

文章编号 1004-924X(2006)01-0058-05

基于径向基函数网络的茶多酚 总儿茶素近红外光谱检测模型的研究

陈华才^{1, 2}, 吕进¹, 陈星旦², 刘辉军¹

(1. 中国计量学院, 浙江 杭州 310018; 2. 中国科学院长春光学精密机械与
物理研究所 应用光学国家重点实验室, 吉林 长春 130033)

摘要:基于径向基函数神经网络(RBFN)建立了茶多酚总儿茶素含量的近红外光谱分析模型。茶多酚光谱采用小波压缩、标准化处理后,进行主成分分解,以主成分光谱作为RBFN的隐层输入函数,并通过改变主成分数对模型进行优化。当主成分数为7时得到了RBFN优化模型,该模型对定标样品集、全样品集和预测样品集的预测值与实际值回归系数 R 分别为0.999, 0.999和0.992,预测均方误RMSEP分别为1.08%, 2.06%和3.68%。

关键词:径向基函数网络;茶多酚;儿茶素;近红外光谱;模型

中图分类号:O657.33 文献标识码:A

Near infrared spectroscopic model for determinating total catechins in tea polyphenol powder based on radical basis function network

CHEN Hua-cai^{1,2}, LU Jin¹, CHEN Xing-dan², LIU Hui-jun¹

(1. *Chinese Jiliang University, Hangzhou 310018, China;*

*2. State Key Laboratory of Applied Optics, Changchun Institute of Optics,
Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China*)

Abstract: A near infrared spectroscopic models for determinating the total catechins in tea polyphenol powder were presented based on radical basis function network. The spectra of tea polyphenol powder samples were pretreated with wavelet transform (sym6), standard normalization and principle components analysis (PCA), the eigenvectors of PCA were used as the input functions of RBFN. The models were optimized by changing the number of PCA and evaluated by the correlation coefficients (R) and root mean square error of prediction (RMSEP) of calibration set and the prediction set. The optimum correlation coefficient (R) of the calibration set, all sample set and prediction set were 0.999, 0.999 and 0.992, respectively, and RMSEP of them were 1.08%, 2.06% and 3.68%, respectively.

Key words: radical basis function network (RBFN); tea polyphenol; catechins, near infrared spectroscopy (NIR); model

收稿日期:2005-09-22;修订日期:2005-11-18.

基金项目:吉林省科技计划重点项目(No. 20040324-2)、中国科学院长春光学精密机械与物理研究所应用光学国家重点实验室开放基金和浙江省自然科学基金项目(No. 202081)

1 引言

茶多酚是从茶叶中提取的由多种儿茶素单体及少量咖啡碱等其他物质组成的混合物。药理实验表明,茶多酚有明显的清除人体自由基,阻断N-亚硝基化合物形成,抑制脂氧合酶活性和脂质过氧化作用,并有防突变、防癌、抗衰老及防辐射等药理活性,其中起主要作用的是儿茶素类物质^[1]。目前已发现茶多酚中主要有以下6种儿茶素单体化合物:儿茶素(+)-Catechin(C)、表儿茶素(-)-epicatechin(EC)、没食子儿茶素(+)-gallocatechin(GC)、表没食子儿茶素(-)-epigallocatechin,(EGC)、表儿茶素没食子酸酯(-)-epicatechin gallate(ECG)、表没食子儿茶素没食子酸酯(-)epigallocatechin gallate(EGCG)。茶多酚中总儿茶素含量(Ct)是指这6种单体的总和,即 $C_t = C + EC + GC + EGC + ECG + EGCG$ 。总儿茶素及其6种单体的含量是茶多酚的重要品质参数,也是茶多酚生产和贸易中重要的检测指标。

目前对茶儿茶素定量分析的方法主要有薄层色谱法^[2]、气相色谱法(GC)^[3]、高效液相色谱法(HPLC)^[4]、毛细管电泳法(CE)^[5]等,其中应用最广泛的是高效液相色谱法。这些方法分析精度高,但是分析过程复杂,速度较慢,无法满足茶多酚生产和贸易过程中对产品质量快速检测和在线成分含量监控的需要。近红外光谱分析技术具有分析速度快、样品处理简单、无需试剂、无污染、多组分检测等特点,已广泛应用于农产品、石油化工产品、食品、纺织品等的成分快速检测。国内外也进行了利用近红外光谱分析技术定性定量分析绿茶品质的研究,建立了绿茶中茶多酚、抗氧化力、儿茶素、咖啡碱等成分定量分析的定标模型^[6-10]。作者也曾以高效液相色谱分析结果为参考值,建立了茶多酚中总儿茶素含量近红外光谱分析的线性模型^[11]。本文在此基础上,利用径向神经网络建立了茶多酚中总儿茶素含量的非线性模型。

