·论著·

基于NIRS技术和PCA-SVM算法快速鉴别国产和进口啤酒花

郭云香[1,2,3]，陈 龙[4,5]，李晓瑾[1,2]，王果平[2]，蒋益萍[3]，辛海量[3]，贾晓光[1,2]（1 新疆医科大学 中医学院，新疆 乌鲁木齐 830011；2 新疆维吾尔自治区中药民族药研究所，国家中医药管理局新疆中药民族药资源重点实验室，新疆 乌鲁木齐 830002；3 海军军医大学药学院生药学教研室，上海 200433；4 襄阳市中心医院 湖北文理学院附属医院，湖北 襄阳 441021；5 湖北中医药大学 中药资源和中药复方教育部重点实验室，湖北 武汉 430065）

［摘要］ **目的** 利用近红外漫反射光谱（near-infrared reflectance spectroscopy，NIRS）法，结合主成分分析（principal component analysis，PCA）和支持向量机（support vector machine，SVM）联用算法，建立PCA-SVM的NIR模式识别模型，用于国产和进口啤酒花的快速鉴别。**方法** 收集上述不同产地的啤酒花样品，制备成均匀粉末，在4 000～12 500 cm^{-1}光谱区，采集各样品粉末的NIR光谱，选取特征谱段9 000～4 100 cm^{-1}为建模谱段，分别采用不同光谱预处理方法进行预处理并分别进行PCA降维。根据2维主成分平面散点图，优选最佳预处理方法。利用最佳预处理方法处理后的光谱PCA降维数据，建立SVM模式识别模型，SVM模型参数采用网格搜索法、遗传算法（GA）、粒子群优化法（PSO）进行寻优。对比不同主成分数所建PCA-SVM模型的预测准确率，确定最佳的主成分数，最终建立PCA-SVM的NIR快速鉴别模型。**结果** 在6500～5400 cm^{-1}谱段，以一阶导数法（first derivative，FD）为最佳光谱预处理方法，PCA提取的光谱前8个主成分为最佳主成分，并经网格搜索法确定最佳SVM内部参数：惩罚因子c=2，核函数参数g=1，建立啤酒花PCA-SVM鉴别模型，该模型五折交叉验证准确率达97.37%，对校正集和测试集样品预测准确率均分别为97.37%和97.44%。**结论** 啤酒花NIRS光谱，进行PCA-SVM算法建模，模型预测准确率高、性能佳，可用于啤酒花样品的快速、无损鉴别。

［关键词］ 啤酒花；近红外漫反射光谱；主成分分析；支持向量机；定性鉴别
［中图分类号］ R282.5 ［文献标志码］ A ［文章编号］ 1006-0111(2019)04-0322-10
［DOI］ 10.3969/j.issn.1006-0111.2019.04.008

Rapid identification of domestic and imported hops based on NIRS technology and PCA-SVM algorithm

GUO Yunxiang[1,2,3], CHEN Long[4,5], LI Xiaojin[1,2], WANG Guoping[2], JIANG Yiping[3], XIN Hailiang[3], JIA Xiaoguang[1,2] (1. Traditional Chinese Medicine College of Xinjiang Medical University, Urumqi 830011, China; 2. Xinjiang Institute of Traditional Chinese Medicine and Ethic Medicine, Key Laboratory of Traditional Chinese Medicine and Ethnic Medicine Resources, State Administration of Traditional Chinese Medicine, Urumqi 830002, China 3. Department of Pharmacognosy, School of Pharmacy, Naval Medical University, Shanghai 200433, China; 4. Xiangyang Central Hospital, Affiliated to Hubei University of Arts and Science, Xiangyang 441021, China; 5. Key Laboratory of Traditional Chinese Resource and Compound Prescription, Hubei University of China Medicine, Wuhan 430065, China)

［**Abstract**］ **Objective** To develop a rapid identification method for domestic and imported hops by the establishment of PCA-SVM model using near-infrared reflectance spectroscopy (NIRS), combined with principal component analysis (PCA) and support vector machine (SVM) algorithm. **Methods** The hop samples from different sources were collected and ground into uniform powder. The NIR spectra of each powder sample were collected in the range of 4000～12500 cm^{-1}. The characteristic spectrum segment was selected from 9000～4100 cm^{-1}, which was pretreated by different spectral pretreatment methods and subjected to PCA dimensionality reduction. According to the 2-dimensional principal component plane scatter plot, the pretreatment method was optimized. The SVM pattern recognition model was established by using the best preprocessing method to process the PCA dimensionality reduction data of the post-spectrum. The SVM model parameters were searched by grid search method, genetic algorithm (GA) and particle swarm optimization (PSO). The prediction accuracy of the PCA-SVM models built

［基金项目］ 国家自然科学基金（U1603283）
［作者简介］ 郭云香，硕士研究生，研究方向：中药药理抗骨质疏松研究，Email：2468463128@qq.com
［通讯作者］ 贾晓光，教授，研究方向：新疆民族药物的研究与开发，Email：xgjia@vip.sina.com.cn；辛海量，博士，副教授，研究方向：中药资源、中药（抗骨质疏松）药理学，Email：hailiangxin@163.com

by different principal component numbers were compared to determine the optimal principal component number. Finally, the rapid NIR identification model of PCA-SVM is established. **Results** In the 6500～5400 cm^{-1} spectral segment, the first derivative (FD) is the best spectral pretreatment method, and the first 8 principal components are the best principal components of the spectrum extracted by PCA. The optimal SVM internal parameters are determined by the grid search method: the penalty factor (c)＝2, the kernel function parameter(g)＝1. The prediction accuracy rate of this hop PCA-SVM identification model was 97.37% for the 5-fold cross validation, 97.37% for the calibration set and 97.44% for test set samples. **Conclusion** This model has high accuracy and consistent performance. It can be used for rapid and non-destructive identification of hop samples.

