基于差分进化算法的模糊核聚类算法及其在故障诊断中的应用

张新萍[1]，张孝远[2]，刘 杰[3]

(1.许继集团有限公司，河南 许昌 461000；2.河南工业大学电气工程学院，河南 郑州 450001；
3.河南职业技术学院，河南 郑州 450000)

摘要：针对模糊核聚类方法中，核函数参数的确定问题以及聚类结果的有效评价问题，提出采用差分进化算法进行核函数参数和聚类中心的同时寻优策略。并将 Xie-Beni 指标推广至核空间，设计了有效的适应度函数以实现聚类效果的提升。对所提出的方法进行数值试验，同时应用在电机轴承的故障诊断中，取得了不错的效果，验证了方法的可行性。

关键词：模糊聚类；核函数；差分进化算法；故障诊断

Fuzzy kernel-clustering algorithm based on differential evolution algorithm and its application in fault diagnosis

ZHANG Xin-ping[1], ZHANG Xiao-yuan[2], LIU Jie[3]

(1. XJ Group Corporation, Xuchang 461000, China; 2. College of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China; 3. Henan Polytechnic, Zhengzhou 450000, China)

Abstract: In allusion to the determination of the kernel parameters and the effective evaluation of the clustering results of Fuzzy Kernel-clustering Algorithm (FKCA), differential evolution algorithm (EA) is used to search the optimal kernel parameter and the clustering centers. Furthermore, the Xie-Beni index is promoted to the kernel space, and a new fitness function is designed to improve the clustering performance. The proposed method is applied in the standard benchmark as well as the motor bearing fault dataset. The results shows that the proposed method is a promising clustering method for fault diagnosis.

This work is supported by High-level Personnel Funds of Henan University of Technology (No. 2013BS059).

Key words: fuzzy clustering; kernel function; differential evolution algorithm; fault diagnosis

中图分类号： TM74 文献标识码：A 文章编号： 1674-3415(2014)17-0102-05

0 引言

滚动轴承是电机系统中连接动静部分的一个媒介，其运行状态直接关系到整个电机系统及其互联生产系统的安全稳定运行。据报道，约有 50%以上的电机故障是与其轴承有关的[1]。因此，对电机轴承开展状态监测与故障诊断研究具有十分重要的现实意义与工程应用前景。采用现代传感技术从轴承获取运行中的振动信号，进而基于振动信号开展故障诊断是该领域的一个研究热点。基于振动信号分析的故障诊断本质上是一类模式识别问题[2-4]。当前的一些诊断方法多采用人工神经网络、支持向量机等有监督的机器学习方法进行故障诊断。这些方法在处理未知新型故障时将束手无策。基于聚类分析的无监督学习方法能够根据故障样本的属性相似性自动聚为不同的类别，因而在故障诊断领域有着广阔的应用前景。

聚类分析是一种无监督的机器学习方法，其在不知样本类别的情况下，通过类内样本具有较高相似度，类间样本具有较大差别的特性将数据样本划分为多个类或簇。当前，聚类分析技术正在蓬勃发展，在诸多领域得到广泛应用，如数据分析、模式识别、图像处理、信息检索、故障诊断、预测分析等[5-7]。有关聚类分析的方法很多，模糊 C 均值聚类算法（Fuzzy C-means, FCM）是其中的一种经典有效的聚类算法。然而经典的 FCM 算法在处理非线性聚类问题时遇到困难，因此学者们将核函数引入到 FCM 中，提出模糊核聚类算法[8-10]。模糊核聚类

基金项目：河南工业大学高层次人才基金项目（2013BS059）

算法采用满足 Mercer 条件[10]的核函数将原始空间的聚类问题映射到高维特征空间进行聚类。模糊核聚类算法凸显了样本间的非线性特征差异，从而能够取得更好的聚类性能。然而，模糊核聚类算法的应用存在两个无法回避的问题需要很好地解决。即核函数参数的选择以及聚类性能的有效评价[11]。

