

DOI: 10.7652/xjtuxb201305021

面向高维特征故障数据的进化软子空间聚类算法

夏虎, 庄健, 于德弘

(西安交通大学机械工程学院, 710049, 西安)

摘要: 针对复杂机械设备故障诊断中特征量众多且对各种故障敏感程度不同的现象,提出了采用软子空间聚类算法来实现故障的识别方法。同时,针对传统软子空间聚类易陷入局部最优,目标函数设计受限的缺点,又提出了采用进化计算实现聚类的方法。利用同类样本在相关特征维上方差小的假设,新的目标函数能更好地评价聚类结果的质量。在该算法中,通过设计类中心和权重值的混合编码以及聚类导向搜索算子,使算法更适于聚类问题的优化,而且设计的修复算子可有效地去除不合理的聚类结果。采用 5 组 UCI 数据集、2 组轴承滚珠故障数据集和 3 组往复式压缩机气阀故障数据集对算法进行了测试,结果表明:该算法明显好于几种的软子空间聚类算法,在 Rand 指标上最多可高出 0.226 6,并且对 2 组不同工况下一级缸气阀故障可实现 100% 的故障识别。

关键词: 故障诊断;软子空间聚类;进化算法;相关特征维

中图分类号: TP391.41 **文献标志码:** A **文章编号:** 0253-987X(2013)05-0115-06

Evolutionary Soft Subspace Clustering Method for Fault Diagnosis of High-Dimensional Features

XIA Hu, ZHUANG Jian, YU Dehong

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

Abstract: Many features can be collected in a complex machine, however, using all these features for fault diagnosis is inappropriate due to different sensitivity of each feature. A soft subspace clustering schedule is proposed. A novel evolutionary algorithm is suggested to avoid the shortages in the traditional soft subspace clustering, such as local optimum and constraint in designing objective function. Under the hypothesis that the samples in the same cluster have small variance in relevant dimensions, a new objective function is constructed in the proposed algorithm to optimize. Considering the special meaning of each gene, a new encoding frame and a cluster-oriented search operator are designed. A repair operator enables to eliminate some meaningless individuals in the evolution process. The experiments on five UCI, two bearing fault and three valve fault of reciprocating compressor datasets demonstrate the better performance of the new algorithm. It is at most 0.226 6 higher than other competitive soft subspace clustering methods on Rand index (RI), especially for two of the valves, fault diagnosis rate gets 100%.

Keywords: fault diagnosis; soft subspace clustering method; evolutionary algorithm; relevant feature dimension

由于多种信号的获取途径和丰富的信号处理技术 可以为机械设备的故障诊断提供大量的特征量,因此

收稿日期: 2012-11-09。 作者简介: 夏虎(1986—),男,博士生;庄健(通信作者),男,副教授。 基金项目: 中央高校基本科研业务费专项资金资助项目(XJJ20100142);广东省战略性新兴产业核心技术攻关项目(2011A091101004);陕西省科技厅工业攻关项目(2011K09-45)。

网络出版时间: 2013-03-15

网络出版地址: <http://www.cnki.net/kcms/detail/61.1069.T.20130315.1137.005.html>

许多研究人员将多传感器信号或多特征量同时用于故障的诊断^[1-3]。这些特征量构成了一个高维数据集,其中多个特征量可能具有相同的故障识别功能,而不同特征量对各种故障的灵敏程度也可能不同,所以该故障数据集包含了不少冗余和无关的特征,并且各个故障通常分布在相关特征量构成的不同子特征空间中。

软子空间聚类技术非常适合于从高维数据中找到相应的类(故障模式),但已有的软子空间聚类技术具有易陷入局部极值的缺点,同时由于需要从目标函数中求解出权重和类中心的显式迭代公式,使得所设计的目标函数的表达形式受到限制。进化计算具有良好的全局寻优能力,并且不要求目标函数的可微性和连续性,因此被许多学者用于数据聚类 and 故障诊断^[4-6]。目前,将进化计算用于软子空间聚类的相关研究工作还很少^[7],并且尚未应用于机械故障模式的识别。为此,本文针对复杂设备故障数据的特点和软子空间聚类的不足,设计了一种面向高维故障特征数据的进化软子空间聚类算法。该算法具有以下特点:①根据样本的相似性,提出在相关特征维上同类样本方差小,而在无关特征维上同类样本方差大的假设,并设计了新的目标函数来评价聚类解的质量;②在进化算法中,采用了类中心和权重的混合编码方式,设计了聚类导向搜索算子和修复算子,使得算法更适于软子空间聚类。此外,通过对多组标准高维数据、球轴承滚珠故障数据和往复压缩机气阀泄漏故障数据的实验,表明本方法的性能优于 k 均值算法和多种软子空间聚类算法。