[**Key words**] Hop; near-infrared diffuse reflectance spectroscopy; principal component analysis; support vector machine; qualitative identification

啤酒花 *Humulus lupulus* L. 为桑科葎草属啤酒花的干燥雌性球穗状花序，不仅是酿造啤酒的重要添加原料[1]。在我国还作为民族药使用。被收录于《新疆药用植物志》《内蒙古植物药志》《宁夏中药志》《哈萨克药物志》《四川中药志》等，具有止咳化痰、健胃、消食、镇静、利尿的功效，为药食两用的新疆特色资源植物[2]。在欧洲，啤酒花提取物用于缓解更年期的潮热不适以及绝经后骨质疏松症[3]。传统啤酒花生药的鉴别方法有显微鉴定、理化鉴定[4]及非线性化学指纹图谱[5-6]等，但这些方法都存在实验复杂、检测时间过长等缺点。近红外漫反射光谱技术(NIR)其无损、快速、准确的优点能够反映样品的综合信息，在植物药[7]、动物药[8]和矿物药[9]中均有涉及，能反映分子中 C-H、N-H、O-H 基团基频振动的倍频吸收与合频吸收。本研究将运用近红外漫反射光谱(NIRS)技术，将主成分分析(PCA)和支持向量机(SVM)等化学计量学算法相结合，建立快速无损的 PCA-SVM 识别模型，用于国产和进口啤酒花的快速鉴别。

1　仪器与材料

1.1　仪器

MPA 傅里叶变换近红外光谱仪(德国布鲁克光学仪器公司，配备固体积分球漫反射附件)，OPUS 7.5 采集和处理软件(德国布鲁克光学仪器公司)，MATLAB R2014b 软件(美国 Math Works 公司)。

1.2　样品

2017 年采集的国内外不同地方的啤酒花样品均经第二军医大学药学院生药教研室辛海量副教授鉴定，并密封存放于干燥阴凉处，详细采集信息见表 1。

表 1　啤酒花样品产地来源及采集时间

编号	产地	采集时间	样品集	编号	产地	采集时间	样品集
PJH-01	新疆昌吉	2017 年 10 月	测试光谱	PJH-19	新疆阜康	2017 年 10 月	测试光谱
PJH-02	甘肃	2018 年 9 月	校正	PJH-20	新疆霍尔果斯	2017 年 10 月	校正
PJH-03	新疆吉木乃	2017 年 10 月	校正	PJH-21	美国	2017 年 4 月	校正
PJH-04	甘肃	2017 年 10 月	测试光谱	PJH-22	捷克	2017 年 4 月	测试光谱
PJH-05	新疆玛纳斯	2017 年 10 月	校正	PJH-23	新疆沙湾	2017 年 10 月	校正
PJH-06	新疆昌吉	2017 年 10 月	校正	PJH-24	新疆阿勒泰	2017 年 10 月	校正
PJH-07	甘肃	2018 年 9 月	测试光谱	PJH-25	新疆昌吉	2017 年 10 月	测试光谱
PJH-08	新疆托里	2017 年 9 月	校正	PJH-26	新疆布克赛尔	2017 年 10 月	校正
PJH-09	甘肃	2018 年 9 月	校正	PJH-27	新疆昌吉	2017 年 3 月	校正
PJH-10	新疆阜康	2017 年 10 月	测试光谱	PJH-28	新疆昌吉	2017 年 3 月	测试光谱
PJH-11	新疆吉木萨尔	2017 年 10 月	校正	PJH-29	新疆塔城	2017 年 10 月	校正
PJH-12	新疆哈密	2017 年 10 月	校正	PJH-30	新疆	2016 年 5 月	校正
PJH-13	德国	2017 年 4 月	测试光谱	PJH-31	新疆塔城	2017 年 10 月	测试光谱
PJH-14	新疆吉木萨尔	2017 年 10 月	校正	PJH-32	新疆布尔津	2017 年 10 月	校正
PJH-15	新疆裕民	2017 年 10 月	校正	PJH-33	新疆吉木萨尔	2017 年 10 月	校正
PJH-16	新疆沙雅	2017 年 10 月	测试光谱	PJH-34	新疆喀什	2017 年 10 月	测试光谱
PJH-17	新疆阜康	2017 年 10 月	校正	PJH-35	新疆博乐	2017 年 10 月	校正
PJH-18	新疆阿勒泰	2017 年 10 月	校正	PJH-36	新疆奇台	2017 年 10 月	校正

（续表 1）

编号	产地	采集时间	样品集	编号	产地	采集时间	
PJH-37	美国	2017 年 4 月	测试光谱	PJH-47	德国	2017 年 4 月	校正
PJH-38	美国	2017 年 4 月	校正	PJH-48	美国	2017 年 4 月	校正
PJH-39	英国	2017 年 4 月	校正	PJH-49	德国	2017 年 4 月	测试光谱
PJH-40	美国	2017 年 4 月	测试光谱	PJH-50	美国	2017 年 4 月	校正
PJH-41	美国	2017 年 4 月	校正	PJH-51	美国	2017 年 4 月	校正
PJH-42	美国	2017 年 4 月	校正	PJH-52	美国	2017 年 4 月	测试光谱
PJH-43	德国	2017 年 4 月	测试光谱	PJH-53	美国	2017 年 4 月	校正
PJH-44	德国	2017 年 4 月	校正	PJH-54	美国	2017 年 4 月	校正
PJH-45	美国	2017 年 4 月	校正	PJH-55	德国	2017 年 4 月	测试光谱
PJH-46	美国	2017 年 4 月	测试光谱	PJH-56	广西	2017 年 3 月	校正

2　方法与结果

2.1　近红外光谱采集

共计 56 批样品，分别取 2g 置于样品瓶中，采用积分球漫反射测试模式扫描 NIR 光谱。光谱扫描范围 4 000～12 500 cm^{-1}，扫描次数 32 次，仪器分辨率为 8 cm^{-1}。每个样品重复扫描 3 次，取平均值作为该样品的分析光谱。所有样品 NIR 光谱见图 1。

