

基于径向基函数网络的激光诱导荧光特征 光谱分离算法*

李宏斌 刘文清 张玉钧 丁志群 赵南京 魏庆农 王玉平 杨立书

(中国科学院环境光学与技术重点实验室, 中国科学院安徽光学精密机械研究所, 合肥 230031)

(2004 年 9 月 30 日收到; 2005 年 1 月 11 日收到修改稿)

应用激光诱导荧光技术测量水中溶解有机物 (DOM) 含量, 具有灵敏度高、检测速度快、可遥测等优点, 其中特征荧光光谱的分离在系统中占有十分重要的地位. 在分析激光诱导荧光光谱特征的基础上, 提出了采用径向基函数网络 (RBFN) 分析荧光光谱数据的数学模型, 应用这种模型从荧光光谱中恢复出了激光、拉曼和 DOM 的荧光等光谱分量成分, 从而得到了水中 DOM 的浓度.

关键词: 径向基函数网络, 激光诱导荧光, 溶解有机物

PACC: 8670L, 8670E, 3350

1. 引 言

激光诱导荧光技术用于测量水中溶解有机物 (DOM) 的含量, 具有灵敏度高、检测速度快、非接触式测量、自动化程度高等优点, 其基本原理是当某种荧光物质的稀溶液, 在一定频率和强度的激发光照射下, 溶液的浓度较小时, 则溶液所产生的荧光的强度与溶液中该种荧光物质的浓度成正比^[1]; 在实际测量中, 荧光物质的荧光强度经过归一化处理后得到荧光因子 $F(\lambda_{ex})$, 荧光因子是与溶液浓度相对应的参量, 可以由它计算出溶液浓度的值^[2].

荧光因子的计算是十分关键的问题, 荧光因子是荧光和拉曼的强度之比 I_n/I_R , I_R 代表拉曼光的强度, I_n 是 DOM 所发荧光的强度. 采集到的光谱数据包含有激光光谱、拉曼光谱和 DOM 的荧光光谱, 只有正确的将它们分离出来才能准确计算出荧光因子. 解决问题的关键是采用什么方法分析光谱分量. 文献[3]采用高斯拟合的方法对光谱分量进行分析, 这种方法是把每一个光谱分量都视为高斯线型的, 只是它们各自的幅度、中心位置和半径不同, 然后通

过优化算法得到它们的一组最佳组合, 即得到各分量数据. 这种方法适用于光谱中的各个分量谱的形状都近似于高斯形状的情况, 但是如果谱分量的形状不是高斯型的, 用此法分析的结果就不够准确. 为解决上述问题, 本文提出了基于径向基函数网络 (radial basis function networks, RBFN) 的激光诱导荧光特征光谱分离算法, 这种方法的优点在于每个分量谱的确切形状和类型不用预先知道.

RBFN 具有网络结构简单、学习速度快等优点, 广泛应用于函数逼近、模式识别、预测和控制等领域. 鉴于 RBFN 具有较强的函数逼近能力的优点, 本文将 RBFN 用于激光诱导荧光遥测系统中的光谱数据分析, 对采集到的光谱数据进行分析处理, 并从原始光谱数据中分离出激光光谱、拉曼光谱和 DOM 的荧光光谱. 通过计算得到拉曼光的强度 I_R 和 DOM 所发荧光的强度 I_n , 从而算出与 DOM 浓度相关的荧光因子 $F(\lambda_{ex})$. 文中首先介绍了 RBFN 结构与特点, 然后给出了基于 RBFN 的激光诱导荧光特性光谱分离算法的数学模型, 并根据激光诱导荧光光谱的特点阐述了基函数的确定方法, 最后给出了实验结果, 并对结果进行了讨论.

* 中国科学院知识创新工程重要方向项目 (批准号: KGCX2-SW-111) 资助的课题.

2. RBFN

2.1. RBFN 结构与特点

RBFN 是一种三层前馈神经网络,由输入层、隐含层和输出层组成,其网络结构如图 1 所示.

