

结合交叠区异点统计和相关分析的 免疫克隆特征选择方法

栗茂林, 梁霖, 王孙安, 刘弹
(西安交通大学机械工程学院, 710049, 西安)

摘要: 为了解决机械故障诊断中的特征选择问题, 利用免疫克隆选择算法, 提出了一种结合交叠区异点统计和相关性分析的免疫克隆特征选择方法, 可有效地去除不相关特征和冗余特征. 基于空间分布的交叠区异点, 设计了交叠区异点统计的优化指标; 基于 J 散度距离, 设计了一种冗余特征的评估指标; 基于免疫克隆选择算法, 结合交叠区异点统计优化指标和冗余特征评估指标, 提出了一种免疫克隆特征选择新算法. 仿真和实际工程应用的结果表明: 文中提出的方法比常用的特征选择方法更加有效, 选出的特征分类精度更高, 特征子集更小, 更能满足故障诊断的需求.

关键词: 故障诊断; 特征选择; 免疫克隆选择算法; 相关分析

中图分类号: TH165.3 **文献标志码:** A **文章编号:** 0253-987X(2012)05-0050-07

Immune Clone Feature Selection Method Combining Different Point Statistics of Overlap Region with Correlation Analysis

LI Maolin, LIANG Lin, WANG Sun'an, LIU Dan
(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: To treat the feature selection for mechanical fault diagnosis, a novel immune clone feature selection method combining different point statistics of overlap region with correlation analysis is proposed with which the irrelevant features and redundant features can be removed effectively. According to the different point of the overlap region in space distribution, an optimization index is designed. An evaluation index of redundant features in correlation analysis following J -divergence is put forward. Then an immune clone feature selection strategy (ICFSS) is constructed with the two indexes. The simulation and industrial applications show its higher precision and smaller feature subset for more effective fault diagnosis.

Keywords: fault diagnosis; feature selection; immune clone selection method; correlation analysis

特征选择就是从原始特征集中选出满足评估标准的特征子集, 更好地进行后续分类识别, 本质是要去除不相关及冗余特征, 这直接关系到故障诊断的准确性和预测的可靠性. 随着现代工业生产设备不断向大型化、复杂化和自动化方向发展, 影响系统正常运行的因素骤增, 设备维护和故障诊断的难度增

大, 使得特征提取和选择成为当前机械故障诊断研究中的一个瓶颈问题.

针对特征选择问题, Yu 等人^[1]提出了一个包括评价、停止条件和验证的基本框架; 廖广兰等人^[2]利用 SOM 网络聚类特点提出了一种特征选择方法; Dash 等人^[3]将特征选择视为组合优化和多峰优化问题; Chen^[4]采用全局最优搜索技术进行特征提

取,但无法实现对特征数量的控制. 随机搜索策略被广泛地应用于特征选择中,如禁忌搜索算法^[5]和遗传算法(GA)^[6-7],但是 GA 易陷入局部最优,不适合多峰优化问题. 模拟生物免疫系统的人工免疫克隆选择算法已被应用于特征选择中^[8-9]. 人工免疫系统(AIS)比 GA 更适合解决多峰优化问题,但特征选择的效果受特征集合评价方法的影响较大.

特征集合的评价策略有缠绕法和过滤法 2 种. 缠绕法将分类器算法嵌入到特征选择中,依据分类的效果来进行选择,分类精度高,但效率较低,且要求数据样本充分. 文献[8-9]均采用了分类精度作为评估指标. 过滤法利用数据的内在特性来进行特征评价和选择,显然更适合于机械故障诊断中的特征选择需求^[10-11]. 机械故障诊断中的特征选择不仅追求分类精度高,同时还希望特征子集尽量小. 针对该多目标优化问题,GA 难以满足,而如果将多目标优化转化为单目标优化问题,即将分类精度的衡量指标(如散布矩阵)和特征数进行加权,则权系数的选择又是一个难题.

本文基于 AIS 的多峰优化能力,提出一种基于交叠区异点统计和相关分析的免疫克隆特征选择方法——ICFSS,设计了 2 个优化指标,剔除了不相关特征和冗余特征,从而实现了分类精度和特征子集大小 2 个方面的优化. 将该方法应用于 UCI 标准数据集和实际机械故障数据集,取得了良好的效果.

