基于代价敏感直推式学习的故障诊断方法

吴 薇[1,2], 胡静涛[1]
(1 中国科学院沈阳自动化研究所 工业信息重点实验室 沈阳 110016; 2 中国科学院研究生院 北京 100039)

摘 要: 针对故障诊断领域存在的不考虑误诊断代价以及提出泛化能力强的诊断规则难等问题,提出了一种代价敏感直推式学习故障诊断方法。基于 Kolmogorov算法随机性理论和代价敏感学习最小期望误分类代价准则提出了代价敏感直推式分类机制,并在此基础上设计了用于故障诊断的 CsTCM-kNN 算法。通过旋转机械轴系故障代价敏感诊断实验,验证了该方法能够有效地降低误诊断代价,且保证较高的诊断准确率。

关键词: 故障诊断; 代价敏感; 直推式学习; 算法随机性理论
中图分类号: TP18, TP206 文献标识码: A 国家标准学科分类代码: 520. 2099

Fault diagnosis based on cost-sensitive transduction inference

Wu Wei[1,2], Hu Jingtao[1]
(1 Key Laboratory of Industrial Informatics, Shenyang Inst of Automation, Chinese Academy of Sciences, Shenyang 110016, China; 2 Graduate School of the Chinese Academy of Sciences, Beijing 100039, China)

Abstract Most researches on fault diagnosis pursue to minimize the error rate without considering the misclassification cost, and are also difficult to propose enough diagnosis rules with good generalization ability. In this paper, a new fault diagnosis method based on cost-sensitive transduction inference is presented, which solves the two problems mentioned above. A cost-sensitive transduction classification machine is proposed based on the Kolmogorov algorithm randomness theory and minimum expected misclassification cost principle, and then a CsTCM-kNN algorithm for fault diagnosis is designed. Experiment results of cost-sensitive fault diagnosis of a rotating machine show that the proposed method can reduce misclassification cost effectively with high diagnosis accuracy.

Key words fault diagnosis; cost-sensitive; transduction inference; algorithm randomness theory

1 引 言

故障诊断是实现有效降低设备全寿命周期成本、增加安全性和稳定性的科学设备维护的关键。在实际的工业生产过程中由于危害程度的不同,设备不同故障状态间的误诊断代价是不对等的,将危害程度大的状态误诊断为危害程度小的状态所需要承担的人身安全和经济损失等代价往往大于相反的情况。另外,由于故障样本的获取是以设备某种程度的损坏为代价的,所以相对于正常样本,故障样本的数量会少得多,这种样本不均衡性会导致传统以分类准确率为性能指标的故障诊断方法的结论更倾向于正常状态的判定,不能有效地避免故障带来的损失。因此,将最小诊断代价作为故障诊断方法的本质目标比最大诊断准确率更为合理。

由于生产流程的复杂化和机械设备的精密化、大型化和自动化,不同设备或设备不同部分间相互联系,耦合非常紧密,使得发生的故障具有多层次性、随机性和不确定性等特点。这样,很难设计某种故障诊断分类器,提取良好泛化能力的诊断规则,使其对未来所有可能样本的预期性能最优。

针对以上目前故障诊断领域存在的问题,需要研究一种可以不局限于某个故障类别闭集上实现的增量学习代价敏感故障诊断方法。

收稿日期: 2008-10 Received Date: 2008-10

代价敏感学习是近年来机器学习和数据挖掘领域的一个新的研究热点,已经在信用卡欺诈识别、网络入侵检测及客户流失预测等领域得到了成功应用[1]。目前的研究思路主要有两种:一是直接改变现有分类算法使其实现代价敏感,即直接法,如 ICET 法[2]、代价敏感决策树[3-4]等;二是不需要改变现有分类算法,只需对算法的输入或输出设计有效的代价敏感转换过程,即元学习法,如 MetaCost法、Costing法和 weighting法等[1]。

统计学习理论领域的直推式学习,不同于经典的归纳学习,不需要建立在整个示例分布上具有低错误率的决策函数,而只是期望在给定的待测示例上达到最好的性能。在学习过程中可以显式地考虑待测示例,从而更能满足故障诊断实际应用的需要[5]。直推式学习在文本识别[6]、异常检验[7]等领域中都已有了不同程度的研究和应用。

