

积分光谱与近红外光谱分析精度关系的研究

耿响 陈斌 朱文静 颜辉

(江苏大学食品与生物工程学院, 江苏 镇江 212013)

摘要 以 MPA 型傅里叶近红外光谱仪和 WQF-400N 型傅里叶近红外光谱仪上检测的奶粉光谱为原始光谱, 利用积分法选择不同的积分区间对原始光谱进行积分, 模拟不同分辨率下的光谱。采用相同的预处理方法——Savitzky-Golay 平滑和一阶微分, 利用偏最小二乘(PLS)回归对奶粉中的脂肪含量建立模型。分别比较两台仪器所建模型的相关系数(R), 预测均方差(f_{RMSEP}), 平均相对误差(E)三个预测参数, 得到两台仪器分别在分辨率为 16 cm^{-1} 和 64 cm^{-1} 时建模效果最好。研究结果表明, 利用积分光谱, 能起到压缩数据, 提高模型预测精度的作用; 同时说明对于具体样品的特定指标要选择合适的分辨率, 在减少工作量的同时获得最佳的定量分析结果, 而单纯的追求高分辨率是没有意义的。

关键词 光谱学; 近红外光谱; 分辨率; 积分光谱; 积分区间; 脂肪

中图分类号 O657.3 **文献标识码** A **doi**: 10.3788/AOS20092912.3551

Study on the Relation Between Integration Spectrum and Accuracy of Near Infrared Spectroscopic Analysis

Geng Xiang Chen Bin Zhu Wenjing Yan Hui

(School of Food and Biological Engineering, Jiangsu University, Zhenjiang, Jiangsu 212013, China)

Abstract The raw spectra of milk powder were detected by means of two Fourier near-infrared spectrometers MPA type and WQF-400N type. Primitive spectra were integrated in different integral intervals to simulate spectra with different resolutions. After being pre-treated with the same methods (Savitzky-Golay smoothing and the first derivative), models for fat content in milk powder were established through partial least square regression (PLS). Comparisons on correlation coefficient R , root mean square error of prediction f_{RMSEP} and mean bias E between the models reveal that MPA type and WQF-400N type spectrometers have the best models at the resolutions of 16 cm^{-1} and 64 cm^{-1} respectively. These results show that integral spectrum could compress data and enhance the model's prediction accuracy, and that an adequate resolution should be selected for special characters of different samples. Both of them ensure the best quantitative analytical result while reducing the work load. It is of no significance to simply pursue higher and higher resolution.

Key words spectroscopy; near-infrared spectroscopy; resolution; integral spectrum; integral interval; fat

1 引 言

脂肪是奶粉的重要营养成分之一^[1], 是国标规定的强制检查项目。但常规分析方法用时较长, 人为误差较大, 设备的维护、使用费用较高, 使得质检

部门无法快速将劣质奶粉从市场上剔除, 导致了如安徽阜阳“大头娃娃”事件、三鹿的“结石婴儿”事件的发生。近红外光谱分析是一种快速、无损、高效的分析方法, 广泛应用于农业、医药、石化、烟草和食品

收稿日期: 2008-12-11; **收到修改稿日期**: 2009-03-23

基金项目: 镇江市社会发展项目(SH2007075)资助课题。

作者简介: 耿响(1981—), 女, 博士研究生, 主要从事近红外光谱分析和农产品无损检测技术等方面的研究。

E-mail: gengxiang2005@sina.com

导师简介: 陈斌(1960—), 男, 博士, 教授, 主要从事近红外光谱分析和农产品无损检测技术等方面的研究。

E-mail: ncp@ujs.edu.cn(通信联系人)

等领域^[2~4]。然而近红外光谱仪的工作状态和扫描时参数的设定对测量过程和分析结果均有一定的影响。已有学者研究过近红外光谱仪器的分辨率对近红外分析的影响^[5~7],得出分辨率对定量模型的预测能力有较大影响的结论。对不同的样品,采集光谱时所用的分辨率越高,所建立的定量模型的效果不一定越好,对于具体样品的特定成分,需要选择合适的分辨率,进而获得最佳的定量分析结果。

