

# 近红外光谱法结合模式识别方法对不同品牌牙膏进行质量监控

林 薇<sup>1,2</sup>, 倪永年<sup>1,2\*</sup>

1. 南昌大学食品科学与技术国家重点实验室, 江西 南昌 330047
2. 南昌大学化学系, 江西 南昌 330031

**摘要** 采用了近红外光谱法测定了4个品牌牙膏样品的近红外光谱, 然后对所得的光谱进行变量预处理, 再运用偏最小二乘法(PLS)、人工神经网络法(ANN)和K-最近邻法(KNN)等几种有监督模式识别法, 以及主成分分析(PCA)和聚类分析两种无监督模式识别法对样品进行了品牌的分类及聚类分析。结果表明采用近红外光谱法所得的光谱变量经多元散射处理后的分类结果比较理想。各品牌牙膏的质量相对是稳定, 但在采集的四种品牌牙膏中, 有两种品牌的牙膏样品间存在较大差异。

**关键词** 近红外光谱法; 模式识别; 牙膏; 质量监控

中图分类号: O657.3 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2011)08-2106-03

## 引言

牙膏是生活的必需品, 它的主要作用是清洁牙齿, 同时还有去污, 消炎的作用。现在市场上牙膏品种较多, 尽管牙膏的成分大致相同, 但是不同品牌牙膏间的质量会存在差异, 而同一品牌的牙膏质量也可能不完全稳定, 因此对不同品牌牙膏进行质量监控是相当必要的。

近红外光谱技术<sup>[1]</sup>具有快速、简便、操作简单、无损伤等特点, 本文运用近红外光谱作为一种快速的鉴定方法, 在将测定光谱变量进行预处理后, 结合模式识别方法<sup>[2, 3]</sup>, 对各个品牌牙膏进行较快速的分类, 以此对各品牌牙膏进行质量监控。

## 1 实验部分

### 1.1 样品准备

从市场上购买四种品牌而批次不同的牙膏样品共69个, 其中包括三种成人牙膏(品牌分别设为A、B和C)和一种儿童牙膏(品牌设为D)。其中A样品20个, B样品19个, C样品20个, 儿童牙膏D样品10个。所有样品在100℃烘箱中干燥48h, 然后将样品用粉碎机(RE-02, 650W, 武汉屹立工具有限公司)粉碎2min, 过40目筛。

### 1.2 光谱测量和预处理

精确称取质量相同的牙膏样品, 置于近红外样品池内(直径22mm, 深2.5mm), 在Hitachi 4100近红外光谱仪上测定800~2500nm范围内近红外光谱, 分辨率为2nm。每个样品扫描三次并取其平均值, 共得到851个光谱变量。

近红外光谱同时受样品的化学和物理因素的影响<sup>[4]</sup>, 因此在模式识别前通常采用数学方法对光谱数据进行预处理以减少噪音的影响<sup>[5, 6]</sup>, 以增强样品中化学组分对光谱的贡献。本文采用多元散射校正(MSC)对所得光谱数据进行预处理<sup>[6-9]</sup>。光谱经过散射校正后, 可以有效消除散射影响<sup>[8, 10]</sup>, 增强与成分含量相关的光谱吸收信息。使用此法时, 先要建立一个待测样品的“理想光谱”, 该光谱的变化与样品中的成分的含量满足线性关系, 然后以该光谱为标准对所有样品的近红外光谱进行校正。在实际应用中, 一般取所有光谱的平均光谱作为理想光谱, 每个样品的近红外光谱与标准光谱进行一元线性回归运算, 再求得的回归常数和回归系数对各红外光谱进行校正。

### 1.3 模式识别区分各品牌牙膏

模式识别是用于样品分类识别的多元分析方法, 是借助数学方法和计算机技术来完成的一种综合技术。本文采用偏最小二乘判别分析法(PLS)、K-最近邻法(KNN)以及人工神经网络(ANN)等三种有监督模式识别法<sup>[3, 11]</sup>对牙膏样品进行了分类。有监督模式识别法常将样品分作为训练集和测试集。PLS是一种基于因子分析的多变量校正方法, 通过找到多元空间的变量趋势来判别在训练集中所建立的类, 最佳组

