

# 基于最小最大核 K 均值聚类算法的水电机组 振动故障诊断

张孝远<sup>1</sup>, 张新萍<sup>2</sup>, 苏保平<sup>1</sup>

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

**摘要:** 基于聚类分析的故障诊断方法能够按照故障样本之间的相似性无监督地将同类故障聚为一簇, 当前已成为一类有效的故障诊断策略。为解决传统聚类算法受初始聚类中心的影响, 易陷入局部最优解的问题, 提出一种最小最大核 K 均值聚类方法。该方法在聚类过程中为簇内方差赋以与其大小成正比的自动修正的权重, 并引入核函数技术以处理低维输入空间的线性不可分问题, 大大提高了聚类的精确性。在标准数据上将所提方法与标准 K-means 及 K-means++ 比较, 显示了所提算法的有效性和优越性。基于这一聚类方法提出了一种具有自学习能力的故障诊断模型。将该诊断模型应用于水电机组振动故障诊断, 实例验证了模型的可行性。

**关键词:** 水电机组; 振动; 故障诊断; 最小最大 K 均值聚类; 核函数

## Vibrant fault diagnosis for hydro-turbine generating unit using minmax kernel K-means clustering algorithm

ZHANG Xiaoyuan<sup>1</sup>, ZHANG Xinping<sup>2</sup>, SU Baoping<sup>1</sup>

(1. College of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China;

2. XJ Group Corporation, Xuchang 461000, China)

**Abstract:** Fault diagnosis methods based on clustering analysis can cluster the fault samples into a certain class according to their similarities without supervision, and thus become one type of effective fault diagnosis strategy. To overcome the problem that traditional clustering methods are susceptible to the initial clustering centers, and thus poor local optima is easily obtained, a MinMax kernel K-means clustering algorithm is introduced. In the proposed method, clusters are assigned weights relative to their variances. And kernel trick is introduced to deal with linear inseparable problem in input space. The proposed method is compared with the traditional K-means and K-means++ in some international standard datasets. The comparison results show its effectiveness and advantage. Then, a fault diagnostic model based on MinMax kernel K-means clustering algorithm is proposed. At last, the fault diagnosis model is applied in fault diagnosis for hydro-turbine generating unit. The results illustrate the effectiveness of the proposed method.

This work is supported by National Natural Science Foundation of China (No. 51409095).

**Key words:** hydro-turbine generating unit; vibration; fault diagnosis; minmax K-means clustering algorithm; kernel function

中图分类号: TM 71 文献标识码: A 文章编号: 1674-3415(2015)05-0027-08

## 0 引言

水电机组是水电能源生产过程中的核心设备, 水电机组的运行健康状况不仅关系到水电厂自身的安全还直接关系到其互联电网的稳定性<sup>[1-2]</sup>。因此,

对水电机组开展故障诊断研究具有十分重要的现实意义与工程应用前景。

当水电机组发生故障时, 从相应部位采集到的振动信号的内部特征也将发生变化。据报道, 水电机组约有 80% 的故障或事故都在振动信号中有所反映<sup>[3]</sup>。通过采集并分析机组振动信号一直都是水电机组故障诊断领域的热点研究问题<sup>[4-7]</sup>。这类故障诊断方法可以被统称为基于信号处理的故障诊断方

**基金项目:** 国家自然科学基金项目 (51409095); 河南工业大学高层次人才基金项目 (2013BS059)

法,其实质是一类模式识别问题。诊断方法的成功与否的关键问题之一在于探寻一种有效的故障推理识别分类模型。

不少学者将在机器学习领域应用较为成熟的人工神经网络<sup>[8-11]</sup>和支持向量机<sup>[6,12-14]</sup>等方法引入到水电机组振动故障诊断领域,取得了不错的诊断效果。然而这类方法属于一种有监督的学习方法,在诊断的学习阶段需要事先给定样本的类别标签。而水电机组在运行中故障样本获取较为困难且故障机理十分复杂,另外随着机组单机容量的增加,运行环境的劣化以及机组材料的不断更新,难免会使水电机组不断出现新的故障。显然,有监督的诊断模型在处理这类问题时存在着一定的局限性。针对这一问题,学者们将无监督的聚类算法应用在水电机组振动故障诊断中来<sup>[15-17]</sup>。无监督的聚类算法能够依据样本自身的属性自动地将属性相似的样本聚类,在诊断的学习阶段无需事先知道样本的类别标签,且可以将新发生的故障根据其属性重新聚类。因此,从诊断方法的实用性角度来看更符合工程实际。

