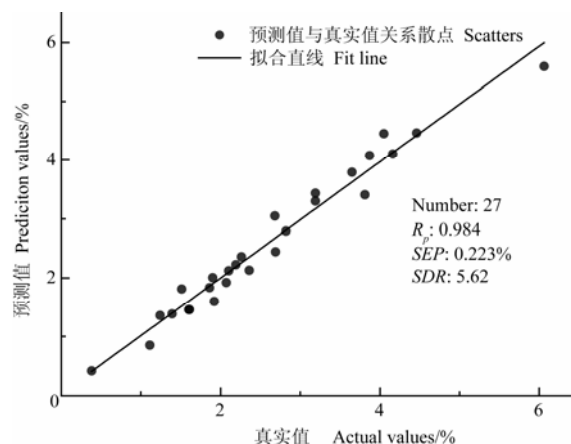
Table 4 Results of different LSSVM prediction models with different input variables

变量选择方法 Method of variable selection		训练方法 Method of training		参数 Parameters			预测集 Prediction set	
方法 Method	变量数 Variable number	方法 Method	时间 Time/s	惩罚因子 γ Penalty factor	核参数 σ^2 Nuclear parameter	相关系数 R_p Correlation coefficient	标准差 Standard error of prediction /%	相对分析误差 Standard deviation ratio/%
无 None	1501	LSSVM	0.081	882	23	0.980	0.253	4.96
MWPLS	470	LSSVM	0.043	882	23	0.981	0.247	5.08
MWPLS-GA	24	LSSVM	0.026	2482	20	0.982	0.240	5.23
MWPLS-SPA	20	LSSVM	0.026	364	26	0.984	0.223	5.62

注：LSSVM (least squares support vector machine) 为近红外光谱技术结合最小二乘支持向量机。

分别以全谱 1501 个变量和筛选的 470、24 和 20 个特征变量作为 LSSVM 的输入向量，以 RBF 为核函数，检验 LSSVM 模型的预测能力，结果见表 4。由表 4 可知，通过筛选特征变量作为 LSSVM 的输入向量，提高了模型的预测能力，其 SDR 均高于 4.96；提高了 LSSVM 的收敛速度，减少了模型训练时间，均低于 0.081 s。对比表 4 和表 3，由具有较高 SDR 模型的预测能力强的评价原则，LSSVM 模型的预测能力均强于 PLS 模型。

采用 27 个未参与建模的样品考察 LSSVM 模型的预测能力。经比较，以 MWPLS 和 SPA 组合筛选的 20 个变量作输入变量，以 γ 为 364 和 σ^2 为 26 的 RBF 核函数的 LSSVM 模型预测能力最强，模型预测相关系数 R_p 为 0.984，预测标准差 SEP 为 0.223%，SDR 为 5.62。LSSVM 预测模型的预测值与知识值之间的相关关系如图 4 所示。



注： R_p 为预测相关系数，SEP 为预测标准差，SDR 为相对分析误差。
Note: R_p , correlation coefficient of prediction, SEP, standard error of prediction; SDR, standard deviation ratio.

图 4 LSSVM 模型预测结果

Fig.4 Prediction results of LSSVM model

3 结 论

采用近红外光谱技术, 结合 LSSVM 机器学习算法, 探讨马铃薯全粉还原糖含量快速检测方法。经过变量筛选和 RBF 核函数参数优化, 建立的马铃薯全粉还原糖含量 LSSVM 校正模型预测结果最优, 模型预测相关系数为 0.984, 预测标准差为 0.223%, SDR 为 5.62。结果表明: MWPLS 结合 SPA 的变量筛选方法, 提高了 LSSVM 校正模型的预测精度, 克服了近红外光谱变量与还原糖含量间非线性因素的影响。

[参 考 文 献]

