

doi:10.3788/gzxb20174611.1130002

# 水质痕量石油类污染物分类识别方法

苑媛媛<sup>1</sup>, 王书涛<sup>1</sup>, 孔德明<sup>2</sup>, 潘钊<sup>2</sup>

(燕山大学 电气工程学院, 河北 秦皇岛 066000)

**摘 要:**基于水质中石油类污染物的强荧光特性,构建了荧光检测实验系统,分别以 50 种不同浓度的汽油、柴油、煤油和机油的水溶液为测量样本,分析了其荧光特性.由于四者荧光峰位置相似,很难直接识别,提出了一种运用神经网络进行模式识别的新方法,将主成分分析法(PCA)和可拓神经网络(ENN)相结合对输入矢量进行降维并进行分类识别.结果表明,该方法与 ENN 和 PCA-BP 算法相比,迭代数由 265 降低到 60;识别离差平方和由 0.236 5 降到 0.014 5;识别效率由 72.50% 提升到 96.25%;识别精度可达  $10^{-6}$  级别.PCA-ENN 算法具有较高的识别精度和识别效率,同时也可用于水质中其它痕量有机物质的光谱识别.

**关键词:**石油类污染物;可拓神经网络;主成分分析;三维荧光光谱;分类识别

中图分类号:O433.4

文献标识码:A

文章编号:1004-4213(2017)11-1130002-7

## Classification of Trace Oil Pollutants in Water Quality

YUAN Yuan-yuan<sup>1</sup>, WANG Shu-tao<sup>1</sup>, KONG De-ming<sup>2</sup>, PAN Zhao<sup>2</sup>

(School of Electrical Engineering, Yanshan University, Qinhuangdao, Hebei 066004, China)

**Abstract:** Based on strong fluorescence property of oil pollutant in water quality, a fluorescence detection system is constructed, gasoline, diesel, kerosene, engine oil with different concentrations in water quality are prepared to analyze their fluorescence characteristics. In order to identify the three oil with the similar fluorescence peaks, the Principal Component Analysis (PCA) combined with Extension Neural Network (ENN) is proposed, which can reduce input vector and identify similar substances. Compare with the ENN and PCA-BP, the results show that the proposed method can make the iteration number dropped from 265 to 60, make the sum of the squares of dispersion decrease from 0.236 5 to 0.014 5, make recognition efficiency increase from 72.50% to 96.25%, and reach the  $10^{-6}$  level of recognition precision. The proposed method possesses high recognition accuracy and recognition efficiency, which can be used in spectral identification of other trace organic materials in water.

**Key words:** Oil pollutant; Extension neural network; Principal component analysis; Three-dimensional fluorescence spectra; Material classification

**OCIS Codes:** 300.2530; 300.6170; 300.6390

## 0 引言

随着社会的发展和工业水平的提高,人类对石油的需求量不断增加,但是在石油的勘探、开采、运输、加工过程中,大量的石油产品被排放到水环境中,由石油造成的水污染日益严重,直接影响水生生物及人类健

**基金项目:**河北省自然科学基金(No.F2017203220)、国家自然科学基金青年基金(No.61501394)、河北省自然科学基金(No.F2016203155)和秦皇岛市科学技术研究与发展计划(No.201502A017)资助

**第一作者:**苑媛媛(1991—),女,博士研究生,主要研究方向为光电检测、环境监测.Email:ysuyuanyuanyuan@163.com

**导师(通讯作者):**王书涛(1978—),男,教授,博士,主要研究方向为光电检测、环境多参量检测、环境监测等.Email:wangshutao@ysu.edu.cn

**收稿日期:**2017-06-12; **录用日期:**2017-07-28

<http://www.photon.ac.cn>

康,因此对水中油类污染物的监测非常重要<sup>[1-2]</sup>.目前,水中油类检测方法主要有红外分光光度法、紫外分光光度法、液相色谱法、气相色谱-质谱联用技术、荧光分析法等<sup>[3-7]</sup>.荧光分析法具有灵敏度高、测量范围大、可操作性强等优点,国内外学者利用荧光法对水中油类及其他污染物质进行了大量的研究<sup>[8-12]</sup>.但当水中油类污染物光谱相似,重叠现象严重时,直接从光谱图上很难实现定性识别,常用的分类识别方法有二阶校正算法<sup>[13-14]</sup>,神经网络算法<sup>[15-18]</sup>等.可拓神经网络(Extension Neural Network, ENN)是可拓学和神经网络相结合的产物,把可拓学中的菱形思维、可拓变换、关联函数等理论融入到神经网络之中,可以有效解决神经网络训练速度慢,容易将结果陷入局部极值等缺点,在模式识别、故障诊断、分类聚类<sup>[19-21]</sup>等领域有广泛的应用.

