

免疫编程的故障特征构造方法

梁霖¹, 徐光华², 刘弹¹

(1. 西安交通大学机械工程学院, 710049, 西安; 2. 西安交通大学机械制造系统工程国家重点实验室, 710049, 西安)

摘要: 在原始特征的多项式组合优化中, 针对遗传编程易陷入局部优化解的缺点, 引入了免疫编程的故障诊断优化策略, 将类内类间散布矩阵判据作为抗体的亲和度, 利用克隆、变异和更新等免疫算子实现抗体的优化, 从而获得了最佳分类能力的复合特征. 对发动机异响信号的6个典型时域特征的应用表明, 克隆选择机制维持了抗体群体的多样性, 同遗传编程方法相比, 优化的复合特征具有更好的识别效果.

关键词: 故障诊断; 免疫编程; 特征构造

中图分类号: TH17 **文献标识码:** A **文章编号:** 0253-987X(2007)09-1036-04

Features Generation in Fault Diagnosis Based on Immune Programming

Liang Lin¹, Xu Guanghua², Liu Dan¹

(1. School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China; 2. State Key Laboratory for Manufacturing Systems Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: Aiming at the shortage of premature convergence in the conventional genetic programming, a new feature generation method based on immune programming is put forward, where the new features are constructed by polynomial expressions of the original features, and the antibody affinity is defined by the ratio of inter-class scatter to the intra-class scatter. Combining cloning, hypermutation and replacement operator, the clonal selection optimal algorithm is adopted to search the best feature with excellent classification performance. The experiment of the six time domain features for engine sound signal shows that due to the maintained diversity of antibodies based on the clonal selection principle, the best compound feature constructed by immune programming is endowed with the better classification ability than the feature optimized by genetic programming.

Keywords: fault diagnosis; immune programming; feature generation

特征选择是故障诊断的关键环节, 由于故障机理不甚清楚以及表现不惟一等特点, 导致在特征提取时带有很大的盲目性, 因此所提取的特征集难以保证识别能力. 随着遗传编程方法(GP)的提出, GP在特征的组合变换中取得了一定的应用, 如 Vafaie^[1]通过特征构造和特征选择的交互迭代, 实现了人脸眼型的自动识别; Chen^[2]利用遗传编程方法, 由8个原始特征产生了最佳复合特征; 文献[3-5]也相继研究了基于遗传编程的特征构造方法, 但由于

GP采用了遗传算法的进化策略, 因此存在着局部优化问题, 并且目前对GP的研究并不深入, 这在一定程度上限制了GP的优化能力.

全局优化能力更强的免疫编程方法为函数表达式的优化提供了一种新的解决方案, Peter和Musilek^[6]在克隆选择原理的基础上, 提出了一种免疫编程(IP)策略, 该方法被认为是由计算机自动程序设计的一个比较系统的方法^[7], 同遗传编程相比, 它具有较强的优化能力. 因此, 本文将免疫编程优化

策略引入到故障诊断的特征构造优化中,通过对原始特征的组合和优化,根据特征分类能力利用克隆选择算法进行择优选择,从而获得了具有最佳识别能力的新特征指标,并给出了实际的应用效果。

