



基于红外光谱和化学计量学的纺织品文物鉴定

郑浩然¹, 周杰¹, 周骏懿¹, 陈栩诗¹, 刘剑², 王秉¹

(1.浙江理工大学材料科学与工程学院, 杭州 310018; 2.中国丝绸博物馆, 杭州 310002)

摘要: 为解决古代纺织品文物的种属鉴别问题, 利用红外光谱和化学计量学的方法进行古代纺织品种属的快速检测。对 $1000 \sim 1800 \text{ cm}^{-1}$ 的古代丝、毛、棉、麻纺织品红外光谱数据进行主成分分析 (Principal component analysis, PCA), 并采用正交偏最小二乘判别分析法 (Orthogonal partial least square-discriminant analysis, OPLS-DA) 建立了 3 个分类模型。结果表明: 采用第一个模型的丝织品、毛织品和纤维素纤维纺织品的正确分类率为 100%, 采用第二个模型的蛋白质纤维纺织品的正确分类率为 98%, 采用第三个模型的丝织品和毛织品的正确分类率为 100%。红外光谱结合化学计量学的方法适用于古代纺织品快速检测。

关键词: 古代纺织品; 红外光谱; 化学计量学; 种属鉴别; 快速识别; 文物鉴定

中图分类号: O657.33

文献标志码: A

文章编号: 1673-3851 (2022) 07-0490-06

Identification of ancient textile relics based on infrared spectroscopy and chemometrics

ZHENG Haoran¹, ZHOU Jie¹, ZHOU Junyi¹,
CHEN Xushi¹, LIU Jian², WANG Bing¹

(1.School of Materials Science & Engineering, Zhejiang Sci-Tech University,
Hangzhou 310018, China; 2.China National Silk Museum, Hangzhou 310002, China)

Abstract: To solve the problem of species identification of ancient textile relics, infrared spectroscopy and chemometrics were used for the rapid detection of ancient textile species. IR spectral data of ancient silk, wool, cotton and linen textiles in the range of $1000 \sim 1800 \text{ cm}^{-1}$ were subjected to principal component analysis (PCA), and three classification models were established using orthogonal partial least square-discriminant analysis (OPLS-DA) method. The results showed that the correct classification rate of silk, wool and cellulose fiber textiles using the first model was 100%, the correct classification rate of protein fiber textiles and cellulose fiber textiles using the second model was 98%, and the correct classification rate of silk and wool fiber textiles using the third model was 100%, indicating that the method of combining infrared spectroscopy with chemometrics is suitable for the rapid detection of ancient textiles.

Key words: ancient textiles; IR spectroscopy; chemometrics; species identification; rapid recognition; cultural relics identification

0 引言

中国古代纺织品几乎贯穿了中国几千年的发展历程,造就了中国古代纺织文化的辉煌篇章。纺织品不仅具有御寒和装饰的作用,还有标识社会阶层和巩固文化认同的功能。纺织品是衡量不同时期人类进步和文明的标尺。大量的纺织品文物已被发掘出来,对它们进行研究,能够深入了解当时政治、经济以及文化的发展情况。

中国古代纺织品材料大致分为丝、毛、棉和麻四类,其中丝织品、毛织品为蛋白质纤维纺织品,棉织品、麻织品为纤维素纤维纺织品,属于有机质文物,很难在墓葬环境中长期保存,极易因温度和湿度的变化以及霉菌侵袭而发生老化降解^[1-2]。

