

文章编号:1004-4213(2010)04-1330-4

基于三维荧光光谱特征分析的油种鉴别技术的研究*

王玉田¹,张艳林^{1,†},王金玉²

(1 燕山大学 测试计量技术及仪器河北省重点实验室,秦皇岛 066004)

(2 中航 577 厂,秦皇岛 066102)

摘要:介绍了运用神经网络进行模式识别的基本原理,将主成分分析法和 BP 神经网络相结合,提出矿物油三维荧光光谱鉴别方案,并进行了系统设计,建立了基本的模型框架.选取矿物油三维荧光光谱的特征参量,组成原始特征向量,采用主成分分析法进行预处理,而后选取主成分运用 BP 神经网络实现油种鉴别.该方法减少了输入变量的维数,消除了各输入变量的相关性,同时简化了网络结构,提高了程序运行的速度.通过实例进行了分析,结果证明该方法有效地实现了矿物油三维荧光光谱的油种鉴别,同时该系统也可用于其它物质的光谱识别技术领域.

关键词:光谱分析;油种鉴别;主成分分析;BP 神经网络

中图分类号: TB99

文献标识码: A

doi: 10.3788/gzxb20103904.1330

0 引言

近年来,工业企业排放的含油污水量不断增加,海上石油运输泄漏事件时有发生,影响了水生生物的正常生存与生长.油类及其分解产物中的多种有毒物质通过食物链进入人体,危害人体健康.污染油的快速检测和分析,鉴别其种类和来源,准确地评估污染程度,是水体环境监测和污染防治的必要手段^[1].每种矿物油样本都有其特有的荧光特征,是鉴别和分析的重要工具,而来源不同、种类繁多的矿物油的三维荧光光谱是多种多样的,难于用简单公式得到种类认定.本文通过对油种的三位荧光光谱进行特征参量提取,采用主成分分析和 BP 神经网络的相结合的方法,完成水环境中污染矿物油的种类鉴别.

1 系统基本原理及其设计

矿物油三维荧光光谱种类鉴别可运用模式识别方法,借助计算机技术,对相同测定条件下的光谱信号进行分析,得出待识别光谱的所属类别.在种类鉴别系统的设计要求该系统不仅具有快速的识别速度、良好的识别效果,而且结构简单、具有良好的扩展性,能够适应系统不断扩充的需要.本文采用基于主成分分析和 BP 神经网络相结合的方法建立三维荧光光谱识别系统.首先对待识别光谱进行特征提取,而后由提取的特征参量经过标准化处理,主成分分析得到独立特征谱,再利用已知标准特征库建立并训

练神经网络,最后运用所建的神经网络对待识别的特征谱进行种类判别,实现未知矿物油荧光光谱的定性分析.在具体实现上还可根据识别样本种类的多少对预分类的特征和神经网络分类的特征进行相应的调整.

矿物油三维荧光光谱实际上是一个二维区域上能量的连续分布,通过对激发波长和发射波长进行扫描得到三维荧光光谱^[2].首先从分析三维荧光光谱的表观统计特征入手,提取矿物油三维荧光光谱的多个统计特征参量,所要求的特征参量需能够表示荧光光谱在荧光强度数值域、在波长空间域、在能量域的分布特征,在能量域的边际累积分布特征.综合考虑,需要提取的特征参量包括平均值、标准差、重心、几何分布散度、相关系数、等效椭圆长轴斜率、峰度和偏度.

主成分分析法从可观测的变量中提取有用信息,舍弃其中无足轻重的部分,尽可能多地保留原变量所包含的信息,同时又用尽可能少的主成分替代原有变量,从而使问题变得简单.主成分分析法的原理^[3]: $X = (X_1, X_2, \dots, X_n)^T$ 是特征参量构成的 n 维向量, Σ 为 X 的协方差矩阵, Σ 的特征值及相应的正交单位化特征向量分别为 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$ 及 e_1, e_2, \dots, e_n , 则 X 的第 i 个主成分 $Y_i = e_i^T X = e_{i1} X_1 + e_{i2} X_2 + \dots + e_{in} X_n$ ($i=1, 2, \dots, n$), 贡献率指某个主成分提取的信息占总信息的份额,前 m 个主成分的贡献率之和称为 Y_1, Y_2, \dots, Y_m 累计贡献率.实际应用中,通常选取 $m < n$, 使前 m 个主成分的累计贡献率达到较高的比例(如 80%~90%), 这样用前 m 个主成分 Y_1, Y_2, \dots, Y_m 代替的原始特征序列 $X_1,$