1 交叠区异点统计优化指标

在克隆选择算法中,抗原为选择的特征子集,亲和度函数用来衡量所选子集的好坏,因此亲和度函数的设计直接关系到特征选择策略的性能. 本文在分析现有文献中亲和度函数使用效果的基础上,设计了交叠区异点统计的优化指标.

首先给出邻域和异点的定义,具体如下.

定义 1 邻域:以点 a 为中心的任何开区间称为点 a 的邻域,记作 $U(a)$. 设 δ 为任一正数,则开区间 $(a-\delta, a+\delta)$ 为 a 的 δ 邻域,记作 $U(a, \delta) = \{x | a-\delta < x < a+\delta\}$, a 为邻域的中心, δ 为邻域的半径.

定义 2 异点:点 x 在点 a 的 δ 邻域里,即 $x \in U(a, \delta)$,如果 x 与 a 不属于同一类,则称 x 为异类点,如果 x 还满足条件 $d(x, A) \leq d(a, A)$,其中 d 定义为两点之间的距离,一般取欧氏距离, A 为 a 所属类别的中心点,即 x 与 a 所属类别中心点的距离小于等于 a 与其所属类别中心点的距离,则定义 x 为 a 的异点,记作 $\tilde{a}_i$.

图 1 所示为二维空间中的邻域和异点示意图,虚线圆画出了点 a 的 δ 邻域范围,邻域内有 4 个点,其中 1 个同类点,3 个异类点,在异类点中,只有左侧实心圆点为异点. 邻域的概念可以拓展到多维空间,即在多维空间,点 a 的 δ 邻域定义为以 a 为中心、 δ 为半径的球体.

根据上述邻域和异点的定义,提出基于空间交叠区异点统计的优化指标 $\tilde{N}$,具体定义为

$$\tilde{N} = \sum_{i=1}^n \tilde{a}_i, \quad i = 1, 2, \dots, n \quad (1)$$

式中: $\tilde{a}_i$ 为数据集中点 x_i 的异点. 统计指标 $\tilde{N}$ 给出的是数据集中所有点的异点总和,值越小就意味着各类数据分得越开,交叠越小.

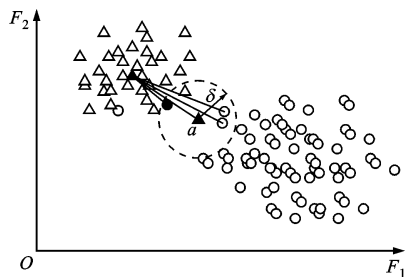


图 1 邻域和异点示意图

2 基于相关性分析的冗余特征判别

J 散度用来度量 2 个概率分布之间的距离,也可以度量 2 个向量之间的相似程度,用于进行运行状态的样本建立和识别^[12]. 这里采用 J 散度距离来度量 2 个特征向量间的相似程度,即

$$J(\mathbf{X}_1, \mathbf{X}_2) = \frac{1}{2N} \sum_{i=1}^N \left[\frac{x_{1i}}{x_{2i}} + \frac{x_{2i}}{x_{1i}} - 2 \right] \quad (2)$$

式中: $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 是待比较的 2 个特征向量; x_{1i} 和 x_{2i} 分别为向量 $\mathbf{X}_1$ 和 $\mathbf{X}_2$ 的特征值, $i=1, 2, \dots, N$, N 为样本点数.

在特征选择算法中,冗余特征可以通过 J 散度相关性分析来剔除. 在这里,设计了基于特征相关性分析的冗余特征评估指标

$$J(\mathbf{S}_1, \mathbf{S}_2) = \frac{1}{2m} \sum_{j=1}^m \left[\frac{S_{1j}}{S_{2j}} + \frac{S_{2j}}{S_{1j}} - 2 \right] \quad (3)$$

将二维特征向量 $\mathbf{S}_1$ 和 $\mathbf{S}_2$ 中所有的特征值依据样本类属关系划分区间,即有 m 个类就划分为 m 个区间,二维特征向量在对应的第 j 个区间内的所有特征值总和是 S_{1j} 和 S_{2j} , $j=1, 2, \dots, m$. 这样做可以有效地消除小特征值随机干扰的影响,提高 J 散度的稳定性.

