

文章编号: 1001 - 9014 (2010) 02 - 0128 - 05

近红外光谱结合小波变换-径向基神经网络 用于奶粉蛋白质与脂肪含量的测定

单 杨^{1,2}, 朱向荣², 许青松³, 梁逸曾¹

(1. 中南大学 化学化工学院, 湖南 长沙 410083; 2. 湖南省食品测试分析中心, 湖南 长沙 410125;
3. 中南大学 数学科学与计算技术学院, 湖南 长沙 410083)

摘要:应用近红外光谱分析技术结合化学计量学方法,建立了奶粉脂肪和蛋白质含量测定的化学计量学建模新方法.首先采用 Kernal-Stone法对校正集样本和预测集样本进行分类,然后利用小波变换滤波技术对样品的近红外光谱进行压缩去噪处理,结合滤波后重构光谱信号建立脂肪和蛋白质的径向基神经网络回归模型,并分别对径向基网络的扩散常数 σ 值及小波变换中的小波基与压缩尺度三个参数进行了详细的讨论.脂肪模型在小波基为 db2 及小波尺度为 4, σ 值为 3.5 时的预测模型精度最好;蛋白质模型在小波基为 db8 及小波尺度为 4, σ 值为 6 时,预测模型精度最好.所建立的脂肪和蛋白质校正模型的预测集相关系数 (R_p) 分别为 0.990 和 0.994, 预测均方根误差分别为 0.007 与 0.004. 预测结果表明, RBF 网络结合小波变换进行建模预测,模型简单、稳健且精度较好,该方法适合奶粉脂肪和蛋白质含量的快速、无损测定.

关键词:近红外光谱;奶粉;径向基神经网络;小波变换;Kernal-Stone法
中图分类号: O657.33 **文献标识码:** A

DETERM INING THE CONTENTS OF FAT AND PROTEIN IN M LK POWDER BY USING NEAR INFRARED SPECTROSCOPY COM BINED W ITH W AVELET TRANSFORM AND RAD ICAL BASIS FUNCTON NEURAL NETWORKS

SHAN Yang^{1,2}, ZHU Xiang-Rong², XU Qing-Song³, LIANG Yi-Zeng¹

(1. College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China;
2. Hunan food test and analysis center, Changsha 410125, China;
3. School of Mathematical Science and Computing Technoby, Central South University, Changsha 410083, China)

Abstract: A new chemometric method for determining the contents of fat and protein in milk powder was established by using near infrared (NIR) spectroscopy combined with chemometric methods. The calibration and prediction sets were partitioned by Kernal-Stone algorithm. Wavelet transform (WT) was used for de-noising and compressing signals. The radical basis function neural networks (RBFNN) model for the contents of fat and protein was built by combining with the reconstruction spectral signal. Three parameters, i.e., the spread value of RBF network, the wavelet functions, and decomposition levels were discussed in detail. The results show that the precision of the prediction model is the best when wavelet function, compression level and spread value are db2, 4, and 3.5 for fat. In the same way, the precision is the best when wavelet function, compression level and spread value are db8, 4, and 6 for protein. Correlation coefficients (R_p) of prediction set for the correction model of fat and protein are 0.990 and 0.994, and root mean square error prediction (RMSEP) is 0.007 or 0.004, respectively. The results also show that the model is easy and robust, and the prediction accuracy is improved by using RBFNN combined with WT for building NIR models. This method is suitable for determining the content of fat and protein in milk powder rapidly and nondestructively.

Key words: near infrared (NIR) spectroscopy; milk powder; radical basis function neural networks; wavelet transformation; Kernal-Stone algorithm

收稿日期: 2009 - 03 - 20, 修回日期: 2009 - 10 - 10

Received date: 2009 - 03 - 20, revised date: 2009 - 10 - 10

基金项目: 国家自然科学基金 (20875104, 10771217) 资助项目, 2008 年湖南省标准化战略资助项目

作者简介: 单 杨 (1963-), 男, 湖南攸县人, 研究员, 主要从事食品安全分析技术与农产品加工领域研究.

引言

奶粉是人们日常生活中不可缺少的乳制品,它是由鲜奶加工而成,保留了鲜奶的主要营养成分。脂肪和蛋白质是评价奶粉产品质量的重要指标。哥特里-罗紫法是目前测定奶制品脂肪含量,使用范围较广、稳定性较好的方法。但此方法需要将样品水解,加入有机溶剂溶解与萃取,操作比较复杂,振荡程度和时间不易掌握,在提取过程中还容易出现乳化现象,造成检测结果偏低。现有检测奶粉蛋白质采用凯氏定氮法。该方法应用范围广、灵敏度高、回收率好且仪器价钱适中,但也存在操作流程繁琐、耗时费力等缺点,且在操作过程中产生大量有害气体污染环境,影响操作人员的健康^[1]。