1 目标函数设计

目标函数是软子空间聚类算法的聚类质量评价指标,同时也是进化算法的优化对象,它决定着聚类结果的质量。软子空间聚类算法的目标函数为

$$J(\mathbf{V}, \mathbf{U}, \mathbf{W}) = J_1(\mathbf{V}, \mathbf{U}, \mathbf{W}) + \sum_i \gamma_i J_A(\mathbf{V}, \mathbf{U}, \mathbf{W}) \quad (1)$$

式中: J_1 为类内所有样本相对于类中心分布的紧致度; J_A 为附加项,表示从其他角度评价聚类质量的指标; γ_i ($\gamma_i \geq 0$) 为权重系数; $\mathbf{V}$ 、 $\mathbf{U}$ 和 $\mathbf{W}$ 分别为聚类中心矩阵、分割矩阵和权重矩阵。

不同的软子空间聚类算法在目标函数的具体内容上有所不同,主要体现在 J_A 的表达形式上,本文设计的目标函数如下

$$J(\mathbf{V}, \mathbf{U}, \mathbf{W}) = \sum_{i=1}^C \sum_{j=1}^N u_{ij} \sum_{k=1}^D w_{ik} (a_{jk} - v_{ik})^2 + \gamma_1 \sum_{i=1}^C \sum_{k=1}^D w_{ik} \lg w_{ik} + \gamma_2 (-A_1/A_2) \quad (2)$$

其中

$$A_1 = \sum_{i=1}^C \sum_{k=1}^D 1/w_{ik} \sigma_{ik}, \quad w_{ik} \leq 1/D$$

$$A_2 = \sum_{i=1}^C \sum_{k=1}^D 1/w_{ik} \sigma_{ik} + \epsilon, \quad w_{ik} > 1/D$$

$$u_{ij} \in \{0, 1\}, \quad \sum_{i=1}^C u_{ij} = 1$$

$$0 \leq w_{ik} \leq 1, \quad \sum_{k=1}^D w_{ik} = 1$$

式中: C 为聚类类数; N 为样本总数; D 为特征维总数; u_{ij} 为第 j 个样本与第 i 个类的从属关系; w_{ik} 为第 k 维特征在第 i 个类中的权重值; v_{ik} 为第 i 个类在第 k 维上的类中心; a_{jk} 为样本 j 在第 k 维的取值; σ_{ik} 为第 i 类样本在第 k 维上的方差; ϵ 为设定的正数,以防止分母为 0。

本文根据子空间聚类的特点认为:相关度紧密的特征维对聚类质量的贡献大,不紧密的特征维则贡献小;同类样本在相关度紧密的特征维上方差小,在不紧密的特征维上方差大。在式(2)中,权重值大于均值 ($1/D$) 的特征维被称为与当前类相关度紧密的特征维。式(2)中的 $\gamma_i (-A_1/A_2)$ 可有效评价选择的特征维以及在该维上的划分是否合理。

2 进化软子空间聚类算法设计

传统的软子空间聚类算法难以通过式(2)推导出 $\mathbf{V}$ 、 $\mathbf{U}$ 和 $\mathbf{W}$ 的表达式,并且易陷入局部最优。因此,本文设计了相应的进化软子空间聚类算法(EASSC)用以优化式(2)。EASSC 采用了类中心和权重混合的编码方式,即

$$x = \{v_{11}, \dots, v_{1D}, v_{21}, \dots, v_{2D}, \dots, w_{11}, \dots, w_{1D}, w_{21}, \dots, w_{2D}, \dots, w_{c1}, \dots, w_{cD}\} \quad (3)$$