本文将代价敏感学习和直推式学习的基本思想和理论相结合,基于 Kolmogorov算法随机性理论和最小期望误分类代价原则提出了代价敏感直推式分类机制(CsTCM, Cost-sensitive Transductive Classification Machine),并设计了代价敏感直推式 k 近邻算法(CsTCM-kNN)。通过旋转机械轴系故障诊断的应用实例,验证了方法的有效性。

2 序列随机性检验

根据 Kolmogorov算法随机性理论,满足独立同分布条件的样本序列 $z=(z_1, \cdots, z_n)$ 的随机性(可以理解为统计规律性),可以由随机性检验函数来判定。

定义 1 随机性检验函数定义为满足如下条件的函数 $t: (9)^n \to [0, \infty)$:

1)对于所有 $r \geqslant 0$, $n \in N$ 及 $P \in P^*$,

$$P^n\{z \in (9)^n: t(z) \leqslant r\} \leqslant r \tag{1}$$

2) t是自上半可计算的

式中:(9)为样本空间,$z=(z_1, \cdots, z_i, \cdots, z_n)$ 为(9)中样本构成的样本序列,n 为样本序列的长度,z_i ($i=1, 2, \cdots, n$),均来自某个可计算的分布为 P 的随机模型。P^* 为(9)中所有可能存在的概率分布的集合[8]。

定义中的条件 1)保证了随机性检验函数 $t(z)$ 的有效性,即根据小概率推断原理,只要样本序列 z 满足独立同分布的弱前提,函数 $t(z)$ 值的大小,就可以反映样本序列 z 的随机性水平。$t(z)$ 值越大,说明样本序列的随机性越明显,反之则越模糊。

3 代价敏感直推式分类学习

3.1 代价敏感直推式分类机制

故障诊断实质上是分类识别问题。对于一个分类问题,存在给定训练样本集合 $\{(z_1, y_{z1}), \cdots, (z_n, y_{zn})\}$,$z_i=(z_{i1}, z_{i2}, \cdots z_{im})$ 为第 i个样本的特征向量,m 为特征数,$y_{zi} \in Y=\{y_1, y_2, \cdots, y_c\}$ 为第 i个样本的类别标记,n 为样本数量,c为类别数。对于独立同分布序列,在常数范围内存在一个最小的随机性检验函数(通用随机性检验函数) $t_{univ}(z)$,即 $t_{univ}(z) \leqslant t(z)+C$, $t(z)$ 为其他任意随机性检验函数,C 为常数。Kolmogorov、Martin-Lof和 Levin都分别证明了该结论[8]。利用 $t_{univ}(z)$ 的含义,可以按照如下假设检验的步骤对待测样本 z_{n+1} 属于不同类别的概率进行估计:

首先,分别建立 z_{n+1} 属于所有可能类别 $y_{z_{n+1}} \in Y=\{y_1, y_2, \cdots, y_c\}$ 的假设;

其次,对于每个假设,将 z_{n+1} 加入该假设类别 $y_{z_{n+1}}$ 的训练样本集合生成新的序列 $z'=(z_1, \cdots, z_n, z_{n+1})$;

利用 $t_{univ}(z)$ 对新序列 z'的随机性水平进行检验;

最后,根据各假设下的序列随机性水平来衡量待测样本属于相应假设类别的可能性大小。

这是一种直推式学习方法,它只建立在样本独立同分布的弱前提下,不需要知道样本分布的具体类型和参数,是根据假设类别情况下置信度间的相对大小来对待测样本分类概率进行评判的,且不一定要在模式类别的闭集上进行。但是,Martin-Lof证明 $t_{univ}(z)$ 是不可计算的[8],因此,在实际应用中需要利用符合定义 1中条件 1)的非通用随机性检测值对其进行近似。

如果存在奇异检验函数 $\alpha(z_i)$ 能够将样本 z_i映射为实数,使得样本序列 $z=(z_1, \cdots, z_n)$ 能够映射到实数空间,即函数 $\alpha: {}^n \to R^n$,且与样本顺序无关,则可以给出引入待测样本 z_{n+1} 后的新样本序列的随机性检验函数近似估计的定义[9]。