当二维特征向量完全相同时,两者间的 J 散度为 0,二维向量之间的差异越大,两者之间的 J 散度也越大.另外, $J(\mathbf{S}_1,\mathbf{S}_2)=J(\mathbf{S}_2,\mathbf{S}_1)$,即表示用 J 散度描述的第 1 维向量对第 2 维向量的相似程度和第 2 维向量对第 1 维向量的相似程度是一致的.

3 面向特征选择的 ICFSS 算法

本文基于抗体克隆选择机理,利用交叠区异点统计优化指标和相关性冗余特征评估指标,采用免疫克隆选择算法,提出一种面向特征选择的 ICFSS 算法,可通过优化交叠区异点统计指标,将数据集维数大大降低,去除不相关或弱相关特征,然后采用基于相关性的冗余特征评估指标去除冗余特征,最终获取分类精度高且特征数少的特征子集,实现特征选择.

算法的基本步骤如下.

步骤 1:根据编码方式产生初始化抗体种群,设定抗体种群数、进化代数、选择概率、变异概率等相关参数,以交叠区异点统计指标作为亲和度函数来计算初始种群的亲和度.

步骤 2:依据亲和度和设定的抗体克隆规模,进行克隆增殖操作、免疫基因操作和克隆选择操作,获得新的抗体种群.

步骤 3:重新计算新抗体种群的亲和度.

步骤 4:迭代运行,若满足终止条件,则输出特征子集并进入下一环节,否则返回步骤 2 继续优化.

步骤 5:对免疫优化后的特征子集进行基于相关性的冗余特征评估,去除冗余特征,输出最终特征子集.

在算法中,抗体采用二进制编码,即基于 $\{0,1\}$ 的一维二进制编码形式,一个抗体表示一个特征子集.对于 m 维特征,抗体为长度 m 的二进制特征矢量,个体中的 0 表示对应特征项不包含于所选特征子集中,1 表示对应特征项包含于所选特征子集中.例如,某个抗体编码为 $\mathbf{A}_k=\{00101011\}$,即表示在 8 个特征中,所选特征子集 $G=\{3,5,7,8\}$.由 n 个抗体组成初始抗体种群 $A=\{A_1,A_2,\cdots,A_n\}$,正整数 n 为抗体种群规模.

免疫基因操作即抗体的变异策略,它直接影响算法的运行速度和效果.由于抗体为二进制编码,个体中不同基因位的改变引起解空间的变化幅度不等,即各个基因位的重要性不相等,因此算法中对于分类较重要的特征,从 1 变异到 0 的概率较小,从 0 变异到 1 的概率较大,即尽量保留对分类较重要的

属性特征,而对于分类不重要的特征,从 1 变异到 0 的概率较大,从 0 变异到 1 的概率较小,从而可以有效地提高优化的效率和质量,减少优化时间.属性重要性的衡量方法很多,本文提出的算法通过对各属性特征计算 K 近邻分类结果来获取属性重要性,进一步实现抗体依据属性重要性的变异.

4 仿真及工程应用

4.1 仿真分析

从 UCI 机器学习数据库中选取 Iris、Wine、Ionosphere 和 Diabetes 4 个数据集进行测试,表 1 为这 4 个数据集的基本信息.

表 1 4 个 UCI 标准数据集的基本信息

参数	数据集			
	Iris	Wine	Ionosphere	Diabetes
样本数	150	178	351	768
数据维数	4	13	34	8
类别数	3	3	2	2

为了对比算法中优化指标的效果,将 ICFSS 算法(不执行第 5 步去冗操作)应用于 UCI 标准数据集,并将所得结果与类内类间优化指标 D 作对比分析,结果见表 2.参数设置:抗体种群数为 20,抗体克隆规模为 10,进化代数为 100,选择概率为 0.2,变异大、小概率分别为 0.8 和 0.2.在优化指标 $\tilde{N}$ 中,不同数据集选择不同的邻域半径,均取数据点之间的平均距离,在指标优化值相同的情况下,选择维数更小的特征子集.