已有的研究分辨率对模型预测精度影响的文献都是选择不同的分辨率对某一类样品进行扫描,得到不同分辨率下的光谱,通过建立模型进行比较。本文给出了另一种比较分辨率对模型精度影响的方法—积分光谱法,即对原始光谱,采用不同的积分区间模拟不同分辨率下的光谱,预处理后建立模型,该方法简单易行,同时也起到对光谱平滑的作用。

2 试验材料与方法

2.1 试验样品组成

样品是从市场上采购的伊利、三鹿、雅士利、雀巢、光明等不同品牌的不同类型的奶粉共 80 种。利用罗兹-哥特里法(GB5413-85)测量每个样品的脂肪含量^[8,9],为了保证化学值的准确,采用 3 次平行检测。数据建模时采用马氏距离法^[10~13],剔除 1 个异常样品,随机选择 24 个样品为预测集,样品中脂肪含量分布如表 1 所示。

表 1 样品中脂肪含量分布
Table 1 Composition of fat in milk samples

Instruments	Content / %	Mean	Standard error
		content / %	
MPA	Calibration set	29.26~8.99	17.23 3.97
	Prediction set	20.96~10.41	16.68 2.38
WQF-400N	Calibration set	29.26~7.87	16.87 4.12
	Prediction set	22.24~11.42	17.15 2.41

2.2 光谱采集

试验所用仪器一台是 MPA 型(德国布鲁克公司)多功能傅里叶变换近红外光谱仪,窗口 $\Phi=2$ cm,带有 RT-PbS 探测器、OPUS QUANT-6 定量分析软件,并附有内置镀金漫反射积分球、样品旋转器和石英样品杯等配件;光谱范围 12800~4000 cm^{-1} ,分辨率 4 cm^{-1} ,以仪器内置镀金为背景。另一台仪器为 WQF-400N 型傅里叶变换近红外光谱仪(北京瑞利公司),光谱范围 10000~3500 cm^{-1} ;分辨率 4 cm^{-1} ;硫化铅(PbS)探测器,使用漫反射附件,以空气为本底,

扫描次数 32 次,每个样品平行测定 3 次取平均值。两台仪器进行光谱采集的试验温度 20 $^{\circ}\text{C}$,相对湿度 45%。图 1 给出了部分样品的光谱图。

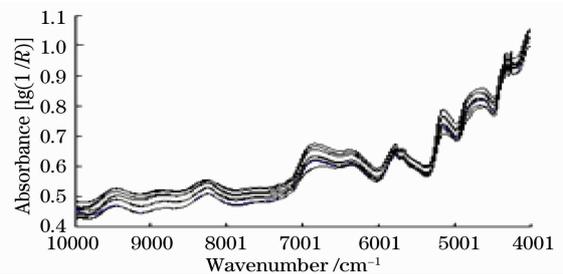


图 1 部分样品的原始光谱

Fig. 1 Original spectrum of some samples

2.3 分辨率的定义

光谱分辨率是指仪器区分两个相邻吸收峰的能力,通常用带宽来表征。高的光谱分辨率虽然给出较为丰富的光谱信息,但会导致收集光谱的时间延长,同时也增加了光谱的噪声。在实际应用中,近红外光谱仪一般不追求高分辨率,通常 16 cm^{-1} 或 10 nm(在 2500nm 处)的分辨率就可以满足绝大多数分析对象的应用要求^[14]。所以针对不同的样品应该选择不同的分辨率以达到较好的建模效果。

