

滚动轴承故障诊断的案例推理方法

付新哲, 张优云, 朱永生

(西安交通大学机械工程学院, 710049, 西安)

摘要: 针对滚动轴承的故障诊断问题,提出了一种采用案例推理(CBR)的诊断方法.为了解决检索相似案例时案例属性多、人工确定关键属性及其权重困难的问题,提出了一种 Filter/Wrapper 复合型特征选择算法,用邻域粗糙集算法粗选属性,用遗传算法进一步精选属性和优化权重,并有效地解决了邻域粗糙集算法中需要人工确定邻域大小的问题.以滚动轴承运行时的振动信号为基本信息,建立了滚动轴承案例库,从案例库中检索与问题案例相似的历史案例,并根据这些历史案例来判定问题案例的故障类别.试验结果表明,故障诊断的正确率达到 100%,故障位置诊断的正确率达到 93.3%,且算法具有较好的稳定性.

关键词: 案例推理;滚动轴承;故障诊断

中图分类号: TH17 **文献标志码:** A **文章编号:** 0253-987X(2011)11-0079-06

Rolling Bearing Fault Diagnosis Approach Based on Case-Based Reasoning

FU Xinzhe, ZHANG Youyun, ZHU Yongsheng

(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: The case-based reasoning approach is introduced into rolling bearing fault diagnosis. To solve the complexity of feature selection and weights optimization, a Filter/Wrapper integrated feature selection algorithm is proposed. Neighborhood rough set algorithm is applied to select essential features from the feature candidate set, then genetic algorithm is applied to refine the essential feature subset. This method solves the problem of determining the size of neighborhood manually in neighborhood rough set algorithm. Genetic algorithm is also used in feature weights optimization. With the runtime vibration signal of rolling bearing as the basic information, a rolling bearing fault case database is constructed. The historical cases similar to the problem case are recalled and chosen to decide the fault type. The database experiment shows the higher efficiency and accuracy for essential attributes and weights in fault diagnosis.

Keywords: case-based reasoning; rolling bearing; fault diagnosis

目前,对于滚动轴承故障诊断方法的研究主要集中在轴承信号处理方法^[1-2]和智能诊断算法上,而智能诊断算法是在支持向量机、流形学习、神经网络等机器学习算法的基础上构建的^[3-5].这类机器学习方法通过建立问题的模型来解决问题,方法本身具有严谨的数学模型或计算过程,将其应用于滚动轴

承的故障诊断中往往能够得到较高的诊断正确率.基于案例的推理(CBR)是一种方法论^[6],只提出了解决问题的指导性步骤,而没有给出实现这些步骤的数学模型或算法过程.然而,将 CBR 方法应用于滚动轴承故障诊断中仍然具有积极的意义.首先, CBR 方法给出与问题案例相似的历史案例,而不仅

仅是诊断结论,这更有利于辅助人工分析过程.由于实际应用的环境复杂多变,因此在试验条件下,性能优良的智能诊断方法投入现场应用后,效果往往较差,但可供对比参考的相似案例往往比诊断结论更有意义.其次,CBR方法通过参考相似案例来解决问题的本质决定了案例库在CBR系统中的核心地位.案例库的建设和维护是所有CBR系统的主要工作之一,CBR系统解决问题的范围也会随着案例库的扩展而扩展,与其他方法的智能诊断系统相比,CBR系统的成长性更好.

本文针对滚动轴承的特点,提出一种结合特征选择和权重优化的CBR故障诊断方法,并利用Case Western Reserve University轴承数据中心提供的数据构建案例库,从而验证了方法的有效性.

1 CBR滚动轴承故障诊断方法模型

针对滚动轴承的特点,本文结合邻域粗糙集算法(NRS)、遗传算法(GA)和 k 邻近算法(k -NN),提出了采用案例推理的滚动轴承故障诊断方法,其基本过程如图1所示.在诊断一个未知故障类型的滚动轴承(问题案例)时,首先计算出备选属性集,然后在数据库中查看该类案例的权重系数是否存在.若存在,则根据权重系数从备选属性集中挑选出关键属性(非关键属性的权重是0).若不存在,则采用NRS算法粗选关键属性,大部分非关键属性在此步中会被剔除.然后,GA算法被用于自动确定属性权重,并进行进一步的属性精选.当关键属性和相应的权重被确定以后,就可以根据 k -NN模型搜索案例库中的相似案例,并根据相似案例的故障类型来判断问题案例的故障类型.

