

利用近红外及中红外融合技术对小麦产地和烘干程度的同时鉴别

邹小波, 封 韬, 郑开逸, 石吉勇, 黄晓玮, 孙 悦

江苏大学食品与生物工程学院, 江苏 镇江 212013

摘 要 小麦是制作馒头的主要原料之一, 小麦中水、蛋白质、淀粉会因产地以及烘干程度的差异而不同, 进而影响到加工成馒头的品质。所以实现对小麦产地和烘干程度的快速鉴别就显得尤为重要。感官评定是鉴别小麦产地和烘干程度常用的方法, 对比感官评定, 光谱分析可以识别样品中的分子结构等信息。基于此, 尝试利用近红外和中红外光谱融合技术实现对不同产地和不同烘干程度的小麦同时鉴别。首先选取了两个不同产地的小麦, 再利用微波干燥法对两个不同产地的小麦做烘干预处理, 使烘干的小麦水含量为 $12\% \pm 0.5\%$, 原麦水含量为 $18\% \pm 0.5\%$ 。分别标记为原麦 A, 烘干 A, 原麦 B, 烘干 B, 再将小麦研磨成粉末, 过 100 目筛网筛选后, 置于自封袋中备用。随后分别采集四种小麦样品的近红外和中红外光谱信息, 在 Matlab 7.10 的环境下使用标准正态变量变换 (standard normal variable transformation, SNVT) 对采集到的原始光谱数据进行预处理, 利用主成分分析对预处理后的数据进行降维处理, 再结合线性判别分析 (linear discriminant analysis, LDA) 和支持向量机 (support vector machine, SVM) 分别建立小麦近红外、中红外光谱数据识别模型。另外利用联合区间偏最小二乘法 (synergy interval partial least square, SiPLS) 筛选出利用标准正态变量变换 (SNVT) 预处理后的小麦近红外和中红外光谱数据特征光谱区间, 将筛选出的近红外和中红外光谱数据特征光谱区间融合后再结合线性判别分析 (LDA) 和支持向量机 (SVM) 建立小麦融合光谱信息的识别模型。然后比较同种光谱数据下利用线性判别分析 (LDA) 和支持向量机 (SVM) 建立的小麦识别模型识别率、比较同种建模方法下近红外和中红外光谱数据建立小麦识别模型识别率、比较同种建模方法下光谱数据融合和单一光谱数据建立小麦识别模型识别率。结果表明, 同种光谱分析方法, 利用 SVM 建立的四种小麦识别模型识别率高于利用 LDA 建立的小麦识别模型识别率。同种建模方法, 近红外光谱数据建立的小麦识别模型识别率优于中红外光谱数据建立的小麦识别模型识别率。而在同种建模方法下, 利用 SiPLS 筛选出近红外和中红外光谱数据的特征光谱区间数据融合后建立小麦识别模型识别率最高, 光谱数据融合后结合 LDA 建立的小麦识别模型校正集识别率为 98.75%, 预测集识别率为 97.50%; 而将此选择的变量结合 SVM 建立的小麦识别模型的校正集和预测集识别率都达到 100.0%。对比利用单一光谱数据建立的小麦识别模型识别率, 光谱数据融合之后建立的小麦识别模型识别率得到显著提高, 该研究从纵向和横向上全面地比较了光谱数据建立的小麦模型识别率, 结果可为更准确地运用光谱融合技术建立小麦产地以及烘干程度识别模型提供参考。

关键词 小麦; 光谱分析技术; 联合区间偏最小二乘法; 线性判别分析; 支持向量机

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)05-1445-06

引 言

馒头是我国居民主要主食之一, 近年来, 人们越来越关注馒头的口感和品质。小麦是制作馒头的主要原料之一, 其水、蛋白质及淀粉含量是影响馒头品质的重要因素。张春庆

等人研究表明, 小麦中蛋白质含量与馒头比容和体积呈显著正相关水平, 其相关系数分别达到 0.980 0 和 0.910 0^[1]。范玉顶等进一步表明小麦中高蛋白质含量有助于增加馒头的比容和体积, 但不利于馒头优良外观、结构和弹韧性的形成^[2]。付奎研究表明, 馒头的硬度、胶着性和咀嚼性随着小麦中淀粉含量的增加而增加, 但比容下降。小麦含水量不仅

