

安防系统光纤信号特征提取与分类算法研究

万遂人, 彭丽成

东南大学生物科学与医学工程学院, 南京 210096

摘要 在宽域全光纤安防系统中,在不影响灵敏度的情况下区分入侵和正常事件是一个关键性的系统性能指标。由于入侵和正常事件以及各个入侵事件的光纤信号在某些情况下极为相似,因此需要对这些信号的特征进行仔细的筛选和识别。本文由此出发,提出了一套光纤信号特征提取方案,即使用小波降噪手段对光纤信号进行去噪;根据信号与噪声的能量在时域分布的不同,提出了一种实用的光纤信号的预分割方法;提取信号在小波空间的能量分布特征,形成特征向量;使用向量机分类器对光纤信号分类。

关键词 光纤信号;特征提取;小波降噪;信号预分割

中图分类号 TP302

文献标识码 A

doi 10.3981/j.issn.1000-7857.2012.36.002

Feature Extraction and Classification of Fiber Signal in Security-Monitoring System

WAN Suiren, PENG Licheng

School of Biological Science & Medical Engineering, Southeast University, Nanjing 210096, China

Abstract The discrimination between the intrusion and the nuisance events without compromising sensitivity is a key performance parameter for any outdoor perimeter intrusion detection system. In certain circumstances, the signals of the intrusion and the nuisance events are almost the same. This is especially the case for the intrusion and the nuisance events which may have a similar impact. In this paper, a series of methods are proposed to extract the fiber-signals: to denoise the input signal based on the wavelet transform; with a new practical algorithm of pre-segmentation of the fiber-signal, based on the hypothesis that the distribution of the energy of the intrusion in the time domain is different from that of the nuisance; to generate the eigenvector extracted from the distribution of the energy in the wavelet space; and to classify the fiber-signal using the support vector machine. With the assumption that different kinds of fiber-signals have different energy distributions in each frequency, the results of this experiment are satisfactory. This method is practical because we may take advantages of the SVM with very small cost.

Keywords fiber signal; feature extraction; wavelet denoising; pre-segmentation

0 引言

宽域全光纤安防系统是一种新型的分布式安全警戒系统。系统在不影响灵敏度的情况下区分入侵和正常事件是一项关键性的系统性能指标。入侵和正常的信号在某些情况下极为相似,因此有必要使用模式识别方法对信号进行区分。

目前,欧美发达国家已开发出一些比较成熟的产品,国外在这方面的研究也开始得较早。比如,Seedahmed等^[1]提出了一套鲁棒性较高的光纤入侵信号识别方法。该方法提取了一些光纤信号的时域特征,并把这些特征形成的特征向量输入神经网络中进行训练和判别;Seedahmed等^[2]提出了一种区

收稿日期:2012-09-18;修回日期:2012-11-01

基金项目:国家高技术研究发展计划(863计划)项目(2002AA245091)

作者简介:万遂人,教授,研究方向为生物医学工程与物理电子学,电子信箱:srwan@seu.edu.cn

分雨水时事件与其他入侵事件的方法。国内产品刚推入市场,相关的研究也是在最近几年才取得的突破。饶云江等^[3]提出了一种利用小波处理对信号进行降噪和分割的方法,陶沛琳^[4]提出了一种分析光纤特征向量的方法,赵杰等^[5]提出了利用小波包变换后的能量谱形成信号的特征向量的方法,倪明等^[6]使用非平衡二叉树向量机作为模式识别的分类器,都取得了良好的效果。

本文根据上述文献的一些研究结果,提出了一套较完整的光纤信号识别方法。具体如下:使用小波降噪手段对光纤信号进行去噪;根据信号与噪声的能量在时域分布的不同,提出了一种实用的光纤信号的预分割方法;提取信号在小波空间的能量分布特征,形成初始的特征向量;使用向量机分类器对光纤信号进行分类。