由图 1 可知，啤酒花生药的 NIRS 主要特征谱段在 9 000～4 100 cm^{-1}，>9 000 cm^{-1} 谱段主要表现为基线漂移和噪声，<4 100 cm^{-1} 为末端吸收，故二者均不用于建模分析，初始建模谱段在 9 000～4 100 cm^{-1}，该谱段的样品的 NIRS 差异较大。利于鉴别国产、进口啤酒花。根据文献可知[10]，啤酒花主要成分包含 CO_3^{2-} 和 HO^{-1}，5 000～4 100 cm^{-1} 范围内会有 C-O 键特征振动峰，在 9 000～7 500 cm^{-1}，6 500～5 400 cm^{-1}，有 H-O 的特征吸收峰。我们可以通过特征谱图差异区分鉴别啤酒花，但因理化性质差异，含氢基团的倍频和合频振动频率差异，在该谱区，吸收峰重叠严重、吸收强度弱、解析困难、无法直观看出等，故采用 NIRS 技术结合化学计量学方法对啤酒花进行优选鉴别。

2.2　样本集划分及类别标签值设定

将 56 批样品按 2∶1 比例随机分为校正集和测试集 2 个子集，校正集训练模型，并以内部交叉验证法验证模型性能，测试集对所建模型进行预测能力评价。采用矢量归一化法（VN）、一阶导数（FD）、二阶导数（SD）对样品进行光谱预处理，运用 PCA-SVM 算法，以 RBF（高斯径向基核函数）为核函数，分别采用网格搜索优化法、遗传算法（GA）、粒子群算法（PSO）并结合五折交叉验证法进行建模，并以五折交叉验证准确率为指标，对 SVM 模型参数组合（c，g）进行寻优，用寻优所确定的最佳参数建立 PCA-SVM 模型，并用所建模型对校正集和测试集样品进行预测，计算预测准确率。具体分集信息见表 1。

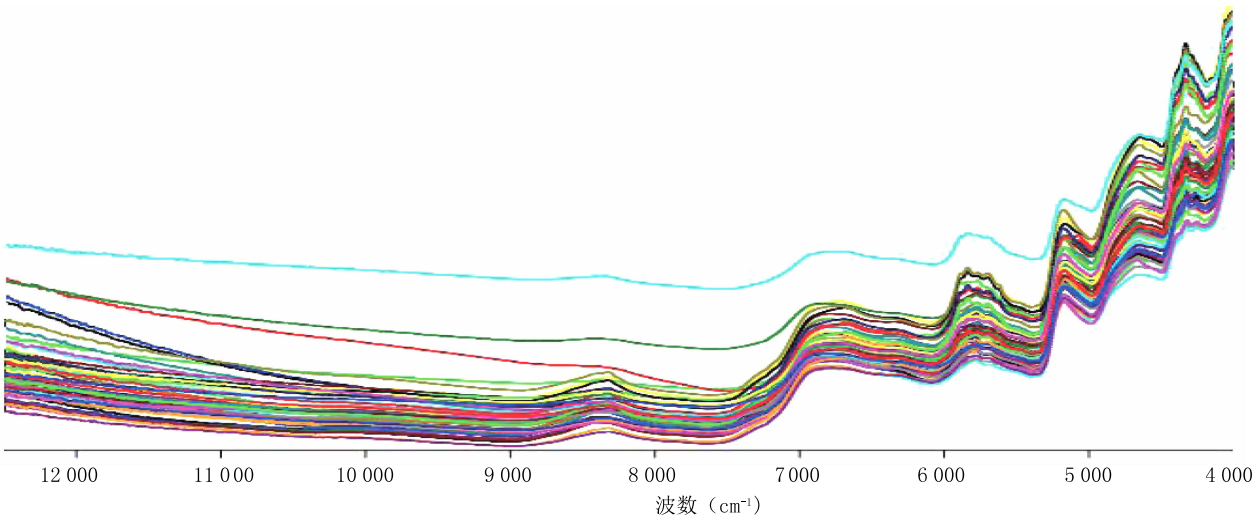

图 1　56 份啤酒花样品的近红外原始光谱叠加图

2.3　PCA 降维及光谱预处理

2.3.1　PCA 降维

PCA 是一种将原来多个具有一定相关性的众多指标，重新组合成一组新的互相无关的综合指标的统计分析方法[11]。运用 PCA 方法对近红外光谱数据特征提取和压缩，对多维数据进行降维，去除输入随机向量之间的相关性，突出原始数据中的隐含特性，可消除众多信息共存中相互重叠的信息部分。选择特征值较大的几个主成分作为特征变量进行模式识别[12]。

在 OPUS 软件中，将校正集样本初始建模谱段的原始光谱数据，进行 PCA 降维处理，提取前两个主成分，利用各样品的第一主成分（PC1）和第二主成分（PC2）得分值绘制平面散点图（图 2）。

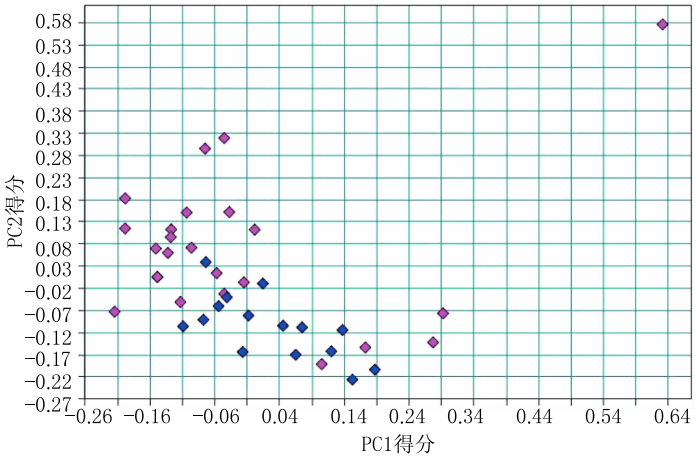

图 2　第一主成分（PC1）和第二主成分（PC2）得分平面散点图

进口：蓝色；国产：红色

2.3.2　光谱预处理

在用近红外漫反射仪进行光谱信息采集时，得到的样品信息包括除自身信息外的无关信息，例如由于仪器、样品粒径大小、装样量、重复测量次数等引起的基线不平、噪音干扰，为了得到可靠的信息，需要对光谱进行预处理以消除干扰，建立更可靠的模型。采用光谱预处理方法有矢量归一化法（VN）、一阶导数法（FD）、二阶导数法（SD）建立定性模型，并以模型效果确定最佳预处理方法。