近红外光谱法用于测定奶粉或牛奶中的指标性成分已有报道^[2-5]。但大多需要采用光谱预处理手段进行预处理,然后采用变量筛选方法筛选变量再建模,步骤比较繁琐。本文采用小波变换直接进行数据压缩与去噪,通过径向基神经网络(radical basis function neural networks, RBFNN)建立奶粉脂肪和蛋白质含量预测模型。结果表明,本方法中的数学模型简单、稳健、可靠,与常用的BP人工神经网络(back propagation neural networks, BPNN)模型及偏最小二乘法(partial least squares, PLS)相比,预测精度更好。

1 材料与方法

1.1 样品来源

研究的奶粉样本来自伊利、蒙牛、雅士利、南山、雀巢和惠氏等11个奶粉厂家,包括婴儿配方奶粉、儿童奶粉、青少年奶粉、中老年奶粉、学生营养奶粉和女士营养奶粉等29类产品。样本数共计100个。

1.2 蛋白质与脂肪的化学测定

脂肪的测定参照哥特里-罗紫法(GB/T 5009.46—2003),在重复性条件下获得的两次独立测定结果的绝对差值不得超过算术平均值的5%。蛋白质的测定参照凯氏定氮法(GB/T 5413.1—1997),根据乳制品中将氮元素含量转化为蛋白质含量的换算系数为6.38,求出所购奶粉中蛋白质含量,在重复性条件下获得的两次独立测定结果的绝对差值不得超过算术平均值的1.5%。以上实验中,每个样品平行测定三次。

1.3 近红外光谱采集

Antaris 傅里叶变换近红外光谱分析仪(美国 Thermo Nicolet 公司),采用积分球和旋转台测定样品 NIR 漫反射光谱。仪器参数设定如下:扫描范围

10000~4000 cm^{-1} ,分辨率 8 cm^{-1} ,扫描次数 64次,每个样品平行测定 3次,取平均光谱。

1.4 方法和原理

1.4.1 径向基神经网络

径向基(RBF)网络是 Powell M. J. D 在 1985 年提出的,以函数逼近理论为基础构造的一类前向网络^[6]。由于它具有简单的结构、快速的训练过程及与初始权值无关的优良特性,在分类、函数逼近和故障诊断等领域有着比较多的应用^[7-9]。

1.4.2 小波变换

小波变换^[10](wavelet transformation, WT)是近几年兴起的一种信号处理方法,可用作分析数据压缩和提取有用信息的工具。在目前的研究中,db族小波基在小波中应用最广泛,具有分析近红外光谱这类平滑信号的特性。其他小波基 symmlet族和 coiflet族等也常被使用。

1.5 数据处理及软件

小波变换、径向基神经网络和 BP神经网络程序由 Matlab 软件工具(Mathworks Inc.)实现,偏最小二乘法和 Kennard-Stone 算法由网络提供的共享算法工具包下载实现^[11]。

1.6 校正模型的评价参数

光谱经过小波变换,再经 RBF 网络建立蛋白质和脂肪的实际含量和 NIR 光谱之间的定量校正模型,以模型的相关系数(R)和预测均方根差(RMSEP)为指标,优化建模参数。

2 结果与讨论

2.1 Kennard-Stone法选取校正集和预测集

奶粉的近红外光谱数据进行建模,选择有代表性的校正集不但可以减少建模的工作量,而且直接影响所建模型的适用性和准确性。本文共 100 个样本,通过 Kennard-Stone 法^[12]依次挑选出 75 个样品作为校正集,余下的 25 个样品作为预测集。

2.2 数据压缩和小波系数选择

在分解尺度为 1, 2, 3 和 4 时,每个尺度下的小波基函数分别为 db1 ~ db8,各自建立了脂肪和蛋白质的近红外光谱模型。通过优化,选择了分解尺度为 4,小波基为 db2 函数,保留 100 个小波系数作为脂肪建模数据的输入。此时的 RMSEP 值为 0.007, R_p 值为 0.990;在分解尺度为 4,小波基为 db4 函数时,预测模型最优,此时的 RMSEP 值为 0.004, R_p 值为 0.994,保留了 111 个小波系数作为蛋白质建模数据的输入。

图 1(a)为奶粉代表性样本的原始光谱,图 1



图 1 奶粉的近红外光谱图 (a)原始光谱图 (b)脂肪的重构光谱信号图

Fig 1 NIR spectrum of milk powder (a) original spectrum (b) reconstructed spectral signal for fat

(b)为基于小波基为 db2函数,分解尺度为 4时的光谱信号重建.采用 100 个小波系数重构得到的光谱信息与样品的原始光谱相比较,可以看出光谱信息几乎没有丢失.

