

一种基于 NNIA 多目标优化的代价敏感决策树构建方法

赵士伟, 卓 力, 王素玉, 沈兰荪

(北京工业大学信号与信息处理研究室, 北京 100124)

摘 要: 本文提出了一种基于非支配邻域免疫算法(NNIA, Nondominated Neighbor Immune Algorithm)多目标优化的代价敏感决策树构建方法. 将平均误分类代价和平均测试代价作为两个优化目标, 然后利用 NNIA 对决策树进行优化, 最终获取了一组 Pareto 最优的决策树. 对多个测试集的测试结果表明, 与 C4.5 算法和 CSDB(Cost Sensitive Decision Tree)算法比较, 本文方法不仅在平均误分类代价和平均测试代价两方面均可以取得优于两者的性能, 而且获得的决策树具有更小的规模, 泛化能力更强.

关键词: 代价敏感; 误分类代价; 测试代价; 多目标优化; 决策树

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2011) 10-2348-05

A Multi-Objective Optimization Based Constructing Cost-Sensitive Decision Trees Method

ZHAO Shi-wei, ZHUO Li, WANG Su-yu, SHEN Lan-sun

(Signal and Information Processing Laboratory, Beijing University of Technology, Beijing 100124, China)

Abstract: A novel method of constructing the cost-sensitive decision trees based on multi-objective optimization is proposed in this paper. The average misclassification cost and the average test cost are treated as the two optimization objectives. NNIA (Nondominated Neighbor Immune Algorithm) is exploited to optimize the decision trees. And some Pareto decision trees are finally obtained. Experimental results show that, compared with the C4.5 algorithm and CSDB (Cost Sensitive Decision Tree) algorithm, the proposed method in this paper can not only outperform these two methods in terms of the two above objectives but also achieve smaller size of the decision trees and stronger generalization ability.

Key words: cost-sensitive; misclassification costs; test costs; multi-objective optimization; decision tree

1 引言

决策树(Decision Tree)作为一种分类方法已经在众多领域得到了应用^[1]. 然而, 现有的决策树算法主要关注的是分类准确率, 而在实际应用中, 除了分类准确率外, 误分类代价和测试代价等也是需要重点考虑的因素^[2]. 文献[2]将不同误分类形式所带来的影响称为误分类代价(Misclassification Costs), 将分类过程中获取需要的指标、特征等(本文统称为属性)所付出的花费称为测试代价(Test Costs).

在分类过程中同时考虑多种代价的问题通常称为多代价敏感分类, 解决这种问题主要有两种策略^[3]: 一种是将多种代价转换为一种综合代价, 另外一种是将每种代价视作一个目标并利用多目标优化技术寻优. 目前

多代价敏感决策树构建方法主要采用第一种方式, 即在构建代价敏感的决策树过程中, 将不同代价在同一尺度下进行衡量^[4~6], 或者将不同代价通过加权的方式合并为一种新的代价^[7]. 这种方法建立的代价敏感决策树能够在各种代价之间获得比较折中的结果, 然而在不同的代价之间, 例如误分类代价和测试代价, 往往很难用同一尺度进行衡量, 也难以根据不同代价的重要程度设置权值^[8]. 为此, 文献[3]提出了一种利用多目标优化技术解决多代价敏感分类问题的方法, 该方法将每种代价定义为一种优化目标, 从而避免了为不同的代价设置统一尺度或者权值.

非支配邻域免疫算法(NNIA, Nondominated Neighbor Immune Algorithm)是通过在人工免疫系统中引入一种新的非支配邻域选择策略提出的一种多目标优化方法^[9].

