

# 应用近红外光谱快速鉴别酸奶品种的研究

何勇, 冯水娟, 李晓丽, 裘正军\*

浙江大学生物系统工程与食品科学学院, 浙江 杭州 310029

**摘要** 提出了一种用近红外光谱技术快速鉴别酸奶品种的新方法。首先应用光谱仪获得5种典型酸奶品种的光谱曲线,用主成分分析法对5种酸奶品种进行聚类分析,再结合人工神经网络技术建立模型进行品种鉴别。主成分分析表明,主成分1和主成分2的累积可信度已达98.96%,前7个主成分的累积可信度达到99.97%。以每一个样品的前7个主成分作为神经网络的输入,品种类型作为神经网络的输出,建立三层BP人工神经网络模型。每个品种各27个样本,5个品种共135个样本用来建立神经网络模型,余下每个品种各5个共25个用于预测。建模品种的拟合率和预测品种的识别率均为100%。说明该方法能快速无损的检测酸奶品种,为酸奶的品种鉴别提供了一种新方法。

**关键词** 近红外光谱; 酸奶; 主成分分析; 人工神经网络; 鉴别

**中图分类号**: S123; TH744.1 **文献标识码**: A **文章编号**: 1000-0593(2006)11-2021-03

## 引言

酸奶起源于保加利亚,我国晋朝已有关于酸奶的记载。由于它所含有的蛋白质经一定程度的分解后易于被人体消化吸收,有较高的营养价值和口味,也具有健美和减肥等功效,深受广大消费群体(儿童、青少年、中青年女性等)喜爱。近年来我国一批大中型乳品企业在市场竞争中奇迹般地崛起,市场超市中的酸奶品种也五花八门,但市场中也不乏各种假冒产品。一些学者利用近红外光谱技术在牛奶的脂肪、蛋白质和乳糖等含量方面做了一些研究<sup>[1-5]</sup>,但是对于酸奶品种鉴别方面的研究还做得很少,主要是在实验室进行。随着市场上酸奶品种的加速发展,品种鉴别将越来越凸现出它的必要性,所以研究一种简单、快速、无损的酸奶品种鉴别方法很有必要。

由于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、测试方便等特点,已经被越来越多地应用于食品工业、石油化工、制药工业等领域<sup>[6-8]</sup>。因而也可用于酸奶的检测。但进行光谱测试后如何从大量的信息中获取有效信息是研究的一个热点。主成分分析是多元统计中的一种数据挖掘技术。在不丢失主要光谱信息的前提下选择为数较少的新变量来代替原来较多的变量,解决了由于谱带的重叠而无法分析的困难。我们选用主成分分析(PCA)和基于误差反向传播算法(back propagation, BP)多层前馈神经网络

建立不同品种酸奶品种的近红外光谱鉴别模型。

## 1 材料与方法

### 1.1 仪器设备

实验使用美国ASD(analytical spectral device)公司的Handheld Field spec光谱仪,其采样间隔(波段宽)1.5 nm,测定范围325~1075 nm,扫描次数30次,分辨率3.5 nm,探头视场角为20°。光源是与光谱仪配套的14.5 V卤素灯。分析软件为ASD View Spec Pro, Unscramble V9.2和DPS(data procession system for practical statistics)<sup>[9]</sup>。

### 1.2 样品来源及光谱的获取

从超市买来生产日期为同一天的五种原味酸奶,分别是蒙牛(内蒙古)、均瑶(上海)、光明(上海)、伊利(内蒙古)和双峰(杭州),每种酸奶各取32个样本,共计160个样本。为减少实验过程中的操作误差,酸奶容器均采用直径为6.5 cm高度为1.4 cm的玻璃器皿,光谱仪置于盛放酸奶的器皿上方,距离酸奶表面150 mm,探头视场角为20°,对每个样品扫描30次,取平均值。

### 1.3 光谱数据预处理

为了消除外界环境的影响,提高光谱数据的有效信息量,我们对光谱数据采用平均平滑法进行预处理,选用平滑窗口大小为9,此时能很好地滤除各种因素产生的高频噪声。由于光谱曲线在首端和末端有较大噪声(如图1所示),所以