1 免疫编程的原理

免疫编程 IP 算法是基于克隆选择原理和低亲和度抗体更新机制而提出的,其算法流程^[6]如图 1 所示。

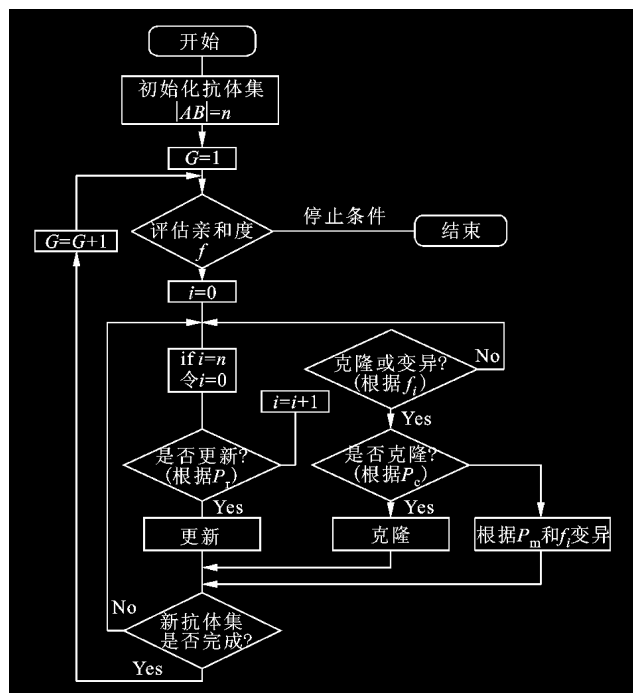


图 1 IP 算法流程图

步骤 1 初始化:初始化抗体种群 A_k ,它由 n 个抗体 A_{k_i} 生成。

步骤 2 评估:提呈抗原 A_g (表征问题的解),计算抗原和所有抗体 $A_{k_i} \in A_k$ 之间的亲和度 f_i 。

步骤 3 更新:在更新概率 P_r 下获得一个新抗体,并置于新抗体集 A'_k 中。

步骤 4 克隆:如果没有生成新的抗体,则进入到克隆阶段,即对当前的抗体集按与其亲和度成比例的克隆概率 P_c 进行抗体克隆,并将克隆体置于 A'_k 中。

步骤 5 高频变异:如果在克隆操作中被选择的高亲和度抗体没有进行克隆,则将它以与其亲和度成反比的变异概率 P_m 进行变异,并将获得的新抗体置于 A'_k 中。

步骤 6 抗体集的迭代:重复执行步骤 3~步骤 5,直到一个完整的 A'_k 生成。

步骤 7 进化代数的迭代:进化代数 $G=G+1$,新抗

体集将返回执行步骤 2 的评估,整个计算过程不断重复直至满足停止条件。

由算法步骤可知,IP 算法通过抗体集的进化来获得最优抗体,在进化初期,随机产生的候选解保证了抗体集的多样性,而更新、克隆以及超变异等免疫操作可促进抗体集的进化,有效维持抗体集的多样性,并且扩大了解的空间范围。通过亲和力的概率选择,使具有高亲和度的抗体通常被选择进行克隆或变异,而低亲和度的抗体通常被更新。

通过对亲和度的评估,可以有效地评价抗体个体的质量,而亲和度和算法中的概率参数可以共同确定新抗体的生成。在每次迭代中,通过更新、克隆和变异操作,最多产生一个抗体,随着新抗体集的迭代进行,算法将对新抗体集进行评估和免疫操作(更新、克隆和变异),以便进行下一轮的进化。若满足停止条件(获得了最大亲和度的抗体),则算法停止,给出最终的搜索结果。

对于 IP 算法的优化性能,文献^[6]进行了详细的分析,并与 GP 算法进行了对比。结果表明,IP 算法的优化质量明显优于 GP 算法,即 IP 算法 100% 搜索到最优解,但 GP 算法最多也只有 90% 的成功搜索率(对于 500 个个体),而在 200 个个体时的搜索率只有 60%。因此,IP 算法的这种强大的优化能力为诊断特征构造的优化搜索提供了有利的工具。

2 基于 IP 算法的特征构造

在诊断特征构造函数表达式的优化中,需要对 IP 算法中的抗体集生成、亲和度和相应的抗体变异环节进行定义。

2.1 抗体编码

在基于遗传编程的函数表达式的优化中,典型的编码方式有树型结构和堆栈结构,其中树型结构可以较好地实现结构与大小的动态自适应调整。因此,本文采用树型层次结构来表达免疫编程中的抗体。树型层次结构的节点可分为两类,其中位于树型结构内部的节点称为运算符,在特征提取中,采用基本运算符(加、减、乘、除)来组成运算符集,而位于树型结构末端的叶子节点称为终止符(操作数),它由故障诊断中描述信号行为的特征组成,如峰值、有效值、方差、偏斜度、峭度、峰峰值等。

2.2 抗体亲和度和变异

在 IP 算法中,亲和度是评价抗体个体的惟一标准,其设计原则应有利于抗体的“优胜劣汰”,保证优秀的个体脱颖而出。在特征选择的应用中,类内类间

散布距离判据可以衡量特征的分类能力,因此若某特征空间或复合特征空间中的各类类间距最大,而类内距最小,则该特征或该复合特征就是优化的目标.根据这一思想,将抗体的亲和力定义为

$$f = \frac{\min(D_{ij}^b, i, j = 1, 2, \dots, m)}{\max(D_j^w, j = 1, 2, \dots, m)} \quad (1)$$

式中: D_{ij}^b 表示第 i 类与第 j 类之间的类间距; D_j^w 表示第 j 类的类内距; m 为类别数.