收稿日期: 2018-03-30, 修订日期: 2018-07-18

基金项目: 国家重点研发计划项目(2016YFD0401104), 江苏省重点研发计划(BE2016306), 国家自然科学基金项目(31772073, 31671844), 江苏省六大人才高峰项目(GDZB-016)资助

作者简介: 邹小波, 1974 年生, 江苏大学食品与生物工程学院教授 e-mail: zou_xiaobo@ujs.edu.cn

对小麦贮藏期限有较大影响,且对小麦的品质如脂肪酸值、面筋指数也有较大影响,进而对馒头的口感和品质产生影响。我国常用的小麦产地和烘干程度鉴别方法有感官评价、传统的物化分析方法。感官评价的结果会受到鉴别人员主观因素的影响;传统的物化分析方法具有耗时长、花费大、破坏样品等缺点;因此有必要研究出一种快速,无损的小麦鉴别方法。

光谱分析具有快速、无损、非接触等优势,被广泛用于小麦的定性和定量检测中。近红外光谱能反映有机分子中含氢键基频振动的倍频和合频信息,其光谱特性与有机物类型 and 含量高度相关。中红外光谱吸收峰是基频、倍频或合频吸收,具有分子结构的特征性,不同化合物有其特异的红外吸收光谱,其谱带的位置、形状、数目和强度均随着化合物及其聚集态的不同而不同^[3]。不同产地的小麦样品,其化合物种类,含量会发生变化,并且小麦在烘干时,其蛋白质,脂肪,糖类等化合物的结构会发生变化,因此近红外光谱和中红外光谱可以用于小麦产地和烘干程度的同时鉴别。近年来,利用光谱分析定性和定量检测小麦成为国内外研究的热点。段国辉等选取地方小麦为材料,利用近红外光谱分析法和国标法对小麦组分进行测定,结果表明近红外光谱分析对小麦蛋白质含量有较高的预测结果^[4]。赵海燕等利用近红外光谱结合偏最小二乘判别分析法对不同产地小麦进行判别分析,结果表明小麦籽粒样品总体正确判别率为 87.50%^[5]。

近红外和中红外光谱的产生机制不同,两种光谱融合分析样品会更加充分利用样品的结构信息进行鉴别。数据融合是将多种传感器获得的数据进行综合分析、优化处理,融合后可以提高模型的预测精度^[6]。国内外的研究多采用单一光谱技术鉴别小麦,且识别率较低。基于此,本研究尝试用特征层融合的方法将近红外和中红外光谱数据融合建模,实现对小麦产地及烘干程度的同时鉴别。

1 实验部分

1.1 样品材料

两个不同产地的小麦(镇江某面粉厂提供),分别将其标记为品种 A(强面筋小麦)、品种 B(中强面筋小麦)。

1.1.1 样品的预处理

首先准确称量品种 A 与品种 B 样品各 100.0 g,并取其中一部分进行微波干燥(南京淮腾机械科技有限公司 YZWZ-18 型微波真空干燥机)。干燥后小麦的水分含量为 $12\% \pm 0.5\%$,原麦的水分含量为 $18\% \pm 0.5\%$ 。分别编号为原麦 A,烘干 A,原麦 B,烘干 B。将其研磨成粉末并过 100 目筛,将筛选出的粉末装入自封袋中备用。

1.2 数据采集方法

(1) 红外光谱数据采集

用近红外光谱仪(美国赛默飞世尔公司)扫描小麦粉末,以仪器内置背景为参比,积分球漫反射,扫描波数为 $10\,000 \sim 4\,000\text{ cm}^{-1}$,扫描次数为 16 次,分辨率为 8 cm^{-1} ,波数间隔 3.856 cm^{-1} 。用中红外光谱仪(美国热电公司)扫描