实验表明,与文献[5]、[6]提出的方法相比,该方法处理得到的各个类之间的类间距更大,分类结果更准确。

1 光纤特征提取算法

本文针对光纤信号的特点,提出了一套完整的光纤信号特征提取算法。光纤信号的特征提取过程可用图1表述,图中虚线框表示在某些情况下不需要使用。

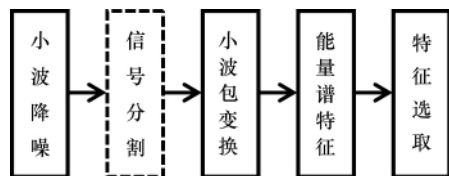


图1 光纤信号特征提取流程

Fig. 1 Process of the fiber signal feature extraction

1.1 信号预处理

光纤信号在进行特征提取之前,需要进行信号降噪和信号分割。这些预处理对于明确特征提取目标有着重要意义。

1.1.1 信号降噪

对光纤信号进行降噪的原因在于,采样得到的光纤信号存在大量的噪声,而噪声的能量对信号的能量谱特征提取将产生影响。在环境噪声能量较大或者输入信号能量较小时,这种影响尤为明显,甚至引起分类器的误判,使得识别正确率下降。由于噪声的能量主要分布在高频部分,故大多数降噪算法均针对信号的高频部分进行处理。但如果高频噪声抑制得过多,也同样会抑制信号本身的高频细节,引起信号失真,从而影响特征提取与信号识别。因此,选取合适的信号降噪算法就显得尤为重要。

目前,小波去噪在一维信号的降噪处理中使用非常广泛,主要是因为其具有以下特点:(1)时频局部化特性;(2)多分辨率特性;(3)解相关特性;(4)选基灵活性。其中,性质(1)和性质(2)决定了小波去噪方法在去除噪声的同时能够

很好地保留信号的突变^[9]。本文选用了一种改进的 VisuShrink 阈值去噪法去除光纤信号中的噪声。

传统的 VisuShrink 阈值方法是针对多维独立正态变量联合分布,在维数趋于无穷时得出的结论,即大于该阈值的系数含有噪声信号的概率趋于零,是基于最小最大估计得出的最优阈值^[8]。阈值 T 的选择满足

$$T = \sigma_n \sqrt{2 \ln N} \quad (1)$$

其中, σ_n 为噪声的标准差(均方差), N 为信号长度或尺寸。这种阈值与信号的长度 N 相关,当 N 较大时,阈值趋向于将所有的小波系数置零,往往会产生“过扼杀”系数的现象,此时小波去噪退化为低通滤波器。虽然该方法有很好的理论支持,但实际应用效果并不好,有分析其根本原因在于这一准则是用渐进分析的手段推出来的。但对于实际问题,信号或图像的复杂性相对于样本尺寸是很重要的。

本文采用的是 VisuShrink 阈值的改进方法^[9],阈值 T' 计算方式为

$$T' = k \sqrt{2 \ln N} \quad (2)$$

$$k = \text{median}(|x|) / 0.6745 \quad (3)$$

其中, x 为输入信号; median 函数为对输入的一段信号值排序,取排序后的中点值。其他参数与传统算法的意义相同。实际应用中,对小波分解的每一层的高频成分进行阈值滤波处理,信号长度 N 为每一层高频成分的长度。这样,得到的阈值便有了局部适应特性。

1.1.2 信号分割

对于包含入侵信息的一段信号来说,入侵信号出现的时间段,信号能量会突然增大,并表现出较强的区别于噪声的局部特性,如图2所示。可以看出,原始信号包含2段入侵扰动,其中1段是完整的入侵扰动,另1段是下一次入侵扰动的开始部分。在提取这段信号之前,必须去除后一段不完整入侵扰动信号的影响,这正是本文进行信号分割的目的。

根据入侵信号的能量分布特征,可通过某些方法将一段信号中的入侵信号段分割出来。进行这一步处理的原因,一个是消除噪声段对小波能量的影响,使得分类与识别的结果更加准确;另一个是由于系统在采样信号的时候,采样时间段内通常含有1—2段入侵信号,甚至某些入侵信号在采样时间段内被截断了,在这种情况下再对输入信号进行小波包能量谱特征提取显然是不正确的。而信号分割可以保证每次进行特征提取的信号中只含有一段入侵信号,从而提高了系统的稳定性和鲁棒性。

信号分割的具体步骤如下:

(1) 对输入信号 x 取平方,均值滤波并进行归一化处理,得到 x_{Norm} 。

(2) 设置噪声的能量上界 LowThresh 和信号的能量下界 HighThresh 。

(3) 一般情况下,完整的入侵信号必包含在两段噪声段之间,因此先取得噪声段的起始位置 stns , 使得当 $\text{id}x < \text{stns}$

时, $xNorm(idx) > LowThresh$, 且 $xNorm(stns) < LowThresh$ 。

(4) 由于入侵信号的能量会突然增大, 因此可以认为入侵信号段开始于一个较大的能量点。取得入侵信号段的起始位置 $stsg$, 使得 $stsg > stns$, 且 $stns < idx < stsg$ 时, $xNorm(idx) < HighThresh$, 而 $xNorm(stsg) > HighThresh$ 。

