

光谱关键变量筛选在农产品及食品品质无损检测中的应用进展

王冬^{1,3}, 吴静珠^{2*}, 韩平^{1,3*}, 王坤²

1. 北京农业质量标准与检测技术研究中心, 北京 100097
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
3. 农业农村部农产品质量安全风险评估实验室(北京), 北京 100097

摘要 农产品及食品的品质与安全一直以来都是人们关注的焦点, 不仅关系着人们的身体健康, 而且关系着社会稳定甚至国家安全。由于农产品及食品的品质不合格引发的安全事件备受社会各界的广泛关注。对农产品及食品的品质监管长久以来都是分析检测领域的重点和难点。我国人口众多, 对农产品和食品的消费量非常大。面对如此大量农产品及食品品质的无损快速检测需求, 光谱法以其快速、无损、高效、环境友好、可现场检测等诸多特点, 为农产品及食品品质的无损快速分析提供了良好的解决方案。然而, 传统的光谱法在检测过程中所使用的数据量十分庞大, 不仅在建立校正模型过程中会消耗大量时间, 而且难以完成大量农产品及食品的品质在线高通量无损快速检测。大量数据的计算成为限制光谱类分析仪器工作效率的主要瓶颈之一, 并且大量数据的计算对仪器设备的硬件配置也提出了非常高的要求, 从而间接地提高了光谱分析技术的应用成本。近年来, 关键变量筛选技术脱颖而出, 并成为光谱分析的一个新热点。通过筛选, 采用少量关键变量建立校正模型即可得到和全谱数据建模准确度相差无几的分析结果, 从而可以有效提高分析仪器的工作效率并间接地降低光谱分析技术的应用成本, 进而为农产品及食品品质的高通量检测提供了可靠的技术支持、为满足人民日益增长的美好生活需要提供科技保障。针对光谱关键变量筛选在粮食及粮食作物、蔬菜、水果、经济作物、肉类、食品品质与安全领域的无损检测应用进行综述, 对光谱关键变量筛选技术的应用从筛选方法、应用范围、应用效果等方面进行了分类总结归纳, 并就光谱关键变量筛选技术在农产品及食品品质无损检测中的应用从变量筛选方法特点及趋势、所选变量的稳定性和可靠性、所选变量的实际意义等方面进行了展望。

关键词 光谱分析; 关键变量筛选; 无损检测; 农产品品质; 食品品质与安全

中图分类号: O657.33 **文献标识码:** R **DOI:** 10.3964/j.issn.1000-0593(2021)05-1593-09

引言

农产品和食品的品质优劣一直是人们关注的热点。农产品及食品品质与安全不仅关系着人们的身体健康, 而且关系着社会稳定甚至国家安全。近年来, 由于农产品及食品品质不合格导致的安全事件依然多发, 造成了不良的社会影响。我国对农产品及食品的消费量非常大, 对大量农产品及食品品质的无损快速检测成为当前亟待解决的问题。长久以来, 对农产品和食品品质的高通量无损快速检测一直是分析检测领域的重点和难点^[1]。

光谱分析法, 尤其近红外光谱分析, 以其无损、快速、高效、环境友好以及可实现在线及现场检测等诸多特点, 为农产品及食品品质无损快速检测提供了良好的解决方案。然而, 传统的光谱分析法在分析过程中所涉及的大量数据成为该方法应用过程中的瓶颈, 主要表现为3个方面: (1)庞大的数据量增加建模过程的计算成本; (2)庞大的数据量对仪器装备的硬件提出了很高的要求, 间接增加了技术应用成本; (3)对样品的预测仍采用全谱数据导致仪器工作效率降低, 无法满足农产品及食品品质高通量无损快速检测的需求。

面对上述困难, 近年来, 尤其近十余年来, 光谱变量筛选算法脱颖而出, 并成为光谱分析的一个新热点^[2]。通过对

收稿日期: 2020-05-26, 修订日期: 2020-10-09

基金项目: 北京市农林科学院科技创新能力建设专项储备性研究课题(KJCX20180409), 北京工商大学食品安全大数据技术北京市重点实验室开放课题(BUBD-2017KF-11), 科技部国家重点研发计划项目(2017YFD0201607)资助

作者简介: 王冬, 1982年生, 北京农业质量标准与检测技术研究中心副研究员 e-mail: wangd@brcast.org.cn

* 通讯作者 e-mail: hanp@brcast.org.cn; pubwu@163.com

光谱数据筛选关键变量, 基于所选的少量关键变量数据即可建立准确度较高的校正模型, 有效提高了工作效率, 并间接地降低光谱分析的应用成本, 进而为农产品及食品品质的高通量检测提供了可靠的技术支持。目前, 常用的关键变量筛选算法主要有以下几个类型: (1) 根据偏最小二乘 (partial least square, PLS) 模型的一些参数进行变量筛选^[3-5], 如无信息变量消除 (uninformative variable elimination, UVE)、竞争自适应重加权采样 (competitive adaptive re-weighted sampling, CARS) 等; (2) 基于智能优化算法进行变量筛选^[6], 如遗传算法 (genetic algorithm, GA)、模拟退火算法 (simulated annealing, SA) 等; (3) 连续投影算法 (successive projection algorithm, SPA); (4) 模型集群分析策略变量筛选算法^[7-8], 如变量组合集群分析 (variable combination population analysis, VCPA)、自举柔性收缩算法 (bootstrapping soft shrinkage, BOSS) 等; (5) 变量区间选择算法^[9], 如区间偏最小二乘 (interval partial least square, iPLS)、移动窗口偏最小二乘 (moving window partial least square, MWPLS)。此外, 为了提高变量筛选算法所选变量稳定性, 近年来, 蒙特卡洛 (Monte-Carlo, MC) 方法^[10-11]正越来越多地被运用到关键变量筛选过程中, 通过多次蒙特卡洛方法抽样选择关键变量, 对所选变量进行频次统计, 以提高所选变量的稳定性。

本文针对光谱关键变量筛选在农产品及食品品质无损快速检测中的应用, 分别就粮食及粮食作物、蔬菜、水果、经济作物、肉类、食品品质与安全等方向进行综述, 对光谱关键变量筛选技术的应用从筛选方法、应用范围、应用效果等方面进行了分类总结归纳, 并就光谱关键变量筛选技术在农产品及食品品质无损检测中的应用从变量筛选方法特点及趋势、所选变量的稳定性和可靠性、所选变量的实际意义等方面进行了展望。

1 光谱关键变量筛选在粮食及粮食作物品质无损检测中的应用

粮食是指烹饪食品中各种植物种子的总称, 富含蛋白质、维生素、膳食纤维、脂肪等营养物质, 是人们获取能量的最主要来源, 是国家之根本, 其重要程度不言而喻。对粮食的品质检测关系到粮食储备、流通、消费等诸多环节。一些学者就粮食及粮食作物品质的无损快速检测过程中关键变量的筛选进行了研究, 并取得了一定的成果。

