

波长优选 BP 神经网络用于近红外光谱分析

尹慧敏¹, 吴文福², 张亚秋²

(1. 吉林农业大学工程技术学院, 吉林 长春 130118; 2. 吉林大学生物与农业工程学院, 吉林 长春 130025)

摘要:近红外光谱分析技术在物质成分分析中得到广泛的应用,在自主研发的滤光片型近红外仪器中应用基于波长优选的 BP 神经网络模型的方法。该方法是采用多元线性回归算法获取最优波长,将最优波长作为 BP 神经网络模型的输入,所得模型的拟合残差为 8.768991×10^{-6} ,建模样品集相关系数和检验样品集相关系数分别为 0.994 和 0.996。试验结果表明,基于波长优选的 BP 神经网络模型方法能够更快获得最优解,减少建模所用变量,建立稳健的定量模型。

关键词:波长优选;BP 神经网络;近红外光谱分析

中图分类号:O657.32 文献标识码:A DOI:10.3969/j.issn.1001-5078.2011.08.010

Application of BP neural network based on wavelength optimization in near infrared spectroscopy

YIN Hui-min¹, WU Wen-fu², ZHANG Ya-qiu²

(1. College of Engineering Technology, Jilin Agricultural University, Changchun 130118, China;

2. School of Biological & Agricultural Engineering, Jilin University, Changchun 130025, China)

Abstract: Near Infrared Spectroscopy has been widely used in material composition analysis. In this self-designed filter NIR device, the article presents a modeling method based on BP neural network with wavelength optimization selection. First, multiple linear regression algorithm modeling is used to analyze the prepared 26 samples to obtain optimal wavelength which is then used as BP neural network model's input. After the calculation, it shows that the fitting residual is 8.768991×10^{-6} and the correlation coefficients of modeling samples and testing samples are respectively 0.994 and 0.996. The experimental result indicates that the BP neural network modeling method based on wavelength optimization can gain the optimal solution more quickly, reduce the variables used in modeling and apparently improve the robustness of the quantitative analysis model, enhance the actual ability of detecting.

Key words: wavelength optimization selection; BP neural network; near infrared spectroscopy

1 引言

近红外(NIR)光谱分析技术在物质成分分析中以快速简便无损等特点得到越来越广泛的应用^[1-3]。它是一种间接测量技术,需要借助化学计量学方法将样品光谱和其质量参数建立数学模型来对未知样品进行检验。由于样品的状态、测量条件、成分的组成等因素对光谱的影响^[4]大多数属于非线性的,因而常规的线性处理方法就不能完全适用于近红外光谱分析中。而作为三大非线性处理方法之一的神经网络模型,随着它的成熟和相关理论的

完善,其在分类、非线性模型方面已经被证明是一种有效的方法^[5-8]。

本文通过对不同蛋白质含量的近红外光谱分析研究,将波长优选和 BP 神经网络相结合,旨在寻找一种检测物质成分含量的快速有效的方法。

基金项目:吉林省科技发展计划国际科技合作项目(No. 20080721)资助。

作者简介:尹慧敏(1983-),女,硕士,助教,主要从事农业信息化技术的研究。E-mail:jlau_yhm@126.com

收稿日期:2011-03-26

2 基本原理

2.1 BP神经网络

神经网络的主要特征是自学习,通过对样本模式的学习模拟信息之间的内在机制,实现输入与输出之间的高度非线性映射。用人工神经网络建立模型不涉及过程的内部机,只要有大量的数据对网络进行训练,网络就能找出输入输出关系,特别适合与对于复杂过程用人工神经网络将能建立精确的数学模型。一个典型的四层BP神经网络模型结构如图1所示。

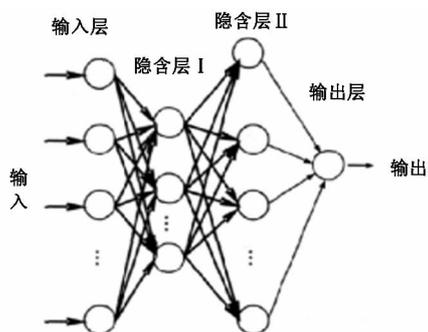


图1 四层BP神经网络模型结构

2.2 主成分信息提取

人工神经网络建模方法的计算量比较大,其输入层节点数不能太多,在采用分立分光元件进行光谱扫描的近红外分析仪器研制中,定标波长中有的对于模型预测能力的提高是有益的,也可能对其预测能力产生干扰,因此,在人工神经网络建模之前,需要将高维的光谱数据变换压缩为低维数据,以作为BP人工神经网络的输入。为此采用定标波长组合的多元线性回归方法对光谱数据进行预处理。