该编码方式与样本规模无关,其复杂度不依赖于样本规模的大小,同时无需解码操作即可直接带入目标函数进行计算。

2.1 交叉和变异算子

先采用二元联赛选择的办法随机从当前种群中选择 2 个个体,其中目标函数值小的个体作为父代个体,然后分别执行该操作 2 次,从而得到 2 个父代个体并用于交叉操作。针对编码所代表的具体物理意义,EASSC 对前半段编码(类中心部分)采用单点交叉操作,后半段编码(权重部分)选择对应的交叉点执行交叉操作。该交叉算子可以保证每个类在各个特征维上的类中心与其对应的特征权重值始终遗传到同一个子代个体上。

对于每一个子代个体,以选择概率 p_s 决定是否进

行变异操作。若个体被选中,则类中心部分和权重部分的每个基因均以变异概率 p_m 进行变异,即

$$x_j = \begin{cases} r_d(0,1) & r_d(0,1) \leq p_m \\ x_j & r_d(0,1) > p_m \end{cases} \quad (4)$$

式中: $r_d(0,1)$ 为服从 0 到 1 之间均匀分布的随机数。

2.2 聚类导向的搜索算子

软子空间聚类假设样本的分布存在隐含特性,即类中心为所有属于该类的样本均值。一个类的样本在相关度紧密的维数上分布紧密,方差取值较小,而对应的权重值较大。基于上述特性,EASSC 设计了聚类导向的搜索算子,用以单步更新类中心和权重值,类中心为

$$v_{ik} = \sum_{j=1}^N u_{ij} a_{jk} / \sum_{j=1}^N u_{ij}$$

$$u_{ij} = \begin{cases} 1, & \sum_{k=1}^D w_{ik} (a_{jk} - v_{ik})^2 \leq \sum_{k=1}^D w_{pk} (a_{jk} - v_{pk})^2 \\ 0, & 1 \leq p \leq C \\ & \text{其他} \end{cases} \quad (5)$$

若某一类中不含任何一个样本,则随机从 $[0,1]$ 中产生一个浮点数作为类中心。权重值为

$$w_{ik} = \exp(\sigma_{ik}) / \sum_{l=1}^D \exp(\sigma_{il}) \quad (6)$$

聚类导向的搜索算子通过提高类中心和权重值的精度,起到加速算法收敛的作用。

2.3 修复算子

在聚类过程中,可能会出现一些不满足要求的个体,这些个体的某个或几个类中心向量对应的类并不包含任何样本,实际的类数少于编码中所表示的类数,即按照距离机制划分样本时,有的类中心和所有样本的距离均不是最小值,所以无法分配到样本。为了剔除这些个体以保证算法的性能,EASSC 在每代进化结束的时候检查每个子代个体,对其中存在空类的个体进行修复。修复算子的操作流程为:假设有 s 个空类,从样本数量最多的 1 个类中随机选择 1 个样本作为第 1 个空类的类中心,权重值则在 $[0,1]$ 中随机生成;然后,再随机选择 1 个和已选择样本不同的样本作为第 2 个空类的类中心,权重值也同样随机生成;重复操作直到所有空类均被修复。修复算子使得每个类至少包含 1 个样本。

2.4 算法流程

算法的工作流程为:①随机生成初始化父代种群 Q ,设置当前代数 $t=0$,最大进化代数为 G ;②对当前的 Q 采用二元联赛法选取父代个体,并执行交

叉操作,得到子代种群 Q_1 ;③以概率 P_s 选择子代个体进行变异操作,得到子代种群 Q_2 ;④对 Q_2 中所有子代个体执行聚类导向搜索算子和修复算子,得到子代种群 Q_3 ;⑤计算 Q_3 中所有个体的目标函数值,并对 $Q \cup Q_3$ 的个体进行升序排列,选择最小的 $|Q|$ 个个体组成新的父代种群;⑥若达到最大进化代数 G ,则停止进化。目标函数值最小的个体为最优解,将样本按照类中心和权重值分配到各个类中,即为最终聚类结果。否则设置 $t=t+1$,转至步骤②继续执行。

在本文设计的算法中,所有的聚类样本在各维上的取值均归一化到 $[0,1]$,算法中个体的各个基因的取值范围也为 $[0,1]$ 。在求解目标函数之前,个体中的权重值编码需要进行归一化处理,以保证每个类的各维权重之和为 1。