杨赛等^[12]采用联合偏度算法 (joint skewness algorithm) 对玉米种子在 438~1 000 nm 范围内的高光谱成像数据筛选关键变量, 筛选了高光谱图像的最优波段, 基于支持向量机分别建立了基于平均光谱、图像熵及二者联合特征的种子分类模型; 结果表明, 在 10 个最优波段条件下, 联合特征分类模型的识别准确度达到 96.28%, 高于全波段联合特征识别模型的 93.47%。Wang 等^[13]采用波段范围 400~1 000 nm 的高光谱成像研究玉米种子硬度、弹性和回弹性的变化, 采用正交信号校正 (orthogonal signal correction, OSC) 对数据进行预处理, 采用 SPA 算法筛选关键变量, 结合 PLS 回归建立校正模型; 结果表明, 针对硬度、弹性、回弹性 3 个指

标分别筛选了 7 个、7 个和 6 个关键变量, OSC-SPA-PLS 回归模型的预测测定系数 (determination coefficient of prediction, R_p^2) 分别为 0.836 5, 0.821 7 和 0.793 0, 预测均方根误差 (root mean square error of prediction, RMSEP) 分别为 0.21, 0.05 和 0.06。Elhadi Adam 等^[14]采用高光谱技术研究了玉米叶斑病的早期检测, 整合了导向规则化随机森林 (guided regularized random forest, GRRF) 算法和传统随机森林 (random forest, RF) 算法用于分类; 结果表明, 健康玉米叶片和处于叶斑病感染早期阶段的叶片之间, 生化浓度存在统计学差异, 有关的关键波长位于 420, 795, 779, 1 543, 1 747 和 1 010 nm; 使用上述变量建立随机森林分类器的总体准确度为 88%, kappa 值为 0.75。Wang 等^[15]使用高光谱遥感技术估算光合色素的状况对于有效评估玉米产量的影响, 采用连续小波变换 (continuous wavelet transform, CWT) 结合 UVE 算法筛选关键变量并建立 PLS 模型; 结果表明, 叶绿素 Chl、类胡萝卜素 Car 的 R^2 分别为 0.943 5 和 0.880 7, 均方根误差 (root mean square error, RMSE) 分别为 0.094 和 0.013, 相对预测性能 (ratio of prediction to deviation, RPD) 分别为 2.65 和 2.38。Gao 等^[16]采用随机森林算法基于高光谱成像数据研究了杂草和玉米的分类方法, 筛选了 30 个关键变量并建立随机森林模型, 对玉米、卷心菜、土大黄和桔梗的平均正确识别率分别为 1.000, 0.789, 0.691 和 0.752。宦克为等^[17]采用 CWT 结合多元散射校正 (multiple scatter correction, MSC) 及支持向量机 (support vector machine, SVM) 优化了小麦蛋白质定量模型, 在优化参数下用 CWT 结合 UVE、SPA 筛选关键变量, 并建立 SVM 模型, RMSEP 达到 0.39。Liang 等^[18]采用最小二乘支持向量回归 (least square support vector regression, LSSVR) 和随机森林回归 (random forest regression, RFR) 估算小麦叶片氮含量, 模型 R^2 不低于 0.72。

由此可见, 关键变量筛选算法在粮食及粮食作物品质无损快速检测方面有一定的应用效果, 可为粮食及粮食作物品质的高通量无损快速检测提供技术支持。其中, 恰当地选择关键变量甚至可以建立准确度更高的校正模型。

2 光谱关键变量筛选在蔬菜品质无损检测中的应用

蔬菜是指可以做菜、烹饪成为食品的植物或菌类, 是人们日常饮食中必不可少的食物之一。蔬菜可提供人体所必需的多种维生素和矿物质等营养物质, 此外, 蔬菜中还有多种多样的植物化学物质, 是人们公认的对健康有效的成分, 对慢性疾病、退行性疾病有很好的预防作用。目前, 近红外技术已在蔬菜品质无损检测中得以应用^[19], 其中不乏一些学者就蔬菜品质无损快速检测过程中关键变量的筛选进行了研究, 并取得了一定的成果。

Liu 等^[20]采用近红外光谱研究了除草剂胁迫下油菜叶片中总氨基酸含量测定模型, 采用直接正交信号校正 (direct OSC, DOSC) 算法进行数据预处理, 采用 SPA 选择关键变量, 所建直接正交信号校正-连续投影算法-最小二乘支持向

量机(DOSC-SPA-least square SVM, DOSC-SPA-LSSVM)模型的相关系数 $r=0.9968$ 。Zhang 等^[21]采用近红外光谱研究了除草剂胁迫下油菜叶片中天冬氨酸含量的定量模型,采用 SPA, MC-UVE 和随机森林算法选择关键变量;其中, SPA-LSSVM 模型具有最佳预测性能,预测相关系数 $r_p=0.9962$ 。罗微等^[22]采用近红外光谱研究了白菜种子的鉴别,采用 SPA 选取特征波长,建立 LSSVM 模型和偏最小二乘判别分析(PLS-discriminant analysis, PLS-DA)模型,其中 LSSVM 模型的训练集和预测集识别率均达到 100%。Anna Luitza Bizerra Brito 等^[23]采用近红外光谱研究了生菜分类,采用 GA 和 SPA 筛选关键变量,建立线性判别分析(linear discriminant analysis, LDA)模型;其中 GA-LDA 模型的预测集正确率达到 95.4%。Su 等^[24]采用近红外高光谱成像和傅里叶变换中红外显微光谱研究甘薯烹饪过程中成分的变化并鉴定甘薯品种,采用 CARS 算法结合 SPA 筛选关键变量,建立了 5 变量 CARS-SPA-PLS 预测模型,预测集 $R_p^2=0.913$ 。Angel Dacal-Nieto 等^[25]将随机森林算法用于马铃薯块茎空心病的近红外高光谱检测,所建 RF-SVM 模型分类正确率达到 89.1%。Cen 等^[26]采用可见-近红外高光谱成像研究了冷害黄瓜的鉴别,采用信息特征选择(mutual information feature selection, MIFS),最大相关最小冗余(max-relevance min-redundancy, MRMR)和顺序前向选择(sequential forward selection, SFS)算法筛选关键波段,结合贝叶斯判别、SVM、K 最临近(K-nearest neighbor, KNN)算法建立判别模型;结果表明, SFS-SVM 模型的判别准确率达到 90.5%~100%。André M K Pedro 等^[27]采用近红外光谱测定番茄产品中的总固形物、可溶性固形物和天然色素含量,采用 SPA 筛选关键变量,并建立 PLS 模型;结果表明,总固形物、可溶性固形物、番茄红素、 β -胡萝卜素的相关系数(r)分别为 0.9998, 0.9996, 0.9996 和 0.9981, RMSEP 分别为 0.42, 0.63, 21.58 和 0.73。Wu 等^[28]采用可见-短波近红外光谱对干紫菜中的蛋白质含量进行定量检测,对数据采用标准正态变量变换(standard normal variate, SNV)预处理,建立了 UVE-SPA-LSSVM, SPA-LSSVM, UVE-SPA-PLS, SPA-PLS, UVE-SPA-MLR, SPA-MLR 模型,其中采用了 UVE 的模型较未采用 UVE 的模型,其 RPD 值平均提高 38.66%。Xiao 等^[29]采用近红外高光谱成像测定双孢蘑菇切片中可溶性固形物含量(soluble solid content, SSC),采用 OSC 进行数据预处理,采用 CARS 算法筛选关键变量,采用 SVM 建立校正模型。结果表明,所建模型 $R_p^2=0.883$, RPD=3.04。Kaveh Mollazade^[30]采用可见-短波近红外高光谱成像研究了纽扣蘑菇褐变的无损鉴别,采用 CARS 算法筛选关键变量,建立 PLS-DA 模型,识别蘑菇褐变的准确率,校正集和验证集分别为 80.6%和 80.3%。