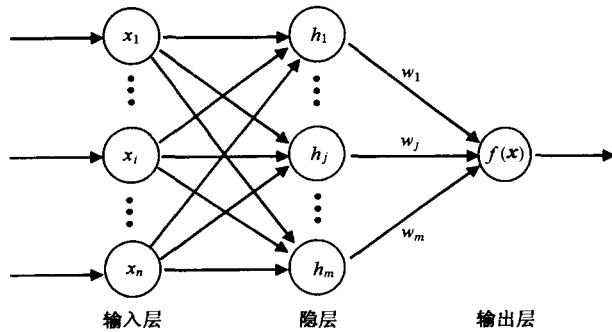


图 1 RBFN 结构图(n 维向量 \mathbf{x} 作为输入, 隐层 m 个节点, 每个节点的传输函数表示为 $h_j(\mathbf{x})$, 输出为 $f(\mathbf{x}) = \sum_{j=1}^m w_j h_j(\mathbf{x})$)

设 H 为具有重建核的 Hilbert 空间, 隐层单元的传输函数 $h_j(\mathbf{x})$ 是 H 的正交归一化基, 则 H 中的任一函数 $f(\mathbf{x})$ 均可由 $h_j(\mathbf{x})$ 来表示^[4], 即

$$f(\mathbf{x}) = \sum_{j=1}^m w_j h_j(\mathbf{x}), \quad (1)$$

w_j 为权值, h_j 通常取高斯函数和柯西函数或者它的变形^[5]. 此类径向基函数 $h(\mathbf{x})$ 是一类特殊的函数, 它们的特征是随着离中心距离的增加, 函数值增大 (或者减小). 文中 $h(\mathbf{x})$ 取高斯函数^[6],

$$h(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{c}\|^2}{r^2}\right), \quad (2)$$

$\|\mathbf{x} - \mathbf{c}\|$ 表示输入向量 \mathbf{x} 与中心向量 \mathbf{c} 间的矢量距离. RBFN 较适合用于函数逼近, 用径向基函数逼近的研究见文献[7].

2.1. RBFN 权值的计算

设有 P 个样本点 $X_i, i = 1, 2, 3, \dots, P$, 根据

$$f(\mathbf{x}) = \sum_{j=1}^m w_j h_j(\mathbf{x}), \text{ 则有方程式}$$

$$\begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \cdots & h_m(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & \cdots & h_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_p) & h_2(\mathbf{x}_p) & \cdots & h_m(\mathbf{x}_p) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_p) \end{bmatrix}. \quad (3)$$

若令

$$\mathbf{H} = \begin{bmatrix} h_1(\mathbf{x}_1) & h_2(\mathbf{x}_1) & \cdots & h_m(\mathbf{x}_1) \\ h_1(\mathbf{x}_2) & h_2(\mathbf{x}_2) & \cdots & h_m(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ h_1(\mathbf{x}_p) & h_2(\mathbf{x}_p) & \cdots & h_m(\mathbf{x}_p) \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_m \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} f(\mathbf{x}_1) \\ f(\mathbf{x}_2) \\ \vdots \\ f(\mathbf{x}_p) \end{bmatrix},$$

则有

$$\mathbf{H}\mathbf{w} = \mathbf{y}. \quad (4)$$

训练集是 $\{(\mathbf{x}_i, \hat{y}_i)\}_{i=1}^p$, 根据最小二乘原理 (least-squares) 使 $E = \sum_{i=1}^p (\hat{y}_i - f(\mathbf{x}_i))^2$ 最小, 可以计算出权值^[5] \mathbf{w} , 表达式为

$$\hat{\mathbf{w}} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \hat{\mathbf{y}}. \quad (5)$$

3. 特征光谱分析

3.1. 荧光光谱数学模型

系统采集到的光谱数据中包含了激光、拉曼和

溶解有机物荧光等分量, 它们每一个都有各自的形状和大小. 光谱总量等于各个分量之和, 其表达式为

$$y(x) = y_L(x) + y_R(x) + y_D(x), \quad (6)$$

其中 y 表示光谱的总量, y_L, y_R 和 y_D 分别表示激光、拉曼和 DOM 荧光分量, x 代表波长, 作为 RBFN 的输入 (输入层只有一个节点).