3 试验材料和方法

3.1 试验仪器

采用自主研发的JD-NIR-01型近红外检测仪进行近红外光谱测定,扫描样品质量为50g。光路体系主要由卤钨灯、滤光片和斩光片等组成。卤钨灯光通过近红外窄带干涉滤光片形成单色后,垂

直照射到样品上被吸收漫反射后由检测器接收,光谱信号通过数据采集卡传入计算机作进一步处理。

3.2 试验材料

所建立的样品集是根据玉米中蛋白质含量范围,样品由淀粉和蛋白粉按比例混合均匀并按梯度配制而得,蛋白质含量范围是10%~16%。总样品数为26个,其中20个样品作为建模样品集,其余样品作为检验样品集。

3.3 化学值测定

近红外分析是以参比分析方法为基础的,其分析的精确度(也叫重现性)与准确度依赖于参比分析法的精确度与准确度,配制成的样品并不能直接用于模型的建立。由于配制过程中难免有误差,因此还需要经过化学方法的测定。选用凯氏定氮法测定样品中蛋白质的含量。

4 建模的结果与讨论

4.1 样品集光谱

已知一组样品(26个样品)的蛋白含量,在本仪器上先扫描空白,然后装样,扫描样品得到的吸光谱如图2所示。

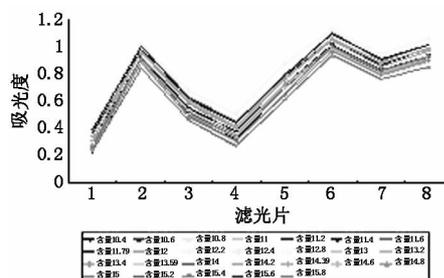


图2 样品集吸收光谱

从样品集吸收光谱可以看出,每个样品的吸收光谱曲线形态相似,具有良好的规律性。

4.2 波长优选

采用以选取各个定标波长组合进行多元线性回归所给出的检验相关系数最大的定标模型所对应的波长组合为最优波长,其结果如表1所示。

表1 线性回归模型的总结表

波长数目	建模样品集		检验样品集						
	平均相关系数	相关系数	相对误差/%						相关系数
			10.60	11.40	12.80	13.20	14.60	15.20	
2	0.350	0.606	29.23	2.54	0.54	2.94	14.18	15.55	0.087
3	0.510	0.824	6.34	4.09	2.59	2.10	19.41	9.29	0.616
4	0.626	0.847	5.89	7.30	1.11	1.25	19.40	9.53	0.707
5	0.713	0.873	14.98	13.54	8.90	3.27	21.29	2.09	0.593
6	0.780	0.877	14.36	12.36	9.47	3.56	18.66	3.77	0.650
7	0.841	0.886	14.23	9.74	6.87	1.50	17.24	7.45	0.665
8	0.887	0.887	13.54	9.19	6.80	0.69	16.68	8.23	0.683

通过对建模样品集的观察,随着参加建立模型的波长的增加,建立的多元线性回归方程的相关系数和平均相关系数都随着增大,说明建立的模型效果增强;通过检验样品集的预测,发现四元线性回归模型预测效果最好,是最经济的。另外,在检验样品集的预测中可以发现含量为 14.60 的点位奇异点,应将其删除。

由上知,最优波长对应的滤光片分别为 1,3,4,8 号滤光片,但是在该仪器中使用多元线性回归算法对蛋白质含量进行检测,预测效果不好,线性度差,因此将采用 BP 神经网络对其建立定量模型。

4.3 BP 神经网络建模

使用神经网络建模方法对 X1, X3, X4 和 X8 最优波长组合进行建模。在训练网络之前将输入样本归一化,然后建立 BP 神经网络模型。

4.3.1 建立 BP 神经网络模型

在 DPS 软件中设定各个参数,确定隐含网络层数为 2,输入层节点数为 4,根据经验值和系统训练结果的比较确定第一隐含节点数为 16,第二隐含节点数为 10;由于训练速率越大,权重变化越快,收敛越快,但是其值过大又会引起系统震荡,所以根据经验值确定最小训练速率为 0.1,动态参数为 0.6,sigmoid 函数的参数为 0.9,允许误差为 0.0000001,迭代次数为 20000 次,数据转换为标准化变换。

4.3.2 训练神经网络

将建模样品集中的 20 个样本作为训练样本训练神经网络,由系统不断地自动调节网络权值和阈值,直到达到允许误差或者迭代次数。图 3 为在建立 BP 神经网络模型训练过程中误差的衰减曲线。