从以上内容可见,变量筛选算法在蔬菜品质无损快速检测中具有较好的应用效果,可为大量蔬菜的品质无损快速检测提供技术支持。其中不乏将多种变量筛选算法相结合的报道,例如将 UVE 与 SPA 相结合,在消除无信息变量的基础上进一步去掉变量间的共线性,其对关键变量的筛选结果优于单一变量筛选算法。

3 光谱关键变量筛选在水果品质无损检测中的应用

水果是指多汁且主要味觉为甜味和酸味、可食用的植物果实。水果不但含有丰富的营养物质,而且具有促进消化等保健作用。水果品质无损快速分级正成为主流趋势,而基于近红外技术对水果进行无损快速检测则可为水果品质无损快速分级提供有效的解决方案。根据水果品质的不同制定差异化价格进行销售,更好地实现物尽其用,在提高水果利用率的同时还可避免浪费。一些学者就水果品质无损检测过程中关键变量的筛选开展了研究,并取得了一定的成果。

王转卫等^[31]采用近红外漫反射光谱研究了富士苹果品质指标的无损快速检测方法,采用主成分分析(principal component analysis, PCA)、SPA 和 UVE 筛选关键变量,并结合 LSSVM 和极限学习机(extreme learning machine, ELM)建立校正模型;结果表明, SPA-ELM 模型预测 SSC, pH 值准确度更高, RMSEP 分别为 0.44 和 0.0068, PCA-ELM 预测硬度、含水率准确度更高, RMSEP 分别为 0.26 和 0.62。Che 等^[32]采用可见-短波近红外光谱研究了苹果淤伤的检测,采用随机森林算法筛选关键变量,平均准确度达到 99.9%,并根据随机森林模型优选出 675 和 960 nm 附近 2 个特征波段。Dong 等^[33]采用近红外高光谱成像对富士苹果在 13 周储存期内的 SSC、硬度、水分和 pH 进行无损检测,采用 SPA、UVE 算法筛选关键变量,并结合 PLS、LSSVM、反向传递网络建模(back propagation network modeling, BPNM)方法建立校正模型;结果表明,全部模型均可准确预测 SSC 和水分, SPA-LSSVM 和全谱 BPNM 可粗略估算 pH 值,而采用上述任何模型预测硬度皆无法得到准确结果。在预测 SSC、水分和 pH 值方面, SPA-LSSVM 模型更具综合性,预测相关系数分别为 0.961, 0.984 和 0.882。Li 等^[34]采用近红外光谱研究了梨中 SSC 的无损检测,采用 MC-UVE 和 SPA 算法相结合筛选关键变量,结合 PLS 算法建立校正模型;结果表明,与 MC-UVE-PLS 和 SPA-PLS 模型相比, MC-UVE-SPA-PLS 模型稳健性更好,而采用 MC-UVE-SPA-PLS 模型的 18 个关键变量所建 SSC 模型的校正集、预测集相关系数(r)分别为 0.88 和 0.88, RMSE 分别为 0.49 和 0.35。进一步地, Li 等^[35]采用可见-近红外光谱研究了多品种梨硬度的无损检测,采用 MC-UVE-SPA 从全谱 1344 个变量中筛选了 17 个关键变量,分别结合 PLS 和 LSSVM 建立校正模型;结果表明, MC-UVE-SPA-LSSVM 模型预测准确度更高,“翠冠”、“黄花”、“清香”三种梨的预测集相关系数(r)分别为 0.94, 0.93 和 0.92, RMSEP 分别为 0.91, 0.92 和 0.96。Zhang 等^[36]采用可见-短波近红外高光谱成像研究了砀山梨糖含量的无损快速检测,采用 MC-UVE, SPA, CARS, GA, CARS-SPA 和 GA-SPA 筛选关键变量,分别结合 PLS、LSSVM、反向传递人工神经网络(back propagation-artificial neural network, BP-ANN)建立校正模型;结果表明, CARS-PLS 和 GA-SPA-PLS 模型准确度更高,预测相关系数(r_{pre})分别为 0.8971 和 0.8969, RM-