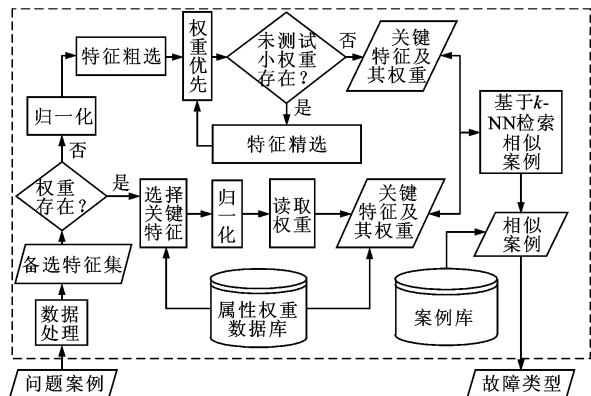


图1 基于案例推理的滚动轴承故障诊断方法

1.1 NRS的属性粗选

从振动信号中可以提取的属性非常多,凭个人

经验挑选有用属性是非常困难和不可靠的,而如果不进行特征选择,把全部属性都用上,不但计算工作量非常大,而且有用属性会被淹没在众多无关属性中,从而降低诊断正确率.特征选择过程中的核心问题是属性的评价标准,根据评价标准的不同,文献[7]把特征选择方法分为Filter型和Wrapper型. Filter型特征选择算法设计一个独立于分类器的评估函数,而Wrapper型特征选择算法直接采用分类正确率作为评估标准.由于直接采用分类成功率作为评价标准,Wrapper型特征选择方法的准确性比Filter型好,但因其计算量决定于分类算法,计算量往往较大,速度比Filter型特征选择方法慢.许多学者提出了复合型的特征选择算法,并结合了Filter和Wrapper型选择器的优点^[8-10].胡清华等人^[11]提出了3种采用NRS的Filter型属性约减算法,本文改进了其中的快速前向属性选择(F2HARNRS)算法,并用于属性的初选. F2HARNRS算法从一个空的属性集开始,在每一轮循环中,会使依赖度增加最大的属性被选入该属性集.当向属性集中添加任何属性,其依赖度增加小于某个给定值时,算法停止.该算法的主要原则是最大化正域、最小化边界.在F2HARNRS算法中,用一个重要的参数 δ 来表示邻域大小,并用 δ 来控制数据分析的粒度,而 δ 不同则计算得到的属性集可能不同,若要获得较好的属性集,则必须设定一个合适的 δ .然而, δ 的合理值与具体的分类问题相关,通常由人根据经验来确定.文献[12]中研究了 δ 对属性选择结果的影响,并给出了 δ 的推荐取值范围为 $[0.1, 0.2]$.为了使本文提出的模型的适用性更好,参考文献[12]给出的范围并结合本文的试验结果,取 δ 为 $0.05 \sim 0.45$,按 0.1 的步长取5个值,所有的由 δ 计算得到的属性集取并集.最后的属性集可能包含一些不必要的属性,但这些属性将在后面的属性精选步骤中被剔除.改进后的算法伪代码为

输入: $\langle U, CUD \rangle$

输出: red

```

1: for  $i = 0 : 4$ 
2:    $\delta = 0.1i + 0.05$ 
3:    $red_i = F2HARNRS(U, CUD, \delta)$ 
4: end