2 原理和算法

目前近红外光谱定量分析中常用的线性校

正方法有主成分回归(PCR),偏最小二乘(PLS)等。这些方法都是以朗伯-比耳(Lambert-Beer)定律为基础,假设样品的吸光度与其成分的浓度具有线性相关性,通过对光谱数据和样品的浓度建立线性校正模型,利用此模型根据光谱数据即可预测样品成分的浓度。而在实际应用中,由于生物样品一般是由多组分组成的混合物体系,其中各组分相互作用、仪器的噪声和基线漂移等使朗伯一比耳定律成立的条件改变,致使吸光度和样品的浓度之间呈现出非线性关系。目前较常用的非线性校正方法是人工神经网络(ANN)方法。在化学计量中常采用的一种人工神经网络结构是反向误差传播(BP)网络,在实际应用中BP网络存在训练速度慢,容易陷入局部极小等问题^[12]。本文采用在逼近能力和学习速度等方面优于BP网络的另一种网络—径向基函数网络(RBFN)来建立茶多酚中儿茶素含量的非线性模型。

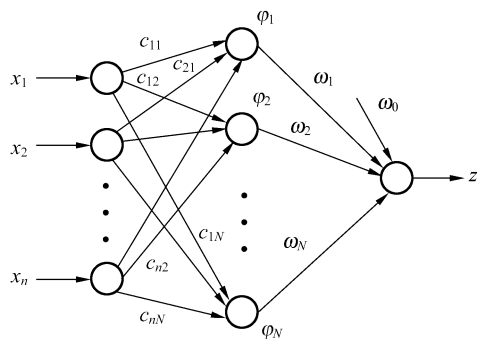


图 1 径向基函数网络的结构图

Fig. 1 Structure of radial basis

RBFN 的典型结构是一个 3 层前向网络,其结构如图 1 所示,其中隐含单元为取径向基函数的局部感受单元。RBFN 可描述为:

$$Z = \frac{\omega_0 + \sum_{i=1}^N \omega_i \varphi_i(\|x - c_i\| / \sigma_i)}{\sum_{i=1}^N \varphi_i(\|x - c_i\| / \sigma_i)}$$

其中, N 为隐层聚类中心数目, $x \in R^n$ 为 n 维输入向量, $c_i \in R^n$ 为第 i 个隐单元中心向量,可看作输入层到第 i 个隐单元中心的权向量, ω_i 为第 i 个隐单元到输出层的权值, σ_i 表示第 i 个隐单元的宽度, $\|\cdot\|$ 为欧几里得范数, φ_i 为径向基函数,通常取高斯函数,即:

$$\varphi_i(\|x - c_i\|/\sigma_i) = \exp\left(-\frac{\|x - c_i\|^2}{\sigma_i^2}\right)$$

3 实验与结果

3.1 实验仪器

Thermo Nicolet NEXUS 870 傅里叶近红外光谱仪, InGaAs 检测器, OMN IC E. S. P. 5. 2a 光谱采集软件, TQ analyst V6 化学计量分析软件。

表 1 茶多酚样品总儿茶素含量实际值

Tab. 1 True values of catechins of tea polyphenol samples

样品集类别 sample set	样品数(n)	最大值 max (%)	最小值 min (%)	平均值 mean (%)
全部样品 all sample	52	98.32	3.64	60.678
定标集 calibration set	39	98.32	3.64	60.539
预测集 prediction set	13	94.96	10.71	61.182

3.3 光谱采集

茶多酚样品不作任何处理直接进行近红外光谱采集。采用漫反射吸收光谱法, 扫描范围 4 000~10 000 cm^{-1} , 分辨率 32 cm^{-1} , 扫描 30 次平均。

3.4 模型结果与分析

近红外光谱仪采集的茶多酚光谱先采用小波压缩(sym6 小波)、标准化处理, 并进行主成分分解, 以分解的主成分向量作为 RBFN 隐层输入函数, 通过改变主成分数来对模型进行优化。通过自行编制的 Matlab 程序实现。

在模型的优化过程中, 主要以预测样品集的预测均方误 RMSEP 为模型优化程度的判据:

$$\text{RMSEP} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

其中: y_i 表示第 i 个样品的真实值, \hat{y}_i 表示第 i 个样品的预测值。

以不同主成分作为输入函数建立的 RBFN 模型对茶多酚样品中总儿茶素含量的预测均方误见表 2。

从表 2 可以看出, 当取主成分为 2 时, RBFN 模型即能获得较好的预测结果。当主成分为 5 时,

3.2 实验样品

具有一定儿茶素含量梯度的茶多酚(喷雾干燥粉) 样品 52 份, 总儿茶素含量范围为 3.64%~98.32%。根据儿茶素含量高低, 间隔取样, 挑选 39 个样品组成定标样品集, 用于定标模型的建立; 剩余的 13 个样品组成预测样品集, 用于对定标模型预测结果的检验(表 1)。儿茶素含量实际值由 HPLC 法测定, 具体测定方法见参考文献。

