

【文章编号】 1004-1540(2005)03-0188-03

# 基于 RBF 网络和 NIRS 的绿茶水分含量分析模型

刘辉军<sup>1</sup>, 吕进<sup>1,2</sup>, 林敏<sup>1</sup>, 陈华才<sup>1</sup>, 于良子<sup>3</sup>

(1. 中国计量学院 计量技术工程学院, 浙江 杭州 310018; 2. 上海理工大学 光电信息工程学院, 上海 200093;  
3. 中国农业科学院 茶叶研究所, 浙江 杭州 310008)

**摘要** 基于径向基函数(RBF)和反向传播(BP)神经网络分别建立了绿茶水分含量的近红外光谱分析模型. 结果表明:RBF 网络预测模型的相关系数  $r(\rho)=0.933$ , 预测标准误 RMSEP=0.528%;BP 网络预测模型的相关系数  $r(\rho)=0.914$ , 预测标准误 RMSEP=0.598%. RBF 网络模型优于 BP 网络模型.

**关键词** 绿茶水分; 径向基函数(RBF); 近红外光谱; 定量分析

**中图分类号** O657.33

**文献标识码** A

## Study on the near infrared spectroscopy model of the moisture content determination based on radial basis function networks

LIU Hui-jun<sup>1</sup>, LU Jin<sup>1,2</sup>, LIN Min<sup>1</sup>, CHEN Hua-cai<sup>1</sup>, YU Liang-zi<sup>3</sup>

(1. College of Metrological Technology and Engineering, China Jiliang University, Hangzhou 310018, China; 2. College of Optics and Electronics Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China;  
3. Tea Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310008, China)

**Abstract:** Mathematic models to determinate the moisture content in green tea by the near infrared spectroscopy (NIR) were built based on radial basis function (RBF) neural networks and back propagation (BP) neural networks. The correlation coefficient ( $r$ ) and root mean square error of predication (RMSEP) of the optimal RBF model were 0.933 and 0.528%, while that of the BP model were 0.914 and 0.598%. The RBF model is better than the BP model.

**Key words:** green tea moisture content; quantitative analysis; radial basis function networks (RBF); near infrared spectroscopy (NIR)

茶叶中水分的含量是主要的茶叶品质指标之一, 茶叶含水量在 5% 以下时较耐贮藏、香气变化小. 当含水量超过 6.5% 时, 茶叶陈化加快, 不利

于其保持原有的品质. 水分含量还影响花茶坯的吸附性能. 茶叶水分含量一般采用减重法、Karl-Fisher 等方法测定, 该类方法测定周期长, 同时

【收稿日期】 2005-05-27

【基金项目】 浙江省自然科学基金资助项目(No. 202081)

【作者简介】 刘辉军(1979-), 男, 湖北黄冈人, 硕士研究生. 主要研究方向为计量测试与仪器.

产生的废弃物易对环境产生污染。

近红外光谱定量分析技术具有分析速度快、多组分同时测定、费用较低、易于实现无损和在线检测等特点。近红外定量分析中,常用逐步回归分析(SRA)、主成分回归(PCR)、偏最小二乘法(PLS)等方法来建立样品成分含量值与样品光谱参数之间的线性模型。但近红外吸收光谱中,光谱参数与样品含量化学测定值之间常具有一定的非线性,此时可采用人工神经网络(ANN)建立非线性模型,以提高模型的预测能力。

目前茶叶中一些成分的近红外光谱定量分析模型主要是采用 PLS 等方法建立的线性分析模型。如 Zhang 等<sup>[1,2]</sup>采用 PLS 建立绿茶表儿茶素没食子酸酯(EGCG)等的近红外光谱定量分析模型,夏贤明<sup>[3]</sup>、孙耀国<sup>[4]</sup>等建立了绿茶中氨基酸、咖啡碱和茶多酚含量的近红外定量分析模型。笔者也尝试建立了绿茶水分含量的线性模型,但模型预测效果并不理想。因此本文采用反向传播(BP)和径向基函数(RBF)神经网络建立了绿茶水分含量的近红外定量分析的非线性模型,并对模型进行了比较。

