

近红外光谱结合化学计量学的常见中国蜂蜜掺杂糖浆鉴别

黄富荣¹, 宋 晗¹, 郭 灏¹, 杨心浩¹, 李立群², 赵红霞^{2*}, 杨懋勋^{3*}

1. 暨南大学光电工程系, 广东 广州 510632
2. 广东省生物资源应用研究所, 广东 广州 510636
3. 珠海大横琴科技发展有限公司, 广东 珠海 519000

摘 要 目前我国蜂蜜市场掺假现象严重, 研究一种快速、准确的方法用于市场流通领域掺假蜂蜜的鉴别具有重要的现实意义。采用近红外光谱(NIR)结合化学计量学方法对常见的天然蜂蜜以及掺假(掺杂常见糖浆)蜂蜜进行建模识别, 并比较偏最小二乘-判别分析(PLS-DA)及支持向量机(SVM)对糖浆掺假蜂蜜鉴别模型的影响。首先, 采集来自中国 10 个省份、20 种常见蜂蜜的 112 个天然纯蜂蜜样品, 以及 6 种常见糖浆样品按不同糖浆含量(10%, 20%, 30%, 40%, 50%, 60%)配制的 112 个掺假蜂蜜样品, 共计 224 个样品; 通过近红外光谱扫描获得所有样品的近红外光谱数据(波长范围 400~2 500 nm); 然后, 分别采用一阶导数(FD)、二阶导数(SD)、多元散射校正(MSC)、标准正态变化(SNVT)四种方式对原始光谱进行预处理; 再结合 PLS-DA 和 SVM 建立天然蜂蜜和糖浆掺假蜂蜜的鉴别模型, 比较不同预处理方法对两种不同建模算法建立的蜂蜜掺假鉴别模型效果。其中 SVM 算法的惩罚参数 c 和核函数参数 g 通过网格搜索法(GS)、遗传算法(GA)、粒子群算法(PSO)三种寻优算法进行优化。分析结果表明: 光谱数据进行预处理后所建立的模型准确率均有明显提升, 而对于 SVM 模型, 惩罚参数 c 和核函数参数 g 对模型准确率的提升效果要比光谱预处理带来的提升效果更明显。在 PLS-DA 算法中, 经 FD 光谱预处理后建立的模型效果最好, 最佳 PLS-DA 模型准确率为 87.50%; 在 SVM 算法中, 经 MSC 预处理后, 再通过 GS 寻优, 获得惩罚参数 c 为 3.0314, 核函数参数 g 为 0.3298 的条件下所建立的模型效果最好, 最佳 SVM 模型准确率为 94.64%。由此可见, 非线性的 SVM 算法结合 NIR 光谱数据所建立的天然蜂蜜与糖浆掺假蜂蜜鉴别模型要优于线性的 PLS-DA 模型, 同时表明 NIR 光谱结合化学计量学方法对常见糖浆掺杂的中国蜂蜜鉴别是可行的。

关键词 蜂蜜掺假; 近红外光谱; PLS-DA; SVM; 糖浆; 鉴别

中图分类号: TS207.3 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)11-3560-06

引 言

蜂蜜是一种富有营养的天然食品, 深受广大消费者的青睐。我国自 20 世纪 90 年代以来一直是世界第一养蜂大国、蜂蜜第一出口大国。然而, 近年由于我国蜂蜜掺假的现象日益猖獗, 严重影响我国蜂蜜声誉及市场竞争力, 出口量大幅减少。为确保蜂蜜的质量, 加强蜂蜜产品的检测显得尤为重要。目前市场上最常见的蜂蜜掺假方式是糖浆掺假, 在蜂蜜中掺入果葡糖浆、玉米糖浆等食用糖浆。现有的糖浆掺假蜂蜜鉴别方法包括: 酶活性法、色谱法、差热分析法、稳定性

碳同位素比值分析法^[1-4]等。其中稳定性碳同位素比值分析法是多年来检测蜂蜜糖浆掺假的标准分析技术, 但该方法只能检测掺入 C4 糖浆(玉米糖浆、蔗糖糖浆、果葡糖浆、麦芽糖浆等)的掺假蜂蜜, 对于掺入 C3 糖浆(大米糖浆、甜菜糖浆等)的掺假蜂蜜无法有效鉴别^[5]。同时上述这些检测方法均存在耗时长、有破坏性、仪器昂贵、操作复杂、需要专业技术人员等缺点, 无法实现现场快速检测。因而, 寻找一种快速、准确的掺假蜂蜜鉴别方法具有重要应用价值。

