第26卷，第1期
2006年1月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 26, No. 1, pp51-53
January, 2006

人工神经网络对 VOCs 的自动识别

刘丙萍[1,2]，李 燕[1*]，张 琳[1]，张黎明[1]，王晓斐[1,3]，王俊德[1]

1. 南京理工大学现代光谱研究室，江苏 南京 210014
2. 曲阜师范大学化学科学学院，山东 曲阜 273165
3. 南京大学化工学院，江苏 南京 210093

摘 要 利用人工神经网络(ANN)对严重混叠的傅里叶变换红外光谱图进行了定性和定量解析。通过大量模拟数据训练神经网络后，引用了新的评价标准——逼近度来选择最优网络模型。利用此优化网络对两类光谱图进行了解析，考察了网络的泛化能力。结果表明：该网络不仅能够对两组分同时存在时的样本进行准确解析，而且对于未知单组分光谱图，也能够进行准确鉴别和定量分析。可见，该研究为人工神经网络在单组分和多组分未知物的定性和定量分析方面提供了一种新思路。

主题词 傅里叶变换红外光谱；人工神经网络；多组分分析；未知物鉴定
中图分类号：O644 **文献标识码**：A **文章编号**：1000-0593(2006)01-0051-03

引 言

随着人们对环境质量的日益重视，对环境大气监测技术的要求也越来越高。理想的监测技术应该能用一台仪器同时监测出多种化合物，并且灵敏度高、选择性好，且能提供实时自动监测。开路傅里叶变换红外光谱(OP-FTIR)技术就满足其中的大多数要求，并被认为是环境大气监测中很有实力的技术[1]。但 OP-FTIR 技术测得的红外光谱图解析复杂，这一点严重限制了该技术的广泛应用[2]。

利用傅里叶变换红外光谱对大气中有害物质进行定量分析时，通常借助化学计量学方法[3-5]，对各种污染物的浓度进行同时定量测定。近年来，人工神经网络作为一种新的化学计量学方法得到了广大科学家和光谱学家的重视，特别是多层反向前馈网络(BP-ANN)，具有很强的非线性建模能力，被认为是一种新型的多元非线性校正方法。

目前，人工神经网络在红外光谱中的主要应用是对已知谱图进行定量解析[6-8]。本研究采用不同形式的训练样本建立网络，通过新的评价标准——逼近度确定最优网络模型，利用该网络不仅对已知谱图进行了定量解析，而且对单组分未知谱图进行了鉴别和定量分析，均得到了比较准确的结果。

1 基本理论

1.1 BP 网络的基本原理

多层反向前馈网络中应用最广泛的是三层 BP 网络，它包括输入层、隐含层和输出层，其基本原理已有许多文献进行了详细阐述[9-11]，本文仅作简单介绍。网络的输入信号为红外光谱中不同波数点的吸光度值，输出信号为各组分的浓度值，隐含层和输出层的传递函数均为双曲正切函数(tgh)，表达式为

$$\mathrm{net}_j = f_j(x) = \frac{2}{1+\exp(-2x)} - 1$$

其中：x 为输入信号。

定义误差函数为均方误差(Mean square error)，表达式为

$$MSE = \frac{1}{N_p}\sum_{p=1}^{N_p}\sum_{i=1}^{N_o}(o_i - t_i)^2$$

其中：N_p 表示光谱个数，N_o 代表输出神经元数，o_i 表示第 i 个神经元的网络输出值，t_i 表示第 i 个神经元的目标输出。由于 MSE 是权值矢量的二次函数，因此本研究利用最速梯度下降准则调节权值，使 MSE 达到最小。

1.2 网络的评价标准

人工神经网络模型的最优标准一般是训练集的预测误差最小，但这样极易产生过拟合现象，即网络过度逼近训练样本，不仅学习了训练样本的共性，而且也很好地学习了样本的个性，导致了网络对预测样本的预测能力较差。也有改用以监控集预测误差最小为标准的，但该法忽略了训练集的预测结果，为此，本研究引用了新的评价标准——逼近度[12]。计算公式如下。

定义逼近误差 e_a

$$e_a = \left(\frac{n_1}{n}\right)e_1 + \left(\frac{n_2}{n}\right)e_2 + |\ e_1 - e_2\ |$$

收稿日期：2004-09-16，**修订日期**：2005-01-31
基金项目：国家自然科学基金(20175008)，中国博士后科学基金和南京理工大学青年学者基金(Njust200303)资助项目
作者简介：刘丙萍，女，1979 年生，南京理工大学化工学院硕士研究生 * 通讯联系人