2.3 RBFNN网络优化

在 RBF网络训练中,隐含层神经元和散布常数的确定是两个关键问题.当输入矢量较少时,可以令隐含层神经元数量等于输入向量的个数,这样在不减少预测精度的同时,可以降低模型的复杂性^[13].在进行 RBF网络训练时,需确定的参数仅为 spread 值.设步长为 0.5,考察了 spread 值在 1~10 范围内,与之对应的不同模型的 RMSEP 值.最终确定了 spread = 3.5 以及 spread = 6 分别作为脂肪和蛋白质的 RBF网络的扩散常数值.

2.4 预测结果

脂肪的近红外光谱模型选择了径向基网络扩散常数 spread 值为 3.5,小波变换分解尺度为 4,小波基为 db2函数,保留 100 个小波系数;蛋白质模型选择了径向基网络扩散常数 spread 值为 6,分解尺度为 4,小波基为 db4函数,保留了 111 个小波系数.采用 RBF网络建立了它们的最优模型并进行了预测.脂肪的 RMSEP 值和 R_p 值分别为 0.007 与

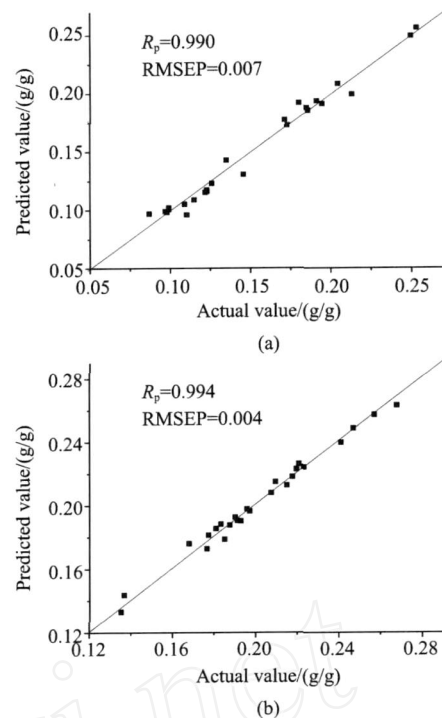


图 2 奶粉定量模型预测集的相关图 (a)脂肪 (b)蛋白质

Fig 2 Prediction value versus actual value for prediction set (a) fat (b) protein

0.990,结果见图 2(a).蛋白质的 RMSEP 值和 R_p 值为 0.004 与 0.994,结果见图 2(b).由图 2 看出,通过上述几种化学计量学方法的结合,脂肪与蛋白质的预测值均能准确地逼近实验实测值.

2.5 建模方法的比较

比较 PLS 法与 BPNN 法 2 种建模方法的结果,由表 1 可以看出,3 种方法中,脂肪与蛋白质是 RBF 模型的 RMSEP 最小, R_p 最大,模型预测效果最好. BPNN 模型的脂肪预测值比 PLS 法的要稍好,但蛋白质模型不及 PLS 法.可能是由于奶粉的近红外光谱数据局局部效应显著,变量个数多,且彼此间常存在严重的复共线性,并与样品组分含量呈非线性关系,所以采用 RBF 网络建模要比 PLS 法好.由于 BPNN 具有自身易陷入局部最小的缺点,结果也较 RBFNN 稍差.

表 1 三种建模方法的比较

Method	Fat		Protein	
	RMSEP	R_p	RMSEP	R_p
PLS	0.010	0.985	0.005	0.992
RBFNN	0.007	0.990	0.004	0.994

Method	Fat		Protein	
	RMSEP	R_p	RMSEP	R_p
BPNN	0.008	0.989	0.006	0.989

3 结论

使用近红外光谱技术对奶粉脂肪和蛋白质含量的检测方法进行了研究. 采用 Kennard-Stone对样本进行了分类, 然后对 1557个变量进行小波变化, 达到压缩去噪的目的. 最后采用 RBF网络进行建模, 通过优化扩散系数 σ 达到较好的效果. 与 BP网络相比, RBF网络具有结构自适应确定、不易陷入局部极小、学习速度快、函数逼近能力强等优点. 近红外光谱结合化学计量学方法能够实现对奶粉脂肪和蛋白质含量快速、无损检测, 为今后奶粉脂肪与蛋白质等成分含量快速无损检测提供了新的化学计量学建模方法. 将进一步研究近红外光谱技术在奶粉成分含量检测的应用, 寻找利于建模的敏感波段, 扩充奶粉品种和样本数量, 并且进一步优化建模方法.

