

应用可见-近红外光谱技术快速无损鉴别婴幼儿奶粉品种

黄敏¹, 何勇¹, 岑海燕¹, 胡兴越^{2*}

1. 浙江大学生物工程与食品科学学院, 浙江 杭州 310029

2. 浙江大学邵逸夫医院, 浙江 杭州 310016

摘要 为了快速无损鉴别婴幼儿奶粉品种, 提出了结合偏最小二乘(PLS)法和人工神经网络(ANN)综合预测婴幼儿奶粉品种的新方法。获取婴幼儿奶粉样本在400~1000 nm波段的漫反射光谱, 采取平均平滑法和多元散射校正(MSC)进行预处理, 用PLS建立校正模型进行模式特征分析及主成分的提取。经过交叉验证法判别, 提取7个主成分作为神经网络的输入变量, 奶粉的品种值作为输出, 建立了三层BP神经网络。9个典型品种的婴幼儿奶粉各取样本30个, 共计270个作为训练集。随机抽取的各个品种的10个样本, 共90个作为预测检验样本, 结果表明, 90个未知样本的品种预测准确率为100%。说明提出的方法具有很好的分类和鉴别作用, 为婴幼儿奶粉的品种快速无损鉴别提供了一种新方法。

关键词 近红外光谱; 偏最小二乘法; 婴幼儿奶粉; 品种; 人工神经网络

中图分类号: TH744.1; TS252.5 文献标识码: A 文章编号: 1000-0593(2007)05-0916-04

引言

婴幼儿奶粉是以牛乳(或羊乳)及其加工制品或谷物、豆类及其加工品为主要原料, 加入适量的维生素、矿物质和其他辅料, 加工制成的供婴幼儿食用的产品, 其营养成分能满足正常婴儿发育的营养需要。近年来, 我国婴儿奶粉市场十分混乱, 一些不法商人为牟取暴利制造“无营养”劣质婴儿奶粉充当优质奶粉出售。劣质奶粉存在的主要问题是: 蛋白质含量严重不足, 几乎没有任何营养, 长期食用会使婴儿缺乏身体发育所需的各种营养要素, 出现造血功能障碍、内脏功能衰竭、免疫力低下等症状, 一旦出现病变, 抢救难度很大。近年来, 湖南、河北、山东、安徽等地相继出现“大头娃娃”, 均系长期食用劣质婴幼儿配方奶粉所致。

奶粉的主要成分有蛋白质、脂肪酸、碳水化合物、维生素、矿物质、多种免疫物质和牛磺酸等。目前, 主要通过化学方法测定奶粉的化学成分来鉴别奶粉的品质^[1, 2]。但是测量过程繁琐、操作复杂, 不利于大范围推广。目前流行的辨别真假奶粉的传统方法主要有试手感、辨颜色、闻气味、尝味道、看溶解速度等。这些方法虽然操作简单、方便, 但是主观性强, 准确性低。如何用简便、快速、准确的现代分析技术和方法定量鉴别奶粉品种已成为维护食品安全中的一项

重要任务。

可见-近红外光谱法是一种快速简便非破坏性的分析方法。它被广泛应用于食品、石油、化工、医药等行业^[3]。国内外很多学者研究利用近红外光谱技术区别物质品种, 如苹果品种^[4]、咖啡品种^[5]、瓜类品种^[6]、大豆品种^[7]、道地山药^[8]、辣椒品种^[9]等。我们通过对不同品种婴幼儿奶粉的光谱分析, 寻找婴幼儿奶粉鉴别的快速有效方法。

人工神经网络是模仿人脑神经细胞的结构和功能的系统^[10]。目前应用较广的是一种基于误差反向传播算法的BP神经网络。偏最小二乘法^[11]是一种目前应用较广泛的光谱分析的数学方法。可以实现全谱或部分谱数据的分析。对于样本数量少, 变量多的情况更具有优越性。本研究采用可见和近红外光谱技术, 提出了一种基于BP神经网络与偏最小二乘法结合的模式识别方法, 对婴幼儿奶粉品种进行鉴别。

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 FieldSpec光谱仪, 其光谱采样间隔为1.5 nm, 分辨率为3.5 nm, 测定范围9302~30769 cm⁻¹, 探头视场角为20度。分析软件采用ASD View Spec Pro, Unscrambler

收稿日期: 2006-03-16, 修订日期: 2006-07-28

基金项目: “十一五”国家科技支撑计划项目(2006BAD10A04), 国家自然科学基金项目(30270773)和高等学校优秀青年教师教学科研奖励计划项目(02411)资助