小麦粉末,采集区间为 $4\,000 \sim 650\text{ cm}^{-1}$,扫描次数为 16 次,分辨率为 2 cm^{-1} 。采集光谱时,每个样品采集 3 次,并取平均值得到 1 条原始光谱,每个样品各采集 30 条原始光谱,共有 120 条原始光谱。

(2) 红外光谱数据预处理

由检测器检测到的光谱信号不仅包含样品待测成分信息,还有各种仪器干扰信息,这些干扰信息会影响所建立模型稳定性和可靠性,因此有必要在数据处理前对采集的光谱数据进行合理的处理,从而减弱甚至消除非目标因素对光谱信息的影响^[7]。本研究用 SNVT 对原始光谱进行预处理,图 1 为 SNVT 处理后的光谱。

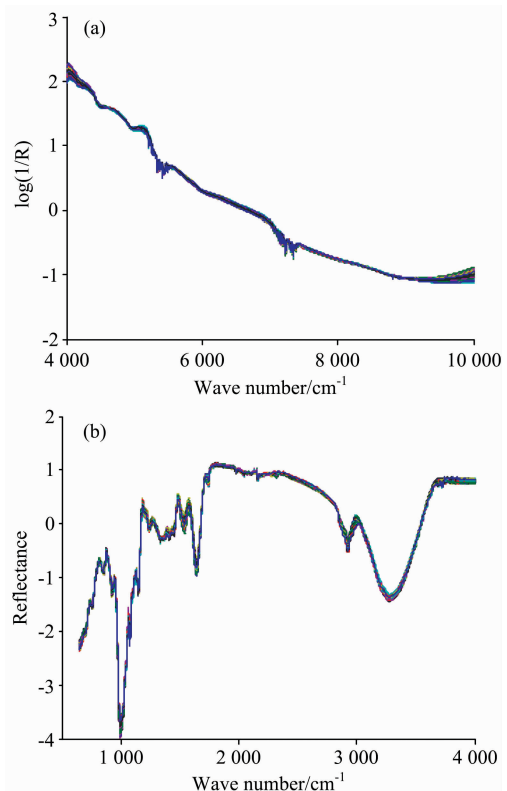


图 1 经 SNVT 处理的 4 种小麦的近红外光谱图和中红外光谱图

(a): 近红外光谱; (b): 中红外光谱

Fig. 1 Near infrared and mid infrared spectra of four wheat samples pretreated by SNVT

(a): Near-infrared; (b): Mid-infrared

2 结果与讨论

2.1 近红外光谱

2.1.1 小麦近红外光谱的主成分分析

图 2 为所得的近红外光谱数据的三维主成分得分图,其中前 3 个主成分的贡献率分别为 82.37%, 16.82% 和 0.380%, 累计贡献率达到 99.57%, 原麦 B 和烘干 B 部分交叉在一起,区分不明显,但原麦 A 和烘干 A 区分较好,为更好地区分样品,需进一步建立识别模型进行分类识别。

