

用于水声目标识别的自适应免疫特征选择算法

杨宏晖^{1,2}, 戴健¹, 孙进才¹, 杜方键¹, 彭圆², 李桂娟²

(1. 西北工业大学航海学院, 710072, 西安; 2. 水下测控技术国防科技重点实验室, 116013, 大连)

摘要: 针对水声目标小样本识别中样本数目有限而特征数目不断增加, 导致分类系统分类性能下降的问题, 提出了一种新的自适应免疫特征选择算法(AIFSA). 该算法先利用先验知识生成初始种群, 接着利用交叉、变异和新的自适应免疫算子指导种群进化, 每代中对分类贡献大且选择特征数目少的个体适应度值高. AIFSA 具有可以利用先验知识、收敛速度快以及优化特征子集维数小的优点. 提取了实测 4 类水声目标的多域特征, 进行特征选择和分类识别仿真实验, 结果表明: AIFSA 可以选择有效特征子集, 在特征维数下降 60% 的情况下, 支持向量机分类器的平均正确分类率下降很小; AIFSA 与标准遗传算法相比, 收敛快、稳定, 所得优化特征子集具有更高的正确分类率和更好的范化性能.

关键词: 自适应免疫特征选择; 水声目标识别; 水声目标多域特征; 小样本识别

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 0253-987X(2011)12-0028-05

A New Adaptive Immune Feature Selection Algorithm for Underwater Acoustic Target Classification

YANG Honghui^{1,2}, DAI Jian¹, SUN Jincai¹, DU Fangjian¹, PENG Yuan², LI Guijuan²

(1. College of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, China;

2. The Key Laboratory for Underwater Scientific Test and Control Technology, Dalian 116013, China)

Abstract: A novel adaptive immune multi-filed feature selection algorithm (AIFSA) for underwater acoustic targets classification is proposed. The AIFSA is proposed to address the problem that the classification performance in classifying underwater acoustic targets declines as the dimension of feature set increases, and that classifying underwater acoustic targets is a small-sample-size classification problem. The AIFSA generates an initial population using prior knowledge, and then generates new generations through repetitive application of mutation, crossover, and adaptive immune operator. In each iteration, individuals with less number of features and with high classification accuracy are given higher fitness values. The advantages of AIFSA include: using of prior knowledge, fast convergence, and small size of optimal feature subset. The multi-field features are extracted from 4 classes of underwater targets and used in feature selection and classification extracted. Experimental results show that the AIFSA can select the subset of efficient features, and there is only a small decline in the accuracy of SVM classifier when the number of features is decreased about 60%. Compared with the genetic algorithm, the AIFSA is more stable and achieves better converge speed, and the feature subset obtained by AIFSA achieves better classification performance and generalizability.

Keywords: adaptive immune feature selection; underwater acoustic target classification; multi-filed feature of underwater acoustic target; small-sample-size classification

长期以来,为了提高水声目标识别的正确率,研究人员利用各种信号处理方法提取了水声目标辐射噪声的多域特征,主要包括时域波形特征、频域分析特征、时频域分析特征、非线性特征和听觉特征^[1-7]。但是,特征数目的增加也带来以下问题^[8-9]:①水声目标样本获取困难,样本数目偏少和特征数目不断增加之间的矛盾使得水声目标识别系统的识别性能下降;②由于多域水声目标特征对目标描述的角度不同,一些特征对于识别任务是冗余的,特征维数过高势必导致识别系统复杂,在线识别速度过慢。为解决上述问题,目前已经研究的水声目标特征选择方法有遗传算法(Genetic Algorithm, GA)^[8,10]、向前向后搜索算法^[10]和粗集理论^[11-12]。基于粗集理论的特征选择算法本质上和向前向后搜索算法一样是一种顺序搜索算法,有可能丢掉最优解。遗传算法理论上可以搜索全局最优解,但是算法随机性大且运行速度过慢。近年来,借鉴人工免疫系统的主要机理的特征选择算法在一些领域取得了较好的结果^[13]。