- [1] 谢庆华, 吴毅歆. 马铃薯品种营养成分分析测定[J]. 云南师范大学学报, 2002, 22(2): 50—52.
Xie Qinghua, Wu Yixin. Analyze to natlition constituent of potato[J]. Journal of Yunnan Normal University, 2002, 22(2): 50—52. (in Chinese with English abstract)
- [2] Nicolai B M, Beullens K, Bobelyn E, et al. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: A review[J]. Postharvest Biology and Technology, 2007, 46(2): 99—118.
- [3] Cecilia S, Luis R. Application of NIR and MIR spectroscopy in quality control of potato chips[J]. Journal of Food Composition and Analysis, 2009, 22(6): 596—605.
- [4] 张小燕, 刘威, 兴丽, 等. 马铃薯加工品质指标近红外预测模型研究[J]. 红外, 2012, 33(12): 33—39.
Zhang Xiaoyan, Liu Wei, Xing Li, et al. An near-infrared prediction model for quality indexes of potato processing[J]. Infrared, 2012, 33(12): 33—39. (in Chinese with English abstract)
- [5] 刘波平, 秦华俊, 罗香, 等. 基于偏最小二乘与广义回归神经网络的近红外光谱测定土豆中 3 种营养成分的研究[J]. 分析实验室, 2007, 26(9): 38—41.
Liu Boping, Qin Huajun, Luo Xiang, et al. Study on the determination of three components in potatoes using near infrared spectroscopy based on partial least squares and generalized regression neural network model[J]. Chinese Journal of Analysis Laboratory, 2007, 26(9): 38—41. (in Chinese with English abstract)
- [6] 刘翠翠, 高红秀, 李赞, 等. 马铃薯块茎钾含量近红外模型的建立[J]. 中国马铃薯, 2011, 25(2): 65—68.
Liu Cuicui, Gao Hongxiu, Li Zan, et al. Near-infrared model establishment for testing potato tubers potassium content[J]. Chinese Potato Journal, 2011, 25(2): 65—68. (in Chinese with English abstract)
- [7] Ni Y, Mei M, Kokot S. Analysis of complex, processed substances with the use of NIR spectroscopy and chemometrics: Classification and prediction of properties: The potato crisps example[J]. Chemometrics and Intelligent Laboratory Systems, 2011, 105(2): 147—156.
- [8] 褚小立. 化学计量学方法与分子光谱分析技术[M]. 北京: 化学工业出版社, 2011.
- [9] 严衍禄, 赵龙莲, 韩东海, 等. 近红外光谱分析基础与应用[M]. 北京: 中国轻工业出版社, 2005.
- [10] 杨辉华, 覃锋, 王义明, 等. NIR 光谱的 Isomap-PLS 非线性建模方法[J]. 光谱学与光谱分析, 2009, 29(2): 322—326.
Yang Huihua, Qin Feng, Wang Yiming, et al. Isomap-PLS nonlinear modeling method for near infrared spectroscopy[J]. Spectroscopy and Spectral Analysis, 2009, 29(2): 322—326. (in Chinese with English abstract)
- [11] 金叶, 杨凯, 吴永江, 等. 基于粒子群算法的最小二乘支持向量机在红花提取液近红外定量分析中的应用[J]. 分析化学, 2012, 40(6): 925—931.
Jin Ye, Yang Kai, Wu Yongjiang, et al. Application of particle swarm optimization based least square support vector machine in quantitative analysis of extraction solution of safflower using near-infrared spectroscopy[J]. Chinese Journal of Analytical Chemistry, 2012, 40(6): 925—931. (in Chinese with English abstract)
- [12] 黄光群, 韩鲁佳. 基于非线性径向基核函数支持向量机的堆肥产品近红外光谱分析研究[J]. 光学学报, 2009, 29(12): 3556—3560.
Huang Guangqun, Han Lujia. Near infrared reflectance spectroscopy analysis of compost products using nonlinear support vector machine with RBF nucleus[J]. Acta Optica Sinica, 2009, 29(12): 3556—3560. (in Chinese with English abstract)
- [13] 朱海霞, 石瑛, 张庆娜, 等. 3,5-硝基水杨酸 (DNS) 比色法测定马铃薯还原糖含量的研究[J]. 中国马铃薯, 2005, 19(5): 266—269.
Zhu Haixia, Shi Ying, Zhang Qingna, et al. Applying 3,5-dinitrosalicylic acid (DNS) method to analyzing the content of potato reducing sugar[J]. Chinese Potato Journal, 2005, 19(5): 266—269. (in Chinese with English abstract)
- [14] Zou X B, Zhao J W, Malcolm J W P, et al. Variables selection methods in near-infrared spectroscopy[J]. Analytica Chimica Acta, 2010, 667(1/2): 14—32.
- [15] 王加华, 李鹏飞, 曹楠宁, 等. 基于 iPLS 原理最优化信息区间的桃糖度组合权重 PLS 模型研究[J]. 红外与红米波学报, 2009, 28(5): 386—341.
Wang Jiahua, Li Pengfei, Cao Nanning, et al. Study on the combination weight PLS model for determining SSC of peach based on the optimal information regions

