

紫外可见吸收光谱结合主成分-反向传播 人工神经网络鉴别真假蜂蜜

欧文娟 孟耀勇* 张小燕 孔 猛

(华南师范大学生物光子学研究院, 广州 510631)

摘 要 研究紫外-可见吸收光谱技术结合化学计量学方法鉴别真假蜂蜜。根据蜂蜜中果糖和葡萄糖的典型质量比 1.2:1.0 配制与真蜂蜜相近的掺假溶液,并以 5%~20% 的比例掺入真蜂蜜中。获取纯正蜂蜜和掺假蜂蜜的紫外-可见吸收光谱,选择最佳敏感波段 250~400 nm 的吸光度值进行主成分分析(PCA),优选主成分作为反向传播人工神经网络(BPANN)的输入向量。输出结果显示,校准集和预测集的准确鉴别率均为 100%;对应的均方根误差分别为 8.523×10^{-3} 和 8.961×10^{-3} 。研究结果表明,基于 PCA-BPANN 的紫外-可见吸收光谱技术能够方便、快速、准确地鉴别真假蜂蜜,为食品质量的快速检测提供可靠参考。

关键词 蜂蜜; 掺假; 紫外-可见吸收光谱; 反向传播人工神经网络; 主成分-反向传播人工神经网络

1 引 言

蜂蜜是一种天然营养食品,具有很好的医疗保健作用和养颜效果。为保证蜂蜜质量,国家标准 GB18796 针对蜂蜜的理化性质作出强制性要求:果糖和葡萄糖含量 $\geq 60\%$,蔗糖含量 $\leq 5\%$,蜂蜜一级品水分 $\leq 20\%$,二级品水分 $\leq 24\%$ ^[1,2]。但一些不法商家对蜂蜜进行掺假,依靠常规的感官物理检测(色泽、滋味、气味、状态)和化学检测(水分、酸度、折射率、果糖、酶值、葡萄糖等)已无法准确鉴别。目前,检测蜂蜜的方法主要有傅里叶变换拉曼光谱法(FT-RS)、花粉分析法、高效液相色谱法(HPLC)、气相色谱-火焰离子化检测技术(GC-FID)、高效阴离子交换色谱-脉冲安培检测技术(HPAEC-PAD)、稳定性碳同位素($^{12}\text{C}/^{13}\text{C}$)比值分析法(SCIRA)及特异性天然同位素分馏核磁共振技术(SNIF-NMR)^[3~5],这些方法应用于检测掺假蜂蜜中,虽然具有很好的可行性,但大都操作繁琐、费时,而且仪器成本高。文献[6~8]研究了傅立叶变换红外-衰减全反射光谱(FTIR-ATR)结合软独立模型类簇法(SIMCA)和偏最小二乘法(PLS)检测掺假蜂蜜,有效解决了蜂蜜检测技术操作繁琐耗时的问题。文献[9]利用近红外光谱(NIR)结合小波变换-最小二乘支持向量机(WT-LV-SVM)检测掺假蜂蜜。对纯正蜂蜜掺假的检测方法主要集中于红外光谱法,尚未见紫外-可见吸收光谱法的研究应用。

本实验采用紫外-可见吸收光谱法,操作简单快捷,无需使用比色皿和毛细管,且仅需 0.5~10 μL 样品,适于检测少量及微量样品。测量时以水为参比,消除了水的羟基峰对蜂蜜检测的干扰。数据分析时采用模拟人脑细胞工作原理的反向传播人工神经网络(BPANN)和主成分分析法(PCA)。结果表明,紫外-可见吸收光谱与化学计量学相结合的方法可快速检测纯正蜂蜜和掺假蜂蜜,为食品质量检测提供可靠的依据。

2 实验部分

2.1 样品制备

D-果糖(生化试剂,国药集团化学试剂有限公司)和 D-葡萄糖(分析纯,广州化学试剂厂)按照质量比 1.2:1.0 配制与真蜂蜜相近的混合溶液^[7]。稀释纯正蜂蜜(紫云英蜂蜜,广州市宝生园有限公司),使得纯正蜂蜜溶液和果糖葡萄糖混合溶液的糖度均为 70 °Brix。在 40 °C 下水浴加热,直至水溶液中的晶体物质完全溶解。参照文献[7]配制掺假 7%、14% 和 21% 的蜂蜜,将果糖葡萄糖混合溶液以 5%、