收稿日期: 2010-10-27, 修订日期: 2011-01-22

基金项目: 国家自然科学基金项目(NSFC21065007)和南昌大学食品科学教育部重点实验室开放基金项目(SKLF-MB-201002, SKLF-TS-200919)资助

作者简介: 林 薇, 女, 1987年生, 南昌大学化学系硕士研究生 \*通讯联系人 e-mail: ynni@ncu.edu.cn

分数可以用交叉验证法确定；KNN 基于训练集样本与未知样本的距离来进行样品判别；ANN 利用由大量简单的处理单元拓展连接而组成的复杂网络，一个神经网络由许多非线性处理基本单元组成，亦称为神经元或节点，神经元是整个神经网络中的一个重要组成部分，以此模拟大脑的神经网络结构和行为。

主成分分析(PCA)<sup>[3, 12]</sup>和聚类分析(CA)<sup>[3, 12]</sup>都属于无监督的模式识别法，PCA 可以分析变量间的相关性，对多变量的光谱数据进行压缩、验证和可视化，在伪品检验和质量监控等方面都有广泛应用；CA 特别适用于样品归属不清楚的情况，聚类分析中，样本与样本之间的距离常用欧式距离表示，而当两种样本合成新类后，样本与类之间或类与类之间的距离表示法就必须确定，本文采用树状聚类分析来对样品进行分类。

## 2 结果与讨论

### 2.1 样品的原始光谱和预处理后的光谱

四个种类牙膏所有样品的原始光谱及多元散射校正后的光谱如图 1，从图可知，所有样品的原始光谱十分相似，并且在所选波长(800~2 500 nm)的吸收变化不大，这可能是牙膏中具有近红外吸收的物质不多，同时，采集的光谱受到样品不均匀、光散射、随机噪音以及基线漂移等影响<sup>[7]</sup>，需要进行光谱前处理来消除噪音，从图可知，经多元散射校正(MSC)处理后的光谱在 1 200~2 500 nm 范围内的特征性有所增强。

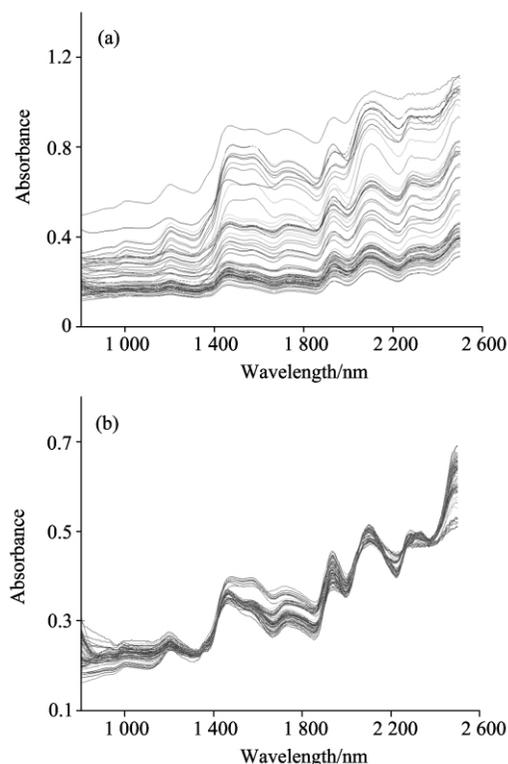


Fig 1 The NIR spectra of 69 toothpaste samples

(a): Original; (b): Preprocessed by MSC

### 2.2 模式识别用于牙膏的分类

#### 2.2.1 有监督模式识别法对样品进行分类

将校正后的光谱数据作为输入变量，并比较各种识别方法的预报结果，以考察各方法的预报能力(预报能力是指模型对预报集样本进行正确分类的比率)。将四个品牌 69 个牙膏样品分成两组，第一组 39 个样品作为训练集，其中包括 11 个 A 品牌牙膏样品，11 个 B 品牌牙膏样品，11 个 C 品牌牙膏样品以及六个 D 品牌牙膏样品；第二组 30 个样品作为预报集，包括九个 A 品牌牙膏样品，八个 B 品牌牙膏样品，九个 C 品牌牙膏样品以及四个 D 品牌牙膏样品。