目前常用的纺织品纤维鉴别方法主要分为显微镜观察法、理化性质鉴别法和光谱法^[3]。显微镜观察法主要是利用光学显微镜和电子显微镜,通过观察纤维横截面和纵截面的特征形态进行种属判定,但是古代出土纺织品往往降解严重,表现形态和纤维结构难以通过显微镜观察^[4]。理化性质鉴别法方法包括燃烧法、着色法、溶解法等,但这些方法会对珍贵的纺织品文物造成一定的损坏,并且往往涉及某些强酸强碱或有毒有害试剂。光谱法主要利用红外光谱(Infrared spectroscopy, IR),通过特征峰的位置、强度来判别种属,但部分降解严重文物样品的红外吸收峰强度较弱,并且红外光谱在测量过程中易受背景与噪声的干扰^[5],影响对样品的种属判断。综上所述,寻找一种有效的方法来鉴别古代纺织品非常必要。近年来,在机器学习基础上发展起来的化学计量学,已广泛运用于光谱领域^[6-7]。化学计量学在红外光谱领域中的诸多方法可以很好地提高模型的稳健性和预测结果的准确性^[8-9]。光谱预处理方法可以很好地改善红外光谱测量中遇到的问题,另外在谱图鉴别方面,化学计量学中的正交偏最小二乘判别分析法(Orthogonal partial least square-discriminant analysis, OPLS-DA)有着众多的应用^[10-12]。

本文将红外光谱结合化学计量学的方法应用于古代纺织品的检测分析中,首先利用衰减全反射傅里叶变换红外光谱法(Attenuated total internal reflectance fourier transform infrared spectroscopy, ATR-FTIR)来获取古代丝、毛、棉、麻纺织品的红外光谱数据,随后使用主成分分析对古代纺织品的分离和聚集趋势进行分析,接着用正交偏最小二乘判别分析法建立古代纺织品种属鉴定模型,通过模型实现对古代纺织

品种属的检测,具有快速无损等特点,为纺织品文物的识别提供了一个全新的、可行的方案,因此在纺织品考古领域有着较好的应用前景。

1 实验部分

1.1 样品集建立

实验样品集由中国丝绸博物馆从各考古遗址收集的220份古代纺织品样品组成。该样品集包括蛋白质纤维纺织品($P; n=210$, n 表示样品数量)和纤维素纤维纺织品($C; n=10$)。其中蛋白质纤维纺织品包括丝织品($S; n=100$)和毛织品($W; n=110$),纤维素纤维纺织品因为样品数较少不再进行细分。

1.2 红外光谱采集与分析

纺织品样品在20℃环境下静置30 min,利用红外光谱仪(Nicolet6700)采集其红外光谱,波数范围650~4000 cm^{-1} ,光谱分辨率为4 cm^{-1} 。数据选择范围1000~1800 cm^{-1} ,最终数据矩阵由415个变量(即波数)和220个观测值(即古代纺织品样品)组成。选取丝织品、毛织品和纤维素纤维纺织品样品中各2个降解程度不同的样品作为代表进行分析。

1.3 模型建立与验证

将原始数据矩阵导入光谱分析软件(The Unscrambler X v.10.4, Camo Software, Oslo, Norway)进行数据预处理。通过标准正态变量变换(Standard normal variate transformation, SNV)对光谱数据进行处理来消除表面散射带来的影响,并进行二阶导数(Second derivative, SD)和 Savitzky-Golay 平滑(Savitzky-Golay smoothing, SG)处理来提高分辨率,增强谱图的特征性。经过预处理后,将数据矩阵导入数据处理软件(SIMCA v.14.1, Umetrics, Umea, Sweden)进行多变量分析。

将预处理后的数据集进行初步的主成分分析(Principal component analysis, PCA),研究样品种间的分离趋势和种内的聚集趋势^[13]。随后对数据集进行正交偏最小二乘判别分析(Orthogonal partial least square-discriminant analysis, OPLS-DA),建立能够根据种属来区分样品的判别模型。通过建立3种不同的OPLS-DA模型来对纤维进行种属判定。第一个模型用来判定丝织品、毛织品和纤维素纤维纺织品,第二个模型用来判定蛋白质纤维纺织品和纤维素纤维纺织品,第三个模型用来判定丝织品和毛织品。