光谱中的每个分量具有不同的形状和大小, 均可由一定数量的高斯函数来逼近, 其表达式为 $y_L(x) = \mathbf{H}_L(x) \mathbf{w}_L, y_R(x) = \mathbf{H}_R(x) \mathbf{w}_R, y_D(x) = \mathbf{H}_D(x) \mathbf{w}_D$, 代入 (6) 式, y 就可以表示为

$$y(x) = \mathbf{H}_L(x) \mathbf{w}_L + \mathbf{H}_R(x) \mathbf{w}_R + \mathbf{H}_D(x) \mathbf{w}_D, \quad (7)$$

$$y(x) = [H_L(x) \quad H_R(x) \quad H_D(x)] \begin{bmatrix} w_L \\ w_R \\ w_D \end{bmatrix}, \quad (8)$$

所以有 $Hw = y$, 其中

$$H = [H_L(x) \quad H_R(x) \quad H_D(x)], w = \begin{bmatrix} w_L \\ w_R \\ w_D \end{bmatrix}.$$

根据(8)式,将本来用于逼近光谱中各个分量的高斯函数组合到一起得到 H , 这样就可以用 H 中所有的高斯函数来逼近光谱总量曲线 y . 根据公式 $\hat{w} = (H^T H)^{-1} H^T \hat{y}$, 计算出权值 w , 由 w 很容易得到 w_L, w_R 和 w_D , 从而即可得到谱分量 y_L, y_R 和 y_D 的值.

3.2. 基函数的中心和半径

隐层中的基函数 ($h_j(x)$ 选择高斯函数) 的中心和半径是 RBFN 的重要参数, 中心和半径的选取决定了网络的性能. 表达式 $\hat{w} = (H^T H)^{-1} H^T \hat{y}$ 中, \hat{y} 是已知的, 要计算 w 就必须给出 H . 那么如何确定 H 就是解决问题的关键所在, 实际上, 只要确定了高斯函数的中心和半径, 就确定了 H .

选取 H 的过程就是确定 RBFN 结构的过程, 只要 H_L, H_R 和 H_D 选取得当, 就能准确分离出各个谱分量. 在确定 H 时, 主要根据光谱的各个分量本身所具有的结构特征, 分别在各分量光谱所占的波段范围内设置一定数量的高斯函数. 具体的方法如下:

1) H_L, H_R 和 H_D 中的基函数分别用于逼近激光、拉曼和 DOM 的荧光光谱, 它们当中的基函数要能够对各自的光谱均具有较好的逼近性能.

2) H_L, H_R 和 H_D 基函数的中心值分布在其各自对应的光谱所覆盖的波长范围内; 半径小于其对应的光谱谱峰的半高宽度, 半径大小要适中不宜过小; 而且它们当中基函数要有一定的数量, 使其覆盖各自对应的光谱所占的波长范围.

3) 激光与 DOM 荧光、拉曼与 DOM 荧光均有重叠, H_L 中的基函数要与 H_D 中的有区别, 同样 H_R 中的基函数也要与 H_D 中的有区别. 激光和拉曼的光谱都比较窄, 而 DOM 的荧光光谱很宽, 所以 H_D 中基函数的半径要远大于 H_L 和 H_R 中基函数的半

径, 并且要小于 DOM 的谱宽, 这样既保证了 H_D 对 DOM 的荧光光谱具有较好逼近性能, 又能确保这样结构的 RBFN 可以从原始数据中准确分离出激光、拉曼和 DOM 的光谱.

图 2 给出了确定基函数方法的示意图.