K 均值是一种最基本的聚类方法,然而其最大的局限在于聚类结果很大程度上取决于初始聚类中心的选取。一些基于 K 均值聚类方法的改进方法大多没有真正意义上解决其局部最优解问题。为此,本文介绍一种改进的 K 均值聚类方法——最小最大 K 均值聚类方法。在聚类过程中为每个簇内方差赋予与之大小成正比的自动修正的权重,并引入核函数技术将低维输入空间映射到高维线性可分性较强的高维空间,大大提高了聚类效果。用国际标准数据集对该方法进行测试,并将其应用在水电机组振动故障诊断中,验证了该方法的有效性。

## 1 最小最大 K 核均值聚类算法

### 1.1 K 均值聚类

将一个数据集  $\mathcal{X}=\{\mathbf{x}_i\}_{i=1}^N, \mathbf{x}_i \in \mathbf{R}^d$  无监督地划分到  $M$  个互无交集的簇  $\{\mathbf{c}_k\}_{k=1}^M$  中, K 均值算法通过最小化式(1)的各个簇的簇内方差之和并不断地更新聚类中心直到满足终止条件来实现聚类。

$$\mathcal{E}_{\text{sum}} = \sum_{k=1}^M v_k = \sum_{k=1}^M \sum_{i=1}^N \delta_{ik} \|\mathbf{x}_i - \mathbf{m}_k\|^2 \quad (1)$$

$$\begin{cases} \delta_{ik} = 1 & \text{if } \mathbf{x}_i \in \mathbf{c}_k \\ \delta_{ik} = 0 & \text{otherwise} \end{cases} \quad (2)$$

本文将簇内方差定义为  $v_k = \sum_{i=1}^N \delta_{ik} \|\mathbf{x}_i - \mathbf{m}_k\|^2$ ,

其与一般意义上的方差  $v'_k$  之间的关系为  $v_k = N v'_k$ ,  $N$

为簇内样本个数。 $\mathbf{m}_k = (\sum_{i=1}^N \delta_{ik} \mathbf{x}_i) / \sum_{i=1}^N \delta_{ik}$  为第  $k$  个簇

的“质心”即聚类中心。 $\delta_{ik}$  是一个指示函数,如式(2)所示,其表征了样本是否属于某个簇。

### 1.2 最小最大 K 均值聚类算法

标准 K 均值聚类算法以最小化所有簇的簇内方差之和为目标,这种做法没有考虑各个簇的簇内方差之间的大小关系。可能会出现某些簇的簇内方差较大,而被另一些具有较小簇内方差的簇所抵消,使得总体上看各个簇的簇内方差之和仍然较小,最终导致算法陷入局部最优解。图 1 示出了用 iris 数据集<sup>[18]</sup>前三个属性值为坐标画出的样本点的空间分布,其中三个黑点代表一次不好的初始聚类中心。这种情况下得到的聚类结果如图 2 所示,iris 数据集的实际类别中心如图 3 所示。可见,由于初始聚类中心选取的不好,导致本来属于一类的样本被聚为两类,而本来的两类被聚为一类。