本文采用 ENN 结合主成分分析(Principal Component Analysis, PCA)实现水质中痕量汽油、柴油、煤油和机油等石油类污染物分类识别.算法采用 PCA 对神经网络输入矢量进行降维,利用 ENN 模型对样品的主成分数据进行分类识别.结果表明,相比 PCA-BP 神经网络和 ENN 算法,PCA-ENN 算法有较高的识别效率.

## 1 算法原理

PCA 是将原有的多项指标简化为少数几个相互无关的综合指标,并尽可能多的反映原有指标信息的一种多元统计方法<sup>[22]</sup>.PCA 的步骤主要包括样本数据标准化,求相关系数矩阵以及其特征值和特征向量等.

ENN 由输入层,隐层和输出层构成.在 BP 神经网络基础上,引入可拓学中扩缩变换的方法,在输出空间中采取一个特定的区域来代替期望输出信号,当网络输出落在此区域时,误差项为零.算法首先提取三维荧光光谱数据中荧光强度平均值( $\bar{z}$ ),标准差( $\sigma$ )等数据进行降维,通过映射和变换法,将高维数据特征转变为低维数据特征.将样本参量标准化处理后,求相关系数矩阵的特征值和特征向量.提取主成分累计贡献率在 90% 到 95% 之间的数据,计算出最终的主成分数据.

很多神经网络应用时会遇到输入变量维数过高的情况,会导致训练速度变慢和过饱和问题.采用主成分分析法对数据进行预处理,减少输入变量维数,消除其相关性,提高了运行速度;可拓神经网络算法训练速度更快,分类识别率更高,结合 PCA 的 ENN 算法,可有效实现痕量油种鉴别,也可用于其他有机物质的光谱识别领域.

PCA-ENN 算法的具体步骤如下:

1) 首先对样本的特征参数进行处理,将样本参数的矩阵表达式为

$$\mathbf{X} = \begin{pmatrix} x_{11} & \cdots & x_{1J} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{NJ} \end{pmatrix} \quad (1)$$

对矩阵中的各个元素作如下标准化变换

$$x'_{nj} = \frac{x_{nj} - \bar{x}_j}{s_j}; n = 1, \dots, N; j = 1, \dots, J \quad (2)$$

式中

$$\bar{x}_j = \frac{1}{N} \sum_{n=1}^N x_{nj} \quad (3)$$

$$s_j = \sqrt{\frac{1}{N} \sum_{n=1}^N (x_{nj} - \bar{x}_j)^2} \quad (4)$$

式中,  $\bar{x}_j$  是矩阵中第  $j$  个特征的平均值;  $s_j$  是矩阵中第  $j$  个特征的标准差.标准化处理之后,矩阵中所有元素均值为 0, 方差为 1, 记为  $\mathbf{X}'$ .

2) 求出自变量矩阵  $\mathbf{X}'$  的相关系数矩阵  $\mathbf{D}$ ;

3) 求矩阵  $\mathbf{D}$  的特征值和特征向量.其中,  $\mathbf{P}$  为特征向量,  $\lambda$  为特征值;

$$\mathbf{D}\mathbf{P} = \mathbf{P}\lambda \quad (5)$$

4) 计算主成分矢量累积贡献率在 90% 到 95% 之间的特征值  $\lambda_1, \lambda_2, \dots, \lambda_m$  相对应的第 1, 2,  $\dots, m$  个主成分;

5) 经主成分分析后的参数向量将作为可拓神经网络的输入向量进行学习和识别.

可拓神经网络的训练过程流程见如图 1.