(5) 入侵信号随着冲击的减小, 会逐渐减小, 直到被噪声淹没, 因此, 入侵信号的终点必在起点 $stsg$ 和下一个信号的起点 $ntstsg$ 之间。因此, 先找到点 $sted$, 使得 $stsg < sted$, 且 $stsg < idx < sted$ 时, $xNorm(idx) > LowThresh$, $xNorm(sted) < LowThresh$,

然后找到点 $ntstsg$, 使得 $sted < ntstsg$, 当 $sted < idx < ntstsg$ 时, $xNorm(idx) < HighThresh$, $xNorm(ntstsg) > HighThresh$ 。最后, 找到 $edsg$, 使得 $sted < edsg < ntstsg$, 且 $edsg < idx < ntstsg$ 时, $xNorm(idx) > LowThresh$, $xNorm(edsg) < LowThresh$ 。

(6) 可认为信号起始于 $stsg$, 结束于 $edsg$ 。因此, 将 $[stsg, edsg]$ 段的信号取出, 作为分割结果。

这一步, 两个关键参数 $LowThresh$ 和 $HighThresh$ 对于信号分割的准确性有重要影响。实验中通常取 $LowThresh=0.05$, $HighThresh=0.8$; 图 2 信号的分割结果如图 3 所示。可以看出,

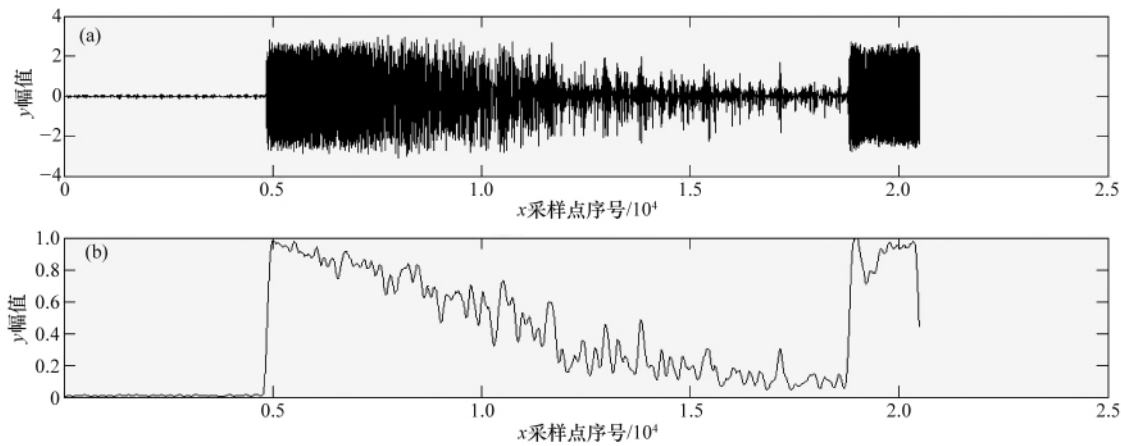


图 2 原始信号 (a) 和其能量信号平均化结果 (b)

Fig. 2 Original signal (a) and the result of mean-filtering energy signal (b)

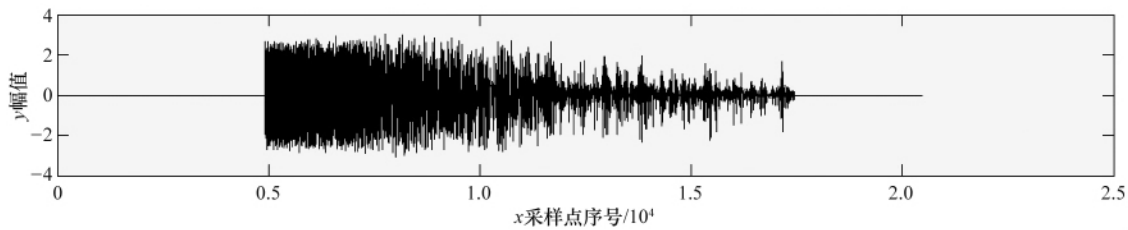


图 3 图 2 信号分割结果

Fig. 3 Segmentation result of the signal from Fig. 2

上述信号分割方法可以得到该时间段内一段完整的入侵扰动。

1.2 特征提取

对于光纤信号来说, 正确提取信号特征对于信号分类的有效性具有重要影响。文献[5]提出, 取原信号的 5 层小波包变换各小波成分的总能量作为信号的特征向量。从概念上讲, 各小波包成分均代表信号在时频域区间上的分布, 且各区间之间包含的信号能量彼此相互独立。因此, 可以用小波包成分的能量代表特征向量的每一维。具体提取步骤如下:

(1) 对信号进行小波变换 (db1 小波), 得到低频成分和高频成分。

(2) 分别对低频成分与高频成分进行 4 层小波包变换, 得到 32 个小波包成分。

(3) 取得各个小波包成分的总能量, 形成一个 32 维的特

征向量。由于信号的细节部分主要存在于高频部分, 高频成分的小波包能量可以适当乘以一个系数, 以增大信号在高频的区分度。

经过此步骤, 即可得到信号的特征向量。

2 光纤信号分类算法

如果要对光纤信号进行分类和识别, 必然包括选择一组样本进行训练和输入未知样本进行识别两个步骤。

本文使用的分类器为支持向量机 (SVM) 分类器。它具有小样本、非线性和高维表达等特性^[10]。SVM 分类器基于这样的思想: 通过适当得到一个足够高维的非线性映射, 属于两类数据总可以被一个超平面分割。

支持向量机主要针对 2 类模式分类问题。设 n 个 d 维样

本 $x_i \in R^d (i \in 1, 2, \dots, n)$, 可分为两类。因此, 可用 $y_i \in \{-1, 1\}$ 表示每个样本所属的类。用点 (x_i, y_i) 表示一个分类空间。在空间中找到一个最优超平面, 使得两类数据完全被超平面分开, 且边沿裕量最大。因此, 问题可转换为在 n 个约束条件:

$$y_i(\mathbf{w}\mathbf{x}_i + b) \geq 1 \quad i \in 1, 2, \dots, n \quad (4)$$

下, 求函数

$$\Phi(\mathbf{w}) = \frac{1}{2} \|\mathbf{w}\|^2 = \frac{1}{2} (\mathbf{w}^T \mathbf{w}) \quad (5)$$

的最小值。其中, \mathbf{w} 为 n 维列向量, b 为超平面的偏移量, Φ 为非线性映射函数, 将 x_i 从 R^d 空间映射到特征空间 R^n 。可以把上面有约束的权向量长度极小化问题转化为无约束的拉格朗日待定因子问题。这样, 需要极大化

$$L(\mathbf{a}) = \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j (\mathbf{x}_i^T \mathbf{x}_j) \quad (6)$$

并满足约束条件

$$\sum_{i=1}^n y_i a_i = 0 \quad a_i \geq 0; i = 1, 2, \dots, n \quad (7)$$

这些方程可用经典二次规划求解。

SVM 是针对二分类问题提出的, 而实际应用中通常需要解决多分类问题。例如, 本系统中的输入信号就分为非入侵信号、剪网信号、爬网信号等。因此, 需要通过将多个向量机进行组合, 实现基于向量机的多分类器^[14], 将多分类 SVM 方法应用于人脸识别问题中, 在较短的处理时间内取得了较好的分类效果。