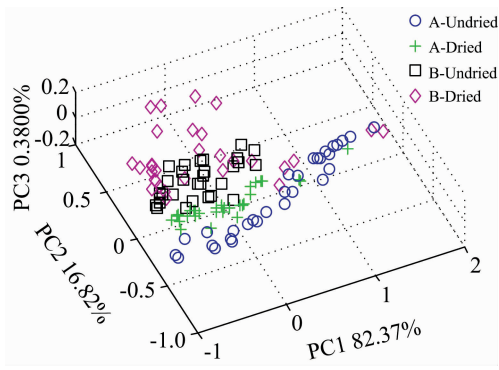


图 2 小麦的近红外三维主成分得分图

Fig. 2 PCA of the four wheat samples from NIR data

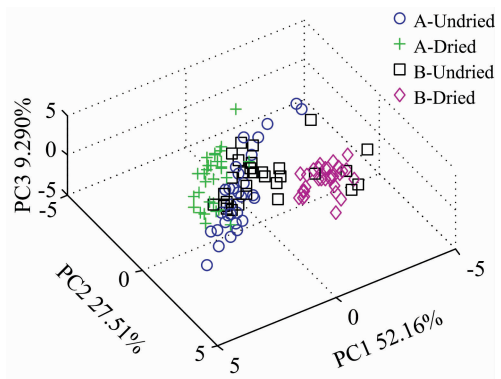


图 3 小麦的中红外三维主成分得分图

Fig. 3 PCA of the four wheat samples from MIR data

2.1.2 基于 LDA 小麦近红外光谱的定性识别

线性判别分析 LDA 是一种常用的数据分类和降维方法。取前 10 个主成分所对应的光谱信息建立 LDA 模型，模型识别率随着主成分数的增加而增加，当主成分数增加到 3 时，校正集识别率为 95.00%，预测集识别率为 90.00%。随着主成分数的再增加，校正集识别率保持不变。这可能是冗余信息降低了模型的识别率。

2.1.3 基于 SVM 小麦近红外光谱的定性识别

从 LDA 模型的识别结果看，该线性算法对小麦品种的分类鉴别并不是十分理想，因此本文考虑使用非线性算法支持向量机(SVM)建立识别模型。运行 SVM 算法时采用径向基函数，并对核函数的参数(惩罚系数 σ 和正规系数 γ)进行优化，采用主成分作为模型的输入向量。以训练集交互验证均方根误差(root AVG square error of cross validation, RM-SECV)值最小为指标。近红外光谱数据建立的识别模型最佳主成分数为 6，最佳参数对(σ, γ)为(0.079 33, 25.94)。结果如表 1 所示，此时模型校正集和预测集识别率都达到 100.00%，但主成分数再增加时，校正集识别率保持不变。趋势与 LDA 类似，但识别率明显高于 LDA。

表 1 近红外光谱的模型识别率对比

模型	最佳主成分数	识别率/%	
		校正集	预测集
LDA	3	95.00	90.00
SVM	6	100.0	100.0

2.2 中红外光谱

2.2.1 小麦中红外光谱的主成分分析

图 3 是中红外光谱的三维主成分得分图，前三个主成分的贡献率分别为 52.16%，27.51%，9.290%，累计贡献率达到 88.96%。烘干 A 和烘干 B 被区别开，但是原麦 A 和原麦 B 未能被区别开。对比近红外光谱，区别不明显，这可能是中红外光谱主要反映化合物含有的官能团、取代基的位置及数目、化合物的类别、立体结构等信息，故在定性检测中没有近红外准确^[8]。还有可能是前 3 个主成分并未能全部反映样品的信息，需进一步建立识别模型进行分类识别^[7]。

2.2.2 基于 LDA 小麦中红外光谱的定性识别

取前 10 个主成分所对应的光谱信息来建立中红外光谱数据的 LDA 模型，根据建立的中红外的 LDA 模型，当主成分数为 6 时，模型的识别率最高，校正集为 91.25%，预测集为 75.00%。综合比较近红外和中红外的模型识别率，中红外光谱的模型识别率低于近红外光谱的模型识别率，表明在定性检测中，中红外光谱的预测精度低于近红外。进一步说明，中红外光谱会受到样品化合物复杂结构的影响，用线性模型无法区分。

2.2.3 基于 SVM 小麦中红外光谱的定性识别

进一步使用 SVM 建立识别模型，模型建立方法与评价指标与 2.1.3 相同。中红外光谱建立的识别模型最佳主成分数是 8，最佳参数对(σ, γ)为(13.19, 15.68)，结果如表 2 所示。此时模型的校正集识别率 100.0%，预测集识别率为 90.00%，对比 LDA 模型，识别率显著提高，但是没有近红外光谱提高的明显；SVM 与线性模型(LDA)相比，精度提高较大，这说明了不同种类样品对中红外光谱的非线性影响可以被 SVM 所建立的非线性模型校正掉，进而实现准确预测。