本文提出了一种新的基于免疫优化的特征选择方法——自适应免疫特征选择算法(Adaptive Immune Feature Selection Algorithm, AIFSA),应用于水声目标多域特征选择与分类识别中,并利用实测 4 类水声目标数据的多域特征进行了特征选择和分类实验,结果证明了新算法能较好地解决水声目标特征选择问题。

1 水声目标多域特征提取

1.1 波形结构特征

(1)过零点分布特征:用 $S(\lambda_i)$ 表示噪声信号样本中波长为 λ_i 的过零点波的个数,则过零点波的波长分布概率函数为

$$P(\lambda_i) = S(\lambda_i) / \sum S(\lambda_i) \quad (1)$$

(2)峰间幅值分布特征:将归一化的峰值幅值等间隔划分成 N 个区间,落在区间 β_i 中的峰间幅值个数为 $S(\beta_i)$,峰间幅值概率分布函数为

$$P(\beta_i) = S(\beta_i) / \sum S(\beta_i) \quad (2)$$

(3)波长差分布特征:信号样本中相邻 2 个过零点波的波长差为 $\Delta\lambda_i$ 的数目为 $S(\Delta\lambda_i)$,则波长差分布概率函数为

$$P(\Delta\lambda_i) = S(\Delta\lambda_i) / \sum S(\Delta\lambda_i) \quad (3)$$

1.2 小波分析特征

(1)各级小波信号的过零点波的波长分布密度的信息熵:对信号进行小波分解,统计原信号和各级

小波高频分量过零点波的波长分布概率,按下式计算信息熵

$$H(\lambda) = - \sum P(\lambda_i) \log P(\lambda_i) \quad (4)$$

式中: λ_i 是样本变量 λ 的一个可能的取值; $P(\lambda_i)$ 是波长分布概率。

(2)各级小波信号的相似性特征:定义 2 个信号样本之间的相似性度量为

$$\delta(z_1, z_2) = \langle z_1, z_2 \rangle / (\|z_1\| \cdot \|z_2\|) \quad (5)$$

式中: z_1 和 z_2 分别表示两个不同的样本; $\langle \cdot \rangle$ 表示样本内积; $\| \cdot \|$ 为样本的模。

(3)小波分解低频包络特征:提取样本信号第 4 级小波分解的低频分量的包络信号,用高斯一阶马尔柯夫过程近似所提取的包络信号,用最小二乘法估计一阶马尔柯夫过程的参数作为水声目标的识别特征。

1.3 听觉特征

(1)听觉谱特征:依次计算声信号的 Bark 域功率谱、临界带听觉谱、响度级谱,最后经过响度级到响度的转换就得到声信号的听觉谱特征^[14]。

(2)Mel 频率倒谱特征(Mel Frequency Cepstral Coefficient, MFCC):将经过分帧处理的声信号的频谱转换为 Mel 频谱,再取对数处理,最后将对数频谱经过离散余弦变换得到 MFCC 特征^[15]。

2 基于自适应免疫的特征选择算法

2.1 算法原理

本文提出的 AIFSA 算法的原理框图如图 1 所示。算法首先利用先验知识生成初始种群,接着利用自适应免疫算子(自适应组合疫苗抽取、自适应接种疫苗和免疫选择)指导种群进化,每代计算每个个体的适应度,对分类贡献大且选择特征数目少的个体适应度值高,进化过程逐代重复,直到满足进化终止

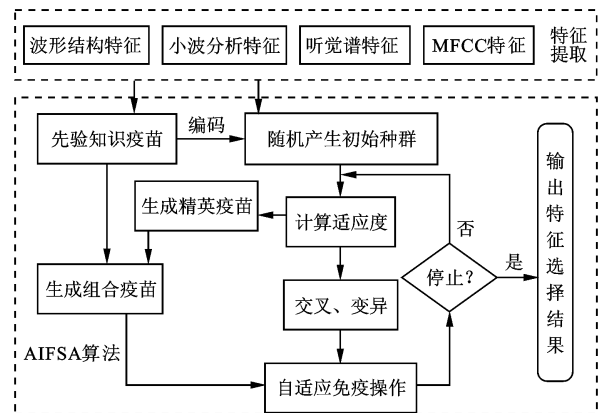


图 1 AIFSA 算法原理框图

条件.