作者简介: 黄敏, 1982年生, 浙江大学生物系统工程与食品科学学院博士研究生 e-mail: yhe@zju.edu.cn * 通讯联系人

V9.2 和 DPS^[12]。

1.2 样品来源及光谱的获取

从超市买来 9 种典型的婴幼儿奶粉(雀巢能多健、伊利、三鹿、国力、贝因美、多美滋、圣元、优博、雅培),各取 30 个样本,共计 270 个样本。9 个品种的婴幼儿奶粉均为 1 至 12 个月婴幼儿食用的。奶粉样本均用直径为 95 mm,高度 14 mm 的培养皿盛装。每个培养皿装满到 5 mm 高度的奶粉作为一个实验样本。全部实验样本随机分成建模集和预测集,建模集有 180 个样本(每个品种 20 个),预测集有 90 个样本(每个品种 10 个)。光谱仪经白板校准后进行测试。光谱仪置于奶粉样本的上方,距奶粉表面 120 mm,对每一个样本扫描 30 次。

1.3 光谱数据预处理

为了去除来自高频随机噪声、基线漂移、样本不均匀、光散射等影响,需要进行光谱预处理。采用平均平滑法,选用平滑窗口大小为 9,此时能很好滤除各种因素产生的高频噪声,再进行 MSC(multiplicative scatter correction)处理。由于光谱曲线在首端和末端有较大噪声,如图 1 所示,所以只取 400~1 000 nm 波段的光谱用于分析^[13,14]。

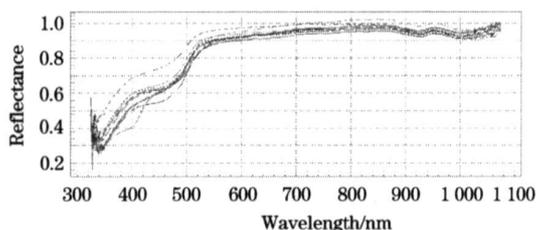


Fig 1 Near infrared reflectance spectra of nine different varieties of infant milk powder

1.4 偏最小二乘法(PLS)

偏最小二乘法(PLS)能够建立光谱数据和成分之间的相互关系,是一种广泛使用的近红外光谱数据处理方法。然而 PLS 是用于建立线性模型的,当出现非线性问题时模型的预测结果会受到影响。因此需要在 PLS 模型的基础上引入非线性的部分,人工神经网络是目前常用的非线性模型。

1.5 BP 神经网络模型

本研究建立了一个三层 BP 神经网络。所有的样本随机分成学习集和预测集数据,其中 180 个样本为学习集样本,90 个样本为预测集样本。网络输入层、隐含层、输出层节点数分别为 7, 8, 1。其中输入层的 7 个节点来自 PLS 分析得

到的主成分。最小训练速率为 0.9,设定训练迭代次数为 2 000 次,对输入样本进行标准化处理。

2 试验结果与分析

2.1 不同奶粉品种的近红外光谱图分析

9 种奶粉品种的典型近红外光谱如图 1 所示。图 1 中横坐标为波长范围 325~1 075 nm,纵坐标为光谱漫反射率。从图 1 中看出,不同品种奶粉的光谱曲线有明显区别,并具有一定的特征性和指纹性。选择波长范围在 400~1 000 nm 的光谱,应用 ASD View Spec Pro 软件,将反射率矩阵进行平均平滑法和 MSC 预处理,再用偏最小二乘法对其聚类。

2.2 模型参数的确定和 PLS 成分的提取

采用 PLS 分析方法,经交互验证法判断,最佳主成分数为 7,主要的模型参数见表 1。选用 7 个主成分模型具有最小残差(RV=0.248),最大相关系数($r=0.9830$),同时,它的校正集标准偏差也最小(SEC=0.4750)。分别用 1, 2, 3, 4, 5, 6, 7, 8, 9 分别代表 9 个不同品种的婴幼儿奶粉品种,将其作为 Y 变量。

Table 1 Main parameters in PLS model

PLS 主成分数	残差(RV)	相关系数/r	校正集标准偏差(SEC)
1	2.165	0.8256	1.4595
2	1.721	0.8642	1.3016
3	1.411	0.8906	1.1764
4	0.746	0.9432	0.8594
5	0.428	0.9700	0.6315
6	0.327	0.9768	0.5538
7	0.248	0.9830	0.4750