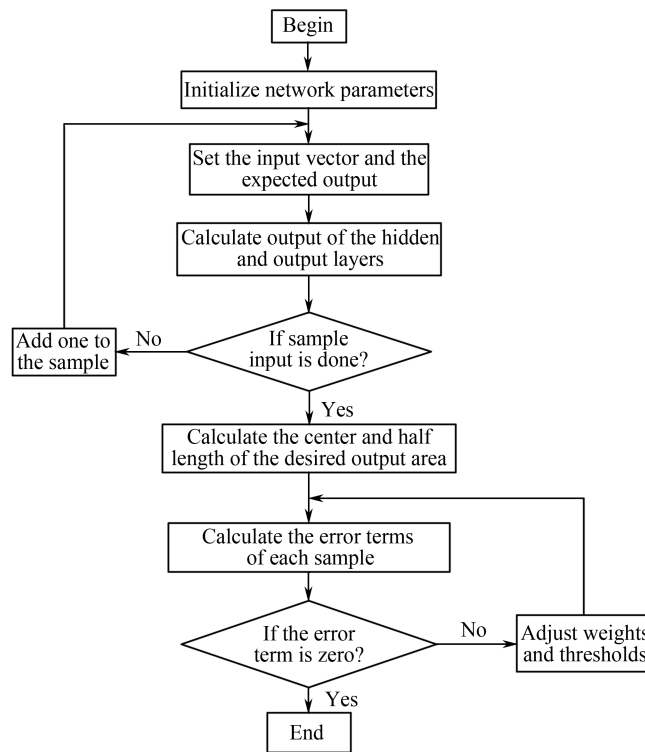


图1 可拓神经网络流程

Fig.1 Flow chart of extension neural network program

## 2 实验部分

### 2.1 荧光检测实验系统

荧光光谱检测系统主要由激发光源、发射器件、样品池、光学探测器件以及显示装置等组成,结构见图2。450 W 高压氙灯发射激发光,经光栅激发单色器滤光,波长大于 200 nm 的紫外光通过滤光器后照射到样品池;样品池中的荧光物质产生的荧光经发射单色器滤除杂光,获得的单色荧光照射到光电倍增管(Photomultiplier Tube, PMT);光电倍增管将荧光信号转变为电信号,放大器对该电信号进行放大,最后经模拟/数字(A/D)转换将电信号转变为数字信号,信号传输到达计算机进行处理。

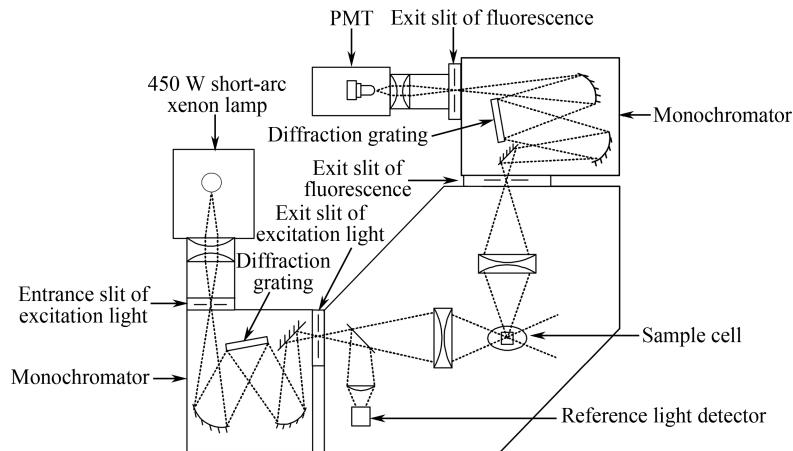


图2 荧光检测实验系统

Fig.2 Structure of fluorescence spectroscope

### 2.2 样品制备

采用移液管取市场上购买的某汽油、柴油、煤油和机油样品各 0.1 mL,分别置于 4 个 100 mL 容量瓶中,加纯净水定容,得到 4 种 1 mL/L 的水溶液,按顺序标记为 1#~4#,并用 HY-5A 回旋振荡器震荡,使其充

分溶解在水中,常温密封保存;分别取 1#~4# 的每种石油样品的水溶液进行稀释成浓度为  $1 \mu\text{L/L} \sim 100 \mu\text{L/L}$  的 50 种不同浓度的样品,分别标记为 G1-G50, D1-D50, K1-K50; E1-E50.