基于近红外光谱(NIR)检测方法具有快速、无损、绿色环保以及操作简单等优点, 近年国内外已开展大量的基于近红外光谱结合化学计量方法鉴别蜂蜜的研究^[6-8]。2011 年,

收稿日期: 2019-07-28, 修订日期: 2019-09-25

基金项目: 国家自然科学基金项目(61975069), 广东省自然科学基金项目(2018A0303131000), 广州市科学技术攻关项目(201604040007, 201604020168)资助

作者简介: 黄富荣, 1979 年生, 暨南大学理工学院光电工程系副研究员 e-mail: furong_huang@163.com

* 通讯联系人 e-mail: hxzh110@126.com; yangmaoxun1980@163.com

Chen 等^[9]利用 NIR 光谱对高果糖玉米糖浆掺杂的中国蜂蜜进行鉴别分析。2015 年, 张妍楠等^[10]将 NIR 光谱与化学计量学结合, 快速、高效的识别出掺入大米糖浆的洋槐蜂蜜。2016 年, Bazar 等^[11]利用光纤浸没探针记录了 4 种不同地区刺槐蜂蜜和高果糖玉米糖浆掺假蜂蜜的 NIR 光谱, 结合水光谱组学发现天然蜂蜜比工业糖浆含有更多高组织度的水分子结构。

以上研究均表明了利用 NIR 分析鉴别掺杂糖浆蜂蜜是可行的, 然而, 目前 NIR 光谱对蜂蜜掺假研究仍然处于方法学研究阶段, 采用的样品过于单一, 多是单一蜜种的蜂蜜掺入单一糖浆, 或者单一蜜种蜂蜜掺入多种糖浆和多蜜种蜂蜜掺入单一糖浆, 然而蜂蜜市场的掺假蜂蜜多种多样, 存在着多种糖浆与多蜜种混掺的情况, 目前建立的蜂蜜鉴别模型的泛化能力较差, 尚未能在蜂蜜市场中得到推广应用。

为此, 收集了来自中国的常见天然蜂蜜样品, 以及常见蜂蜜掺假的糖浆样品, 配制成较为齐全的天然蜂蜜和糖浆掺假蜂蜜样品, 采集样品的近红外光谱, 结合化学计量学方法, 以期建立一个具有较强的普适性蜂蜜掺假鉴别模型, 用于市场流通环节的蜂蜜鉴别。

1 实验部分

1.1 样品

1.1.1 蜂蜜及糖浆样品

112 个天然蜂蜜样品由广东省生物资源应用研究所提供, 均取自各地的养蜂场以确保未加入任何物质。为了使掺假蜂蜜鉴别具有普适性, 在蜂蜜选取方面, 选择 20 种不同蜜种的蜂蜜, 包括有乌柏蜜、冬蜜、半枫荷蜜、山杏蜜、枸杞蜜、桔梗蜜、枇杷蜜、夏枯草蜜、北芪蜜、五味子蜜、火炭母蜜、益母草蜜、银条蜜、洋槐蜜、椴树蜜、桂花蜜、龙眼蜜、荔枝蜜、荆条蜜、红花蜜。产地源自全国 10 个省份, 包括有广东、广西、湖南、湖北、云南、甘肃、黑龙江、福建、安徽、江苏, 蜂蜜样品的具体信息见表 1。糖浆样品购自当地超市, 包括 2 种 C3 植物糖浆(甜菜糖浆和玉米糖浆)以及 4 种 C4 植物糖浆(果葡糖浆、玉米糖浆、麦芽糖浆和蔗糖糖浆), 共 6 种常见的蜂蜜掺假糖浆种类。

