

基于白术 FTIR 的径向基函数神经网络鉴别研究

金文英^{1,2}, 程存归^{1*}, 吴小华¹

1. 浙江师范大学化学与生命科学学院, 浙江 金华 321004

2. 义乌工商职业技术学院, 浙江 义乌 322000

摘要 为了鉴定白术及其伪品, 采用径向基函数神经网络(RBF)分别测试了白术及其伪品的傅里叶变换红外光谱。采用36个样本作训练集, 27个样本作检验集, 用各种模式的BPF进行了监督性训练。当训练目标误差平方和定为0.01时, 各类RBF对训练集中白术样本识别的正确率均为100%, 但对检验集样本识别的结果各不相同, 其识别的正确率与隐含层节点数 S_1 有关。发现当 S_1 较大时, 识别正确率反而下降, 可能此时网络的非线性程度过高, 使其不适用于该类样本集的训练。线性—线性型RBF识别的结果随 S_1 的变化不是很大, 但识别的正确率不高, 基本在85%左右。非线性—线性型RBF识别的结果最佳。当 S_1 为3时, 其识别正确率超过了97%。因此该法可用于简便、快速、准确地识别白术及其伪品。

主题词 傅里叶变换红外光谱; 径向基函数神经网络; 白术

中图分类号: O657.3 **文献标识码:** A **文章编号:** 1000-0593(2006)12-2210-04

引言

白术为菊科 Compositae 植物白术 *Atractylodes macrocephala* Koidz. 的干燥根茎, 为多年生草本, 栽培历史悠久, 主产东阳、磐安一带, 是著名的“浙八味”之一。近年来白术的开发研究日益受到重视, 然而市售白术饮片多有混杂, 在全国药材市场上发现了多种白术的伪品, 最常见的是菊三七及白芍根头片等。对于中药材白术的鉴别, 目前常用的是显微鉴别、理化鉴别及薄层色谱分析等^[1]。傅里叶变换红外光谱法能给出复合体系物质的信息, 用于中药材的鉴别已有报道^[2-6]。但所得到的信息是所含所有物质的红外光谱的叠加, 对于同科同属中药材有时通过比较 FTIR 谱图的差异很难对正品及伪品做出准确的识别。而人工神经网络处理信息不是运用数学方法精密的计算, 它是通过信息样本对神经网络的训练, 使其具有类似人脑的记忆、辨识能力, 完成各种信息处理功能, 其中 BP 网络已有报道^[7-9]。径向基函数神经网络(radial basis Function network, 简称 RBF) 由于采用高斯基函数, 具有表示形式简单、径向对称、光滑性好及便于进行理论分析等特点^[10]。本文利用各种样品不同部位的红外光谱, 通过相同部位的 FTIR 的特征吸收峰位进行径向基函数神经网络训练分析, 从而区别白术的正伪品。实验结果表明采用人工神经网络技术可鉴别白术及其伪品。

1 RBF 模型 原理部分

径向基函数(radial basis function, RBF)神经网络由3层组成, 其结构如图1所示。输入层结点只传递输入信号到隐层, 隐层结点由象高斯函数那样的辐射状作用函数构成, 而输出层函数通常是简单的线性函数。

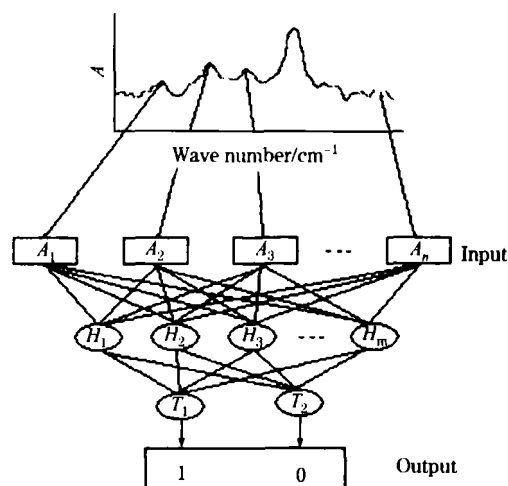


Fig 1 The pattern recognition of radial basis function

收稿日期: 2005-09-06, 修订日期: 2005-11-16

基金项目: 浙江省自然科学基金项目(301468)资助

作者简介: 金文英, 女, 1975年生, 义乌工商职业技术学院计算机工程系讲师 * 通讯联系人

隐层结点中的作用函数(基函数)对输入信号将在局部产生响应,也就是说,当输入信号靠近基函数的中央范围时,隐层结点将产生较大的输出,由此看出这种网络具有局部逼近能力,所以径向基函数网络也称为局部感知场网络。