### 2.3 实验过程

实验中,将激发波长设置为  $230 \sim 600 \text{ nm}$ ,发射波长设置为  $250 \sim 750 \text{ nm}$ ,扫描水的荧光光谱,得到其瑞利散射出现在  $E_m \geq 600 \text{ nm}$ ,对所测油类无干扰.根据石油类污染物的荧光特性,我们将激发和发射波长范围分别设置为  $200 \sim 450 \text{ nm}$  和  $250 \sim 450 \text{ nm}$ ,步长为  $2 \text{ nm}$ ,冷却温度  $-19.7 \text{ }^\circ\text{C}$ ,狭缝宽度  $2.78 \text{ mm}$ ,积分时间为  $0.1 \text{ s}$ ,发射波长始终滞后激发波长  $20 \text{ nm}$ ,以避免瑞利散射干扰.

## 3 结果与讨论

### 3.1 水质中痕量汽油、柴油、煤油和机油的荧光光谱分析

分别对制备好的水中汽油、柴油、煤油和机油溶液进行荧光扫描,得到部分光谱图见图 3.

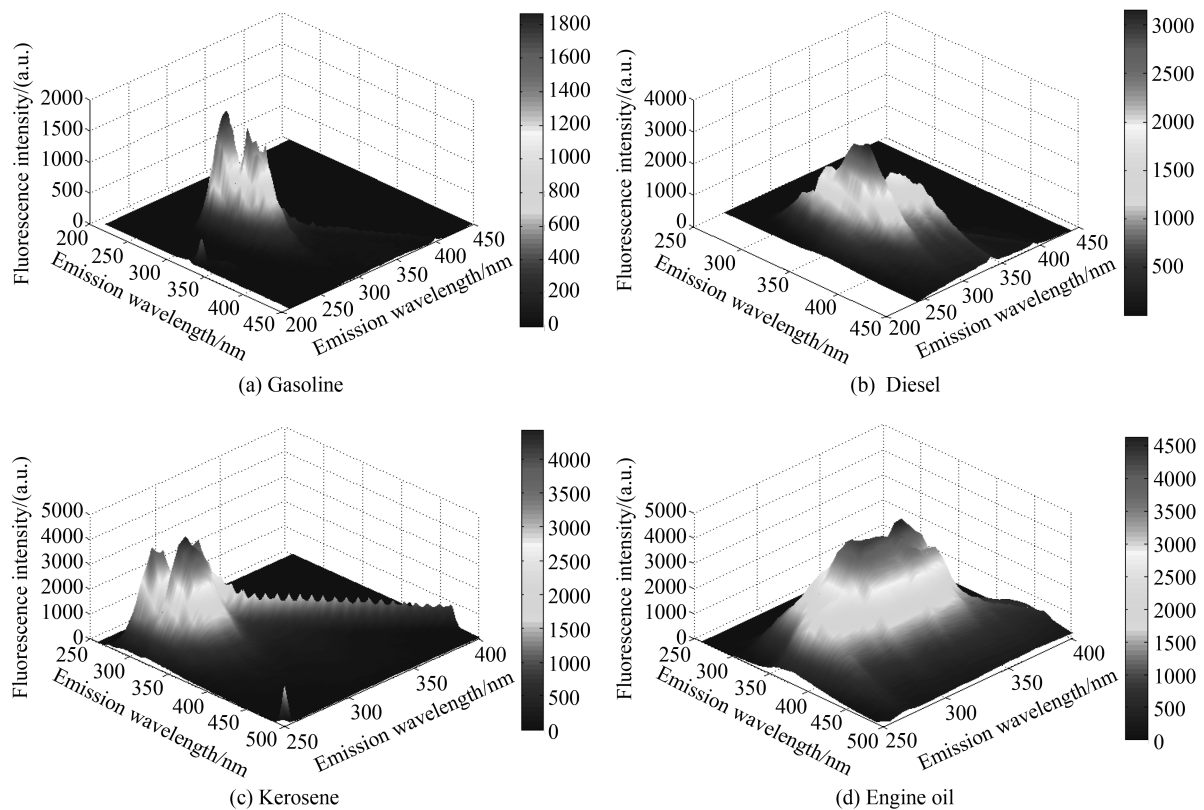


图 3 水质中四种油类的三维荧光光谱

Fig.3 Three-dimensional fluorescence spectra maps of three oils in water quality

由图 3 中可看出,汽油、柴油、煤油和机油荧光峰区域见表 1,荧光峰位置有一定程度的重叠,无法通过光谱图直接识别出石油类污染物种类.采用 PCA-ENN 算法对测量样本进行分类识别.

