

采用 NITS 检测乳品中蛋白质、脂肪含量

郭中华, 王磊, 刘三亚, 唐燕薇
(宁夏大学物理电气信息学院, 宁夏 银川 750021)

摘要:应用近红外透射光谱法对乳制品中蛋白质和脂肪含量进行检测。首先对光谱进行平滑等预处理,然后使用小波基为 db3,分解尺度为 6 的小波进行数据压缩,最后以压缩后光谱数据作为输入,采用径向基函数人工神经网络(RBF-ANN)建立四种乳制品蛋白质和脂肪含量的预测模型,并试验得出最佳扩散常数 spread 值,其中,对蛋白质建模时最佳 spread 值为 135,此时,相关系数和预测集均方差分别达到 0.9999 和 0.0301;对脂肪建模时最佳 spread 值为 105,此时,相关系数和预测集均方差分别达到 0.9997 和 0.0968。结果表明,结合 RBF-ANN 和小波压缩建立的定量模型更稳定、精度更高,能够对乳制品品质进行快速无损检测。

关键词:近红外光谱;乳制品;小波压缩;RBF 人工神经网络

中图分类号:TP399;O657.33 **文献标识码:**A **DOI:**10.3969/j.issn.1001-5078.2013.07.007

Determining the content of the protein and fat in dairy product using near infrared transmittance spectroscopy

GUO Zhong-hua, WANG Lei, LIU San-ya, TANG Yan-wei
(School of Physics and Electronic Information Engineering, Ningxia University, Yinchuan 750021, China)

Abstract: The content of protein and fat in dairy products is determined by the near infrared transmittance spectroscopy (NITS). The spectrum is preprocessed by methods such as smoothing, then, the data is compressed by wavelet function db3 and compression level 6. The prediction models of protein and fat in the four dairy products are established by radial basis function artificial neural network (RBF-ANN) using the compressed spectrum data as the inputs. The best spread value is obtained by experiments. When the spread is 135, the prediction accuracy of prediction set of protein is the highest and the Correlation coefficient and Mean square error are 0.9999 and 0.0301 respectively. In the same way, when the spread is 105, the prediction accuracy of prediction set of fat is the highest and the Correlation coefficient and Mean square error are 0.9997 and 0.0968 respectively. The results show that the model based on RBF-ANN combined with wavelet is more stable and more accurate. It can be used to test the qualities of dairy products quickly and nondestructively.

Key words: NITS; dairy product; wavelet compression; RBF-ANN

1 引言

鲜牛奶及其制品具有很高的营养价值,随着人们健康意识的提高,具有丰富营养成分的乳制品在人们日常生活中的地位越来越重要。在牛奶营养成分中,脂肪和蛋白质的含量是衡量其质量的核心指标。各个国家对牛奶的理化指标,包括脂肪、蛋白质等含量都有非常明确规定^[1]。从乳业生产角度来

看,牛奶中脂肪和蛋白质等主要成分的含量可为奶厂管理决策提供重要信息,如根据其成分含量科学调控饲料和精饲料配比,蛋白质含量能反映食物中

基金项目:宁夏回族自治区自然科学基金(No. NZ1103)。

作者简介:郭中华(1973-),男,博士,副教授,研究方向为信号处理与检测,图像处理,网络通信。E-mail: guozhh@nxu.edu.cn

收稿日期:2012-12-12;修订日期:2013-04-07

提供的能量充足与否等。因此,研究一种快速检测乳制品主要成分含量的方法,是可以实际生产和生活提供重要的理论依据。

近年来,以光的吸收、衍射、散射、折射、偏振等性质为基础的光谱分析法在食品质量无损检测中极具优势^[2-4],且光谱检测技术的发展为牛奶成分含量的检测提供了新手段。其中近红外光谱分析法作为光谱测量技术和化学计量学有机结合的分析方法,又因具有高效、快速、成本低和绿色环保等优点而应用较为广泛,主要有采用近红外光谱技术检测奶粉(固态)和原奶(液态)中的主要成分含量以及掺假奶、还原奶的检测^[5-8],对于物态较为复杂且经过加工的乳制品主要成分检测还相对较少,因此本文选取几种市场占有率较高的乳制品作为检测对象。