为确定最佳的光谱预处理方法，提取有效的主成分。利用 Matlab R2014b 软件，在训练集样品的 9 000～4 100 cm^{-1}谱段，分别对 VN，FD 或 SD 预处理后的光谱进行 PCA 降维，并提取不同预处理条件下的前两个主成分，利用各样品的第一主成分（PC1）和第二主成分（PC2）得分值绘制平面散点图（图 3）。

由图 3 可知，校正样品经 VN、FD、SD 预处理后可见鉴别趋势，但 VN、SD 预处理后部分同类别样品分布较近，易混淆。而校正集样品的光谱经 FD 预处理后，其主成分得分的散点图上，同类样品彼此靠近，异类样品彼此分离，相比于其他预处理方法，其分类效果最佳，故确定 FD 为最佳光谱预处理方法。光谱在 9 000～4 100cm^{-1}谱段的光谱经 FD 预处理后，消除了基线漂移，同时光谱峰差异得到显著放大，更有利于进行品种鉴别。

2.4　特征谱段筛选

在 9 000～4 100 cm^{-1}谱段中还包括水的特征吸收 7 500～6 500 cm^{-1}，5 400～5 000 cm^{-1}。此外还有尚不明确的干扰信息，因此，为简化模型，消除干扰，提高模型稳定性，对建模谱段进行筛选。在不降低模型鉴别能力的情况下，尽可能缩小建模谱段的范围。

由图 4，在排除水分的干扰后，初始建模谱段可被分为 3 部分：9 000～7 500 cm^{-1}，6 500～5 400 cm^{-1}，5 000～4 100 cm^{-1}，故将上述 SD 预处理后的三个谱段，分别进行 PCA 降维，提取前两个主成分，绘制主成分得分散点图（图 5）。由图 5 可见6 500～5 400 cm^{-1}谱段效果最佳，该谱段条件佳，PCA 得分散点图上，同类样品点相对集中，异类样品点能较好分离。但 3 批样品出现类别混乱。尚需进一步的优化。

因此，选取 6 500～5 400 cm^{-1}谱段，以 FD 为最佳预处理方法进行光谱预处理，以 PCA 提取主成分，获得各样品光谱的主成分得分，作为 SVM 模型的输入变量，建立国产和进口啤酒花的近红外光谱 PCA-SVM 定性分析模型。