3 数据处理

3.1 积分光谱

为比较积分光谱法的效果,两台仪器均采用 10000~4000 cm^{-1} 范围的光谱数据。首先将原始光谱进行归一化,再以积分区间为 2, 4, 8, 16, 32 进行积分,得到分辨率为 8, 16, 32, 64, 128 cm^{-1} 的模拟光谱。

积分光谱所用的方法是数值积分中最常用的梯形法^[15],即将积分区间按采样间隔 $\Delta\bar{\nu}$ 等分为 n 个小区,则令积分区间中中间位置的波数点对应的吸光度为

$$\Delta\bar{\nu} \left[(a_0 + a_n) / 2 + \sum_{i=1}^{n-1} a_i \right]$$

式中 a_i 是 ($i = 0, 1, 2, \dots, n$) 积分区间中第 i 个波数点对应的吸光度值。此公式相当于利用插入点 i 将所选定区间进行等间隔划分,将划分后的小区间视为以 $\Delta\bar{\nu}$ 为底, a_i 为高的小长方形,求得每一个小长方形的面积再求和。

3.2 光谱数据处理

使用 Matlab7.0 软件和自行编制的 NIRSA2.2 软件对两台仪器的试验数据先进行归一化再积分、

然后对积分后的光谱进行预处理并建模。

预处理方法:一阶微分,窗口宽度为 7;五点多项式平滑即五点 Savitzky-Golay 滤波处理。

建模方法:偏最小二乘回归(PLS)^[16]。

模型评价指标:校正与预测相关系数(R_r/R)、回判和预测均方差(f_{RMSEC}/f_{RMSEP})、校正与预测的平均相对误差(E_r/E)

4 结果分析

4.1 MPA 型仪器的数据建模及结果分析

从表 2 中可以看出,随着分辨率的降低,波数点(N)减少,除分辨率为 128 cm^{-1} 时,其他分辨率下的主成分数变化不大。虽然分辨率为 4 cm^{-1} 时的校正相关系数较高,回判平均相对误差和回判均方

表 2 MPA 型傅里叶近红外光谱仪分辨率对模型预测精度的影响

Table 2 Influence of MPA Fourier near-infrared spectrometer's resolution on the prediction accuracy of models

Resolution / cm^{-1}	N	Factor	R_r/R	E_r/E	f_{RMSEC}/f_{RMSEP}
4	1555	4	0.975 8/0.796 5	4.395 2/7.519 7	0.860 6/1.455 8
8	777	3	0.920 2/0.860 5	8.220 7/6.522 5	1.533 8/1.247 0
16	388	5	0.936 9/0.887 9	8.845 5/5.701 9	1.375 1/1.083 5
32	194	4	0.902 7/0.887 6	8.646 5/5.868 3	1.692 3/1.085 2
64	97	4	0.888 3/0.884 7	9.121 6/5.968 6	1.806 0/1.088 7
128	48	10	0.937 2/0.881 6	6.526 1/6.251 1	1.369 5/1.240 8

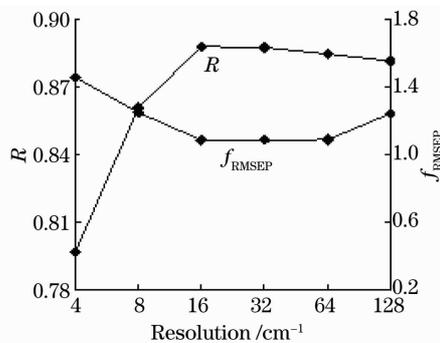


图 2 R 和 f_{RMSEP} 与 MPA 型仪器分辨率的关系

Fig. 2 R and f_{RMSEP} change tendency versus different resolutions of MPA spectrometer

2) R , f_{RMSEP} 和 E_r 这三个参数说明,虽然分辨率越高,光谱数据所包含的有效信息越多,但同时含有的随机噪声产生的影响也越大,只有较好的除噪后建立的模型效果才会比较好;若能够先对采集的光谱数据进行积分,减少数据量的同时也对光谱进行了平滑,除掉了一部分随机噪声,在建立模型前只需简单的预处理就可以得到较高的预测精度。