对原有变量进行线性组合提取新变量。PLS 成分是从原有自变量的样本数据矩阵 X 中提取的相互正交的成分,既保留尽量多与因变量的相关性,又保留了较多的方差,从而在消除原有自变量共线性的同时,使建立的回归模型仍能充分地反映出自变量与因变量之间的相互关系。从表 2 可以看出, Y 变量与 PC9 的相关性最大,而贡献率 PC1 最大。前 7 个主成分的贡献率达到 99.69%,而 Y 变量与 PC7 的相关性超过 98%。第 8 和第 9 个主成分与 Y 变量的相关性只比第 7 个主成分提高 0.0015 和 0.0042,而且这 2 个主成分的贡献率很低,说明 PLS 提取的前 7 个主成分已经足够解释原始光谱数据所包含的大部分信息。

Table 2 Regression coefficient and contribution rate of principal components

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Y(r^2)	0.8256	0.8642	0.8906	0.9432	0.9700	0.9768	0.9830	0.9845	0.9872
贡献率	47.65%	22.71%	25.64%	1.78%	0.91%	0.78%	0.22%	0.10%	0.03%

2.3 基于特征变量建立 BP 神经网络模型

如果将全部的光谱波段作为神经网络的输入,大大加大了神经网络的计算量。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一定的相关性。而 PLS 推荐的前面 7 个主成分已经包含了大部分光谱信息。因此,将这 7 个特征变量作为 BP 神经网络的输入变量。设置网络输入层、

隐含层、输出层节点数分别为 7, 8, 1。训练集和预测集样本数为 180 和 90。对 90 个未知样本的预测结果见表 3。表 3 中 1 至 10 号、11 至 20 号、21 至 30 号、31 至 40 号、41 至 50 号、51 至 60 号、61 至 70 号、71 至 80 号和 81 至 90 号样本

的真实值依次为 1, 2, 3, 4, 5, 6, 7, 8, 9。依次代表雀巢能多健、伊利、三鹿、国力、贝因美、多美滋、圣元、优博、雅培这 9 个品种。拟合残差为 0 000 097, 90 个样本的预测相对偏差均在 5% 以下, 即预测的正确率达到了 100%。

Table 3 Prediction results for unknown samples by BP model

样本	预测值	真值	样本	预测值	真值	样本	预测值	真值
1	1 014 90	1	31	4 190 46	4	61	6 70 084	7
2	1 012 93	1	32	4 190 92	4	62	6 978 22	7
3	1 008 00	1	33	4 026 14	4	63	6 656 55	7
4	1 028 79	1	34	4 179 56	4	64	6 679 94	7
5	1 008 03	1	35	3 938 92	4	65	6 681 34	7
6	1 019 86	1	36	4 024 57	4	66	6 658 64	7
7	1 040 71	1	37	4 110 98	4	67	6 715 86	7
8	1 037 28	1	38	4 028 49	4	68	6 693 13	7
9	1 031 88	1	39	4 068 95	4	69	6 895 68	7
10	1 037 93	1	40	4 107 58	4	70	6 684 01	7
11	2 011 29	2	41	4 929 76	5	71	7 952 71	8
12	2 091 67	2	42	5 096 11	5	72	8 021 33	8
13	2 013 97	2	43	5 068 78	5	73	8 023 65	8
14	2 083 97	2	44	5 140 03	5	74	8 038 54	8
15	2 092 12	2	45	5 141 37	5	75	8 075 77	8
16	2 080 01	2	46	4 915 24	5	76	8 099 07	8
17	2 051 58	2	47	5 141 62	5	77	8 060 39	8
18	2 094 24	2	48	4 894 18	5	78	8 030 91	8
19	2 094 57	2	49	5 011 96	5	79	7 622 79	8
20	2 017 06	2	50	5 172 22	5	80	8 108 07	8
21	2 948 50	3	51	6 100 84	6	81	8 871 04	9
22	3 026 62	3	52	6 138 22	6	82	8 889 90	9
23	3 003 20	3	53	6 156 55	6	83	8 889 90	9
24	2 862 50	3	54	6 179 94	6	84	8 937 52	9
25	2 892 50	3	55	6 281 34	6	85	8 812 43	9
26	2 855 84	3	56	6 238 64	6	86	8 915 84	9
27	2 932 67	3	57	6 215 86	6	87	8 868 89	9
28	2 942 67	3	58	6 113 13	6	88	8 925 94	9
29	2 912 15	3	59	6 195 68	6	89	8 889 14	9
30	2 944 24	3	60	6 214 01	6	90	8 948 64	9