在进行OPLS-DA分析之前,将数据集按7:3进行划分,分别创建训练集($n=154$)和测试集($n=$

66),其中训练集被用来构建校准模型,测试集则用于模型的外部验证。

使用7倍交叉验证(Cross-validation,CV)将得到的OPLS-DA模型进行内部验证,并通过X轴方向模型的累计解释率(R^2X_{cum})和模型累计预测率(Q^2_{cum})来对模型进行评价,其中 R^2X_{cum} 可以用来评价模型的拟合优度, Q^2_{cum} 可以用来评价模型的预测优度。Y轴方向模型的累计解释率(R^2Y_{cum})、内部交叉验证均方根误差(Root mean square error from cross-validation, RMSECV)、校正模型均方根误差(Root mean square error of estimation, RMSEE)也是评估模型优劣的重要指标^[14]。通过置换检验来检查模型是否存在过度拟合的情况,当 R^2 和 Q^2 的y轴截距值分别小于等于0.4和0.05的时候,该模型可认为检测结果较为可靠^[15]。

最后通过外部测试集验证,通过预测均方根误差(Root mean square error of prediction, RMSEP)和判别准确率来评估模型的判别性能。

2 结果与讨论

2.1 红外光谱分析和初步化学计量分析

含丝、毛、纤维素纤维的6个古代纺织品样品在650~4000 cm⁻¹范围内的红外光谱图如图1所示。从图1中可以看出:各种纺织品的特征峰集中在1000~1800 cm⁻¹范围内,因此后续建模用到的红外数据也选择这一波数范围;对比同样为丝织品的红外光谱图,样品b的特征峰十分明显,而样品a的特征峰强度较弱,是由于样品本身降解较为严重;同样样品f也存在吸收峰吸收信号较弱的问题,并且所有纺织品的红外光谱均存在重叠吸收带问题。

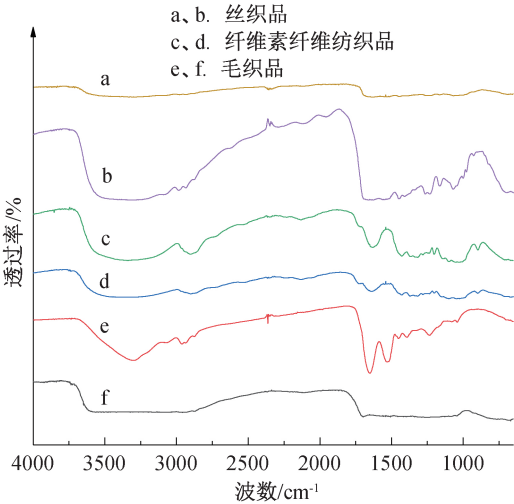


图1 古代纺织品样品的红外光谱图

通过进行SNV、SD、SG对红外光谱进行处理改善重叠吸收带和特征峰吸收峰较弱问题,丝织品、纤维素纤维纺织品、毛织品等6个古代纺织品样品的在650~4000 cm⁻¹范围内的二阶导数红外光谱图如图2所示。图2表明:样品a和样品b的二阶导数红外光谱图中均显示出了较强的特征峰吸收信号。在丝织品样品的红外光谱中,均在1697、1448、1166、1231 cm⁻¹附近出现强烈的吸收峰,其中,1697 cm⁻¹处的吸收峰是由酰胺I区的转角结构所产生的,1448 cm⁻¹处的吸收峰是由—CH₂的弯曲振动所产生的,1166 cm⁻¹处的吸收峰是由C—O伸缩振动所产生的,1231 cm⁻¹处的吸收峰是由酰胺III区的N—H弯曲和C—N伸缩振动组合吸收所产生的^[16-17]。在毛织品样品的红外光谱中,较为明显的吸收峰出现在1517、1450 cm⁻¹和1238 cm⁻¹附近,分别由酰胺II区N—H弯曲和C—N伸缩振动组合吸收、—CH₂的弯曲振动、酰胺III区的N—H弯曲和C—N伸缩振动组合吸收所产生的^[18]。对于纤维素纤维纺织品,1317 cm⁻¹和1161 cm⁻¹处的吸收峰分别是由—OH面内变形振动和纤维素中碳氧反对称伸缩振动所产生的^[19]。