2.2 问题的编码与初始种群的生成

特征选择问题解编码为 $x=(\alpha_1,\alpha_2,\cdots,\alpha_j\cdots,\alpha_n)$,其中 n 为特征数目, $\alpha_j=1$ 表示第 j 个特征被选中, $\alpha_j=0$ 表示未被选中. 根据长期积累的经验,已知上述水声目标多域特征中的某些特征具有较强的分类能力和推广性能,本文称这些特征为先验知识特征. 用这些特征编码,生成先验知识个体 H , H 中代表先验知识特征的基因为 1,其余的基因为 0. 初始种群由先验知识基因串 H 和其邻近的随机点组成.

2.3 适应度函数

适应度函数如下式所示

$$f(x_i)=\lambda acc(x_i)+(1-\lambda)d(x_i) \tag{6}$$

式中: $d(x_i)$ 是 1 除以第 i 个个体选择的特征个数,则 $1/n\leq d(x_i)\leq 1$; $acc(x_i)$ 是某次循环中第 i 个个体 x_i 的分类性能,用 SVM 分类器^[16]的分类性能表示, SVM 分类器的分类性能用交叉确认法计算, $acc(x_i)$ 取值范围在 0~1 之间;参数 λ 可以调整 2 个参数 $acc(x_i)$ 和 $d(x_i)$ 在适应度计算中所占的权重,一般识别系统对识别精度的要求高于对运行时间的要求,因此 λ 可以在 0.7~1 之间取值,本文取 $\lambda=0.7$.

2.4 自适应免疫算子

- (1)自适应组合疫苗抽取. 将 2.2 节中所述的先验知识个体作为先验知识疫苗 H ,从每代群体中抽取适应度最佳的个体作为精英疫苗,组合疫苗由第 k 代精英疫苗 B_k 和先验知识疫苗 H 组成.
- (2)自适应疫苗接种. 若先验知识疫苗的适应度低于当代个体适应度的平均值,则用精英疫苗接种;若先验知识疫苗的适应度高于当代适应度的平均值,则 2 个疫苗都用于接种. 接种是依照接种概率,按疫苗修改个体上的基因状态.
- (3)免疫退火选择. 计算接种了疫苗个体的适应

度,若不如父代,说明出现了退化,则该个体将被父代所取代. 在第 k 代群体 $A_k=(x_1,x_2,\cdots,x_n)$ 中以概率 $P(x_i)$ 选择个体 x_i 进入新的父代群体, $P(x_i)$ 按下式计算

$$P(x_i)=\exp(f(x_i)/T_k)/\sum_{i=1}^n\exp(f(x_i)/T_k),$$
$$T_k=\ln(T_0/k+1) \tag{7}$$

式中: T_0 取值 100.

3 实验与讨论

3.1 实验数据

本文所用的实测水声目标分为 A、B、C、D 4 类,每类 480 个样本,样本总数为 1 920,每个样本提取 71 维多域特征. 实验比较了特征选择前后 SVM 对 4 类水声目标的正确分类率,并比较了所提出的 AIFSA 与 GA 的性能,2 种算法的参数设置如表 1 所示.

表 1 2 种算法的参数设置

算法	种群大小	选择概率	交叉概率	变异概率	接种概率	进化代数
AIFSA	100	0.8	0.8	0.08	0.25	500
GA	100	0.8	0.8	0.08		500

3.2 特征选择与分类实验

对水声目标数据,分别用 AIFSA 和 GA 进行特征选择实验,比较了 2 种算法在特征选择过程中的适应度变化、特征选择结果及特征选择前后 SVM 的分类性能. 实验结果及讨论如下.