收稿日期: 2005-09-30, 修订日期: 2006-01-16

基金项目: 国家自然科学基金, 高等学校优秀青年教师教学科研奖励计划(02411)和浙江省自然科学基金人才基金(RC02067)资助项目

作者简介: 何勇, 1963年生, 浙江大学生物系统工程与食品科学学院教授 \*通讯联系人

取 400 ~ 1 000 nm 波段的光谱数据用于本实验分析。

#### 1.4 人工神经网络模型

全部样本随机分成建模集和预测集,建模集有五种酸奶各 27 个样本,共计 135 个。预测集有五种酸奶各 5 个样本,共 25 个。建立了一个三层的人工神经网络结构,各层传递函数都用 S 型(Sigmoid)函数。经主成分分析后得到的每个样本的前 7 个主成分作为神经网络的输入,即输入层节点数为 7,经多次实验确定隐含层节点数为 11,输出层节点数为 1(酸奶不同的品种值)。设定系统允许误差为 0.000 1,网络指定参数中学习速率为 0.4,设定训练迭代次数为 3 000 次。

## 2 试验结果与分析

### 2.1 酸奶样本的近红外漫反射光谱

5 种酸奶品种的典型近红外光谱曲线如图 1 所示。图 1 中横坐标为波长,范围是 325 ~ 1 075 nm,纵坐标为光谱漫反射率。从图 1 中可以看出,不同品种酸奶品种的光谱曲线有明显区别,并具有一定的特征性和指纹性,这为酸奶的品种鉴别奠定了基础。我们选择波长范围在 400 ~ 1 000 nm 的光谱,应用 ASD View Spec Pro 软件,把测得的光谱曲线做平均处理,并转换成 ASCLL 码,形成反射率矩阵并进行平滑预处理,然后利用 Unscramble V9.2 软件对光谱数据进行主成分分析(PCA)。

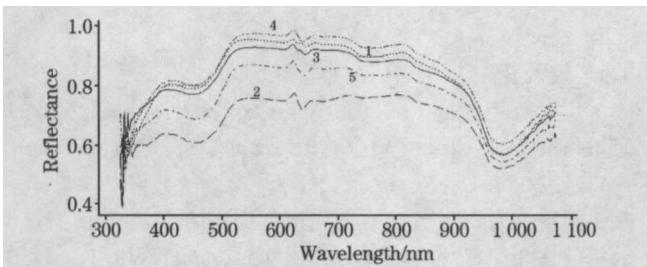


Fig 1 Reflectance spectra of five different varieties of yogurt  
1: Mengniu; 2: Junyao; 3: Guangming; 4: Yili; 5: Shuangfeng

### 2.2 不同品种酸奶的主成分分析

对蒙牛、均瑶、光明、伊利和双峰 5 种酸奶各 32 个样本,共 160 个样本进行主成分分析聚类。前 7 个主成分的特征值及累计可信度如表 1 所示。

Table 1 PCs and reliabilities

主成分	PC1	PC2	PC3	PC4	PC5	PC6	PC7
累计可信度/ %	96.04	98.96	99.61	99.85	99.93	99.95	99.97

### 2.3 基于 7 个主成分建立的 BP 酸奶品种预测模型

波段从 325 ~ 1 075 nm 共有 750 个点,如果全部用来建立人工神经网络模型不仅计算量大,而且有些区域样品的光谱信息很弱,与样品的组成或性质间缺乏相关关系。为此我们在主成分分析的基础上,选取对酸奶品种贡献率大的前 7 个主成分作为输入建立品种预测模型。

从 400 ~ 1 000 nm 有效光谱范围中取每个试验样本的前 7 个主成分作为神经网络的输入元,品种值作为神经网络的输出元(将蒙牛、均瑶、光明、伊利和双峰的品牌值分别设为 1, 2, 3, 4, 5)。建立了一个 7 个输入单元, 11 个隐含单元和一个输出单元的三层 BP 神经网络模型。5 个品种 135 个建模样本的拟合残差为  $9.1836 \times 10^{-5}$ ,对要鉴别的 25 个样本进行品种识别。结果表明,建模样本的品牌拟合率和预测品牌的识别率均为 100%(见表 2)。