3 实验设计和分析

本文采用 5 组 UCI 数据集来验证所设计算法的有效性,还通过实际的机械故障数据来验证算法在工程应用中的性能。这些数据包括 2 组轴承滚珠划伤故障数据集和 3 组往复压缩机气阀泄漏故障数据集。同时,选择传统的 k 均值算法和 3 种优秀的软子空间聚类算法(模糊子空间聚类算法(FSC)^[8]、熵权重 k 均值算法(EWKM)^[9] 和强化软子空间聚类算法(ESSC)^[10])与本文算法进行对比分析,每种算法分别测试 10 次,采用 RI 指标评价算法的性能,记录所得 RI 指标的均值 M_R 和方差 V_R 。所有算法的参数设置为:EASSC 的最大迭代代数 $G=100$,种群规模 $P=10$,变异算子的选择概率 $p_s=0.5$,变异概率 $p_m=1/D$, $\epsilon=1.0 \times 10^{-7}$;EWKM 系数 γ 的取值为 1、2、5、10、50、100 和 1 000,迭代终止条件为 $(J_{t-1} - J_t)^2 \leq 10^{-9}$ 。ESSC 的迭代终止条件为 $G=1\ 000$,或 $\|x_{t-1} - x_t\|^2 \leq 10^{-9}$,系数 γ_1 的取值为 1、2、5、10、50、100 和 1 000,系数 γ_2 的取值为 0、0.01、0.05、0.1、0.2、0.3、0.5、0.7 和 0.9。FSC 的迭代终止条件为 $G=1\ 000$,或 $(J_{t-1} - J_t)^2 \leq 10^{-9}$, γ_1 的取值为 0、 10^{-10} 、 10^{-5} 、 10^{-4} 、 10^{-3} 和 10^{-2} ,系数 β 的取值为 1.05、1.2、1.5、2.5、10、20 和 50; k 均值算法的迭代终止条件为 $G=1\ 000$,或 $\|x_{t-1} - x_t\|^2 \leq 10^{-9}$ 。

3.1 UCI 数据测试

用于测试的 UCI 数据集为 Wine、Australian、Glass、Vehicle 和 Ionosphere,测试结果如表 1 所示。表 1 的测试结果表明,采用本文设计的基于进

化搜索的软子空间聚类算法,能够在多个 UCI 数据集上获得比传统软子空间聚类算法更好的性能,其中在 Wine 数据集上 Rand 指标最高,比 FSC 算法高出 0.226 6,因此进化算法比传统算法采用的期望最大化方法更易获得全局最优解。表 1 中 EAEWK 算法的测试结果是采用本文所提方法优化 EWKM 中的目标函数所得到的,它相当于本文所提目标函

表 1 各种算法对 UCI 数据集测试结果的对比

数据集	EASSC		FSC		EWKM		ESSC		EAEWK	
	M_R	V_R	M_R	V_R	M_R	V_R	M_R	V_R	M_R	V_R
Wine	0.950 7	0.008 5	0.724 1	0	0.933 1	0.00	0.945 9	0.000 0	0.933 9	0.000 0
Australian	0.751 8	0.000 0	0.751 8	0	0.712 2	0.00	0.705 2	0.037 2	0.751 8	0.000 0
Glass	0.687 3	0.026 3	0.670 7	0	0.672 7	0.00	0.636 1	0.036 8	0.662 6	0.011 5
Vehicle	0.654 0	0.004 8	0.650 5	0	0.637 8	0.01	0.554 3	0.000 0	0.653 3	0.000 1
Ionosphere	0.644 8	0.041 6	0.623 2	0	0.588 9	0.00	0.588 9	0.000 0	0.584 1	0.000 0

3.2 轴承滚珠故障数据测试

根据美国西储大学轴承数据中心提供的轴承滚珠故障数据,采用文献[2]所示的频域、时域指标提取了 24 个特征量,共得到 2 组数据集(Bdata1、Bdata2)。Bdata1 包含 100 个正常工况的样本,50 个滚珠轻微损伤故障(损伤处直径为 0.178 mm,深度为 0.279 mm)的样本,50 个中度损伤(损伤处直径为 0.356 mm,深度为 0.279 mm)的样本。Bdata2 比 Bdata1 多一类数据样本,即 50 个严重损伤(损伤处直径为 0.533 mm,深度为 0.279 mm)的样本。表 2 为各种算法对滚珠故障的测试结果。