表 2 中红外光谱的模型识别率对比

模型	最佳主成分数	识别率/%	
		校正集	预测集
LDA	6	91.25	75.00
SVM	8	100.0	90.00

2.3 基于近红外和中红外光谱融合的小麦样品定性识别

近红外和中红外光谱均可以快速、无损地对小麦品种进行定性识别，但它们的预测机制不同^[9]，因此本研究近红外和中红外光谱数据融合建立定性识别模型，常用的数据融合方法有数据层融合和特征层融合。数据层融合是原始数据层上的融合，直接对所获得的各种传感器原始数据进行综合分析，这种融合方法虽然可以尽可能保持多的数据，但原始数据包含较多的无关信息，处理起来费时费力，且抗干扰能力差。对比数据层融合，特征层融合首先提取各传感器原始数

据的特征信息, 然后对提取的特征信息进行综合分析和处理^[10], 这种融合方法首先进行特征提取, 压缩数据, 有利于实时快速处理数据, 且所提取的特征信息和决策分析相关, 融合结果的模型识别率得到较大提高。基于此, 本研究采取特征层融合, 具体是通过联合区间偏最小二乘法 (SiPLS) 优选出近红外和中红外光谱数据的特征光谱区间, 对优选出的特征光谱区间进行综合分析和处理, 从而得到更准确的预测结果。

2.3.1 基于 SiPLS 的近红外光谱特征区间优选

SiPLS 是一种常用的特征变量筛选方法, 是联合同一次区间划分后的子区间建立的 PLS 模型, 最终筛选出精度较高的特征变量^[11]。为了得到最佳的筛选结果, 需要对子区间的划分数及联合区间数进行优化, 以均方根误差 (root mean square error, RMSE) 为评判标准。经反复比较, 当划分区间数为 20, 联合区间数为 4 时, 得到的 RMSE 最小为 0.156 5。此时联合的子区间数为 [4 5 7 14], 子区间的波数范围分别为 5 500~4 894.4, 6 094~5 797 和 8 196~7 899 cm^{-1} 。如图 4 所示, 根据实测值和预测值建立的校正集和预测集的散点图的相关系数分别为 0.990 9 和 0.967 4, 相关性较好, 说明可以用优选出来的变量代替全局变量来进行数据建模。根据文献[9, 12], 筛选出的三个波段范围对应于样品中蛋白质, 脂肪和无机盐的信息。O—H 基团合频位于 5 000 cm^{-1} 附近。C—H 键一次倍频在 5 700 cm^{-1} 附近。筛选出的这些

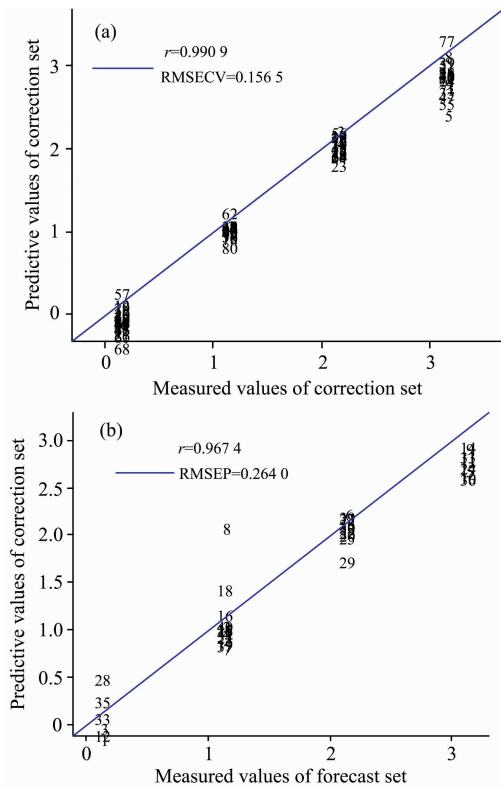


图 4 SiPLS 校正集和预测集的实测值和预测值散点图
(a): 校正集; (b): 预测集

Fig. 4 Scatter plot of measured and predictive values for correction sets and forecast sets (SiPLS)
(a): Correction set; (b): Forecast set