表 1 水质中石油类的光谱特征对比表

Table 1 Compared table of spectral characteristics of the petroleum in the water quality

Species	Peak number	The largest Ex/nm position	The largest Em/nm position
Gasoline	2	274~362	286~356
Diesel	2	260~330	320~370
Kerosene	2	270~300	280~350
Engine oil	2	300~340	350~390

### 3.2 PCA-ENN 对水质中石油类污染物分类识别

主成分数据反应了样品的特征信息.为加快运行速度,本文已将不同样本三维荧光光谱数据的荧光强度值标准变换,标准化处理之后,矩阵中所有元素均值为 0,方差为 1.求取标准化后荧光强度的平均值( $\bar{z}$ )、标

准差( $\sigma$ )、重心( $m_x, m_y$ )、相关系数( $\rho$ )、长轴斜率( $k$ )、几何方差( $g$ )、偏度( $ske$ )、峰度( $kur$ )共9个值,充分代表了光谱图中光强大小及等高线形状.从而得到这9项指标的相关系数矩阵,如表2所示.在MATLAB平台上调用pcacov函数对其进行主成分分析,这9个主成分的特征值分布如图4所示,具体特征值及贡献率如表3所示,其中,前4个主成分的累计贡献率达到93.1242%.因此,我们选用前4个主成分进行后续分析,虽然会有一定的损失,但损失不大,将这四项目主成分特征值对应的特征向量作为神经网络模型油种鉴别的输入样本,实现降维处理.

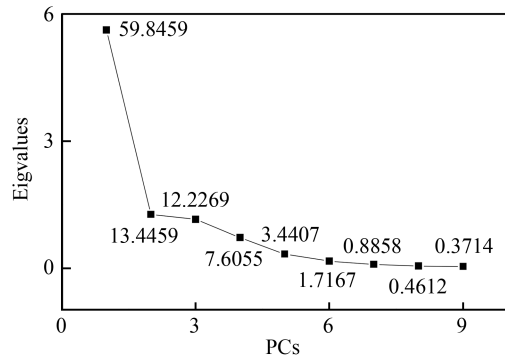


图4 主成分分析的特征值

Fig.4 Eigenvalues plot of principle component analysis

表2 相关系数矩阵

Table 2 The characteristic parameters of the correlation matrix

	$z$	$\sigma$	$m_x$	$m_y$	$kur$	$ske$	$\rho$	$g$	$k$
$z$	1.000 0	0.846 9	0.352 4	0.423 7	-0.901 5	-0.931 4	-0.138 5	0.967 6	-0.479 2
$\sigma$	0.846 9	1.000 0	0.673 8	0.376 5	-0.962 0	-0.957 0	-0.183 0	0.880 5	-0.624 7
$m_x$	0.352 4	0.373 8	1.000 0	0.127 6	-0.624 3	-0.578 4	-0.115 7	0.505 7	-0.587 1
$m_y$	0.423 7	0.376 5	0.127 6	1.000 0	-0.415 0	-0.135 5	-0.055 8	-0.032 0	0.054 2
...	...	...	...	...	...	...	...	...	...
$g$	0.967 6	0.880 5	0.505 7	-0.032 0	-0.917 6	-0.944 5	0.100 0	1.000 0	-0.461 2
$k$	-0.479 2	-0.624 7	-0.587 1	0.054 2	0.584 9	0.745 6	0.408 3	-0.461 2	1.000 0

表3 主成分分析特征值和贡献率

Table 3 The eigenvalues and contribution rates of principle component analysis

Principal component	Eigenvalue	Contribution rate/%
PC1	5.631 4	59.845 9
PC2	1.265 2	13.445 9
PC3	1.150 5	12.226 9
PC4	0.715 7	7.605 5
PC5	0.323 8	3.440 7
PC6	0.161 5	1.716 7
PC7	0.083 4	0.885 8
PC8	0.043 4	0.461 2
PC9	0.034 9	0.371 4

在对四种水质中石油类样品共200组光谱数据进行PCA分析的基础上,选取每种石油类样品30组光谱数据作为训练集输入网络,对PCA-ENN网络模型进行训练.利用每种样品剩余的20组光谱数据作为测试样本进行分类识别.识别结果如表4所示,其中,机油的识别正确率达到了100%,汽油、柴油和煤油的识别率分别达到了95%、95%和95%,分析原因可能是汽油、柴油和煤油均出现了两个荧光峰,且位置及其相似,因此在预测时出现了少量误报情况.对比采用ENN和PCA-BP进行预测如表5所示,可得出采用PCA-ENN后,迭代数明显降低,训练时间缩短,预测结果的离差平方和由0.2365降低至0.0145,平均识别成功率也由72.50%提升至96.25%.