近红外光谱是指波长在 780 ~ 2526 nm 范围内的电磁波,波数范围为 3500 ~ 13000 cm^{-1} ,从波长上可以分为短波近红外(780 ~ 1100 nm)和长波近红外(1100 ~ 2526 nm)。其分析特点是利用近红外光照射被测样品,由于分子团的振动,样品会吸收一部分能量。不同分子团的振动,吸收光的波段不同,主要信息是含 H 基团倍频和合频的吸收^[9]。由于大多数有机化合物都含有 H 基团,近红外光谱分析非常适合分析有机样品。获取近红外光谱一般分为两种技术:透射光谱技术和反射光谱技术,而透射一般分为直接透射和漫透射,反射又分为直接反射和漫反射。由于本文研究对象为液态乳制品,故采用透射法获取样品光谱。

2 实验工作

样本选自宁夏银川市几家大型超市的某品牌液态乳制品,包括纯牛奶、酸牛奶、麦香奶和枸杞红枣奶四个品种共计 320 个,其中每个品种挑选 3 个以上不同生产批次的 80 个样本。每个品种 80 个样本分别按照约 3:1 的比例划分校正集和预测集。

实验时,每个乳制品样品被分为两份,一份用于光谱采集,一份用于化学参考值的测定。样本光谱由 Antaris II 傅里叶变换近红外光谱分析仪的漫透射附件获得。在实验室温度 25 $^{\circ}\text{C}$ 左右条件下,每次光谱仪预热时间为 30min ~ 1h。样品使用光程 1 mm 的石英比色皿装盛约容积 4/5。光谱仪参数设定如下:扫描范围为 10000 ~ 40000 cm^{-1} ;分辨率 8 cm^{-1} ;扫描次数 32 次,每个样本平行测定 3 次,取平均光谱。以麦香奶为例,原始光谱取

平均后如图 1 所示。为了消除背景噪声,每 100 分钟获取一次背景光谱。

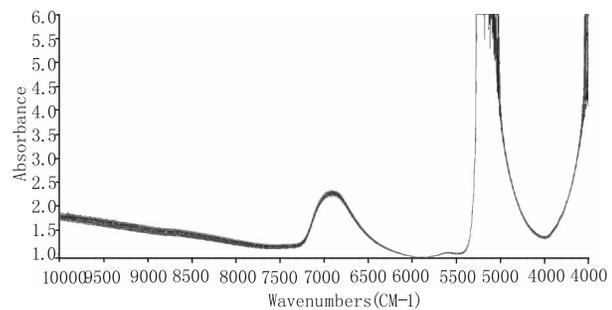


图1 麦香奶原始光谱图

样品蛋白质和脂肪的实际含量(即化学参考值)使用 MilkoScan FT120 多功能乳成分分析仪测量得到。每个样品平行测定 3 次后取平均。以麦香奶为例,蛋白质和脂肪含量信息如表 1 所示。

表1 麦香奶蛋白质、脂肪含量信息

品种	成分	最小值/ ($\text{g} \cdot \text{dL}^{-1}$)	最大值/ ($\text{g} \cdot \text{dL}^{-1}$)	平均值/ ($\text{g} \cdot \text{dL}^{-1}$)	标准差/ ($\text{g} \cdot \text{dL}^{-1}$)
麦香奶	蛋白质	2.27	2.52	2.35	0.04
	脂肪	2.57	2.87	2.67	0.07

3 建模、结果与分析

3.1 原理和方法

3.1.1 RBF 神经网络(RBF-ANN)

RBF(径向基函数)神经网络是由输入层、隐含层和输出层组成的前向神经网络^[10]。RBF-ANN 的基本思想是:用径向基函数作为隐单元的“基”构成隐含层空间,隐含层对输入矢量进行变换,将低维的模式输入数据变换到高维空间内,使得在低维空间内的线性不可分的问题在高维空间内线性可分。与 BP-ANN 相比,RBF-ANN 在结构上具有隐含层到输出层的线性权值关系。