表 2 2 种优化指标的特征选择对比结果

优化指标	特征子集			
	Iris	Wine	Ionosphere	Diabetes
D	$\{S_3,S_4\}$	$\{S_2,S_6,S_7,S_8,S_{10},S_{11},S_{12}\}$	$\{S_1,S_2\}$	$\{S_2\}$
$\tilde{N}$	$\{S_3,S_4\}$	$\{S_1,S_5,S_7,S_{10},S_{13}\}$ $\{S_7,S_{10},S_{13}\}\{S_1,S_7\}$	$\{S_7,S_8,S_{28}\}$ $\{S_9,S_{15},S_{16}\}$	$\{S_2,S_6,S_7,S_8\}$

由表 2 可知,由于 Iris 数据集在第 3 和第 4 维上的数据分布是类内较集中、类间较分散,分类性能好,在其他维上分布较差,因此 2 种优化指标的结果相同,但其他数据集的约减结果相差较大.

为了对比算法中冗余特征评估指标的效果,用

J 散度距离衡量特征间的相关性,将上述 UCI 数据集的优化指标 $\tilde{N}$ 的结果进行冗余特征删除,结果如表 3 所示.

表 3 相关性分析结果

数据集特征子集			
Iris	Wine	Ionosphere	Diabetes
$\{S_4\}$	$\{S_1, S_7, S_{10}\}$ $\{S_7, S_{10}\} \{S_1, S_7\}$	$\{S_7, S_8, S_{28}\}$ $\{S_9, S_{15}, S_{16}\}$	$\{S_2, S_7, S_8\}$

对全部特征集、类内类间指标优化子集及删除冗余特征前、后的交叠区异点统计指标优化子集采用 K 近邻分类来进行比较,结果见表 4,近邻数选为 5.

表 4 特征子集的分类正确率及方差比较

数据集	Iris		Wine		Ionosphere		Diabetes	
	正确率/%	方差	正确率/%	方差	正确率/%	方差	正确率/%	方差
全集	95.94	1.73	95.55	1.69	83.39	2.51	72.56	1.67
指标 D 优化集	96.07	1.64	87.35	3.18	75.09	1.47	66.36	2.58
指标 $\tilde{N}$ 优化集	96.07	1.64	96.91	1.50	87.81	2.30	74.28	1.67
			95.13	1.70	86.70	2.13		
			91.33	2.59				
指标 $\tilde{N}$ 优化集 (去冗后)	95.73	1.71	94.37	1.94	87.81	2.30	73.03	1.74
			92.57	2.05	86.70	2.13		
			91.33	2.59				

4.2 发动机异响数据的工程应用

4.2.1 对象介绍 CA141 型汽油发动机异响信号的采样点数为 2 048,采样频率为 6 kHz,包含正常、活塞敲缸、活塞销响、气缸漏气和气门挺杆响 5 种状态,每种状态选择 50 组数据样本,样本总数为 250. 每个样本包含提取的 6 个典型时域特征:均方幅值 x_{rms} 、波形指标 K 、峰值指标 C 、方差 V 、峭度 k 和偏斜度 S ,具体见图 2~图 7,图中量值均已作归一化处理. 由图可知,发动机由于运动部件较多,故障复杂,不同状态样本在时域特征值上都有所波动,但由于信号信噪比较低,在特征空间中的分布存在不同程度的交叠现象,使得单独依靠某时域指标值无法完全区分 5 种状态. 相比而言,波形指标 K 特征空间中的数据分布较好,有 3 种状态分布都非常集中,且很容易区分,但对于正常和活塞敲缸状态依然很难区分. 偏斜度 S 特征空间中的数据分布最差,前 3 种状态交叠无法区分,后 2 种状态虽有所区分,但依然存在交叠.

4.2.2 优化结果 为了对比异点统计和冗余特征评估 2 个指标的优化效果,将算法分开进行,先看优

从表 4 可以看出:通过免疫克隆算法优化指标 $\tilde{N}$ 得到的特征子集比全部特征集合的分类精度都高,且方差更小,说明优化算法的性能更稳定;与优化指标 D 得到的子集相比,除 Iris 数据外,分类精度都有很大提高,同时除 Ionosphere 数据外,方差也都更小,说明了优化指标 $\tilde{N}$ 的有效性. 另外,对比优化指标 $\tilde{N}$ 去冗前、后的效果可知,由于采用了相关性冗余特征评估指标,去除冗余特征后的特征子集在分类正确率上虽然比去冗前有所降低,但却获得了维数更低的特征子集,这对故障诊断将更有意义.