式中，e_1 为训练集标准预测误差，e_2 为监控集标准预测误差。n_1 为训练集样本点数，n_2 是监控集样本点数，n 是已知样本点数。逼近度 d_a 为

$$d_a = \frac{100}{e_a}$$

这里以逼近度为标准进行选优，综合考虑了训练集和监控集的预测结果，显然比较合理。

2 实验部分

2.1 运行软件和硬件

所用人工神经网络算法程序由 Matlab 6.5 编写，网络由 Newff 函数创建，Traingdm 函数训练，Sim 函数仿真，程序在 PC 机上 Windows 操作系统下运行。

2.2 数据的准备

本文所用的气体 FTIR 光谱数据取自美国环境保护协会标准谱库(EPA)。所选物质是苯乙烯和 1,3-丁二烯，其特征吸收峰分别为 1 499，909，695 cm^{-1} 和 1 596，1 014，908 cm^{-1}。本研究共合成样本 51 组，其中两组分的混合光谱 25 组，苯乙烯和 1,3-丁二烯的单组分光谱图各 13 组。

2.3 网络的建立

51 个模拟样本分为两组，其中 35 个作为训练样本，包括 15 个两组分混合光谱和两种单组分光谱各 10 个。16 个作为监控样本，包括 10 个两组分混合光谱和两种单组分光谱各 3 个。网络训练时，数据的取点方式为特征吸收峰处取点，文献[6]已证明这种方法效果较好。

3 结果和讨论

3.1 隐含层的节点数

本研究通过比较不同节点数时神经网络的逼近度来确定隐含层节点数。图 1 表示了隐含层节点数在 2～13 之间时所对应苯乙烯的标准预测误差和逼近度。从图 1 可知，当隐含层节点数为 5 时，网络的标准预测误差最小，逼近度最大。故本研究取隐含层节点数为 5。

3.2 学习速率 *Ir* 和动量因子 *mc*

为确定学习速率，在 0～2 之间每隔 0.1 取一个数据进行实验，结果发现，当学习速率为 1.8 时，神经网络的性能最佳，逼近度最大。此后，在 0～1 之间选取了不同的动量因子进行实验，结果显示，动量因子为0.45时逼近度达到最大，网络性能最好。

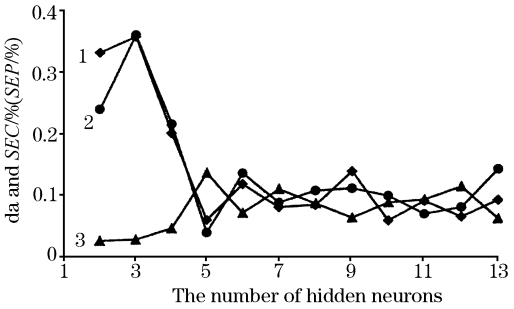

Fig. 1 Effect of hidden neurons on network

1, SEC/% of styrene; 2, SEP/% of styrene;
3, The degree of approximation(da)

3.3 迭代次数(Epochs)

由经验可知，当预测结果达到要求后，增加迭代次数对网络性能的影响不是很大，只会延长计算时间，并且有时迭代次数太多反而会造成神经网络的过度训练，导致结果产生较大偏差。图 2 为迭代次数与逼近度和苯乙烯标准预测误差之间的关系曲线。

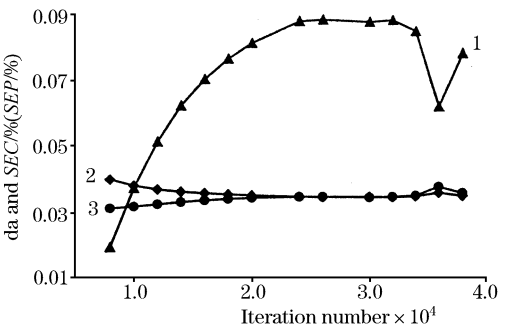

Fig. 2 Effect of different iteration numbers

1, The degree of approximation(da);
2, SEC/% of styrene; 3, SEP/% of styrene

从图中可以看出，当迭代次数大于 26 000 时，苯乙烯的标准预测误差变化不是很大，而且当迭代次数达到 34 000 时，预测误差反而增加，逼近度明显降低，说明网络存在过拟合现象。因此，本实验的最佳迭代次数确定为 26 000。