定义 2 对于待测样本 z_{n+1} 相对于假设类别 y 的 p 值定义为:

$$p(\alpha(z_{n+1})) = \frac{\#\{i: \alpha(z_i) \geqslant \alpha(z_{n+1})\}}{n+1} \tag{2}$$

式中:# 表示集合的基数,z_i为 y类别的训练样本,$\alpha(z_i)$ 为奇异检验函数。

文献[9]给出了 p 值函数符合定义 1中条件:1)的定理证明。统计学理论中,p 值为数据所提供证据的强度。直观地,p 值可以理解为待测样本属于假设类别 y 的概率,$1-p$ 为判定假设成立的风险。所以,p 值越大,表示 z_{n+1} 属于 y 类的可能性越大。

在分类的过程中,不仅要考虑分类的准确性,还要考虑由于错误分类而产生的代价,代价的最小化才是实际分类应用的本质目标。代价敏感分类问题中,类别之间的误分类代价一般用形如表 1的代价矩阵表示,矩阵元素 $c(y_i, y_j)$ 表示将 y_i 类别的样本误分类为 y_j类所产生的代价。代价矩阵可以由领域专家预先给出,或者利用其

他方法学习得到的, 在代价敏感分类方法研究过程中, 通常假定已知[1]。给定代价矩阵, 根据最小期望误诊断代价原则, 将待测样本 z_{n+1} 分为 y类的判定函数为:

$$y = \arg\min_{y_i \in Y} \sum_{j=1}^{c} p(y_j \mid z_{n+1}) c(y_i, y_j) \quad (3)$$

这里, $p(y_j \mid z_{n+1})$ 为待测样本 z_{n+1} 属于 y_j 的后验概率, $c(y_i, y_j)$ 是代价矩阵的元素。

一般情况下, 直接获取样本的后验概率需要预先知道样本的分布类型和分布参数, 但是表征工业设备运行状态的这种复杂特征数据, 它的分布情况是很难预先得到。根据序列随机性检验 p 值的涵义, 待测样本相对于某待测类样本空间的 p 值越大, 则表明待测样本属于该类样本空间的可能性越大, 所以, 以 p 值函数的输出作为样本的后验概率映射, 可以将代价敏感因子引入到前述的基于随机性检验函数的直推式学习机制中, 进而利用下式实现待测样本 z_{n+1} 属于 y类的判定:

$$y = \arg\min_{y_i \in Y} \sum_{j=1}^{c} p_j(\alpha(z_{n+1})) c(y_i, y_j) \quad (4)$$

这里, $p_j(\alpha(z_{n+1}))$ 表示待测样本 z_{n+1} 相对于假设类别 y_j 的 p 值。

3.2　CsTCM-kNN算法

按照上一节提出的代价敏感直推式分类机制设计具体分类算法, 首先要确定奇异检验函数的形式。奇异检验函数是要根据具体的分类算法而选定的。复杂工业过程设备运行状态存在层次性、随机性和不确定性, 状态特征属性值的分布情况是很难获得的先验知识, 而且能够实现增量学习的方法更有利于提高诊断系统的准确率和自学习能力, 由此, 本文选取最近邻方法作为核心分类算法, 相应地选取了距离奇异检验函数。

定义 3　给定 y类样本 z_i, 距离奇异检验函数 $\alpha(z_i)$ 定义为:

$$\alpha(z_i) = \frac{\sum_{j=1}^{k} D_{ij}^{y}}{\sum_{j=1}^{k} D_{ij}^{-y}} \quad (5)$$

式中: k为近邻数, D_{ij}^{y}表示 z_i 与其在类别 y样本中的第 j个近邻间的距离, D_{ij}^{-y}则表示 z_i与除类别 y样本外所有样本组成的集合中第 j个近邻间的距离。

同类样本由于具有相似性, 样本间距离通常会相对小于非同类样本间的距离。因此, 当待测样本 z_i 越接近同类样本或越远离非同类样本时, $\alpha(z_i)$ 的值会越小, 反之则越大。奇异检验函数值反映了待测样本与待测类别间的亲和度。