与其他具有代表性的进化多目标优化算法 NSGA-II、SPEA2 及 PESA-II 相比, NNIA 运算复杂度低、更为有效可行。

本文提出了一种基于 NNIA 的代价敏感决策树构建方法,将平均误分类代价和平均测试代价作为两个优化目标,利用 NNIA 对决策树进行优化.将决策树看作是免疫进化中的抗体,通过研究决策树抗体的随机构建方法、剪枝策略、以及变异操作等,最终建立了代价敏感的决策树.对部分 UCI 公共测试集^[10]的测试结果表明,与 C4.5 算法^[11]及 CSDB (Cost Sensitive Decision Tree)算法^[8]相比,本文算法在误分类代价和测试代价两个指标上均取得了优于两者的性能,并且本文算法得到的决策树规模更小,泛化能力更强。

2 基于 NNIA 算法的代价敏感决策树构建

NNIA 是一种基于人工免疫系统的多目标优化算法.该算法模拟了免疫响应中多样性抗体共生、少数抗体激活的现象,通过采用非支配邻域的个体选择、根据拥挤距离的比例克隆等策略加强了对当前 Pareto 前沿面中较稀疏区域的搜索^[9].因此, NNIA 能够获得分布均匀的 Pareto 最优解,并且同现有的进化多目标优化方法相比, NNIA 计算复杂度低.本文采用该算法来构建代价敏感决策树。

NNIA 算法的基本步骤包括:初始抗体群的建立、优势抗体群更新、终止条件判断、活性抗体群的选择、比例克隆,以及重组和超变异操作^[9].为了使 NNIA 更适应决策树抗体并且进一步降低复杂度,本文在构建代价敏感决策树过程中对 NNIA 算法进行了改进,这种基于 NNIA 算法的代价敏感决策树的构建框图如图 1 所示。

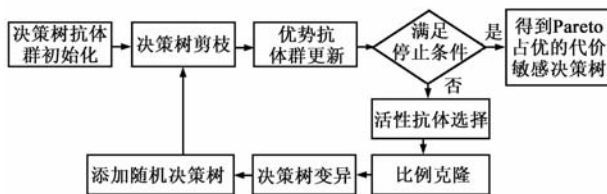


图1 基于NNIA的代价敏感决策树构建框图

改进的 NNIA 算法在每次迭代过程中均增加了对决策树抗体的剪枝操作,以得到更为精简鲁棒的决策树.由于本文提出的变异操作需要利用决策树的测试结果,如果变异之前进行重组操作势必会破坏原有的决策树结构,这时需要重新对训练集进行测试.因此,为了降低算法的计算复杂度,本文算法去除了 NNIA 中的重组操作.另外,为了增加抗体的多样性,在每次迭代过程中都加入了新的随机决策树抗体。

NNIA 算法中首先需要确定的是待优化目标,本文

将决策树的平均误分类代价和平均测试代价作为优化的两个目标,这是构建代价敏感决策树的基础.另外从图 1 中可以看出,决策树抗体初始化、决策树剪枝、决策树变异操作以及添加随机决策树是本文算法的四个关键步骤.决策树抗体初始化与添加随机决策树步骤都需要随机生成决策树,因而接下来将对决策树优化目标的确定、决策树抗体的随机生成、决策树的剪枝和变异等操作进行详细介绍.至于优势抗体群更新、活性抗体选择及比例克隆等操作与标准的 NNIA 相同,详细步骤请参考文献[9]。

2.1 优化目标的确定

本文采用的优化目标与文献[6]类似,都同时考虑了误分类代价和测试代价,即构建的决策树对误分类代价和测试代价敏感.不同的是文献[6]将误分类代价和测试代价用统一单位(美元)衡量,并利用平均总代价作为优化目标.而本文考虑到这两种代价是人们关心的两个不同的方面,不能在同一尺度下衡量,因而将平均误分类代价和平均测试代价作为两个优化目标,并利用多目标优化的方法构建决策树。