3.4 网络的应用

为考察网络的泛化能力，本研究人工合成了两类光谱，一类包括两个组分，一类包括一个“未知”组分，将其数据输入已训练好的网络进行定量，预测结果见表 1。

Table 1 Results of the prediction samples

Compounds		1	2	3	4	5	6	7	8	%*SEP*
1,3-Butadiene	c_{meas}	8.17	6.49	7.25	9.58	5.91	6.25	0.00	9.52	2.14
	c_{real}	8.00	6.50	7.00	9.50	5.80	6.40	—	9.4	
Styrene	c_{meas}	6.72	9.19	5.92	8.50	7.27	0.00	5.85	0.00	2.99
	c_{real}	6.60	9.20	6.00	8.60	7.50	—	5.60	—	

表中 1～5 表示两组分的混合物，6～8 表示单组分的未知光谱，“——”表示物质不存在，c_{real} 表示物质的真实浓度，c_{meas} 表示预测浓度。由表 1 可知，该算法的预测结果与各组分的实际浓度非常接近，两种物质的标准预测误差(%*SEP*)

分别为 2.14 和 2.99，而且当样本中不包含某组分时，网络对应的输出值是 0.00，说明该物质不存在，与实际情况相符，这充分证明了人工神经网络具有很好的非线性校正能力。由此可以推断，如果训练集中包含更多种组分时，也可以利用该方法，通过增加训练集样本数，来优化网络，以对单组分未知物进行鉴别和定量。

4　结　论

本研究通过大量数据训练网络，以逼近度为评价标准确定了最优模型，并利用两类光谱检验了网络的泛化能力，结果发现：不仅两组分同时存在时能够得到准确结果，而且对于单组分未知光谱图，网络也能进行准确鉴别和定量解析，证明了人工神经网络是一种有效实用的多元非线性校正方法。同时可以推断，当训练集中包含大量组分时，可用同样的方法对谱图进行解析，从而鉴别出单组分未知物并进行定量分析。

参　考　文　献

[1] Marshall T L, Chaffin C T, Hammaker R M. Environ. Sci. Tech., 1994, 28: 224A.

[2] Newman A R. Anal. Chem., 1997, 69: 43A.

[3] Wang Junde, Clench M R, Wang Tianshu, et al. Spectrosc. Lett., 1997, 30: 99.

[4] Gu Binghe, Wang Junde, Wang Lianjun, et al. Spectrosc. Lett., 1998, 31: 1451.

[5] LIU Fang, WANG Jun-de(刘　芳，王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2001, 21(5): 607.

[6] Li Yan, Wang Junde, Chen Zuoru, et al. Anal. Lett., 2001, 34(12): 2203.

[7] Richardson R L, Yang H S, Griffiths P R. Appl. Spectrosc., 1998, 52: 565.

[8] Hadjiiski L, Geladi P, Hopke P. Chem. Intell. Lab. Syst., 1999, 49: 91.

[9] ZHANG Li-ming(张立明). The Model of Artificial Neural Network and Its Application(人工神经网络的模型及其应用). Shanghai: Fudan University Press(上海：复旦大学出版社), 1993. 34.

[10] ZHU Er-yi, YANG Peng-yuan(朱尔一，杨芃原). Chemometrics and Its Application(化学计量学技术及应用). Beijing: Science Press(北京：科学出版社), 2003. 92.

[11] LI Yan, SUN Xiu-yun, WANG Jun-de(李　燕，孙秀云，王俊德). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2000, 20(6): 773.

[12] LIU Ping, LIANG Yi-zeng, WANG Su-guo, et al(刘　平，梁逸曾，王素国，等). Acta Chimica Sinica(化学学报), 1997, 55: 386.

Automated Recognition of VOCs Using Artificial Neural Networks

LIU Bing-ping[1,2], LI Yan[1*], ZHANG Lin[1], ZHANG Li-ming[1], WANG Xiao-fei[1,3], WANG Jun-de[1]

1. Laboratory of Advanced Spectroscopy, Nanjing University of Science and Technology, Nanjing　210014, China

2. Department of Chemistry, Qufu Normal University, Qufu　273165, China

3. Department of Chemistry, Nanjing University, Nanjing　210093, China

Abstract　Quantitative analysis of FTIR spectra, which are seriously overlapped in the spectral bands, was studied by artificial neural networks. The optimum network was chosen by a new criterion, i.e. the degree of approximation. After the network was established, two kinds of spectra were resolved. It was demonstrated that accurate results could be obtained when two components were both included. In addition, the unknown spectrum could be identified and quantified. It was showed that the artificial neural network has excellent non-linear ability of solution. Meanwhile, the method provides an efficient approach to the identification and quantification of the unknown samples.

Keywords　FTIR; Artificial neural network; Multi-component analysis; Identification

(Received Sep. 16, 2004; accepted Jan. 31, 2005)

* Corresponding author