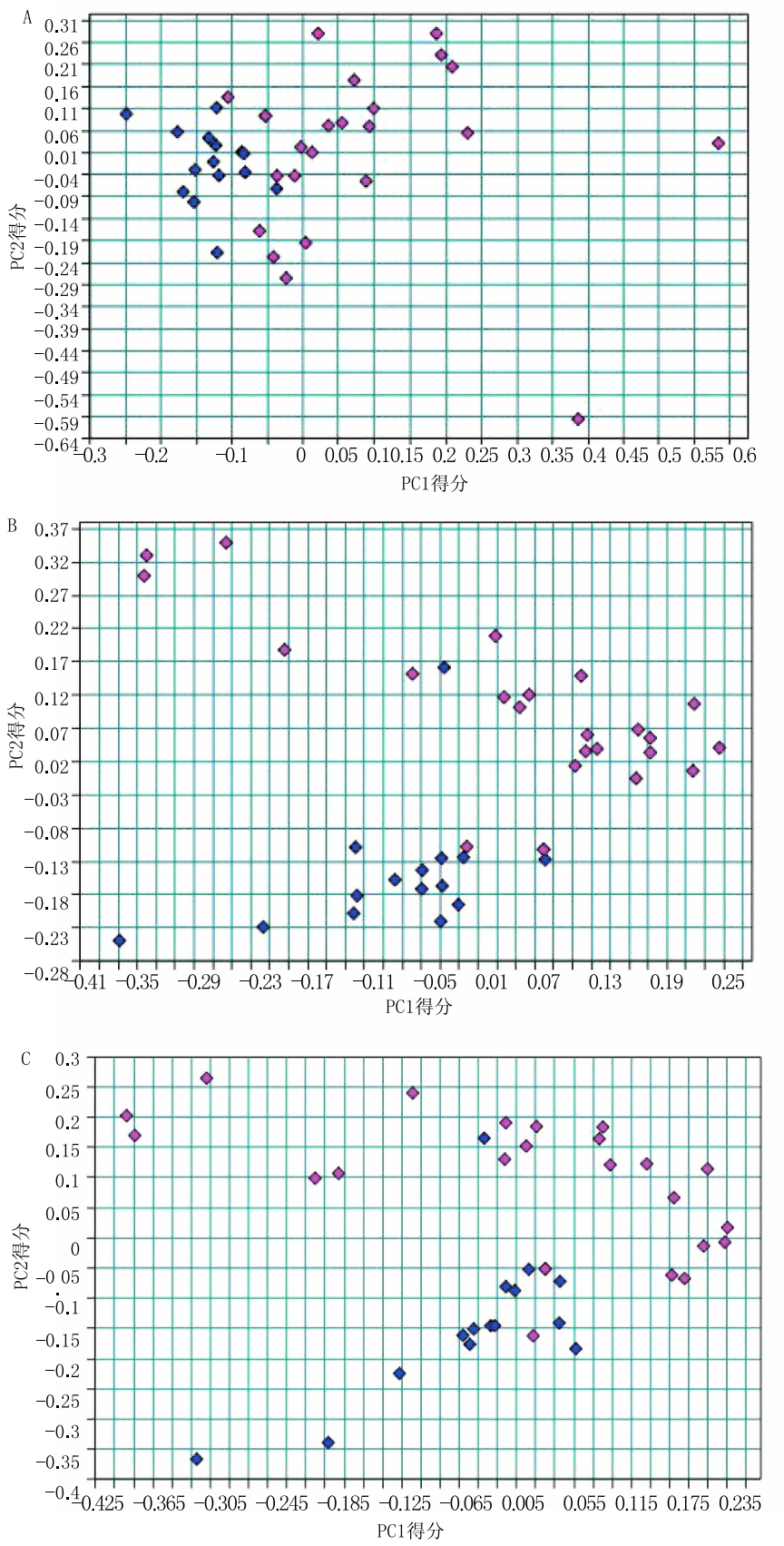

图 3　光谱预处理 VN、FD、SD 散点图

A. 矢量归一化法散点图；B. 一阶导数法散点图；C. 二阶导数法散点图

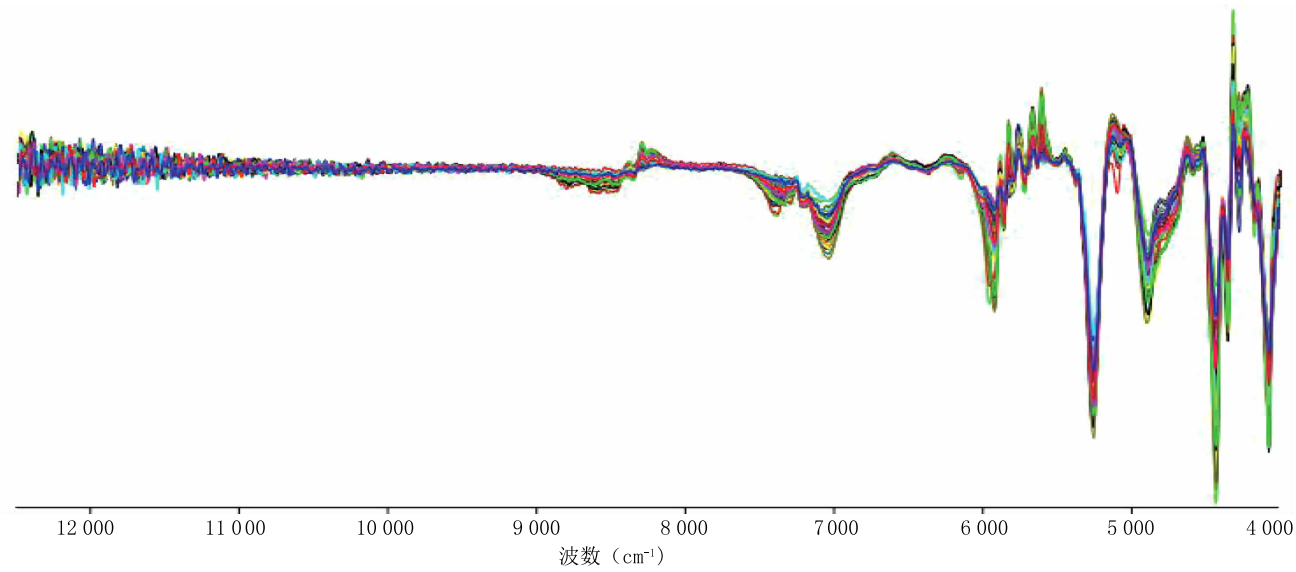

图 4　啤酒花样品的一阶导数 NIRS

2.5　SVM 建模

2.5.1　SVM 算法

SVM 算法[13]是一种基于统计学理论的新的机器学习方法，SVM 通过寻求结构化风险最小来提高学习机泛化能力，实现经验风险和置信范围的最小化，达到在统计样本量较少的情况下，获得良好统计规律的目的。在解决小样本、非线性、高维数据时具有很大优势，在很大程度上能够克服“过学习”和“维数灾难”等问题。SVM 常用核函数有多项式、Sigmoid 感知核和高斯径向基核（RBF）。其中，RBF 核函数[14]是应用最广泛的核函数，适用于低纬、高维、小样本或大样本等情况，是较为理想的分类依据函数。

2.5.2　内部参数优化

在 RBF 为核函数的 SVM 算法中有 2 个重要的参数：惩罚因子 c 和核函数参数 g，不同参数所建立的模型的预测能力不同，故参数优化的方法在建模过程中有着很大的影响。网格搜索法（GS）是 SVM 问题上应用最为普遍的参数寻优算法，它是将参数（c，g）在一定的空间范围中划分成网格，从网格中全部的点中找到最优参数[15]。此外，遗传算法（GA）和粒子群算法（PSO）是近年来迅猛发展起来的智能算法，GA 算法是借鉴生物界自然选择和遗传机制，利用选择、交换和突变等算法的操作，随着不断的遗传迭代，保留目标数据较优的变量，最终达到最优结果的一种方法[16]。PSO 算法模拟鸟群飞行觅食的行为，通过鸟之间的集体协作使群体达到最优目的。在 PSO 算法系统中，每个备选解被称为一个粒子，多个粒子共存、合作寻优，每个粒子根据其自身的经验和相邻粒子群的最佳经验在问题空间中向更好的位置飞行，搜索最优解[17]。

基于上述原理，本文使用了 RBF 核函数建立国产和进口啤酒花的 SVM 模式分类模型。模型以 FD 预处理的样品光谱（6500－5400cm^{-1}）经 PCA 提取的前两个主成分得分为 SVM 输入变量，以各类样品的类别标签值为输出，分别采用网格搜索优化法、GA、PSO 并结合五折交叉验证法，以五折交叉验证准确率为指标，对 SVM 模型参数组合（c，g）进行寻优，用寻优所确定的最佳参数建立 PCA-SVM 模型，并用所建模型对校正集和测试集样品进行预测，计算预测准确率。综合考虑五折交叉验证准确率、校正集预测准确率和测试集预测准确率，对 PCA-SVM 模型进行评价。不同寻优方法的寻优过程见图 6 所示，所得参数建立的 SVM 模型效果见表 2。

综合对比上述预测准确率，准确率越高，模型越好，据此判断最佳 SVM 模型。由表 2，不同寻优方法下，两个主成分所建 PCA-SVM 模型的效果一致。但 PSO 寻优所确定的 c 值偏大而 g 值偏小，该选优方法不合适。对比网格搜索法和 GA 寻优过程，发现 GA 寻优操作更为复杂，且具有一定随机性，故确定网格搜索法为最佳。依据网格寻优结果，确定最佳 c 和 g 值均为 1，利用前 2 个主成分得分值建立的 PCA-SVM 模型对校正集和测试集预测准确率均较高（>85%）。但考察主成分累计贡献率变化（图 7），前两个主成分累计贡献率为 91.70%（<95%），2 个主成分包含原光谱数据的信息量有限，是否为最佳建模主成分数尚不明确，故对主成分数