SEP 分别为 0.39% 和 0.35%。Guo 等^[37]采用近红外光谱研究了 4 个品种桃的无损鉴别,采用 PCA, UVE 和 SPA 分别从全谱 2 074 个变量中筛选出 8 个主成分、1 067 个、10 个特征波长,分别结合 LSSVM 和 ELM 建立桃品种鉴别模型;结果表明,PCA-LSSVM, UVE-LSSVM 模型以及 PCA-ELM 模型的准确率可达到 100%,其他模型准确率皆不低于 96%。Zhang 等^[38]采用可见-短波近红外高光谱成像检测桃表面缺陷,采用 MC-UVE 和 SPA 筛选关键变量,结合 PLS 算法建立判别模型;结果表明,基于少量特征波长可建立人工缺陷、非人工缺陷模型,人工缺陷、非人工缺陷、对照和样品总准确度分别为 87.5%, 96.7%, 95.0% 和 93.3%。Yu 等^[39]采用可见-短波近红外高光谱研究枇杷缺陷的识别,采用 CARS 算法选择了 12 个关键变量,结合 PLS-DA 算法建立判别模型;结果表明, CARS-PLS-DA 模型对有缺陷枇杷的总体识别准确率为 92.3%。Huang 等^[40]采用可见-近红外高光谱成像研究了桑葚中总花青素含量和抗氧化活性的检测,采用 SPA, UVE 和 CARS 三种波长选择算法筛选关键变量,结合 PLS 和 LSSVM 建立校正模型;结果表明,9 变量 CARS-LSSVM 模型预测总花青素含量的准确度更高,交互验证 $R^2=0.959$, RPD=4.964, 而 18 变量 CARS-LSSVM 模型预测抗氧化活性的准确度更高,交互验证 $R^2=0.995$, RPD=14.255。Zhao 等^[41]采用可见-短波近红外高光谱成像研究桑葚中总可溶性固形物(total soluble solid, TSS)的检测,采用随机蛙跳(random frog, RF)算法从 512 个变量中筛选关键变量,结合 PLS, LSSVM 建立 TSS 的校正模型;结果表明,RF-LSSVM(径向核函数)模型的校正集、交互验证集、预测集相关系数(r)分别为 0.999, 0.958 和 0.956,校正均方根误差(root mean square error of calibration, RMSEC)、交互验证均方根误差(root mean square error of cross validation, RMSECV)、RMSEP 分别为 0.061, 0.453 和 0.430。Elfatih Abdel-Rahman 等^[42]采用可见-近红外高光谱数据研究甘蔗叶中氮浓度的检测,采用随机森林(random forest, RF)算法筛选关键变量,结合逐步多元回归算法建立校正模型;结果表明,非线性 RF 回归模型测定系数 $R^2=0.67$,验证均方根误差(root mean square error of validation, RMSEV)=0.15%。Zhang 等^[43]采用可见-短波近红外光谱建立哈密瓜 SSC 校正模型,采用 CARS, UVE, CARS-SPA, UVE-SPA 筛选关键变量,分别结合 PLS 和 LSSVM 建立校正模型;结果表明,哈密瓜赤道区域多光谱模型略优于总体多光谱模型,UVE-SPA-PLS 模型和 CARS-SPA-LSSVM 模型预测相关系数(R_p)分别为 0.914 3 和 0.914 3, RMSEP 分别为 0.835 9 和 0.895 8。Hu 等^[44]采用可见-短波近红外光谱对哈密瓜 SSC 进行定量测定,采用 SPA, MC-UVE, CARS 和 MC-UVE-SPA 筛选关键变量,结合多元线性回归(multiple linear regression, MLR), PLS 和 LSSVM 建立校正模型;结果表明,MC-UVE-SPA 筛选的 18 个变量建模准确度更高,MC-UVE-SPA-PLS, MC-UVE-SPA-LSSVM, MC-UVE-SPA-MLR 模型预测哈密瓜 SSC 的 RMSEP 在 0.95~0.99 之间。Mithun 等^[45]采用高光谱数据结合 RGB 数据对天然成熟和人工催熟香蕉进行识别,采用随机森林

(random forest, RF)筛选关键变量,结合多层感知前向神经网络建立校正模型,自然成熟和人工催熟香蕉的识别准确度分别达到 98.74% 和 89.49%。

从以上内容可见,变量筛选算法在水果品质无损快速检测中具有较好的应用效果,可为大量水果的品质无损快速检测提供技术支持,将为水果收购入库、精品出库、分级销售过程中的品质无损快速检测提供重要技术支持。

4 光谱关键变量筛选在经济作物品质无损检测中的应用

经济作物亦称“工业原料作物”,一般指为工业,特别是轻工业提供原料的作物。我国纳入人工栽培的经济作物种类繁多,包括纤维作物(如棉、麻等)、油料作物(如芝麻、花生等)、糖料作物(如甘蔗、甜菜等)、三料(饮料、香料、调料)作物、药用作物、染料作物、观赏作物、水果和其他经济作物等。近年来,近红外技术越来越多地被应用于经济作物品质的无损检测^[46],其中一些学者就关键变量筛选进行了探索并取得了一定的成果。

Zheng 等^[47]基于 CARS 算法提出稳定 CARS(stable CARS, SCARS)算法,用于建立烟草品质模型过程中关键变量的筛选;结果表明,与 MWPLS, MCUVE 和 CARS 相比,SCARS 可选择更少的变量,所建 PLS 模型的潜变量数更少且 RMSECV 更小。买书魁等^[48]采用近红外光谱建立酿酒高粱中支链淀粉和直链淀粉的定量分析模型,采用 UVE, SPA, UVE-GA 和 UVE-SPA 筛选关键变量,结合 PLS 算法建立校正模型;结果表明,基于 UVE-GA 筛选关键变量所建模型具有更高的准确度,支链淀粉、支链淀粉 R^2 分别为 0.952 3 和 0.941 7, RMSEP 分别为 1.28 和 0.09, RPD 分别为 12.1 和 34.18。陈斌等^[49]采用近红外光谱建立油菜籽含油率定量分析模型,采用 CARS, GA, SPA, UVE, 向后区间偏最小二乘(backward interval partial least squares, BI-PLS)、协同区间偏最小二乘(synergy interval partial least squares, SIPLS)算法筛选关键变量,结合 PLS、LSSVM 建立定量分析模型;结果表明,对于 PLS 模型,BIPLS-GA 所选的 26 关键变量建模效果最好,预测相关系数(r_p)、RMSEP 分别为 0.933 0 和 0.007 5,对于 LSSVM 模型,SIPLS-GA 所选的 13 关键变量建模效果最好,预测相关系数(r_p)、RMSEP 分别为 0.919 2 和 0.005 5。Sun 等^[50]采用近红外高光谱成像研究茶叶中水分的分布情况,分别采用 SPA, CARS, SPA-逐步回归(stepwise regression, SR), CARS-SR 算法筛选关键变量,结合 MLR 建立校正模型以实现茶叶中水分分布的可视化;结果表明,采用 Savitzky-Golay 平滑结合多元散射校正(multiple scattering correction, MSC)预处理,结合 CARS-SR-MLR 所建模型具有更好的预测性能, R_p^2 和 RMSEP 分别为 0.863 1 和 0.016 3。Cecilia Masemola 等^[51]将蒙特卡洛交互验证 CARS(Monte-Carlo cross validation CARS, MCCV-CARS)用于桉树叶片可见-近红外光谱的关键变量筛选,结合 PLS 算法建立氮含量校正模型;结果表明,鲜叶片、干叶片氮含量模型 R_p^2 分别为 0.82 和 0.88,

