

文章编号: 1001-9014(2010)03-0216-04

用最小二乘支持向量机的可见-近红外光谱 测定蜂花粉贮存时间

金航峰¹, 黄凌霞¹, 吴迪², 金佩华¹, 楼程富¹

(1. 浙江大学 动物科学学院, 浙江 杭州 310029;
2. 浙江大学 生物系统工程与食品科学学院, 浙江 杭州 310029)

摘要:为了探索一种快速有效的蜂花粉新鲜程度检测方法,利用可见-近红外光谱技术结合最小二乘支持向量机(LS-SVM)对蜂花粉的贮存时间进行了检测.选择常温环境中贮存时间为 4~50 天(共计 47 天)的茶花蜂花粉作为研究对象,将全光谱数据作为输入变量建立了 LS-SVM 模型.结果显示,该 LS-SVM 模型预测效果较好,预测相关系数 r_p 达到了 0.996,预测标准误差(SEP)和预测均方根误差(RMSEP)的值分别为 1.310 和 1.308,优于偏最小二乘法(PLS)和主成分回归(PCR)的预测结果,说明基于 LS-SVM 的可见-近红外光谱技术能够很好地对花粉贮存时间进行检测.同时对不同贮存时间段花粉的预测效果进行了比较,发现该 LS-SVM 模型适用于对第 11~50 天范围的贮存时间进行检测.

关键词:可见-近红外光谱;贮存时间;最小二乘支持向量机

中图分类号:0657.3;S896.8 文献标识码:A

STORAGE PERIOD DETERMINATION OF BEE POLLEN BY VISIBLE-NEAR INFRARED SPECTROSCOPY WITH LEAST SQUARES-SUPPORT VECTOR MACHINES

JIN Hang-Feng¹, HUANG Ling-Xia¹, WU Di², JIN Pei-Hua¹, LOU Cheng-Fu¹

(1. College of Animal Sciences, Zhejiang University, Hangzhou 310029, China;
2. College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China)

Abstract: In order to investigate a fast and efficient method determining the freshness of bee pollen, visible and near-infrared (Vis-NIR) reflectance spectroscopy with least squares-support vector machines (LS-SVM) was applied to determine storage period of bee pollen. The Camellia bee pollens stored for 4~50(47) days at room temperature were investigated. Spectra were collected by an ASD Fieldspec spectrometer as the input variables to build the LS-SVM model. Results show that the prediction performance of LS-SVM model is better than partial least square (PLS) and principal component regression (PCR). Its correlation coefficient of prediction set (r_p) is 0.996, standard error of prediction (SEP) is 1.310, and root mean square error of prediction (RMSEP) is 1.308. It is concluded that Vis-NIR spectroscopy with LS-SVM is a feasible method to determine the storage period of bee pollen. Moreover, the results for different storage periods were compared. It is shown that the storage periods between 11~50 can be well determined by LS-SVM.

Key words: Vis-NIR spectroscopy; storage period; least squares-support vector machines (LS-SVM)

引言

蜂花粉是一种天然营养食品,富含蛋白质、碳水化合物、矿物质、维生素和其它活性物质^[1].蜂花粉具有一定的医疗作用,如调节内分泌、治疗糖尿病、补肾安神和抗肿瘤等,被誉为“浓缩的天然药库”和

“全能的营养库”^[2].但是随着蜂花粉贮存时间的延长,营养成分会随着时间的推移而逐渐减少,相应的医疗效果也会有所降低.1989 年制定的国家标准中规定了过氧化氢酶作为蜂花粉新鲜度的检测指标,目前过氧化氢的检测方法还是以在特定的 pH 值、温度和底物浓度下产生的分子氧的体积来表示过氧

收稿日期:2009-02-21,修回日期:2009-12-18

Received date: 2009-02-21, revised date: 2009-12-18

基金项目:浙江省新苗人才计划(2007R40G2010051);中国博士后科学基金(20080431301);浙江省教育厅科研基金(20070125);2009 年浙江大学青年教师交叉研究种子基金

作者简介:金航峰(1986-),男,浙江杭州人,博士研究生,主要研究领域为近红外无损检测, jinhangfeng@zju.edu.cn.