- obtained from iPLS methods[J]. *J. Infrared Millim. Waves*, 2009, 28(5): 386—341. (in Chinese with English abstract)
- [16] 邹小波, 赵杰文. 用遗传算法快速提取近红外光谱特征区域和特征波长[J]. *光学学报*, 2007, 27(7): 1316—1321.
- Zou Xiaobo, Zhao Jiewen. Methods of Characteristic wavelength region and wavelength selection based on genetic algorithm[J]. *Acta Optica Sinica*, 2007, 27(7): 1316—1321. (in Chinese with English abstract)
- [17] 陈斌, 王豪, 林松, 等. 基于相关系数法与遗传算法的啤酒酒精度近红外光谱分析[J]. *农业工程学报*, 2005, 21(7): 99—102.
- Chen Bin, Wang Hao, Lin Song, et al. Analysis of near infrared spectroscopy of beer alcohol content by correlation coefficients and genetic algorithms[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2005, 21(7): 99—102. (in Chinese with English abstract)
- [18] 吴迪, 吴洪喜, 蔡景波, 等. 基于无信息变量消除法和连续投影算法的可见-近红外光谱技术白虾种分类方法研究[J]. *红外与毫米波学报*, 2009, 28(6): 423—427.
- Wu Di, Wu Hongxi, Cai Jingbo, et al. Classifying the species of *exopalaemon* by using visible and near infrared spectra with uninformaton variable elimination and successive projections algorithm[J]. *J. Infrared Millim. Waves*, 2009, 28(6): 423—427. (in Chinese with English abstract)
- [19] 钱海波, 孙来军, 王乐凯, 等. 基于连续投影算法的小麦湿面筋近红外校正模型优化[J]. *中国农学通报*, 2011, 27(18): 51—56.
- Qian Haibo, Sun Laijun, Wang Lekai, et al. Near infrared spectroscopy calibration model optimizing of wet gluten based on successive projections algorithm[J]. *Chinese Agricultural Science Bulletin*, 2011, 27(18): 51—56. (in Chinese with English abstract)
- [20] 宦克为, 刘小溪, 郑峰, 等. 基于蒙特卡罗特征投影法的小麦蛋白质近红外光谱测量变量选择[J]. *农业工程学报*, 2013, 29(4): 266—271.
- Huan Kewei, Liu Xiaoxi, Zheng Feng, et al. Variable selection of near-infrared spectroscopy for measuring wheat protein based on MC-LPG[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(4): 266—271. (in Chinese with English abstract)
- [21] Suykens J A K, Gestel T V, Brabanter J D, et al. Least squares support vector machines[M]. Singapore: World Scientific Publishing Co., 2002.
- [22] Chauchard F, Cogdill R, Roussel S, et al. Application of LS-SVM to non-linear phenomena in NIR spectroscopy: development of a robust and portable sensor for acidity prediction in grapes[J]. *Chemometrics and Intelligent Laboratory Systems*, 2004, 71(2): 141—150.
- [23] 谈爱玲, 毕红卫. 基于 KPCA 和 LSSVM 的蜂蜜近红外光谱鉴别分析[J]. *激光与红外*, 2011, 41(12): 1331—1336.
- Tan Ailing, Bi Hongwei. Identification of honey by NIR spectroscopy technology based on KPCA and LSSVM[J]. *Laser and Infrared*, 2011, 41(12): 1331—1336. (in Chinese with English abstract)
- [24] McGlone V A, Jordan R B, Martinsen P J. Vis/NIR estimation at harvest of pre- and post-storage quality indices for 'Royal Gala' apple[J]. *Postharvest Biology and Technology*, 2002, 25(2): 135—144.
- [25] 曹楠宁, 王加华, 李鹏飞, 等. 基于 GA 和 SCMWPLS 算法的 NIR 光谱信息变量提取研究[J]. *光谱学与光谱分析*, 2010, 30(4): 915—919.
- Cao Nanning, Wang Jiahua, Li Pengfei, et al. Study on variable selection of NIR spectral information based on GA and SCMWPLS[J]. *Spectroscopy and Spectral Analysis*, 2010, 30(4): 915—919. (in Chinese with English abstract)
- [26] McGlone V A, Kawano S. Firmness, dry-matter and soluble-solids assessment of postharvest kiwifruit by NIR spectroscopy[J]. *Postharvest Biology and Technology*, 1998, 13(2): 131—141.
- [27] 吴桂芳, 何勇. 基于主成分分析和支持向量机的山羊绒原料品种鉴别分析[J]. *光谱学与光谱分析*, 2009, 29(6): 1541—1544.
- Wu Guifang, He Yong. Identification of varieties of cashmere by vis/NIR spectroscopy technology based on PCA-SVM[J]. *Spectroscopy and Spectral Analysis*, 2009, 29(6): 1541—1544. (in Chinese with English abstract)
- [28] 虞科, 程翼宇. 一种基于最小二乘支持向量机算法的近红外光谱判别分析[J]. *分析化学*, 2006, 34(4): 561—564.
- Yu Ke, Cheng Yiyu. Discriminating the genuineness of Chinese medicines with least square support vector machines[J]. *Chinese Journal of Analytical Chemistry*, 2006, 34(4): 561—564. (in Chinese with English abstract)
- [29] 梁栋, 管青松, 黄文江, 等. 基于支持向量机的冬小麦叶面积指数遥感反演[J]. *农业工程学报*, 2013, 29(7): 117—123.
- Liang Dong, Guan Qingsong, Huang Wenjiang, et al. Remote sensing inversion of leaf area index based on support vector machine regression in winter wheat[J]. *Transactions of the Chinese Society of Agricultural Engineering (Transactions of the CSAE)*, 2013, 29(7): 117—123. (in Chinese with English abstract)