表 2 各种算法对轴承滚珠故障数据的测试结果

数据集	M_R				
	EASSC	FSC	EWKM	ESSC	k 均值算法
Bdata1	1.000 0	0.967 3	0.954 8	0.884 9	0.797 1
Bdata2	0.940 3	0.898 8	0.883 7	0.799 8	0.795 5

表 2 结果显示 EASSC 所得结果优于其他算法,说明进化算法和新的目标函数可有效提高滚珠故障的识别率。k 均值算法所得结果差于其他算法,说明不同的故障模式分布于子空间中的假设与实际故障的分布情况相吻合。Bdata1、Bdata2 是故障源相同但故障程度不同的两组数据集,利用 EASSC 可有效识别此类故障。

3.3 压缩机气阀泄漏故障数据测试

由一套两级往复式压缩机故障实验平台,获得了压缩机故障数据集 Cdata1、Cdata2 和 Cdata3,其中加速度信号的采集频率为 40 kHz,温度和压力信

数中 $\gamma_2 = 0$ 时的情况,因此将该算法称为进行 EWKM 算法(简称 EAEWK)。可见,EAEWK 相对于 EWKM 在性能上有一定的提高,但却在多个数据集上(Wine、Glass、Vehicle 和 Ionosphere)比 EASSC 的性能差。本文所设计的目标函数能够更有效地引导算法,获得相关度紧密的特征维,从而得到更优的分割结果。

号的采集频率为 1 kHz,每 0.6 s 的信号为一个样本源,共提取了 57 个特征量。针对加速度信号提取了 9 个时域指标^[3]、11 个高频(大于 7.5 kHz)频域指标、11 个低频(小于 7.5 kHz)频域指标和 11 个 3 层 dB10 小波重构后的低频频域指标,以及温度信号统计均值和最大最小值的差值、阀腔压力信号统计均值、最大和最小值、储气罐压力信号统计均值和最大最小值的差值。该实验平台的工作过程为:先运行压缩机,对储气罐进行储气;当储气罐压力达到 1 MPa 时,打开阀门,气体带动气动马达转动,磁粉制动器则阻碍马达转动,从而模拟带载运行的情况,此时储气罐压力逐渐下降,最后达到一个稳定值。Cdata1 为空载下储气罐储气时的数据,包含 30 个正常工况的样本,50 个一级缸排气阀阀片严重泄漏(5 mm 通孔)的样本和 50 个一级缸进气阀阀片严重泄漏(5 mm 通孔)的样本。Cdata2 为带载下储气罐开始排气但压力未达到稳定前的数据,包含 30 个正常工况的样本,20 个一级缸排气阀片严重泄漏的样本,20 个一级缸进气阀阀片严重泄漏的样本和 30 个一级缸进气阀阀片轻微泄漏的样本。Cdata3 为带载下储气罐的排气压力达到稳定后的数据,包含与 Cdata2 同样的故障,样本数依次为 40、30、30 和 30。表 3 为各种算法对气阀泄漏故障的测试结果。

通过表 3 可以得到与轴承故障类似的结论,但 ESSC 的结果明显差于其他几种算法,这可能是因为其目标函数的设计不适于压缩机气阀故障数据所造成的。此外,Cdata1、Cdata2 和 Cdata3 是气阀故障在不同工况下的数据集,后 2 种工况的聚类质量

比空载工况下的结果好,说明带载情况下压缩机的故障更易通过各种特征显现出来,从而提高了识别准确率。

表 3 各种算法对压缩机气阀泄漏故障的测试结果

数据集	M_R				
	EASSC	FSC	EWKM	ESSC	k 均值算法
Cdata1	0.889	0.841 2	0.849 9	0.672 7	0.823 1
Cdata2	1.000	0.918 4	0.880 9	0.814 7	0.878 1
Cdata3	1.000	0.959 3	0.935 5	0.803 8	0.871 4

图 1 是 Cdata3 各个类的特征权重分配图,可见各个类对所有特征的敏感程度不同,即使类 3、类 4 为同源故障,但由于故障程度不同,使得它们对一些特征的反应存在着明显的差别,例如特征 14、15、40 和 41。