波段对应的吸收基团也说明了烘干对小麦蛋白质、淀粉和脂肪中基团产生的影响, 从而能对小麦的产地和烘干程度作出鉴别。

2.3.2 基于 SiPLS 的中红外光谱特征区间优选

通过 SiPLS 对中红外光谱数据进行特征区间筛选, 处理方法与 2.3.1 相同。结果显示, 当联合的子区间数为 [10 11 12 19] 时, RMSE 值最小为 0.322 1。子区间的波数范围分别为 2 659~2 157 和 3 823~3 665.1 cm^{-1} 。如图 5 所示, 根据实测值和预测值建立的校正集和预测集的散点图的相关系数分别为 0.961 0 和 0.949 5, 相关性较好, 说明可以用优选出来的变量代替全局变量来进行数据建模。根据文献[9, 12], 筛选出的波段范围 2 659~2 157 cm^{-1} 反映样品中蛋白质中基团的振动信息, 波段范围 3 823~3 665.1 cm^{-1} 反映样品中脂肪和淀粉中基团的信息。中红外光谱筛选出的波段范围对应的化合物信息与近红外筛选出的波段范围对应的化合物信息相同。O—H 和 N—H 基团的弯曲振动和伸缩振动位于 3 000~3 400 cm^{-1} 。筛选出的这些波段对应的吸收基团也说明了烘干对小麦蛋白质、淀粉和脂肪中基团的产生的影响, 从而能够对小麦的产地和烘干程度作出鉴别。

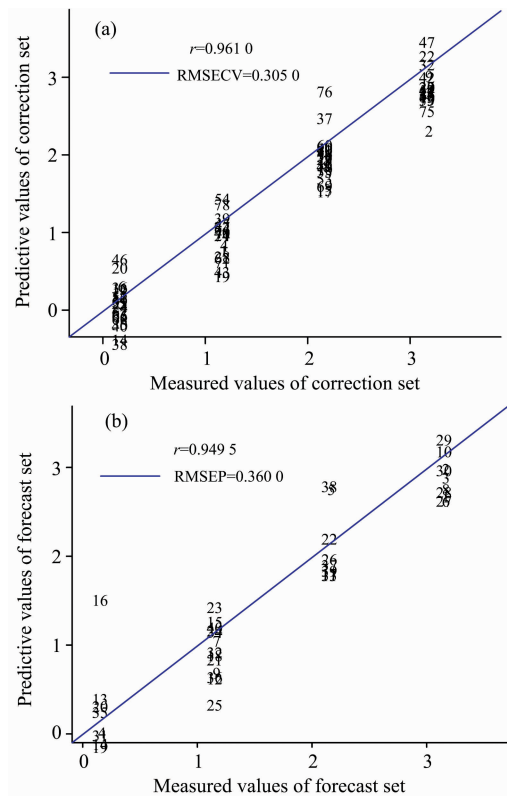


图 5 SiPLS 校正集和预测集的实测值和预测值散点图
(a): 校正集; (b): 预测集

Fig. 5 Scatter plot of measured and predictive values for correction sets and forecast sets (SiPLS)
(a): Correction set; (b): Forecast set

2.3.3 基于光谱数据融合的定性识别

(1) 基于 LDA 小麦光谱数据融合的定性识别

将 2.3.1 和 2.3.2 中优选出的特征光谱区间进行融合来

建立 LDA 模型。取前 10 个主成分所对应的光谱信息来建立融合后光谱数据的 LDA 模型。结果显示：随着主成分数的增加，LDA 模型的识别率增加，当主成分数增加到 7 时，模型的识别率最佳，校正集和预测集的识别率分别是 98.75% 和 97.50%，随着主成分数的再增加，校正集识别率基本保持不变，对比于近红外、中红外光谱数据建立的 LDA 模型，识别率均有所上升，结果表明光谱融合可以提高模型识别率。

表 3 近、中红外光谱融合的模型识别率对比
Table 3 Comparison of model recognition rates for near and middle spectral fusion