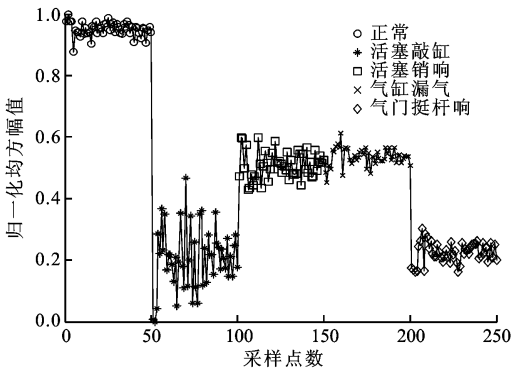


图 2 均方幅值 x_{rms} 特征空间中的数据分布

化指标的效果,再看去冗指标的效果. 参数设置:抗体种群数为 20,抗体克隆规模为 10,进化代数为 100,选择概率为 0.2,变异大、小概率分别为 0.8 和 0.2,邻域选择为数据点间的平均距离. 优化的结果有 $\{K, V\}$ 、 $\{x_{rms}, K\}$ 、 $\{x_{rms}, C\}$ 、 $\{C, V\}$ 、 $\{K, C\}$ 、 $\{x_{rms}, K, V\}$ 和 $\{x_{rms}, K, C\}$. 从结果可以看出, K 等指标能较好区分各类状态数据的特征,因此被选中,而 S 等指标不能很好区分各类状态数据的特征,因此未被选中. 图 8~图 12 给出了部分优化结果,各