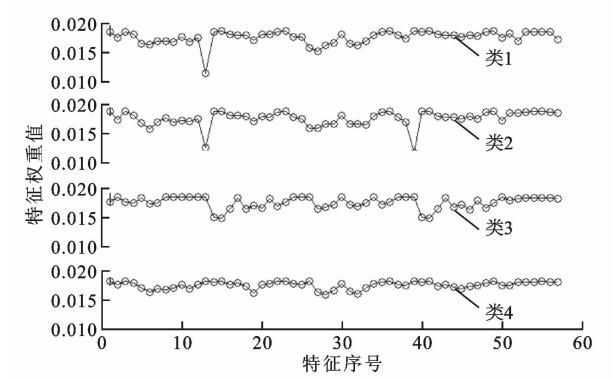


图 1 Cdata3 的 4 个类对 57 个特征量的权重分配图

3.4 EASSC 的参数敏感性

为了研究目标函数中的参数取值对算法结果的影响,本文测试了各种算法对 Cdata2 在不同参数组合下的聚类结果,如图 2 所示。从图 2 中可以发现,本文算法和 ESSC 算法相对其他算法在性能上对参数的依赖较小,而本文算法的性能在各种参数组合下普遍高于其他几种算法。此外,本文算法在最后 9 种参数组合下($\gamma_1=1\ 000$ 时, $\gamma_2=0,0.01,0.05,0.1,0.2,0.3,0.5,0.7,0.9$),其性能有所下降。当参数组合中 $\gamma_2=0$ 或 0.01 时,目标函数的最后一项几乎被忽略,算法所得 Rand 指标约为 0.94 ,比 γ_2 取其他值时所得的结果差。因此,在选取参数时, γ_1 不应过大,可在 $[1,100]$ 中取值; γ_2 不应过小,可在 $[0.05,0.9]$ 中取值。这样可以保证本文提出的目标函数的熵权重和权重比值能够产生合理量级的取值。

3.5 EASSC 的噪声敏感性

本文以 Wine、Bdata1 和 Cdata1 为测试对象,分

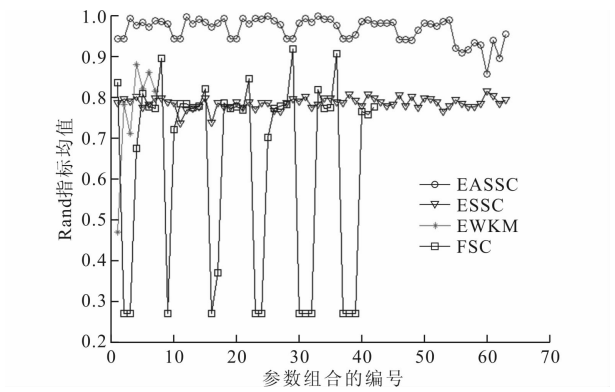


图 2 不同算法的 Rand 指标均值的测试结果

别施加 V_R 为 $0.05,0.1,0.15,0.2,0.25$ 和 0.3 的高斯噪声,然后用 EASSC 对带噪声的数据进行聚类,结果如表 4 所示。当方差为 $0.05,0.15$ 和 0.2 时,以带噪声数据的 Rand 指标下降幅度不超过无噪声时的 10% 为标准,Wine、Bdata1、Cdata1 的 Rand 指标最多下降幅度分别为 $7\%,6\%$ 和 3.6% 。可见,本文算法对数据中包含的噪声不敏感。

表 4 EASSC 对噪声数据的测试结果

V_R	M_R		
	Wine	Bdata1	Cdata1
0.00	0.950 7	1.000 0	0.889 4
0.05	0.950 5	0.985 8	0.881 7
0.10	0.923 2	0.988 9	0.860 0
0.15	0.883 7	0.963 8	0.835 9
0.20	0.826 9	0.941 0	0.810 3
0.25	0.773 9	0.910 9	0.786 1
0.30	0.712 1	0.890 2	0.764 5

4 结 论

针对复杂机械设备中的高维故障特征数据集,本文提出了一种基于进化计算的软子空间聚类算法。该算法不仅在标准 UCI 数据集上取得了比传统软子空间聚类算法、k 均值算法好的聚类结果,而且还成功地实现了对轴承滚珠故障、压缩机气阀泄漏故障的高准确率识别。在目标函数设计方面,本文设计的目标函数着重检验同类样本在相关度紧密的特征维上是否具有正确的划分,取得了比其他目标函数(EASSC 和 EAEWKM)更好的结果。在算子设计方面,本文设计的算子考虑了聚类问题中编码所含的具体物理意义,因此更适于驱动算法搜索最优解。实验结果表明,本文设计的进化方法优于传统软子空间采用的期望最大化求解方法,采用该