1.1.2 糖浆掺假蜂蜜样品

配制不同掺假度[糖浆和纯蜂蜜的质量比($W/W\%$)]掺假蜂蜜样品。其中: 单种糖浆掺假样品 60 个: 设定 6 个浓度梯度, 分别为 10%, 20%, 30%, 40%, 50%, 60%。每种糖浆配一个浓度梯度, 每个浓度配制 10 个不同的糖浆掺假样品, 共配制 6×10 共 60 个单种糖浆掺假样品。混掺蜂蜜样品 52 个: 52 个掺假度均为 60% 的糖浆混掺蜂蜜样品。其中包括 2 种不同糖浆混掺样品 15 个, 3 种不同糖浆混掺样品 20 个, 4 种不同糖浆混掺样品 10 个, 5 种不同糖浆混掺样品 6 个及所有 6 种糖浆的混掺样品 1 个。

考虑到部分蜂蜜在低于 25 °C 时所含的葡萄糖容易结晶而使得内部成分分布不均匀, 从而导致采集的光谱信噪比变低。因此所有的蜂蜜样品及糖浆掺假样品均先在 50 °C 下水浴 1 h, 以充分去除结晶物或结晶核, 再在 30 °C 下水浴 48 h

以确保完全去除气泡, 处理完毕后所有样品在试管上填好标签后, 放入试管架中在室温下静置待用。

表 1 蜂蜜样品种类、产地及数量

Table 1 Types, origin and numbers of honey samples

蜜种	产地	样品数	蜜种	产地	样品数
乌柏	广东	34	火炭母	广西	4
冬蜜	广东	26	益母草	云南	4
半枫荷	湖南	4	银条	云南	2
山杏	广西	2	洋槐	甘肃	2
枸杞	湖北	2	椴树	黑龙江	2
桔梗	广东	2	桂花	广西	2
枇杷	江苏	2	龙眼	福建	2
夏枯草	安徽	4	荔枝	广东/湖北	8
北芪	甘肃	4	荆条	湖北	2
五味子	湖北	2	红花蜜	安徽	2

1.2 光谱采集

采用丹麦福斯公司 XDS Rapid Content 光栅型近红外光谱分析仪及其透射附件, 光谱采集范围为 400~2 500 nm, 探测器为 Si(400~1 100 nm)、PbS(1 100~2 500 nm)。

采集掺杂天然蜂蜜样品和糖浆掺杂的蜂蜜样品的光谱, 每隔 2 nm 采样一次, 获得 400~2 500 nm 范围的光谱。所有样品的光谱数据均采集 3 次取平均, 得到纯天然蜂蜜光谱 112 条, 糖浆掺杂的蜂蜜样品 112 条, 共 224 条光谱。

1.3 PLS-DA

PLS-DA 是最常用的一种线性判别算法。PLS-DA 的基本思想是通过 PLS 分类技术在特征提取过程中获取样本的分类信息, 然后建立自变量和分类变量之间的回归模型, 进而有效地提取出与分类有关的特征变量, 实现数据的分类识别。主要步骤如下: 首先对光谱矩阵 X 和类别向量 Y 进行正交分解, 得到光谱矩阵和类别向量的得分矩阵 T 和 U , 见式(1)和式(2)。然后对得分矩阵 T 和 U 进行线性回归, 求得回归系数 B , 如式(3)和式(4)所示。最后通过式(5)求得未知样品光谱矩阵 X_{test} 的预测类别 Y_{test} 。其中, P 和 Q 分别为光谱矩阵 X 和类别向量 Y 的载荷矩阵; E 和 F 分别为光谱矩阵 X 和类别向量 Y 的 PLS 拟合残差矩阵。

$$X = TP + E \quad (1)$$

$$Y = UQ + F \quad (2)$$

$$U = TB \quad (3)$$

$$B = (T^T T)^{-1} (T^T U) \quad (4)$$

$$Y_{\text{test}} = X_{\text{test}} BQ \quad (5)$$

1.4 SVM

SVM 算法原理如下: 训练数据集由 $\{(x_i, y_i), i = 1, 2, \dots, N\}$ 表示, 其中 $x_i \in R^D$ 是输入向量, $y_i \in \{-1, 1\}$ 是其对应的期望输出。SVM 可以找到在两类数据之间的最佳超平面 ($\omega * x + b = 0$ (ω 表示平面的法向矢量, 而 b 是从平面到原点的距离)。对于线性可分的情况, 在分类后的平面将数据分成两类后, 两类数据之间的差值为 $2 / \|\omega\|$ 。分类器是