Table 2 The Discrimination results for unknown samples by BP model

预测样本序号	真实值	预测值
(1)	1	1.001 99
(2)	1	1.006 90
(3)	1	1.008 19
(4)	1	1.005 52
(5)	1	1.003 76
(6)	2	2.024 90
(7)	2	1.981 73
(8)	2	2.075 44
(9)	2	1.920 87
(10)	2	2.041 23
(11)	3	3.128 09
(12)	3	3.127 87
(13)	3	3.070 92
(14)	3	3.211 99
(15)	3	3.108 06
(16)	4	4.155 83
(17)	4	4.200 99
(18)	4	4.006 95
(19)	4	4.131 80
(20)	4	3.974 08
(21)	5	4.979 34
(22)	5	4.994 03
(23)	5	4.995 95
(24)	5	4.977 01
(25)	5	4.993 52

## 3 结 论

提出了一种应用光谱技术并结合主成分分析和人工神经网络技术对酸奶品种进行快速无损检测的新方法,该方法使用方便,准确度高。试验表明对五种酸奶品种的识别率达到 100%。说明运用近红外光谱技术可以快速、准确、无损的对酸奶品种进行鉴别。提出的基于主成分分析和人工神经网络的光谱数据分析方法可大大地减少计算量,加快了分析和识别速度,同时得出的酸奶品种的主成分变量为快速鉴别酸奶种类和开发酸奶快速识别仪器提供了依据,也为其他液体食品品种的鉴别提供了有效的方法。

## 参 考 文 献

- [1] Lobodan, Yukihiro Ozaki. *Analytical Chemistry*, 2001, 73(1): 64.
- [2] Tsenkova R, Atanassova S, Itoh K, et al. *J. Anim. Sci.*, 2000, 78: 515.
- [3] LI Qing-bo, WANG Yan, XU Ke-xin(李庆波, 汪 燕, 徐可欣). *Food Science(食品科学)*, 2002, 23(6): 122.
- [4] L ÜLi-na, ZHANG Yue, ZHOU Ding-wen(吕丽娜, 张 钊, 周定文). *Journal of Tianjin University(天津大学学报)*, 2004, 37(12): 1093.
- [5] LIU Jie, YU Chang-qing, LI Jia-ze, et al(刘 杰, 于常青, 李家泽, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2001, 21(6): 769.
- [6] TANG Qi-yi, FENG Ming-guang(唐启义, 冯明光). *DPS Data Processing System for Practical Statistics(实用统计分析及其 DPS 数据处理系统)*. Beijing: Science Press(北京: 科学出版社), 2001.
- [7] Antihus H G, HE Yong, Annia G P. *Journal of Food Engineering*, 2006, 77(2): 313.
- [8] HE Yong, LI Xiao-li, SHAO Yong-ni(何 勇, 李晓丽, 邵咏妮). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2006, 26(5): 850.
- [9] ZHANG Yun, HE Yong. *Bioresource Technology*, 2006, 97(16): 2024.

## Study on Fast Discrimination of Varieties of Acidophilous Milk Using Near Infrared Spectra

HE Yong, FENG Shui-juan, LI Xiao-li, QIU Zheng-jun \*

College of Biosystems Engineering and Food Science, Zhejiang University, Hangzhou 310029, China

**Abstract** A new method for the discrimination of varieties of near acidophilous milk by means of near infrared spectroscopy (NIRS) was developed. Firstly, through the principal component analysis (PCA) of spectroscopic curves of 5 typical kinds of acidophilous milk, the clustering of acidophilous milk varieties was processed. The analysis results showed that the cumulate reliabilities of PC1 and PC2 (the first two principal components) reached 98.96%, and the cumulate reliabilities of PC1 to PC7 (the first seven principal components) were 99.97%. Secondly, a discrimination model of artificial neural network (ANN-BP) was set up. The first seven principal components of the samples were applied as ANN-BP inputs, and the values of type of acidophilous milk were applied as outputs, then the three layer ANN-BP model was build. In this model, every variety of acidophilous milk includes 27 samples, the total number of samples is 135, and the rest 25 samples were used as prediction set. Calculation results showed that the distinguishing rate of the five acidophilous milk varieties was 100%. This model is reliable and practicable. So a new approach to the rapid and lossless discrimination of varieties of acidophilous milk was put forward.

**Keywords** Near infrared spectral; Acidophilous milk; Principal component analysis (PCA); Artificial neural network; Discrimination

(Received Sep. 30, 2005; accepted Jan. 16, 2006)

\* Corresponding author