

不同厂家冬凌草片的近红外光谱主成分聚类分析

石杰¹, 李长滨¹, 吴拥军², 黄伟¹, 严会会¹, 王丽杰¹

(1. 郑州大学 化学系 河南 郑州 450001; 2. 郑州大学 公共卫生学院 河南 郑州 450001)

摘要: 利用近红外光谱对不同厂家的冬凌草片进行鉴别归属, 并进行快速分类. 采集不同厂家不同批次冬凌草片样品的近红外漫反射光谱, 通过不同计量学方法进行处理. 结合主成分聚类分析法对不同厂家冬凌草片进行聚类判别分析, 分类识别正确率达 100%, 从而推断出不同厂家冬凌草片成分及生产工艺上的差别. 结果表明, 该方法准确、快速、简便, 可用于不同厂家冬凌草片的分类鉴别和质量控制.

关键词: 近红外光谱; 主成分; 聚类分析; 冬凌草片

中图分类号: O 657.3 **文献标志码:** A **文章编号:** 1671-6841(2011)04-0067-04

0 引言

作为一种新兴的检测手段, 近红外漫反射光谱技术可用于对样品的直接测定, 无需破坏样品及制样, 操作简便、快速. 近红外光谱(NIRS)技术是利用全谱或部分波段的光谱数据进行定性或定量分析的无损检测技术, 广泛应用于烟草^[1]、食品^[2]、石油化工^[3]等领域. 近年来, 近红外光谱技术已发展成为中药定性判别和定量分析的新兴方法^[4-6]. 由于不同厂家的生产工艺和原辅料组成均存在差异, 反映在近红外光谱图上使得对其进行分类鉴别成为可能.

冬凌草为唇形科香茶属多年生草本或亚灌木, 自然分布于太行山南部. 其味甘苦, 性微寒, 具有清热解毒、消炎止痛、健胃活血之功效. 对急慢性咽炎、扁桃体炎、口腔炎有显著疗效^[7]. 冬凌草片由冬凌草浸膏制成, 收载于卫生部药品标准. 作者采用近红外光谱结合主成分聚类分析法, 实现了对不同厂家冬凌草片快速、准确的分类鉴别, 对于冬凌草生产工艺的改进和假药的判别有一定的借鉴意义.

1 实验部分

1.1 仪器与样品

Antaris 6700 型近红外光谱仪(美国 Thermo Nicolet 公司), 配有铟镓砷(InGaSn)检测器、积分球漫反射采样系统、样品旋转器和石英样品杯; 石英研钵; OMNIC 光谱采集软件和 TQ 8.0 分析软件; 德国 Bruker 公司开发的 OPUS 5.5 分析软件; 药典筛; BP210S 型万分之一分析天平(德国赛多利斯公司).

河南省济源市济世药业有限公司生产的冬凌草片(糖衣片) 22 批和冬凌草片(复方含片) 22 批, 河南省安阳市华安药业有限公司生产的冬凌草片 23 批, 河南省安阳市路德药业有限公司生产的冬凌草片 25 批, 河南广宇博科生物制药有限公司生产的冬凌草片 6 批, 共计 98 个批次.

1.2 实验方法

分别取不同批次的冬凌草片剂 20 片, 磨碎并过 100 目药典筛, 混匀后取 8 g 放入测量杯, 扣除背景采集其近红外光谱. 采样方式为积分球漫反射, 光谱范围为 12 000 ~ 4 000 cm^{-1} , 分辨率为 8 cm^{-1} , 扫描次数为 32 次, 温度 25 ~ 27 $^{\circ}\text{C}$, 相对湿度 45% ~ 50%. 每个样品重复测定 3 次, 求其平均值作为样品的标准光谱. 每次扫描前都要对样品进行混匀, 并用同一重物压实以保证每次的密实程度一致. 预处理方法为一阶导数 + 矢量

收稿日期: 2011-04-29

基金项目: 国家自然科学基金资助项目, 编号 J0830412; 河南省重大公益科研项目, 编号 081100912500.