2010-10-14 收稿; 2011-01-18 接受

本文系国家自然科学基金(No. 60411130595)和广东省中医药局项目(No. 2008233)资助

* E-mail: menggy@tom.com

10%、15%和20%分别掺到34个纯正蜂蜜溶液中。为对比分析全部为真蜂蜜溶液和全部为假蜂蜜溶液的不同光谱特征,同时配备了11个100%掺假样品(即果糖葡萄糖混合溶液)将制得的45个假蜂蜜样品标记为 $F_{n,i}$ (n 为样品序号; $i=5\%, 10\%, 15\%, 20\%, 100\%$,为掺假百分比)。其余16个未掺假的蜂蜜溶液记为纯正蜂蜜样品 H_n 。最后利用XW-80A涡旋混合器(广州泰通仪器有限公司)将每个样品溶液混合均匀。

2.2 仪器与光谱测量

ND-1000超微量分光光度计(美国NanoDrop公司),波长范围为220~750 nm,分辨率为3 nm,吸光率精确度为0.003(1 mm光程),吸光率范围0.02~75(相当于10 mm光程);移液枪(0~10 μL);AR1140电子天平(奥豪斯公司);MATLAB 7.0数据分析软件。

测量时以蒸馏水为参比,即消除了水在UV-Vis波段的吸收对液体样品检测的干扰。吸取2 μL 样品置于分光光度计测量基座上,采用UV-Vis模式测量样品的紫外-可见吸收光谱。每完成一个样品的测量,用蒸馏水清洗上下基座,以避免残余液对下一个样品测量的影响。每个样品重复测量3次后取均值。

3 结果与讨论

3.1 样品的紫外-可见吸收光谱图分析

本实验采集了紫外-可见波段220~750 nm的吸收光谱。由于仪器检测范围的原因,导致谱图两端220~250 nm和700~750 nm光谱噪声较大,因此,截取了250~700 nm波段进行分析。图1为纯正蜂蜜样品 H_n 和掺假样品 $F_{n,i}$ 截取波段的紫外-可见吸收光谱图。由图1可知,纯正蜂蜜样品与掺假的 $F_{n,5\%}$ 、 $F_{n,10\%}$ 、 $F_{n,15\%}$ 和 $F_{n,20\%}$ 样品的光谱图极其相似,在257 nm处出现波谷;273 nm处有强吸收峰;在406 nm处吸光度值近似为0,即主要为水的吸收;在440~700 nm区间,吸光度值基本保持不变,说明此波段内,主要为蜂蜜主要成分(果糖和葡萄糖)的吸收。而 $F_{n,100\%}$ 光谱谱线明显不同于纯正蜂蜜样品的光谱谱线,在250~385 nm之间谱线较平缓且吸收峰位于295 nm处。表明识别纯正蜂蜜样品和100%掺假样品,不需要任何数据处理,由光谱图即可鉴别。但鉴别掺假5%~20%的蜂蜜样品需进行统计处理才能辨别。

3.2 模型的建立及评价指标

选取图1中差别较大的敏感波段250~400 nm的吸光度值进行数据处理。每隔3 nm取一个数据点,此波段内共计51个光谱数据,因此,所有样品光谱组成了一个 61×51 维的矩阵。建立两种模型:原始光谱数据作为输入向量,“1”代表真蜂蜜样品,“0”代表假蜂蜜样品,作为目标向量(即样品的真实值),建立BPANN模型;对该波段的原始数据进行标准化处理后,提取主成分,分析二维主成分得分图,将代表光谱特征信息的主成分作为BPANN的输入,“1”代表真蜂蜜样品,“0”代表假蜂蜜样品作为目标向量,构建PCA-BPANN模型。两种模型中的BPANN