表4 PCA-ENN分类识别结果

Table 4 Classification results of PCA-ENN

	Gasoline	Diesel	Kerosen	Engine oil
Identify the samples number	20	20	20	20
Correctly identify the samples number	19	19	19	20
Recognition accuracy	95%	95%	95%	100%

表 5 各方法分类识别结果  
Table 5 Each method of classification results

Methods	Epoch limits	Sum of deviation square ( $\eta^2$ )	Average recognition success rate
ENN	265	0.085 6	90.00%
PCA-BP	122	0.236 5	72.50%
PCA-ENN	60	0.014 5	96.25%

## 4 结论

通过构建荧光检测实验系统,获得 200 组水质中汽油、柴油、煤油与机油四种石油类污染物的三维荧光光谱.由于四者荧光峰位置相似,很难进行直接识别,本文采用主成分分析法对光谱图中代表样品信息的 9 个特征参数向量进行降维,降低特征参量维数,采用 ENN 对主成分数据进行分类识别.结果表明 PCA-ENN 算法对机油的识别正确率达到了 100%,对汽油、柴油和煤油的识别率均达到了 95%;通过与 ENN 和 PCA-BP 算法对比,识别效率由 72.50% 提升到 96.25%,迭代数由 265 降低到 60,识别离差平方和( $\eta^2$ )由 0.236 5 降到 0.014 5,该方法提高了神经网络分类识别速度和识别效率,检测精度可达  $10^{-6}$  级别,可以用于水质中石油类污染物快速准确的鉴别.

### 参考文献

- [1] SPIER C, STRINGFELLOW W T, HAZEN T C, *et al.* Distribution of hydrocarbons released during the 2010 MC252 oil spill in deep offshore waters[J]. *Environmental Pollution*, 2013, **173**(1): 224-230.
- [2] COHEN M A. Water pollution from oil spills[J]. *Encyclopedia of Energy Natural Resource & Environmental Economics*, 2013, **3**(3): 121-126.
- [3] RUIZ M D, BUSTAMANTE I T, DAGO A, *et al.* A multivariate calibration approach for determination of petroleum hydrocarbons in water by means of IR spectroscopy[J]. *Journal of Chemometrics*, 2010, **24**(7-8): 444-447.
- [4] UOZUMI J, PEIPONEN K E, TURUNEN P, *et al.* Robust sensor for turbidity measurement from light scattering and absorbing liquids[J]. *Optics Letters*, 2009, **34**(23): 3743-5.
- [5] GAULIER F, GIBERT A, WALLS D, *et al.* Mercury speciation in liquid petroleum products: Comparison between on-site approach and lab measurement using size exclusion chromatography with high resolution inductively coupled plasma mass spectrometric detection (SEC-ICP-HR MS)[J]. *Fuel Processing Technology*, 2015, **131**: 254-261.
- [6] BERNABEU A M, FERNANDEZ-FERNANDEZ S, BOUCHETTE F, *et al.* Recurrent arrival of oil to Galician coast: the final step of the Prestige deep oil spill[J]. *Journal of Hazardous Materials*, 2013, **250-251**(2): 82-90.
- [7] MOONKOO K, UNHYUK Y, HONG S H, *et al.* Hebei spirit oil spill monitored on site by fluorometric detection of residual oil in coastal waters off Tae-an, Korea[J]. *Marine Pollution Bulletin*, 2010, **60**(3): 383-9.
- [8] WANG Chun-yan, LI Wen-dong, LUAN Xiao-ning. Species identification and concentration quantification of crude oil samples in petroleum exploration using the concentration-synchronous-matrix-fluorescence spectroscopy[J]. *Talanta*, 2010, **81**(1-2): 684-691.
- [9] JIANG F, LEE F S, WAN X, *et al.* The application of excitation/emission matrix spectroscopy combined with multivariate analysis for the characterization and source identification of dissolved organic matter in sea water of Bohai sea[J]. *China Marine Chemistry*, 2008, **110**(1-2): 109-119.
- [10] CALIMAG-WILLIAMS K, KNOBEL G, GOICOECHEA H C, *et al.* Achieving second order advantage with multi-way partial least squares and residual bi-linearization with total synchronous fluorescence data of monohydroxy-polycyclic aromatic hydrocarbons in urine samples[J]. *Analytica Chimica Acta*, 2014, **811**: 60-69.
- [11] WANG Yu-tian, ZHAO Xu, PAN Zhao, *et al.* Processing method of five-dimensional data based on fluorescence spectra[J]. *Chinese Journal of Luminescence*, 2016, **37**(11): 1436-1443.  
王玉田,赵煦,潘钊,等. 基于荧光光谱的五维数据处理方法[J]. 发光学报, 2016, **37**(11): 1436-1443.
- [12] HAN Zhong-zhi, WAN Jian-hua, LIU Jie, *et al.* Multispectral imaging detection using the ultraviolet fluorescence characteristics of oil[J]. *Chinese Journal of Luminescence*, 2015, **36**(11): 1335-1341.  
韩仲志,万剑华,刘杰,等. 利用油品紫外荧光特性的多光谱成像检测[J]. 发光学报, 2015, **36**(11): 1335-1341.
- [13] WANG Shu-tao, SHAO Xiao-qing, PAN Zhao, *et al.* Application of three-dimensional fluorescence spectrum method and PARAFAC in petroleum pollutant measurement and identification[J]. *Advanced Materials Research*, 2011, **311-313**: 1213-1216.
- [14] GOICOECHEA H C, CALIMAG-WILLIAMS K, CAMPIGLIA A D. Multi-way partial least-squares and residual bi-linearization for the direct determination of monohydroxy-polycyclic aromatic hydrocarbons on octadecyl membranes via room-temperature fluorescence excitation emission matrices[J]. *Analytica Chimica Acta*, 2012, **717**(5): 100-109.
- [15] LIU Qian-qian, WANG Chun-yan, SHI Xiao-feng, *et al.* Identification of spill oil species based on low concentration