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^N a_i y_i (x_i \times x) + b \right\} \quad (6)$$

对于非线性情况, SVM 将低维空间中的数据映射到高维空间。分类器是

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^N a_i y_i K(x_i \times x) + b \right\} \quad (7)$$

其中 $\text{sgn}\{\}$ 是符号函数, a_i 是拉格朗日乘子, x_i 是训练样本, x 是要分类的样本, $K(x_i \times x)$ 是核函数。选择合适的核函数是建立性能良好的 SVM 模型最重要的一步, 通常包括两部分工作: 一是选择合适的核函数类型; 二是确定核函数类型后优化重要的参数。研究表明核函数参数选择高斯核函数(RBF)所建立的模型具有良好的学习能力^[12]。因此本实验使用 RBF 核函数实现 SVM 的建模, 其中 RBF 核函数的两个重要参数为惩罚参数 c 和核函数参数 g , 该两参数对于控制模型的复杂度、逼近误差及模型的测量精度有重要影响, 因此对该两参数进行优化很有必要。常用的参数优化算法有网格搜索法(grid search, GS)、遗传算法(genetic algorithm, GA)和粒子群优化算法(particle swarm optimization, PSO)^[13]。

1.5 样品集划分

采用 Sample Set Partitioning Based on Joint X-Y Distance(SPX Y)法^[14]将 224 个蜂蜜及糖浆掺假蜂蜜样品按照 3:1 划分训练集和预测集。训练集有 168 个蜂蜜及糖浆掺假蜂蜜样品, 纯蜂蜜和掺假蜂蜜样品各 84 个; 预测集有 56 个蜂蜜及糖浆掺假蜂蜜样品, 纯蜂蜜和掺假蜂蜜样品各 28 个, 样品集具体划分如表 2 所示。

表 2 样品集划分
Table 2 Division of sample sets

类别	样品数		总计
	训练集	预测集	
天然蜂蜜样品	84	28	112
掺假蜂蜜样品	84	28	112
总计	168	56	224

1.6 模型评估指标

分类模型性能一般由预测集的准确率(accuracy)、灵敏度(sensitivity)及特异性(specificity)进行评价。准确率、特异性和灵敏度越接近于 1, 表明模型的分类效果越好。分类准

确率是指在分类模型中, 通过预测集来对已建立的模型进行测试, 统计样品被正确判别的数目占样品总数的比值。灵敏度和特异性是临床诊断中很重要的两个指标, 灵敏度是指被正样本被正确分类的百分率, 特异性是指负样本被正确分类的百分率。

$$\text{Accuracy} = \frac{n_{\text{correct}}}{n_{\text{total}}} \quad (8)$$

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (9)$$

$$\text{Specificity} = \frac{FP}{FP + TN} \quad (10)$$

式中, n_{correct} 表示预测集数据被模型正确分类的所有样本个数, n_{total} 表示预测集数据的所有样本个数。TP 表示预测集数据被正确分类的正样本个数, FN 表示预测集数据被错误分类的正样本个数, FP 表示预测集数据被错分的负样本个数, TN 表示预测集数据被模型正确分类的负样本个数。

2 结果与讨论

2.1 光谱分析

天然蜂蜜和糖浆掺假蜂蜜样品的 NIR 原始光谱及平均光谱如图 1 所示。图 1(a)中可以看出天然蜂蜜和糖浆掺假蜂蜜的 NIR 光谱峰形基本一致, 但光谱重叠严重, 无法仅凭肉眼将两者区分开。可以观察到在 1 452, 1 771, 1 939, 2 100, 2 276, 2 326 和 2 437 nm 处有明显的特征吸收峰。其中 1 452 和 1 939 nm 是近红外区域水的两大吸收峰, 1 452 nm 处归属于 O—H 伸缩振动的一级倍频, 1 939 nm 处归属于 O—H 伸缩和弯曲振动的组合频; 1 771 nm 处归属于 C=O 的二倍频吸收, 与蜂蜜糖分有关; 2 100 nm 处归属于 O—H 变形振动和 C—O 伸缩振动; 2 276 nm 处归属于 C—H 伸缩和变形振动的组合频; 2 326 nm 处归属于 CH₂ 的伸缩和变形带^[12]。从图 1(b)的平均光谱图中可以看出在 1 939, 2 100 和 2 276 nm 特征峰处两者的平均光谱强度有所差异, 这些差异说明 NIR 光谱可用于天然蜂蜜和糖浆掺假蜂蜜的鉴别。下面利用 NIR 光谱结合化学计量学算法来建立天然蜂蜜和糖浆掺假蜂蜜的高效判别模型。