均采用三层结构,以Sigmoid函数作为连接隐含层和输出层的传递函数,按照校准集与预测集2:1的比例关系^[2]将40个样品作为校正集,21个样品作为预测集,训练次数设为2000,目标误差设为0.0001,输入层节点由MATLAB软件自动计算,隐含层节点需经过多次尝试才能确定,输出层节点设为1。

若样品的网络输出结果在0.9~1.1范围内,表示此样品为真蜂蜜;在-0.1~0.1范围内,表示此样品为假蜂蜜。换言之,样品的预测值与真实值之差,即绝对偏差值,在阈值 ± 0.1 内,则为正确识别;超出阈值范围,则为错误识别。评价模型优劣,主要依据准确鉴别率,同时比较校正集的均方根误差(RMSEC)与预测集的均方根误差(RMSEP)。若RMSEC和RMSEP近似相等,说明所建立的BPANN模型已有效逼近训练样本所蕴含的规律,具有很强的逼近能力,即泛化能力很好。均方根误差计算公式为^[2,10]:

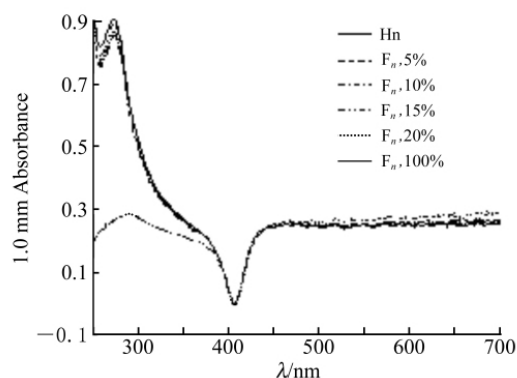


图1 纯正蜂蜜样品与掺假样品的紫外-可见吸收光谱图

Fig. 1 UV-Vis absorption spectra of authentic honey sample H_n and adulterated samples $F_{n,i}$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (1)$$

其中 $\hat{y}$ 为预测值, y_i 为真实值。

3.2.1 BPANN 模型 以 40 个样品的原始光谱数据 (40×51 维矩阵) 为输入向量, 经多次尝试, 隐含层设为 5 时网络最优化。迭代 39 次后, 网络输出误差为 6.4125×10^{-5} , 小于收敛误差界值 0.0001, 完成了目标。输出结果显示: 训练集中每个样品的绝对偏差都未超出阈值 ± 0.1 , 即准确率为 100%, $\text{RMSEC} = 8.008 \times 10^{-3}$ 。其余的 21 个样品作测试集, 预测结果的绝对偏差如图 2 所示。有 2 个样品 ($n = 2, 12$) 的绝对偏差超过 0.1, 测试集的准确率为 90.48%, $\text{RMSEP} = 6.1539 \times 10^{-2}$ 。可知, RMSEC 与 RMSEP 相差一个数量级, 说明由原始光谱所建立的 BPANN 模型未能完全有效地逼近训练样本所蕴含的规律, 逼近能力较弱。