基于代价敏感直推式分类机制和距离奇异检测函数, 本文设计了代价敏感直推式 k近邻(CsTCM-kNN)算法。下面给出算法的形式化描述。

算法 1 CsTCM-kNN算法

输入: 待测样本 z_{new}; 类别数 c; 训练样本集 $Dataset$; 近邻数 k; 代价矩阵 $\boldsymbol{CM}$, 元素为 $c(y_i, y_j)$

输出: 类别标号 class_label

```
Begin
  For each z_i in Dataset {
      D_zi^{y_i} = traditionnal_kNN(y_i, z_i);
      D_zi^{-y_i} = traditionnal_kNN(-y_i, z_i);
    compute_save(α(z_i));
    compute_save(d(z_i, z_new));
  }
  For j = 1 to c {
      append(z_new, Dataset_{y_j});
      For each z_i in Dataset_{y_j}
      If d(z_i, z_new) < D_zik^{y_j}
        compute_save(α(z_i));
      For each z_i in Dataset_{-y_j}
      If d(z_i, z_new) < D_zik^{-y_j}
        compute_save(α(z_i));
      compute_save(α_j(z_new));
      compute_save(p_j(α(z_new)));
  }
  class_label = arg min_{y_i∈Y} Σ_{j=1}^{c} p_j(α(z_new)) c(y_i, y_j);
  End
```

算法首先利用传统 k近邻算法, 分别求得训练样本集中所有样本与同类样本和非同类样本的最短的 k个距离, 并升序排列, 得到序列 $D_{zi}^{y_i}$和 $D_{zi}^{-y_i}$, 进而得到每个样本的奇异检验函数值; 然后, 穷尽假设待测样本属于每个待测类别, 在每个原假设情况下, 计算新生成的样本序列相应待测样本的奇异检验函数值和 p值, 其中 $Dataset_{y_j}$ 表示 y_j 类的样本集, $Dataset_{-y_j}$ 表示非 y_j 类样本组成的样本集; 最后以每种分类假设下待测样本的 p值和代价矩阵, 以最小期望总误分类代价为目标判定待测样本的最终分类标识。

其中, 第一部分关于训练样本的 k近邻距离和奇异检验函数值的计算, 可以通过离线方式预先计算, 在每次分类时, 只需要在每个分类假设下, 新的样本序列的随机性有相应变化的时候, 才进行个别调整。另外, 样本的距离计算, 可以根据实际样本特征向量的特点, 选取适合的距离, 如 Euclidean距离、动态弯曲距离、编辑距离及最大公共子序列等。

4　旋转机械轴系故障振动诊断应用实例

为了验证本文提出的基于代价敏感直推式学习的故

障诊断方法的有效性和实用性,以旋转机械轴系振动故障为例进行了代价敏感故障诊断。

4.1 数据准备

实验采用 ZHS-2型双跨多功能转子实验台模拟旋转机械的运行状态。该实验台采用额定电流 2 A、输出功率 300 W 的永磁式直流伺服电动机经联轴器直接驱动转子。通过调速器可以手动实现 0~10 000 r/min范围的无级调速。实验中,模拟旋转机械四种运行状态的升速过程:无故障(y_1)、转子不平衡(y_2)、转子不对中(y_3)、动静摩擦(y_4)。转子转速由 300 r/min上升到 2 400 r/min,每增加约 70 r/min,水平 X 和垂直 Y 方向的电涡流传感器进行一次采样,共 30次。每种运行状态每个特定转速采集 100组振动位移数据,每组包含 X 和 Y 两个方向各 1 024个采样点,采样频率为 1 024 Hz。选取的特征频率点包括:1x工频;0.2x、0.4x、0.6x、0.8x四个分倍频;2x~8x七个高倍频。

4.2 特征提取

设备运行状态特征提取的精炼与准确,是提高故障诊断质量和效率的关键前提。旋转机械的大多数常见故障,都有与之对应明显不同的全息谱特征[10]。这些特征主要集中于不同阶次(频率)椭圆(直线和正圆看作椭圆的特殊形式)的大小和旋转角度等形状特征以及初相点方位和进动方向等轨迹运动特征。实验借鉴全息诊断融合分析设备振动全貌的思想,利用一种按转速和频率顺序排列椭圆特征向量生成的类多维时间序列,对全息诊断用于判定设备状态的知识进一步扩展和提取表示。