本文正是采用这样的分类方法处理光纤信号的多分类问题, 并取得了良好的效果。

3 实验结果与分析

本文的输入信号来源为无锡永大集团的分布式光纤传感网络安全系统。输入的信号主要分为 3 类: 非入侵信号、剪网入侵信号和爬网入侵信号, 其典型信号如图 4 所示。

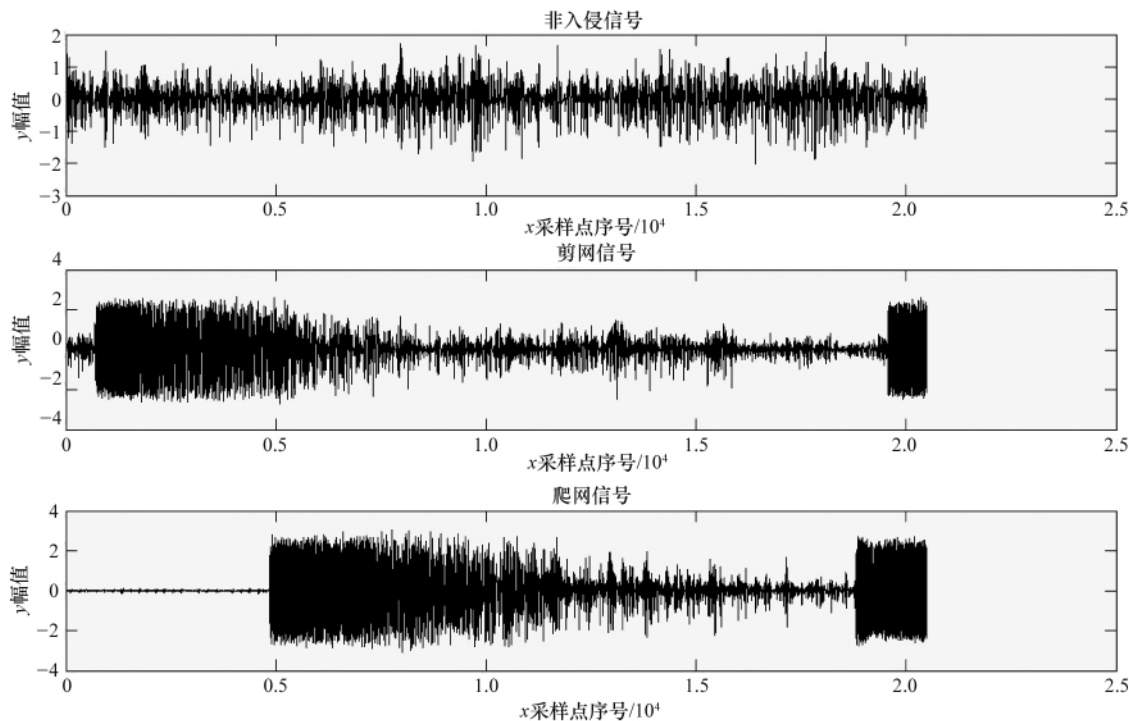


图 4 3 类信号的典型样本示例

Fig. 4 Samples of 3 typical signals

因此, 需要训练两个支持向量机。

样本训练过程如下: 首先, 将样本分为两类, 非入侵信号一类, 包含 $2n$ 个样本; 剪网入侵信号和爬网入侵信号一类, 两种信号分别包含 n 个样本。在对两类信号进行降噪处理后输入到向量机 SVM1 进行训练。然后, 再取 $2n$ 个剪网入侵信号样本和 $2n$ 个爬网入侵信号样本各作为一类, 进行降噪和信号分割后输入向量机 SVM2 进行训练。在训练 SVM2 的时候, 需要注意, 由于信号是采样系统在一定时间段内提取的,

因此有时候会出现这样的情况, 即在某一时间段内, 入侵信号由前一次入侵冲击的后半部分和后一次入侵冲击的前一部分组成, 如图 5 所示。

从图 5 可以看出, 该时间段内不存在一段完整的入侵信号, 因此, 只有当信号分割提取的信号的长度大于某一长度时, 才将其输入向量机中进行训练, 否则舍弃该信号。

信号识别过程如下: 对未知信号进行降噪和特征提取后输入 SVM1 中进行识别, 如果信号属于非入侵信号类, 则输出

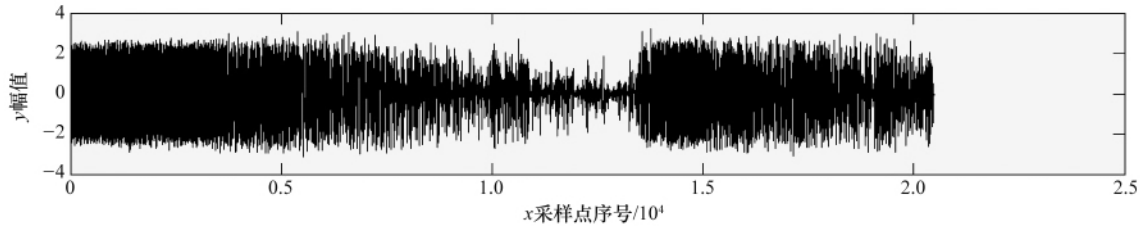


图5 由前一次入侵冲击的后半部分和后一次入侵冲击的前一部分组成的入侵信号

Fig. 5 Intrusion signal which consists of the last part of the former signal and the front part of latter signal

结果;如果信号属于另一类,则再对信号进行分割及特征提取后输入 SVM2 进行识别,并输出最后的分类结果。

SVM1 分类器的分类结果如表 1 所示。

表 1 光纤信号识别结果(SVM1)

Table 1 Recognition result of fiber signal by SVM1

光纤信号	输入信号数	识别正确数	准确率/%
非入侵信号	100	97	97
剪网信号和爬网信号	100	100	100

SVM2 分类器的分类结果如表 2 所示。

表 2 光纤信号识别结果(SVM2)