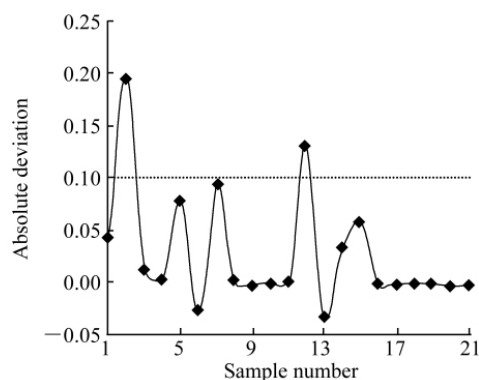


图 2 BPANN 模型中预测集的绝对偏差图

Fig. 2 Absolute deviation values of prediction set by BP-ANN model

3.2.2 PCA-BPANN 模型 对所有样品在 250 ~ 400 nm 的光谱数据 (61×51 维矩阵) 进行标准化处理后, 调用函数 princomp(X) 进行主成分分析。所有样品的第一、二主成分得分的运行结果如图 3 所示。 $F_{n,100\%}$ 全部分布在图右侧, 明显远离纯正蜂蜜样品和掺假 5% ~ 20% 的样品, 这是因为纯正蜂蜜中虽然 60% 以上是果糖和葡萄糖, 但还含有氨基酸、蛋白质、维生素、矿物质和微量元素等成分; 掺假 5% ~ 20% 的样品依据掺假量的不同, 与纯正蜂蜜样品分布的距离远近也不同。比如, 掺假 5% 的样品与纯正蜂蜜样品分布在同一区域, 并且第二主成分出现交叉重叠现象。由各点在横轴上的投影可见, 掺假含量越多, 第一主成分得分越大, $F_{n,100\%}$ 得分最大。但是掺假量相对少的样品与纯正蜂蜜样品在横轴上的投影却非常接近, 无法准确区分。所以, 本研究联合了 BPANN 方法共同分析, 利用提取的主成分作为 BPANN 的输入向量并建立模型。

为验证 PCA-BPANN 模型的可靠性及准确性, 优化了不同数量的主成分作为输入变量: (1) 由主成分分析可知, 前 3 个主成分的累积贡献率达到 99.27%, 可代表原始光谱^[10], 因此选取前 3 个主成分作为 BPANN 模型的输入。输出结果显示, 校准集和预测集的准确率均为 100%, 相关参数及其它性能指标见表 1; (2) 根据数理统计中选取主成分的一般规律, 即主成分的累积贡献率超过 85% ~ 95%, 就可解释大部分原始数据方差, 代表原始光谱^[11], 而第一主成分的贡献率达到 96.74%, 因此选取第一主成分作为 BPANN 模型的输入。输出结果显示, 校准集和预测集的准确率均为 100%。

从两种 PCA-BPANN 模型的预测结果 (表 1) 可知, 所有预测样品的绝对偏差都未超过阈值, 表明两种方法的预测结果都 100% 准确, 从而验证了所建立的 PCA-BPANN 模型的可靠性及准确性。由两者的其它性能指标可见, 以前 3 个主成分作输入变量时, 校准集的 RMSEC 约为预测集的 RMSEP 的 2.39 倍; 而以第一主成分作输入变量时, RMSEC 和 RMSEP 接近, 即测试样本的误差接近于训练样本的误差, 说明建立的网络模型已有效逼近训练样本所蕴含的规律, 具有很好的泛化能力。相比而言, 只用一个主成分建模, 性能更好。

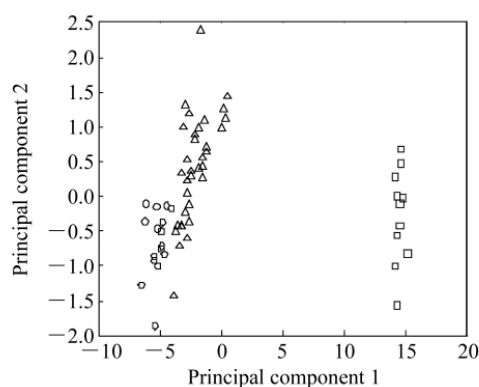


图 3 所有样品的第一主成分与第二主成分得分散点图

Fig. 3 Scores plot of all samples on principal component 1 and 2

Δ : $F_{n, 5\% \sim 20\%}$; $\circ$: H_n ; $\square$: $F_{n, 100\%}$.

表 1 两种 PCA-BPANN 模型的参数与结果比较

Table 1 Parameters and results of two PCA-BP-ANN models

输入变量 Input variables	隐含层节点数 Neuron No. of hidden layer	输出层节点数 Neuron no. of output layer	迭代次数 Iterations	校准集 Calibration set ($n = 40$)		预测集 Prediction set ($n = 21$)	
				准确率 Correct identification rates	RMSEC	准确率 Correct identification rates	RMSEP
PC1 , PC2 , PC3	5	1	16	100%	0.008138	100%	0.003401
PC1	6	1	9	100%	0.008523	100%	0.008961

PC: Principal Component.