进化方法解除了目标函数设计的限制,可使更多合理的目标函数运用于软子空间聚类中。

参考文献:

- [1] 张周锁,侯照文,孙闯,等.应用粒计算的混合智能故障诊断技术研究[J].西安交通大学学报,2011,45(1):48-53.
ZHANG Zhouuo, HOU Zhaowen, SUN Chuang, et al. Hybrid intelligent diagnosis technology based on granular computing [J]. Journal of Xi'an Jiaotong University, 2011, 45(1): 48-53.
- [2] LEI Yaguo, HE Zhengjia, ZI Yanyang, et al. New clustering algorithm-based fault diagnosis using compensation distance evaluation technique [J]. Mechanical Systems and Signal Processing, 2008, 22(2): 419-435.
- [3] LEI Yaguo, HE Zhengjia, ZI Yanyang, et al. Fault diagnosis of rotating machinery based on multiple AN-FIS combination with GAs [J]. Mechanical Systems and Signal Processing, 2007, 21(5): 2280-2294.
- [4] 张云,冯博琴,麻首强,等.蚁群-遗传融合的文本聚类算法[J].西安交通大学学报,2007,41(10):1146-1150.
ZHANG Yun, FENG Boqin, MA Shouqiang, et al. Text clustering based on fusion of ant colony and genetic algorithms [J]. Journal of Xi'an Jiaotong University, 2007, 41(10): 1146-1150.
- [5] 董红斌,杨宝迪,刘佳媛,等.协同演化算法在聚类中的应用[J].模式识别与人工智能,2012,25(4):676-683.
DONG Hongbin, YANG Baodi, LIU Jiayuan, et al. A co-evolutionary algorithm for clustering [J]. Pattern Recognition and Artificial Intelligence, 2012, 25(4): 676-683.
- [6] WU Qi, LAW R, WU Shuyan. Fault diagnosis of car assembly line based on fuzzy wavelet kernel support vector classifier machine and modified genetic algorithm [J]. Expert Systems with Applications, 2011, 38(8): 9096-9014.

- [7] ZHU Lin, CAO Longbing, YANG Jie. Multiobjective evolutionary algorithm-based soft subspace clustering [C]//Proceedings of the 2012 IEEE Congress on Evolutionary Computation. Piscataway, NJ, USA: IEEE, 2012:1-8.
- [8] GAN Gaojun, WU Jianhong, YANG Zijiang. A fuzzy subspace algorithm for clustering high dimensional data [C]//Proceedings of the 2nd International Conference on Advanced Data Mining and Applications. Berlin, Germany: Springer-Verlag, 2006: 271-278.
- [9] JING Liping, NG M K, HUANG J Z. An entropy weighting k -means algorithm for subspace clustering of high-dimensional sparse data [J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(8): 1026-1041.
- [10] DENG Zhaohong, CHOI K S, CHUNG F L, et al. Enhanced soft subspace clustering integrating within-cluster and between-cluster information [J]. Pattern Recognition, 2012, 43(3): 767-781.

[本刊相关文献链接]

- 王跃飞,张振涛,张波,等.利用声发射的往复空压机环状阀泄漏故障诊断试验.2012,46(9):26-30.
- 陈景龙,訾艳阳,何正嘉,等.自适应冗余多小波及其在故障诊断中的应用.2012,46(7):44-49.
- 徐宇亮,孙际哲,陈西宏,等.采用多目标粒子群算法的模拟电路故障诊断研究.2012,46(6):92-97.
- 栗茂林,梁霖,王孙安,等.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.2012,46(5):50-56.
- 薛士明,曹军义,林京,等.分数阶阻尼裂纹转子的非线性动力学特性分析.2012,46(1):76-80.
- 杨攀,桂小林,田丰,等.一种高效的用于话题检测的关键词元聚类方法.2012,46(10):24-28.
- 邵必林,边根庆,张维琪,等.采用 k -均值聚类算法的资源搜索模型研究.2012,46(10):55-59.
- 杨清宇,孙凤伟,张墨,等.利用测地线距离的改进谱聚类算法.2012,46(8):1-7.

(编辑 管咏梅)