作者简介: 石杰(1950-), 男, 教授, 主要从事食品与药物分析研究, E-mail: shijie@zzu.edu.cn; 通讯作者: 李长滨(1983-), 男, 硕士研究生, 主要从事近红外技术在药物分析和新药开发方面的研究, E-mail: lichangbin911@163.com.

©1994-2021 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

归一化,聚类方法为 Ward 算法.98 个批次冬凌草片的近红外原始光谱叠加图和一阶导数光谱叠加图见图 1.

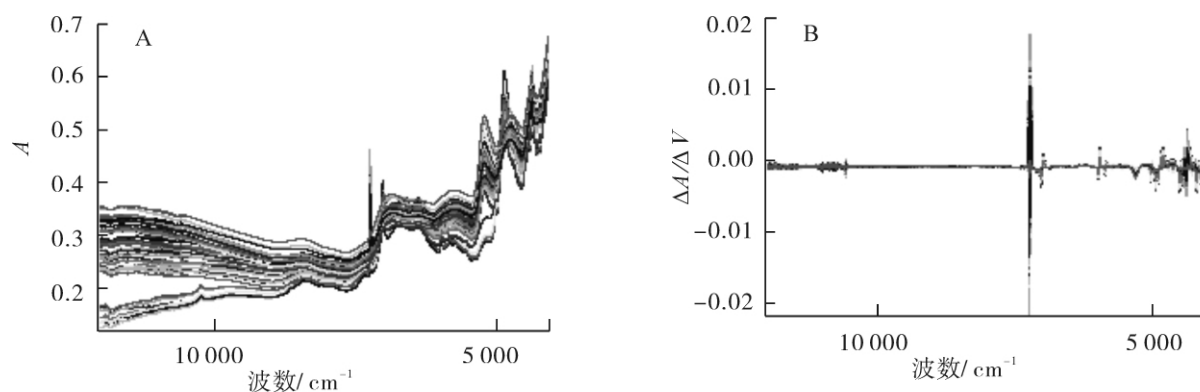


图1 98 个批次冬凌草片的近红外原始光谱(A) 及一阶导数光谱(B)

Fig.1 The original NIR spectrum (A) and first derivative spectrum (B) of donglingcao tablets of 98 batches

2 参数优化

2.1 建模波段的选择

建模前对光谱波段进行筛选,既可避免引入过多冗余信息,也可避免由于波段选择较窄而丢掉一些有用的信息.冬凌草片的光谱曲线在高频段噪声较大,通过谱区比较,发现谱图在 $7\ 283 \sim 4\ 008\ \text{cm}^{-1}$ 范围内信息量丰富,各种特征信息表现明显,故取此波段进行数据处理.

2.2 光谱预处理方法的选择

高频随机噪声、基线漂移、样本不均匀、光散射等因素都会对模型的预测准确性和稳定性产生较大的影响,因此需要对光谱进行预处理以消除这种影响对预测带来的误差.由图 1 可知,对样品的原始谱图进行一阶导数处理可以有效地消除样品由于颜色差别引起的基线漂移,使谱图特征信息更加明显地表达出来.通过比较样品的近红外原始光谱及一阶导数光谱,发现一阶导数光谱的鉴别效果优于原始光谱.矢量归一化用以消除光程变化或样品颗粒大小、密实程度等变化对光谱产生的影响,选择样品的近红外谱图 9 点平滑一阶导数 + 矢量归一化方法进行优化处理.

3 样品的主成分聚类分析及评价

3.1 主成分分析

为了充分利用光谱信息,必须对光谱数据进行降维处理.主成分分析是常用的有效降维方法,保留多少主成分取决于保留部分的方差和在方差总和中所占的百分比(即累积贡献率),它标志着前几个主成分概括信息的多寡^[8].本实验对一阶导数 + 矢量归一化法处理后的光谱数据进行主成分分析,主成分数对光谱的影响见表 1,可以看出,不同主成分数对鉴别结果起着关键的作用,前 5 个主成分的累积贡献率已经达到 98.08%,主成分数增加到 6,累积贡献率几乎不变.综合考虑选择前 5 个主成分来概括表示样品近红外光谱的主要信息,此时可以把不同厂家的样品区分开,分类结果与实际归属一致.