3.1.2 小波变换

小波变换^[11](Wavelet Transform, WT)是一个时间和频率的局域变换,能够有效地从信号中提取信息。在近红外光谱分析中,小波变换可用于压缩样本光谱数据从而达到光谱数据被降维的目的。

3.1.3 校正模型的评价参数

光谱数据经过预处理,然后使用 RBF-ANN、PLS 和 BP-ANN 分别建立蛋白质和脂肪实际含量和光谱之间的定量校正模型,模型优劣则由相关系数(R)和预测均方差(RMSEP)来确定,最后反向优化模型参数。

3.2 光谱数据预处理

采用光谱分析软件 OMNIC7.0 将光谱转化为 CSV 格式的数据,然后使用 TQ7.0 对光谱使用多元散射校正(MSC)、导数算法(Derivative)、平滑(Smoothing)以及相结合等 11 种方法进行预处理。最后分别对每个品种的蛋白质和脂肪建立 PLS 模

型,纯牛奶建模结果如表 2、3 所示。分析表 2 中 R 和 RMSEP 值,得出蛋白质建模的最佳光谱预处理方法为二阶导数(SD)加 S-G 多点平滑,脂肪的最佳光谱预处理方法为 MSC、一阶导数(FD)加 S-G 多点平滑。以麦香奶为例,蛋白质和脂肪建模的预处理光谱图分别如图 2(a)和(b)所示。

表 2 纯奶蛋白质 PLS 建模结果

纯奶蛋白质			
预处理方法	相关系数 R	预测均方差 RMSEP	最佳波段范围/cm ⁻¹
MSC	0.79490	0.0308	4210 - 4220,6395 - 6400,7300 - 7368
SNV	0.92470	0.0304	4740.27 - 4502.40,5254 - 5282,9249.31 - 8762
FD	0.96046	0.0274	4190 - 4770,6809.22 - 6858.80,7000 - 8400
SD	0.90992	0.0272	4170.10 - 4820,7000 - 8500
S - G	0.96817	0.0297	5278 - 5644,9521 - 10208
SD + S - G	0.96859	0.0301	4270.51 - 4879.43,5311.04 - 6427.63,7169.78 - 8297.03
FD + MSC	0.85383	0.0246	8644 - 7069,5854 - 6933
SD + MSC	0.93271	0.0245	4201 - 4829,5325 - 6827,7000 - 8500
FD + S - G + MSC	0.9270	0.0289	4321 - 4400,5623 - 6769,7088 - 8500
SD + S - G + MSC	0.96093	0.0278	4170 - 4900,6470 - 7000,7120 - 9500
S - G + MSC	0.57163	0.0587	4519 - 4522,5230 - 6582

表 3 纯奶脂肪 PLS 建模结果

纯奶脂肪			
预处理方法	相关系数 R	预测均方差 RMSEP	最佳波段范围/cm ⁻¹
MSC	0.98507	0.106	4269 - 4801,5478 - 6296,7006 - 9495
SNV	0.76371	0.0979	4159.84 - 4910,5424 - 6404.31,7020 - 8150
FD	0.95380	0.119	4160 - 4948,7012.71 - 9520
SD	0.98870	0.111	4201.09 - 4902,5304 - 6660.48,7000 - 9621
S - G	0.81580	0.109	4127 - 4945,6220 - 6400,7006.36 - 8130
SD + S - G	0.98970	0.101	4200.21 - 4839.83,5301.47 - 6387.43,7135.54 - 8303.03
FD + MSC	0.96046	0.113	4796.82 - 4159.77,5379 - 6772,6960 - 9200
SD + MSC	0.98923	0.131	4201.09 - 4942,7000 - 10020
FD + S - G + MSC	0.99746	0.0728	4261.51 - 4782.43,5381.00 - 6380.63
SD + S - G + MSC	0.99358	0.0926	4139 - 4920,5995 - 6765,7000 - 10005
S - G + MSC	0.98423	0.105	4220 - 4820,5404 - 6730,7006 - 10200