表 2 不同主成分 RBFN 模型对定标样品集、全部样品和预测样品集的预测均方误

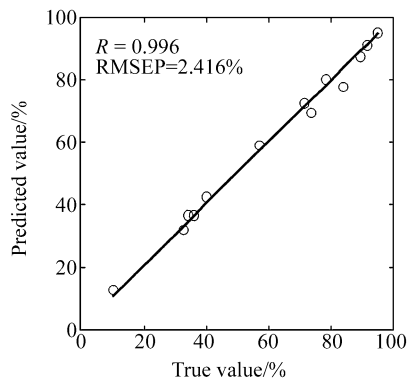
Tab. 2 RMSEP of calibration set, total sample set, and prediction set by RBFN with different PCA numbers

主成分 PCA(n)	RMSEP(%)		
	Calibration set	Total samples set	Prediction set
1	23.03	22.96	22.75
2	4.00	4.36	5.21
3	1.80	3.23	5.65
4	1.48	3.35	6.19
5	1.12	1.94	3.37
6	1.18	1.58	2.42
7	1.21	1.81	2.94
8	1.32	2.90	5.32
9	1.32	2.90	5.32

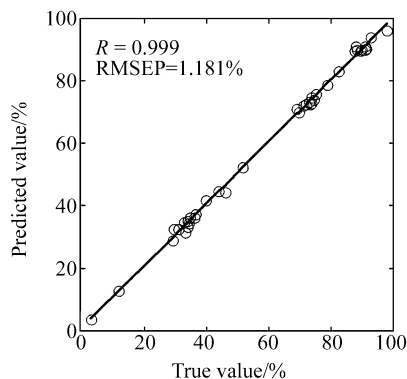
模型对定标样品集预测的均方误(SEP)最小, 表明此时模型的拟和程度最高, 但全样品集的预测均方误和预测样品集均方误不是最优, 表明模型存在一定的过拟合现象(Over-fitting)。当主成分取 6 时, 预测样品集和全样品集的 RMSEP 有最小值, 定标样品集的 RMSEP 也接近最小, 此时可认为是最佳模型。

以该模型分别通过茶多酚样品的近红外光谱值预测样品中总儿茶素含量。图 1 为该值与实际

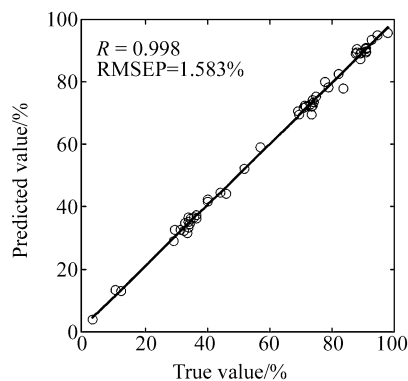
值的回归系数。可以看出,该模型的拟合程度和预测准确度都十分理想,无论是对参加建模的定标样品集,还是对未参加建模的预测样品集,预测值与实际值的回归系数都在 0.99 以上,预测均方误差小于 3%。



(a) 预测样品集
(a) Prediction set



(b) 定标样品集
(b) Calibration set



(c) 全样品集
(c) All samples set

图 2 RBFN 模型对不同样品集预测结果散点图
Fig. 2 Scattering plots of RBFN optimum model prediction to different sample sets

4 结 论

结果表明,RBFN 模型具有较好的稳健性。茶多酚光谱采用小波压缩,标准化处理后,进行主成分分解,以主成分向量作为 RBFN 的隐层输入函数,并对模型进行优化。优化模型对定标样品集、全部样品集和预测样品集中儿茶素预测值与实际值的回归系数 R 分别为 0.999,0.999 和 0.992,预测均方误差 RMSEP 分别为 1.08%,2.06%和 3.68%。RBFN 模型充分考虑了光谱值与儿茶素含量之间的非线性,因此模型的拟合能力大大增强,但同时也存在过拟合现象的风险。为避免过拟合风险,实际应用中必须采用预测样品集对模型进行检验,只有对预测样品集和定标样品集同时具有较高的回归系数和较低的预测标准误差的模型才是稳健的模型。

参考文献:

- [1] 范娟,许士凯. 茶多酚药理作用的研究进展[J]. 现代中西医结合杂志,2002,11(22):2301-2302.
FAN J, XU SH K. Advances of pharmacological studies on tea polyphenols[J]. *Modern Journal of Integrated Traditional Chinese and Western Medicine*, 2002,11(22):2301-2302. (in Chinese)
- [2] 于海宁,沈生荣,臧荣春. 茶多酚中儿茶素类的 HPLC 分析方法学考察[J]. 茶叶科学,2001,21(1):61-64.
YU H N, SHEN S R, ZANG R C. Study on the method of HPLC analysis of catechins in tea polyphenols[J]. *J. Tea Sciences*, 2001,21(1):61-64. (in Chinese)
- [3] 李秉定,汤蓬,汤坚,等. 用气相色谱和高效液相色谱测定茶叶中抗氧化剂的组成[J]. 南京农业大学学报,1992,15:73.
LI B D, TANG P, TANG J, et al. Determination of antioxidants in tea by gas chromatography and high performance

- liquid chromatography[J]. *J. Nanjing Agri. Univ.*, 1992,15:73. (in Chinese)
- [4] 戴军,王洪新,陈尚卫,等. 茶叶及茶多酚中儿素的高效液相色谱分析方法研究[J]. 色谱, 2001,19(5):398-402.
DAI J,WANG H X,CHEN S W,*et al.* Study on the analytical methods of catechins in tea and green tea polyphenol samples by high performance liquid chromatography[J]. *Chromtography*, 2001,19(5):398-402. (in Chinese)
- [5] 曹玉华,张欣,丁祥欢,等. 毛细管电泳-电化学检测测定茶叶中咖啡因、表儿茶素和抗坏血酸[J]. 分析化学, 2001,29(9):1072-1075.
CAO Y H,ZHANG X,DING X H,*et al.* Determination of caffeine,epicatechin and ascorbic acid in tea samples by capillary zone electrophoreses with electrochemical detection[J]. *Chinese Journal of Analytical Chemistry*,2001,29(9):1072-1075.
- [6] 夏贤明,丁宁. 用近红外光谱法检测绿茶中品质成分的研究[J]. 分析化学,1991,19(8):945-948.
XIA X M,DING N. Study on the quality constituents in green tea with near infrared spectroscopy[J]. *Chinese Journal of Analytical Chemistry*,1991,19(8):945-948. (in Chinese)
- [7] SCHULZ H,ENGELHARDT U H,WEGENT A, *et al.* Application of near-infrared reflectance spectroscopy to the simultaneous prediction of alkaloids and phenolic substances in green tea leaves[J]. *J of Agricultural and Food Chemistry*, 1999,47(12):5064-5067.
- [8] LUYPACERT J,ZHANG M H,MASSART D L. Feasibility study for the use of near infrared spectroscopy in the qualitative and quantitative analysis of green tea,*Camellia sinensis*(L.)[J]. *Analytica Chimica Acta*, 2003, 487: 303-312.
- [9] ZHANG M H,LUYPACERT J,FERANDEZ P J A,*et al.* Determination of total antioxidant capacity in green tea by near-infrared spectroscopy and multivariate calibration[J]. *Talanta*, 2004, 62: 25-35.
- [10] 孙耀国,林敏,吕进,等. 近红外光谱法测定绿茶中氨基酸、咖啡碱和茶多酚的含量[J]. 光谱实验室,2004,21(5):939-943.
SUN Y G,LIN M,LV J,*et al.* Determination of the contents of amino acids,caffeine,and polyphenols in green tea by near infrared spectroscopy[J]. *Chinese J. Spectro. Lab.*, 2004,21(5):939-943. (in Chinese)
- [11] 陈华才,吕进,佟春红,等. 近红外光谱法测定茶多酚中总儿茶素含量[J]. 中国计量学院学报,,2005,16(1):17-20.
CHEN H C,LV J,FENG C H,*et al.* Feasibility study on quantitative analysis of total catechnins in tea polyphenol powder by near infrared spectroscopy[J]. *J. Chin. Jiliang Univ.*,2005,16(1):17-20. (in Chinese)
- [12] 张乃尧,阎平凡. 神经网络与模糊控制[M]. 北京:清华大学出版社,1998.
ZHANG N R,YAN P F. *Neural networks and fuzzy control*[M]. Beijing:Publishing House of Stinghua University, 1998

作者简介:陈华才(1968—),男,博士,副研究员,中国计量学院生命科学院教师,中国科学院长春光学精密机械与物理研究所应用光学国家重点实验室博士后。主要从事生物这全检测、生物光学检测方面的科研与数学工作。E-mail:huacaichen@cjl.u. edu. cn