

# 小波包变换广义回归神经网络同时分辨三种有机化合物的重叠光谱

高玲<sup>1</sup>, 李小平<sup>2</sup>, 任守信<sup>1</sup>

1. 内蒙古大学化学化工学院, 内蒙古 呼和浩特 010021

2. 内蒙古产品质量检验所, 内蒙古 呼和浩特 010024

**摘要** 文章结合两种化学计量学技术, 研制了一种小波包变换广义回归神经网络(WPTGRNN)方法, 对光谱严重重叠的三种有机化合物进行同时测定。该法结合小波包变换(WPT)和广义回归神经网络(GRNN)改进了除噪质量和预测能力。通过最佳化, 选择了小波函数、小波包分解水平及GRNN的平滑因子。偏最小二乘(PLS)法用于比较研究, 编制了三个程序(PWPTGRNN, PGRNN和PPLS)进行相关计算。结果表明, WPTGRNN法是成功的且优于GRNN及PLS方法, 与GRNN方法比较所有组分质量浓度的预测值与实际值的相对预测标准误差由4.0%降低为2.3%。

**关键词** 小波包变换; 广义回归神经网络; 同时测定; 有机化合物

中图分类号: O6.04 文献标识码: A DOI: 10.3964/j.issn.1000-0593(2008)10-2392-04

## 引言

具有重叠信号的多种化合物同时测定一直是分析化学难点之一。近年来化学计量学的发展为不经分离解析重叠信号提供了某些可行办法<sup>[1-6]</sup>。小波分析是近年发展起来的良好时频分析工具, 可同时在时域和频域具有良好分辨率<sup>[7-11]</sup>。小波变换可对光谱信号进行多分辨分解, 有效地滤除噪声, 但它只对低频信号进一步分解, 对高频信号不予考虑<sup>[8]</sup>, 从而导致小波时间分辨率高时, 频率分辨率低的缺陷, 应用小波包变换(WPT)<sup>[12-14]</sup>可以克服这一缺陷, 它对上一层的低频部分和高频部分同时进一步分解。

人工神经网络由于其独特的结构和信息处理方式而得到日渐广泛的应用。误差后传多层前馈网络(BPMLFN)是化学中应用最普遍的技术<sup>[15-18]</sup>。但后传(BP)方法因其收敛速度慢, 易陷入局部最小应用上受到限制。径向基函数(RBFN)<sup>[19]</sup>因其局部逼近性能可克服局部最小和改进收敛速率。广义回归神经网络(GRNN)<sup>[20]</sup>是RBFN的一个分支。本文结合WPT和GRNN两种化学计量学技术研制了小波包变换广义回归神经网络(WPTGRNN)方法同时测定光谱严重重叠的对硝基苯胺、 $\alpha$ -萘胺和联苯胺三种有机化合物, 实验结果令人满意。

## 1 方法原理

### 1.1 小波包变换(WPT)

小波包分析是小波分析的重要扩展。与小波分析不同之处在于小波包分析同时分解逼近(Approximations)和细节(Details)两部分, 因此可更为灵活地表示不同类型的信号。一个快速小波包变换原理(FWPT)用于实现小波包变换。这个原理类似于快速离散小波变换(FDWT)。FWPT, FDWT算法和小波包除噪原理已有详述<sup>[1, 12]</sup>。

### 1.2 广义回归神经网络(GRNN)

广义回归神经网络是径向基函数网络(RBFN)的一个分支。每一个GRNN网络由四层组成: 输入层、模式层、求和层和输出层。GRNN结构简图如图1所示, 输入层不处理信息, 只为模式层分配输入信息。输入层与模式层全部相连。模式层中每一神经元代表一个径向基函数。一个通用的高斯核心函数应用于模式层中。

$$H(X) = \exp\left[-\frac{(X-C)^T(X-C)}{2\sigma^2}\right]$$

这里矢量 $X$ 用于放置输入数据, 矢量 $C$ 代表核心函数的中心位置。平滑因子( $\sigma$ )是控制函数平滑性质的一个参数, 求和层神经元的处理过程十分简单。权重 $W_i$ 和模式层神经元的输出信号 $H_i(X)$ 依下式进行点积计算

