

用于水声目标识别的自适应遗传样本选择算法

戴 健, 杨宏晖, 王 芸, 孙进才

(西北工业大学航海学院, 西安 710072)

摘要: 针对训练样本集中含有噪声样本、冗余样本以及无关样本, 导致分类系统分类性能下降、不稳定的水声目标识别问题, 提出了一种新的自适应遗传样本选择算法(Adaptive Genetic Instance Selection Algorithm, AGISA)。算法先随机生成初始种群, 接着利用设计的遗传算子(跨代选择、自适应交叉和简化最近邻变异)指导种群进化, 每代中对分类贡献大且选择样本数目少的个体适应度值高。提取了实测 3 类水声目标的多域特征, 进行样本选择和分类识别仿真实验, 结果表明: AGISA 可以选出有效样本子集, 在样本维数下降约 73% 的情况下, 支持向量机分类器的正确分类率能提高约 2.5%; 并且 AGISA 具有较好的收敛性、稳定性, 所得优化样本子集具有较好泛化能力且能明显减少分类的时间。

关键词: 自适应遗传样本选择; 水声目标识别; 样本选择; 分类识别

中图分类号: TP312

文献标识码: A

文章编号: 1000-3630(2013)-04-0332-04

DOI 编码: 10.3969/j.issn1000-3630.2013.04.015

An adaptive genetic instance selection algorithm for underwater acoustic target classification

DAI Jian, YANG Hong-hui, WANG Yun, SUN Jin-cai

(School of Marine Engineering, Northwestern Polytechnical University, Xi'an 710072, China)

Abstract: In this paper, a new adaptive genetic instance selection algorithm (AGISA) is proposed for underwater acoustic target classification. The AGISA is proposed to address the problem that the classification performance in classifying underwater acoustic targets declines and becomes unstable as the training instance set contains noise samples, redundant samples and irrelevant samples. The AICISA generates an initial population randomly, and then generates new generations through designed genetic operators (cross-generational selection, adaptive crossover and reduced nearest neighbor mutation). In each generation, antibodies with less number of features and with high classification accuracy are given higher fitness values. The multi-field features are extracted from 3 classes of underwater targets, and used in instance selection and classification experiments. Experimental results show that AGISA can select the subset of efficient instances, and there is about 2% increase in the accuracy of SVM classifier when the number of features decreases about 70%. AGISA has good convergence and stability, and the instance subset obtained by AGISA achieves good generalization ability, which can reduce the classification time obviously.

Key words: adaptive genetic instance selection, underwater acoustic target classification, instance selection, sample classification

0 引言

基于有监督学习理论的水声目标识别系统主要包括特征提取、特征选择和分类器设计这几个部分。长期以来, 为了提高水声目标识别的正确率, 研究人员从不同的角度对水声目标辐射噪声原始信号进行了分析和研究, 并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^[1-9], 主要有: 时域波形特征、频域分析特征、时频域分析特征、

非线性特征和听觉谱特征。但是, 提取的水声特征样本集未经过样本选择处理就用于训练分类器会带来如下问题: (1) 样本集含有很多冗余样本; (2) 样本集含有与应用无关的不相关样本; (3) 样本集含有在采集录制过程中掺入的错误样本和噪声样本; (4) 样本集含有在采集录制过程中缺失的某些目标样本; (5) 样本集过大、样本之间关系过于复杂。这些样本参与训练分类器, 必将导致分类器的分类识别性能下降、不稳定。然而在水声目标识别研究中, 研究工作主要集中在特征提取、特征选择^[10]和分类器设计等方面, 关于样本选择方法的研究报道甚少, 因此, 探索用于水声目标识别的样本选择方法具有重要的理论意义和实际价值。

收稿日期: 2012-05-14; 修回日期: 2012-08-30

作者简介: 戴健(1988-), 男, 江苏人, 硕士研究生, 研究方向为模式识别、声信号信息处理、水声目标识别。

通讯作者: 杨宏晖, E-mail: hhyang@nwpu.edu.cn

本文提出了一种新的样本选择方法——自适应遗传样本选择算法(Adaptive Genetic Instance Selection Algorithm, AGISA)，应用于水声目标样本选择与分类识别中。利用实测 3 类水声目标数据进行了样本选择和分类实验，实验结果证明了新算法具有较好的收敛性、稳定性和泛化性能，能够选出有效样本子集，去除无关样本、冗余样本和噪声样本，较好地解决了水声目标样本选择问题。