* 国家自然科学基金(60974115)资助

† Tel: 13780333124

Email: panpanhq@163.com

收稿日期: 2009-08-20

修回日期: 2009-11-26

X_2, \dots, X_n ,不但可以使原始特征向量的维数降低,而且也不会损失原始影响因素中太多的信息。

BP神经网络是由大量具有非线性映射能力的神经元组成的前馈分层网络结构。标准BP算法存在收敛速度慢、局部极值和难以确定隐层及隐结点个数等缺点,所以在实际应用中受到一定的限制。动量法^[4]可以降低网络对于误差曲面局部细节的敏感性,有效抑制网络陷入局部极小,标准BP算法收敛速度慢的一个重要原因是学习速率选择不当,而自适应调整学习速率有利于缩短学习时间^[5],为了改善BP网络的收敛性,缩短学习时间,采用动量法和自适应调整学习率的改进BP算法为

$$\omega(k+1) = \omega(k) + \alpha(k)[(1-\eta)D(k) + \eta D(k-1)] \quad (1)$$

式中, $\omega(k)$ 既可表示单个权值,也可以表示权值向量, $D(k)$ 为 k 时刻的负梯度, α 为学习率, η 为动量

因子, $0 \leq \eta < 1$ 。

本文采用具有一个隐含层的BP神经网络用于矿物油的种类鉴别。根据实验安排利用主成分分析法所得到的特征主成分作为输入,输入层节点设置为符合要求的主成分数,由于矿物油中荧光成分多环芳香烃主要为一环、二环、三环和四环结构。隐层为四个节点,输出节点是要区分的种类数,这里设为4,包括柴油、机油、煤油、原油。输入层和隐含层采用选择Tan-Sigmoid作为传输函数,输出层采用线性函数,以保持输出的范围^[6]。

2 实验及结论

利用MATLAB工具对40个矿物油样本的三维荧光光谱进行了特征参量的提取,其中 $S_1, S_3, S_5, S_6, S_7, S_8, S_{13}, S_{14}$ 为机油, S_2, S_9, S_{10}, S_{12} 为煤油, $S_4, S_{15}, S_{16}, S_{17}, S_{20}, S_{21}, S_{22}, S_{24}$ 为原油,其余均

表1 部分原始输入样本

Table 1 Initial data of partial input sample

Sample	z	σ	m_x	m_y	kur	ske	ρ	g	k
S_1	5.078	13.86	85.15	318.6	3.751	1.036	-0.148	37.16	12.41
S_2	2.462	8.268	77.08	296.3	8.561	2.289	-0.124	40.31	6.611
S_3	5.034	12.52	80.86	326.1	4.548	1.189	-0.271	41.82	5.145
S_4	5.539	10.57	108.1	328.2	9.241	1.814	-0.152	46.74	9.845
S_5	5.436	12.99	93.47	299.8	9.097	2.221	-0.301	43.86	3.631
S_6	5.457	12.56	95.37	301.1	8.634	2.121	-0.298	44.85	3.474
S_9	2.746	8.522	80.98	293.0	9.517	2.146	-0.186	46.69	6.722
S_{10}	8.464	16.33	98.78	314.3	6.943	1.709	-0.219	44.63	3.174
S_{15}	13.97	22.17	117.4	311.6	2.137	0.636	-0.317	49.97	2.556
S_{16}	13.89	21.83	118.1	312.7	2.031	0.563	-0.298	49.83	2.729
S_{17}	14.51	23.01	119.2	315.0	2.078	0.621	-0.320	49.85	2.575
S_{20}	15.33	21.41	149.1	312.9	2.287	0.174	-0.103	53.01	8.001
S_{21}	4.376	9.956	114.5	297.0	2.987	0.289	-0.214	53.29	0.202
S_{25}	8.159	18.39	83.59	324.4	2.571	1.005	-0.353	39.44	2.509
S_{26}	6.688	16.14	76.54	325.6	2.556	0.982	-0.199	37.17	4.540
S_{27}	6.438	14.49	106.2	295.0	2.505	0.904	-0.354	40.07	3.471
S_{28}	6.870	15.21	104.1	293.7	2.925	0.988	-0.325	40.41	3.756
S_{29}	7.267	15.43	97.68	296.4	3.447	1.087	-0.275	41.95	3.655
S_{30}	7.963	16.15	96.47	299.7	4.086	1.229	-0.247	42.94	3.459