对于模糊核聚类算法，一般选择高斯核函数，即式（1）。

$$K(x,y)=\exp(-\frac{\|x-y\|^2}{2\sigma^2}) \qquad (1)$$

在高斯核函数中核函数宽度σ需要确定。有关σ值的确定文献中多采用试错法或专家经验进行确定，带有较强的随机性。因此，有必要针对σ值的确定开展专门的研究。另一方面，由于模糊核聚类算法属于一类无监督机器学习方法，自然地，其聚类结果的合理性有待有效的评价。

针对模糊核聚类方法的上述两个问题，本文提出一种基于差分进化算法（Differential Evolution Algorithm, DE）的模糊核聚类算法。该方法将Xie-Beni指标[12]推广到核空间，以核Xie-Beni指标为聚类目标函数，同时将聚类中心和核函数参数作为优化变量进行聚类。用所提出的方法对国际标准测试数据集进行数值试验，并应用于电机轴承故障诊断的工程实际验证了方法的可行性。

1　模糊核聚类算法

1.1 模糊C均值聚类

设有N个未知标号的样本$(x_1,x_2,\cdots,x_N)$，聚为K个互无交集的模糊类，假设第k类的样本数目为N_k，$N=\sum_{k=1}^{K}N_k$，算法通过最小化如下的聚类目标函数来自适应地变化选取新的聚类中心。

$$J_{\mathrm{MSE}}=\sum_{i=1}^{K}\sum_{j=1}^{N}(\mu_{ij})^m d(x_j,v_i) \qquad (2)$$

式中：v_i为第i个聚类中心，i=1,2,⋯,K；μ_{ij}为模糊隶属度；$\sum_{i=1}^{K}\mu_{ij}=1$，$\mu_{ij}\in(0,1)$，$\forall j, d(x_j,v_i)=\|x_j-v_i\|^2$。模糊C均值算法在目标函数中增加了模糊权重指数$m$，$1<m<+\infty$。聚类中心和隶属度按照式（3）、式（4）进行更新。

$$v_i=\frac{\sum_{j=1}^{N}\mu_{ij}^m x_j}{\sum_{j=1}^{N}\mu_{ij}^m},i=1,2,\cdots,K \qquad (3)$$

$$\mu_{ij}=\frac{1}{\sum_{k=1}^{K}(\frac{d_{ij}}{d_{kj}})^{1/(m-1)}},i=1,2,\cdots,K;j=1,2,\cdots,N \qquad (4)$$

FCM 算法计算简单，具有较为直观的几何意义，因此在众多聚类算法中应用最为广泛，也较为成功。然而，当输入数据为线性不可分、类分布为非高斯或非球形分布时，其聚类效果较差。引入核函数的技巧能够解决这一问题。

1.2 模糊核聚类

模糊核聚类算法的基本思想是利用非线性映射函数$\varphi(x)$将原始输入空间映射到高维线性近似可分空间再进行聚类。因此，上述式（2）变为如式（5）的形式。

$$J_{\mathrm{MSE}}=\sum_{i=1}^{K}\sum_{j=1}^{N}(\mu_{ij})^m d(\varphi(x_j),\varphi(v_i)) \qquad (5)$$

这种非线性映射$\varphi(x)$本身比较复杂，因此在实际应用中利用满足 Mercer 条件的核函数隐式的计算，则式（5）中的

$$d(\varphi(x_j),\varphi(v_i))=\|\varphi(x_j)-\varphi(v_i)\|^2=\varphi(x_j).\varphi(x_j)-2\varphi(x_j).\varphi(v_i)+\varphi(v_i).\varphi(v_i)=K(x_j,x_j)-2K(x_j,v_i)+K(v_i,v_i) \qquad (6)$$

式中：$K(\cdot,\cdot)$为核函数；$K(x,y)=\varphi(x).\varphi(y)$。考虑到高斯核函数中$K(x,x)=1$则上式变为

$$d(\varphi(x_j),\varphi(v_i))=\|\varphi(x_j)-\varphi(v_i)\|^2=2(1-K(x_j,v_i)) \qquad (7)$$