3.3 小波数据压缩

使用 MATLAB R2009a 对经过预处理的最佳光谱在分解尺度为 1~6,小波基函数为 db1~db6 时,分别进行硬阈值法小波压缩,然后建立蛋白质和脂肪的校正模型,通过优化,选择分解尺度为 6、小波

基为 db3 的函数。经过小波压缩后的光谱数据点由原始的 2179 变为 38,从而达到了数据降维的目的,也为 RBF 和 BP 神经网络建模提供条件,可以有效地提高网络训练速度和精度。小波压缩后的光谱图如图 2(c)所示。

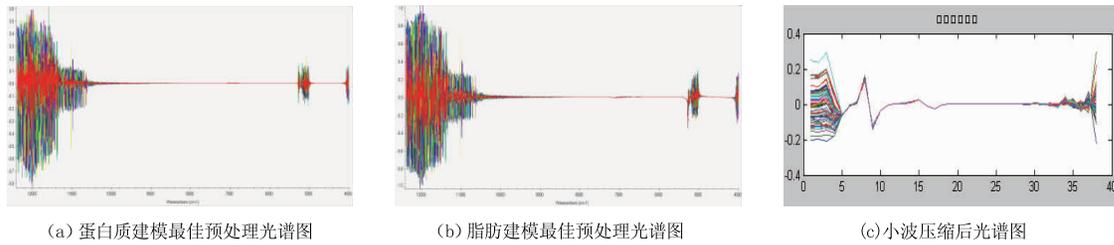


图 2 预处理光谱图

3.4 神经网络建模

3.4.1 RBF-ANN 建模

将小波压缩后的光谱数据作为网络输入变量分别建立蛋白质和脂肪的预测模型,通过对网络的反复训练,最终确定最佳的网络训练参数。蛋白质建模时,输入层节点数为 38、隐含层节点数为 10、输出层节点数为 1,扩散常数 SPREAD 为 135;脂肪建模时,输入层节点数为 38、隐含层节点数为 8、输出层节点数为 1,扩散常数 SPREAD 为 105。

3.4.2 BP-ANN 建模

为了模型性能的可比性,网络输入和 RBF-ANN 保持一致,通过对网络的反复训练,最终确定最佳网

络训练参数。蛋白质建模时,隐含层节点数为 11、输出层节点数为 1、最大训练次数为 860、动量因子为 0.38、学习速率为 0.001、目标误差为 0.001;脂肪建模时,隐含层节点数为 7、输出层节点数为 1、最大训练次数为 1000、动量因子为 0.45、学习速率为 0.001、目标误差为 0.001。

以麦香奶为例,使用 PLS、RBF-ANN 和 BP-ANN 建立蛋白质和脂肪定量模型预测集的相关图分别如图 3(a)、(b)、(c)、(d)、(e)和(f)所示。四种牛奶蛋白质和脂肪的三种建模结果分别如表 4、5 所示。