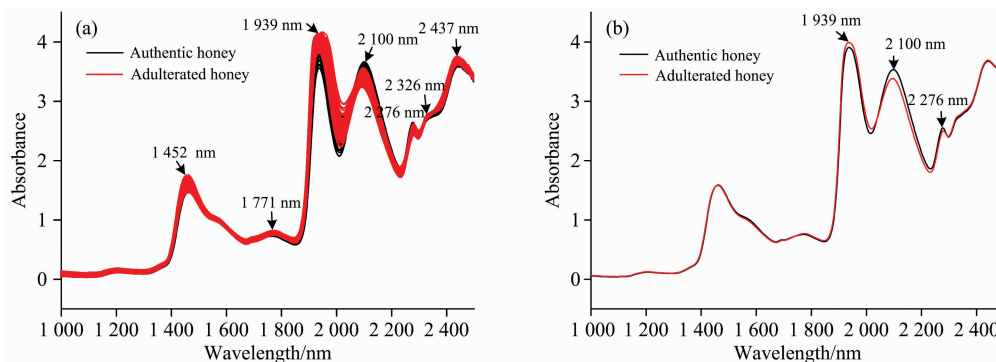


图 1 天然蜂蜜及糖浆掺假蜂蜜样品的近红外光谱

(a): 原始光谱; (b): 平均光谱

Fig. 1 NIR spectra of natural honey and syrup adulterated honey samples

(a): Original spectrum; (b): Average spectrum

2.2 PLS-DA 模型

PLS-DA 算法属于有监督学习, 在建立天然蜂蜜和糖浆掺假蜂蜜的 PLS-DA 判别模型之前, 要给划分好的训练集和预测集人工设定标签。将天然蜂蜜样品赋值为 1, 糖浆掺假蜂蜜样品赋值为 -1。样品赋好类别值后, 再采用 FD, SD, MSC 以及 SNVT 四种方法进行光谱预处理。之后对训练集数据进行 K 折交叉验证(K 取 5)选出最佳的 PLS 因子数, 再

根据最佳因子数将光谱数据和类别值相关联, 建立糖浆掺假蜂蜜的判别模型。最后利用预测集数据对建立的模型进行评估, PLS-DA 模型预测集的分类结果如表 3 所示。原始 NIR 光谱数据经 FD 预处理后鉴别天然蜂蜜和糖浆掺假蜂蜜的结果最好, 因此经 FD 预处理后建立的模型是最佳 PLS-DA 模型, 最佳 PLS 因子数为 10, 准确率为 87.50%, 灵敏度和特异性分别为 96.43% 和 78.57%。

表 3 PLS-DA 模型对糖浆掺假蜂蜜的分类结果

Table 3 Model classification results of syrup adulterated honey by PLS-DA

预处理方法	最佳因子数	灵敏度	特异性	准确率
Original data	5	100%(28/28)	60.71%(17/28)	80.36%(45/56)
FD	10	96.43%(27/28)	78.57%(22/28)	87.50%(49/56)
SD	8	85.71%(24/28)	78.57%(22/28)	82.14%(46/56)
MSC	5	100%(28/28)	60.71%(17/28)	80.36%(45/56)
SNVT	10	100%(28/28)	67.86%(19/28)	83.93%(47/56)

2.3 SVM 模型

分两步建立糖浆掺假蜂蜜的最佳 SVM 判别模型: (1) 在惩罚参数 c 和核函数参数 g 均取默认值的条件下(惩罚参数 c 的默认值为 1, 核函数参数 g 的默认值为 $1/k$, k 是类别数), 选出最佳的预处理方式。(2) 利用寻优算法(GS, GA 和 PSO)进行 c 、 g 参数寻优, 将寻得的最佳 c 、 g 参数值建立 SVM 模型并对预测集进行预测, 选出天然蜂蜜和糖浆掺假蜂蜜的最佳 SVM 判别模型。