化氢底物消耗的微克分子数,步骤繁琐,重复性较差,而且会对被检样品造成破坏.国内有人曾研究维生素 C 和黄酮类物质在蜂花粉贮存过程中的含量变化^[1,3],但是这些化学物质并不十分稳定,检测结果存在一定的误差.所以研究一种简单、快速、非破坏性的蜂花粉新鲜度的检测技术很有必要.

可见-近红外光谱分析技术是一种快速简便非破坏性的分析检测方法,可以充分利用全谱或多波长下的光谱数据进行定性或定量分析.由于可见-近红外光谱分析技术具有速度快、效率高、成本低、测试重现性好和测量方便等特点,已经被越来越多地应用于食品、石油、化工和医药等多种领域^[4-6].在蜂产品产业中,已有研究表明可见-近红外光谱技术可以用于蜂蜜和蜂花粉的化学组分分析^[7,8]、蜂蜜的品牌鉴别^[9]等. Giunchia 等采用傅里叶近红外光谱技术建立了鸡蛋新鲜度的无损检测模型^[10].侯瑞锋等采用近红外漫反射光谱建立了挥发性盐基氮的预测模型,并通过聚类分析方法对不同新鲜度的猪肉近红外光谱数据进行了分类处理^[11].但是基于可见-近红外光谱技术的蜂花粉贮存时间检测研究还鲜有报道.

本研究通过对常温下不同贮存时间茶花蜂花粉的可见-近红外反射光谱进行分析,目的在于寻找一种预测蜂花粉贮存时间的快速有效的方法.采用全光谱数据作为输入变量建立最小支持向量机(LS-SVM)模型,预测结果将与光谱分析常用的偏最小二乘法(PLS)以及主成分回归法(PCR)模型进行比较.并且分析了 LS-SVM 模型对于不同贮存时间段花粉的预测效果,确定模型适用的贮存时间检测范围.

1 材料与方法

1.1 样品来源和数据采集

本实验选用由浙江省养蜂场提供的新鲜茶花(*Camellia*)蜂花粉样本,常温下置于阴暗处保存,并在贮存的第 4 天至第 50 天每天采集样品的光谱数据.光谱采集使用美国 ASD(Analytical Spectral Device, Boulder, USA)公司的可见-近红外光谱仪(Handheld FieldSpec).光谱采样间隔为 1.5 nm,测定范围为 325 ~ 1075nm,扫描次数为 30 次,分辨率为 3.5nm,探头视场角为 45°.光源采用与光谱仪配套的 14.5V 卤素灯.盛样容器为直径 65mm,高 10mm 的玻璃皿.反射光谱采集时,将茶花蜂花粉取出置于盛样容器中,每次采集 30 个样本的光谱数据.对光谱仪进行标准白板光谱校正后,进行光谱描

1.2 光谱数据预处理

首先采用 Savitzky-Golay 平滑法,平滑点数取 7,用以去除光谱曲线中的高频成分,保留有用低频信息,然后进行附加散射校正(multiplication scatter correction, MSC)处理,以降低颗粒状茶花蜂花粉样本的散射作用对光谱数据造成的影响.预处理部分采用 Unscrambler V9.6(CAMO AS, Oslo, Norway)软件.为了消除光谱数据在首尾两端产生的较大噪音,只取 400 ~ 1000nm 波段的数据进行分析,以提高测量数据的信噪比.

1.3 最小二乘支持向量机(LS-SVM)

支持向量机(support vector machines, SVM)是建立在统计学习理论之上的一种具有良好性能的学习工具,因其具有强大的非线性处理能力和良好的推广能力而得到广泛的应用.最小二乘支持向量机(least square-support vector machines, LS-SVM)是在经典 SVM 的基础上进行改进后得到的算法,它能够进行线性和非线性的多元建模,并且降低计算的复杂性,加快求解速度,在少量的训练样本中进行高维特征空间学习.由于 RBF 核函数作为非线性函数能够减少训练过程中计算的复杂性,本文选择 RBF 核函数进行 LS-SVM 计算,得到 LS-SVM 的函数估计为:

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (1)$$

在 LS-SVM 模型建立过程中,需要对调整参数 γ 和 RBF 核函数参数 σ^2 进行选择. γ 对改进 LS-SVM 模型非常重要,而 σ^2 则控制函数回归误差,并且直接影响初始的特征值和特征向量. σ^2 过小会导致大量的回归量的产生,并且最终导致过拟合.相反, σ^2 过大会导致回归量的减少,模型过于简单,从而影响预测精度.此外, σ^2 还关系到 LS-SVM 模型对输入变量噪声的灵敏度. LS-SVM 计算采用 Matlab V7.6(The Math Works, Natick, USA).