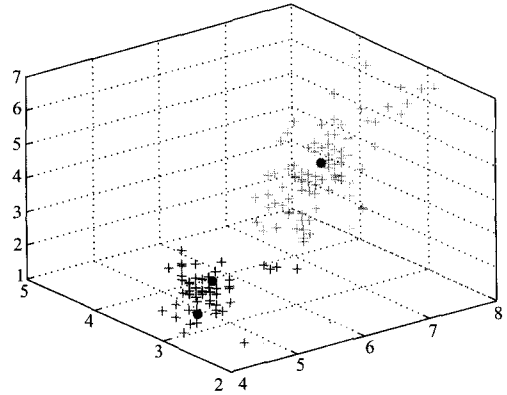


图 1 iris 数据集一次差的随机初始化的聚类中心

Fig. 1 Bad initialization clustering centers for iris data

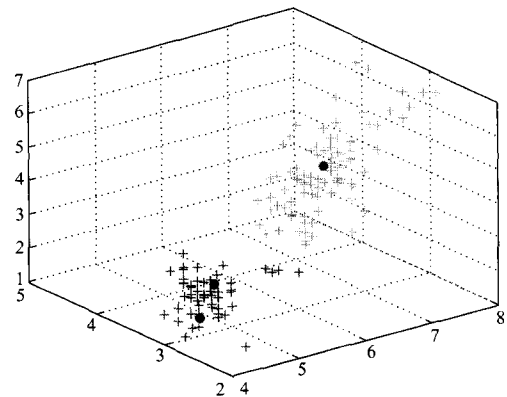


图 2 iris 数据集在不好的初始聚类中心情况下的 K 均值聚类结果

Fig. 2 A poor K-means solution for iris data set under bad initialization

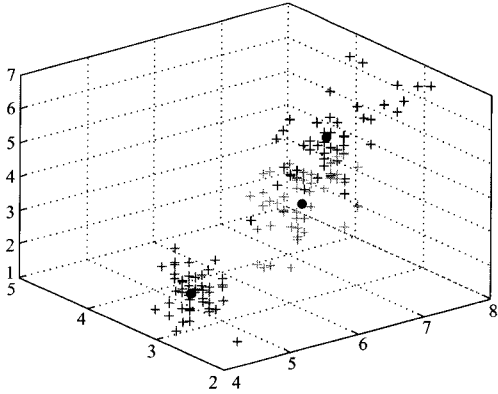


图 3 iris 数据集实际的类别中心

Fig. 3 Actual class centers of iris data set

基于上述考虑, 应建立一种综合考虑各个簇之间簇内方差值关系的聚类目标函数。文献[19]提出通过最小化  $K$  个簇中的最大簇内方差值来进行聚类, 即最小化下述式(3)的  $\varepsilon_m$  实现聚类, 被称为最小最大 K 均值聚类方法(The MinMax K-means clustering algorithm)。

$$\varepsilon_{\max} = \max_{1 \leq k \leq M} v_k = \max_{1 \leq k \leq M} \left\{ \sum_{i=1}^N \delta_{ik} \|x_i - m_k\|^2 \right\} \quad (3)$$

直接最小化式(3)的  $\varepsilon_m$  的求解十分困难, 为了能够类似于传统 K 均值那样通过迭代的方式来完成聚类, 将式(3)松弛化为式(4)。

$$\begin{aligned} \varepsilon_w &= \sum_{k=1}^M w_k^p v_k = \sum_{k=1}^M w_k^p \sum_{i=1}^N \delta_{ik} \|x_i - m_k\|^2 \\ w_k &\geq 0, \sum_{k=1}^M w_k = 1, 0 \leq p < 1 \end{aligned} \quad (4)$$

式中:  $v_k$ 、 $\delta_{ik}$  的意义同式(1)和式(2);  $w_k$  为各个簇的簇内方差的权重;  $p$  是  $w_k$  的指数, 是一个常量, 其控制着权重对不同簇方差差异的敏感性。 $p$  值越大则越敏感。

综上, 最小最大 K 均值算法可以描述为式(5)。

$$\begin{aligned} \min \max_{\{c_k\}_{k=1}^M, \{w_k\}_{k=1}^M} \varepsilon_w \\ \text{s.t. } w_k \geq 0, \sum_{k=1}^M w_k = 1, 0 \leq p < 1 \end{aligned} \quad (5)$$

式(5)是一个非凸的约束优化问题, 可以采用拉格朗日乘子法求解, 经过一系列的运算可以得到如下的解。

$$\begin{aligned} w_k &= v_k^{1/(1-p)} / \sum_{k'=1}^M v_{k'}^{1/(1-p)} \\ v_k &= \sum_{i=1}^N \delta_{ik} \|x_i - m_k\|^2 \end{aligned} \quad (6)$$