$$\sum_{i=1}^N W_i H_i(X), N \text{ 为训练}$$

收稿日期: 2007-03-28, 修订日期: 2007-06-29

基金项目: 国家自然科学基金项目(20667002, 60762003)资助

作者简介: 高玲, 女, 1957年生, 内蒙古大学化学化工学院高级工程师 e-mail: lingyuxi@hotmail.com

对的数目。在求和层中  $\sum_{i=1}^N H_i(X)$  经由单个特别神经元计算求得。模式层与求和层亦全部相连。GRNN 的输出层执行实际运算。求和层经常比输出层多一个神经元。GRNN 的标准方程可用下式表示:

$$\hat{y}(X) = \frac{\sum_{i=1}^N W_i \exp\left[-\frac{(X - C_i)^T (X - C_i)}{2\sigma^2}\right]}{\sum_{i=1}^N \exp\left[-\frac{(X - C_i)^T (X - C_i)}{2\sigma^2}\right]}$$

根据算法原理编制三个程序(PWTGRNN, PGRNN 和 PPLS) 执行相关计算。

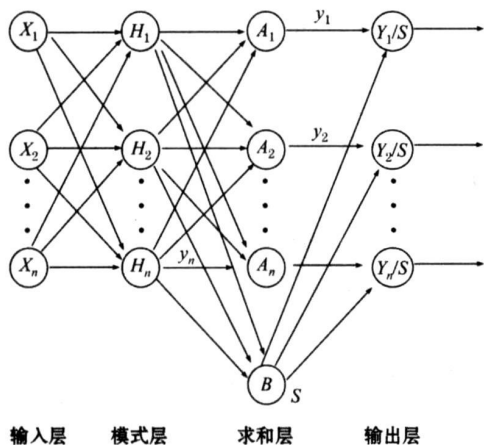


Fig 1 The simple plot of GRNN architecture

## 2 实验部分

### 2.1 仪器和试剂

带有 OPF2 辅助功能的 UV-240 型紫外可见分光光度计及 UV-120 型紫外可见分光光度计(日本岛津公司); pH-S3B 型精密 pH 计(上海雷磁仪器厂); 奔腾 IV 型计算机(中国联想)。2.000 mg·mL<sup>-1</sup> 的对硝基苯胺、 $\alpha$ -萘胺和联苯胺标准储备液均用 1:1 的乙醇溶解配制; 0.200 0 mg·mL<sup>-1</sup> 的工作液用二次蒸馏水经适当稀释制备; pH 6.30 的缓冲液用 0.1 mol·mL<sup>-1</sup> 的乙酸和 0.1 mol·mL<sup>-1</sup> 的乙酸钠溶液按 3:100 的比例混合, 并调节其 pH 6.30。试剂均为分析纯。双蒸馏水和去离子水用于全部试验。

### 2.2 实验方法

以 L<sub>16</sub>(4<sup>5</sup>) 正交表进行正交设计, 分别移取 16 份不同量的对硝基苯胺、 $\alpha$ -萘胺和联苯胺三种标准工作液于 25 mL 容量瓶中, 加入 10.00 mL pH 6.30 的 HAc-NaAc 缓冲溶液, 以水稀释至刻度。用 1 cm 石英比色皿, 以试剂空白作参比, 在 250~450 nm 波长范围内, 以 2 nm 波长间隔测量其吸光度, 并以此数据构筑标准吸光度矩阵  $D$ 。合成样品溶液按上述方法构筑吸光度矩阵  $D_u$ 。

## 3 结果与讨论

### 3.1 $\alpha$ -萘胺和对硝基苯胺和联苯胺的吸收光谱

以试剂空白为参比, 测得对硝基苯胺和  $\alpha$ -萘胺和联苯胺的吸收光谱曲线。如图 2 所示, 在 250~450 nm 波长范围内每种化合物均有较强的吸收峰, 并且吸收光谱严重重叠, 其中对硝基苯胺的最大吸收波长为 380 nm,  $\alpha$ -萘胺的最大吸收波长为 302 nm, 联苯胺的最大吸收波长为 278 nm。