- synchronous fluorescence spectra and rbf neural network[J]. *Spectroscopy and Spectral Analysis*, 2012, **32**(4): 1012-1015.
- 刘倩倩,王春艳,史晓凤,等. 基于 RBF 神经网络的较低浓度下同步荧光光谱的溢油鉴别[J]. *光谱学与光谱分析*, 2012, **32**(4):1012-1015.
- [16] DJEMILI R, BOUROUBA H, KORBA M C A. Application of empirical mode decomposition and artificial neural network for the classification of normal and epileptic EEG signals[J]. *Biocybernetics & Biomedical Engineering*, 2016, **36**(1): 285-291.
- [17] BIEROZA M, BAKER A, BRIDGEMAN J. Classification and calibration of organic matter fluorescence data with multiway analysis methods and artificial neural networks: an operational tool for improved drinking water treatment[J]. *Environmetrics*, 2011, **2**(3): 256-270.
- [18] WANG Shu-tao, CHEN Dong-ying, WANG Xing-long, *et al.* A New Method for the Determination of Potassium Sorbate Combining Fluorescence Spectra Method with PSO-BP Neural Network [J]. *Spectroscopy and Spectral Analysis*, 2015, **35**(12): 3549-3554.
- [19] CHEN Huang-cheng, GU Feng-chang, WANG Meng-hui. A novel extension neural network based partial discharge pattern recognition method for high-voltage power apparatus[J]. *Expert Systems with Applications*, 2012, **39**(3): 3423-3431.
- [20] SHATNAWI Y, Al-KHASSAWENEH M. Fault diagnosis in internal combustion engines using extension neural network[J]. *IEEE Transactions on Industrial Electronics*, 2013, **61**(3): 1434-1443.
- [21] WANG Zhen-mu, WU Ming-ju, CHEN Jian-hong, *et al.* Extension neural network approach to classification of brain MRI[C] Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IEEE Computer Society, 2009: 515-517.
- [22] CHANG Yu-qing, WANG Shu, WANG Fu-li, *et al.* Process monitoring method based on multiple PCA models[J]. *Chinese Journal of Scientific Instrument*, 2014, **35**(4): 901-908.
- 常玉清,王姝,王福利,等. 基于多 PCA 模型的过程监测方法[J]. *仪器仪表学报*, 2014, **35**(4):901-908.