假设椭圆特征向量为 $v_f = (da, db, \theta, \phi)$,其中,$a$ 为椭圆的长半轴,b为短半轴,θ为长轴倾角,ϕ为初相点的方位角,d为符号变量,当椭圆为正进动时为 1,反之为 -1,各向量元素值可以利用特定转速下转子同支承截面相互垂直的 X, Y 两个方向振动位移信号同借此频率简谐波合成的轨迹方程系数计算得到;$V = \{v_{f1}, v_{f2}, \cdots, v_{fi}, \cdots, v_{fi}\}$ 为依据频率大小将旋转机械转子同一支承截面同一转速下的全息谱中椭圆的特征向量在频率轴上顺序排列,抽象得到的序列,f_i为频率标识,k为观测阶次频率数,旋转机械运行状态特征由序列 $P = \{p(t') \mid i = 0, 1, 2, \cdots, n\}$ 表示,其中序列元素 $p(t'_i) = V_{t'_i} = \{v_{t'_1 f_1}, v_{t'_1 f_2}, \cdots, v_{t'_1 f_k}\}$,$n$为序列 P 的长度。

给定长度分别为 m 和 n、对应元素频率标识相同的两个全息序列 P 和 Q 的距离为:

$$d(P, Q) = 1 - Similarity(P, Q) \tag{6}$$

其中,

$$Similarity(P, Q) = \begin{cases} simE(P, Q), & \text{若 } m = n = 1; \\ LCSS_\delta(P, Q)/\min(m, n), & \text{其他} \end{cases} \tag{7}$$

$$LCSS_\delta(P, Q) = \begin{cases} 0\text{若 } P \text{ 或 } Q \text{ 长度为 } 0; \\ 1 + LCSS_\delta(Head(P), Head(Q)), \\ \text{若 } simE(p(t'_m), q(t'_n)) > \delta; \\ \max[LCSS_\delta(Head(P), Q), LCSS_\delta[P, Head(Q)]], \\ \text{其他} \end{cases} \tag{8}$$

$$simE(p(t'_1), p(t'_2)) = \frac{1}{k}\sum_{i=1}^{k}\sigma(v_{hfi}, v_{tfi}) \tag{9}$$

$$\sigma(v_{f1}, v_{f2}) = \begin{cases} 1\text{若 } \Delta v_f \text{ 各元素小于 } v_\varepsilon \text{ 中对应元素} \\ 0\text{其他} \end{cases} \tag{10}$$

$\Delta v_f = v_{f1} - v_{f2}$, $v_\varepsilon = (a_\varepsilon, b_\varepsilon, \theta_\varepsilon, \phi_\varepsilon)$ 为椭圆特征向量元素误差阈值组成的阈值向量。

利用上述方法,将采集到的振动数据转化为全息序列,得到长为 30的全息序列共 400组,每个转子运行状态包含 100组。

4.3 代价敏感直推式故障分类

为了评价 CsTCM-kNN的性能,实验选取文献[11]的 TCM-kNN法和传统 kNN方法在相同全息序列数据集中进行准确率和总误诊断代价的相关结果对比。

另外,在实际的工业生产过程中,旋转机械故障所带来的代价涉及人力、物力等很多因素,需要领域专家实地评估和统计才能得到。在本文实验的过程中,直接引用了文献[12]中给出的对应故障的误分类代价数据,如表1所示。

表 1 代价矩阵

Table 1 Cost matrix

真实类	预测类			
	y_1	Y_2	y_3	y_4
y_1	0	1	1	1
y_2	4	0	2	2
y_3	5	3	0	3
y_4	10	7	7	0

图 1和图 2分别显示了近邻数 k分别选取 1到 8,阈值向量 $v_\varepsilon = [0.1, 0.1, 30, 30]$, $\delta = 5$,三种方法采用十折交叉验证(ten fold cross-validation)得到的准确率和总误诊断代价的情况。