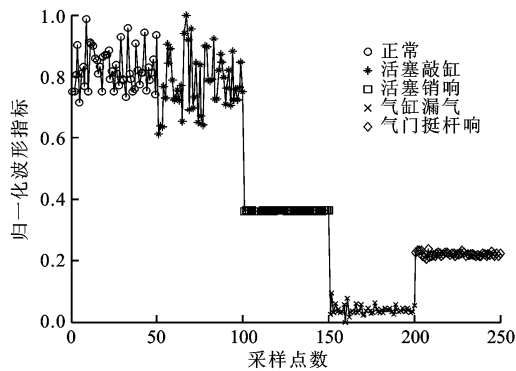


图 3 波形指标 K 特征空间中的数据分布

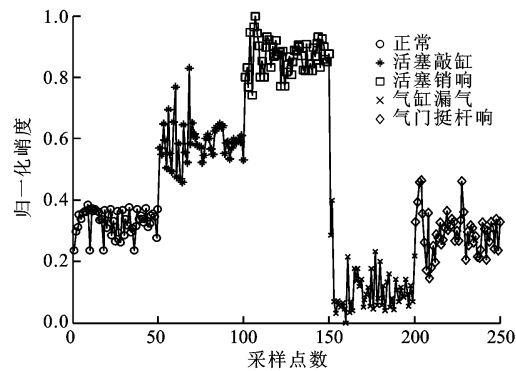


图 6 峭度 k 特征空间中的数据分布

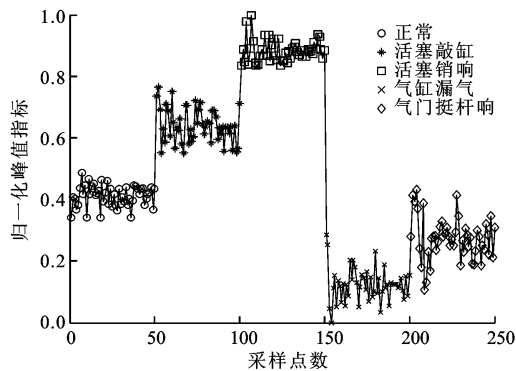


图 4 峰值指标 C 特征空间中的数据分布

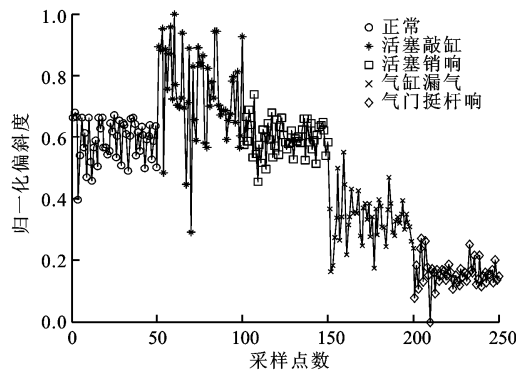


图 7 偏斜度 S 特征空间中的数据分布

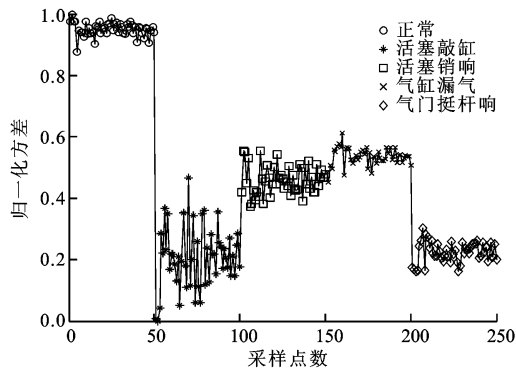


图 5 方差 V 特征空间中的数据分布

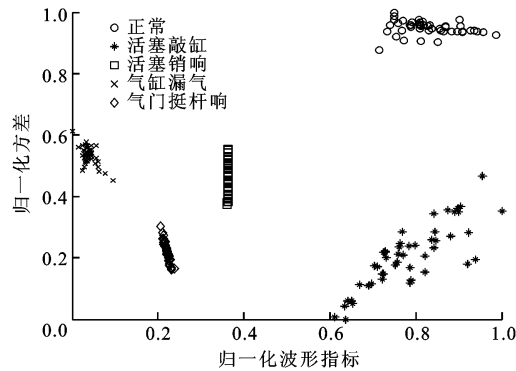


图 8 K 和 V 二维空间中的数据分布

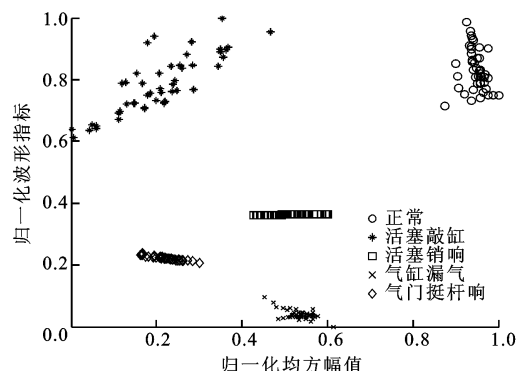


图 9 x_{rms} 和 K 二维空间中的数据分布

类都很好地实现了分离。

4.2.3 去冗效果 采用相关性冗余特征评估指标对发动机数据各维特征间的相关性进行度量,结果如表 5 所示.由表 5 可知, x_{rms} 和 V 二维特征非常相似, C 和 k 二维特征也很相似,而 x_{rms} 、 K 和 V 之间的相似性较差.因此,去冗后的结果有 $\{K, V\}$ 、 $\{x_{rms}, K\}$ 、 $\{x_{rms}, C\}$ 、 $\{C, V\}$ 、 $\{K, C\}$. 经过相关性分析约减后的特征子集不仅能完全区分各类样本,分类正确率达到 100%,而且更加精简,更适合于故障诊断的需求。