表 1 主成分数与累积贡献率和错判例数的关系

Tab.1 The relationship of principal components, cumulative reliabilities and misjudge number

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7
累积贡献率/%	78.24	96.68	97.56	97.86	98.08	98.23	98.35
错判例数	23	0	0	0	0	0	0

3.2 聚类分析

将98份样品中的76份作为校正集用于建立判别分析模型,其余22份作为验证集用于检验所建模型的性能.校正集样品的原始光谱经过一阶导数+矢量归一化法处理后,结合判别分析在 $7\ 283\sim 4\ 008\ \text{cm}^{-1}$ 波段内建立判别分析模型,并对验证集进行判别预测.结果表明,该模型对校正集样品的判别和验证集样品的预测鉴别都与实际分类完全一致,判别分析结果见图2.

聚类分析是一种无管理模式识别方法,常用于目标观测对象的分类,即利用表征观测对象的一组变量对目标进行分类^[9].因此,在主成分分析的基础上引入聚类分析,可排除众多化学信息中相互重叠的部分而又不丢失原始数据信息,使聚类分析的计算简化.本实验采用Ward算法对冬凌草片进行聚类分析,结果5个厂家的冬凌草片被清晰地聚为5类.由此可见,建立的模型性能较好,能够准确、快速地判别不同厂家的冬凌草片.由树状图可知,聚类分析能够通过光谱间距离的大小,准确地把同一厂家的样品先聚为一类,然后再把与此厂家光谱性质相近的样品聚类,最后与其他厂家的产品聚类,依次进行聚类直至把所有产品聚类完成.图3中冬凌草片的聚类分析结果与实际分类结果一致.通过聚类分析,找出不同厂家样品之间的信息特征,从而达到了鉴别的目的.

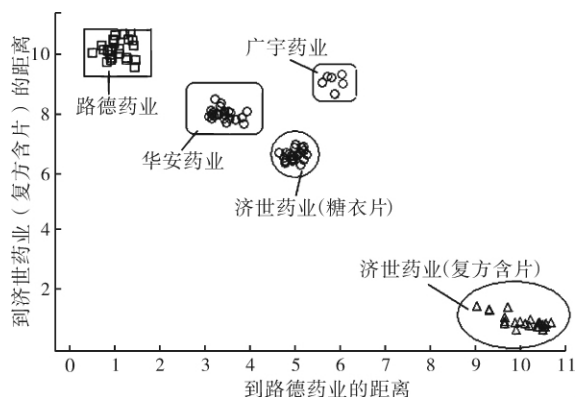


图2 冬凌草片的判别分析图

Fig. 2 Discriminate analysis diagram of donglingcao tablets

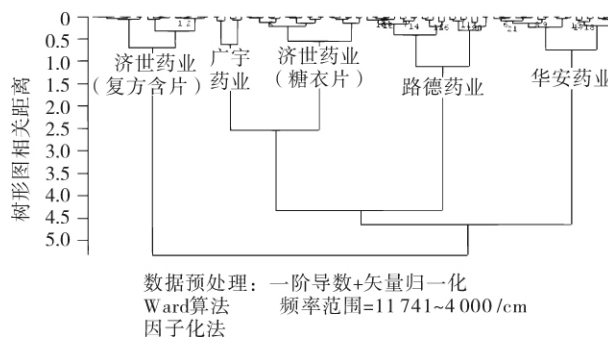


图3 冬凌草片的聚类分析图

Fig. 3 Clustering analysis diagram of donglingcao tablets

4 结论

1) 中药的化学成分比较复杂,且由于绝大部分中药辅料都含有淀粉、糊精、蔗糖等,成分相似,NIRS信息重叠严重,各样品的特征信息很难从峰位、峰形、峰强等方面直观鉴别,因此,必须借助于化学计量学的方法对谱图进行优化处理,突出样品之间化学成分的细小差异,从而达到鉴别的目的.不同厂家的冬凌草片由于其化合物的种类和含量、药品采集以及制备方法存在差异,结合近红外光谱技术和计量学处理技术,使得不同厂家的样品分类判别成为可能.