作为基函数的形式,最常用的是高斯函数:

$$R_i(x) = \exp\left[-\frac{\|x - c_i\|^2}{2\sigma_i^2}\right],$$

$$i = 1, 2, \dots, m$$

其中 x (图 1 中即为吸光度 A 值) 是 n 维输入向量; c_i 是第 i 个基函数的中心, 与 x 具有相同维数的向量, a_i 是第 i 个感知的变量(可以自由选择参数), 它决定了该函数围绕中心点的宽度; m 是感知单元的个数。 $\|x - c_i\|$ 是向量 $x - c_i$ 的范数, 它通常表示 x 和 c_i 之间的距离, $R_i(x)$ 在 c_i 处有一个一维的最大值, 随着 $\|x - c_i\|$ 增大, $R_i(x)$ 迅速衰减到零。对于给定的输入 $x \in R_n$ 只有一小部分靠近 x 的中心被激活。

从图 1 可以看出, 输入层实现 $x \rightarrow R_i(x)$ 的非线性映射, 输出层实现从 $R_i(x)$ 到 y_k 的线性映射, 即

$$y_k = \sum_{i=1}^m W_{ik} R_i(x),$$

$$k = 1, 2, \dots, p$$

其中 p 是输出结点数。

其连接权的学习修正仍可采用 BP 算法。由于 $R_i(x)$ 是高斯函数, 因而对任意 x 均有 $R_i(x) > 0$, 从而失去局部调整的权值的优点, 而事实上, 当 x 远离 c_i 时 $R_i(x)$ 值已经非常的小, 因此可以作为零对待。因此实际上只当 $R_i(x)$ 大于某一数值(例如 0.05)时才对相应的权值 W_{ij} 进行修改。经过这样的处理后 RBF 网络也同样具备局部逼近网络学习收敛快的优点。同时这样近似处理, 可在一定程度上克服跟高斯函数不具备紧密性的缺点。

考虑到提高网络精度和减少隐层结点数, 也可以将网络基函数改成多变量正态密度函数

$$R_i(x) = \exp\left[-\frac{1}{2}(x - c_i)^T k (x - c_i)\right]$$

其中 $k = E[(x - c_i)(x - c_i)^T]^{-1}$ 是输入协方差阵的逆。

2 实验部分

2.1 材料与仪器

材料: 正品白术采自浙江磐安药材市场, 为菊科 *Compositae* 植物白术 *Atractylodes macrocephala* Koidz 的干燥根茎。菊三七为菊科植物菊三七 *Gynura segetum* (Lour.) Merr 的干燥根茎; 白芍根头片为毛茛科植物芍药 *Paeonia lactiflora* Pall 的干燥根头部的切片, 均由中国药品生物制品鉴定所中药标本馆提供。所有样品均经浙江省金华市药检所中药科及浙江师范大学植物学教研室鉴定。

仪器: 美国 NICOLET 公司生产的 NEXUS 670 型傅里叶变换红外光谱仪, DTGS 检测器, OMNIC E.S.P. 5.1 智能操作软件, OMNI 采样器, 光谱范围 4000~650 cm^{-1} , 分辨率 4 cm^{-1} , 扫描累加次数 32 次。

2.2 实验方法

采用单面刀分别切取样品不同部位置于红外光谱仪的

OMNI 采样器中的锗晶片与校正压力装置之间, 按照测试条件直接测定样品的傅里叶变换红外光谱(横坐标为波数, 纵坐标为吸光度)。

3 结果与讨论

3.1 白术及其伪品木质部的 FTIR 比较

由于根茎类中药材的有效药用成分大多集中在木质部, 故本文采用木质部作为 FTIR 测定的部位。图 2 为正品白术及其伪品菊三七、白芍根头片木质部的红外光谱。由图中可看出三者主要区别是在包括指纹区在内的 1700~700 cm^{-1} 范围及饱和 C-H 键的伸缩振动区域。

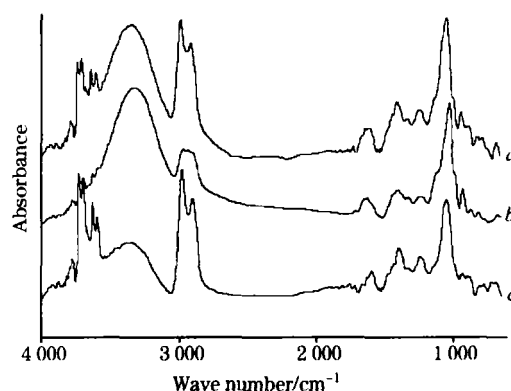


Fig 2 FTIR of xylem of (a) *Atractylodes macrocephala* Koidz; (b) *Gynura segetum* (Lour.) Merr; (c) *Paeonia lactiflora* Pall