RMSEP 分别为 0.14 和 0.13。耿响等^[52]采用近红外光谱建立棉、麻样品鉴别模型,采用 UVE, SPA 和 CARS 算法筛选关键变量,结合 PLS 建立判别模型;结果表明,CARS 算法将原光谱 1 475 个变量减少至 67 个,所建 PLS 模型较另两种算法更佳,校正测定系数(R_c^2)、RMSEC、交互验证测定系数(R_{cv}^2)、RMSECV 分别为 0.987, 1.24%, 0.980 和 1.57%。Yuan 等^[53]将 4 种变量筛选算法 CARS, MC-UVE, SPA, 特征投影图(latent projective graph, LPG)用于茯苓近红外光谱关键变量的筛选,并建立 PLS-DA 模型;结果表明,PCA 模式空间中,白茯苓和茯苓皮表现出明显差异,针对白茯苓、茯苓皮所选关键变量数分别为 13 和 10,根据变量筛选结果所建 Fisher 判别函数模型可靠且有效。Xie 等^[54]采用傅里叶变换近红外光谱(Fourier-transform near-infrared spectroscopy, FT-NIRS)测定发酵冬虫夏草菌丝体中精氨酸含量,采用 CARS 和 SPA 筛选关键变量,结合 LSSVM 算法建立校正模型;结果表明,CARS-LSSVM 模型的 R_p^2 , RMSEP 和 RPD 分别为 0.837 0, 0.084 1 和 2.474 1,经关键变量筛选,输入变量数仅占原变量数的 2.04%。

从以上内容可见,变量筛选算法在经济作物品质无损快速检测中具有较好的应用效果,可为经济作物的品质无损快速检测提供技术支持,进而可为提高产品附加值、增强产品市场竞争力提供技术保障。

5 光谱关键变量筛选在肉类品质无损检测中的应用

肉类是指动物的皮下组织和肌肉,可以提供丰富的蛋白质、脂肪和热量。我国肉类消费总量近年来稳居世界前列;在采用近红外技术研究肉类品质的无损快速检测^[55-56]中,针对肉类品质无损检测过程中的关键变量筛选,一些学者进行了探索并取得了一定的成果。

马世榜等^[57]采用可见-近红外光谱对生鲜牛肉挥发性盐基氮进行定量分析,采用 UVE 和 SPA 筛选关键变量,结合 LSSVM 建立校正模型;结果表明,Savitzky-Golay 平滑后,UVE-SPA-LSSVM 建模使变量数减少了 99.5%,预测相关系数和标准差分别为 0.925 和 $4.615 \text{ mg} \cdot (100 \text{ g})^{-1}$ 。Pu 等^[58]采用近红外高光谱成像研究羔羊肉品质无损检测,采用 UVE, SPA, 克隆选择算法(clonal selection algorithm, CSA)筛选关键变量,并将三种方法结合为 UVE-SPA-CSA,经过合并与优化,对蛋白质、脂肪、水含量采用 7 个关键变量建立 MLR 模型;结果表明,脂肪、蛋白质、水含量的校正相关系数(r_c)分别为 0.95, 0.80 和 0.91,RPD 分别为 4.13, 1.31 和 2.53。王文秀等^[59]采用可见-短波近红外、长波近红外及二者融合数据建立猪肉脂肪、蛋白质定量分析模型,采用 CARS、随机蛙跳(random frog, RF)、MC-UVE-SPA 算法筛选关键变量;结果表明,融合数据基于 CARS 筛选变量数据建模效果最佳,脂肪、蛋白质 PLS 模型相关系数(r)分别为 0.950 5 和 0.951 0。Jia 等^[60]采用可见-短波近红外高光谱成像研究鲜鸡胸肉 pH 值,采用 CARS 算法筛选了 20 个关键变量,建立 PLS 回归模型;结果表明,基于所选关键变

量建模的验证集测定系数(R_v^2), RMSEV 和 RPD 分别为 0.94, 0.06 和 3.55,作为对比,未选择变量建模的 R_v^2 , RMSEV 和 RPD 分别为 0.87, 0.16 和 2.02。崔腾飞等^[61]采用近红外高光谱成像技术研究富硒鸡蛋、无公害鸡蛋、普通鸡蛋的鉴别,从卷积平滑、基线校正、SNV、标准化四种预处理中优选卷积平滑作为数据预处理方法,采用 SPA, UVE 和 BIPLS 分别筛选了 8, 107 和 155 个关键变量,并分别建立全谱(full spectra, FS), SPA, UVE, BIPLS 的 PLS-DA 模型;结果表明,BIPLS-PLS-DA 模型识别准确率最高,校正集、预测集识别率分别为 95.24% 和 78.18%。Wu 等^[62]采用高光谱成像研究了虾中掺入明胶的检测,采用 UVE 结合 SPA 筛选关键变量,采用 LSSVM 算法建立校正模型;结果表明,UVE-SPA 从原光谱 462 个变量中筛选出 13 个关键变量,基于关键变量所建 LSSVM 模型预测测定系数(R_p^2)达到 0.965。He 等^[63]采用高光谱成像建立三文鱼中假单胞菌分布模型,分别采用 SPA 和 CARS 算法筛选关键变量,结合 PLS 回归算法建立校正模型;结果表明,CARS-PLS 模型更优,其预测测定系数(R_p^2)和 RMSEP 分别达到 0.91 和 0.49。Cheng 等^[64]采用可见-短波近红外高光谱成像研究针对鱼片活菌计数的可行性,采用连续投影算法筛选了 7 个关键波长变量,分别结合 PLS 和 LSSVM 算法建立校正模型;结果表明,SPA-PLS 模型效果更优,其预测测定系数 R_p^2 , RMSEP 和 RPD 分别为 0.90, 0.57 和 3.13。

从以上内容可见,变量筛选算法在肉类品质无损快速检测中具有较好的应用效果,其中不乏采用所选关键变量建立模型优于全谱建模的例子,可为肉类品质无损快速检测提供技术支持,进而可为提高产品附加值、增强产品市场竞争力提供技术保障。

6 光谱关键变量筛选在食品品质与安全无损检测中的应用

食品一直以来是人们获取能量的重要来源,对人类的重要性不言而喻。食品安全(food safety)指食品无毒、无害,符合应当有的营养要求,对人体健康不造成任何急性、亚急性或者慢性危害。食品安全不仅关系着人们的身体健康,还关系到社会稳定甚至国家安全。在食品品质与安全无损检测方面,近红外技术近年来得以广泛应用;面对巨大的检测工作量,一些学者就光谱变量筛选在食品品质与安全无损检测中的应用开展了研究,并取得了一定的成果。

Li 等^[65]采用傅里叶变换近红外光谱建立苹果汁掺假的鉴别模型,采用 SPA 结合遗传算法(genetic algorithm, GA)、粒子群优化算法(particle swarm optimization, PSO)、组搜索优化器(group search optimizer, GSO)、萤火虫算法(firefly algorithm, FA)四种群智能优化算法筛选关键变量,并结合 PLS 算法建立校正模型;结果表明,SPA-PSO 算法可将建模变量数降低至 4 个,基于 4 个关键变量数据的 PLS 模型预测测定系数(R_p^2)、RMSEP 分别为 0.998 6 和 0.062 8。Wu 等^[66]采用近红外、中红外光谱建立奶粉中钙含量的定量校正模型,采用 UVE 结合 SPA 筛选关键变量,分