进而建立 PLS 法、ANN 法和 KNN 法校正模型，这三种方法均首先需要对模型的参数进行训练，获得最佳参数后可对未知样品进行预报，预报结果列于表 1 中。从结果可知，ANN 对牙膏的品牌识别率优于其他两种方法。由此可以大致的推断，采用近红外光谱法所得的光谱变量，在一定程度上反应了各品牌牙膏的内在质量，可用于牙膏的质量监控。

Table 1 The prediction results by different methods

样品种类	准确分类样品的数目		
	PLS <sup>a</sup>	ANN <sup>b</sup>	KNN <sup>c</sup>
A	7	9	8
B	6	8	7
C	6	9	7
D	3	4	4
识别率/%	73	100	87

注：a: PLS 法的因子数为 4；b: ANN 法的训练次数和最小均方差误差分别为 400 和 0.000 1；c: KNN 的 K 值为 5

#### 2.2.2 无监督模式识别法对样品进行分类

采用主成分分析(PCA)和聚类分析法(CA)对所有样品进行分类。图 2 为 69 个样品在 PCA 的 PC1~PC2 主成分图上的得分，可以明显的看到样品可分为四类，其中儿童牙膏 D 独自聚为一类，此类样品在 PC1 上得分都为负值，与成人牙膏明显不同。而成人牙膏分为三类，但这三类样品在 PC1 上的得分相近，特别 A 和 B 类成人牙膏聚类紧凑，它们之间的差异主要表现在第一主成分上，说明品牌不同的成人牙膏的主要成分比较一致，但一些次要成分不同。

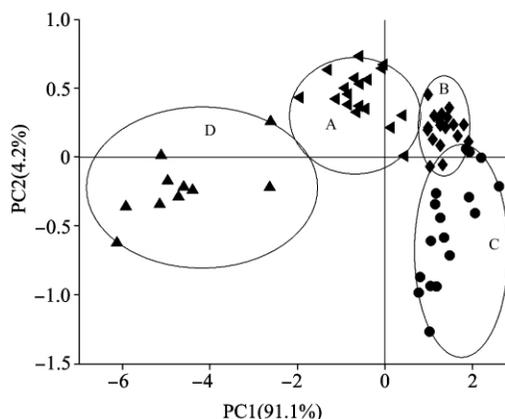


Fig 2 Scores plot of PCA of 69 toothpaste samples with NIR data as input variables

图 3 为 69 个牙膏样品的树状分类图。从图可以看出, 四种品牌牙膏能很好的被分类, 只有个别的样品分类错误(分类错误的样品图中已标出, 共七个), 其中儿童牙膏 D 很明显的与成人牙膏的距离较远而聚为一类。同时, 儿童牙膏 D 和成人牙膏 C 样品间的距离较大, 这与主成分分析的结果一致。

### 3 结 论

通过近红外光谱仪得到吸收光谱, 采用模式识别来对各品牌牙膏进行质量监控。结果表明, 各个品牌牙膏的质量是相对稳定的, 但在购买的样品中, 有两个品牌的牙膏质量依然存在较大的差异。本文所建立的研究方法快速、简便, 为牙膏的快速质量监控提供了参考。