本文采用了分类代价矩阵的方法来计算平均误分类代价^[2,6].对于一个 K 分类问题,分类代价矩阵 C 是一个 $K \times K$ 阶矩阵,其中元素 $C_{i,j}$ ($0 \leq i, j < K$) 是将类别为 i 的样本预测为类别 j 时付出的代价.代价矩阵中元素的值由领域专家确定,但一般情况下主对角线上的元素值为 0,即如果预测正确则不会因误分类而增加代价.给定决策树 t 和训练样本集 D ,则平均误分类代价 $f_1(t)$ 可以用式(1)定义:

$$f_1(t) = \frac{1}{|D|} \sum_{d \in D} C_{I_d, h(t, d)} \quad (1)$$

其中, $|D|$ 为训练样本集 D 中所包含的样本数目, d 为训练样本集 D 中的样本, I_d 为样本 d 的实际类别, $h(t, d)$ 为决策树 t 对样本 d 的预测类别。

文献[4,5]在构建代价敏感决策树时虽然都考虑了测试代价,但是仅简单地为每个测试属性定义了固定代价,并没有考虑实际可能遇到的复杂情况^[12].在很多实际应用中每种属性的获取都需要付出一定的代价,例如医疗诊断中每项指标的获取需要花费一定的金钱、模式识别中每个特征的提取需要一定的时间等.并且,这些属性之间可能会共享若干操作,比如多项血液指标测量时仅需要一次抽血等.为此,本文在计算平均测试代价时考虑了这种比较常见的情况,如式(2)所示。

给定属性集合 A 和公有操作集合 P ,假设集合 A 中的每个属性 α 的获取需要先进行集合 P_α 中的所有操作,其中 P_α 是集合 P 的子集,则获取属性 α 所需代价 $T_A(\alpha)$ 的定义如式(2)所示:

$$T_A(\alpha) = T_\alpha U(\alpha) + \sum_{p \in P_\alpha} T_p U(p) \quad (2)$$

其中 T_α 是在所需预操作已完成的情况下获取属性 α 所需代价, T_p 是完成公有操作 p 所需代价, 这些代价都需要领域专家给出. $U(\alpha)$ 和 $U(p)$ 是指示函数, 若判决过程中还未获得属性 α 或者还未进行操作 p 则 $U(\alpha)$ 或 $U(p)$ 的值分别为 1, 否则值为 0.

对于给定的决策树 t , 其所对应的叶节点组成集合 L_t . 从决策树的根节点到每一个叶节点的路径组成一条判决规则, 即决策树中的每一个叶节点对应一条判决规则. 若规则 r 有 N 个内部节点, 每个内部节点对应的判决属性是 $\alpha_i (0 \leq i < N)$, 由根节点至叶节点顺序累加获取属性 α_i 所需代价即为规则 r 的测试代价 $T_R(r)$, 定义如式(3)所示:

$$T_R(r) = \sum_{i=0}^{N-1} T_A(\alpha_i) \quad (3)$$

其中 $T_A(\alpha_i)$ 表示获取属性 α_i 所需要的代价, 计算方法如式(2)所示.

决策树 t 的平均测试代价 $f_2(t)$ 的定义如式(4)所示:

$$f_2(t) = \sum_{l \in L_t} \frac{|D_l|}{|D|} T_R(r_l) \quad (4)$$

其中 D_l 是训练数据集 D 中满足规则 r_l 的决策属性条件的样本子集, $|D_l|$ 是集合 D_l 的数据样本数目, r_l 是叶节点 l 对应的判决规则, $T_R(r_l)$ 是规则 r_l 的测试代价.

确定了平均误分类代价 $f_1(t)$ 和平均测试代价 $f_2(t)$ 后, 本文提出的方法归结为利用 NNIA 算法解决式(5)中的两个目标的优化问题:

$$\begin{cases} \min & F(t) = (f_1(t), f_2(t))^T \\ t \in \Omega \end{cases} \quad (5)$$

其中 Ω 为整个决策树空间, t 为决策树空间中的变量.