在 IP 算法中,由于抗体为树型结构,因此对变异位进行任意一种的变异操作:①同类元素替换,即终止符元素变异为随机挑选的终止符,运算符元素变异为随机选择的运算符;②子树替代,随机产生一棵新的子树来替代变异位置以下的原有子树.

3 应用实例分析

3.1 原始数据

对发动机异响特征进行构造,实验对象为 CA141 型汽油发动机,本文选择了正常、活塞敲缸、活塞销响、气缸漏气和气门挺杆响 5 种状态^[8],每种状态选择 50 组样本数据.其中,每个异响样本提取 6 个典型的时域特征指标:均方幅值 x_{rms} 、方差 D 、峰值指标 C 、波形指标 K 、峭度 k 和偏斜度 S .图 2~图 5 为 5 种状态样本在上述部分时域特征空间中的分布.

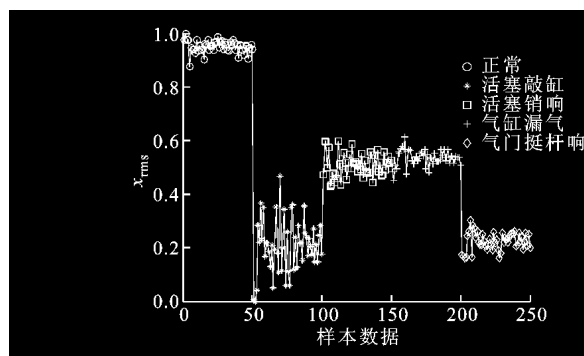


图2 均方幅值特征空间中的数据分布

图中数据分布表明,由于发动机运动部件多、故障复杂,以及异响信号的信噪比较低,因此样本数据的特征值有较大的波动,造成在特征空间中有不同程度的重叠现象.例如,图 2 中活塞销响故障(第 100 点至第 150 点)和气缸漏气故障(第 150 点至第 200 点)的均方幅值指标的变化范围基本相同,而活塞敲缸故障(第 50 点至第 100 点)和气门挺杆响故障(第 200 点至第 250 点)的均方幅值也相互重叠.在 6 个特征空间中, K 中的数据分布相对较好.

3.2 优化结果对比

IP 算法的参数设置为:终止符集合选择原始特征集,即 $\{x_{\text{rms}}, D, C, K, k, S\}$, 抗体集的大小 $n = 120$,其余参数的选择采用文献^[6]中的优化结果,即 $P_r = 0.5, P_c = 0.3, P_m = 0.2, G = 200$. IP 算法获得的最优复合特征指标为

$$F_{\text{best}} = \frac{K}{x_{\text{rms}} + K + k} \quad (2)$$

式(2)充分体现了原始特征中相对较好的特征,即波形指标.图 6 所示为 5 种状态在该特征空间中的数据分布,由图中可见,不同状态的数据分布在不同的空间中,完全实现了数据分离.同各原始特征空间中的数据分布对比表明,在优化的复合特征空间中,各类样本数据的类内距离非常小,表现了良好的聚类特点,说明了该复合特征的有效性.