技术	模型	最佳主成分数	识别率/%	
			校正集	预测集
光谱融合	LDA	7	98.75	97.50
	SVM	5	100.0	100.0

(2) 基于 SVM 小麦光谱数据融合的定性识别

为保证结果的一致性，使用 SVM 建立识别模型，模型建立方法与评价指标与 2.1.3 相同。经处理得到模型的最佳主成分数为 5，最佳参数对 (σ, γ) 为 $(1.639, 17.13)$ ，结果如表 3 所示，此时模型校正集和预测集的识别率均为 100.0%。主成分再增加时，校正集的识别率保持不变。但此光谱处理

方法建立的 SVM 模型的识别率均优于以上建立的 SVM 模型的识别率，结果表明通过特征层数据融合，再结合非线性算法支持向量机(SVM)建立识别模型，模型的识别率显著提高。

3 结 论

采集了不同产地和不同烘干程度小麦的光谱信息并进行了不同的分析，结合 LDA 和 SVM 建立了快速识别模型，主要结论如下：

(1) 近红外和中红外光谱都可以用于小麦产地和烘干程度的同时鉴别，建立模型的识别率较高。

(2) 中红外光谱建立的识别模型，无论是线性识别模型还是非线性识别模型，识别率都低于近红外光谱建立的识别模型，说明近红外光谱对小麦产地和烘干程度鉴别的准确率高于中红外光谱。

(3) 将筛选出的近红外和中红外特征光谱区间融合之后建立的小麦产地和烘干程度 LDA 模型和 SVM 模型识别率均比单一光谱技术有所提高。这表明光谱融合在定性检测中是可行的。

研究表明，通过融合近红外和中红外特征光谱数据可更准确地对小麦产地、烘干程度同时做出鉴别。

References

- [1] ZHANG Chun-qing, LI Qing-qi(张春庆, 李晴祺). *Scientia Agricultura Sinica*(中国农业科学), 1993, 26(2): 39.
- [2] FAN Yu-ding, LI Si-shen, SUN Hai-yan, et al(范玉顶, 李斯深, 孙海艳, 等). *Journal of the Chinese Cereals and Oils Association*(中国粮油学报), 2005, 20(2): 16.
- [3] Machado J C, Faria M A, Ferreiraimpl V O, et al. *Talanta*, 2018, 180: 69.
- [4] DUAN Guo-hui, ZHANG Yuan, GAO Hai-tao, et al(段国辉, 张 园, 高海涛, 等). *Jiangsu Agricultural Sciences*(江苏农业科学), 2011, 39(5): 422.
- [5] ZHAO Hai-yan, GUO Bo-li, WEI Yi-min, et al(赵海燕, 郭波莉, 魏益民, 等). *Scientia Agricultura Sinica*(中国农业科学), 2011, 44(7): 1451.
- [6] DENG Hai-yan, SONG Xiang-zhong, XIONG Yan-mei(邓海雁, 宋相中, 熊艳梅). *Spectroscopy and Spectral Analysis*(光谱学与光谱分析), 2016, 36(10): 183.
- [7] SHI Ji-yong, HU Xue-tao, ZHU Yao-di, et al(石吉勇, 胡雪桃, 朱瑶迪, 等). *Transactions of the Chinese Society for Agricultural Machinery*(农业机械学报), 2016, 47(2): 245.
- [8] ZHANG Ju-hua, ZHU Xiang-rong, SHANG Xue-bo, et al(张菊华, 朱向荣, 尚雪波, 等). *Food Industry Technology*(食品工业科技), 2010, 31(10): 421.
- [9] ZOU Xiao-bo, CUI Xue-ping, SHI Ji-yong, et al(邹小波, 崔雪平, 石吉勇, 等). *Transactions of the Chinese Society for Agricultural Machinery*(农业机械学报), 2018, 49(3): 341.
- [10] WU Shuang, WANG Jie, YU Ya-ru, et al(吴 双, 王 杰, 俞雅茹, 等). *Journal of the Chinese Cereals and Oils Association*(中国粮油学报), 2017, 32(11): 158.
- [11] WANG Sheng-peng, ZHENG Peng-cheng, GONG Zi-ming, et al(王胜鹏, 郑鹏程, 龚自明, 等). *Journal of Huazhong Agricultural University*(华中农业大学学报), 2018, 37(1): 89.
- [12] Hell J, Preckle R M, Danne R L, et al. *Food Control*, 2016, 60(5): 365.