- (1)特征选择过程中适应度变化的比较. 图 2 给出了 AIFSA 和 GA 运算 500 代的每代归一化平均适应度值的变化曲线. 从图 2 可以看出, AIFSA 的适应度曲线上升较快而且上升后趋于稳定,而 GA 的适应度曲线上升趋势不明显且波动很大,这说明

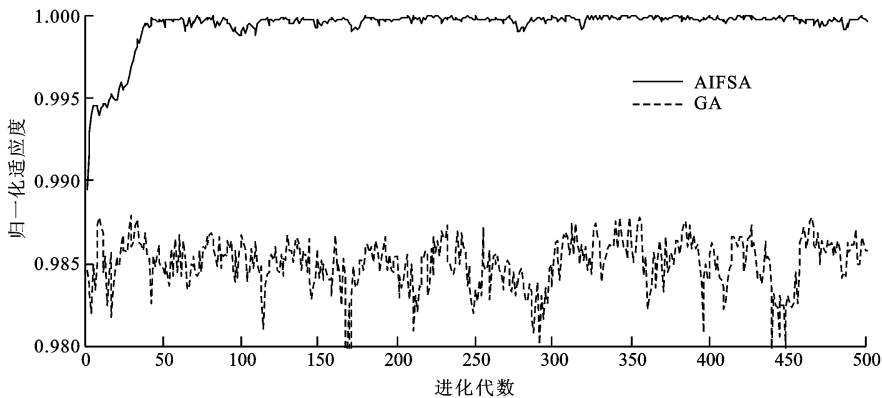


图 2 AIFSA 和 GA 的适应度曲线

AIFSA 比 GA 收敛性好,能更快地找到一定条件下的最优解,进化过程稳定,可以抑制优化过程中出现的退化现象,而 GA 则随机性较强,收敛速度慢,较易出现退化现象。

(2)特征选择结果的比较. 分别将 AIFSA 和 GA 运行 K 代,统计每代的最佳个体. 定义每个特征在全部 K 代的最佳个体中的选择次数除以 K 为每个特征的特征重要性指数. 用 AIFSA 和 GA 得到的 71 维特征的特征重要性指数如图 3 和图 4 所示. 从图 3 和图 4 可以看出,在 AIFSA 运行过程中,大约有 15 个特征被选择的次数明显高于其他特征,但是在 GA 运行过程中,仅有个别特征的重要性指数明显高于其他特征的重要性指数,其余特征的重要性指数没有显著差别. 这说明,本文提出的 AIFSA 充分利用了先验知识,所设计的适应度函数及自适应免疫操作有效地指导了最优特征子集的搜索。