## 1 RBF 网络

### 1.1 向基函数神经元模型

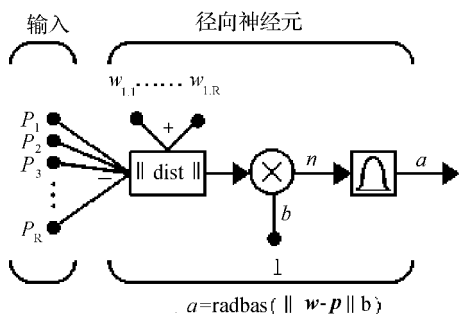


图 1 基函数神经元模型

一个具有  $R$  维输入的基函数神经元模型,如图 1,图中  $\| \text{dist} \|$  模块表示求取输入矢量和权值矢量的距离。此模型中采用高斯函数  $\text{radbas}$  作为径向基神经元的传递函数,其输入为  $n$  维矢量  $p$  和权值矢量  $w$  的距离乘以阈值  $b$ 。

### 1.2 径向基函数网络的结构

图 2 包括两层(隐层和输出层)。网络的输入

维数为  $R$ 、隐层神经元个数为  $S^1$ 、输出个数  $S^2$ 。

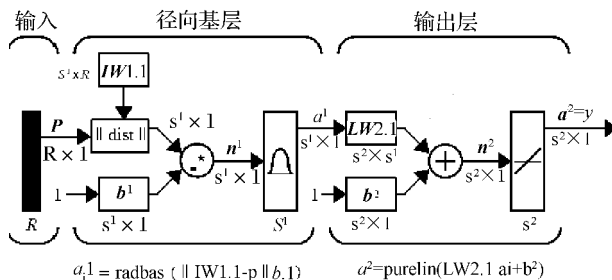


图 2 径向基函数网络的结构图

其中,  $a_i^1$  表示隐层输出矢量  $a^1$  的第  $i$  个元素,  $w_i^1$  表示第  $i$  个隐层神经元的权值矢量。

RBF 网络是以函数逼近理论为基础而构造的一类前向网络,它的学习等价于在多维空间中寻找训练数据的最佳拟合平面。而 BP 网络则是典型的全局逼近网络。由于二者的构造有本质的不同,RBF 与 BP 网络相比,不仅学习速度快,并且网络的函数逼近能力、模式识别与分类能力都优于后者。

## 2 实验部分

### 2.1 样品来源

70 个绿茶样品及其水分含量值,由中国农业科学院茶叶科学研究所提供,水分含量的实测值如表 1,采用减重法测定。样品等级分一级、二级、三级和特级,产地分布浙江、云南、福建、安徽等省,具有较好的代表性。

表 1 实验样品水分含量实测值

样品集类	最大值/%	最小值/%	平均值/%
全部样品(70)	10.2	3.9	5.62
定标集(48)	10.2	3.9	5.58
预测集(22)	9.7	4.9	5.68

### 2.2 近红外光谱仪器

光谱采集仪器为美国 Thermo Nicolet 公司生产的 Nexus870 型傅里叶近红外光谱仪。

绿茶近红外光谱的采集采用漫反射方式,谱区采集范围:780 ~ 2 500 nm,光谱分辨率:4  $\text{cm}^{-1}$ ,室温。每个样品在不同位置进行 8 次扫描,取其平均光谱。同时,在实验过程中,严格控制环境湿度,防止在光谱的采集过程中水分含量的变化。

### 3 模型和结果

#### 3.1 光谱预处理

由于在近红外光谱区水分的吸收光谱重叠严重,吸收峰不明显,并存在很强的干扰,因此在建立模型的过程中对谱区的选择和预处理非常重要.我们选择了光谱信息较丰富的合频以及一级倍频区的光谱,去掉了二级及以上倍频部分,首先进行归一化处理,在采用BP和RBF网络前分别使用了一阶导数(First derivative)、二阶导数(Second derivative)和Detrend等三种预处理方法.