2.2 决策树抗体的随机生成

NNIA 算法框架首先需要对决策树抗体进行编码, 本文直接用二叉决策树表示抗体^[13]. 由于二叉树和二叉树之间可以相互转换, 因而即使仅考虑二叉树, 搜索的决策树空间也没有因此减少. 本文随机建立的二叉决策树为满二叉树, 并且建立过程与传统的采用启发式策略自顶向下建立决策树的方法不同, 其内部节点

决策属性和分裂点都是随机选择的, 这有利于在整个决策空间搜索.

如果内部节点选择的决策属性为离散型, 设该属性的取值集合为 V_D , 分裂集合则为在集合 V_D 中随机选择的不为空的子集 V_D^* , 且 V_D^* 与 V_D 不相等. 若测试样本的该属性取值属于集合 V_D^* , 则转向决策树左分支, 否则转向右分支.

对于连续属性, 需要预先根据训练样本集 D 来设置候选分裂值集合 V_c , 分裂值为在集合 V_c 中随机选取的元素 v . 若测试样本的该属性值不大于 v , 则转向左分支, 否则转向右分支. 由于以集合 D 中类边界位置的值作为连续属性的分裂点能够取得局部最大信息增益^[13], 并且也能够最大限度地降低误分类代价, 因而本文将连续属性的类边界位置处的值设置为候选分裂值.

叶节点类别的指派方法为: 对于叶节点 l , 训练样本集 D 经决策树 t 测试后, 得到符合叶节点 l 对应规则的数据样本子集 D_l , 将使 D_l 误分类代价最小的类别指定为叶节点 l 的类别, 如式(6)所示.

$$l_c(x) = \arg \min_{d \in D_l} C_{l_d, x} \quad (6)$$

其中 x 为叶节点 l 应指派的类别, $C_{l_d, x}$ 含义与式(1)相同.

2.3 决策树剪枝策略

考虑到随机生成的决策树某些分支可能无效以及构建代价敏感的决策树的需要, 本文采用了两个剪枝策略: (1) 最小支持项数目; (2) 基于分类错误代价的剪枝方法.

最小支持项数目的剪枝策略与 C4.5 算法生长过程中控制树规模的方法类似, 由于本文并非按启发式规则生成决策树, 因而将此作为剪枝策略. 剪枝前需预先设定决策树的每分支最小支持项的数目 δ , 然后自顶向下考察每棵子树. 利用该剪枝方法可在获取决策树后剪除基本无效的子树, 能够降低树的规模.

具体剪枝方式为: (1) 若子树 t_b 的左右子树支持项的数目都小于 δ , 则剪除子树 t_b , 并在 t_b 的位置用叶节点代替, 如图 2(a) 所示; (2) 若子树 t_b 的左(右)子树支持项数目小于 δ , 而右(左)子树的支持项数目不小于 δ , 则将子树 t_b 的树根及左(右)子树剪除, 将右(左)子

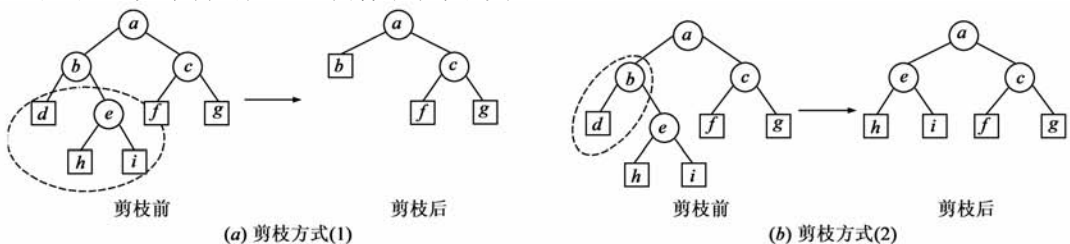


图2 基于最小支持项数目的决策树剪枝方式

树替代原子树 t_b 的位置,如图 2(b)所示。

基于分类错误代价的剪枝策略则是利用训练样本集 D 评估决策树的误分类代价,并自顶向下对子树是否被剪枝作出决定。对于内部节点 N ,假设以该节点作为根节点的子树为 t_n ,若剪除 t_n 后的决策树的平均误分类代价不大于剪枝前的平均误分类代价,则对子树 t_n 进行剪枝;否则保留该子树。误分类代价的计算方法与式(1)类似。