别从近红外、中红外光谱的 2 756 个、3 727 个变量中筛选了 11 个、15 个关键变量, 结合 LSSVM 算法建立校正模型; 结果表明, 基于近红外 11 个关键变量所建模型的测定系数 R^2 , RPD, RMSEP, RSEP 分别为 0.636, 1.587, 78.815 和 13.40%, 基于中红外 15 个关键变量所建模型的测定系数 R^2 , RPD, RMSEP 和 RSEP 分别 0.930, 3.703, 30.162 和 5.22%。温珍才等^[67]采用可见-近红外光谱建立油茶籽油中掺入大豆油、菜籽油的鉴别模型, 采用 CARS, UVE 和 SPA 筛选关键变量, 分别结合 PLS 回归算法建立校正模型; 结果表明, CARS-PLS 模型优于全谱 PLS, UVE-PLS, SPA-PLS 模型, 掺入大豆油、菜籽油及两种油的 CARS-PLS 模型相关系数 (r) 分别为 0.950, 0.928 和 0.980, RMSEP 分别为 24.5, 29.0 和 30.8。Jiang 等^[68]采用傅里叶变换近红外光谱建立发酵度分析模型, 采用 CARS 算法和稳定 CARS (SCARS) 算法筛选关键波长变量, 并分别结合 PLS-DA 建立判别模型; 结果表明, 从 1 557 个光谱波长变量中, CARS 和 SCARS 算法分别筛选了 58 个、47 个关键变量, 与全谱 PLS-DA 模型相比, 基于所选关键变量建模的正确率皆更高, 其中 SCARS-PLS-DA 模型更优, 正确识别率达到 91.43%。熊雅婷等^[69]采用近红外光谱检测黄酒中非糖固形物、酒精度、总酸、氨基酸态氮, 采用 SPA 算法筛选关键变量, 上述 4 个指标经 SPA 分别筛选了 20, 23, 19 和 13 个关键波长, 并将所选关键变量位置与分子中有关的化学键进行指认, 结合 PLS、MLR 算法建立校正模型; 结果表明, SPA-MLR 模型效果最好, 上述 4 个指标的测定系数 R^2 分别为 0.881, 0.998, 0.983 和 0.924。Zhao 等^[70]采用近红外高光谱成像对全麦面粉中掺入的花生粉、核桃粉进行定量识别研究, 采用 UVE、SPA 筛选关键变量, 结合 PLS 算法分别针对花生粉、核桃粉、花生粉及核桃粉建立定量校正模型; 结果表明, UVE-PLS 算法所建模型普遍优于 SPA-PLS 算法所建模型, 花生粉及核桃粉的 UVE-PLS 模型的预测测定系数 (R^2_p)、RMSEP 分别为 0.987 和 0.373%。Wang 等^[71]采用近红外光谱测定腐竹品质, 采用 MC-UVE、随机森林 (random forest, RF)、CARS 算法筛选关键变量, 并结合 PLS 算法建立校正模型; 结果表明, 蛋白质、脂质、水分分别采用 CARS-PLS, RF-PLS 和 CARS-PLS 建模效果更优, 测定系数 (R^2) 分别为 0.958, 0.966 和 0.976, RMSEP 分别为 0.656%, 0.442% 和 0.123%。

综上所述, 变量筛选算法在食品品质与安全无损检测领域具有一定的应用成果, 恰当地筛选关键变量可以使用较少的变量建立和全谱模型效果相近甚至超越全谱模型的校正模型, 这将为食品品质与安全的无损快速检测、提高仪器工作效率等提供技术参考。

7 结 论

农产品及食品是人们获取能量的主要来源, 其品质与质量安全不仅关系到百姓生活品质与身体健康, 而且关系着社会稳定甚至国家安全。面对大量农产品及食品的品质检测工作, 通过对样本采集全谱数据并筛选关键变量, 从而简化校

正模型、提高仪器工作效率是一种较为有效的技术方案。

就本文综述而言, 关键变量筛选工作可主要归纳为以下两方面趋势。(1) 多种关键变量筛选算法相结合, 取长补短。很多学者将 UVE、CARS 等算法与 SPA 算法相结合, 克服了 UVE、CARS 算法第一轮筛选后所选变量仍较多的缺点, 并充分发挥了 SPA 去共线性的功能。(2) 关键变量筛选研究过程越来越多地引入蒙特卡洛 (Monte-Carlo, MC) 方法, 为所选变量稳定性提供了保证。在样本数有限的前提下, 根据 MC 方法随机生成多个子校正集并根据各子校正集筛选关键变量, 在此基础上统计关键变量出现的频次, 从而为稳健关键变量的筛选提供了可能, 进而克服了基于不同校正集所选关键变量有差异的困难。

然而, 光谱关键变量的筛选目前仍存在问题, 主要体现在以下三个方面。(1) 光谱数据预处理对关键变量筛选的影响尚不明确。恰当的数据预处理可以增强光谱质量, 有利于光谱信息的提取; 然而, 光谱数据预处理对光谱关键变量的筛选有何影响, 目前尚无定论。因此, 在将来的研究中, 针对不同状态、不同化学环境的样品, 采用不同的光谱数据预处理对关键变量筛选结果的影响将会是下一步工作中值得研究的内容之一。(2) 所选变量的可靠性以及方法的普适性仍有待提高。光谱关键变量的筛选在精简建模变量的同时剔除了冗余变量是目前对光谱变量筛选算法的共识。然而, 在关键变量筛选过程中, 根据所选变量建立的校正模型的准确度是否能达到实际工作需要, 亦即所选变量的可靠性, 以及变量筛选方法的普适性仍然是需要注意的问题。一般而言, 对于统一体系, 随着所选变量数目的减少, 基于所选关键变量数据所建校正模型的准确度大体上呈下降趋势, 但也不乏基于所选关键变量所建校正模型的预测准确度接近甚至优于全谱模型的例子; 目前普遍认为基于所选关键变量建模的准确度和全谱建模准确度接近或能够满足实际工作需要即是可行的。值得注意的是, 虽然一些学者的研究表明, 恰当地筛选关键变量并基于所选关键变量数据建立校正模型的准确度完全可以满足实际工作需要, 接近甚至优于全谱建模结果, 但是上述结论对于种类、样式繁多的农产品及食品, 以及各种各样品质指标是否具有普适性, 仍存在一定的不确定性。因此, 对于关键变量筛选工作, 所选变量的稳定性、可靠性以及方法的普适性仍是需要继续研究的内容之一。(3) 所选变量的理化含义有待进一步解析。以近红外光谱为例, 近红外光谱主要来源于分子中含氢基团的合频与倍频吸收, 而对近红外光谱筛选关键变量的过程很少有学者从分子角度讨论所选波长变量的归属或其理化意义。虽然一些学者就所选关键变量的归属做了简要的分析, 但仍然停留在化学键的层面, 没有上升为分子层面。而对所选关键波长变量从分子层面进行解析不仅可以间接验证变量筛选算法的正确性、有效性和变量筛选算法的普适性, 而且有利于从分子角度揭示所建模型的机理。因此, 对所选光谱关键变量从分子角度解释其理化意义将成为今后有待进一步研究的内容之一。