1.3 核空间的Xie-Beni指标

Xie-Beni指标[12]兼顾了簇内的紧致性和簇间的分离性，其定义如式（8）。

$$V_{\mathrm{XB}}=\frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^m\|x_j-v_i\|^2}{N}\cdot\frac{1}{\min_{j\neq k}\|v_j-v_k\|^2}=\frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^m\|x_j-v_i\|^2}{N(\min_{j\neq k}\|v_j-v_k\|^2)} \qquad (8)$$

其中：$\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^m\|x_j-v_i\|\Big/N$评价了簇内样本的紧致性；$1\Big/\min_{j\neq k}\|v_j-v_k\|^2$评价了簇间样本的分离性，因此，Xie-Beni指标V_{XB}取极小值时对应了簇内紧致，簇间分离的最优的聚类结果。

核Xie-beni指标是指将核函数引入到Xie-beni指标，核函数采用一种隐性的映射 $\phi(\bullet)$ 将数据从低维线性不可分空间映射到高维线性可分空间。从而，式（8）的V_{XB}变为式（9）的形式。

$$V_{\mathrm{KXB}}=\frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{m}\left\|\phi(x_j)-\phi(v_i)\right\|^2}{N(\min\limits_{j\neq k}\left\|\phi(v_j)-\phi(v_k)\right\|^2)} \tag{9}$$

考虑到式（7）的简化形式，则式（9）变为

$$V_{\mathrm{KXB}}=\frac{\sum_{i=1}^{K}\sum_{j=1}^{N}\mu_{ij}^{m}(1-K(x_j,v_i))}{N(\min\limits_{j\neq k}(1-K(x_j,v_i))} \tag{10}$$

2 基于差分进化算法的模糊核聚类算法

2.1 差分进化算法

差分进化算法是一种并行的直接实值最优化搜索算法，最初由 Storn 和 Price[13]于 1995 年提出。和其他进化算法一样，差分进化算法基于种群和随机化搜索。假定种群规模为 NP，维数为 D，x_i^t 代表第 t 代种群的第 i 个向量，差分进化算法通过差分变异、交叉、选择算子实现个体（向量）的进化。

1）变异。差分进化算法的变异策略比较多，常用的是“DE/rand/1”方法，可描述为

$$V_i^t=X_{r_1^i}^t+F\cdot(X_{r_2^i}^t-X_{r_3^i}^t) \tag{11}$$

其中：$V_i^t=\{v_{i1}^t,v_{i2}^t,\cdots,v_{iD}^t\}$ 为突变向量；r_1^i,r_2^i,r_3^i 为互不相同并且和 i 不同的随机分布在[1, NP]区间中的整数；F 是收缩因子。

2）交叉。差分进化算法采用一种十分简单的交叉算子，如式（12）。

$$u_{ij}^t=\begin{cases}v_{ij}^t & \text{if}(\mathrm{rand}_j[0,1]\le CR)\text{or}(j=j_{\mathrm{rand}})\\ x_{ij}^t & \text{otherwise}\end{cases} \tag{12}$$

其中：$i=1,2,\cdots,NP$，$j=1,2,\cdots,D$。$U_i^t=\{u_{ij}^t,u_{ij}^t,\cdots,u_{iD}^t\}$ 被称为目标向量，CR 为交叉概率，其范围一般为[0,1]，其控制着目标向量由变异向量直接拷贝的概率。j_{rand} 是一个随机分布在[1,D]中的整数。

3）选择。差分进化算法的选择算子同样十分简单。采用“贪婪”策略，经过变异和交叉操作后的试验个体与目标个体进行竞争，只有当前者的适应度值不大于后者，才替换目标个体进入下一代。这一操作按式（13）进行。

$$X_i^{t+1}=\begin{cases}U_i^t & \text{if } f(U_i^t)\le f(X_i^t)\\ X_i^t & \text{otherwise}\end{cases} \tag{13}$$