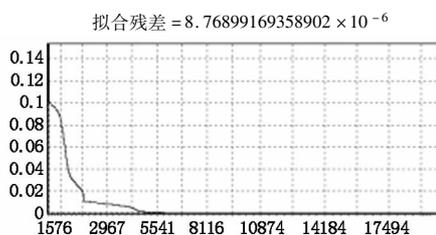


图 3 BP 神经网络训练过程中误差的衰减曲线

建模样品集经 BP 神经网络模型预测的结果如图 4 所示,预测值与化学值的相关系数 R 为 0.994,相对误差平均值为 1.27%。

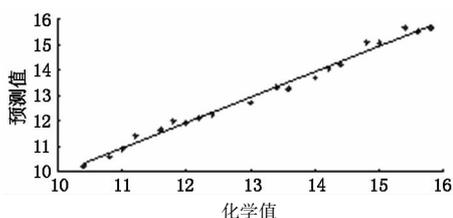


图 4 BP 神经网络模型建模样品集预测值与化学值关系

4.3.3 检验

BP 神经网络模型对检验样品集进行检验,其结果如图 5 所示。

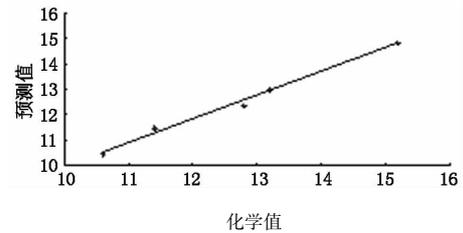


图 5 BP 神经网络模型检验样品集预测值与化学值关系

检验集样品经 BP 神经网络模型预测的预测值与化学值的相关系数 R 为 0.996,平均相对误差值为 1.97%。

5 结论

综上,通过在自主研发的 JD-NIR-01 型近红外检测仪上对蛋白质含量的标定过程中,可以发现 NIR 分析是一种非线性处理方法。采用多元线性回归方法确定最优波长后,再利用最优波长建立 BP 神经网络模型的这种方法,能够明显的提高定量分析模型的稳健性,增强实际检测能力,并减少建模变量。另外,本文提出的波长优选法与 BP 神经网络算法的标定方法虽然是以自制样品作为训练和检验样本,但也同样适用于其他粉状样品。

参考文献:

- [1] Soumya Roy, Ramaswamy Ananteswaran, John Shenk, et al. Determination of moisture content of mushrooms by spectroscopy [J]. J. Sci. Food Agric, 1993, 63 (3): 355 - 360.
- [2] Stephen Delwiche, Kents Mckenzie, Billd Webb. Quality characteristics in rice by NIR reflectance analysis of whole-grain milled samples [J]. Cereal Chem, 1996, 73 (2): 257 - 263.
- [3] Yang Chunmei, Li Ning, Han Donghai, et al. Recent developments of NIRS technique for nondestructive measurement of fruit and vegetable interior quality [J]. Laser & Infrared, 2009, 39(11): 1137 - 1141. (in Chinese)
杨春梅, 李宁, 赵学珏, 等. 用于果蔬内部品质无损检测的 NIRS 技术新进展 [J]. 激光与红外, 2009, 39 (11): 1137 - 1141.
- [4] Yan Yanlu, Zhao Longlian, Han Haidong, et al. Near infrared spectroscopy basic and applications [M]. Beijing: China Light Industry Press, 2005. (in Chinese)
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.

- [5] Bai Yingkui, Meng Xianjiang. Improving partial least square regression precision in NIR multi-component analysis using artificial neural network [J]. Spectroscopy and Spectral Analysis, 2005, 25(3): 381 - 383. (in Chinese)
白英奎, 孟宪江. 利用神经网络提高偏最小二乘法的 NIR 多组分分析精度 [J]. 光谱学与光谱分析, 2005, 25(3): 381 - 383.
- [6] Lin Min, Lü Jin. Determination on components of corns based on neural networks and near infrared spectrum [J]. Infrared Technology, 2004, 26(3): 78 - 81. (in Chinese)
林敏, 吕进. 基于神经网络与近红外光谱的玉米成分检测方法 [J]. 红外技术, 2004, 26(3): 78 - 81.
- [7] Feng Shangkun, Xiu Haiju. Study on content of beer alcohol with near infrared spectroscopy based on back-propagation neural network [J]. Infrared Technology, 2008, 30(1): 58 - 60. (in Chinese)
冯尚坤, 徐海菊. 基于 BP 神经网络的啤酒酒精精度近红外光谱快速检测 [J]. 红外技术, 2008, 30(1): 58 - 60.
- [8] Zhang Guangjun, Suranjan Panigrahi. On-line near-infrared transmittance type inspecting system for protein content of grain [J]. Opto Electronic Engineering, 2001, 28(2): 19 - 22. (in Chinese)
张广军, Suranjan Panigrahi. 近红外透射式谷物蛋白质含量在线监测系统 [J]. 光电工程, 2001, 28(2): 19 - 22.