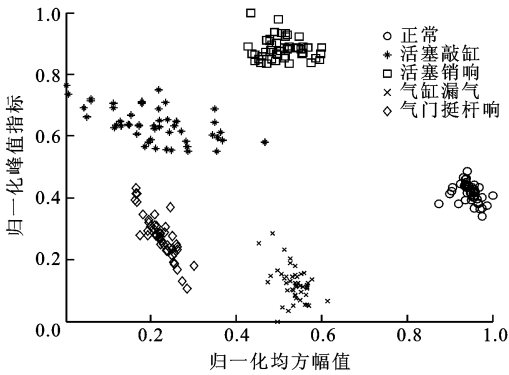


图 10 x_{rms} 和 C 二维空间中的数据分布

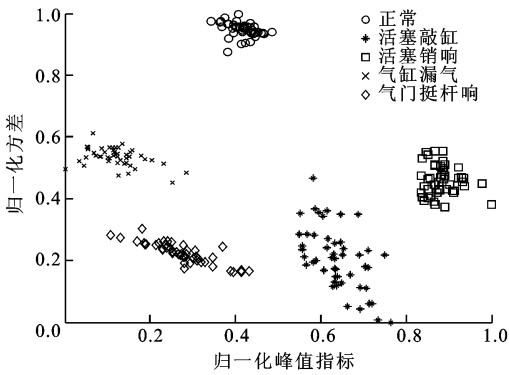


图 11 C 和 V 二维空间中的数据分布

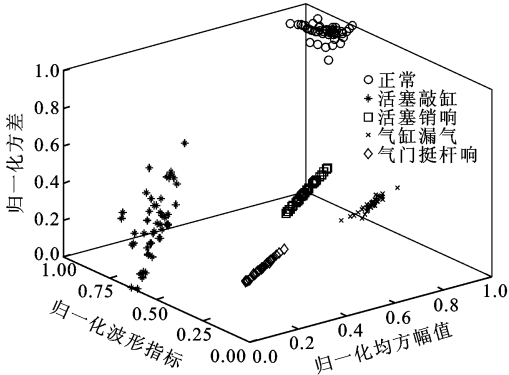


图 12 x_{rms} 、 K 和 V 三维空间中的数据分布

表 5 发动机数据各维特征间的 J 散度距离

特征	J 散度距离					
	x_{rms}	K	C	V	k	S
x_{rms}	0.000 0	1.388 4	0.510 2	0.001 0	0.601 5	0.252 0
K	1.388 4	0.000 0	0.288 1	1.382 3	0.296 6	0.706 3
C	0.510 2	0.288 1	0.000 0	0.523 3	0.008 3	0.171 4
V	0.001 0	1.382 3	0.523 3	0.000 0	0.614 1	0.256 2
k	0.601 5	0.296 6	0.008 3	0.614 1	0.000 0	0.240 3
S	0.252 0	0.706 3	0.171 4	0.256 2	0.240 3	0.000 0

4.3 轴承数据的工程应用

4.3.1 数据介绍 选用文献[11]中的一组滚动轴承故障数据进行特征选择,该数据集包含 55 个样本,其中 34 个为正常状态样本,21 个为不同程度的故障状态样本,由 12 个原始特征组成,具体见表 6.

表 6 轴承数据的具体特征含义

特征维数	特征含义
1	0.5~2.0 kHz 范围内的噪声水平(S_1 ,dB)
2	1.0~11.0 kHz 范围内的噪声水平(S_2 ,dB)
3	0.5~2.0 kHz 范围内的噪声水平(S_3 ,MPa)
4	1.0~11.0 kHz 范围内的噪声水平(S_4 ,MPa)
5	0.1~1.0 kHz 范围内的轴向振动(S_5 ,m/s ²)
6	1.0~11.0 kHz 范围内的轴向振动(S_6 ,m/s ²)
7	0.1~1.0 kHz 范围内的垂直方向振动(S_7 ,m/s ²)
8	1.0~11.0 kHz 范围内的垂直方向振动(S_8 ,m/s ²)
9	0.1~1.0 kHz 范围内的水平方向振动(S_9 ,m/s ²)
10	1.0~11.0 kHz 范围内的水平方向振动(S_{10} ,m/s ²)
11	$S_{11}=(S_5^2+S_7^2+S_9^2)^{1/2}$ (m/s ²)
12	$S_{12}=(S_6^2+S_8^2+S_{10}^2)^{1/2}$ (m/s ²)

4.3.2 特征选择结果 采用 ICFSS 算法对轴承数据进行特征提取,参数设置如下:抗体种群数为 20,抗体克隆规模为 10,进化代数 为 100,选择概率为 0.2,变异大、小概率分别为 0.8 和 0.2,邻域选择为数据点间的平均距离.优化的结果有 $\{S_5, S_{12}\}$ 、 $\{S_{11}, S_{12}\}$ 、 $\{S_5, S_6\}$ 和 $\{S_7, S_{12}\}$.这 4 个优化结果中都不包含噪声水平,但却都包含了影响轴承性能的低、高频分量,结果完全符合工程实际.另外,在优化结果的数据分布图中可以看出,轴承数据在优化特征下实现了正常状态和故障状态的很好分离,用较少的特征获取了更好的分类效果,而且还提供了多个特征组合可供选择,对实际故障诊断应用具有重要意义.

5 结 语

(1)为了提高特征的分类能力,本文基于克隆选择优化算法,设计了交叠区异点统计优化指标和相关性冗余特征评估指标,提出了一种结合交叠区异点统计和相关性分析的免疫克隆特征选择方法(ICFSS),为故障诊断特征选择提供了一种新方法.