综上所述, 光谱关键变量筛选在精简光谱变量数、提高建模和仪器工作效率方面可以提供良好的解决方案, 可为大量农产品及食品品质与安全检测工作提供有效的技术保障;

此外, 光谱关键变量的筛选在精简输入变量方面具有十分重要的作用, 因此还可为专用型仪器的研发提供可靠的技术支持, 从而可为降低光谱分析技术的应用成本、扩展光谱分析技术的应用范围奠定坚实的理论基础。

References

- [1] Cortés V, Blasco J, Aleixos N, et al. *Trends in Food Science & Technology*, 2019, 85: 138.
- [2] Naidoo L, Cho M A, Mathieu R, et al. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2012, 69: 167.
- [3] Xu S, Zhao Y, Wang M, et al. *Catena*, 2017, 157: 12.
- [4] Tang G, Hu J, Yan H, et al. *Optik*, 2016, 127(3): 1405.
- [5] Song X Z, Huang Y, Tian K D, et al. *Optik*, 2020, 203: 164019.
- [6] Liu D, Sun D W, Zeng X A. *Food & Bioprocess Technology*, 2014, 7: 307.
- [7] Yun Y H, Wang W T, Deng B C, et al. *Analytica Chimica Acta*, 2015, 862: 14.
- [8] Deng B C, Yun Y H, Cao D S, et al. *Analytica Chimica Acta*, 2016, 908: 63.
- [9] Song X, Huang Y, Yan H, et al. *Analytica Chimica Acta*, 2016, 948: 19.
- [10] WANG Zhong-yi, HOU Rui-feng, HUANG Lan, et al(王忠义, 侯瑞锋, 黄 岚, 等). *Transactions of the Chinese Society of Agricultural Engineering(农业工程学报)*, 2007, 23(5): 1.
- [11] LI Xiao-qin(李晓勤). *Journal of Agricultural Mechanization Research(农机化研究)*, 2009, 31(7): 104.
- [12] YANG Sai, ZHU Qi-bing, HUANG Min(杨 赛, 朱启兵, 黄 敏). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2017, 37(3): 990.
- [13] Wang Lu, Pu H, Sun D W, et al. *Food Analytical Methods*, 2015, 8(6): 1535.
- [14] Adam E, Deng H, Odindi J, et al. *Journal of Spectroscopy*, 2017, 2017: 1.
- [15] Wang Z, Chen J, Fan Y, et al. *Computers and Electronics in Agriculture*, 2020, 169: 105160.
- [16] Gao J, Nuyttens D, Lootens P, et al. *Biosystems Engineering*, 2018, 170: 39.
- [17] HUAN Ke-wei, LIU Xiao-xi, WANG Xin, et al(宦克为, 刘小溪, 王 欣, 等). *Journal of Changchun University of Science and Technology · Natural Science Edition(长春理工大学学报·自然科学版)*, 2014, 37(5): 146.
- [18] Liang L, Di L, Huang T, et al. *Remote Sensing*, 2018, 10: 1940.
- [19] Su W H, Sun D W. *Institute of Food Technologists*, 2018, 17: 220.
- [20] Liu F, Jin Z L, Naeem M S, et al. *Food and Bioprocess Technology*, 2011, 4: 1314.
- [21] Zhang Chu, Kong W, Liu F, et al. *Heliyon*, 2015, 2(1): e00064.
- [22] LUO Wei, DU Yan-zhe, ZHANG Hai-liang(罗 微, 杜焱焱, 章海亮). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2016, 36(11): 3536.
- [23] Brito A L B, Araújo D A, Pontes M J C, et al. *Analytical Methods*, 2015, 7(5): 1890.
- [24] Su W H, Bakalis S, Sun D W. *Biosystems Engineering*, 2019, 180: 70.
- [25] Dacal-Nieto A, Formella A, Carrión P, et al. *Lecture Notes in Computer Science*, 2011, 6855: 180.
- [26] Cen H, Lu R, Zhu Q, et al. *Postharvest Biology and Technology*, 2016, 111: 352.
- [27] Pedro A M K, Ferreira M M C. *Analytical Chemistry*, 2005, 77(8): 2505.
- [28] Wu D, Chen X, Zhu X, et al. *Analytical Methods*, 2011, 3: 1790.
- [29] Xiao K, Liu Q, Wang L, et al. *LWT—Food Science and Technology*, 2020, 122: 109030.
- [30] Mollazade K. *Food Analytical Methods*, 2017, 10: 2743.
- [31] WANG Zhuan-wei, CHI Qian, GUO Wen-chuan, et al(王转卫, 迟 茜, 郭文川, 等). *Transactions of the Chinese Society for Agricultural Machinery(农业机械学报)*, 2018, 49(5): 348.
- [32] Che W, Sun L, Zhang Q, et al. *Computers and Electronics in Agriculture*, 2018, 146: 12.
- [33] Dong J, Guo W. *Food Analytical Methods*, 2015, 8(10): 2635.
- [34] Li J, Tian X, Huang W, et al. *Food Analytical Methods*, 2016, 9(11): 3087.
- [35] Li J, Zhang H, Zhan B, et al. *Infrared Physics & Technology*, 2020, 104: 103154.
- [36] Zhang D, Xu L, Liang D, et al. *Food Analytical Methods*, 2018, 11: 2336.
- [37] Guo W, Gu J, Liu D, et al. *Computers and Electronics in Agriculture*, 2016, 123: 297.
- [38] Zhang B, Li J, Fan S, et al. *Computers and Electronics in Agriculture*, 2015, 114: 14.
- [39] Yu K Q, Zhao Y R, Liu Z Y, et al. *Food & Bioprocess Technology*, 2014, 7: 3077.
- [40] Huang L, Zhou Y, Meng L, et al. *Food Chemistry*, 2017, 224: 1.
- [41] Zhao Y R, Yu K Q, He Y. *Journal of Analytical Methods in Chemistry*, 2015, 2015(2): 343782.
- [42] Abdel-Rahman EM, Ahmed FB, Ismail R. *International Journal of Remote Sensing*, 2013, 34(2): 712.
- [43] Zhang D, Xu L, Wang Q, et al. *Food Analytical Methods*, 2019, 12: 136.
- [44] Hu R, Zhang L, Yu Z, et al. *Infrared Physics and Technology*, 2019, 109: 102999.