首先利用原始 NIR 光谱数据及 4 种不同预处理(FD, SD, MSC, SNVT)后的数据建立 SVM 判别模型, 比较不同预处理方法建立的 SVM 模型对预测数据集的预测准确率的高低来选取最佳的预处理方式。此次建模的惩罚参数 c 和核函数参数 g 均取默认值, 不同预处理下的 SVM 模型分类结果如表 4 所示。四种预处理建立的 SVM 模型准确率相对于原始光谱均有所提高, 但准确率提高的都不显著, 准确率提升最多的预处理方式是 MSC, 相比于原始光谱的预测集准确率提高了 3.57%。

表 4 不同预处理下的 SVM 模型分类结果

Table 4 The classification results of SVM under different pretreatment

预处理方法	寻优参数		训练集准确率	预测集准确率
	c	g		
Original data	Default	Default	89.88%(151/168)	78.57%(44/56)
FD	Default	Default	92.26%(155/168)	80.36%(45/56)
SD	Default	Default	93.45%(157/168)	80.36%(45/56)
MSC	Default	Default	91.67%(154/168)	82.14%(46/56)
SNVT	Default	Default	91.07%(153/168)	80.36%(45/56)

确定最佳预处理方式后, 分别利用 GS, GA 和 PSO 对惩罚参数 c 和核函数参数 g 进行优化, 参数寻优过程及交互验证结果如图 2 所示: 图 2(a) 是 GS 寻优结果的三维显示, 随着交互验证准确率的提高, 由不同 c 、 g 值交织而成的方格颜色由冷色调(深蓝)向暖色调(明黄)变化, 同时方格各顶点所

在水平面相应升高, 当惩罚参数 c 是 3.031 4 和核函数参数 g 是 0.3298 时, 交互验证准确率达到最高为 98.21%; 图 2(b) 是 GS 参数寻优结果的等高线图, 由图 2(a) 投影到二维平面所得。图 2(c) 是 GA 寻优结果图, 由最佳适应度曲线可知, 在迭代次数为 0~50 时, 交互验证准确率不断提高, 当迭代次数为 50 时准确率达到饱和为 98.21%, 寻得的最佳惩罚参数 c 是 8.057 4、核函数参数 g 是 0.292 7。图 2(d) 是 PSO 的寻优结果图, 相比于 GS 和 GA, 其交互验证的初始准确率相对较高(82.74%), 在经过 55 次迭代后交互验证准确率稳定在 97.62%, 寻得的最佳惩罚参数 c 是 38.703 1、核函数参数 g 是 0.189 7。

经 GS, GA, PSO 三种寻优算法优化 c 、 g 参数后的交互验证准确率最低达到了 97.62%, 下一步利用寻得的最佳 c 、 g 参数值分别建立 SVM 模型并用预测集检测模型的好坏以选出最佳 SVM 模型, 预测集结果如表 5 所示: 经过 c 、 g 参数寻优后模型准确率均显著提升, 准确率最高达到了 94.64%, 灵敏度即对天然蜂蜜的识别准确率均达到了 100%, 特异性即对糖浆掺假蜂蜜的识别准确率最佳为 89.29%。相比于默认 c 、 g 值建立的最优 SVM 模型, 经 GS 寻优后的准确率提升了 12.5%; 经 GA 寻优后的准确率提升了 10.72%; 经 PSO 寻优后的准确率提升了 7.15%。经 GS 寻得惩罚参数 $c=3.031 4$, 核函数参数 $g=0.329 8$ 的条件下, 所建的 SVM 模型效果最好, 预测集准确率为 94.64%、灵敏度为 100%、特异性为 89.29%。

3 结论

开展了近红外光谱结合化学计量学方法分析鉴别天然蜂蜜和糖浆掺假蜂蜜。实验结果表明: 基于天然蜂蜜和糖浆掺假蜂蜜的 NIR 光谱数据结合 PLS-DA 和 SVM 算法所建立的蜂蜜鉴别模型均可实现对掺假蜂蜜的鉴别, 但由于蜂蜜样品成分复杂, 以及光谱预处理无法完全消除各种噪声所造成的影响, 从而使得光谱数据与蜂蜜中是否掺入糖浆的线性关系