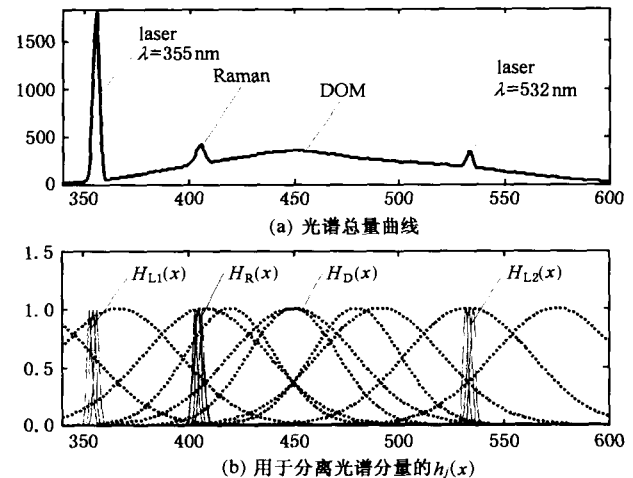


图 2 $h_j(x)$ 的中心和半径选取方法示意图

3.3. 权值非负

根据最小二乘原理, $w = (H^T H)^{-1} H^T y$, w 是方程 $Hw = y$ 的较优解, 但是它并没有考虑到各个分量的方程式的约束条件, 所以其对应的 w_L, w_R 和 w_D 则不一定就是 $y_L = H_L w_L, y_R = H_R w_R, y_D = H_D w_D$ 的一组解, 所以这样得到的各个分量的结果很有可能是错的.

为了解决上述问题, 需对计算 w 的算法加以改进. 解决办法, 对于方程 $Hw = y$, 按照非负最小二乘准则^[8] (nonnegative least-squares) 来求解 w . 这样做, 保证了每一个高斯函数对光谱总量与所有光谱分量的贡献均是非负的, 这与实际情况相符, 具有一定的物理意义. 实际上, 谱线值 (无论是总量还是分量) 都是非负的, 并且谱线中各个分量的谱峰位置不重叠, 而且它们有其各自不同的半径, 只要 H 选取的合适, 采用这种方法就能得到与实际相符的谱分量. 实验结果证明, 采用非负最小二乘准则求解 w , 计算得到的谱分量 y_L, y_R 和 y_D 与实际值接近, 效果很好, 而且比较稳定.

3.4. 荧光因子

荧光因子是计算溶液浓度的十分重要的参量,它是溶液中荧光物质所发荧光的强度与拉曼光的强度的比值^[2],其表达式为

$$F(\lambda_{ex}) = \frac{I_f}{I_R} = \frac{\int_0^{\infty} y_f(x) dx}{\int_0^{\infty} y_R(x) dx}, \quad (9)$$

所以有

$$F_{DOM}(\lambda_{ex}) = \frac{I_{DOM}}{I_R} = \frac{\int_0^{\infty} y_D(x) dx}{\int_0^{\infty} y_R(x) dx} = \frac{\sum_{i=1}^p y_{Di}}{\sum_{i=1}^p y_{Ri}}, \quad (10)$$

I_f 和 I_R 代表光信号的强度, p 表示谱信号的样本点数, y_D 是 DOM 分量, y_R 是拉曼分量.

3.5. 浓度计算

浓度计算及数据处理的工作流程如图 3 所示. 首先, 计算荧光因子, 利用 RBFN 对原始的光谱数据处理, 准确分离出各个光谱分量, 计算出荧光因子; 然后, 进行浓度校准, 利用标准样品, 找出荧光因子与浓度值之间的对应关系, 作出一条浓度校准曲线; 最后, 测待测样品的浓度, 利用 RBFN 分析待测样品的光谱, 计算荧光因子, 再由浓度校准曲线算出浓度值.