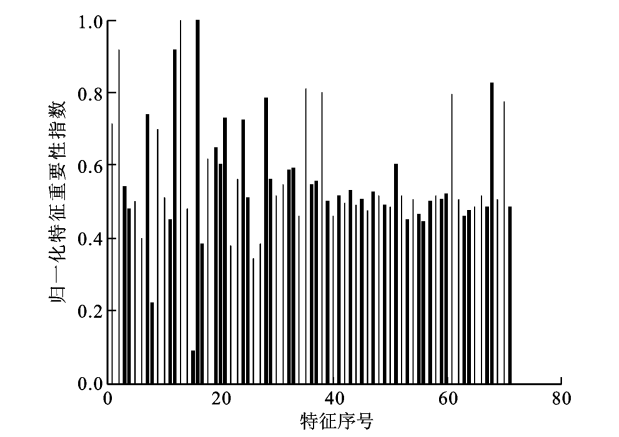


图 3 AIFSA 的特征选择结果

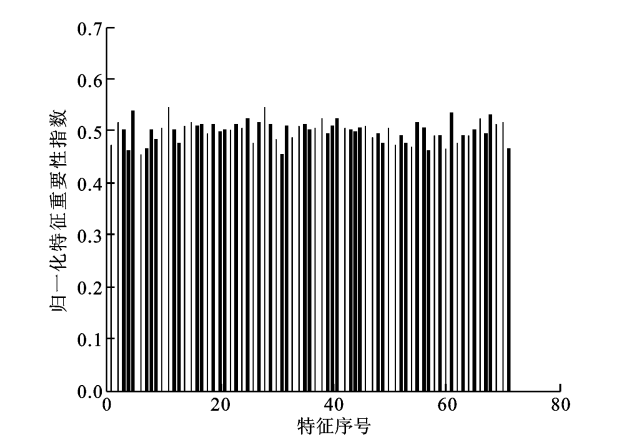


图 4 GA 的特征选择结果

(3)特征选择前后 SVM 分类性能的比较. 为了验证 AIFSA 与 GA 特征选择算法的有效性,分别用 SVM 分类器对两种算法得到的 5 维到 70 维优化特

征子集进行分类实验. 用 10 折交叉确认法来估计 SVM 对水声目标的分类性能. 特征选择前后 SVM 分类性能由 3 个参数表达——平均正确分类率、正确分类率的方差以及特征选择前后 SVM 所需的分类时间. 实验结果如图 5 和图 6 所示。

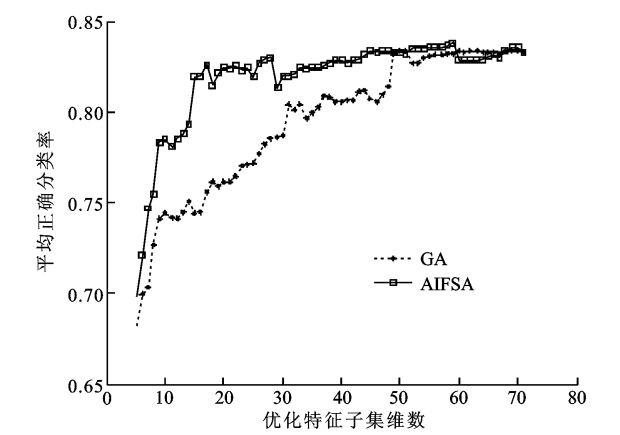


图 5 平均正确分类率

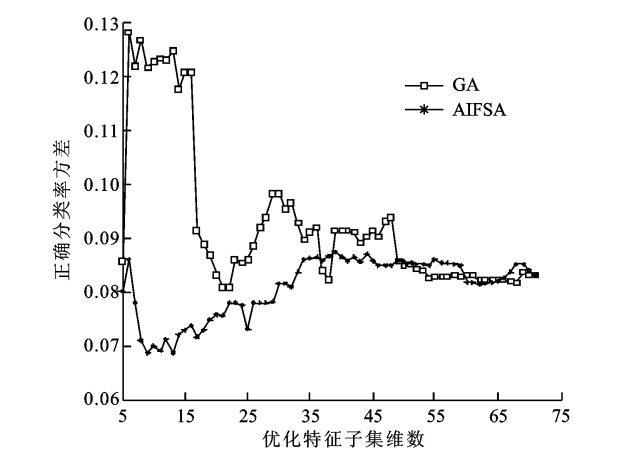


图 6 正确分类率的方差

从图 5 可知,当优化特征子集维数低于 50 维时,AIFSA 选出的特征子集比 GA 算法所选特征子集的平均正确分类率要高,而且 AIFSA 在保持正确分类率下降很小的情况下,可以减少近 60% 的特征. 从图 6 可以看到:在特征子集维数相同的情况下,AIFSA 选出的特征子集正确分类率的方差比 GA 的要小,这说明 AIFSA 选出的特征子集具有更好的范化能力和更好的稳健性;当优化特征子集维数高于 50 维后,AIFSA 的平均正确分类率变化不大,但是方差增大,这是由于随着特征维数的增高,特征子集中含有冗余特征,特征子集范化能力降低,此时,AIFSA 与 GA 选出的特征子集的分类性能没有明显差别. 从实验中还可知,用 15 维特征子集比用 71 维特征子集做分类的时间可以减少 30%。

(4)敏感特征分析. 从上面的实验结果及分析可知,利用 AIFSA 选择出的 15 个重要性指数较高的特征训练的分类器具有较高的正确分类率和范化能力,这说明,这些特征对目标分类贡献较高,具有更敏感的分类信息. 表 2 列出了前 5 名最敏感的特征.