REFERENCES

- [1] Reid L M, Woodcock T, O'Donnell C P, *et al* Differentiation of apple juice samples on the basis of heat treatment and variety using chemometric analysis of M R and N R data[J]. *Food Research International*, 2005, **38** (10): 1109—1115.
- [2] LV Li-Na ZHANG Yue, ZHOU Ding-Wen Composition analysis of milk by using near infrared diffuse reflection spectroscopy[J]. *Journal of Tianjin University* (吕丽娜, 张玥, 周定文. 采用近红外漫反射光谱法分析牛奶成分. *天津大学学报*), 2004, **37** (12): 1093—1096.
- [3] JWU Di, He Yong, FENG Shui-Juan, *et al* Application of infrared spectra technique based on LS-support vector machine to the non-destructive measurement of fat content in milk powder[J]. *Journal of Infrared and Millimeter Waves* (吴迪, 何勇, 冯水娟, 等. 基于 LS-SVM的红外光谱技术在奶粉脂肪含量无损检测中的应用. *红外与毫米波学报*), 2008, **27** (3): 180—184.
- [4] JWU D, He Y, Feng S J. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment[J]. *Analytica Chimica Acta*, 2008, **610** (2): 232—242.
- [5] Borin A, Ferrero M F, Mello C, *et al* Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk[J]. *Analytica Chimica Acta*, 2006, **579** (1): 25—32.
- [6] Bishop C M, *Neural Networks for Pattern Recognition* [M]. Clarendon Press, Oxford, 1995.
- [7] Pulido A, Ruiz nchez I, Rius F X. Radial basis functions applied to the classification of UV visible spectra[J]. *Analytica Chimica Acta* 1999, **388** (3): 273—281.
- [8] Fidêncio P H, Poppi R J, Andrade J C. Determination of organic matter in soils using radial basis function networks and near infrared spectroscopy[J]. *Analytica Chimica Acta*, 2002, **453** (1): 125—134.
- [9] SHAO Yong-Ni, He Yong Method for predicting acidity of bayberry juice by using vis/near infrared spectra[J]. *Journal of Infrared and Millimeter Waves* (邵咏妮, 何勇. 可见/近红外光谱预测杨梅汁酸度的方法研究. *红外与毫米波学报*), 2006, **25** (6): 478—480.
- [10] Feam T, Davies A M C. Near infrared reflectance spectroscopy as a tool for the determination of dichloromethane extractable matter and moisture content in combed wool slivers[J]. *Journal of Near Infrared Spectroscopy*, 2003, **11** (5): 333—339.
- [11] Kennard R W, Stone L A. Computer aided design of experiments[J]. *Technometrics*, 1969, **11**: 137—148.
- [12] Daszykowski M, Semeels S, Kaczmarek K, *et al* TOM-CAT: A MATLAB toolbox for multivariate calibration techniques[J]. *Chemometrics and intelligent laboratory systems*, 2007, **85** (2): 269—277.
- [13] Center for the Research and Development of FeiSi Technology. *The theory artificial of neural networks and its realization in Matlab7* [M]. Beijing: Electronics Industry Press (飞思科技产品研发中心编. *神经网络理论与 Matlab7 实现*. 北京: 电子工业出版社), 2006, 67.
- [3] Dobbs R, Hyttinen M, Steer B, *et al* Rugged and efficient Ka-Band extended interaction klystrons for satellite communication system[C]. *NEC2007*. 2007: 107—108.
- [4] Editorial Board of Vacuum Tube Design Manual *High-power klystron design manual* [M]. Beijing: National Defense Industry Press(电子管设计手册编辑委员会. *大功率速调管设计手册*. 北京: 国防工业出版社), 1979: 157—162.
- [5] Nguyen K T, Pershing D, Wright E L, *et al* Sheet-Beam 90GHz and 220 GHz extend-interaction-klystron designs [C]. *NEC2007*. 2007: 193—194.
- [6] LU Zhi-Gang Research on three-mode overlapped two-gap coupled cavity type output circuit[D]. Beijing: Institute of Electronics, Chinese Academy of Sciences(刘志刚. *三模重*
- [7] LN Fu-Min, DNG Yao-Gen, ZHANG Zhi-Qiang, *et al* Simulation computation method for calculating the impedance matrix of double gap cavity of klystron[J]. *Journal of Electronics and Information technology* (林福民, 丁耀根, 张志强, 等. 速调管双间隙腔的间隙阻抗矩阵频率特性的模拟计算方法. *电子与信息学报*), 2004, **26** (9): 1480—1486.
- [8] Shin Y M, Park G S, Scheitrum G P, *et al* Circuit analysis of an extended interaction klystron[J]. *Journal of the Korean Physical Society*, 2004, **44** (5): 1239—1245.
- [9] Ko K, Lee T G, Kröll N, *et al* A three-gap klystron output cavity at X-band[R]. *SLAC-PUB_5760*, April 1992.

(上接 108页)

叠双间隙耦合腔输出回路的研究. 北京: 中国科学院电子学研究所), 2007: 31—37.