4.2 WQF-400N 型仪器的数据建模及结果分析

由表 3 可以看出,主成分数(factor)除分辨率为 $4, 8\text{ cm}^{-1}$ 时波动较大,其他分辨率时的主成分数变

差都较小,但是 R_r 与 R , f_{RMSEC} 与 f_{RMSEP} 均相差较大,说明此模型的稳定性不好。预测集的平均相对误差随着分辨率的降低呈先减小后增大的变化趋势,在分辨率为 16 cm^{-1} 时达到最小。 R , f_{RMSEP} 两个参数在不同分辨率下的变化趋势在图 2 中给出。可得出如下结论:

1) 图 2 可以看出随着分辨率逐渐降低,相关系数 R 先增大后减小,分辨率从 16 cm^{-1} 到 128 cm^{-1} 变化过程中, R 变化缓慢,分辨率为 16 cm^{-1} 时的 R 最大; f_{RMSEP} 当分辨率为 $16, 32$ 和 64 cm^{-1} 时 f_{RMSEP} 的值相差不大,在分辨率为 16 cm^{-1} 时达到最小。 R 和 f_{RMSEP} 的变化趋势表明分辨率为 16 cm^{-1} 时模型的预测精度最好,与表 2 中给出的预测平均相对误差最小值所对应的分辨率是一致的。

化不大;预测平均相对误差的变化趋势是随着分辨率的降低先增加,再减小最后增大,当分辨率为 128 cm^{-1} 达到最大值 1.5290 ,最小值是 64 cm^{-1} 时的 0.9365 ;虽然 4 cm^{-1} 时的校正平均相对误差和回判均方差都是最小,但是其预测相关系数不高,且与回判相关系数相差较大,说明此模型稳定性不好。模型预测参数 R , f_{RMSEP} 的变化趋势如图 3 所示。

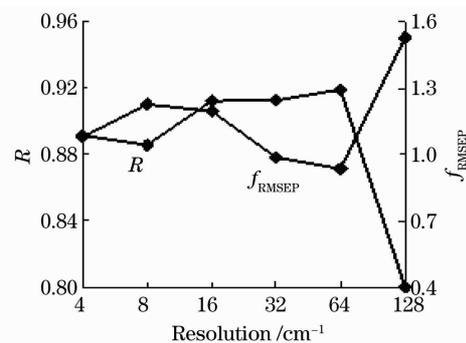


图 3 R 和 f_{RMSEP} 与 WQF-400N 型仪器分辨率的关系

Fig. 3 R and f_{RMSEP} change tendency versus different resolutions of WQF-400N spectrometer

从图 3 看出 R 在分辨率为 $16, 32, 64\text{ cm}^{-1}$ 时变化缓慢,分辨率为 64 cm^{-1} 时达到最大为 0.9190 , 128 cm^{-1} 时最小为 0.8000 ; f_{RMSEP} 的变化趋势比较

表 3 WQF-400N 型傅里叶近红外光谱仪分辨率对模型预测精度的影响

Table 3 Influence of WQF-400N Fourier near-infrared spectrometer's resolution on the prediction accuracy of models

Resolution / cm^{-1}	N	Factor	R_r/R	E_r/E	$f_{\text{RMSEC}}/f_{\text{RMSEP}}$
4	1555	4	0.968 2/0.891 7	5.118 4/5.383 1	0.984 2/1.079 1
8	777	10	0.993 4/0.885 9	2.189 5/5.898 9	0.450 8/1.227 1
16	388	8	0.950 0/0.912 5	6.355 6/5.252 2	1.228 4/1.197 2
32	194	6	0.909 0/0.912 8	8.834 3/4.817 6	1.639 6/0.987 4
64	97	6	0.903 1/0.919 0	8.919 5/4.647 5	1.689 5/0.936 5
128	48	6	0.881 3/0.800 0	9.148 4/7.455 0	1.859 6/1.529 0