荧光因子与浓度之间存在一定的对应关系, 这种对应关系就是浓度校准曲线. 浓度校准, 可以用多项式回归的方法, 也可以用神经网络的方法(比如 RBFN 就具有很好的函数逼近性能), 或者用其他的方法. 一般情况下, 浓度校准曲线的线形不是很复杂, 采用多项式回归的方法就能够基本满足要求.

4. 实验结果与讨论

激光诱导荧光遥测系统的实验装置如图 4 所示. 激光器选用的是法国 QUANTEL 公司 BRILLIANT 系列的 Nd:YAG 激光器, 发出的 1064 nm 波长的光经过三倍频, 得到 355 nm 波长的光作为激发光; 激光的脉冲能量不小于 80 mJ, 脉冲重复频率在 10 Hz.

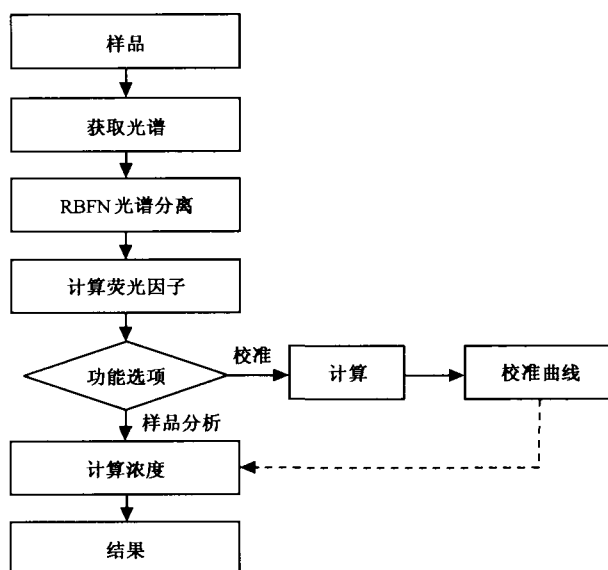


图 3 浓度计算及数据处理的工作流程

由望远镜接收到的光经过一个单色仪分光后, 再由光电倍增管接收. 测量时, 在计算机的控制下, 设定单色仪的波长, 然后控制激光器出光激发, 在激发同时采用外同步的方式对光电倍增管的信号进行采集, 按此步骤以一定的步进进行扫描完成一条谱线的采集过程.