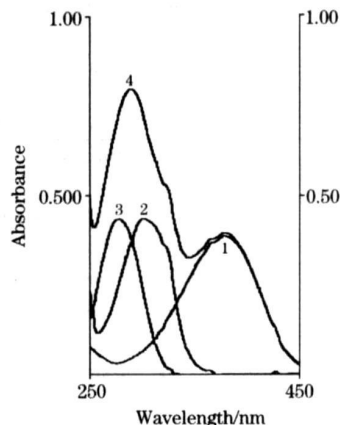


Fig 2 Absorbance spectra of samples

- 1: 对硝基苯胺(4.000  $\mu\text{g} \cdot \text{mL}^{-1}$ );
- 2:  $\alpha$ -萘胺(12.00  $\mu\text{g} \cdot \text{mL}^{-1}$ );
- 3: 联苯胺(3.200  $\mu\text{g} \cdot \text{mL}^{-1}$ );
- 4: 同浓度三种化合物的混合液

### 3.2 小波包变换及小波包除噪

为探讨除噪效果, 在原始吸收光谱的平均信号中加入 1%, 2%, 5% 的随机噪声, 经小波包除噪后的结果见图 3, 其第 1 列为原始光谱、重构光谱和它们之间的差别, 由此可见, 重构光谱和原始光谱相似, 没有边缘效应和信息的丢失; 第 2~4 列为加入随机噪声后的原始光谱、重构光谱和它们的差别, 由此可见, 当随机噪声小于等于 2% 时, 重构光谱和原始光谱相似, 不会引起信息的丢失; 从第 4 列可看出, 当随机噪声为 5% 时, 与原始光谱比重构光谱显示部分失真。

### 3.3 广义回归神经网络(GRNN)

GRNN 是基于非线性回归理论的前向神经网络, 它是径

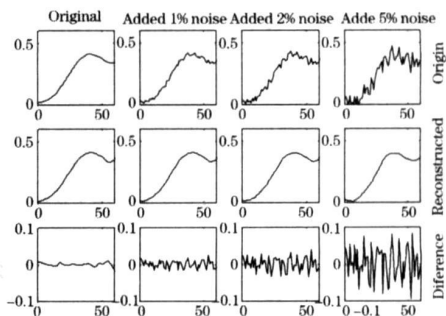


Fig 3 Original and reconstructed spectra as well as their estimated difference by means of WP denoising for different noise added

- 排: (1) 原始光谱
- (2) 重构光谱
- (3) 差值
- 列: (1) 原始平均光谱
- (2) 加入 1% 噪声
- (3) 加入 2% 噪声
- (4) 加入 5% 噪声

向基函数网络(RBFN)的一个分支。训练数据是简单的读入网络的输入层。网络的中心仅被训练数据矢量确定,不需要最小二乘和遗传原理等无监督学习过程。模式层和求和层之间的权重仅和目标矢量相同,所以输出是目标值的简单权重和。因此 GRNN 方法不需要迭代训练手续。GRNN 唯一可调参数是高斯核心函数的宽度,它可控制此高斯函数的平滑(Smoothing)性质,故又称为平滑因子( $\sigma$ ),  $\sigma$  的最佳化对

GRNN 方法性能优劣最为关键。选取从 0.01 到 100 的平滑因子进行试验,计算所有组分的预测标准误差(SEP)<sup>[8]</sup>,当  $\sigma = 1.0$  时,SEP 值最低。当  $\sigma$  小于 0.05 时,此网络完全不能工作。应用 GRNN 方法测量对硝基苯胺、 $\alpha$ -萘胺和联苯胺的浓度,其实际浓度和百分回收列于表 1,SEP 和相对预测标准误差(RSEP)<sup>[8]</sup>列于表 2。