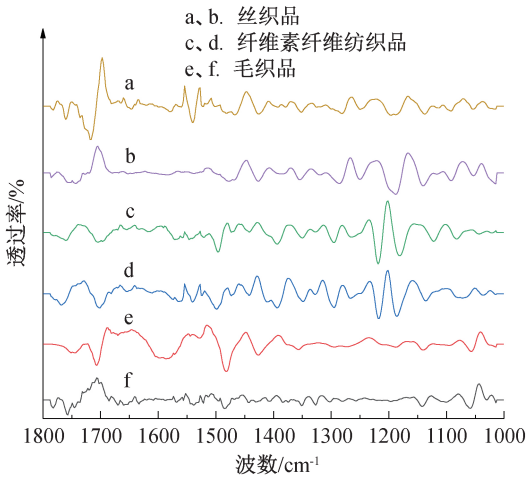


图2 古代纺织品样品的二阶导数红外光谱图

选择经过SNV-SD-SG预处理过后的古代纺织品样品的IR光谱(1000~1800 cm⁻¹)进行PCA分析,古代纺织品样品的主成分得分散点图如图3所示。图3表明:不同种属的纺织品都存在一定的聚类趋势,但是丝织品与毛织品、毛织品与纤维素纤维纺织品组别之间都存在重叠现象,难以根据种属进行样品区分。因此,通过主成分得分来对古代纺织品样品进行判别是不合适的。为得到优秀的判别模型,通过运用有监督的OPLS-DA模型对样品进行进一步分类。

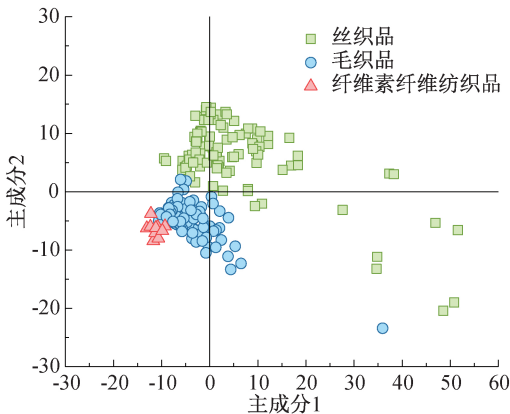


图 3 古代纺织品样品的主成分得分的散点图

2.2 丝织品、毛织品和纤维素纤维纺织品的 OPLS-DA 分析

第一个模型是利用 2 个预测主成分和 2 个正交主成分建立的能够区分丝织品、毛织品、纤维素纤维纺织品的分类模型。为检验模型的优劣,使用计算机辅助系统进行 CV 检验, CV 检验结果如表 1 所示。一般来说 R^2X_{cum} 、 Q^2_{cum} 的值越接近 1 越好, 高于 0.5 则视为较好^[20]。模型拥有 52% 以上的数据集变异性 ($R^2X_{cum}=0.528$), 预测能力约为 90% ($Q^2_{cum}=0.897$), 两个参数的数值均较高(表 1), 表明模型的拟合准确性较好。为检测模型建立过程中是否存在过度拟合的问题, 通过置换检验得到 R^2 和 Q^2 的 y 轴截距值分别为 0.173 和 -0.217, 证实了模型的有效性 (R^2y 轴截距值 ≤ 0.4 和 Q^2y 轴截距值 ≤ 0.05)。图 4(a) 是丝织品、毛织品、纤维素纤维纺织品的 OPLS-DA 得分散点图, 从图 4(a) 可以看出: 3 组样品在二维空间中呈现完全分离的状态, 丝织品样品沿第 1 预测主成分正向分布, 而毛织品和纤维素纤维纺织样品沿第 2 预测主成分负向分布, 并且毛织品和纤维素纤维纺织样品两组样品又分布在第 2 预测主成分正负两侧。此外, 3 组样品中丝织品样品组和纤维素纤维纺织样品组分别在水平方向和竖直方向上没有紧密分组, 但这并不影响 3 组样品的正确分组。通过基于测试集的外部检测来进一步验证 OPLS-DA 模型的准确性, 表 2 是根据外部验证所得判别结果。表 2 表明: 用来判别丝