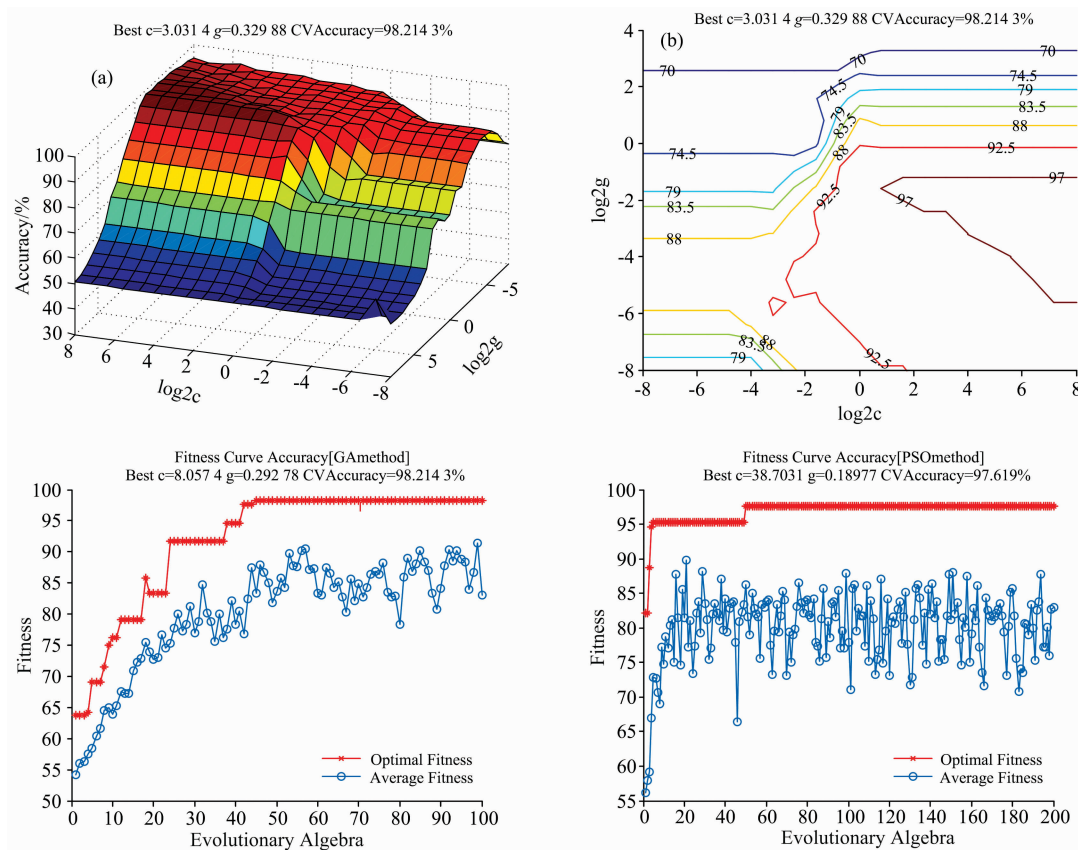


图 2 SVM 模型 c 、 g 参数寻优结果图

(a): GS 寻优结果三维显示; (b): GS 寻优结果等高线图; (c): GA 寻优结果; (d): PSO 寻优结果

Fig. 2 The results of c 、 g parameter optimization

(a): 3D display of GS optimization results; (b): Contour diagram of GS optimization results; (c): GA optimization results; (d): PSO optimization results

表 5 参数寻优后 SVM 模型分类结果

Table 5 Model classification results of SVM by parameter optimization

寻优算法	寻优参数		灵敏度	特异性	准确率
	c	g			
GS	3.031 4	0.329 8	100%(28/28)	89.29%(25/28)	94.64%(53/56)
GA	8.057 4	0.292 7	100%(28/28)	85.71%(24/28)	92.86%(52/56)
PSO	38.703 1	0.189 7	100%(28/28)	78.57%(22/28)	89.29%(50/56)

受到严重干扰, 因此非线性的 SVM 算法用于蜂蜜掺假效果更好。综上所述, 基于 NIR 光谱数据结合 PLS-DA 和 SVM 算法可以实现常见中国蜂蜜糖浆掺假鉴别, 有望为中国蜂蜜