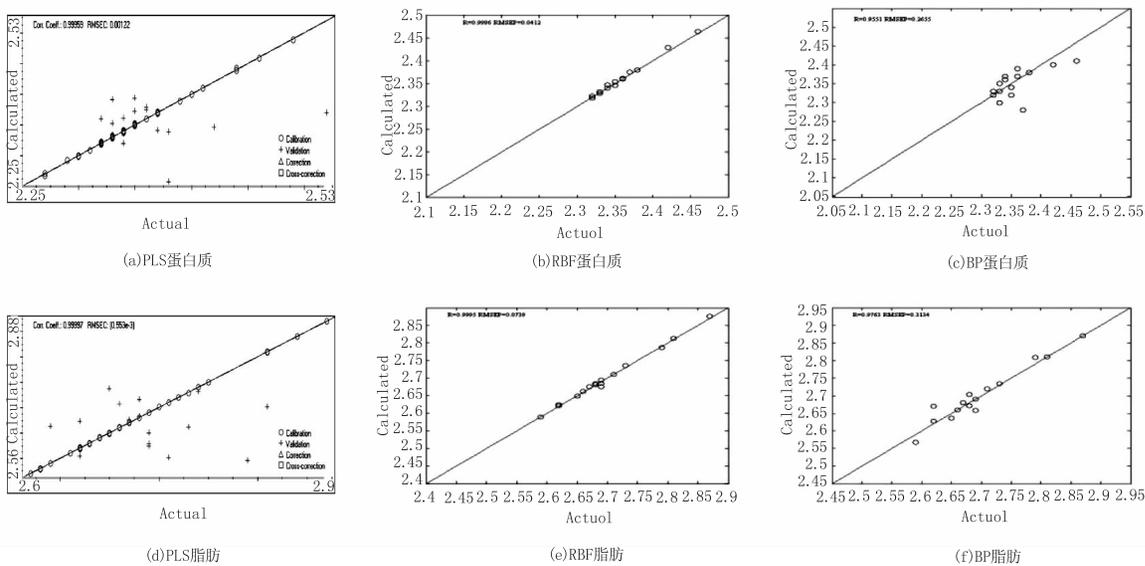


图 3 定量模型预测集相关图

表 4 蛋白质定量模型结果对比

	蛋白质					
	PLS		RBF		BP	
	相关系数 R	预测集均方差 RMSEP	相关系数 R	预测集均方差 RMSEP	相关系数 R	预测集均方差 RMSEP
纯牛奶	0.96859	0.0301	0.9999	0.0301	0.9384	0.277
酸牛奶	0.98787	0.0587	0.9972	0.06	0.9334	0.2502
麦香奶	0.99959	0.0587	0.9996	0.0412	0.9551	0.2655
枸杞奶	0.99998	0.0409	0.9997	0.0331	0.9478	0.2142

表 5 脂肪定量模型结果对比

	脂肪					
	PLS		RBF		BP	
	相关系数 R	预测集均方差 RMSEP	相关系数 R	预测集均方差 RMSEP	相关系数 R	预测集均方差 RMSEP
纯牛奶	0.99958	0.0728	0.9997	0.0968	0.9252	0.3144
酸牛奶	0.99995	0.0654	0.998	0.0688	0.9655	0.3728
麦香奶	0.99997	0.0771	0.9995	0.0739	0.9763	0.3134
枸杞奶	0.97805	0.0308	0.9996	0.065	0.9427	0.3581

3.4.3 建模结果比较

首先,从图 3(a)、(b)、(d)和(e)中可以看出,通过几种化学计量学方法的有效结合,麦香奶蛋白质和脂肪含量的预测值能够准确的逼近实测值。其次,从表 1、2 中可以看出,三种建模方法对于脂肪的预测结果均好于蛋白质的。对于蛋白质和脂肪,预测效果最好的是 RBF-ANN 模型,相关系数 R 最大,且 RMSEP 较小。PLS 模型对脂肪的预测效果接近于 RBF-ANN 模型,除枸杞奶的略差以外。而由于 BP-ANN 本身易陷入局部最小的缺点,模型的预测效果最差。再次,区分建模方法和乳制品品种,观察建模结果蛋白质和脂肪两种成分含量的建模结果发现:采用 PLS 法对枸杞奶的建模结果最好;采用 RBF-ANN 法对纯牛奶的建模结果最好;采用 BP-ANN 法对麦香奶的建模结果最好。最后,对于四个品种的乳制品,麦香奶的预测效果最佳。