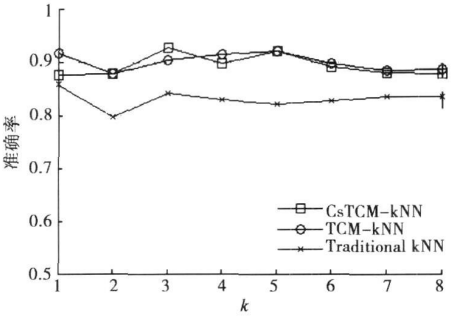


图 1 诊断准确率比较

Fig. 1 Diagnostic accuracy comparison

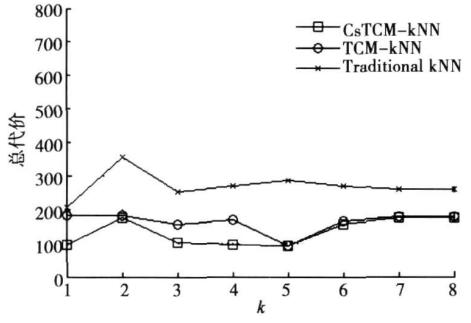


图 2　误诊断总代价比较

Fig. 2　Total misclassification cost comparison

从图 1的实验结果可以看出, 对于不同 k值, 传统 kNN算法的诊断准确率都明显低于 CsTCM-kNN算法和 TCM-kNN算法, 而且后两种算法的诊断准确率基本相当, 都能满足故障诊断实际需要。图 2给出的总误诊断代价实验结果显示, 在 k值在 1~4的范围内, CsTCM-kNN算法的误诊断代价明显少于 TCM-kNN算法, 而传统 kNN算法的总代价明显高于前两者。这是由于 CsTCM-kNN算法借鉴了基于 Kolmogorov算法随机性理论的置信度分类机制, 在传统 kNN算法的基础上有效提高了分类准确率, 并且同时考虑了误诊断的代价, 将分类边界向靠近误诊断代价小的类别样本空间偏移。

表 2和表 3分别给出了 $k=3$时, 每个设备状态具体的准确率和误诊断代价的统计结果。从表中可以看出, CsTCM-kNN与 TCM-kNN算法的准确率都高于传统 kNN算法, 而误诊断代价明显低于传统 kNN算法。具体地, CsTCM-kNN算法与 TCM-kNN算法相比, 在动静摩擦和不对中两个状态数据上的诊断准确率有所提高, 而其他两类状态稍有降低。可见, CsTCM-kNN算法在实现故障诊断的过程中, 虽然损失了少许低代价故障类别的诊断准确率, 但对于误诊断代价相对高的实例分类性能有了明显提升, 且有效地降低了误诊断总代价。

表 2　$k=3$时, 不同设备状态的准确率

Table 2　Accuracy for different machine conditions when $k=3$

	y_1	y_2	y_3	y_4
CsTCM-kNN	0.91	0.92	0.93	0.95
TCM-kNN	0.92	0.91	0.89	0.90
Traditional kNN	0.87	0.86	0.83	0.81

表 3　$k=3$时, 不同设备状态的误诊断代价

Table 3　Misclassification costs for different machine conditions when $k=3$

	y_1	y_2	y_3	y_4
CsTCM-kNN	9	22	31	41
TCM-kNN	8	26	43	79
Traditional kNN	13	38	55	148

5　结　论

本文提出了一种代价敏感直推式学习故障诊断方法。该方法以结合代价敏感学习和直推式学习基本理论和思想的 CsTCM-kNN算法来实现故障识别分类。不同于传统故障诊断方法, 它的分类判定准则是最小误诊断代价而不是最大诊断准确率, 而且该算法基于样本独立同分布的弱假设前提, 不需要预先知道样本数据的分布类型和相关参数, 实现简单且具有增量学习的能力。通过旋转机械轴系故障的代价敏感直推式学习故障诊断实验表明, 本文提出的方法明显地降低了误诊断总代价, 且保证较高的诊断准确率。然而, 随着样本数据的不断积累, 数据集的规模会不断增长, 该方法的实现效率会受到制约, 所以, 如何对训练样本集进行约减以及如何在训练样本集中建立高效地搜索策略, 将作为本文的下一步研究内容。

参考文献

[1] 凌晓峰, SHENG V S. 代价敏感分类器比较研究 [J]. 计算机学报. 2007, 30(8): 1203-1211.
LING X F, SHENG V S. A comparative study of cost-sensitive classifiers [J]. Chinese Journal of Computers, 2007, 30(8): 1203-1211.