表 2 水下目标分类敏感特征

序号	特征域	是否与目标的物理量相对应
1	小波分解信息熵	否
2	小波分解信息熵	否
3	听觉谱	舰船螺旋桨引起的划水声和节奏声
4	听觉谱	某类目标在运动时引起的特定频段噪声
5	波形结构	舰船螺旋桨引起的划水声和节奏声

表 2 中:第 1 个特征是小波分解信号的过零点波的波长分布密度的信息熵;第 2 个特征是小波信号的相似性特征,由于小波分析特征计算的复杂性,尽管这 2 个小波分析特征都包含着重要而敏感的分类信息,但目前仍无法和目标具体的物理属性建立非常明确的联系,这需要今后进一步深入研究;第 3 个和第 5 个特征是听觉谱和波形结构特征,这 2 个特征从不同域描述了各类水声目标在运动过程中螺旋桨引起的划水声和节奏声;第 4 个特征是听觉谱特征,这个特征描述了某类目标在运动时会引起特定频段噪声的特性,这是这类目标特有而其他目标不具备的特性. 实验结果证明,AIFSA 可以在保持 SVM 正确分类率下降很小的情况下,选择出对分类重要而敏感的特征,降低特征维数,特征选择后的特征子集具有更好的范化性能,减少了分类时间.

4 结 语

本文提出了用于水声目标识别的自适应免疫特征选择算法,提出了可利用先验知识的初始种群生成方法以及自适应免疫算子和适应度函数,并采用本文算法对 4 类实测水声目标数据进行特征选择和分类仿真实验. 结果表明,AIFSA 算法应用于水声目标识别领域具有有效性和鲁棒性,相对于 GA 算法具有收敛快、稳定、特征选择结果更优的特点,能较好地解决水声目标多域特征选择问题.

参考文献:

[1] 史广智,胡均川,笪良龙. 基于模型匹配的舰船螺旋桨噪声特征精细分析[J]. 声学学报,2009,34(5): 401-407.
SHI Guangzhi, HU Junchuan, DA Lianglong. Feature fine analysis of noise radiated from ship propeller based

on model matching [J]. Acta Acustica, 2009, 34(5): 401-407.
[2] 谢骏,胡均川,笪良龙,等. 舰船目标噪声调制包络谱融合算法研究[J]. 系统仿真学报, 2009, 21(1):62-64.
XIE Jun, HUN Junchuan, DA Lianglong, et al. Research on fusion arithmetic for modulation envelope spectrum of navy vessel's noise [J]. Journal of System Simulation, 2009, 21(1): 62-64.
[3] 樊养余,孙进才. 基于高阶谱的舰船辐射噪声特征提取[J]. 声学学报,1999, 24(6): 611-616.
FAN Yangyu, SUN Jincai. Feature extraction of ship-radiated noise using higher-order spectrum [J]. Acta Acustica, 1999, 24(6): 611-616.
[4] 李雪耀,彭圆,林良骥,等. 基于小波与三次样条插值的包迹谱的水下目标分类研究[J]. 声学学报, 2004, 29(1):63-67.
LI Xueyao, PENG Yuan, LIN Liangji, et al. Study on classification of underwater targets based on modulation spectrum by wavelet transforms and cubic spline technique [J]. Acta Acustica, 2004, 29(1):63-67.
[5] 王娜,陈克安. 水下噪声音色属性回归模型及其在目标识别中的应用 [J]. 物理学报,2010, 59 (4):2873-2881.
WANG Na, CHEN Ke'an. Regression model of timbre attribute for underwater noise and its application to target recognition [J]. Acta Physica Sinica, 2010, 59 (4):2873-2881.
[6] TUCKER S, BROWN G J. Classification of transient sonar sounds using perceptually motivated features [J]. IEEE Journal of Oceanic Engineering, 2005, 30 (3):588-600.
[7] VACCARO R J. The past, present and future of underwater acoustic signal processing [J]. IEEE Signal Processing Magazine, 1998, 15(4): 21-51.
[8] 杨宏晖,孙进才,袁骏. 基于支持向量机和遗传算法的水下目标特征选择算法[J]. 西北工业大学学报, 2005, 23(4): 512-515.
YANG Honghui, SUN Jincai, YUAN Jun. A new method for feature selection for underwater acoustic targets [J]. Journal of Northwestern Polytechnical University, 2005, 23 (4): 512-515.
[9] YANG Honghui, SUN Jincai. Feature selection for multi-class underwater acoustic targets via SVMs and genetic algorithms [EB/OL] (2005-10-16)[2011-01-10]. http://apps.webofknowledge.com/full_record.do? qid=5&log_event=no&viewType=full Record&product=WOS &SID=P2e517AOA66IcjaecmB&page=1&doc=1.
[10] ZHAO Zheng, LIU Huan. Searching for interacting