2) 外界环境和外部因素对光谱的重现性影响比较大,因此,测样时间要尽量集中并保持外部环境一致.作者对药品颗粒度、装样量、密实程度进行了统一处理,以增强不同厂家样品光谱间的可比性,减少了外部因素引起的实验误差.

3) NIRS聚类分析结果与实际分类完全吻合,说明NIRS分析技术具有较好的重现性和可靠性.实验通过将NIRS与主成分聚类分析法相结合,建立了鉴别不同厂家冬凌草片的新方法.结果表明,该方法准确、快速、无损,可用于不同厂家冬凌草片的判别分析,同时对冬凌草片生产工艺的改进和冬凌草片假药的初步判别也有一定的借鉴意义.

参考文献:

[1] 张鼎方.近红外技术应用于成品卷烟的内在质量稳定性控制[J].福建分析测试,2009,18(2):39-44.
 (C)1994-2022 China Academic Electronic Journal Publishing House. All rights reserved. http://www.cnki.net

- [2] 孙通,徐惠荣,应义斌. 近红外光谱分析技术在农产品/食品品质在线无损检测中的应用研究进展[J]. 光谱学与光谱分析, 2009, 29(1): 122-126.
- [3] 钱平,孙国琴,张存洲. 基于近红外光谱技术的石油组分定量分析新方法[J]. 光谱学与光谱分析, 2008, 28(12): 2851-2854.
- [4] 李艳英,白雁,陈志红,等. 近红外漫反射光谱法测定不同厂家一清颗粒的含量[J]. 药物分析杂志, 2009, 29(7): 1126-1129.
- [5] 吴拥军,李伟,相秉仁,等. 近红外光谱技术用于白芷类中药的鉴定研究[J]. 中药材, 2001, 24(1): 26-28.
- [6] 聂黎行,王刚力,李志猛,等. 近红外光谱法在中药辅料质量控制中的应用[J]. 中国中药杂志, 2009, 34(17): 14-17.
- [7] 刘俊鹏,于玲,王亚娟. 冬凌草片质量标准研究[J]. 药物研究, 2008, 46(13): 31-32.
- [8] 王丽,何鹰,王颜萍,等. 近红外光谱技术结合主成分聚类分析判别海面溢油种类[J]. 海洋环境科学, 2004, 23(2): 58-60.
- [9] 刘木清,周德成,徐新元,等. 聚类算法用于中药材的近红外光谱分析[J]. 光谱学与光谱分析, 2007, 27(10): 1985-1988.

Clustering Analysis for Principal Components of Donglingcao Tablets from Different Manufacturers by Near Infrared Spectroscopy

SHI Jie¹, LI Chang-bin¹, WU Yong-jun², HUANG Wei¹, YAN Hui-hui¹, WANG Li-jie¹
(1. Department of Chemistry, Zhengzhou University, Zhengzhou 450001, China;
2. College of Public Health, Zhengzhou University, Zhengzhou 450001, China)

Abstract: Near infrared spectroscopy (NIRS) was used in the identification and rapid classification of donglingcao tablets from different manufacturers. NIRS technology was employed to collect near infrared spectra of donglingcao tablets. Different chemometrics methods and clustering analysis for principal components were used to discriminate donglingcao tablets. The correct classification rate was 100%, and it could be concluded that the composition and production process of donglingcao tablets were different. The results indicated that the method was accurate, rapid, simple, and it could be applied to classification and quality controlling of donglingcao tablets produced by different manufacturers.

Key words: near infrared spectroscopy; principal component; clustering analysis; donglingcao tablet