在聚类过程中簇及其中心不断被更新, 样本按照式(7)不断地被划分到新的簇中。显然, 随着权重  $w_k$  的不断增大, 只有非常接近聚类中心  $m_k$  的样本才被划分到簇  $k$  中。即样本被划分到簇中心与其之间的权重距离最近的那个簇中。聚类中心按照式(8)更新。

$$\delta_{ik} = \begin{cases} 1 & k = \arg \min_{1 \leq k' \leq M} w_{k'}^p \|x_i - m_{k'}\|^2 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

$$m_k = \frac{\sum_{i=1}^N \delta_{ik} x_i}{\sum_{i=1}^N \delta_{ik}} \quad (8)$$

由于  $0 \leq p < 1$ ,  $1/(1-p) > 0$ , 所以簇内方差  $v_k$  越大则其所对应的权重  $w_k$  越大。也就是说, 对于一个给定的数据划分, 最小最大 K 均值算法根据簇内方差值的大小来相应的设置权重, 方差值越大其对应的权重越大。随着算法的迭代, 簇及其中心不断更新, 新的样本按照其与簇中心的最近加权距离, 即式(7), 划分到最近的簇中。显然, 具有较高权重的簇, 其中心距离该簇中的样本的加权距离也相应增加。因此, 具有较大簇内方差的簇将失去一部分距离其中心较远的样本。同时, 具有较小簇内方差的簇, 由于其权重也比较小, 将得到一部分距离其中心较近的样本, 使得其簇内方差值增加。可见, 整个算法的迭代过程正是对拥有大方差的簇进行惩罚, 而对拥有小方差的簇进行补偿, 从而使得各个簇的簇内方差向差别减小的方向进行。

根据式(8), 任意两个簇的权重之比为:  $\frac{w_k}{w_{k'}} =$

$$\left( \frac{v_k}{v_{k'}} \right)^{1/(1-p)}, \quad \frac{w_k^p}{w_{k'}^p} = \left( \frac{v_k}{v_{k'}} \right)^{p/(1-p)}, \quad \text{由于 } 0 \leq p < 1, \text{ 随着}$$

$p$  的增加, 权重之比将变大, 也就是说, 不同的簇之间的簇内方差的差异被强化。 $p$  越大则不同簇的权重差异越大。换句话说,  $p$  值的大小控制了簇内方差的差异被反映在权重上的程度。过大的  $p$  值将导致方差大的簇被过度的惩罚, 使得其中的样本被过度的划分到其他簇, 显然不合适。过小的  $p$  值, 比如  $p=0$ , 则使得权重都为 1, 则算法变回了传统 K 均值算法。显然也是不合适的。本质上来讲,  $p$  的值控制了算法对大簇内方差的簇的容忍程度。 $p$  值应该设置的适中。经过分析, 一般  $p=0.5$  会取得不错的结果。

为了增加算法的稳健性, 在权重中设置了一个记忆效应, 如式(9)所示。

$$w_k^{(t)} = \beta w_k^{(t-1)} + (1-\beta) \left( v_k^{1/(1-p)} / \sum_{k'=1}^M v_{k'}^{1/(1-p)} \right), \quad 0 \leq \beta \leq 1 \quad (9)$$