4 结论

使用近红外光谱技术对乳制品蛋白质和脂肪含量的检测方法进行了研究。对预处理后的样品光谱数据进行小波压缩,达到了降维和除噪的目的。最后分别采用 PLS、RBF-ANN 和 BP-ANN 建立蛋白质和脂肪含量的预测模型,结果表明,较 PLS 和 BP-ANN 而言,RBF-ANN 具有自适应确定、不易陷入局部最小,学习速度快和函数逼近能力强的优点,其模型精度最佳,最稳定。近红外光谱结合化学计量学分析方法能够实现对乳制品蛋白质和脂肪含量的快速、无损检测,为实际的生产和生活提供了重要的理论依据。下一步将探讨校正集和预测集的划分方法来优化建模结果;增加检测的主要营养成分,如乳糖。

参考文献:

[1] The national food industry standardization technical committee secretariat. Milk and dairy products national standard application guidelines[M]. Beijing: Chinese Standard

Press,2001;25-65. (in Chinese)

全国食品工业标准化技术委员会秘书处编. 乳与乳制品国家标准应用指南[M]. 北京: 中国标准出版社, 2001;25-65.

- [2] Etzion Y, linker R, Cogan U. Determinations of protein concentration in raw milk by mid-infrared Fourier transform infrared/attenuated total reflectance spectroscopy [J]. Dairy Sci,2004,87(9):2779-2788.
- [3] Pan Yuanyuan, Liu Yande, Kong Xudong. Research progress of on-line detection of fruit quality using near infrared spectroscopy [J]. Laser & Infrared, 2010, 40(8): 805-810. (in Chinese)
潘圆媛,刘燕德,孙旭东. 水果内部品质近红外动态在线检测研究进展[J]. 激光与红外, 2010, 40(8): 805-810.
- [4] Ke Zhao, Yan Xiong, Min Zhao. Rapid non-destructive testing of navel orange based on near infrared spectroscopy[J]. Laser & Infrared, 2011, 41(6):649-652. (in Chinese)
赵珂,熊艳,赵敏. 基于近红外光谱技术的脐橙快速无损检测[J]. 激光与红外,2011,41(6):649-652.
- [5] Nagarajan R, Singh Parul, Mehrotra Ranjana. Direct determination of moisture in powder milk using near infrared spectroscopy[J]. Journal of Automated Methods & Management in Chemistry,2006(3):1-4.
- [6] Tianzi Wang, Limin Zheng, Lijun Tian, et al. Analysis of milk and dairy products by near-infrared spectroscopy: a review[J]. Spectroscopy and Spectral Analysis, 2010, 30(12):3208-3212. (in Chinese)
王田子,郑丽敏,田立军,等. 近红外在乳及乳制品质量检测中的研究发展[J]. 光谱学与光谱分析, 2010, 30(12):3208-3212.
- [7] Yun Xu, Yiming Wang, Jingzhu Wu, et al. Detecting the melamine of pure milk by near Infrared spectra[J]. J. Infrared Millim. Waves, 2010, 29(1):53-56. (in Chinese)
徐云,王一鸣,吴静珠,等. 用近红外光谱检测牛奶中的三聚氰胺[J]. 红外与毫米波学报, 2010, 29(1):

- 53 – 56.
- [8] Zhu Xiangrong, Shan Yang, Li Gaoyang, et al. Determination of fat and protein content in liquid milk by near infrared spectroscopy[J]. Food Science, 2011, 32(12): 191 – 195. (in Chinese)
朱向荣, 单阳, 李高阳, 等. 近红外光谱法快速测定液态奶中蛋白质和脂肪含量[J]. 食品科学, 2011, 32(12): 191 – 195.
- [9] Lu Wanzhen, Yuan Hongfu, Xu Guangtong. Modern near infrared spectral analysis technology[M]. 2nd ed. Beijing: China Petrochemical Press, 2007: 15. (in Chinese)
陆婉珍, 袁洪福, 徐广通. 现代近红外光谱分析技术[M]. 2版. 北京: 中国石化出版社, 2007: 15.
- [10] Bishop C M. Neural Networks for Pattern Recognition[M]. Oxford: Clarendon Press, 1995.
- [11] Fearn 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.