2.2 基于差分进化算法的模糊核聚类

1）编码。在差分进化聚类算法中，种群中的一个个体表示待求解的聚类中心和高斯核宽度 σ 的集合，即

$$z_i=(c_{i1},\ldots,c_{ij},\ldots,c_{ik},\ \sigma) \tag{14}$$

式中：k 表示聚类中心个数；c_{ij} 表示第 i 个个体的第 j 个聚类中心。

2）适应度函数设计。以式（10）所示的核空间的 Xie-Beni 指标为目标函数，以聚类中心和核参数为优化变量，通过 DE 搜索式（10）的函数极值，则适应度函数设计为

$$F=\frac{r_0}{eps+V_{\mathrm{KXB}}} \tag{15}$$

式中，eps 是一个微小的正数。

3）算法流程

Step1：设置模糊指数 m，最大迭代次数 T，最大聚类数 $K_{\max}$，并令 K=2；初始化 DE 群体，粒子位置向量表示的聚类中心和核参数 σ 等。

Step2：通过粒子位置向量解码聚类中心向量和 σ，计算核矩阵和模糊矩阵，由式（10）计算 V_{KXB}。

Step3：依据式（15）计算每个个体的适应度值。

Step4：依据式（11）~式（13）进行变异、交叉和选择操作。群体进入下一代。

Step5：重复执行 Step2~Step4 直到满足终止条件。

Step6：计算 $V_{\mathrm{KXB}}(k)$，若 $k<K_{\max}$，则 $k\leftarrow k+1$，转入执行 Step2~Step4；否则，结束算法，并选择对应的 k 为最佳聚类数，相应的聚类结果为最佳聚类结果。

3 故障诊断模型与数值仿真

采用所提方法进行故障诊断时，首先利用已知样本进行上述算法的聚类分析以获取标准故障模式（F_1，F_2，…，F_K）以及核参数 σ_F。则未知故障样本 S 可通过下式来判定其类别归属。

$$F_t=\min\{2(1-\exp(-\frac{(F_i-S)}{2\sigma_F^2}))\},i=1,2,\cdots,K \tag{16}$$

为评估诊断模型的有效性，对本算法和标准 FCM 算法进行了比较。实验所用数据集来源于 uci 数据集[14]，选用了其中的 iris，wine 和 Breast Cancer。对这三个数据集分别剔除其类别标签，将数据随机分为训练集合测试集，对这个三个数据集的描述如

表 1 所示。

表 1 数据集描述

Table 1 Description of data sets

数据集	样本数	类别数	属性数	训练样本	测试样本
Iris	150	3	4	100	50
Wine	178	3	13	102	76
Breast Cancer	683	9	2	444	239

对于所提方法参数设置如下：DE 的种群大小：NP=30，算法的最大迭代次数 T=200，交叉概率 CR=0.9，收缩因子 F=0.5，模糊度 m=2。FCM 算法的迭代停止阈值设置为1×10^{-6}，模糊度 m=2，聚类数为 3。两种方法对训练样本和测试样本得到的聚类精度如表 2 所示。

表 2 两种算法的比较

Table 2 Comparison of two methods

算法	训练精度/%		测试精度/%	
	FCM	DEFCM	FCM	DEFCM
Iris	91.0	100	96.0	98.0
Wine	76.5	78.4	78.9	92.1
Breast Cancer	90.1	92.3	96.2	97.9

由表 2 可见，与标准 FCM 相比，所提方法可以根据训练样本建立更加有效的聚类模型，训练精度和测试精度均优于 FCM。

4　应用于电机轴承故障诊断

滚动轴承作为电机最重要的部件，其运行正常与否对整个电机系统有着重要的影响。本文将上述诊断模型应用于西储大学（Case Western Reserve University，CWRU）轴承数据中心提供的滚动轴承试验数据中[15]。该试验数据获取至一个 6205-2RS JEM SKF 深沟球轴承，在测试轴承上采用电火花加工技术布置单点故障，故障直径为 0.533 4 mm，深度为 0.279 4 mm。一共加工了常见的三种故障：内圈故障（Inner race，IR）、外圈故障（Outer race，OR）和滚动体故障（Ball，B），采集到的数据除了这三种故障外还有正常样本。因此，我们取上述 4 种运行状态的振动信号共 30 组，数据长度为 2 400，采样频率为 12 kHz。图 1 示出此四种状态振动信号的时域波形。