3.3 小结

利用紫外-可见吸收光谱法检测纯正蜂蜜样品、掺假 5% ~ 20% 的蜂蜜样品和 100% 掺假液的样品。选取最佳敏感波段 250 ~ 400 nm 的吸光度值进行反向传播人工神经网络(BPANN) 建模,其预测集准确率为 90.48% ,RMSEP 为 6.1539×10^{-2} ;而对选取的吸光度值主成分分析后,再构建 BPANN 模型,其准确率为 100% ,RMSEP 为 8.961×10^{-3} 。说明 PCA-BPANN 模型比单独利用 BPANN 建模效果更好。而且,经过主成分分析后,输入变量得到大大压缩,进而显著提高了人工神经网络的收敛速度和泛化能力。因此可以考虑增加样品数量及掺假百分含量进一步验证模型的准确性。研究结果表明,基于 PCA-BPANN 的紫外-可见吸收光谱技术能够方便、快速、准确地鉴别真假蜂蜜,为食品质量的快速检测提供可靠参考。

References

- 1 GB 18796-2005 *Honey*(蜂蜜) . National Standards of the People's Republic of China(中华人民共和国国家标准) ,2005
- 2 TU Zhen-Hua , ZHU Da-Zhou , JI Bao-Ping , MENG Chao-Ying , WANG Lin-Ge , QING Zhao-Kun(屠振华 , 朱大洲 , 籍保平 , 孟超英 , 王林舸 , 庆兆坤) . *Chinese J. Anal. Chem.* (分析化学) ,2010 ,38(1) : 45 ~ 50
- 3 Cotte J F , Casabianca H , Chardon S , Lheritier J , Grenier-Loustalot M F. *J. Chromatogr. A* ,2003 ,1021(1) : 145 ~ 155
- 4 Cordella C , Militão J S L T , Clément M , Dragnudel P , Cabrol-Bass D. *Anal. Chim. Acta* ,2005 ,531(2) : 239 ~ 248
- 5 ZHU Qing , WANG Bi-Ni , LIANG Yan , CHEN Wei-Jun , CAO Wei (朱青 , 王毕妮 , 梁艳 , 陈卫军 , 曹伟) . *Apic. China*(中国蜂业) ,2008 ,59(10) : 35 ~ 37
- 6 Gallardo-Velázquez T , Osorio-Revilla G , Loa M Z , Rivera-Espinoza Y. *Food Res. Int.* ,2009 ,42(3) : 313 ~ 318
- 7 Kelly J F D , Downey G , Fouratier V. *J. Agric. Food Chem.* ,2004 ,52(1) : 33 ~ 39
- 8 Kelly J D , Petisco C , Downey G. *J. Agric. Food Chem.* ,2006 ,54(17) : 6166 ~ 6171
- 9 Zhu X R , Li S F , Shan Y , Zhang Z Y , Li G Y , Su D L , Liu F. *J. Food Eng.* ,2010 ,101(1) : 92 ~ 97
- 10 TANG Shou-Peng , YAO Xin-Feng , YAO Xia , TIAN Yong-Chao , CAO Wei-Xing , ZHU Yan(汤守鹏 , 姚鑫锋 , 姚霞 , 田永超 , 曹卫星 , 朱艳) . *Chinese J. Anal. Chem.* (分析化学) ,2009 ,37(10) : 1445 ~ 1450
- 11 CHEN Gui-Ming , QI Hong-Yu , PAN Wei(陈桂明 , 戚红雨 , 潘伟) . *MATLAB Mathematical Statistics*(6. x) (MATLAB 数理统计(6. x)) . Beijing(北京) : Science Press(科学出版社) ,2002: 199 ~ 208

Application of UV-Visible Absorption Spectroscopy and Principal Components-Back Propagation Artificial Neural Network to Identification of Authentic and Adulterated Honeys

OU Wen-Juan¹ , MENG Yao-Yong^{*1} , ZHANG Xiao-Yan¹ , KONG Meng¹¹(College of Biophotonics , South China Normal University , Guangzhou 510631)

Abstract UV-visible (UV-vis) absorption spectroscopy in combination with chemometrics was used to identify the authentic and adulterated honeys. Adulterant solution prepared using *D*-fructose and *D*-glucose following the mass ratio of typical of honey composition (1.2: 1.0) was close to the real honey and added to individual