进行优选，进一步提高 PCA-SVM 模型性能。

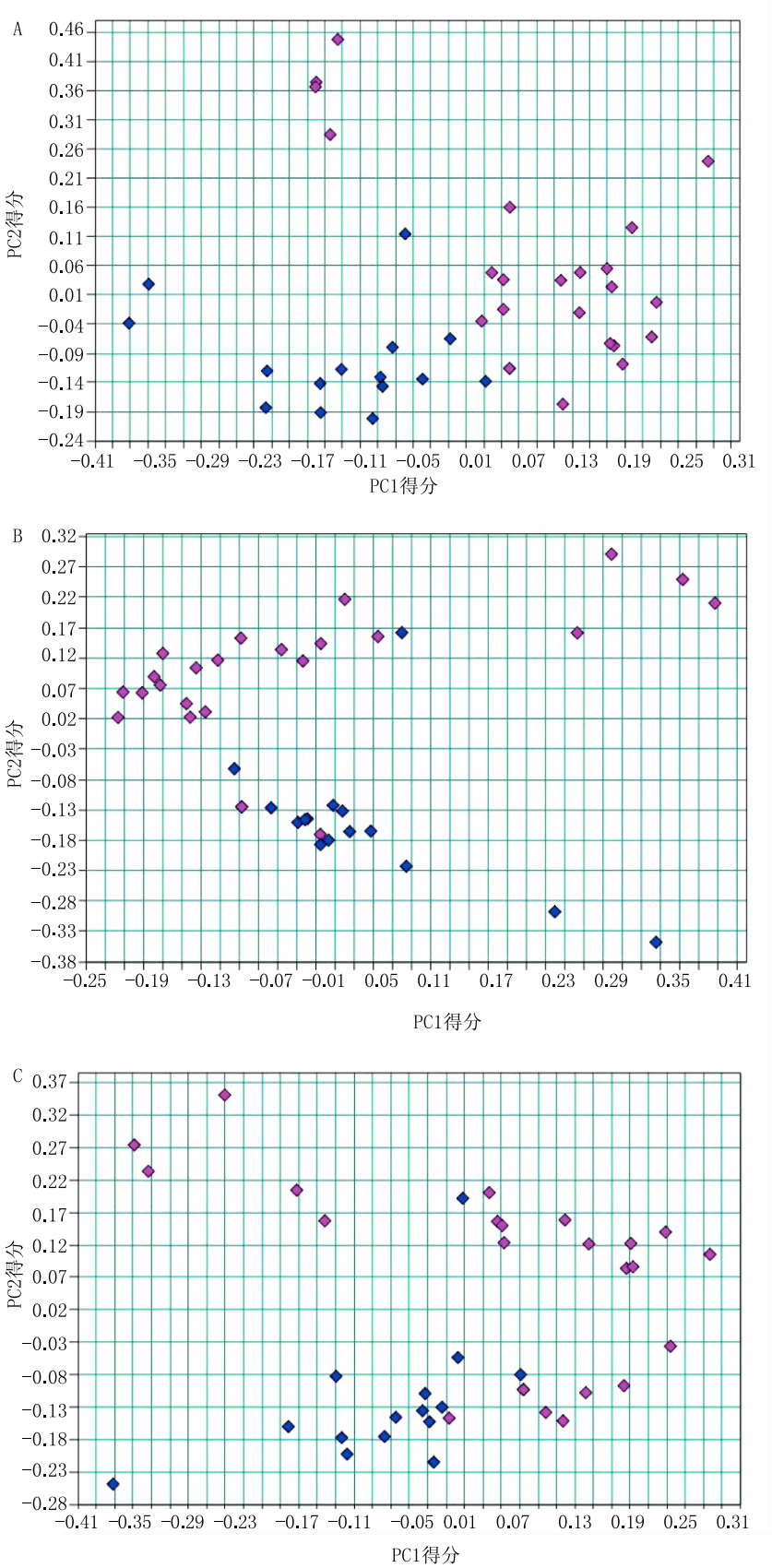

图 5　不同建模波段的一阶导数 PCA 得分散点图

A. 9 000～7 500 cm^{-1}；B. 6 500～5 400 cm^{-1}；C. 5 000～4 100 cm^{-1}

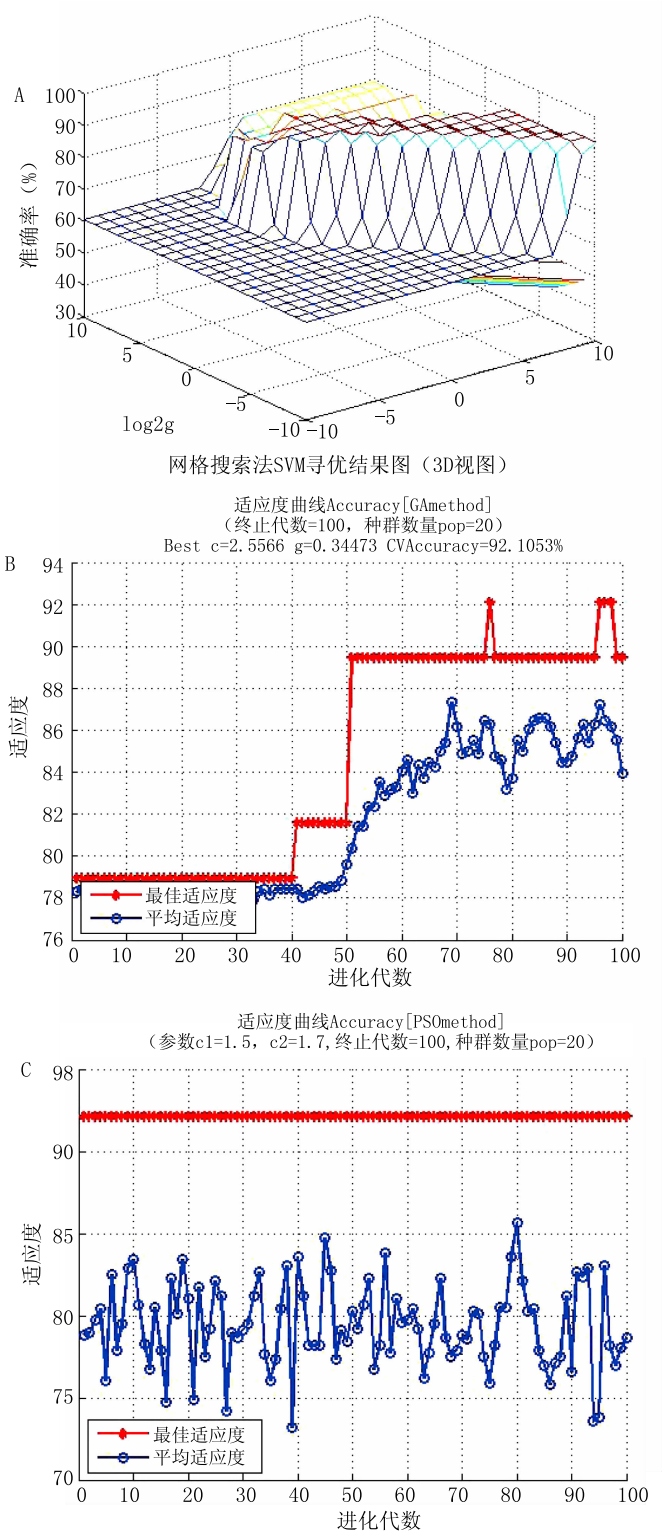

图 6　SVM 内部参数寻优过程图

A. 网格搜索优化法；B. 遗传算法；C. 粒子群优化算法

表 2　相同主成分数的 PCA-SVM 模型的寻优过程建模参数及验证、评价效果

寻优算法	主成分数	c	g	准确率/%		
				五折交叉验证	校正集	测试集
网格	2	1	1	92.11	92.11(35/38)	88.89(16/18)
GA	2	2.5566	0.3447	92.11	92.11(35/38)	88.89(16/18)
PSO	2	778.6877	0.001	92.11	92.11(35/38)	88.89(16/18)

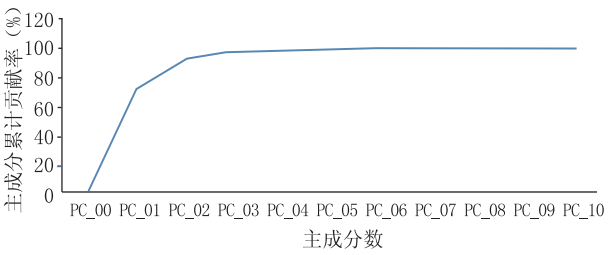

图 7　主成分累计贡献变化图

2.5.3　主成分数进一步优选

经原始光谱（9 000～4 100 cm^{-1}）FD 预处理、PCA 降维得到前 2 个主成分，并提取特征谱段数据（6 500～5 400 cm^{-1}）进行 SVM 建模，前 2 个是否为必要或最佳主成分尚不明确，且对原始光谱数据信息的代表性不强，故需对主成分数进行优选。在前 2 个主成分的基础上，增加建模的主成分个数，将 PCA 提取前 3、4、5、6、7、8、9、10 个主成分，以防数据丢失。但主成分数增加会使模型稳定性减低，故需对建模的主成分数进行筛选。根据上述 SVM 算法建模及寻优过程依次建立 8 个 PCA-SVM 分类模型，由图 7 可得，前 10 个主成分累计贡献率达 99.74%，前 10 个主成分的贡献率相对较大，对原数据的代表性较强，故本研究在这前 10 个主成分中进行筛选，依次建立不同主成分数的 PCA-SVM 模型，并对比建模效果见表 3。

表 3　不同主成分数的 PCA-SVM 模型的建模参数及验证、评价效果

主成分	主成分数	寻优算法	c	g	准确率（%）		
					五折交叉验证	校正集	测试集
前 3 个	3	网格	0.5	2	92.11	92.11(35/38)	88.89(16/18)