3 结 论

应用可见-近红外光谱技术建立了基于偏最小二乘法和人工神经网络的婴幼儿奶粉品种鉴别模型。该模型的预测效果好, 鉴别速度快, 对未知样品的预测相对偏差均在 5% 以下, 识别率达到 100%。说明运用可见-近红外光谱技术可以快速、准确的对婴幼儿奶粉品种进行鉴别。在模式识别

上, 采用偏最小二乘法和 BP 神经网络相结合的方法, 优化了常规的模式识别方法, 大大提高了识别的精确度。将从 PLS 分析中得到的特征变量作为 BP 神经网络的输入, 不但减少了神经网络的计算量, 加快了训练速率, 而且去除了干扰信息, 也提高了预测的正确率。因此, 用偏最小二乘法结合 BP 神经网络的模式识别和光谱技术研究婴幼儿奶粉品种鉴别是可行的。

参 考 文 献

- [1] HOU Dong-yan, HUI Rui-hua, LI Tie chun, et al(侯冬岩, 回瑞华, 李铁纯, 等). Journal of Anshan Normal University(鞍山师范学院学报), 2004, 6(4): 37.
- [2] GU Ri-xu, ZHANG Jian-guo(谷日旭, 张建国). Shanxi Journal of Preventive Medicine(山西预防医学), 1997, 6(3): 198.
- [3] LE Jun-ming, CHEN Ying, DING Ying(乐俊明, 陈 鹰, 丁 映). Guizhou Agricultural Sciences(贵州农业科学), 2005, 33(3): 62.
- [4] He Yong, Li Xiao-li, Shao Yong-ni. Lecture Notes in Artificial Intelligence, 2005, 3809: 1053.
- [5] Esteban Diez I, Gonzalez-Saiz J M, Pizarro C. Analytica Chimica Acta, 2004, 514: 57.
- [6] Zsolt Seregely, Tamas Deak, Gyorgy Denes Bisztray. Chemometrics and Intelligent Laboratory Systems, 2004, 72: 195.
- [7] Turza S, Toth A, Varadi M. Near Infrared Spectroscopy: Proceedings of the 8th International Conference, Chichester, UK: NIR Publications, 1998. 183.
- [8] SUN Sh-qin, TANG Jun-ming, YUAN Zi-min, et al(孙素琴, 汤俊明, 袁子民, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(2): 258.
- [9] YU Wei, YONG Ke-lan(俞 蔚, 雍克岚). Food Science(食品科学), 2003, 24(11): 105.
- [10] BAO Yi-dan, WU Yan-ping, HE Yong(鲍一丹, 吴燕萍, 何 勇). Journal of Agricultural Mechanization Research(农机化研究), 2004, 3: 162.
- [11] He Yong, Song Hai-yan, Pereira A G, et al. Lecture Notes in Computer Science, 2005, 3644: 859.
- [12] 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), 2002.
- [13] QI Xiao-ming, ZHANG Lu-da, DU Xiao-lin, et al(齐小明, 张录达, 杜晓林, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(5): 870.
- [14] YIN Qiu, SHU Xiao-zhou, XU Zhao-an, et al(尹 球, 疏小舟, 徐兆安, 等). Journal of Infrared and Millimeter Waves(红外与毫米波学报), 2004, 23(6): 427.

Fast Discrimination of Varieties of Infant Milk Powder Using Near Infrared Spectra

HUANG Min¹, HE Yong¹, CEN Hai-yan¹, HU Xing-yue^{2*}

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

2. Sir Rum Run Shaw Hospital, Zhejiang University, Hangzhou 310016, China

Abstract A new method for discrimination of varieties of infant milk powder by means of visible/near infrared spectroscopy (Vis/NIRS) (325-1 075 nm) was developed. Partial least square (PLS) was used to analyze the characteristics of the pattern. PLS compressed thousands of spectral data into a small quantity of principal components and described the body of spectra. The first seven principal components were confirmed as the best number of principal components. Then, these seven principal components were applied as the input to a back propagation neural network with one hidden layer. The infant milk powder varieties data were applied as the output of BP neural network. One hundred eighty samples containing nine typical varieties of infant milk powder were selected randomly, and they were used as a training set of the BP neural network model, and the remainder samples (total 90 samples) formed the prediction set. With a proper network training parameter, the recognition accuracy of 100% was achieved. This model is reliable and practicable. So the present paper could offer a new approach to the fast discrimination of varieties of infant milk powder.

Keywords Near infrared spectrum; Partial least square; Infant milk powder; Varieties; BP neural network

(Received Mar. 16, 2006; accepted Jul. 28, 2006)

* Corresponding author