3.2 特征吸光度值的提取及数据预处理

从白术及其伪品木质部典型的红外光谱图中可看出, 在 1700~700 cm^{-1} 范围内, 其吸收峰位置和吸收峰强度均有较明显的差异, 这一差异为白术的识别奠定了一定的数学基础。因此本文在 1700~700 cm^{-1} 之间, 每隔 8 cm^{-1} , 选取一个红外吸光度值 A_i , 共取 41 个数据点, 这 41 个吸光度值分别可反映出不同白术的特征吸收, 作为 RBF 的输入值。

在 RBF 训练前, 合理地数据进行预处理是非常重要的, 合理的数据可改善网络性能, 缩短收敛时间。为此, 对原始数据作以下处理: 对吸光度 A_i 进行归一化处理, 归一化公式为 $A_i = (A_i - A_{\min}) / (A_{\max} - A_{\min})$, 其中 $A_{\max}$ 为同一光谱图中所选取的吸光度值中的最大吸光度值, $A_{\min}$ 为最小吸光度值; 其目的是缩小样本在测定时所造成红外吸光度的差异, 使样本数据具有一致性。

3.3 RBF 的设计

MATLAB 是 MathWorks 公司推出的一套高性能可视化数值计算软件, 十分适用于科技计算。本研究基于 MATLAB 的神经网络工具箱中的 Solverb 函数对网络进行编辑, 设计出了白术的 RBF 识别程序, 其主要语句及训练参数如下, 其中 Minmaxp 为 A' 中每一行的极小和极大值构成的 $n \times 2$ 阶矩阵; S_1 为隐含层节点数; T_{F1} , T_{F2} 分别为第 1 和第 2 层转换函数的代号, 其具体形式可为非线性函数 Tansig 和线性函数 Sigmoid 两种。

```
net = newff ( minmaxp , [ S12 ], { TF1 TF2 } ) ;
% 创建 41 × S1 × 2 型的三层 RBF;
net.trainparam.epochs = 100;
% 设置最大训练批次数为 100。
net.trainparam.goal = 0.01;
% 设置训练目标误差平方和为 0.01。
net.trainparam.show = 5;
% 每隔 5 个训练批次显示一次训练结果。
net = train(net, A', T')
```

% 用 Levenberg-Marquardt 学习规则训练网络。

3.4 识别结果

本研究采用了 3 种不同类型的人工神经网络, 即非线性-线性、线性-线性、非线性-非线性。采用 36 个样本作训练集, 27 个样本作检验集, 分别对不同模式的 RBF 进行了监督性训练。当训练目标误差平方和定为 0.01 时, 各类 RBF 对训练集中白术样本识别的正确率均为 100%, 但对检验集

样本识别的结果均不一样, 并且识别的正确率与隐含层节点数有关。采用非线性-非线性型 RBF 识别, 当隐含层节点数较多时, 识别正确率不是升高而是下降; 采用线性-线性型 RBF 识别时, 结果随 S_1 的变化不是很大, 但识别的正确率不是很高, 基本在 85% 左右; 采用非线性-线性型 RBF 识别时, 该类型 RBF 识别效果最佳, 特别是当隐含层节点数为 3 时, 识别正确率达到了 97%。

4 结 论

采用傅里叶变换红外光谱直接测定法可以直接快速地测定植物样品的傅里叶变换红外光谱, 应用 3 种不同模式的人工神经网络对白术的红外光谱进行了识别, 从监督集识别结果来看, 隐含层节点数为 3 的非线性-线性型人工神经网络的识别能力最强, 其识别正确率可达 97%。该法具有简便、快速、准确等特点。