对上述振动信号提取两个时域特征式式（17）、式（18）和两个频域特征式（19）、式（20）。

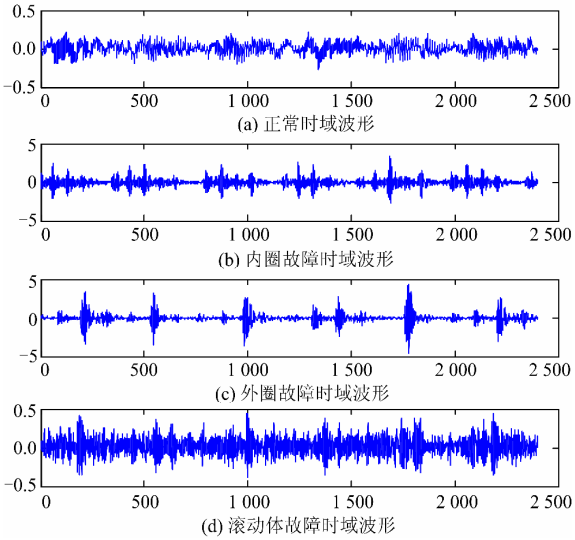

图 1 正常和不同故障的轴承振动信号时域波形

Fig. 1 Time-domain waveform of bearing vibration under normal state and different fault

$$T_1=\frac{\sum_{n=1}^{N}x(n)}{N} \tag{17}$$

$$T_2=\sqrt{\frac{1}{N-1}\sum_{n=1}^{N}(x(n)-T_1)^2} \tag{18}$$

式中，$x(n)$是一个信号，$n=1,2,\cdots,N$，N 是采样点数。

$$F_1=\frac{\sum_{k=1}^{K}s(k)}{K} \tag{19}$$

$$F_2=\frac{\sum_{k=1}^{K}(s(k)-F1)^2}{K-1} \tag{20}$$

此处 $s(k)$ 某信号的频域变换，$k=1,2,\cdots,K$。

经过对每种运行状态的振动信号的特征提取，我们可以得到一个 30×4 的矩阵，即对每一种运行状态得到 30 个样本，四种状态共得到 120 个样本。从每个样本中随机抽取 20 个作为训练样本，其余 10 个样本作为测试样本。用前述诊断模型进行诊断发现诊断率为 100%，验证了所提方法的可行性。

5　结论

本文提出了一种基于差分进化算法的模糊核聚类算法，采用差分进化算法同时对核函数参数及聚类中心进行寻优，将 Xie-Beni 指标推广至核函数空间，并基于核 Xie-Beni 指标设计了适应度函数。然后，基于新的聚类方法构建了一种滚动轴承故障诊断模型，将该模型应用在国际标准数据集和滚动

轴承故障诊断上取得了较好的结果，为聚类方法提供了一种新的思路。

参考文献

[1] NANDI S, TOLIYAT H A, LI Xiao-dong. Condition monitoring and fault diagnosis of electrical motors-a review[J]. IEEE Trans on Energy Conversion, 2005, 20(4): 719-729.

[2] 肖剑, 周建中, 李超顺, 等. 基于混合蜂群算法特征参数同步优化支持向量机的水电机组轴心轨迹识别方法研究[J]. 电力系统保护与控制, 2013, 41(21): 26-32.
XIAO Jian, ZHOU Jian-zhong, LI Chao-shun, et al. Identification of shaft orbit of hydropower unit by simultaneous optimization of feature parameters and support vector machine based on hybrid artificial bee colony[J]. Power System Protection and Control, 2013, 41(21): 26-32.

[3] 郭艳平, 颜文俊, 包哲静, 等. 基于经验模态分解和散度指标的风力发电机滚动轴承故障诊断方法[J]. 电力系统保护与控制, 2012, 40(17): 83-87.
GUO Yan-ping, YAN Wen-jun, BAO Zhe-jing, et al. Fault diagnosis of bearing in wind turbine based onempirical mode decomposition and divergence index[J]. Power System Protection and Control, 2012, 40(17): 83-87.