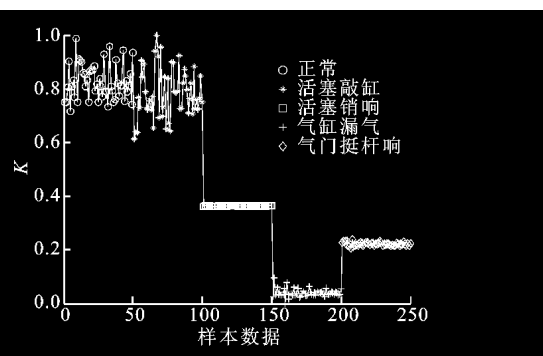


图3 波形指标特征空间中的数据分布

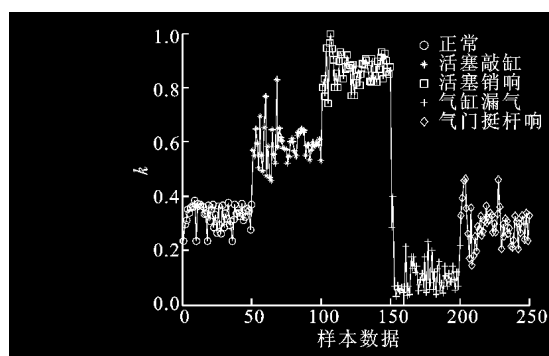


图4 峭度特征空间中的数据分布

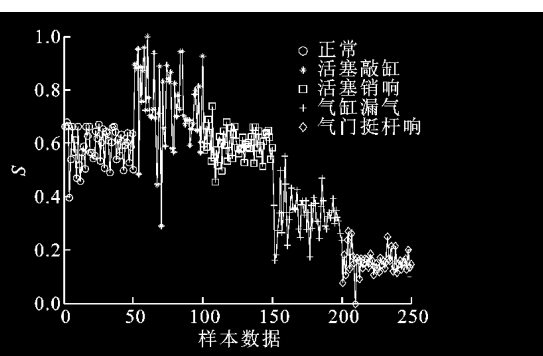


图5 偏斜度特征空间中的数据分布

为验证基于 IP 算法的特征构造方法的有效性,采用 GP 方法^[9]进行对比,算法的参数设置:群体个数为 120,杂交率为 0.9,变异率为 0.1,进化代数为 2 000,经过 20 次独立运行,获得的最优复合特征指标为

$$F_{\text{best}} = x_{\text{ms}} - 2K - S \quad (3)$$

在该复合特征中,同样包含了原始特征中具有较好分类效果的波形指标.图 7 所示为样本数据在该特征空间中的分布,同各原始特征空间中的数据分布对比表明,在该复合特征空间中,5 种不同状态的数据也获得了有效的分离,并且各类状态样本的类内距离也明显得到改善,因此该特征也是一个优化的复合特征.

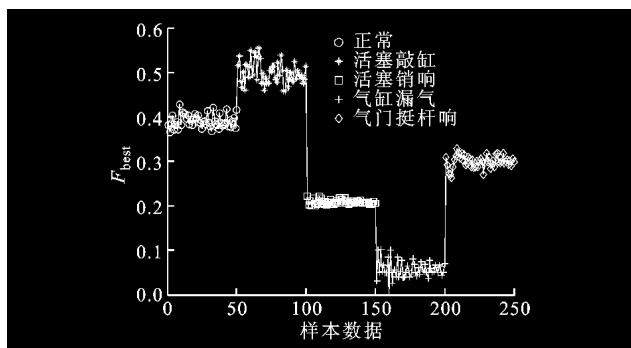


图 6 基于 IP 算法的优化特征空间中的数据分布

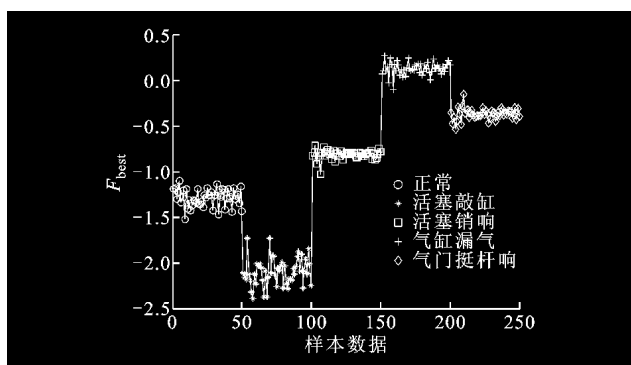


图 7 基于 GP 算法的优化特征空间中的数据分布

对由 IP 和 GP 算法获得的 2 个优化结果进行比较(见表 1),由此可见,在前 3 种状态中,IP 算法获得的复合特征数据分布方差明显小于 GP 算法,而在气门挺杆响故障中,2 个算法的能力却相当,只是在气缸漏气故障中,IP 算法优化特征中的数据方差大于 GP 算法.因此,综合来看,5 种状态在 IP 算法优化特征式(2)中的离散程度相对较小,但 IP 算法优化的复合特征分类性能却优于 GP 算法.造成 GP 算法优化质量不高的原因在于其具有和遗传算法相同的缺点,即早熟问题,因此当进化到一定的局部优化解后,就很难再进一步提高算法的优化质量

了.