表2 相关系数矩阵

Table 2 Correlation matrix

Sample	z	σ	m_x	m_y	kur	ske	ρ	g	k
z	1.000 0	0.888 6	0.846 1	0.412 8	-0.702 2	-0.739 9	-0.138 7	0.540 2	-0.480 5
σ	0.888 6	1.000 0	0.912 2	0.376 7	-0.842 8	-0.887 2	-0.181 0	0.580 6	-0.624 8
m_x	0.846 1	0.912 2	1.000 0	0.127 3	-0.673 7	-0.833 5	-0.115 9	0.737 7	-0.587 7
m_y	0.412 8	0.376 7	0.127 3	1.000 0	-0.415 0	-0.135 5	-0.055 3	-0.034 0	0.054 5
kur	-0.702 2	-0.842 8	-0.673 7	-0.415 0	1.000 0	0.867 3	0.226 7	-0.146 3	0.584 7
ske	-0.739 9	-0.887 2	-0.833 5	-0.135 5	0.867 3	1.000 0	0.092 6	-0.502 8	0.745 5
ρ	-0.138 7	-0.181 0	-0.115 9	-0.055 3	0.226 7	0.092 6	1.000 0	0.100 0	0.408 5
g	0.540 2	0.580 6	0.737 7	-0.034 0	-0.146 3	-0.502 8	0.100 0	1.000 0	-0.460 9
k	-0.480 5	-0.624 8	-0.587 7	0.054 5	0.584 7	0.745 5	0.408 5	-0.460 9	1.000 0

为柴油. 部分原始输入样本见表 1, 其中 z 、 σ 、 kur 、 ske 、 ρ 、 g 、 k 分别为平均值、标准差、峰度、偏度、相关系数、几何方差、长轴斜率. m_x 、 m_y 分别为 x 和 y 方向的重心. 相关系数矩阵见表 2. 利用 MATLAB 编程, 对输入样本的特征参量标准化采用标准差变换, 得到相关系数矩阵 R , 求得 R 的特征值和特征向量, 在程序中设置贡献率为 85%, 可得到符合要求的主成分.

由载荷矩阵(见表 3)可知, 第一主成分中标准差、重心(m_x)等指标占有较大的载荷, 说明标准差和重心(m_x)等指标是鉴别矿物油荧光谱的主要因素, 第二主成分中重心(m_y)等占有较大载荷, 第三主成分中相关系数占有较大载荷. 主成分相关系数矩阵见表 4.

表 3 载荷矩阵

Table 3 Component matrix

Item	Y_1	Y_2	Y_3
z	-0.889 3	0.096 8	0.210 0
σ	-0.976 3	0.070 0	0.098 3
m_x	-0.927 9	-0.227 9	0.088 7
m_y	-0.301 4	0.747 3	0.451 5
kur	0.842 2	-0.378 7	0.040 0
ske	0.930 3	0.072 8	0.043 7
ρ	0.221 4	-0.367 8	0.796 0
g	-0.619 4	-0.615 7	0.197 6
k	0.736 7	0.162 1	0.506 2

表 5 油种鉴别输出

Table 5 Oil identification output

Sample	Target output				NN output			
	1	0	0	0				
S_1	1	0	0	0	0.996 7	0.002 1	-0.001 8	0.000 8
S_2	1	0	0	0	0.989 3	-0.001 7	0.001 3	0.002 5
S_3	1	0	0	0	0.992 3	-0.000 7	0.003 1	-0.000 3
S_4	0	1	0	0	0.001 1	0.994 2	-0.000 5	0.001 9
S_5	0	1	0	0	-0.000 5	0.987 7	-0.002 2	0.001 0
S_6	0	0	1	0	0.002 7	0.002 0	0.988 7	-0.000 9
S_7	0	0	1	0	-0.001 5	0.003 3	0.991 1	-0.000 6
S_8	0	0	0	1	0.002 4	0.001 3	-0.000 8	0.989 3
S_9	0	0	0	1	-0.001 4	0.001 1	-0.000 4	0.991 5
S_{10}	0	0	0	1	0.001 3	-0.000 9	-0.001 1	0.993 6

3 结论

采用主成份分析和神经网络相结合的方法建立了油种鉴别的分类模型, 主成分分析法可以有效地去除数据相关性, 减少输入样本维数, 大大简化了网络结构, 降低了神经网络训练的难度. 同时, 采用动量法和自适应调整学习速率相结合的改进 BP 算法进行神经网络训练, 实现了对矿物油种类的有效鉴别, 在本文中, 仅选取了体现三位荧光谱特征的 9 个

表 4 主成分相关系数矩阵

Table 4 Correlation matrix of principal components

Item	Y_1	Y_2	Y_3
Y_1	1.000 0	-0.017 2	-0.025 8
Y_2	-0.017 2	1.000	-0.043 5
Y_3	-0.025 8	-0.043 5	1.000 0

选用这三项主成分来代替原先的九项特征参量. 对主成分进行相关性分析, 验证主成分之间互不相关, 用这 3 项主成分的特征值对应的特征向量, 即经过变换后的综合指标, 标准化后作为鉴别矿物油 BP 神经网络模型的输入样本.