```

```

5: return  $red = red_0 \cup red_1 \cup red_2 \cup red_3 \cup red_4$ 

```

1.2 GA的权重优化和属性精选

灰色关系分析、粗糙集、遗传算法等^[13-15]都被用

于获取属性权重,其中遗传算法在权重优化方面表现得尤为突出.文献[15-16]在 CBR 系统中采用遗传算法优化属性权重,取得了较好的效果.把 GA 算法应用于现实问题的主要工作是设计编码/解码方法、适应度函数和遗传算子,如图 2 所示.由于本文需要优化的对象是属性的权重系数,因此每一个染色体就是一组权重系数.采用二进制编码方法,假设有 n 个关键属性,它们的权重是 $(\omega_1, \omega_2, \dots, \omega_n)$,其中 $0 \leq \omega_n \leq 1, \sum_n \omega_n = 1, 0 \leq n \leq 1$.每一个权重占 8 位二进制数,则每个染色体的长度是 $8n$ 位.由于初始种群是随机产生的二进制码,实际上编码误差并不存在,因此解码误差是由表示精度引起的,其值小于 0.002.

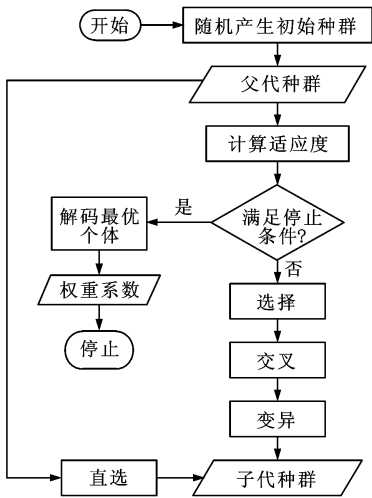


图 2 GA 的流程图

染色体的适应度表示其适应环境的能力,本文中的染色体是一组权重系数,因此采用这组权重系数进行故障诊断,诊断结果的正确率即为该染色体的适应度值.设案例库中存在相关历史案例 N 个,依次以这 N 个历史案例中的一个为问题案例,解码被评估的染色体,得到一组权重系数并进行故障诊断,通过 N 次诊断的正确率即为该染色体的适应度值.选择算子采用轮盘赌算法,交叉算子采用随机双点交叉算法,变异算子采用随机单点变异算法,变异概率为 0.02.试验结果表明,通常情况下,经过小于 100 次的迭代就能够获得满意的优化结果,因此停止条件是某个染色体的分类正确率为 100%,或迭代次数达到 100 次.

在属性粗选阶段,为了避免 δ 取值的影响,取了多组 δ 值,并合并各次特征选择的结果,这样可能会让最后的属性集中产生一些不必要的属性.这些属

性的存在可能会降低分类成功率,并增加计算时间.因此,设计了一个 Wrapper 型的属性精选步骤,基本思想是逐个测试属性,删除不必要的属性,其过程如图 3 所示.将经过属性粗选和权重优化的属性集设为精选属性集,当属性的权重值小于平均值的 20%时,认为该属性是小权重属性.从精选属性集中剔除一个小权重属性,然后重新用 GA 优化其余属性的权重.若 GA 得到的最优个体的适应度下降了,则恢复被剔除的小权重属性,否则继续测试下一个.逐个测试所有小权重属性,直到精选属性集中不存在未测试的小权重属性时,算法结束.

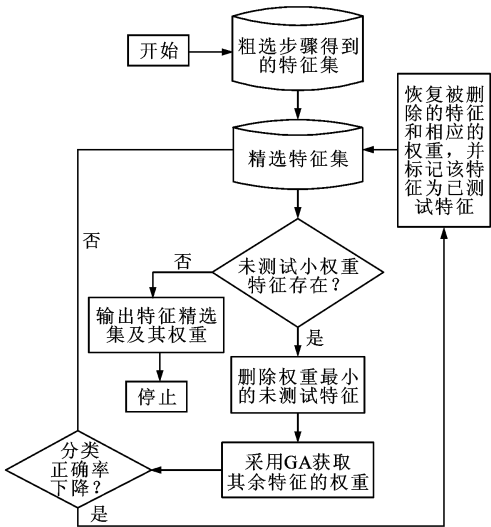


图 3 采用 GA 的属性精选过程

1.3 基于 k -邻近模型的故障判断

k -邻近模型根据 2 个案例之间的距离来度量它们之间的相似性程度,通常案例 C_i 、 C_j 之间的距离 $D(C_i, C_j)$ 可以表示为

$$D(C_i, C_j) = \left(\frac{\sum_{r=1}^n \omega_r (c_{ir} - c_{jr})^2}{\sum_{r=1}^n \omega_r} \right)^{1/2}$$