2.4 决策树抗体的变异操作

决策树抗体的结构并不固定,无法采用普通的变异操作进行变异。为此,本文采用了 5 种决策树变异操作^[13],分别是:(1) 利用随机创建的子树替代原决策树中随机选取的子树;(2) 对决策树进行随机剪枝;(3) 随机改变内部节点决策属性的分裂点;(4) 随机改变内部节点的决策属性和分裂点;(5) 随机分裂叶节点。

现有的决策树变异操作在选择决策树的变异点时并没有考虑决策树中各节点的性能,也没有考虑变异后对决策树性能可能会产生的影响,为此,本文提出了一种新的变异点选择策略。考虑到决策树自根节点至叶节点的决策过程,本文将每个叶节点对应的决策规则视为决策树抗体的基因,并根据决策规则的分类性能设置该基因被选择为变异点的概率。在选定变异基因后,进一步随机选择变异操作,并根据选定的变异操作选择需要变异的节点。这种选择策略会以较高的概率选择决策树中分类性能较差的分支进行变异,有利于决策树变异后性能的提高。

决策树 t 中决策规则 r 的分类性能利用该决策规则的误分类代价确定,计算方式如式(7)所示

$$Cost(t, r) = \frac{1}{|D|} \sum_{d \in D_r} C_{I_d, h(t, d)} + \epsilon \quad (7)$$

其中 $|D|$ 和 $C_{i,j}$ 的含义与式(1)相同, D_r 为满足规则 r 的 D 的样本子集。式(7)中 ϵ 为一较小的正数,用以确保 $Cost(t, r)$ 不为 0,从而可使决策规则 r 有被选择为变异基因的概率。

规则 r 被选择为变异基因的概率如式(8)所示:

$$p_r = \frac{Cost(t, r)}{Cost_{max}} \quad (8)$$

其中 $Cost_{max}$ 为决策树 t 中所有规则的最大误分类代价。

由式(7)和式(8)可以看出,决策树中分类性能最差也即误分类代价最大的规则,最有可能会被选择为变异的规则,这能够使决策树向降低误分类代价的目标变异。

3 实验结果及分析

为了验证所提出方法的有效性,本文在 UCI 公共测试数据集上进行了测试,并与经典的 C4.5 算法^[11]和

CSDB 算法^[8]做了对比研究。C4.5 算法是 Quinlan 于 1992 年提出的方法,已成为目前决策树算法研究的一个基准^[13],在对比实验中采用的是 C4.5R8 源码*。CSDB 是文献[8]中提出的一种新的代价敏感决策树算法,为了在不同尺度上衡量误分类代价和测试代价,该算法采用了如式(9)所示的属性分裂准则^[8]。

$$Performance(A_i) = \frac{(2^{GainRatio(A_i)} - 1)}{(TestCost(A_i) + w)} \times Redu_Mc(A_i) \quad (9)$$

在测试过程中,从 UCI 数据仓库中随机选择了 4 个用于分类的数据集(adult, car, breast-cancer, wine),这 4 个数据集的具体信息如表 1 所示。考虑到所有的属性只要付出一定的代价均可以获取,因此本文算法并没有考虑缺失值的情况,去除了数据集中的缺失项。实验中将各数据集的 2/3 作为训练数据,剩余的 1/3 作为测试数据。另外,由于 UCI 数据仓库并没有指定各数据集的分类代价矩阵和属性代价列表,本文对实验中选择测试的 4 个数据集的分类代价矩阵和属性代价列表分别进行了随机指定。