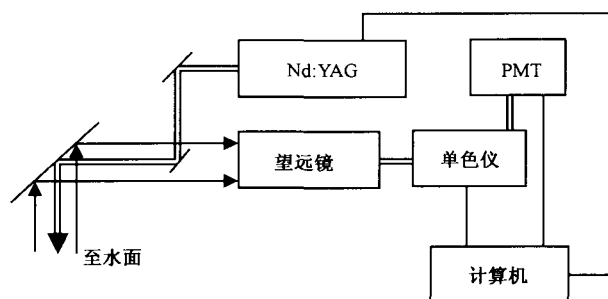


图 4 激光诱导荧光遥测系统实验装置

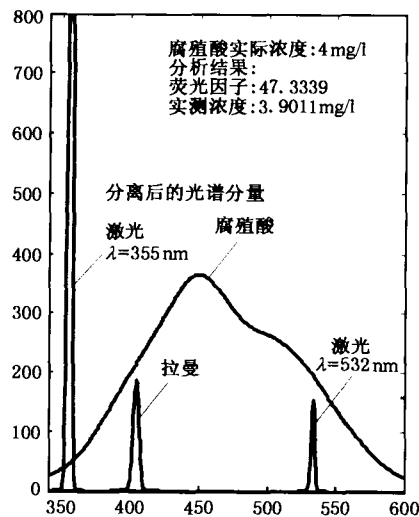
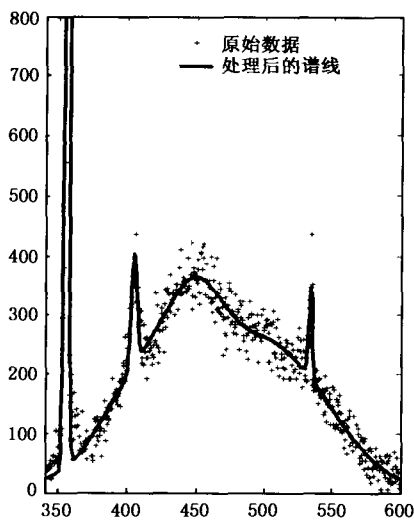
腐殖酸是水中 DOM 中具体代表性的物质, 所以我们选用腐殖酸作为实验样品, 配置不同浓度的腐殖酸溶液, 对其进行了实验. 光谱的扫描步进设为 0.5 nm, 波长的扫描范围从 340 nm 到 599.5 nm. RBFN 参数设置, 在 $\lambda = 355$ nm 附近设定了 3 个高斯函数, 半径均为 1.5 nm; 在 $\lambda = 405$ nm 附近设定了 4 个高斯函数, 半径为 1.5 nm 的 3 个、半径为 3 nm 的 1 个; 在 $\lambda = 450$ nm 附近设定了 6 个高斯函数, 半径

为 83.3 nm 的 3 个、半径为 30 nm 的 3 个;在 $\lambda = 532$ nm 附近设定了 3 个高斯函数,半径为 3.3 nm.

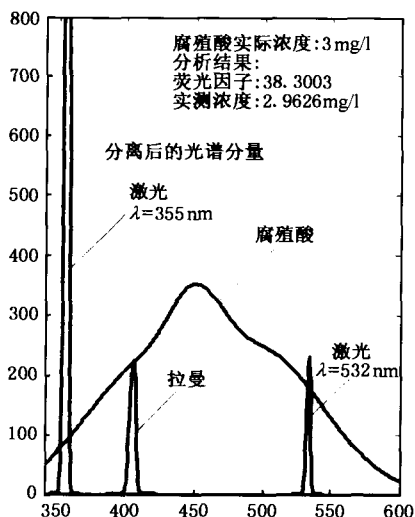
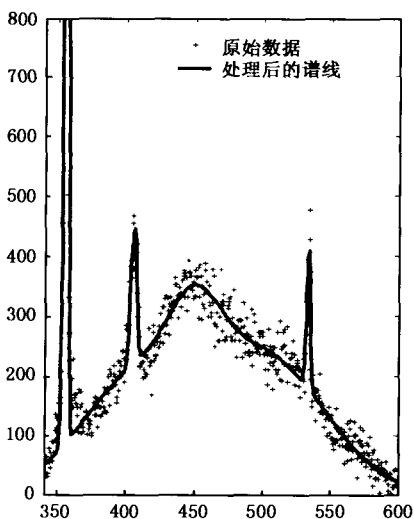
在对腐殖酸溶液进行分析之前,需对本实验系统进行校准.分别配置 10 种不同浓度的腐殖酸溶液(浓度分别为 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5 mg/L)作为校准样品,通过计算求得校准曲线.待系统校准后,再配置一些一定浓度的腐殖酸溶液,作为待测样品.按照图 3 所描述数据处理的流程,分别对四种不同浓度的腐殖酸溶液进行分析,实验结果如图 5 所示.

从图 5 中的原始光谱数据可以看出,实验系统的信号稳定性并不好.造成这种现象的主要原因有两个:一个原因是激光器输出的激光脉冲能量有 $\pm 6\%$ 的误差;另一个原因是实验系统本身的随机噪声.在这样的条件下,四种不同浓度腐殖酸溶液的实测浓度与实际浓度值都很接近,由此说明 RBFN 具有很好的噪声抑制性能、具有很好的函数逼近性能.