表 1 优化特征空间中算法各状态数据的分布对比

算法	方差*				
	正常	活塞敲缸	活塞销响	气缸漏气	气门挺杆响
IP	0.026 5	0.044 2	0.009 2	0.034 5	0.024 5
GP	0.035 1	0.056 5	0.018 4	0.025 8	0.023 2

*:归一化样本.

3.3 IP 算法总结

通过算法的对比实验可知:①在诊断信息的特征构造中,由于免疫克隆选择机制的多样性,IP 算法不仅获得的优化解的质量优于 GP,而且优化解的稳定性也优于 GP 算法;②从优化特征解中可见,通常不包含分类能力较弱的特征,而过多的原始特征将大大增加解空间的搜索难度,降低优化质量,因此在特征构造之前有必要对原始特征进行筛选;③基于 IP 算法优化的新特征并不能保证将原始样本数据相互完全分离,但可以提高原始特征的数据分类能力,因此在故障的诊断识别中,可进一步利用多特征信息来完成数据样本的分离.

4 结 论

基于克隆选择的免疫编程是一种新型的免疫算法,同 GP 算法的思想相同,它可以实现抗体的自动设计,具有较好的优化能力.因此,本文基于免疫编程的优点,将其应用于诊断信息的特征构造中,通过原始特征的组合形成新的特征指标,并根据特征的分类能力采用克隆选择机制进行优化选择,最终获得了分类能力较好的复合特征.通过多个典型故障案例的分析表明,虽然 GP 算法也可获得具有较好分类能力的复合特征,但由于 IP 算法可以有效地维护抗体群的多样性,因此在一定程度上克服了 GP 算法的缺点,从而获得分类能力更优的新特征,进一步改善和提高了特征的分类能力.

参考文献:

- [1] Vafaie H, de Jong K. Feature space transformation using genetic algorithms[J]. IEEE Intelligent System, 1998,13(2):57-65.
- [2] Chen Peng, Nasu M, Toyota T. Self-reorganization of symptom parameters in frequency domain for failure diagnosis by genetic algorithm[J]. Journal of Intelligent and Fuzzy Systems, 1998,6(1):27-37.
- [3] 李良敏,屈梁生. 遗传编程在无量纲指标构建中的应

- 用[J]. 西安交通大学学报, 2005, 36(7): 736-739.
- Li Liangmin, Qu Liangsheng. Genetic programming for dimensionless parameters construction[J]. Journal of Xi'an Jiaotong University, 2005, 36(7): 736-739.
- [4] Durga P, Nikhil R, Jyotirmoy D. Genetic programming for simultaneous feature selection and classifier design[J]. IEEE Trans on System, Man, and Cybernetics, B: Cybernetics, 2006, 36(1): 106-117.
- [5] Sun Ruixiang, Fugee T. Combing bootstrap and genetic programming for feature discovery in diesel engine diagnosis[J]. International Journal of Industrial Engineering, 2004, 11(3): 273-281.
- [6] Peter M, Adriel L, Marek R, et al. Immune programming[J]. Information Sciences, 2006, 176(8): 972-1002.
- [7] Koza J R, Streeter M J, Keane M. Routine high-return human-competitive machine learning[C]//Proceedings of the International Conference on Machine Learning and Applications. Boston: Kluwer Academic Publishers, 2003: 6-12.
- [8] 林京. 基于连续小波技术的机械动态信号处理[D]. 西安: 西安交通大学机械工程学院, 1999.
- [9] 王小平, 曹立阳. 遗传算法——理论、应用与软件实现[M]. 西安: 西安交通大学出版社, 2002: 92-103.

(编辑 管咏梅)

本刊实行在线投稿、免费阅读
<http://www.jdxb.cn>