(2)在 UCI 和实际故障诊断案例的应用中,通过克隆选择优化算法对交叠区异点统计指标进行优

化,获取的特征子集的分类正确率更高,且方差更小,说明了交叠区异点统计指标的有效性.

(3)在应用中,利用相关性分析对特征冗余性进行评估,去除冗余特征,获取了维数更低特征子集,证明了相关性冗余特征评估指标设计的有效性,这对故障诊断具有更加重要的意义.

参考文献:

- [1] YU Lei, LIU Huan. Efficient feature selection via analysis of relevance and redundancy [J]. Journal of Machine Learning Research, 2004, 5(10): 1205-1224.
- [2] 廖广兰,史铁林,姜南,等. 基于 SOM 网络的特征选择技术研究[J]. 机械工程学报, 2005, 41(2): 46-50.
LIAO Guanglan, SHI Tielin, JIANG Nan, et al. Feature selection based on SOM networks [J]. Chinese Journal of Mechanical Engineering, 2005, 41(2): 46-50.
- [3] DASH M, LIU H. Feature selection for classification [J]. International Journal of Intelligent Data Analysis, 1997, 1(3): 131-156.
- [4] CHEN Xuewen. An improved branch and bound algorithm for feature selection [J]. Pattern Recognition Letters, 2003, 24(12): 1925-1933.
- [5] 张文化,刘素华,侯惠芳. 一种用于特征选择的禁忌搜索算法[J]. 计算机应用与软件, 2010, 27(5): 125-127.
ZHANG Wenhua, LIU Suhua, HOU Huifang. A taboo search algorithm for feature selection [J]. Computer Applications and Software, 2010, 27(5): 125-127.
- [6] JACK L B, NANDI A K. Genetic algorithms for feature selection in machine condition monitoring with vibration signals [J]. IEE Proceedings: Vision, Image and Signal Processing, 2000, 147(3): 205-213.
- [7] 贾要勤,徐光华,孙瑞祥,等. 粗糙集属性约简的基因算法[J]. 中国机械工程, 2000, 11(7): 797-800.
JIA Yaoqin, XU Guanghua, SUN Ruixiang, et al. Reduction of rough set attributes using genetic algorithm [J]. China Mechanical Engineering, 2000, 11(7): 797-800.
- [8] 朱虎明,焦李成. 并行免疫克隆特征选择算法[J]. 西安电子科技大学学报, 2008, 35(5): 853-857.
ZHU Huming, JIAO Licheng. Parallel immune clonal selection for feature selection [J]. Journal of Xidian University, 2008, 35(5): 853-857.
- [9] 尚荣华,焦李成,吴建设,等. 用于非监督特征选择的免疫克隆多目标优化算法[J]. 西安电子科技大学学报, 2010, 37(1): 18-22.
SHANG Ronghua, JIAO Licheng, WU Jianshe, et al. Immune clonal multi-objective algorithm for unsupervised feature selection [J]. Journal of Xidian University, 2010, 37(1): 18-22.
- [10] 梁霖,徐光华. 基于克隆选择的粗糙集属性约简方法[J]. 西安交通大学学报, 2005, 39(11): 1231-1235.
LIANG Lin, XU Guanghua. Reduction of rough set attribute based on immune clone selection [J]. Journal of Xi'an Jiaotong University, 2005, 39(11): 1231-1235.
- [11] 刘弹,徐光华,梁霖,等. 基于邻接区域交叠概率的特征选择方法[J]. 机械工程学报, 2009, 45(2): 1-5.
LIU Dan, XU Guanghua, LIANG Lin, et al. Feature selection method based on overlapped probability of intersection area [J]. Chinese Journal of Mechanical Engineering, 2009, 45(2): 1-5.
- [12] MITRA P, MURTHY C A, PAL S K. Unsupervised feature selection using feature similarity [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(4): 301-312.

[本刊相关文献链接]

- 利用膜粒子群优化和信息熵的医学文本特征选择. 2012, 46(4): 45-51.
- 用于水声目标识别的自适应免疫特征选择算法. 2011, 45(12): 28-32.
- 滚动轴承故障诊断的案例推理方法. 2011, 45(11): 79-84.
- 广义邻域粗集下的集成特征选择及其选择性集成算法. 2011, 45(6): 34-39.
- 基于 Fisher 判据时频分析的运动相关脑电特征选择及优化. 2008, 42(8): 1026-1030.

(编辑 葛赵青)