表 1 实验中选择的数据集信息

	adult	car	breast-cancer	wine
样本数目	45222	1728	683	178
属性数目	14	6	10	13
类别数目	2	4	2	3

NNIA 算法应用过程中需要指定初始抗体群规模、迭代次数等参数,实验中采用的控制参数如表 2 所示。

表 2 本文 NNIA 算法采用参数

参数	值
初始抗体群规模	50
优势抗体群规模	20
活性抗体群规模	10
克隆种群规模	100
变异率	0.3
迭代次数	1000

平均误分类代价和平均测试代价是本文重点考虑的两个优化目标。在对比实验中,将本文算法的指标与 C4.5 和 CSDB 算法的这两个指标进行了对比,对比结果如图 3 所示。由于本文算法得到的解是一组 Pareto 最优解集,因此图 3 给出的是 4 个不同数据集的 Pareto 前沿。

C4.5 和 CSDB 是利用不同的启发式规则自顶向下生成的决策树,仅能得到一组解,在图 3 中用点表示。从图 3 中可以看出,C4.5 和 CSDB 得到的解在 4 个测试数据集中均位于本文方法得到的 Pareto 前沿面的外侧,也就是说,采用本文方法得到的决策树能获得更优的平

* <http://www.rulequest.com/Personal/c4.5r8.tar.gz>.

均误分类代价和平均测试代价指标.另外,本文算法得到的并非单棵决策树,在应用中可根据具体情况选择合适的决策树,这也是本文算法的另一个优点.

决策树叶节点数目对应着结构风险最小化中的 VC 维,在一定程度上反映了决策树的泛化能力^[13].表 3 对

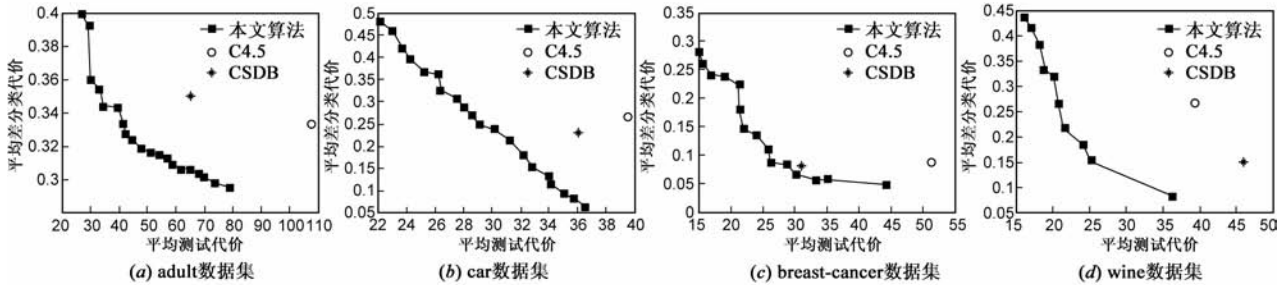


图3 本文算法与C4.5和CSDB算法的平均误分类代价及平均测试代价对比

表 3 叶节点数目对比列表

	Adult	car	breast-cancer	wine
本文算法	12.2	13.7	5.7	5.8
C4.5	736	154	31	9
CSDB	621	150	21	13

从图 3 和表 3 中可以得出如下结论:本文算法不仅能够获得在平均误分类代价和平均测试代价方面均优于 C4.5 和 CSDB 算法的决策树,而且获得的决策树具有更小的规模,更强泛化能力.分析其原因,主要包括:

(1) 本文算法将决策树的平均误分类代价和平均测试代价视为优化的两个目标,而 C4.5 算法并没有考虑误分类代价和测试代价问题,因而本文算法能够在这两方面取得好于 C4.5 算法的性能;

(2) CSDB 虽然同时考虑了误分类代价和测试代价,但作为基于启发式规则的决策树构建方法,在选择分裂属性时仅考查了一次分裂之后的情况,而本文算法是对整个决策树进行评价的,因而能够取得更优的性能;

(3) 由于本文算法采用了最小支持项数目和基于分类错误代价的剪枝方法,并且决策树抗体采用了随机生成和变异策略,这使得生成的决策树能够保持较小的规模,而 C4.5 和 CSDB 都是基于某种启发式规则自顶向下构建决策树,仅采用最小支持项数目控制树的规模,这使得在训练样本增大时树的规模增大迅速.