这里  $\beta$  控制着上次迭代得到的权值对当前权值更新的影响。经过试验,  $\beta=0.3$  可以取得较好的结果。

最小最大 K 均值聚类算法的流程如图 4 所示。

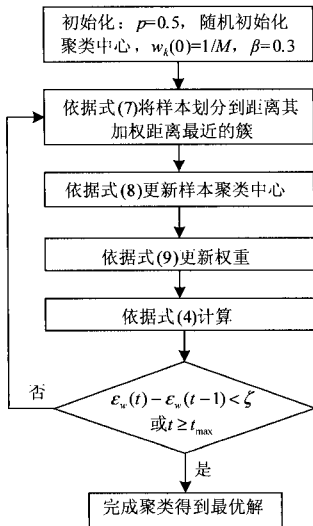


图 4 最小最大 K 均值聚类的算法流程

Fig. 4 Flowchart of minmax K-means clustering algorithm

算法开始时, 随机初始化聚类中心。参数设置  $p=0.5$ ,  $\beta=0.3$ , 初始权重  $w_k(0)=1/M$ 。将样本划分到距离其最近的簇, 更新聚类中心, 更新权重直到算法满足终止条件。终止条件中  $\zeta$  是一个事先给定的很小的正数,  $t_{\max}$  是事先给定的最大迭代次数。

### 1.3 引入核函数

核函数可以将样本从一个线性不可分的低维输入空间隐含地映射到线性可分性较强的高维特征空间。因此, 利用核函数可以改善聚类结果对样本分布的依赖性, 提高对线性不可分的样本的聚类效果。

样本集  $\mathcal{X}=\{\mathbf{x}_i\}_{i=1}^N$ ,  $\mathbf{x}_i \in \mathbf{R}^d$  在非线形映射  $\phi$  的作用下可以被映射到特征空间  $H$  中, 对应样本变为:  $\{\phi(\mathbf{x}_i)\}_{i=1}^N$ 。特征空间的点积可以用核函数来表示, 即  $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j)$ 。一般来讲, 满足 Mercer 条件的函数都可以作为核函数。本文选用应用较为广泛的高斯核函数, 如式(10)所示。

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (10)$$

引入高斯核函数后在特征空间中样本与簇中心的距离变为

$$\begin{aligned} \|\phi(\mathbf{x}_i) - \phi(\mathbf{m}_k)\|^2 &= \\ [\phi(\mathbf{x}_i) - \phi(\mathbf{m}_k)][\phi(\mathbf{x}_i) - \phi(\mathbf{m}_k)]^T &= \\ [K(\mathbf{x}_i, \mathbf{x}_i) - 2K(\mathbf{x}_i, \mathbf{m}_k) + K(\mathbf{m}_k, \mathbf{m}_k)] &= \\ 2[1 - K(\mathbf{x}_i, \mathbf{m}_k)] \end{aligned} \quad (11)$$

将式(11)替换 1.2 节中相应的样本与簇中心的距离  $\|\mathbf{x}_i - \mathbf{m}_k\|$  即可得到最小最大核 K 均值聚类方法。

### 1.4 数值试验

本部分将最小最大 K 均值算法与标准 K 均值算法以及 k-means++ 算法<sup>[20]</sup>相比较以验证算法的有效性。

k-means++ 算法的基本思想是在  $i-1$  个聚类中心 ( $i=1, \dots, k$ ) 确定的情况下, 对所有样本点选取下一个样本点加入初始点的集合, 每个点被选取的概率正比于它们的  $D$  值。 $D$  值是每个点到距离它们最近的初始点的距离的平方。k-means++ 算法经过对初始聚类中心细化处理, 使得其能够将初始聚类中心充分地分布在整个样本空间, 并且降低了 K 均值算法的迭代次数。因此被认为是对 K 均值算法的重大改进。为更有说服力地证明最小最大 K 均值算法的有效性, 本文将最小最大 K 均值聚类算法与 k-means++ 算法进行比较。

试验采用 iris<sup>[18]</sup>、wine<sup>[18]</sup>、segment<sup>[21]</sup> 和 pendigits<sup>[18]</sup> 数据集。这些数据集的基本信息在表 1 中列出。这些数据集本身带有类别标签, 试验时首先将其类别标签去掉, 分别应用 k-means、k-means++ 和本文的最小最大 K 均值算法对各个数据集进行聚类。然后比较各个算法的聚类召回率指标。召回率按照式(12)计算。其中  $S_{\text{correct}}$  表示类别  $i$  的实际的样本个数,  $S_{\text{comoute}}$  表示聚类到第  $i$  类的样本个数。全局召回率按照各个类别的召回率的平均值来计算。试验结果在表 2 中给出。