Table 2 Recognition result of fiber signal by SVM2

光纤信号	输入信号数	识别正确数	准确率/%
剪网信号	100	95	95
爬网信号	100	98	98

由实验结果可知,3类信号均有较高的识别率。由表 1 可知,所有入侵信号都能被正确地识别出来。实际应用中,安防系统最重要的是识别出入侵和非入侵信号,而入侵信号属于哪一类并不重要。因此,本文提出的一整套光纤信号分类和识别方法,对于实际的光纤安防系统是有效的。

4 结论

本文提出的光纤信号提取与分类算法之所以有效,是因为不同种类的扰动信号在不同频率下的能量分布不同。通过小波包变换可以得到信号在各个频率段的能量分布,并且支持向量机可以在短时间、小样本条件下对特征向量分类。因此,本文提出的一整套算法也适用于实际的应用环境。由于输入的特征向量维数还是过多(32 维),其中存在一些冗余特征,因此,在特征向量输入到 SVM 之前,还可以对这些特征进行统计分析,消去冗余特征等。

参考文献 (References)

[1] Mahmoud S S, Katsifolis J. Robust event classification for a fiber optic perimeter intrusion detection system using level crossing features and artificial neural networks[J]. *Proceedings of SPIE*, 2010, 7677: 767708.
 [2] Seedahmed S M, Jim K. Elimination of rain-induced nuisance alarms in

distributed fiber optic perimeter intrusion detection systems [J]. *Proceedings of SPIE*, 2009, 7316: 731604.
 [3] 饶云江, 吴敏, 冉曾令, 等. 基于准分布式 FBG 传感器的光纤入侵报警系统[J]. *传感技术学报*, 2007, 20(5): 45-49.
 Rao Yunjiang, Wu Min, Ran Zengling, et al. *Chinese Journal of Sensors and Actuators*, 2007, 20(5): 45-49.
 [4] 陶沛琳, 延凤平, 刘鹏, 等. 基于 Mach-Zehnder 干涉仪的光纤入侵行为识别系统[J]. *量子电子学报*, 2011, 28(2): 183-190.
 Tao Peilin, Yan Fengping, Liu Peng, et al. *Chinese Journal of Quantum Electronics*, 2011, 28(2): 183-190.
 [5] 赵杰, 丁吉, 万遂人, 等. 全光纤安防系统模式识别混合编程的实现[J]. *东南大学学报: 自然科学版*, 2011, 41(1): 41-46.
 Zhao Jie, Ding Ji, Wan Shuren, et al. *Journal of Southeast University: Natural Science Edition*, 2011, 41(1): 41-46.
 [6] 倪明, 陈智慧, 万遂人, 等. 基于神经网络和支持向量机的模式识别应用研究[R]. 东南大学校庆研究生学生报告会, 南京: 东南大学, 2011.
 Ni Ming, Chen Zhihui, Wan Suren, et al. Based on neural networks and support vector machine application of pattern recognition [R]. Postgraduate's Seminar of Southeast University's Anniversary. Nanjing: Southeast University, 2011.
 [7] Vidakovic B L, Ozoya C B. On time-dependent wavelet denoising[J]. *IEEE Transactions on Signal Processing*, 1998, 46(9): 2549-2551.
 [8] Donoho D L, Johnstone I M. Adapting to unknown smoothness via wavelet shrinkage [J]. *Journal of the American Statistical Association*, 1995, 90(432): 1200-1224.
 [9] Sachs L. Applied statistics: A handbook of techniques [M]. New York: Springer-Verlag, 1984.
 [10] Duda R O, Hart P E, Stork D G. Pattern classification [M]. Beijing: China Machine Press, 2009.
 [11] 夏思宇, 潘泓, 金立左, 等. 非平衡二叉树多类支持向量机分类方法[J]. *计算机工程与应用*, 2009, 45(17): 167-169.
 Xia Siyu, Pan Hong, Jin Lizuo, et al. *Computer Engineering and Applications*, 2009, 45(17): 167-169.

(责任编辑 朱宇)

《科技导报》“卷首语”栏目征稿

“卷首语”栏目每期邀请一位中国科学院院士和中国工程院院士就重大科技现象、事件,以及学科发展趋势、科学研究热点和前沿问题等,撰文发表个人的见解、意见和评论。本栏目欢迎院士投稿,每篇文章约 2000 字,同时请提供作者学术简历、工作照和签名电子文档。投稿邮箱:kjdbbjb@cast.org.cn。