4 结束语

本文提出了一种基于 NNIA 多目标优化的代价敏感决策树构建方法.首先分析了决策树的平均误分类代价和平均测试代价,并将这两种代价作为 NNIA 优化的两个目标.然后针对决策树抗体,研究了 NNIA 进化过程中随机决策树的构建、决策树剪枝、决策树变异等操作,最终获取了一组对误分类代价和测试代价敏感的 Pareto 最优决策树.本文方法获得的决策树在平均误

比了上述三种算法的叶节点数目,其中本文算法的叶节点数目是所得到的 Pareto 最优决策树叶节点的均值.从表 3 中可以看出,本文方法构建的决策树的叶节点数目远小于 C4.5 和 CSDB 中的算法,尤其是训练数据样本量较大时,例如 adult 数据集和 car 数据集.

分类代价和平均测试代价方面均明显好于 C4.5 和 CSDB 算法,并且决策树规模较小,泛化能力更强.另外,相对 C4.5 和 CSDB 算法,本文得到的是一组决策树,便于决策者根据具体情况选择合适的决策树.

参考文献

- [1] 李光,王亚东,苏小红.隐私保持的决策树分类挖掘[J].电子学报,2010.38[1]:204-212.
- [2] P D Turney. Types of cost in inductive concept learning[A]. Proc of the Workshop on Cost-Sensitive Learning at the 17th International Conference on Machine Learning[C]. Stanford University, Stanford, California, USA, 2000
- [3] 马光志.多代价敏感的反神经网络优化研究[D].武汉华中科技大学博士论文,2009.
- [4] C X Ling, V S Sheng, Q Yang. Test strategies for cost-sensitive decision trees[J]. IEEE Transactions on Knowledge and Data Engineering, 2006. 18[8]: 1055-1067.

(下转第 2396 页)

作者简介



赵士伟 男,1980 年出生,山东高唐人.博士研究生.主要研究方向为网络多媒体信息处理.

E-mail: zhao_shiwei@emails.bjut.edu.cn

卓力 女,1971 年出生,江苏徐州人.博士,北京工业大学教授.主要研究方向为图像/视频编码与传输,网络多媒体分析,无线视频传感器网络.

王素玉 女,1976 年出生,河北丰润人.讲师,在站博士后.主要研究方向为图像/视频信号压缩编码和超分辨率复原,智能视觉监控等.

沈兰荪 男,1938 年出生,江苏苏州人.教授,博士生导师.主要研究方向为图像/视频信号处理、传输、压缩与应用.