Simultaneous Identification of Wheat Origin and Drying Degree Using Near-Infrared and Mid-Infrared Fusion Techniques

ZOU Xiao-bo, FENG Tao, ZHENG Kai-yi, SHI Ji-yong, HUANG Xiao-wei, SUN Yue
School of Food and Biological Engineering, Jiangsu University, Zhenjiang 212013, China

Abstract Wheat is one of the main raw materials for making steamed bread, and the water, protein and starch in wheat vary depending on the place of production and the degree of drying, which in turn affects the quality of processed steamed bread. Therefore, it is particularly important to quickly identify the place of origin and degree of drying of wheat. Sensory evaluation is a common method used to identify the origin and degree of drying in wheat, in contrast to sensory evaluation, spectral analysis techniques can identify information such as molecular structure in a sample. Based on this, this paper attempts to use the near-infrared and mid-infrared spectral fusion technology to achieve the simultaneous identification of wheat from different producing areas and different degrees of drying. In this study, wheat from two different origins were selected and microwave drying was used to pretreat the wheat from the two different origin so that the moisture content of the dried wheat decreased to $12\% \pm 0.5\%$, while the moisture content of the undried wheat was $18\% \pm 0.5\%$. They were marked as undried wheat A, dried A, undried wheat B, dried B and then ground into powder, sieved with a 100 mesh screen and placed in a sealed bag for use. Subsequently, the near infrared and mid-infrared spectral information of four wheat samples were collected, and then the raw spectral data collected were pre-processed using the standard normal variate transformation (SNVT) using Matlab 7.10 version. The principal component analysis was used to reduce the dimension of the preprocessed data, and then the linear and short-infrared (NIR) and mid-infrared (MIR) spectral data were identified using linear discriminant analysis (LDA) and support vector machine (SVM), respectively to create a recognition model. In addition, using the synergy interval partial least square (SiPLS) method, the characteristic spectral ranges of the near-infrared and mid-infrared spectral data of the wheat pretreated with the standard normal variate transformation (SNVT) were screened out. After the fusion of the characteristic spectral ranges of the near-infrared and mid-infrared spectral data, a linear discriminant analysis (LDA) and support vector machine (SVM) were used to establish the identification model of the fusion spectral information of wheat. The recognition rate of wheat identification model established by linear discriminant analysis (LDA) and support vector machine (SVM) under the same spectral data were compared and the near-infrared and mid-infrared spectral data of the same modeling method established the wheat identification model. Recognition rate, comparison of spectral data fusion under the same modeling method and single spectral data were used to establish the recognition rate of the wheat identification model. The results showed that using the same kind of spectral analysis method, the recognition rates of the four wheat identification models established using SVM were higher than those of wheat identification models established using LDA. The recognition rate of wheat identification model established by near-infrared spectral data using the same modeling method was better than that of wheat identification model established by mid-infrared spectral data. Under the same modeling method, the identification rate of the wheat identification model established by the fusion of the characteristic spectral interval data of the near-infrared and mid-infrared spectral data filtered by SiPLS was the highest. After the fusion of spectral data, the wheat identification model established with LDA was integrated. The recognition rate of the correction set was 98.75%, and the recognition rate of the prediction set was 97.50%. The recognition rate of the correction set and the prediction set of the wheat identification model established by combining this selected variable with the SVM reached 100.0%. Comparing the recognition rate of wheat identification model established by using single spectral data, the recognition rate of wheat identification model established after fusion of spectral data was significantly improved. This study compared the wheat model's recognition rate established by spectral data from both vertical and horizontal directions. The results can provide a reference for the more accurate use of spectral fusion technology to establish wheat production areas and drying degree identification model.

Keywords Wheat; Spectral analysis technology; Synergy interval partial least square (SiPLS); Linear discriminant analysis (LDA); Support vector machine (SVM)

(Received Mar. 30, 2018; accepted Jul. 18, 2018)