2 实验结果与分析

2.1 光谱谱图分析

图 1 为常温下第 4 天至第 50 天茶花蜂花粉样品的光谱图,横坐标为波长,范围在 400 ~ 1000nm,纵坐标为光谱吸光度值.从图 1 可以看出,不同贮存时间茶花蜂花粉的光谱交错重叠.由于实验条件和仪器的限制,光谱数据存在不同程度的平移,还会引入一些噪声,要消除这些因素的影响,需要结合化学计量学方法对所得的光谱数据进行处理,从而建立

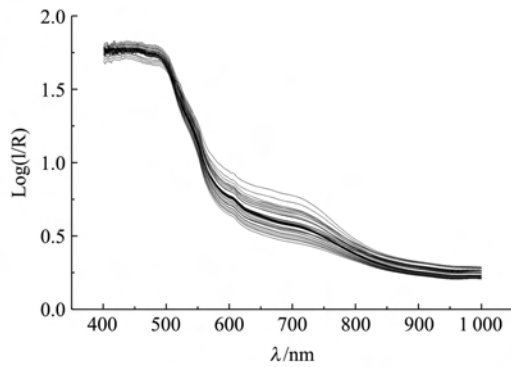


图1 不同贮存时间茶花粉的可见-近红外光谱曲线图
Fig.1 Vis-NIR absorbance spectra for Camellia pollen stored in different periods

起常温下蜂花粉贮存时间的检测模型。

2.2 LS-SVM 预测模型的建立

将实验得到的 47 天共 1410 个花粉样本随机分为两部分,其中 940 个样本(每一天的花粉 20 个样本)用于 LS-SVM 模型的建立,剩余 470 个样本(每一天的花粉 10 个样本)作为预测样本,用 400 ~ 1000nm 范围内的 600 个光谱数据作为输入变量进行 LS-SVM 建模和预测,模型预测参数见表 1。从表中可以看出,以全光谱数据作为输入变量建立的 LS-SVM 模型的预测效果很好(见图 2),预测相关系数 r_p 达到 0.996,且预测标准误差(standard error of prediction, SEP)和预测均方根误差(root mean square error of prediction, RMSEP)均较小,分别为 1.310 和 1.308,说明模型的预测性能较好。

同时分别选用偏最小二乘法(PLS)以及主成分回归法(PCR)两种常用方法建立的模型对茶花粉的贮存时间进行预测(表 1),可以看到 LS-SVM 模型的预测效果均好于其它两种模型。说明 LS-SVM 模型能够更好地提取光谱中的有用信息,从而更加精确地预测蜂花粉贮存时间。

2.3 不同时间段花粉的 LS-SVM 模型预测

将第 4 至 50 天共 47 天的预测样本按照 5 个时间段划分为 5 个预测集,具体分为:第 4 ~ 10 天、第 11 ~ 20 天、第 21 ~ 30 天、第 31 ~ 40 天、第 41 ~ 50 天。用已建立的 LS-SVM 模型分别对 5 个预测集的

表 1 基于不同模型的预测结果

Table 1 Prediction results for different models

建模方法	r_p	SEP	RMSEP
LS-SVM	0.996	1.310	1.308
PLS	0.978	2.798	2.796
PCR	0.976	2.937	2.937

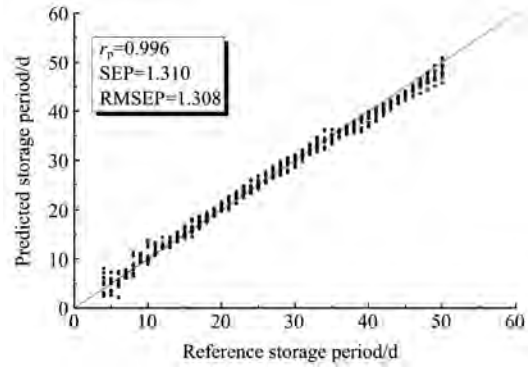


图2 LS-SVM 模型预测蜂花粉贮存时间结果
Fig.2 Pollen storage period prediction result based on the LS-SVM model