前 4 个	4	网格	1	1	92.11	92.11(35/38)	88.89(16/18)
前 5 个	5	网格	1	1	92.11	92.11(35/38)	88.89(16/18)
前 6 个	6	网格	1	1	92.11	92.11(35/38)	88.89(16/18)
前 7 个	7	网格	32	0.125	94.74	92.11(35/38)	94.44(17/18)
前 8 个	8	网格	2	1	97.37	97.37(37/38)	94.44(17/18)
前 9 个	9	网格	4	0.5	94.74	97.37(37/38)	94.44(17/18)
前 10 个	10	网格	4	0.5	94.74	97.37(37/38)	94.44(17/18)

由表 2 和表 3 可知：随主成分数的增加，校正集和测试集预测准确率均增加，其中当主成分数为 8 时，校正集预测准确率达到最大，其后保持稳定；当主成分数为 7 时，测试集预测准确率达到最大，并保持稳定。此外，五折交叉验证准确率先增大，后减小，当主成分数为 8 时，五折交叉验证准确率最大。故确定啤酒花样品最佳主成分数为 8。即以 PCA 提取的前 8 个主成分得分为 PCA-SVM 的输入变量。

2.6　SVM 评价

综上所述，在 6 500～54 00 cm^{-1} 建模谱段，确定最佳光谱预处理方法为一阶导数法（FD），FD 预处理光谱 PCA 降维后，确定最佳主成分前 8 个主成分（PC1，PC2，...，PC8）。经网格搜索法确定最佳 SVM 建模参数组为：c=2，g=1，所建 PCA-SVM 模型对校正集和测试集样品预测正确率均分别为 97.37%和 97.44%，预测准确率高。模型五折交叉验证准确率亦达 97.37%，模型性能最佳。该模型对校正集和测试集预测结果见图 8。

通过图 8 可以看出，在寻优过程中发现，网格搜索算法寻优原理简单，具有可重复性，而 PSO 算法和 GA 寻优的智能算法，其运算过程具有一定的随机性，对复杂问题的解决能力更强。建立光谱数据经 FD 预处理后进行降维，使得建模复杂度相对简化，且不同样品的趋向度不一致，故确定最优内部寻优方法为网格搜索算法。以网格搜索算法确定的模型 8 对训练集和测试集样品的预测可知，校正集的第 9 个样品 PJH-14 预测错误，测试集的第 17 个样品 PJH-51 预测错误，可能样品的基数过小，可达标原始光谱的绝大多数信息。