明显,当分辨率为 64 cm^{-1} 时达到最小值 0.9365, 128 cm^{-1} 时最大为 1.5290。

另外,从 R , f_{RMSEP} 及 E_r 与分辨率的变化趋势看到,分辨率为 64 cm^{-1} 时所建模型的预测精度最好;对光谱进行不同区间积分起到了对光谱进行不同窗口宽度平滑的效果,对具体样品要选取适当的积分区间,积分区间过大会使得光谱发生畸变,从而建模效果不好。

5 结 论

通过对两台仪器测得的奶粉的近红外光谱数据进行不同积分区间的积分,得到了不同分辨率的模拟光谱,从采用相同的预处理和相同的建模方法得到的不同分辨率下的预测精度比较可以得到以下结论:

1) 分辨率对模型的预测精度有很大影响。但并不是分辨率越高所建模型就越好,针对不同的仪器,不同的检测样品,有一个最佳分辨率存在。本文比较了进口 MPA 型仪器和国产 WQF-400N 型仪器,试验结果表明,对于奶粉脂肪所建模型,两台仪器的最优分辨率相差较大分别为 16 cm^{-1} 和 64 cm^{-1} ,原因可能是两者仪器稳定性的差异:MPA 型仪器自稳定性较好,外界环境的较小变化对其影响不大;而 WQF 型仪器自稳定性较差,抗外界干扰能力较差。

2) 采用的积分光谱法,可以作为一种光谱的预处理方法。该方法既能起到压缩数据,减少工作量,也具有对原光谱平滑的作用,因此可以提高校正模型的预测建模效果。

3) 在选择仪器时高分辨率不是必要条件。过高的分辨率会使数据量增大,对数据的预处理要求提高、处理繁琐。在选择仪器时应根据自身需要,不能一味的追求高分辨率。

参 考 文 献

1 Chang Min, Chu Pengjiao, Xu Kexin. Study on noninvasive

detection using NIR diffuse reflectance spectrum for monitoring protein content in milk powder[J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(1): 43~45

常 敏, 褚鹏蛟, 徐可欣. 近红外漫反射光谱无损检测乳粉蛋白质的研究[J]. *光谱学与光谱分析*, 2007, **27**(1): 43~45

2 Zhao Jiewen, Zhang Haidong, Liu Muhua. Preprocessing methods of near-infrared spectra for simplifying prediction model of sugar content of apples[J]. *Acta Optica Sinica*, 2006, **26**(1): 136~140

赵杰文, 张海东, 刘木华. 简化苹果糖度预测模型近红外光谱预处理方法[J]. *光学学报*, 2006, **26**(1): 136~140

3 J. M. Amigo, J. Cruz, M. Bautista *et al.*. Study of pharmaceutical samples by NIR chemical-image and multivariate analysis [J]. *Trac-Trend. Anal. Chem.*, 2008, **27**(8): 696~713

4 Han Donghai, Wang Jiahua. Review of non-destructive measurement of fruit quality by means of near infrared spectroscopy[J]. *Chinese J. Lasers*, 2008, **35**(8): 1123~1131

韩东海, 王加华. 水果内部品质近红外光谱无损检测研究进展[J]. *中国激光*, 2008, **35**(8): 1123~1131

5 Duan Yanqing, Yang Tao, Kong Xiangyong *et al.*. Effects of sample granularity and spectral resolution on tobacco nicotine NIR predictive model [J]. *J. Yunnan University (Natural Sciences Edition)*, 2006, **28**(4): 340~344

段焰青, 杨 涛, 孔祥勇等. 样品粒度和光谱分辨率对烟草烟碱 NIR 预测模型的影响[J]. *云南大学学报(自然科学版)*, 2006, **28**(4): 340~344