- features in subset selection [J]. Intelligent Data Analysis, 2009, 13(2):207-228.
- [11] SWINIARSKI R W, SKOWRON A. Rough set methods in feature selection and recognition [J]. Pattern Recognition Letters, 2003, 24(6): 833-849.
- [12] 詹艳梅,孙进才. 基于粗糙集理论的特征选择方法[J]. 自然科学进展, 2004, 12(12):1483-1487.
ZHAN Yanmei, SUN Jincai. The feature selection method based on rough theory [J]. Nature Science Progressing, 2004, 12(12):1483-1487.
- [13] 朱虎明,焦李成. 基于免疫记忆克隆的特征选择[J]. 西安交通大学学报, 2008, 42(6): 679-682.
ZHU Huming, JIAO Licheng. Feature selection via immune memory clone [J]. Journal of Xi'an Jiaotong University, 2008, 42(6): 679-682.
- [14] 杨宏晖,候宏,曾向阳,等. 基于声信号人耳听觉谱特征的风机故障诊断[J]. 仪器仪表学报, 2009, 30(1): 175-179.
YANG Honghui, HOU Hong, ZENG Xiangyang, et al. Fault diagnosis for fan based on auditory spectrum feature of sound signal [J]. Chinese Journal of Scientific Instrument, 2009, 30(1): 175-179.
- [15] DAVIS S, MERMELSTEIN P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences [J]. IEEE Trans ASSP, 1980, 28(4): 357-366.
- [16] VAPNIK V. The nature of statistical learning theory [M]. New York, USA: Springer-Verlag, 2000.

[本刊相关文献链接]

- 应用水波弥散特性的舰船航迹三维可视化新方法. 西安交通大学学报, 2011, 45(8):49-53.
- 强干扰环境水下运动阵列方位估计的新方法. 西安交通大学学报, 2010, 44(12):71-75.
- 频谱包络滤波器及其应用. 西安交通大学学报, 2010, 44(8):48-52.
- 利用粒子群优化的人脸特征提取识别算法. 西安交通大学学报, 2010, 44(4):48-51.
- 图像分割的谱聚类集成算法. 西安交通大学学报, 2010, 44(6):93-98.
- 一种考虑环境作用的协同免疫遗传算法. 西安交通大学学报, 2009, 43(11):80-84.
- 基于特征层和二代曲波变换的多模生物特征融合识别方法. 西安交通大学学报, 2009, 43(10):32-36.
- 三维子空间约束的遮挡点恢复方法. 西安交通大学学报, 2009, 43(4):10-13.
- 采用二代曲波变换和反向传播神经网络的人脸识别方法. 西安交通大学学报, 2008, 42(10):1213-1216.

(编辑 刘杨)