织品、麻织品、纤维素纤维纺织品的 OPLS-DA 模型的正确判别率达到了 100%, 测试集中没有一个样品被错误分类, 并且显示出较低的 RMSEP 值 (0.147347)。

2.3 蛋白质纤维纺织品和纤维素纤维纺织品的 OPLS-DA 分析

考虑到丝织品和毛织品同为蛋白质纤维纺织品, 因此设计了第二个模型, 该模型根据蛋白质纤维纺织品和纤维素纤维纺织品两类来进行纺织品种属判别。首先根据 CV 检验来评估模型的分类能力, 从表 1 中可以看出: 相比于第一个模型, 该模型的 R^2X_{cum} 、 Q^2_{cum} 、 R^2Y_{cum} 的值都稍低, 仍保持在一个较好的水平。图 4(b) 表明: 纤维素纤维纺织样品和蛋白质纤维纺织样品两组之间有明显的分离现象, 蛋白质纤维纺织样品主要在第 1 预测主成分负侧分布, 纤维素纤维纺织样品则主要分布在第 1 预测主成分正侧。在外部测试集验证中, 模型的正确判别率为 98.48%, 唯一的分类错误出现在纤维素纤维纺织样品判别中, 其原因可能是因为纤维素纤维纺织样品数量太少, 模型在建立时缺乏足够的样本。

2.4 丝织品和毛织品样品的 OPLS-DA 分析

目前出土的纺织品纤维纺织品主要为丝织品和毛织品, 因此设计了第三个 OPLS-DA 模型来对丝织品和毛织品进行种属判别。对总共由 147 个样品组成的训练集进行 OPLS-DA, 其中丝织品样品数为 70, 毛样品织品数为 77。模型的 Q^2_{cum} 和 R^2Y_{cum} 值都是 3 个模型中最高的, 表明模型的类分离表现良好。毛织品样品组在第 1 预测主成分上沿正向分布, 丝织品样品沿第 1 预测主成分负向分布, 两组样品之间没有重叠现象(见图 4(c))。丝织品样品组的样本分布并不紧密, 表明存在一定的组内变异性, 但这并不影响通过沿水平方向的第 1 预测主成分来实现样品的正确分类。外部测试集的验证结果也表现了模型的准确性, 外部测试结果如表 2 所示, 表 2 表明, 该模型对测试集中的 30 个丝织品样品和 33 个毛织品样品均实现了正确分类。

表 1 OPLS-DA 模型计算的交互验证统计信息

OPLS-DA 模型	交叉验证				置换检验			
	NC	R^2X_{cum}	R^2Y_{cum}	Q^2_{cum}	RMSECV	RMSEE	R^2y -Intercept	Q^2y -Intercept
S vs. W vs. C	2+2	0.528	0.915	0.897	0.141648	0.123399	0.173	-0.217
P vs. C	1+2	0.472	0.893	0.868	0.075649	0.069003	0.242	-0.270
S vs. W	1+2	0.486	0.945	0.923	0.138360	0.118570	0.235	-0.227

注: NC 为 OPLS 成分数(预测+正交); R^2y -Intercept, Q^2y -Intercept 为置换检验中通过执行 200 次交互计算得出的 R^2y 轴截距值和 Q^2y 轴截距值; S 为丝织品; W 为毛织品; C 为纤维素纤维纺织品; P 为蛋白质纤维纺织品。