参 考 文 献

- [1] The Pharmacopoeia Committee of the People's Republic of China(国家药典委员会). Chinese Pharmacopoeia(中国药典)(2005, Vol. I). Beijing: Chemical Industry Publishing House(北京: 化学工业出版社), 2005. 68.
- [2] CHENG Cui gui, SUN Cui rong, PAN Yuan jiang(程存归, 孙翠荣, 潘远江). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(9): 1055.
- [3] CHENG Cui gui(程存归). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2003, 23(2): 282.
- [4] CHENG Cui gui, RUAN Yong ming, LI Bing lan(程存归, 阮永明, 李冰岚). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(11): 1355.
- [5] CHENG Cui gui, YING Tao kai(程存归, 应桃开). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2005, 25(1): 36.
- [6] CHENG Cui gui, GUO Shui liang, CHEN Jia hua(程存归, 郭水良, 陈建华). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2002, 22(6): 954.
- [7] TANG Yan feng, ZHANG Zhuo yong, FAN Guo qiang(汤彦丰, 张卓勇, 范国强). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2004, 24(11): 1348.
- [8] ZHANG Zhuo yong, ZHOU Hua lan, LIU Si dong(张卓勇, 周化岚, 刘思东). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2002, 22(2): 341.
- [9] XU Yong qun, SUN Si qin, ZHOU Qun, et al(徐永群, 孙素琴, 周群, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2002, 22(6): 945.
- [10] WEN Xin, ZHOU Lu, WANG Dai li, et al(闻新, 周露, 王丹力, 等). Neural Network Application Design for MATLAB (MATLAB 神经网络应用设计). Beijing: Science Press(北京: 科学出版社), 2000. 244.

Identification of Rhizoma Atractylodes Based on FTIR Spectra and Radial Basis Function Network

JIN Wei ying^{1, 2}, CHENG Cui gui^{1*}, WU Xiao hua¹

1. College of Chemistry and Life Science, Zhejiang Normal University, Jinhua 321004, China

2. Department of Computer Science and Engineering, Yiwu Industrial and Commercial College, Yiwu 322000, China

Abstract In order to recognize the *atractylodes macrocephala* Koidz. (rhizoma atractylodes) and its confusable varieties, three kinds of models of radial basis function network(RBF), nonlinear linear, linear linear, and nonlinear nonlinear model, were used combined with their Fourier transform infrared spectra (FTIR). Rhizoma atractylodes models were collected by Fourier transform infrared spectra, 36 samples were gathered as a training target, and 27 samples as a test set, then their supervision training was performed using three models each. When the summation of error square of the training target was selected as 0.01, the

correct rate for recognition of Fourier transform infrared spectra using each RBF was 100% for the training set, but was different for the test set, which depended on the number of mode in hidden layer, S_1 . It was found that with the increase of S_1 , the correct rate would decrease oppositely. This may be caused by the high degree of the nonlinearity of the networks, so that the models of networks were not fit for the training of this kind of sample set. When using linear linear model of RBF, the correct rate varied with S_1 to some extent, but was generally about 85%. Recognizing ability obtained using nonlinear linear model of RBF was the best. Its correct rate of recognition was > 97%. When $S_1=3$, and so this method can be used to recognize *atractylodes macrocephala* Koidz. (rhizoma atractylodes) and its confusable varieties simply, rapidly and accurately.

Keywords FT IR; Radial basis function network(RBF); *Atractylodes macrocephala* Koidz. (rhizoma atractylodes)

(Received Sep. 6, 2005; accepted Nov. 16, 2005)

* Corresponding author

《中国学术期刊文摘》中文版和英文版 2007 年征订启事

《中国学术期刊文摘》分中文版(简称 CSAC)和英文版(简称 CSAE)两种,各自收录了我国高水平学术期刊中基础科学、医学、农业科学和工程技术领域约 40 个学科的论文文摘,全景展现我国的科研成果与进展。

作为综合性科技类检索刊物,《中国学术期刊文摘》致力于将我国科学技术各领域的原创性学术成果全面、快速地向科技工作者交流、传播,其中 CSAE 是我国第一份综合性英文版科技类学术检索刊物。

《中国学术期刊文摘》由中国科学技术协会主管,科技导报社主办并负责编辑、出版、发行,对科研单位、高等院校、图书馆以及广大科技工作者检索和了解我国的科技研究成果、学术研究动向具有重要的参考价值。

《中国学术期刊文摘(中文版)》刊号为 CN 11-3501/N, ISSN 1005-8923, 2007 年为半月刊,大 16 开,国内定价 38.00 元/册,全年定价 912 元,邮发代号: 82-707。

《中国学术期刊文摘(英文版)》刊号为 CN 11-5411/N, ISSN 1673-4084, 2007 年改为月刊,大 16 开,国内定价 15.00 元/册,全年定价 180 元,邮发代号: 80-487。

欢迎广大科技工作者、科研单位、高等院校、图书馆订阅。

通讯地址: 北京市海淀区学院南路 86 号科技导报社(邮编 100081)

联系电话: 010-62103122

联系人: 姚玉琴

征订信箱: wzjb@cast.org.cn

单位主页: <http://www.csac.org.cn>

户 名: 科技导报社

账 号: 0200001409089017271

开户银行: 工商银行百万庄支行