- [45] Mithun B S, Sujit S, Karan B, et al. Proc. SPIE, 2018, 10665: 106650T.
- [46] BAO Yi-dan, CHEN Na, HE Yong, et al(鲍一丹, 陈 纳, 何 勇, 等). Optics and Precision Engineering(光学精密工程), 2015, 23(2): 349.
- [47] Zheng K, Li Q, Wang J, et al. Chemometrics and Intelligent Laboratory Systems, 2012, 112: 48.
- [48] MAI Shu-kui, YANG Yang, ZHAO Xiao-bo, et al(买书魁, 杨 洋, 赵小波, 等). Food Science and Technology(食品科技), 2019, 44(2): 301.
- [49] CHEN Bin, LU Bing, LU Dao-li(陈 斌, 卢 丙, 陆道礼). Modern Food Science and Technology(现代食品科技), 2015, 31(8): 286.
- [50] Sun J, Zhou X, Hu Y, et al. Computers and Electronics in Agriculture, 2019, 160: 153.
- [51] Masemola C, Cho M A. International Journal of Remote Sensing, 2019, 40(17-18): 6841.
- [52] GENG Xiang, GUI Jia-xiang, SUN Tong, et al(耿 响, 桂家祥, 孙 通, 等). Shanghai Textile Science and Technology(上海纺织科技), 2015, 43(8): 80.
- [53] Yuan T, Zhao Y, Zhang J, et al. Scientific Reports, 2018, 8: 89.
- [54] Xie C, Xu N, Shao Y, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2015, 149: 971.
- [55] Kamruzzaman M, ElMasry G, Sun D W, et al. Food Chemistry, 2013, 141(1): 389.
- [56] Xu J L, Riccioli C, Sun D W. Journal of Food Engineering, 2016, 196: 170.
- [57] MA Shi-bang, PENG Yan-kun, XU Yang, et al(马世榜, 彭彦昆, 徐 杨, 等). Journal of Jiangsu University • Natural Science Edition(江苏大学学报 • 自然科学版), 2013, 34(1): 44.
- [58] Pu H, Sun DW, Ma J, et al. Journal of Food Engineering, 2014, 143(Dec.): 44.
- [59] WANG Wen-xiu, PENG Yan-kun, ZHENG Xiao-chun, et al(王文秀, 彭彦昆, 郑晓春, 等). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2017, 48(9): 303.
- [60] Jia B, Yoon SC, Zhuang H, et al. Journal of Food Engineering, 2017, 208(Sep.): 57.
- [61] CUI Teng-fei, YANG Xiao-yu, DING Jia-xing, et al(崔腾飞, 杨晓玉, 丁佳兴, 等). Science and Technology of Food Industry(食品工业科技), 2018, 39(5): 13.
- [62] Wu D, Shi H, He Y, et al. Journal of Food Engineering, 2013, 119(3): 680.
- [63] He H J, Sun D W. LWT—Food Science and Technology, 2015, 62: 11.
- [64] Cheng J H, Sun D W. LWT—Food Science and Technology, 2015, 62(2): 1060.
- [65] Li Y, Guo Y, Liu C, et al. Food Analytical Methods, 2016, 10(6): 1.
- [66] Wu D, Nie P, He Y, et al. Food & Bioprocess Technology, 2012, 5(4): 1402.
- [67] WEN Zhen-cai, SUN Tong, XU Peng, et al(温珍才, 孙 通, 许 朋, 等). Journal of Jiangsu University • Natural Science Edition(江苏大学学报 • 自然科学版), 2015, 36(6): 673.
- [68] Jiang H, Zhang H, Chen Q, et al. Molecular and Biomolecular Spectroscopy, 2015, 149: 1.
- [69] XIONG Ya-ting, LI Zong-peng, WANG Jian, et al(熊雅婷, 李宗朋, 王 健, 等). Food and Fermentation Industries(食品与发酵工业), 2015, 41(3): 185.
- [70] Zhao X, Wang W, Ni X, et al. Applied Sciences, 2018, 8(7): 1076.
- [71] Wang J, Wang J, Zhang X, et al. Journal of Food Quality, 2018, 2018: 3136516.

Application of Spectral Key Variable Selection in Non-Destructive Detection of the Qualities of Agricultural Products and Food

WANG Dong^{1,3}, WU Jing-zhu^{2*}, HAN Ping^{1,3*}, WANG Kun²

1. Beijing Research Center for Agricultural Standards and Testing (BRCAT), Beijing Academy of Agricultural and Forestry Sciences, Beijing 100097, China
2. Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing Technology and Business University (BTBU), Beijing 100048, China
3. Laboratory of Quality & Safety Risk Assessment for Agro-Products (Beijing), Ministry of Agriculture and Rural Affairs, Beijing 100097, China

Abstract The quality of agricultural products and food has always been one of the focuses of attention. The quality and safety of agricultural products and food are related to people's health and related to social stability and even national security. In recent years, the safety incidents caused by the unqualified quality of agricultural products and food have attracted all social circles' attention. The supervision of the quality of agricultural products and food has been the key point even difficulty in analysis and detection for a long time. Given a large population, the consumption of agricultural products and food is enormous in China. In the face of such a large number of the needs of non-destructive and rapid detection of agricultural products and food quality, spectroscopy analysis can provide a good solution for the non-destructive and rapid detection for agricultural products and food with the characteristics of fast, non-destructive, efficient, environmentally friendly, on-site testing. However, due to the large amount of data used in the traditional spectral analysis, it is time-consuming in developing calibration models and difficult to complete the online, high-throughput, non-destructive and rapid detection of the large number of agricultural products and food quality. On the other hand, the calculation of such a large number of data has also become one of the main bottlenecks limiting the efficiency of spectral analysis instruments, and the calculation of a large number of data also puts forward very high requirements for the hardware configuration of the instruments, which will increase the application cost of spectral analysis technology indirectly. In recent years, key variable selection has emerged and become a new hotspot of spectral analysis. According to the selection, calibration models can be developed by a few numbers of the key variables, which are of almost the same accuracy to the models developed by the full spectra. Thus it can improve the analytical instruments' working efficiency effectively and reduce the application cost of the spectral analysis technology. It will also provide reliable technical support for the high-throughput detection of agricultural products and food quality and provide the scientific and technological support for meeting the increasing demand of the people for a better life. In this paper, the applications of spectral key variable selection in the non-destructive detection of grain and grain crops, vegetables, fruits, cash crops, meat, food quality and safety were reviewed. With summarizing others' works in recent years, the applications of spectral key variable selection technology were summarized from the aspects of selection method, application scope, application effect, and so forth. Finally, the application of spectral key variable selection technology in non-destructive detection of agricultural products and food quality prospected from the aspects of the characteristics and trends of the variable selection methods, the stability and reliability and the practical significance of the selected variables.

Keywords Spectroscopic analysis; Key variable selection; Non-destructive detection; Agricultural products quality; Food quality and safety

(Received May 26, 2020; accepted Oct. 9, 2020)

* Corresponding authors