6 Wang Yibing, Wang Hongyu, Zhai Hongju *et al.*. Influence of near infrared spectral resolution on the quantitative analysis[J]. *Chinese J. Anal. Chem.*, 2006, **34**(5): 699~701

王一兵, 王红宇, 翟宏菊等. 近红外光谱分辨率对定量分析的影响[J]. *分析化学研究简报*, 2006, **34**(5): 699~701

7 Xie Lijuan, Liu Donghong, Zhan Yuhuan *et al.*. Study on the influence of resolution on near infrared spectra and quantitative analysis[J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(8): 1489~1492

谢丽娟, 刘东红, 张宇环等. 分辨率对近红外光谱和定量分析的影响研究[J]. *光谱学与光谱分析*, 2007, **27**(8): 1489~1492

8 The State Bureau of Quality and Technical Supervision, GB/T5413.1-5413.32, The Universal Method of Infant Formulas and Milk Powder[S]. Chinese Standard Press, 1997. 1~9

中华人民共和国国家技术监督局, GB/T 5413.1-5413.32, 1997 婴幼儿配方食品和乳粉通用检验方法[S]. 中国标准出版社, 1997. 1~9

9 Wang Yun, Xu Kexin, Chang Min. Study on multi-band NIR spectroscopy for the determination of fat and protein contents in milk[J]. *Optical Instruments*, 2006, **28**(3): 3~7

王 云, 徐可欣, 常 敏. 近红外光谱技术检测牛奶中脂肪及蛋白质含量校正模型的建立[J]. *光学仪器*, 2006, **28**(3): 3~7

10 Liu Qiang, Luo Changbing, Chen Shaojiang *et al.*. An outlier diagnosis on near infrared spectroscopy analysis of NDF content in corn silage feeds[J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(8): 1514~1518

- 刘 强, 罗长兵, 陈绍江 等. 近红外光谱分析青贮玉米 NDF 中判别异常光谱的研究[J]. *光谱学与光谱分析*, 2007, **27**(8): 1514~1518
- 11 D. Ververidis, C. Kotropoulos. Gaussian mixture modeling by exploiting the mahalanobis distance [J]. *IEEE T. Signal Process.*, 2008, **56**(7): 2797~2811
- 12 Zhang Xin, Guo Jia, Ni Lijun . Recognition of plant parts of tobaccos based on infrared and near infrared spectra [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(12): 2437~2440
- 张 鑫, 郭 佳, 倪力军. 基于红外与近红外光谱的烟叶部位识别[J]. *光谱学与光谱分析*, 2007, **27**(12): 2437~2440
- 13 Chen Quansheng, Zhao Jiewen, Zhang Haidong *et al.*. Identification of authenticity of tea with near infrared spectroscopy based on support vector machine[J] . *Acta Optica Sinica*, 2006, **26**(6): 933~937
- 陈全胜, 赵杰文, 张海东 等. 基于支持向量机的近红外光谱鉴别茶叶的真伪[J]. *光学学报*, 2006, **26**(6): 933~937
- 14 Yan Yanlu, Zhao Longlian, Han Donghai *et al.*. The Basis and Application of the Analysis Technique for Near Infrared Spectra [M]. Beijing : Chinese Light Industry Press, 2005. 121~137
- 严衍禄, 赵龙莲, 韩东海 等. 近红外光谱分析基础与应用[M]. 北京:中国轻工业出版社, 2005. 121~137
- 15 Wu Jinguang. The Technology And Application of Modern Fourier Transform Infrared Spectroscopy [M]. Beijing: Scientific and Technical Document Press, 1994. 129~130
- 吴瑾光. 近代傅里叶变换红外光谱技术及应用[M]. 北京:科学技术文献出版社, 1994. 129~130
- 16 A. D. Burns, W. E. Ciurzak. Handbook of Near Infrared Analysis[M]. New York; CRC Press, 1992. 189~203