Table 1 Actual concentration and recovery of the synthetic samples

Sample No.	Actual concentration / ( $\mu\text{g} \cdot \text{mL}^{-1}$ )			Recovery / %								
	I	II	III	WPTGRNN			GRNN			PLS		
				I	II	III	I	II	III	I	II	III
1	1.600	3.200	0.800	92.2	101.1	111.1	112.1	79.9	156.3	108.4	119.8	103.6
2	1.600	12.00	4.000	86.5	99.0	108.3	79.3	98.2	113.7	96.7	106.6	97.8
3	1.600	20.00	6.400	84.1	99.3	106.3	77.6	98.7	109.7	99.0	99.2	99.9
4	4.000	3.200	4.000	99.5	100.2	100.4	98.3	102.1	100.0	96.0	82.2	101.3
5	4.000	12.00	6.400	91.7	98.2	108.5	103.6	93.5	110.0	99.4	104.8	97.3
6	4.000	20.00	0.800	94.9	100.2	121.5	100.4	99.9	101.2	100.7	100.6	112.4
7	6.400	3.200	6.400	99.9	100.1	100.1	98.3	105.4	99.1	100.3	71.1	99.9
8	6.400	20.00	4.000	99.7	100.0	100.3	100.9	98.8	104.6	101.1	97.7	102.4

I: 对硝基苯胺; II:  $\alpha$ -萘胺; III: 联苯胺

Table 2 SEP and RSEP (%) values for synthetic samples by three methods

Method	SEP/ ( $\mu\text{g} \cdot \text{mL}^{-1}$ )				RSEP / %			
	I	II	III	Total compounds	I	II	III	Total compounds
WPTGRNN	0.30	0.17	0.45	0.20	4.5	0.74	5.9	2.3
GRNN	0.33	0.64	0.67	0.35	4.8	2.9	8.8	4.0
PLS	0.08	0.59	0.09	0.35	2.0	4.3	1.9	4.0

I: 对硝基苯胺; II:  $\alpha$ -萘胺; III: 联苯胺

### 3.4 小波包变换广义回归神经网络(WPTGRNN)

每一个小波函数具有不同的特点。小波函数对某一给定信号是最佳的,不一定对其他类型信号是最好的。因此,选择小波函数对这一方法是至关重要的。我们应用 SEP 和 RSEP 选择最佳小波函数,本实验选取 Db 2, 4, 6, 20; Coiflet 5; Symmlet 4, 8 等 20 种小波函数进行试验,发现 Sym7 小波优于其他函数。分解水平(L)的选择也是重要的,选取不同分解水平进行试验发现  $L = 2$  时体系的 SEP 和 RSEP 最低。其他需要优化的函数为 GRNN 的平滑因子  $\sigma$ 。经最佳化,选择  $\sigma = 1.0$ 。用 WPTGRNN 方法测量合成样中对硝基苯胺、 $\alpha$ -萘胺和联苯胺的浓度,其实际浓度和百分回收率列于表 1,

SEP 和 RSEP 列于表 2。从表 2 可以看出与 GRNN 方法比较,用 WPTGRNN 方法所有组分的 SEP 和 RSEP 分别从  $0.35 \mu\text{g} \cdot \text{mL}^{-1}$  和 4.0% 降低至  $0.20 \mu\text{g} \cdot \text{mL}^{-1}$  和 2.3%。

## 4 结 论

WPTGRNN 方法结合 GRNN 和 WPT 两种技术的优点,提高了除噪能力,克服了通用 BP-MLFN 方法易陷于局部最小所导致的问题。本例中应用 BP-MLFN 方法计算,超过 5 000 次仍不收敛。实验结果表明 WPTGRNN 方法优于 GRNN 及 PLS<sup>[21]</sup> 方法。

## 参 考 文 献

- [1] GAO Ling, SHI Jur xian, REN Shou xin (高玲, 石俊仙, 任守信). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2006, 26(1): 117.
- [2] SHAO Xue guang, WANG Fang, CHEN Da, et al. Anal. Biochem., 2004, 378: 1382.
- [3] NI Yong nian, QI Zheng bao (倪永年, 齐正保). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2006, 26(7): 1364.
- [4] DONG Lin, XU Lu (董林, 许禄). Chemical Journal of Chinese Universities (高等学校化学学报), 2005, 26(6): 1039.
- [5] SHAO Xue guang, WANG Gu o qing, WANG Si fang, et al. Anal. Chem., 2004-76: 5143.
- [6] Dinc E, Ozdemir A, Baleanu D. J. Pham Biomed. Anal., 2005, 37: 569.
- [7] Daubechies I. Communication on Pure and Applied Mathematics, 1988, 41(7): 909.