#### 3.2 模型优化

为了简化网络结构,采用先对光谱数据矩阵进

行PLS降维,并利用交互验证法确定最佳主成分数,然后将PLS的最佳主成分作为RBF和BP网络的输入.计算程序通过自行编制的Matlab程序实现,不同的RBF和BP网络模型预测结果如表2.

在模型的优化选择过程中,我们主要考虑了定标集和预测集中水分含量真实值与其计算值之间的相关系数 $r(t)$ 和 $r(p)$ 以及对应的均方根误差RMSEC和RMSEP.

$$\text{RMSEC} = \sqrt{\frac{1}{48} \sum (e_i)^2}, \text{RMSEP} = \sqrt{\frac{1}{22} \sum (e_j)^2}.$$

其中: $e_i$ 表示定标集中第*i*个样品水分含量的真实值和模型计算值之差, $e_j$ 表示定标集中第*j*个样品水分含量的真实值和模型预测值之差.

表2 RBF和BP网络预测的结果

预处理	因子数	BPN				RBF			
		定标集		预测集		定标集		预测集	
		$r(t)$	RMSEC/%	$r(p)$	RMSEP(%)	$r(t)$	RMSEC/%	$r(p)$	RMSEP%
无	6	0.784	0.781	0.637	1.431	0.917	1.458	0.817	1.124
一阶导数	6	0.920	1.002	0.914	0.598	0.966	1.646	0.886	0.646
二阶导数	6	0.901	0.758	0.889	0.613	0.943	0.854	0.883	0.729
Detrend	6	0.916	0.895	0.880	0.707	0.963	0.428	0.933	0.528

由表2可以看出,优化的RBF网络预测模型 $r(p)=0.933$ , $\text{RMSEP}=0.528\%$ ;BP网络预测模型的 $r(p)=0.914$ , $\text{RMSEP}=0.598\%$ ,可见,在茶叶水分的近红外光谱分析中RBF网络模型优于BP网络模型,它们对应的真实值与预测值之间的回归曲线分别如图3(a)和3(b).

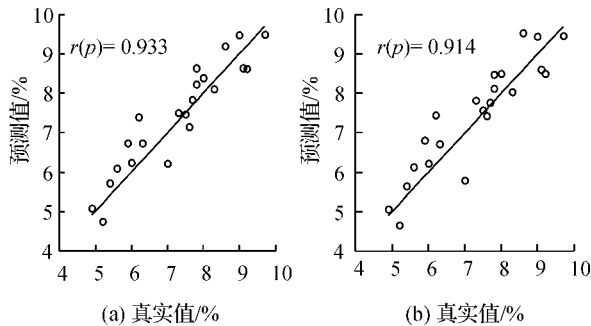


图3 模型的预测回归曲线

### 4 结论

实验结果表明,RBF和BP神经网络模型都有较好的稳健性,模型具有较高的相关系数 $r(t)$

和较低的RMSEC.两模型对预测样品集进行预测,预测回归系数 $r(p)$ 和RMSEP也比较理想.其中,RBF网络模型更优于BP网络模型.同时,当考虑环境温度和湿度的变化时,建立更稳健的绿茶水分预测模型,还需进一步对绿茶的近红外光谱参数与其含水分化学测定值之间的非线性进行研究.

#### 【参考文献】

- [1] LUYP AET 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, 478:303-312.
- [2] ZHANG M H, LUYP AERT J, FERNANDEZ PIENIA J A, et al. Determination of total antioxidant capacity in green tee by near infrared spectroscopy and multivariate calibration[J]. *Talanta*, 2004, 62: 25-35.
- [3] 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.
- [4] 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.