经主成分分析后得到的新输入样本, 包含 3 个输入变量即经主成分分析后提取的 3 项主成分, 从 40 组数据中随机选取 10 组数据作为鉴别样本, 将测试样本的特征谱输入训练好的网络, 进行矿物油种类鉴别的仿真实验, 这样神经网络的执行效率有了很大的提高, 结果如表 5.

由于未经主成分分析的原始输入样本包含 9 个输入变量, 即矿物油三维荧光谱的 9 个特征参量, 而经过主成分分析的新样本输入节点数为 3, 对大量数据分析而言, 网络节点数的减少简化了网络结构, 可使得程序运行的速度明显地加快, 对一般 BP 网络和 PCA-BP 网络鉴别速度作比较, 实验结果表明, 鉴别时间从 113 s 减少到 65 s, 对神经网络的执行效率有了很大的提高.

特征参量, 在今后的研究中, 可进一步探索更加具有价值的特征, 挖掘三维矿物油荧光谱所体现的更多信息, 进而完善鉴别系统, 实现对矿物油油种快速、准确的鉴别.

参考文献

- [1] SHANG Li-ping, XU Xiao-xuan, XU Jing-jun, *et al.* Fluorescence measurement of oil micro-contamination in water with optical fiber technique and its application [J]. *Acta Metrologica Sinica*, 2005, **26**(1): 81-85.
尚丽平, 徐晓轩, 许京军, 等. 水中微量矿物油污染的光纤荧光监测技术及应用研究[J]. 计量学报, 2005, **26**(1): 81-85.

- [2] TIAN Guang-jun, SHANG Li-ping, SHI Jin-shan. Measurement of oil's three-dimension fluorescence spectrum [J]. *Chinese Journal of Scientific Instrument*, 2004, **25**(4): 819-823.
田广军,尚丽平,史锦珊. 油类三维荧光光谱测量及其指纹图统计特征[J]. *仪器仪表学报*, 2004, **25**(4):819-823.
- [3] XU Jing-cai, WANG Hua. The neural network prediction method for the limit parameters of coal self-ignition [J]. *Journal of China Coal Society*, 2002, **27**(4): 366-370.
徐精彩,王华. 煤自燃极限参量的神经网络预测方法[J]. *煤炭学报*, 2002, **27**(4):366-370.
- [4] BAI Peng, WANG Jian-hua, WANG Hong-ke, et al. A method of mixed gas component Infrared spectrum recognition based on SVM regression model[J]. *Acta Photonica Sinica*, 2008, **37**(4):754-757.
白鹏,王建华,王宏柯,等. 基于 SVM 回归模型的混合气体组分种类光谱识别方法[J]. *光子学报*, 2008, **37**(4):754-757.
- [5] XIONG Yu-hong, WEN Zhi-yu, WANG Ming-yan, et al. Design and analysis of spectral recognition system based on neural network [J]. *Spectroscopy and Spectral Analysis*, 2007, **27**(1):139-142.
熊宇虹,温志渝,王命延,等. 基于神经网络的光谱识别系统的设计与分析[J]. *光谱学与光谱分析*, 2007, **27**(1):139-142.
- [6] ZHAO Jing, XIA Liang-zheng. Face recognition using neural network based on continuous wavelet transform [J]. *Acta Photonica Sinica*, 2005, **34**(9):1425-1430.
赵静,夏良正. 基于连续小波变换的神经网络人脸识别研究[J]. *光子学报*, 2005, **34**(9):1425-1430.

Oil Identification Technique Based on Analysis of Three-dimensional Fluorescence Spectra Feature

WANG Yu-tian¹, ZHANG Yan-lin¹, WANG Jin-yu²

(1 Hebei Key Laboratory of Measurement Technology and Instrumentation, Yanshan University, Qinhuangdao, Hebei 066004, China)

(2 China National Aviation Industrial Company 5711, Qinhuangdao, Hebei 066102, China)

Abstract: The basic pattern recognition theory of artificial neural network is introduced. The identification system for the oil's three-dimensional fluorescence spectra is established based on the principal component analysis (PCA) and BP neural network method, and the system design and the basic frame of model are made. PCA was inducted to pre-analyze the feature parameters of oil's three-dimensional, the principal components of original variables were used as the input of network, then the network output realized the identification of oil's three-dimensional fluorescence spectrum. This method can cut down the dimensions of original input, eliminate the relativity between variables, at the same time simplify the network structure, and improve the convergence speed. Actual instance was tested effectively that the PCA-BP neural network compared with the normal neural network reduced training time and possesses better performance. Results showed that the method can be used to realize the identification of oil's three-dimensional fluorescence spectrum.

Key words: Spectral analysis; Oil identification; Principal component analysis; BP neural network



WANG Yu-tian was born in 1952. He received the Ph. D. degree from Harbin Institute of Technology in 1995. Now he is a professor and Doctoral supervisor, and his research interests focus on optical testing and optical fiber sensor.