- [ 8 ] REN Shou xin, GAO Ling. *Talanta*, 2000, 50: 1163.
- [ 9 ] Esteban Diez I, Gonzalez Saiz J M, Pizarro C. *Anal. Chim. Acta*, 2004, 515: 31.
- [ 10 ] Vogt F, Cramer J, Booksh K. *J. Chemometrics*, 2005, 19: 510.
- [ 11 ] Esteban Diez I, Gonzalez Saiz J M, Pizarro C. *Anal. Chim. Acta*, 2005, 544: 89.
- [ 12 ] GAO Ling, REN Shou xin. *Spectrochim. Acta Part A*, 2005, 61: 3013.
- [ 13 ] Sengur A, Turkoglu I, Ince M C. *Expert Systems with Applications*, 2007, 32(2): 527.
- [ 14 ] Zhang S, Mathew J, Ma L, et al. *Mechanical Systems and Signal Processing*, 2005, 19(2): 357.
- [ 15 ] Bishop C M. *Neural Network for Pattern Recognition*. Oxford: Clarendon Press, 1995.
- [ 16 ] Ni Yong nian, Liu Chao, Kokot S. *Anal. Chim. Acta*, 2000, 419: 185.
- [ 17 ] Chan Wing hong, Lee A W M, Liang Yizeng, et al. *Analyst*, 1997, 122: 657.
- [ 18 ] Ni Yong nian, Chen Sou hui, Kokot S. *Anal. Chim. Acta*, 2002, 463: 305.
- [ 19 ] Ros F, Pintore M, Chretien J R, et al. *Chemometr. Intell. Lab. Syst.*, 2007, 87: 26.
- [ 20 ] Specht D. *IEEE Trans Neural Networks*, 1991, 2: 568.
- [ 21 ] Gao Ling, Ren Shou xin. *Journal of Automatic Chemistry*, 1998, 20(6): 179.

## Simultaneous Resolution of Overlapped Spectra of Three Kinds of Organic Compounds Using a Wavelet Packet Transform Based Generalized Regression Neural Network

GAO Ling<sup>1</sup>, LI Xiaoping<sup>2</sup>, REN Shou xin<sup>1</sup>

1. College of Chemistry and Chemical Engineering, Inner Mongolia University, Huhhot 010021, China

2. The Institute of Products Quality Inspection, Inner Mongolia, Huhhot 010024, China

**Abstract** A wavelet packet transform based generalized regression neural network (WPTGRNN) was developed to perform simultaneous spectrophotometric determination of *p*-nitroaniline,  $\alpha$ -naphthylamine and benzidine. This method combines wavelet packet transform (WPT) with generalized regression neural network (GRNN) for improving the quality of noise removal and enhancing the ability of prediction. Wavelet packet representations of signals provided a local time-frequency description and separation ability between information and noise. The quality of noise removal can be further improved by using best basis algorithm and thresholding operation. Generalized regression neural network (GRNN) was applied for overcoming the convergence problem encountered in back propagation training and facilitating nonlinear calculation. The GRNN is also advantageous in that the training process is much faster and without making any assumption about the form of the prediction model. By optimization, the wavelet function, decomposition level and smoothing factor of GRNN were selected. The partial least squares (PLS) method was used for comparative study. PLS method uses both the response and concentration information to enhance its ability of prediction. Three programs, PWPTGRNN, PGRNN and PPLS, were designed to perform relative calculations. Experimental results showed WPTGRNN method to be successful and better than others. Compared with GRNN method, the relative standard errors of all components between the actual and estimated values of mass concentration for WPTGRNN method decreased from 4.0% to 2.3%. Aniline type compounds are widely applied in industries such as chemistry, printing and pharmacy, and are one of the most important raw materials for synthetic medicine, dye, insecticides, polymer and explosives. Aniline type compounds are highly poisonous, and can also cause cancer. Simultaneous determinations of aniline type compounds are very important in environmental and industrial analysis.

**Keywords** Wavelet packet transforms; Generalized regression neural network; Simultaneous determination; Organic compounds

(Received Mar. 28, 2007; accepted Jun. 29, 2007)