表 2 不同时间段花粉的 LS-SVM 模型预测结果

Table 2 Prediction results for pollen stored in different periods with the LS-SVM model

时间段(天)	r_p	SEP	RMSEP
4 ~ 10	0.840	1.520	1.510
11 ~ 20	0.968	0.902	0.897
21 ~ 30	0.958	0.908	0.904
31 ~ 40	0.926	1.232	1.225
41 ~ 50	0.952	1.880	1.870

样本进行预测,预测结果见表 2。从表中可以看出,第 4 ~ 10 天花粉样本的贮存时间预测精度相对较低。原因可能是贮存初期花粉中物质比较不稳定,含量变化相对较大,模型预测较困难。其余四个时间段的预测精度均达到了 0.9 以上,说明建立的 LS-SVM 模型能够检测 11 ~ 50 天时间段的花粉样本的贮存时间。

3 结论

采用可见-近红外光谱技术对常温下蜂花粉贮存时间进行检测。利用全光谱数据作为输入变量建立了 LS-SVM 模型,其预测结果优于 PLS 和 PCR 模型。通过比较不同时间段蜂花粉的 LS-SVM 模型的预测结果,得知该模型适用的蜂花粉样本贮存时间的检测范围为第 11 ~ 50 天。结果表明,运用基于 LS-SVM 的可见-近红外光谱技术可以快速、无损地检测常温下蜂花粉的贮存时间,为今后进一步研究光谱技术应用于蜂花粉新鲜度的检测提供了理论依据。

REFERENCES

- [1] ALMEIDA-MURADIAN L B, PAMPLONAA L C, COIMBRA S, *et al.* Chemical composition and botanical evaluation of dried bee pollen pellets[J]. *Journal of Food Compo-*

- sition and Analysis, 2005, **18**(1):105—111.
- [2] LADD P G. Pollen presenters in the flowering plants-form and function[J]. *Botanical Journal of the Linnean Society*, 1994, **115**(3):165—195.
- [3] CAMPOS M G, WEBBY R F, MARKHAM K R, et al. Age-Induced diminution of free radical scavenging capacity in bee pollens and the contribution of constituent flavonoids [J]. *J. Agric. Food Chem.*, 2003, **51**:742—745.
- [4] LI Ming-zan. *Spectroscopy Analysis Technology and Application* [M]. Beijing: Science Press(李民赞. 光谱分析技术及其应用. 北京: 科学出版社), 2006.
- [5] LIU Fei, HE Yong, WANG Li, et al. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy [J]. *Analytica Chimica Acta*, 2008, **610**(2):196—204.
- [6] ZUDEA M, PFLANIA M, KAPRIELIAN C, et al. NIRS as a tool for precision horticulture in the citrus industry [J]. *Biosystems Engineering*, 2008, **99**(3):455—459.
- [7] GARCIA-ALVAREZ M, HUIDOBRO J F, HERMIDA M, et al. Major components of honey analysis by near-infrared transreflectance spectroscopy [J]. *J. Agric. Food Chem.*, 2000, **48**(11):5154—5158.
- [8] GONZALEZ-MARTIN I, HERNANDEZ-HIERRO J M, BARROS-FERREIRO N, et al. Use of NIRS technology with a remote reflectance fibre-optic probe for predicting major components in bee pollen [J]. *Talanta*, 2007, **72**(3):998—1003.
- [9] SHAO Yong-ni, HE Yong, BAO Yi-dan. Application of visible/Near infrared spectroscopy to discriminating honey brands based on independent component analysis and BP neural network [J]. *Spectroscopy and Spectral Analysis*(邵咏妮, 何勇, 鲍一丹. 基于独立组分分析和 BP 神经网络的可见/近红外光谱蜂蜜品牌的鉴别. *光谱学与光谱分析*), 2008, **28**(3):602—605.
- [10] GIUNCHIA A, BERARDINELLI A, RAGNI L, et al. Non-destructive freshness assessment of shell eggs using FT-NIR spectroscopy [J]. *Journal of Food Engineering*, 2008, **89**(2):142—148.
- [11] HOU Rui-feng, HUANG Lan, WANG Zhong-yi, et al. The preliminary study for testing freshness of meat by using near-infrared reflectance spectroscopy [J]. *Spectroscopy and Spectral Analysis*(侯瑞锋, 黄岚, 王忠义, 等. 用近红外漫反射光谱检测肉品新鲜度的初步研究. *光谱学与光谱分析*), 2006, **26**(12):2193—2196.