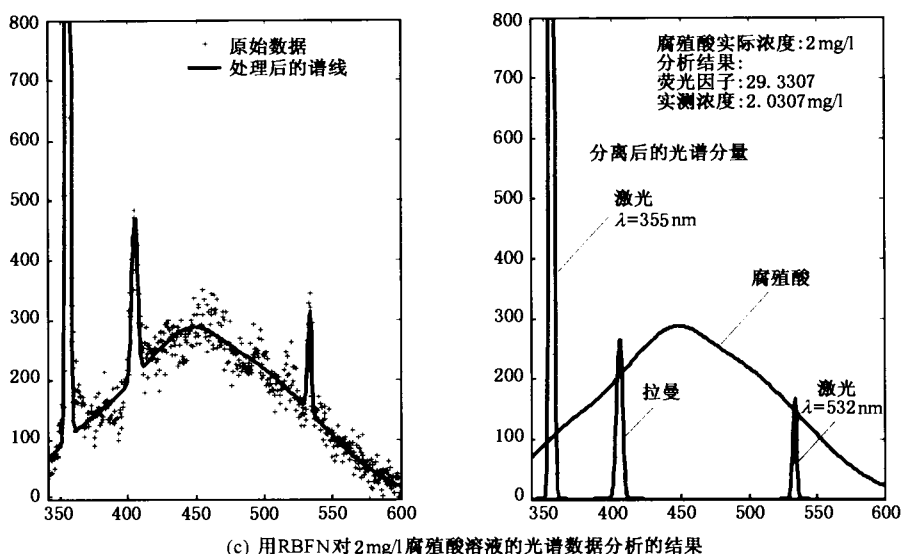
荧光法测水溶液浓度受到一定的限制,主要表现在以下几个方面:



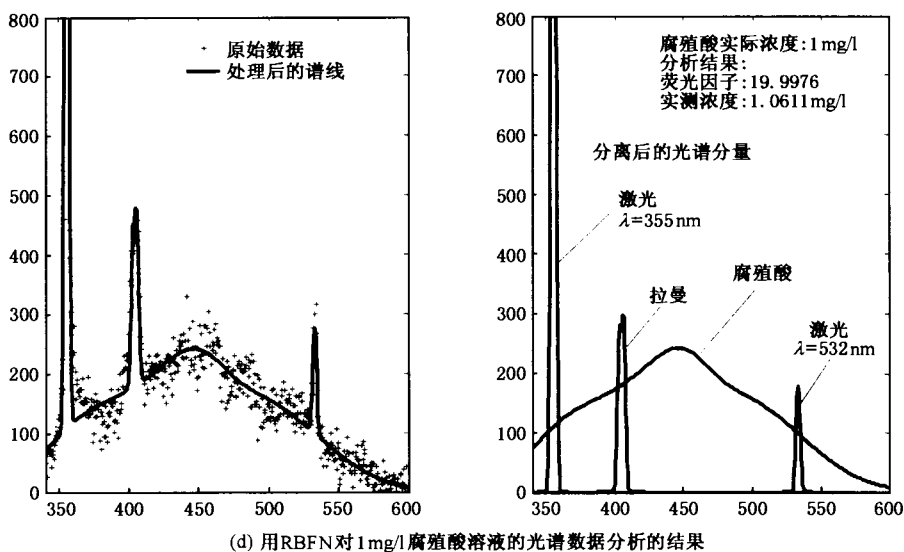
(a) 用RBFN对4 mg/l腐殖酸溶液的光谱数据分析的结果



(b) 用RBFN对3 mg/l腐殖酸溶液的光谱数据分析的结果



(c) 用RBFN对2mg/l腐殖酸溶液的光谱数据分析的结果



(d) 用RBFN对1mg/l腐殖酸溶液的光谱数据分析的结果

图 5 用 RBFN 光谱分析和处理结果

1) 溶液浓度不能过高. 当溶液浓度较低时, 荧光强度随着浓度的增加而增加, 呈线性变化的趋势. 当溶液浓度的增加到一定程度时, 呈非线性变化趋势, 荧光强度不再随着浓度的增加而增大, 就会造成荧光信号的饱和. 用荧光法测浓度时, 首先要保证荧光信号不饱和.