- Place Cells for Small Mobile Robots[A]. International Conference on Artificial Neural Networks (ICANN) [C]. Springer, Berlin, 2009. 747 – 756.
- [61] Machine Learning Group at Ghent University, Reservoir Computing Toolbox [DB/OL]. <http://snn.elis.ugent.be/node/59>, 2007-10-09.
- [62] Jaeger H. A Simple Toolbox of Echo State Networks [DB/OL]. http://www.faculty.jacobs-university.de/hjaeger/esn_research.html, 2007-06-10.
- [63] Georg Holzmann. Efficient C++ library for analog reservoir computing neural networks (Echo State Networks) [DB/OL]. <http://aureservoir.sourceforge.net/>, 2007-10-07.
- [64] Benjamin Schrauwen, Michel D'Haene, David Verstraeten, Jan Van Campenhout. Compact hardware liquid state machine on FPGA for real-time speech recognition [J]. Neural Networks, 2008, 21(2-3): 511 – 523.
- [65] H Jaeger. Reservoir riddles: Suggestions for echo state network research [A]. International Joint Conference on Neural Networks, IJCNN 2005 [C]. Piscataway, NJ 08855-1331, 2005. 1460 – 1462.
- [66] J Hertzberg, H Jaeger. Learning to ground fact symbols in behavior-based robots [A]. Proceedings of the 15th European Conference on Artificial Intelligence [C]. Amsterdam, 1000 AE, Netherlands, 2002. 708 – 712.
- [67] W Maass, T Natschlag, H Markram. A model for real-time computation in generic neural microcircuits [A]. Advances in Neural Information Processing Systems (NIPS) [C]. MIT Press, 2003. 15. 229 – 236.
- [68] Michal Cernansky. Novel recurrent connectionist approaches – echo state networks [J]. ICIC Express Letters, 2008, 2(3): 219 – 224.
- [69] Georg Holzmann. Reservoir computing: A powerful black-box framework for nonlinear audio processing [A]. Proceedings of the 12th International Conference on Digital Audio Effects [C]. Kluwer Academic Publishers, 2009. 1 – 8.
- [70] David Reid, Mark Barrett-Baxendale. Glial reservoir computing, second UKSIM european symposium on computer modeling and simulation [A]. Computer Modeling and Simulation, 2008, EMS '08 [C]. Piscataway, NJ 08855-1331, 2008. 81 – 86.

作者简介



彭宇男, 教授, 博士生导师, 1973年6月生于陕西西安, 哈尔滨工业大学自动化测试与控制研究所副所长, 主要研究方向为测试诊断技术、无线传感器网络技术和数据挖掘技术等。

E-mail: pengyu@hit.edu.cn



王建民男, 1976年生于吉林长春, 现为哈尔滨工业大学自动化测试与控制系博士研究生, 主要研究方向为时间序列分析、数据挖掘、故障诊断。

E-mail: wjmfuzzy@126.com

(上接第 2352 页)

- [5] Q Yang, C Ling, X Chai, et al. Test-cost sensitive classification on data with missing values [J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18[5]: 626 – 638
- [6] P D Turney. Cost-sensitive classification: empirical evaluation of a hybrid genetic decision tree induction algorithm [J]. Journal of Artificial Intelligence Research, 1995, 2: 369 – 409
- [7] J V Davis, J Ha, C J Rossbach, et al. Cost-sensitive decision tree learning for forensic classification [A]. Proc of 17th European Conference on Machine Learning [A]. Berlin GERMANY, SEP 18 – 22, 2006: 622 – 629
- [8] X Liu. Cost-sensitive decision tree with missing values and multiple cost scales [A]. Proc of the first IITA International Joint Conference on Artificial Intelligence [C]. Hainan, CHINA, APR 25-MAY 26, 2009: 294 – 297
- [9] M Gong, L Jiao, H Du, et al. Multiobjective immune algorithm with nondominated neighbor-based selection. Evolutionary Computation [J]. 2008, 16[2]: 225 – 255
- [10] A Asuncion, D J Newman, UCI machine learning repository [http://www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, CA: University of California, School of Information and Computer Science, 2007.
- [11] R Quinlan. C4. 5: programs for machine learning. Morgan Kaufmann Publishers, 1992.
- [12] F Min, Q Liu. A hierarchical model for test-cost-sensitive decision systems [J]. Information Sciences, 2007, 179: 2442 – 2452.
- [13] D Kim. Structural risk minimization on decision trees using an evolutionary multiobjective optimization [A]. Proc of the 7th European Conference on Genetic Programming [C]. University Coimbra, Coimbra, PORTUGAL, APR 05 – 07, 2004. 338 – 348.