式中: c_{ir} 表示 C_i 的第 r 个属性; n 表示属性的个数; $\omega_r (r=1, 2, \dots, n)$ 表示权重,它反映了属性之间的相对重要性.文献[17]指出采用不等权重的 CBR 系统比采用等权重的 CBR 系统的性能更好.

对于给定的轴承型号和诊断类型,获得关键属性及其权重后,将权重代入 k -NN 模型计算问题案例与案例库中各案例之间的距离.与问题案例距离越小的案例,相似度越大,选择相似度最大的 3 个最为相似的案例,当 3 个相似案例中的 2 个或 2 个以上属于同一类故障时,判定问题案例的故障类型为

多数相似案例所属的故障类型. 若 3 个案例分别属于不同类型的故障, 则认为问题案例故障类型应为相似度最大案例的故障类型.

2 应用实例

2.1 案例库

利用美国 Case Western Reserve University 轴承数据中心提供的数据构建案例库, 来验证本文提出的方法. 案例库由 84 个滚动轴承故障案例组成, 其中包含 28 个滚动体、28 个内圈和 24 个外圈故障案例, 以及 4 个正常案例. 在数据的获取试验中, 驱动端的测试由滚动轴承支撑的马达主轴作振动, 制造故障的方法是用电火花机床在试验轴承的滚动体、内圈或外圈上打盲孔.

2.2 备选属性集

对于每个案例, 均从振动信号中提取 55 个特征量作为备选属性集. 这 55 个备选属性分为 11 个时域属性、12 个频域属性和 32 个小波包能量值(5 层小波包分解). 其中, 11 个时域指标的计算公式在表 1 中列出, 运用对振动信号进行傅里叶变换得到的频域序列 $f(t)$ 取代原时域信号 $x(t)$, 并代入表 1 的公式中, 得到了前 11 个频域属性. 最后一个频域属性是频域重心, 其公式在表 1 的最后一项列出. 55 个备选属性分别为均值、方差、方根幅值、均方根、最

表 1 属性的计算公式

属性	计算公式
均值	$\bar{x} = \frac{1}{n} \sum_{t=1}^n x(t)$
方差	$x_v = \frac{1}{n} \sum_{t=1}^n (x(t) - \bar{x})^2$
方根幅值	$x_r = \left(\frac{1}{n} \sum_{t=1}^n x(t) ^{1/2} \right)^2$
均方根	$x_{rms} = \left(\frac{1}{n} \sum_{t=1}^n x(t)^2 \right)^{1/2}$
峰峰值	$\hat{x} = E\max(x(t))$
偏斜度	$x_{ske} = \frac{1}{n} \sum_{t=1}^n (x(t) - \bar{x})^3$
峭度指标	$x_{kur} = \frac{1}{n} \sum_{t=1}^n (x(t) - \bar{x})^4$
波形指标	$S = x_{rms} / \bar{x}$
峰值因子指标	$C = \hat{x} / x_{rms}$
脉冲指标	$C = \hat{x} / \bar{x}$
裕度指标	$L = \hat{x} / x_r$
频域重心	$F_c = \sum f_s(f) / \sum s(f)$

大值、偏斜度、峭度、脉冲指标、峰值指标、裕度指标、波形指标、频域均值、频域方差、频域方根幅值、频域均方根、频域最大值、频域偏斜度、频域峭度、频域脉冲指标、频域峰值指标、频域裕度指标、频域波形指标、频域重心和第 1 至第 32 个小波包能量值. 其中, 第 i 个小波包能量值(P_e)表示为第 iP_e .