[2] TURNEY P D. Cost-sensitive classification - empirical evaluation of a hybrid genetic decision tree induction algorithm [J]. Journal of Artificial Intelligence Research (1076-9757). 1995: 369-409.

[3] DRUMMOND C, HOLTE R C. C4.5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling[C]. Proceedings of the International Conference on Machine Learning (ICML 2003) Workshop on Learning from Imbalanced Data Sets II. Washington, DC, USA, 2003.

[4] LING C X, SHENG V S, YANG Q. Test strategies for cost-sensitive decision trees[C]. IEEE Transactions on Knowledge and Data Engineering. IEEE Educational Activities Department, Piscataway, NJ, USA, 2006, 18(8): 1055-1067.

[5] 周志华, 王珏. 机器学习及其应用 2007 [M]. 北京: 清华大学出版社, 2007.
ZHOU Z H, WANG Y. Machine learning and its applications 2007 [M]. Beijing: Tsinghua University Press, 2007.

[6] 陈毅松, 汪国平, 董士海. 基于支持向量机的渐进直推式分类学习算法. [J]. 软件学报. 2003, 14(3): 451-460.

CHEN Y S, WANG G P, DONG SH H. A progressive transductive inference algorithm based on support vector machine [J]. Journal of Software, 2003, 14(3): 451-460.

[7] BARBAR D, DOMENICONI C, ROGERS J P. Detecting outliers using transduction and statistical testing[C]. Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. Philadelphia, PA, USA: 2006: 55-64.

[8] GAMMERMAN A, VOVK V. Prediction algorithms and confidence measures based on algorithmic randomness theory[J]. Theoretical Computer Science (0304-3975). 2002: 209-217.

[9] 邱德红, 陈传波, 金先级. 基于算法随即理论和奇异描述的置信学习机器[J]. 计算机研究与发展. 2004, 41(9): 1586-1592.
QIU D H, CHEN CH B, JIN X J. Confidence learning machine based on algorithmic theory of randomness and dissimilarity description [J]. Journal of Computer Research and Development. 2004, 41(9): 1586-1592.

[10] 屈梁生. 机械故障的全息诊断原理[M]. 北京: 科学出版社, 2007.
QU L SH. Holospectrum and holobalancing technique in machinery diagnosis [M]. Beijing: Science Press, 2007.

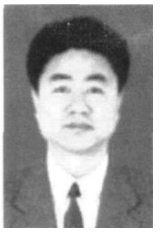
[11] PROEDROU K, NOURETDINOV I, VOVK V, et al. Transductive confidence machines for pattern recognition [C]. Proceedings of the 13th European Conference on Machine Learning. Heidelberg: Springer-Verlag, 2002: 381-390.

[12] 刘金福, 于达仁, 胡清华, 等. 基于加权粗糙集的代价敏感故障诊断方法[J]. 中国电机工程学报. 2007, 27(23): 93-99.
LIU J F, YU D R, HU Q H, et al. Cost-sensitive fault diagnosis based on weighted rough sets [J]. Proceedings of the Chinese Society for Electrical Engineering. 2007, 27(23): 93-99.

作者简介

吴薇, 2003年于辽宁大学获学士学位, 现为中国科学院沈阳自动化研究所博士研究生, 主要研究方向为旋转机械故障诊断与预测, 时间序列挖掘等。
E-mail: wuwe@sia.cn

Wu Wei received B. Sc. from Liaoning University in 2003. She is now a Ph. D candidate in Shenyang Institute of Automation, Chinese Academy of Science. Her research interests are rotating machinery fault diagnosis and prognosis, and time series data mining.



胡静涛, 分别于1985年和1988年于大连理工大学获得学士学位和硕士学位, 现为中国科学院沈阳自动化研究所研究员, 中国科学院研究生院教授, 博士生导师, 主要研究方向为远程设备监测与故障诊断。
E-mail: hujingtao@sia.cn

Hu Jingtao received B. Sc. from Dalian Institute of Technology in 1985 and M. Sc. from Dalian University Science and Technology in 1988. He is now a professor both in Shenyang Institute of Automation and Graduate School of Chinese Academy of Science. His main research interest covers the areas of remote monitoring and fault diagnostics of equipment.