### References

- [1] Luypaert J, Massart D L, Vander Heyden Y. *Talanta*, 2007, 72: 865.
- [2] ZHANG Shun-wei, WANG Zhi-jun, ZHU Zhong-liang(张顺玮, 王志军, 朱仲良). *Computers and Applied Chemistry(计算机与应用化学)*, 2008, 25(3): 333.
- [3] Berrueta L A, Alonso-Sclces R M, Heberger K. *J. Chromatogr A*, 2007, 1158: 196.
- [4] WU Jing-zhu, WANG Yi-ming, ZHANG Xiao-chao, et al(吴静珠, 王一鸣, 张小超, 等). *Modern Scientific Instruments(现代科学仪器)*, 2006, 1: 69.
- [5] Rinnan A, van den Berg F, Engelsen S B. *Trac-Trend Anal. Chem.*, 2009, 28(10): 1201.
- [6] Candolfi A, De Maesschalck R, Jouan-Rimbaud D, et al. *J. Pharm. Biomed. Anal.*, 1999, 21: 115.
- [7] WANG Yuan-hui, LI Qing-tao, JIANG Yi, et al(王远辉, 黎庆涛, 姜毅, 等). *Food Science and Technology(食品科技)*, 2009, 34(8): 283.
- [8] NI Zhen, HU Chang-qin, FENG Fang(尼珍, 胡昌勤, 冯芳). *Chin. J. Pharm. Anal.(药物分析杂志)*, 2008, 28(5): 824.
- [9] ZHU Xiao-li, YUAN Hong-fu, LU Wan-zhen(褚小立, 袁洪福, 陆婉珍). *Progress in Chemistry(化学进展)*, 2004, 16(4): 528.
- [10] Liu F, He Y. *Food Ses. Int.*, 2008, 41: 562.
- [11] Zhao Y, Wang J, Lu Q J, et al. *Innovative Food Science & Emerging Technologies*, 2010, 11(3): 520.
- [12] NI Yong-nian(倪永年). *The Application of Chemometrics in Analysis Chemistry(化学计量学在分析化学中的应用)*. Beijing: Science Press(北京: 科学出版社), 1999. 5: 161.

## The Application of Near-Infrared Spectroscopy and Pattern Recognition to Quality Assessment of Toothpaste Samples of Different Brands

LIN Wei<sup>1,2</sup>, NI Yong-nian<sup>1, 2\*</sup>

1. The State Key Laboratory of Food Science and Technology of Nanchang University, Nanchang 330047, China
2. Department of Chemistry, Nanchang University, Nanchang 330031, China

**Abstract** The near-infrared spectroscopy (NIR) was combined with pattern recognitions method and applied to the quality assessment of toothpaste samples of four different brands. Several chemometrics approaches, such as principal component analysis (PCA), clustering analysis (CA), partial least squares (PLS), artificial neural networks (ANN) and K-nearest neighbor (kNN) were used to investigate the quality of toothpastes samples. The obtained results showed that the four clustering groups can be observed after the pretreatment of multiple scatter correction for the NIR data. It was also found that the quality of toothpastes of all the four brands was relatively stable, however, there is a significant difference in the quality between two brand kinds of toothpaste samples.

**Keywords** Near-infrared spectrometry; Pattern recognition; Toothpastes; Quality assessment

\* Corresponding author

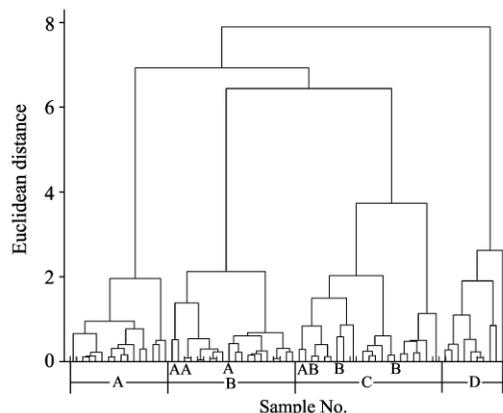


Fig 3 Cluster of all toothpaste samples with NIR spectra as input variables