2.3 单次训练结果分析

从案例库中随机抽取 15 个案例作为测试案例, 其余案例作为训练案例, 保持各类案例的抽取比例相近, 分别诊断有、无故障及故障位置. 属性的粗选结果如图 4 和图 5 所示. 其中, 第 $2P_e$ 表示第 2 个小波包的能量值, 其他依次类推. 从图中可以看出, 当诊断是否有故障时, 进行属性粗选得到的关键属性只有 4 个, 且多次重复计算得到的结果有较大差别, 这说明诊断轴承是否有故障相对容易, 存在多组可选关键属性. 诊断故障位置时, 属性粗选得到的关键属性较多, 说明诊断故障位置相对复杂, 需要的属性更多.

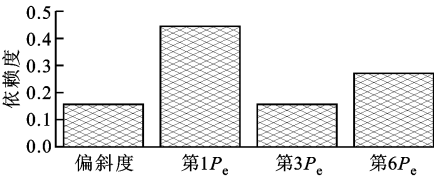


图 4 诊断有、无故障时的属性粗选结果

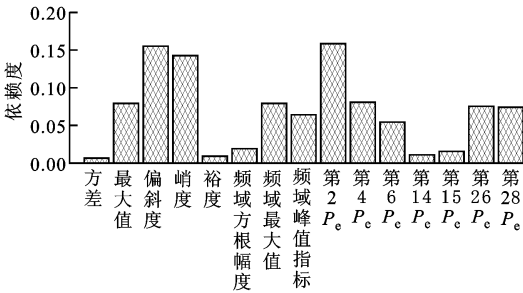


图 5 诊断故障位置时的属性粗选结果

图 6 和图 7 是权重优化和属性的精选结果, 从图中可以看出, 诊断是否有故障时, 偏斜度, 第 1、3、6 个小波包的能量值被选择为关键属性. 诊断故障位置时, 方差, 最大值, 第 3、4、16、26 个小波包的能量值被选择为关键属性. 相对于属性粗选阶段的结果, 关键属性从 14 个减少到 6 个, 有 8 个属性在属性精选阶段被剔除.

在 2 种诊断方法下, 属性之间的权重差别较大, 说明各种属性重要性的差别也较大. 在被选中的关键属性中, 小波包能量值都占到较大的比例和较高

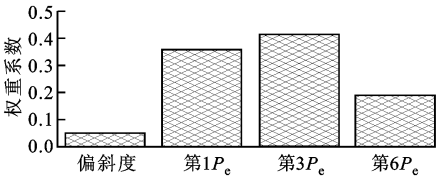


图 6 诊断有、无故障时得到的关键属性及其权重

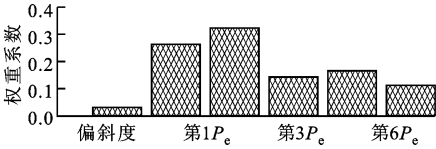


图 7 诊断故障位置时得到的关键属性及其权重

的权重,但 32 个小波包能量值中只有少数被选中.可见,对于给定型号的轴承,部分频带上能量值的属性是很敏感的,这对诊断轴承故障具有重要的价值,但多数频带上能量值的属性是不重要的.

采用图 6、图 7 中的关键属性及其权重,对 15 个测试案例进行故障诊断,诊断有、无故障的正确率为 100%,诊断故障位置的正确率为 93.3%.对比 2 种诊断方法,诊断故障时被选中的属性较少,且正确率更高.可见,判断轴承有、无故障比判断故障位置更容易.