[4] 杨伟, 顾明星, 彭静萍. 证据理论在电机故障诊断中的应用[J]. 电力系统保护与控制, 2010, 38(2): 64-67.
YANG Wei, GU Ming-xing, PENG Jing-ping. Application of evidence theory in fault diagnosis for electric machine[J]. Power System Protection and Control, 2010, 38(2): 64-67.

[5] 苏海锋, 张建华, 梁志瑞, 等. 基于改进均值聚类随机粒子群算法的变电站 LCC 规划[J]. 电工技术学报, 2012, 27(4): 209-215.
SU Hai-feng, ZHANG Jian-hua, LIANG Zhi-rui, et al. Substation LCC planning based on refined mean clustering random particle swarm algorithm[J]. Transactions of China Electrotechnical Society, 2012, 27(4): 209-215.

[6] 王贺, 胡志坚, 张翌晖, 等. 基于聚类经验模态分解和最小二乘支持向量机的短期风速组合预测[J]. 电工技术学报, 2014, 29(4): 237-245.
WANG He, HU Zhi-jian, ZHANG Yi-hui, et al. A hybrid model for short-term wind speed forecasting based on ensemble empirical mode decomposition and least squares support vector machines[J]. Transactions of China Electrotechnical Society, 2014, 29(4): 237-245.

[7] 周国亮, 宋亚奇, 王桂兰, 等. 状态监测大数据存储及聚类划分研究[J]. 电工技术学报, 2013, 28(增刊 2): 337-344.
ZHOU Guo-liang, SONG Ya-qi, WANG Gui-lan, et al. Research of condition monitoring big data storage and clustering[J]. Transactions of China Electrotechnical Society, 2013, 28(S2): 337-344.

[8] KAMEL M S, SELIM S Z. New algorithms for solving the fuzzy clustering problem[J]. Pattern Recognition, 1994, 27(3): 421-428.

[9] 李侃, 刘玉树. 模糊核聚类的自适应算法[J]. 控制与决策, 2004, 19(5): 595-597.
LI Kan, LIU Yu-shu. Fuzzy kernel clustering self-adaptive algorithm[J]. Control and Decision, 2004, 19(5): 595-597.

[10] SCHOLKOPF B, MIKA S, BURGES C J C, et al. Input space versus feature space in kernel-based methods[J]. IEEE Trans on Neural Networks, 1999, 10(5): 1000-1017.

[11] 李超顺, 周建中, 肖剑, 等. 基于引力搜索核聚类算法的水电机组振动故障诊断[J]. 中国电机工程学报, 2013, 33(2): 98-104.
LI Chao-shun, ZHOU Jian-zhong, XIAO Jian, et al. Vibration fault diagnosis of hydroelectric generating unit using gravitational search based kernel clustering method[J]. Proceedings of the CSEE, 2013, 33(2): 98-104.

[12] XIE Xuan-li, BENI G. A validity measure for fuzzy clustering[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 1991, 13(8): 841-847.

[13] STORN R, PRICE K. Differential evolution-a simple and efficient heuristic for global optimization over continuous space[J]. Journal of Global Optimization, 1997, 11(4): 341-359.

[14] ASUNCION A, NEWMAN D. UCI machine learning repository[M].

[15] 郑近德, 程军圣, 杨宇. 多尺度排列熵及其在滚动轴承故障诊断中的应用[J]. 中国机械工程, 2013, 24(19): 2641-2646.
ZHENG Jin-de, CHENG Jun-sheng, YANG Yu. Multi-scale permutation entropy and its applications to rolling bearing fault diagnosis[J]. Chinese Journal of Mechanical Engineering, 2013, 24(19): 2641-2646.

收稿日期：2013-07-31； 修回日期：2013-10-25

作者简介：

张新萍(1962-)，男，高级工程师，主要研究方向为电力电子、继电保护控制及电力系统一次设备；

张孝远(1981-)，男，讲师，通讯作者，主要从事发电机组故障诊断、控制方面的研究工作。E-mail: xyz_1981@tom.com