市场流通环节提供一种快速、可靠现场蜂蜜鉴别方法。但本实验中部分蜂蜜样本数少, 实验结果鲁棒性以及泛化能力将有一定影响, 仍需进一步提高模型的性能。

References

[1] Arroyo-Manzanares N, Garcia-Nicolás M, Castell A, et al. Talanta, 2019, 205(1): 120123.
 [2] Puscas A, Hosu A, Cimpoi C, et al. Journal of Chromatography A, 2013, 1272(1): 132.
 [3] Gan Z L, Yang Y, Li J, et al. Journal of Food Engineering, 2016, 178: 151.
 [4] Çinar S B, Eksi A, Coskun I. Food Chemistry, 2014, 157: 10.
 [5] Guler A, Kocaokutgen H, Garipoglu A V, et al. Food Chemistry, 2014, 155: 155.
 [6] Li S, Zhang X, Shan Y, et al. Food Chemistry, 2017, 218: 231.
 [7] Naila A, Flint S H, Sulaiman A Z, et al. Food Control, 2018, 90: 152.

- [8] Guelpa A, Marini F, Plessis A D, et al. Food Control, 2017, 73(8): 1388.
- [9] Chen L Z, Xue X F, Ye Z H, et al. Food Chemistry, 2011, 12(4)8: 1110.
- [10] ZHANG Yan-nan, CHEN Lan-zhen, XUE Xiao-feng, et al(张妍楠, 陈兰珍, 薛晓峰, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2015, 35(9): 2536.
- [11] Bazar G, Romvari R, Szabo A. Food Chemistry, 2016, 194: 873.
- [12] Akay M F, Abut F, Lu M, et al. Neural Computing and Applications, 2016, 27(6): 1785.
- [13] Gao X, Hou J. Neurocomputing, 2015, 174: 906.
- [14] Galvão R K H, Araujo M C U, José G E, et al. Talanta, 2005, 67(4): 736.

Determination of Chinese Honey Adulterated with Syrups by Near Infrared Spectroscopy Combined with Chemometrics

HUANG Fu-rong¹, SONG Han¹, GUO Liu¹, YANG Xin-hao¹, LI Li-qun², ZHAO Hong-xia^{2*}, YANG Mao-xun^{3*}

1. Opto-electronic Department of Jinan University, Guangzhou 510632, China

2. Guangdong Institute of Applied Biological Resources, Guangzhou 510636, China

3. Zhuhai Dahengqin Science and Technology Development Co., Ltd., Zhuhai 519000, China

Abstract To find a fast, accurate, and effective method for the identification of honey adulteration, near-infrared spectroscopy combined with chemometrics was used to analyze natural honey and adulterated honey in this paper. First, 224 samples were collected for the study, including 112 natural pure honey samples from 20 common honeys in China, and 112 adulterated honey sample were prepared with 6 different syrup samples according to different syrup contents(10%, 20%, 30%, 40%, 50%, or 60%). Near infrared spectral data (wavelength range of 400~2 500 nm) of all samples were obtained by near infrared light instrument scanning. Then, first derivative (FD), second derivative (SD), multiple scattering correction (MSC), and standard normal variation (SNVT) pre-processing of the original spectra combined with PLS-DA (linear algorithm) and SVM (non-linear algorithm) modeling, respectively, were adopted to establish a differential model of natural honey and syrup-adulterated honey and compare the effects of different pretreatment methods on the honey adulteration identification model established by the two different modeling algorithms. The penalty parameter c and the kernel function parameter g of the SVM algorithm were optimized by three optimization algorithms: grid search, genetic algorithm, and particle swarm optimization. The analysis results showed that the PLS-DA model established by the FD preprocessing had the best effect, and the accuracy of the best PLS-DA model was 87.50%. After MSC pre-processing, the SVM model with the penalty parameter c of 3.031 4 and the kernel function parameter g of 0.329 8 was the best. The accuracy of the best SVM model was 94.64%. It can be seen that the non-linear SVM algorithm combined with the NIR spectral data natural honey and syrup-adulterated honey identification model is better than the PLS-DA model.

Keywords Honey adulterated; Near-infrared spectroscopy; PLS-DA; SVM; Syrups; Determination

(Received Jul. 28, 2019; accepted Sep. 25, 2019)

* Corresponding authors