2.4 多次训练结果分析

从 84 个案例中依次抽取 10,15,20,...,70 个案例作为测试案例,则剩余的训练案例数量依次为 74,69,64,...,14 个,取案例的方法是随机获取,但要求保证各类抽取比例相近.用训练案例集作为案例库进行训练,将得到的关键属性和权重用于诊断测试案例的故障位置,每个数量的测试案例独立进行 3 次试验.图 8 为诊断正确率随训练案例数量的变化趋势,从图中可以看出,当训练案例数目小于 49 时,诊断正确率随着训练样本的数量增加而提高.但是,当训练案例数目达到 49 以上时,诊断正确率上下波动,没有明显的上升趋势.造成这种现象的

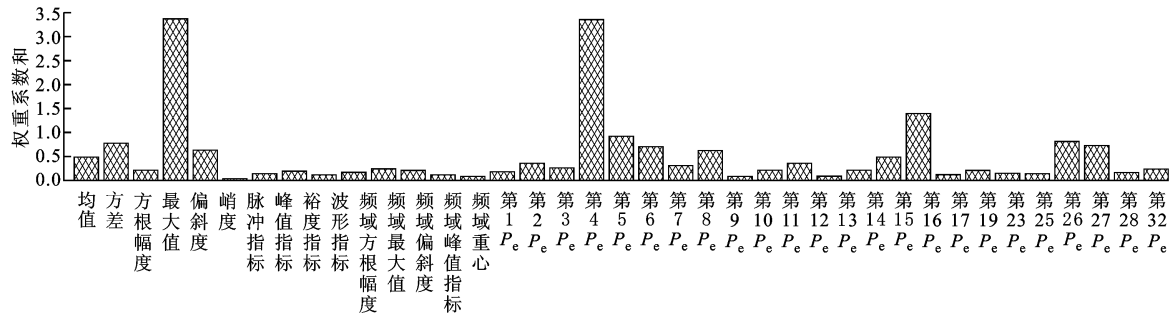


图 9 多次训练得到的权重值累加图

原因是训练案例越多,本文提出的模型获取关键属性及其权重就越合理,同时案例库的覆盖范围也越大,从而能够获取更好的相似案例.当训练案例增加到一定数量后,案例库分布已经能够很好地表征该类问题的期望分布,继续增加案例数量,并不能显著地提高 CBR 系统的性能,这也证明了建立案例库修剪机制的可行性.

从图 9 中可以看出,在累加训练样本数量为 49~74 之间的 18 组属性权重中,有 37 个属性在至少一次训练中被选择为关键属性,而通常单次训练得到的关键属性不会多于 15 个.这说明多次训练得到的关键属性具有差异性,其原因为训练案例不同、各个属性之间并非完全独立,例如,脉冲指标通常与峰峰值正相关.从图中还可以看出,权重值分布比较集中,最大值、第 3 小波包能量、第 16 小波包能量在 18 次训练中得到的权重和较大,而大部分属性的权重和则很小(权重和大于 0.25 的属性只有 15 个).这说明本文提出的获取关键属性及其权重的算法有较好的稳定性,并且多次训练的结果相似.

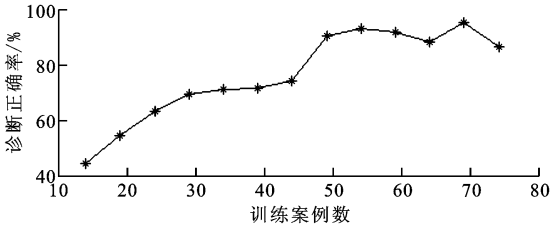


图 8 训练案例数量-诊断正确率的关系曲线

3 结 论

本文提出了采用案例推理的滚动轴承故障诊断方法,设计了 Filter、Wrapper 复合型的特征选择策略,用邻域粗糙集算法从备选属性集中粗选属性,用遗传算法进一步精选属性,并同时优化属性权重.结合 Filter、Wrapper 类型的属性选择算法的优点,解决了邻域粗糙集算法中需要人工确定邻域大小的问

题.应用实例的试验结果表明,大部分非关键属性在属性粗选阶段被剔除,少数非关键属性在属性精选阶段被剔除.采用本文提出的模型获取的关键属性及其权重进行滚动轴承故障的诊断,诊断是否有故障的正确率达到100%,诊断故障位置的正确率达到93.3%.多次训练的结果表明,本文提出的算法有较好的稳定性,同时证明案例库中的案例数量并非越多越好,当案例数达到一定数量后,诊断正确率不会随训练案例数的增加而明显提高.