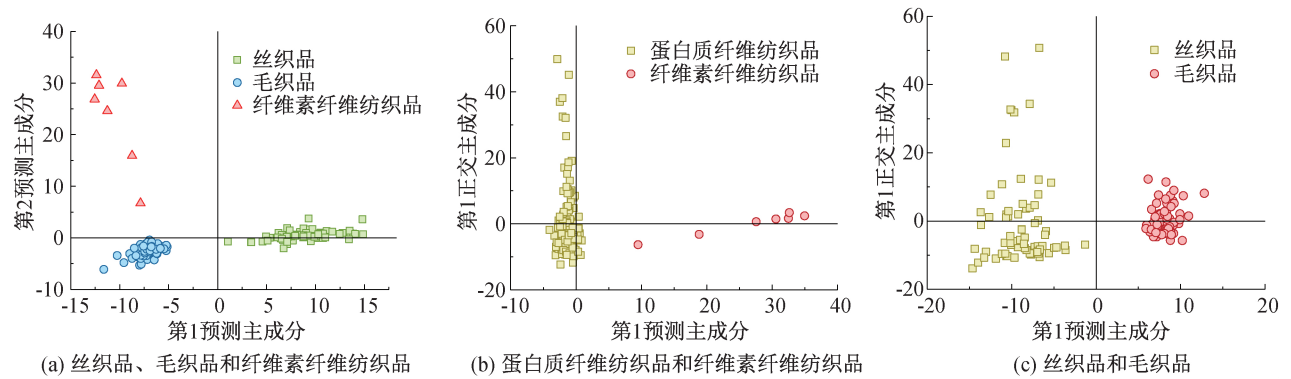


图 4 古代纺织品 OPLS-DA 模型得分散点图

表 2 根据外部验证所得判别结果				
OPLS-DA 模型	类别	单一判别率/%	RMSEP	判别率/%
S vs. W vs. C	S	100	0.147347	100
	W	100		
	C	100		
P vs. C	P	100	0.092467	98.48
	C	66.67		
S vs. W	S	100	0.143467	100
	W	100		

注:S为丝织品;W为毛织品;C为纤维素纤维纺织品;P为蛋白质纤维纺织品。

3 结 论

本文为检测古代纺织品,采用化学计量学结合红外光谱的方法,通过 OPLS-DA 分析,构建了无监督的定性检测模型,并对所建立的定性模型进行验证,实现了对纺织品文物的种属判定。主要结论如下:

a) OPLS-DA 用于红外光谱技术中构建的 OPLS-DA 模型能够准确地识别丝织品、毛织品和纤维素纤维纺织品,并且模型通过 7 倍 CV 进行内部验证和独立的测试集进行外部验证,统计输出的结果表明了模型优异的预测判别能力以及模型本身并不存在过度拟合的现象。

b) OPLS-DA 结果表明,即使纺织品降解程度不同,OPLS-DA 模型都能很好地区分丝织品、毛织品和纤维素纤维纺织品,可以实现降解较为严重和保存情况良好的纺织品文物检测。

通过本文的研究发现,OPLS-DA 结合红外光谱的方法在古代纺织品种属鉴别方面有着不错的应用潜力。与传统的分析方法相比,这种方法具有快速、简单等优点。然而要想通过该方法来建立优异的分类模型,仍然存在一些挑战,其中模型样本数据库的规模是最为关键的问题。在后续的研究中,将

继续扩充数据库的规模,为古代纺织品种属鉴定提供数据支持。

参考文献:

[1] Li J, Zheng H L, He Y J, et al. Ultrasensitive electrochemical immunosensor reveals the existence of silk products on the maritime silk road [J]. ACS Sensors, 2019, 4(12): 3203-3209.

[2] Liu J, Guo D H, Zhou Y, et al. Identification of ancient textiles from Yingpan, Xinjiang, by multiple analytical techniques[J]. Journal of Archaeological Science, 2011, 38(7): 1763-1770.