(上接第 204 页)

的,而在实际应用当中还要考虑自然环境因素对目标偏振度的影响,如水体在波浪影响下的偏振信息校正方法等,这些问题还需要进一步的探讨.总之,利用偏振信息来反演盐度最佳波段对偏振传感器的设计与水体检测遥感技术的发展提供了可靠的依据.

REFERENCES

- [1] TAN Xiao-Sheng, GE Cheng-Hui. A method for calculating solar angles and its application on remote sensing [J]. *Remote Sensing for Land & Resources*(谈小生, 葛成辉. 太阳角的计算方法及其在遥感中的应用. *国土资源遥感*), 1995, **7**(2):48—50.
- [2] Bradley G, Henderson, James Theiler, et al. The polarized emissivity of a wind-roughened sea surface: A Monte Carlo model [J]. *Remote Sensing of Environment*, 2003, **88**(4):453—467.
- [3] MEI An-Xin, PENG Wang-Lu, QIN Qi-Ming, et al. *Introduction of Remote Sensing* [M]. Beijing: Higher Education Press(梅安新, 彭望禄, 秦其明, 等. *遥感导论*. 北京: 高等教育出版社), 2002:35—41.
- [4] CAO Han-Jun, QIAO Yan-Li, YANG Wei-Feng, et al. Characterization and analysis of the polarization images in remote sensing [J]. *Chinese Journal of Quantum Electornics*(曹汉军, 乔延利, 杨伟锋, 等. 偏振遥感图像特性表征分析. *量子电子学报*), 2002, **19**(4):373—378.
- [5] ZHANG Chao-Yang, CHENG Hai-Feng, CHEN Zhao-Hui, et al. Detecting low reflectivity camouflage net by using polarization remote sensing [J]. *J Infrared Millim. Waves*(张朝阳, 程海峰, 陈朝辉, 等. 偏振遥感识别低反射率伪装网研究. *红外与毫米波学报*), 2009, **28**(2):137—140.
- [6] WU Tai-Xia, YAN Lei, XING Yun, et al. Polarization reflection effect of plane rough surface under vertical observation [J]. *J Infrared Millim. Waves*(吴太夏, 晏磊, 相云, 等. 垂直观测时水平粗糙地表偏振反射作用研究. *红外与毫米波学报*), 2009, **28**(2):151—155.
- [7] Csiszar G I, Gutman P, Romanov, et al. Using ADEOS/POLDER data to reduce angular variability of NOAA/AVHRR reflectances [J]. *Remote Sensing of Environment*. 2001, **76**(3):399—409.
- [8] Marc Leroy, Bréon F M. Angular signatures of surface reflectances from airborne POLDER data [J]. *Remote Sensing of Environment*, 1996, **57**(2):97—98.
- [9] Egan W G, Duggin M J. Comparative merits of multispectral optical polarization to microwave [J]. *Proceeding of SPIE*, 2002, **4481**:292—298.
- [10] Shaw J A. Degree of liner polarization in spectral radiances from water-viewing infrared radiometers [J]. *Applied Optics*, 1999, **38**(15):3157—3165.
- [11] ZHAO Yun-Sheng, JIN Lun, SONG Kai-Shan, et al. Study on the characteristics of polarized reflectances on liquid surface [J]. *Journal of Northeast Normal University*(赵云升, 金伦, 宋开山, 等. 液体表面偏振反射特征研究. *东北师大学报自然科学版*), 2000, **32**(4):103—106.
- [12] LUO Yang-Jie, ZHAO Yun-Sheng, WU Tai-Xia, et al. Research and application of multi-angle polarization characteristics of water body mirror reflection [J]. *Science in China (Series D: Earth Sciences)*(罗杨洁, 赵云升, 吴太夏, 等. 水体镜面反射的多角度偏振特性研究及应用. *中国科学 D 辑:地球科学*), 2007, **37**(3):411—416.
- [13] JIN Xi-Feng, QIAO De-Lin, ZHOU Su-Xiang. The measuring device for bidirectional polarized reflectance [P]. Patent NO. 96239489.0, Mar. 4, 1998(金锡峰, 乔德林, 周素香. 二向性偏振反射测量装置: 96239489.0 专利权公告日, 1998-03-04).