参考文献:

- [1] 王晓冬,何正嘉,訾艳阳.滚动轴承故障诊断的多小波谱峭度方法[J].西安交通大学学报,2010,44(3):77-81.
WANG Xiaodong, HE Zhengjia, ZI Yanyang. Spectral kurtosis of multiwavelet for fault diagnosis of rolling bearing[J]. Journal of Xi'an Jiaotong University, 2010,44(3):77-81.
- [2] 毋文峰,陈小虎,苏勋家.基于经验模式分解的单通道机械信号盲分离[J].机械工程学报,2011,47(4):12-16.
WU Wenfeng, CHEN Xiaohu, SU Xunjia. Blind source separation of single-channel mechanical signal based on empirical mode decomposition[J]. Journal of Mechanical Engineering,2011,47(4):12-16.
- [3] 汤宝平,李锋,陈仁祥.基于Littlewood-Paley小波支持向量机的故障诊断[J].振动与冲击,2011,1(30):128-131.
TANG Baoping, LI Feng, CHEN Renxiang. Fault diagnosis based on Littlewood-Paley wavelet support vector machine[J]. Journal of Vibration and Shock, 2011,1(30):128-131.
- [4] 栗茂林,王孙安,梁霖.利用非线性流形学习的轴承早期故障特征提取方法[J].西安交通大学学报,2010,44(5):45-49.
LI Maolin, WANG Sun'an, LIANG Lin. Feature extraction for incipient diagnosis of rolling bearings based on nonlinear manifold learning[J]. Journal of Xi'an Jiaotong University,2010,44(5):45-99.
- [5] 吴元修.神经网络在风电机组机械传动系统故障诊断中的应用研究[J].制造业自动化,2010(10):130-131.
WU Yuanxiu. Application research of the fault diagnosis on the wind turbine mechanical transmission system with neural networks[J]. Manufacturing Automation, 2010(10):130-131.
- [6] WATSON I. Case-based reasoning is a methodology not a technology[J]. Knowledge-Based Systems, 1999, 12(1/6):303-308.
- [7] GUYON I, ELISSEEFF A. An introduction to variable and feature selection[J]. The Journal of Machine Learning Research, 2003(3):1157-1182.
- [8] UNCU O, TURKSEN I. A novel feature selection approach: combining feature wrappers and filters[J]. Information Sciences, 2007, 177(2):449-466.
- [9] ZHU Z, ONG Y S, DASH M. Wrapper/filter feature selection algorithm using a memetic framework[J]. Systems Man and Cybernetics: Part B Cybernetics, IEEE Transactions on Systems, 2007, 37(1):70-76.
- [10] PENG H, LONG F, DING C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(8):1226-1238.
- [11] HU Q, YU D, LIU J, et al. Neighborhood rough set based heterogeneous feature subset selection[J]. Information Sciences, 2008, 178(18):3577-3594.
- [12] 胡清华.混合数据知识发现的粗糙计算模型和算法[D].哈尔滨:哈尔滨工业大学,2008.
- [13] LIN R, WANG Y, WU C, et al. Developing a business failure prediction model via RST, GRA and CBR[J]. Expert Systems with Applications,2009,36(2):1593-1600.
- [14] JIANG Y, CHEN J, RUAN X. Fuzzy similarity-based rough set method for case-based reasoning and its application in tool selection[J]. International Journal of Machine Tools and Manufacture, 2006,46(2):107-113.
- [15] SHIN K, HAN I. Case-based reasoning supported by genetic algorithms for corporate bond rating[J]. Expert Systems with Applications, 1999, 16(2):85-95.
- [16] AHN H, KIM K. Bankruptcy prediction modeling with hybrid case-based reasoning and genetic algorithms approach[J]. Applied Soft Computing, 2009, 9(2):599-607.
- [17] NUNEZ H, SANCHEZMARRE M, CORTES U, et al. A comparative study on the use of similarity measures in case-based reasoning to improve the classification of environmental system situations[J]. Environmental Modelling & Software, 2004,19(9):809-819.

(编辑 管咏梅)