Rapid detection of reducing sugar for potato granules by near infrared spectroscopy

Sun Xudong¹, Dong Xiaoling²

(1. School of Mechatronics Engineering, East China Jiaotong University, Nanchang 330013, China; 2. School of Foreign Language, East China Jiaotong University, Nanchang 330013, China)

Abstract: Reducing sugar content is one of the important indicators for evaluating the quality of potato granules. Near-infrared (NIR) spectroscopy has been attempted to determine reducing sugar content in potato granules using near-infrared (NIR) spectroscopy combined with least squares support vector machine (LSSVM) algorithm. NIR spectra were recorded in the wavenumber range of 10 000~4 000 cm^{-1} at a 4 cm^{-1} interval. The 110 samples were divided into calibration and prediction sets in terms of their respective actual value for avoiding bias in subset division. One of every four samples was divided into the prediction set according the range of actual value in calibration set covering the range in the prediction set. The calibration set contained 83 samples, and the remaining 27 samples constituted the prediction set. Three different variable selection methods, namely the moving windows partial least square (MWPLS), MWPLS-genetic algorithm (MWPLS-GA), and MWPLS-successive projection algorithm (MWPLS-SPA), were performed comparatively to choose spectral variables associated with reducing sugar content distributions. The partial least square (PLS) models were developed with these selection spectral variables with the number of PLS components optimized according to root mean square error of cross validation (RMSECV) in the calibration set. The results derived by variable selection techniques were then compared with the performance of PLS models with new samples in the prediction set. The PLS calibration model exhibited a higher correlation coefficient of prediction (R_p) of 0.976, lower standard error of prediction (SEP) of 0.273%, and ratio of SEP and standard deviation (SDR) of 4.593, which was built using 20 spectral variables selected by the MWPLS-SPA method. Nonlinear models of the least squares support vector machine (LSSVM) were developed using different spectral variables selected by MWPLS, MWPLS-GA, and MWPLS-SPA. The main parameters of penalty factor (γ) and nuclear parameters (σ^2) of the nuclear function for the radial basis function (RBF) were optimized by a two-step search method. Through comparison the performance of LSSVM models with new samples, the optimal LSSVM models for reducing sugar content were obtained with R_p of 0.984, SEP of 0.223%, and SDR of 5.62, which were developed with 20 spectral variables selected by the MWPLS-SPA method. The results indicated that: 1) the accuracy of the quantitative analysis conducted by NIR spectroscopy can be improved through appropriate wavelength selection with the MWPLS-SPA method; and 2) the implementation of LSSVM nonlinear models could predict reducing sugar content in potato granules more accurately than a linear model of PLS. It was concluded that NIR spectroscopy combined with MWPLS-SPA and LSSVM methods has significant potential to quantitatively analyze reducing sugar content in potato granules, and this real time, in situ measurement will significantly improve the efficiency of quality control and assurance.

Key words: near infrared spectroscopy, nondestructive examination, algorithm, variable selection, potato, reducing sugar

(责任编辑: 刘丽英)