1 自适应遗传样本选择算法

1.1 算法原理

本文提出的自适应遗传样本选择算法的原理框图如图 1 所示。算法首先随机生成初始种群，接着利用跨代选择、自适应交叉和简化最近邻变异^[11]操作指导种群进化，每代计算每个个体的适应度，对分类贡献大且选择样本数目少的个体适应度高，进化过程逐代重复，直到满足进化终止条件。

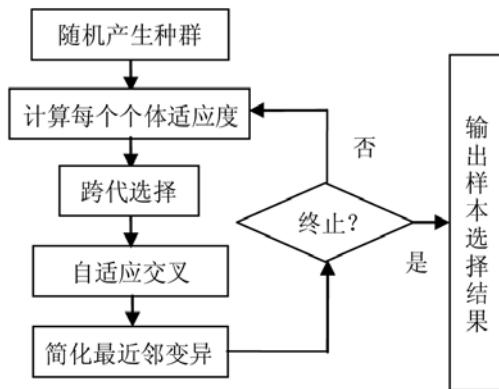


图 1 算法原理框图
Fig.1 Block diagram of algorithm principle

1.2 问题的编码

首先将样本选择问题解编码成二进制基因链码，编码为 $y=(\beta_1, \beta_2, \dots, \beta_j, \dots, \beta_n)$ ，其中， n 为样本数目， $\beta_j=1$ 表示第 j 个样本被选， $\beta_j=0$ 表示第 j 个样本未被选中。

1.3 适应度函数

为了充分考虑识别系统对识别精度和运行时间的要求，算法所用适应度函数^[8]如式(1)所示：

$$f(y_i) = \alpha(1 - \varepsilon_w(y_i)) + (1 - \alpha)\gamma(y_i) \quad (1)$$

其中 $\varepsilon_w(y_i)$ 是某次循环中第 i 个个体 y_i 的分类性能，用 SVM 分类器(Support Vector Machine, SVM)^[12,13] 的分类错误率表示，取值范围在 0~1 之间； $\gamma(y_i)$ 是 1 除以第 i 个个体 x_i 选择的样本个数，取值范围为 $1/n \leq \gamma(y_i) \leq 1$ 。参数 α 可以调整两个参数 $\varepsilon_w(y_i)$ 和

$\gamma(y_i)$ 在适应度计算中所占的权重，一般情况下，识别系统对识别精度的要求高于对运行时间的要求，因此，本文取 $\alpha=0.8$ 。

1.4 遗传算子

遗传算法主要由三种遗传算子指导种群进化，三种算子即选择算子、交叉算子和变异算子，本文提出的自适应遗传样本选择算法所用的三种遗传算子如下：

(1) 跨代选择：在本文提出的自适应遗传样本选择算法中，将上代种群与通过交叉和变异方法产生的个体群混合在一起，依据个体的适应度，从中选择较优的个体遗传到下一代。在进化过程中，该选择算子既能保证个体的多样性，也能有效抑制进化过程中出现的退化现象。

(2) 自适应交叉：算法首先设定一阈值 HD，接着计算两个父代个体间的海明距离 d ，即两个父代个体位值相异的位数为 d ，若 $d \geq HD$ ，则从两个父代中随机选取 $d/2$ 个对应位置，实行两个父代个体位值的互换；若 $d < Hd$ ，则两个父代个体不进行交叉操作；并且在种群进化收敛的同时，逐渐地减小该阈值。该交叉算子能有效避免过早收敛。

(3) 简化最近邻变异：设初始个体 y ，首先取 $T=y$ ，接着从样本集 T 中按序删除一个样本 β_j ，用 SVM 分类器对样本集 T 进行分类，如果没有导致 y 中样本被 T 中样本错误地分类，则删除该样本，循环往复，直到 y 中所有原先被选择的样本都被扫描一遍。该操作在算法中可以避免过早收敛或搜索停滞，在样本选择过程中能有效删除无关样本、冗余样本以及噪声样本，减少样本选择数目。原理框图如图 2 所示。