[3] 姚敏. 纤维鉴别在纺织品修复保护中的作用[J]. 文物鉴定与鉴赏, 2021(6): 104-106.

[4] 王钟元, 宋珊珊, 翟雨洁, 等. 免标记型免疫传感器的构建及其对毛织品微痕迹的检测[J]. 浙江理工大学学报(自然科学版), 2021, 45(4): 478-482.

[5] 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.

[6] Gao R, Chen C, Wang H, et al. Classification of multicategory edible fungi based on the infrared spectra of caps and stalks [J]. PLoS One, 2020, 15(8): e0238149.

[7] Sharma S, Chopri R, Singh R. Forensic discrimination of menstrual blood and peripheral blood using attenuated total reflectance (ATR)-Fourier transform infrared (FT-IR) spectroscopy and chemometrics [J]. International Journal of Legal Medicine, 2020, 134(1): 63-77.

[8] 付建华. 近红外光谱分析技术在纺织品检测中的应用[D]. 杭州: 中国计量大学, 2016: 8-15.

[9] Jandrić Z, Haughey S A, Frew R D, et al. Discrimination of honey of different floral origins by a combination of various chemical parameters[J]. Food Chemistry, 2015, 189: 52-59.

- [10] Tigabu M, Daneshvar A, Wu P F, et al. Rapid and non-destructive evaluation of seed quality of Chinese fir by near infrared spectroscopy and multivariate discriminant analysis[J]. *New Forests*, 2020, 51(3): 395-408.
- [11] Qin H W, Wang Y Z, Yang W Z, et al. Comparison of metabolites and variety authentication of *Amomum tsao-ko* and *Amomum paratsao-ko* using GC - MS and NIR spectroscopy[J]. *Scientific Reports*, 2021, 11(1): 15200.
- [12] Jolayemi O S, Tokatli F, Buratti S, et al. Discriminative capacities of infrared spectroscopy and e-nose on Turkish olive oils [J]. *European Food Research and Technology*, 2017, 243(11): 2035-2042.
- [13] 林先凯, 黄小凯, 王宁, 等. 红外光谱法结合主成分分析快速识别 3 种天然皮革[J]. *理化检验—化学分册*, 2019, 55(2): 152-155.
- [14] Ghidini S, Varrà M O, Dall'Asta C, et al. Rapid authentication of European Sea bass (*Dicentrarchus labrax* L.) according to production method, farming system, and geographical origin by near infrared spectroscopy coupled with chemometrics [J]. *Food Chemistry*, 2019, 280: 321-327.
- [15] Varrà M O, Fasolato L, Serva L, et al. Use of near infrared spectroscopy coupled with chemometrics for fast detection of irradiated dry fermented sausages[J]. *Food Control*, 2020, 110: 107009.
- [16] 苑萌萌, 陈长洁, 岳静, 等. 基于红外光谱的古代丝绸蛋白质分子结构变化分析[J]. *丝绸*, 2013, 50(12): 7-10.
- [17] 陈茹茹. 基于蛋白质组学和免疫学技术的丝绸文物分析研究[D]. 杭州: 浙江理工大学, 2020: 19-21.
- [18] 王敏. 红外光谱对混纺纤维的定性和定量方法研究[D]. 杭州: 浙江理工大学, 2014: 7-27.
- [19] 罗曦芸, 杜一平, 沈美华, 等. 红外光谱在纤维质文物材料鉴别中的应用研究[J]. *光谱学与光谱分析*, 2015, 35(1): 60-64.
- [20] Zhou P, Hu O, Fu H Y, et al. UPLC-Q-TOF/MS-based untargeted metabolomics coupled with chemometrics approach for Tieguanyin tea with seasonal and year variations [J]. *Food Chemistry*, 2019, 283: 73-82.

(责任编辑: 廖乾生)