2) 信号采集系统的限制. 在信号采集时, 采集卡本身的动态范围受到一定限制, 浓度较高时, 即使荧光信号没有饱和, 也有可能造成采集卡的信号饱和.

3) 影响浓度测量精度的因素还有整个实验系统

的其他参数、环境因素等.

测量时, 要避免出现以上 1) 和 2) 两种情况的信号饱和, 这两种饱和决定了测量范围的上限; 测量范围的下限主要取决于待测物的荧光效率和系统的信噪比. 本实验系统对腐殖酸溶液测量的浓度范围在 0.4—10 mg/L.

5. 结 论

径向基函数网络用于激光诱导荧光遥测系统中

的光谱数据分析,可以有效的将各个光谱分量从原始数据中分离出来.实验表明,利用此方法得到的光谱分量计算出荧光因子,再通过荧光因子和校准曲线计算 DOM 溶液浓度,计算的浓度值与实际的浓度

很接近,效果比较好.运用 RBFN 进行特征光谱分离,不但能抑制一定的噪声,而且运算速度快,算法复杂度不高、容易实现,具有一定的实用价值.

- [1] Wei F S, Xu X B, Yan J C et al 1997 *Manual of Analytical Method for Water and Waste Water Monitoring* (III) (Beijing: Chinese Environmental Sciences press) p153 (in Chinese) [魏复盛、徐晓白、阎吉昌等 1997 水和废水监测分析方法指南(下册)(北京:中国环境科学出版社出版)第 153 页]
- [2] Drozdowska V, Walczowski W, Hapter R et al 2004 *EARSeLe Proceedings* 3 138
- [3] Drozdowska V 2002 *OCEANOLOGIA* 44 339
- [4] Yan P F and Zhang C S 2002 *Artificial Neural Networks and Simulative Genetic Algorithms* (Beijing: Tsinghua University press) p40—41 (in Chinese) [阎平凡、张长水 2000 神经网络与模拟进化计算(北京:清华大学出版社)第 40—41 页]
- [5] Mark J L Orr 1996 *Introduction to Radial Basis Function Networks* (Edinburgh: Centre for Cognitive Science University of Edinburgh) p9, 11, 12
- [6] Liu D, Ren H P and Kong Z Q 2003 *Acta Phys. Sin.* 52 531 (in Chinese) [刘丁、任海鹏、孔志强 2003 物理学报 52 531]
- [7] PoWell M J D 1992 *Advances in Numerical Analysis* (W. A. Light, ed.: Oxford University Press) p105—210
- [8] Lawson, C L and Hanson R J 1974 *Solving Least-Squares Problems* (New Jersey: Prentice-Hall) Chapter 23 p161

A method of characteristic LIF spectral signatures separation based on radial basis function networks^{*}

Li Hong-Bin Liu Wen-Qing Zhang Yu-Jun Ding Zhi-Qun

Zhao Nan-Jing Wei Qing-Nong Wang Yu-Ping Yang Li-Shu

(Key Laboratory of Environmental Optics and Technology, Anhui Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Hefei 230031, China)

(Received 30 September 2004; revised manuscript received 11 January 2005)

Abstract

The separation of characteristic spectral signatures plays an important role in the laser induced fluorescence (LIF) system used for monitoring dissolved organic matter (DOM) in polluted water, which has advantages of high sensitivity, fast detection, remote measurement, etc. In this paper, a method of characteristic spectral signatures separation based on radial basis function networks (RBFN) will be presented. Using this method, we separated the spectral components of laser, Raman and fluorescence of DOM from the LIF spectral. The concentration of DOM in polluted water can be retrieved from the separated spectrums.

Keywords: radial basis function networks, laser induced fluorescence, dissolved organic matter

PACC: 8670L, 8670E, 3350

^{*} Project supported by the Foundation of Knowledge Innovation of Chinese Academy of Sciences (KG CX2-SW-111).