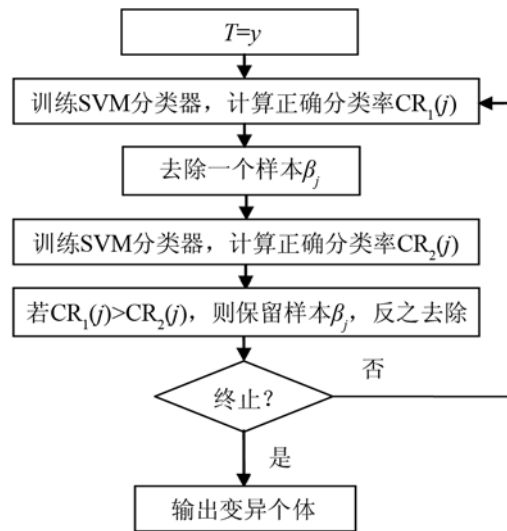


图 2 简化最近邻变异原理框图
Fig.2 Block diagram of Reduced Nearest Neighbor mutation principle

2 实验与讨论

2.1 实验数据

实验所用的实测水声目标信号分为 A、B、C 三类，一类 1315 个样本，二类 1849 个样本，三类 2366 个样本，总共 5530 个样本；每个样本提取了 3 类共 35 维水声目标多域特征，分别是波形结构特征(过零点分布特征、峰间幅值分布特征和波长差分布特征)、小波分析特征(各级小波信号的过零点波的波长分布密度的信息熵、各级小波信号的相似特征和小波分解低频包络特征)以及听觉谱特征^[14]；每类选取 300 个共 900 个样本作为训练样本，总样本作为测试样本。实验比较了样本选择前后 SVM 对 3 类水声目标的正确分类率，分析了提出的自适应遗传样本选择算法的性能，参数设置如表 1 所示。

表 1 参数设置
Table 1 Parameters setting

进化代数	种群大小	海明距离	自适应交叉概率	简化最近邻变异概率
100	100	450	0.8	0.08

2.2 样本选择与分类实验

对水声目标数据，用自适应遗传样本选择算法进行样本选择实验，分析了算法的样本选择过程中样本选择结果和样本选择前后 SVM 的分类性能，实验结果及讨论如下。

(1) 样本选择结果

将 AGISA 运行 K 代，统计每代的最佳个体。将每个样本在全部 K 代的最佳个体中的选择次数除以 K 作为每个样本的归一化样本重要性指数。用 AGISA 得到的 900 维样本的归一化样本重要性指数如图 3 所示。

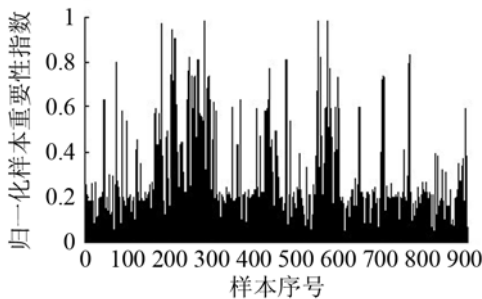


图 3 AGISA 样本选择结果
Fig.3 Result of instance selection of AGISA

从图 3 可以看出，在 AGISA 运行过程中，大约有 1/4 的样本被选择的次数明显高于其它样本，这些样本对 SVM 分类器正确分类的贡献更高。这

说明，提出的 AGISA 充分利用了所设计的遗传算子有效地指导了最优样本子集的搜索。

(2) 样本选择前后 SVM 的分类性能

为了验证 AGISA 的有效性，用 SVM 分类器对算法得到的 10 维、20 维、30 维...900 维优化样本子集进行分类实验。样本选择前后 SVM 的分类性能由每类样本及总样本的正确分类率和样本选择前后 SVM 分类所需时间 2 个参数表达。实验结果如图 4 和表 2 所示。