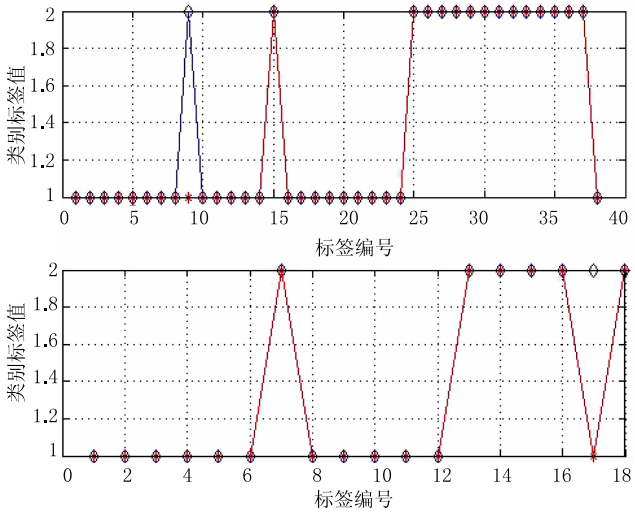

图 8　模型 8 对样品的预测效果

A. 预测集；B. 校正集

3　讨论

本实验通过收集国产和进口啤酒花样品，在收集的 56 个样品中，有 21 个进口啤酒花，35 个国产啤酒花，利用 NIRS 光谱，建立了啤酒花中药材的 PCA-SVM 模式识别模型，该模型对预测集和校正集样品的预测准确率高。模型五折交叉验证准确率亦达 97.37%，模型性能最佳。可用于啤酒花样品的快速鉴别。建模过程中，本文采用光谱 PCA 降维所得的主成分得分平面散点图，对光谱预处理主成分进行优选，根据样品的趋势，确定样品的最佳预处理方法。然后对 SVM 的内部参数进行 GA 算法、PSO 算法寻优，建立 PCA-SVM 算法，快速鉴别啤酒花样品。

本文首次将 NIRS 技术应用于啤酒花中药材的鉴别，证明了其具有可能性，为啤酒花的鉴别提供了新的方法。但由于样本量及产地的限制，本文在对啤酒花样品进行模型五折交叉验证中校正集第 9 个样品 PJH-14 预测错误，测试集第 17 个样品 PJH-51 预测错误。后期需对样品量和产地、品种进行扩增，对模型进行完善。使该方法可以快速鉴别啤酒花，提高测试正确率和准确率。该方法也可用于其他中药材的鉴别，例如矿物类和树脂类中药材[18-19]。本研究方法较新，收集样品量大，建立方法具有一定的应用和推广价值。

【参考文献】

[1] AGHAMIRI V, MIRGHAFOURVAND M, MOHAMMAD-ALIZADEH-CHARANDABI S, et al. The effect of Hop (*Humulus lupulus* L.) on early menopausal symptoms and hot flashes: A randomized placebo-controlled trial[J]. Complement Ther Clin Pract, 2016, 23: 130-135.

[2] ZANOLI P, ZAVATTI M. Pharmacognostic and pharmacological profile of *Humulus lupulus* L. [J]. Journal of Ethnopharmacology, 2008, 116(3): 383-396.

[3] HOYLES R K, ELLIS R W, WELLSBURY J, et al. A multicenter, prospective, randomized, double-blind, placebo-controlled trial of corticosteroids and intravenous cyclophosphamide followed by oral azathioprine for the treatment of pulmonary fibrosis in scleroderma[J]. Arthritis Rheum, 2006, 54(12): 3962-3970.

[4] 王春阳，罗正东，赵丽．中药啤酒花的生药鉴定[J]. 中医药信息，1997(3)：20.

[5] 张娟．非线性化学指纹图谱在啤酒和啤酒花鉴别及定量分析中应用[D]. 长沙：中南大学，2014.

[6] 郭沙沙，王志沛，骆学雷，等．非线性化学指纹图谱在啤酒花鉴别评价中的应用[J]. 酿酒科技，2013(9)：71-74.

[7] 丁念亚，黎薇，冯昕韡，等．近红外漫反射光谱在中药分类及真伪鉴别中的应用[J]. 计算机与应用化学，2008，25(4)：499-502.

[8] CHO C H, WOO Y A, KIM H J, et al. Rapid qualitative and quantitative evaluation of deer antler (Cervuselaphus) using near-infrared reflectance spectroscopy [J]. Microchemical Journal, 2001, 68(2): 189-195.

[9] 陈龙，张晓冬，孙杨波，等．基于近红外漫反射光谱和 PCA-SVM 算法快速鉴别炉甘石[J]. 中国实验方剂学杂志，2019(4)：25-27

[10] 尼珍，胡昌勤，冯芳．近红外光谱分析中光谱预处理方法的作用及其发展[J]. 药物分析杂志，2008，28(5)：824-829.

[11] DUBUISSON-JOLLY M P, GUPTA A. Color and texture fusion: application to aerial image segmentation and GIS updating[J]. Image and Vision Computing, 2000, 18(10): 823-832.

[12] KWITT R, MEERWALD P, UHL A. Lightweight detection of additive watermarking in the DWT-domain [J]. IEEE Trans Image Process, 2011, 20(2): 474-484.

[13] VAPNIK V N. The nature of statistical learning theory[M]. New York : Springer, 1999: 988-999.

[14] 李盼池，许少华．支持向量机在模式识别中的核函数特性分析[J]. 计算机工程与设计，2005，26(2)：302-304.

[15] 王健峰，张磊，陈国兴，等．基于改进的网格搜索法的 SVM 参数优化[J]. 应用科技，2012，39(3)：28-31.

[16] 杨旭，纪玉波，田雪．基于遗传算法的 SVM 参数选取[J]. 辽宁石油化工大学学报，2004，24(1)：54-58.

[17] 姚全珠，蔡婕．基于 PSO 的 LS-SVM 特征选择与参数优化算法[J]. 计算机工程与应用，2010，46(1)：134-136，229.

[18] 张晓冬，陈龙，黄必胜，等．基于 NIRS 和 PCA-SVM 算法快速鉴别 4 种含铁矿物药[J]. 中成药，2018，40(2)：404-410.

[收稿日期] 2019-01-09　**[修回日期]** 2019-05-14
[本文编辑] 陈盛新