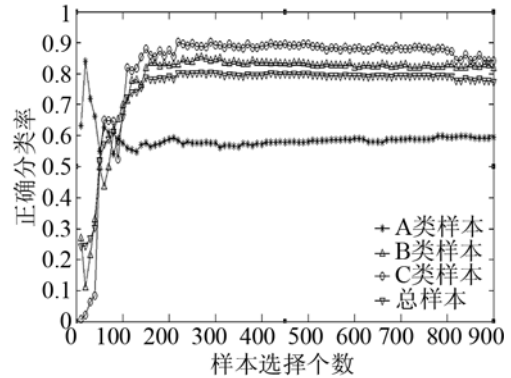


图 4 正确分类率
Fig.4 Correct classification rate

由图 4 可知，每类正确分类率随着样本维数的增加，先上升再慢慢趋于平稳后略有下降，并且在此过程中会达到一峰值(样本子集大约在 250 维左右时，总样本的正确识别率达到峰值)，这说明随着样本维数的增加，样本子集中含有的无关样本、冗余样本以及噪声也随之增加，样本子集的泛化能力则随之降低。

表 2 样本选择前后对比
Table 2 Classification results contrast

	正确分类率	分类时间/s	样本维数
样本选择前	0.7741	0.5425	900
样本选择后	0.7991	0.1940	242

从表 2 可以看出，AGISA 在样本维数减少约 73% 的情况下，大概能有效提高 2.5% 的正确分类率；并且用 242 维样本子集比用 900 维样本子集做分类的时间可以减少 64% 左右。这说明，AGISA 能选出有效的样本子集，去除无关样本、冗余样本以及噪声样本，减少学习算法的计算量，提高分类的正确率。

3 结论

本文提出了用于水声目标识别的自适应遗传样本选择算法，设计了新的遗传算子指导种群进化。利用实测 3 类水声目标数据进行样本选择和分

类实验。实验结果证明了自适应遗传样本选择算法具有较好的收敛性、稳定性和泛化性能, 以及应用于水声目标识别领域的有效性和鲁棒性, 能够选出有效样本子集, 去除无关样本、冗余样本和噪声样本, 较好地解决水声目标样本选择问题。

参 考 文 献

- [1] 史广智, 胡均川, 笄良龙. 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, 笄良龙, 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 Envelop Spectrum of Navy Vessels 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 Liangyi, 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, **4**(59): 2873-2881.
WANG Na, CHEN Kean. Regression model of timbre attribute for underwater noise and its application to target recognition[J]. Acta Physica Sinica, 2010, **4**(59): 2873-2881.
- [6] TUCKER S. Auditory analysis of sonar signals[R]. Ph. D. Transfer Report, 2001.
- [7] CHEN T. The past, present and future of underwater acoustic signal processing[J]. IEEE Signal Processing Magazine, 1996, 67-94.
- [8] 史广智, 胡均川. 基于小波包和 $1\frac{1}{2}$ 维谱的舰船辐射噪声频域特征提取及融合[J]. 声学技术, 2004, **23**(1): 4-7.
SHI Guangzhi, HU Junchuan. Extraction and fusion of frequency-domain features from ship-radiated noise based on wavelet packet and $1\frac{1}{2}$ dimension spectrum[J]. Technical Acoustics, 2004, **23**(1): 4-7.
- [9] 周有, 相敬林. 基于混合基函数分解的舰船辐射噪声分解算法[J]. 声学技术, 2007, **26**(5): 942-945.
ZHOU You, XIANG Jinglin. Wide-band signal analysis of ship-radiated noise based on mixed basis function[J]. Technical Acoustics, 2007, **26**(5): 942-945.
- [10] 杨宏晖, 戴健, 孙进才, 等. 用于水声目标识别的自适应免疫特征选择算法[J]. 西安交通大学学报, 2011, **45**(12): 28-32.
YANG Honghui, DAI Jian, SUN Jincai, et al. Underwater acoustic target recognition based on Adaptive Immune Feature Selection Algorithm[J]. Journal of Xi'an Jiaotong University, 2011, **45**(12): 28-32.
- [11] Nicolas Garcia-Pedrajas. Constructing ensembles of classifiers by means of weighted instance selection[J], IEEE Transactions on Neural Networks, 2009, **20**(2): 258-277.
- [12] VAPNIK V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 2000.
- [13] 赵亚楠, 李钢虎, 曾渊. 基于最小均方无失真相应和支持向量机的被动声纳目标识别[J]. 声学技术, 2011, **30**(3): 223-226.
ZHAO Ya'nian, LI Ganghu, ZENG Yuan. A method of passive sonar target recognition based on Minimum Variance Distortionless Response and Support Vector Machine[J]. Technical Acoustics, 2011, **30**(3): 223-226.
- [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.