表 1 数据集的基本信息

Table 1 Basic information of data sets

数据集	样本数目	特征数目	类别数目	各类是否对称
iris	150	4	3	是
wine	178	13	3	否
segment	2 310	19	7	是
pendigits	7 494	16	10	否

$$\text{Re}(i) = \frac{S_{\text{correct}} \cap S_{\text{compute}}}{S_{\text{correct}}} \quad (12)$$

由表 2 可见, 最小最大 K 均值算法在 iris、segment 和 pendigits 数据集上的聚类效果均好于其他两种方法, 在 wine 数据集上与 k-means++ 算法的结果一样, 但都明显优于标准 k-means 算法。

表2 不同方法的聚类全局召回率比较  
Table 2 Comparison of global recall rate for  
different algorithms

算法	数据集			
	iris	wine	segment	pendigits
k-means	0.92	0.95	0.89	0.86
k-means++	0.95	0.96	0.92	0.89
MinMax	0.96	0.96	0.93	0.90

## 2 基于最小最大K均值聚类算法的水电机组振动故障诊断

### 2.1 水电机组的振动故障

水力发电系统是一个复杂的非线性动力系统,其故障多以振动的形式表现出来,机组振动可能是电气、机械或水力中的一种或者几种相互耦合引起。国内外已有不少研究者对水电机组的常见故障进行了系统的分析。

文献[22]从电气、机械和水力三个方面分析了影响水电机组振动的因素。电气方面的因素主要包括:发电机静态气隙不均匀、动态气隙不均匀、分数槽产生的次谐波、定子铁芯冲片松动及铁芯瓢曲、不对称三相负荷运行以及发电机出口突然短路等。机械方面的因素主要包括:机组轴线不正或对中不良、转动部分质量不平衡、机组支撑结构或者轴系刚度不足、推力轴承制造或调整不良、导轴承缺陷或间隙调整不当、轴密封调整不当等。水力方面的因素主要有:叶道涡、卡门涡、尾水管涡带、小开度压力脉动、水力自激振动及过渡过程中的不稳定现象等。

另外,根据水电机组的结构,可以将机组的常见故障分为五大类<sup>[7]</sup>,包括导轴承故障、大轴故障、推力轴承故障、发电机故障、过流部件故障等。每一大类有更具体的分类。其中导轴承包括上导、下导及水导3个轴承及其相应的支架;大轴包括发电机轴、水轮机轴及联轴节,部分水轮机还在上述2轴之间安装1个中间轴;推力轴承包括推力头、瓦块、弹性油箱及推力支架;发电机包括定子、转子及相关结构;过流部件包括导水机构、转轮及尾水管等。这些部位都有可能出现不同程度的故障,有些还会相互影响。

上述是一般意义上的水电机组振动故障的分类。事实上,由于水电机组工作的特殊性,机组振动是水、机、电共同作用的结果,机组故障可能出现在不同的部件和部位。不同的结构,不同的工作条件可能出现不同的新故障。因此,探求合适的无

监督的故障诊断方法对于有效识别机组故障尤其是新型故障具有重要的意义。

### 2.2 水电机组的最小最大核K均值聚类诊断模型

用最小最大核K均值算法对已知故障样本进行聚类,获得 $M$ 个聚类中心矢量:  $(C_1, C_2, \dots, C_M)$ 。设 $X$ 是一个未知的待诊断样本,则样本 $X$ 依据式(13)被诊断为第 $i$ 类故障。

$$i = \arg \min_{1 \leq i' \leq M} 2(1 - K(X, C_{i'})) \quad (13)$$

即比较待诊断样本与各个聚类中心的欧式距离,将样本划分到特征空间中距离其最近的聚类中心所代表的类别中。更新聚类中心,进一步完善诊断模型。如此,随着诊断工作的进展,诊断模型将不断地被添加新的样本,即诊断模型不断地学习新的“知识”,从而调整聚类中心,使得诊断模型不断完善。

### 2.3 工程应用实例

本部分采用所提出的聚类诊断模型运用于两个工程实例。

#### 2.3.1 实例1

大量故障实例表明:水电机组不同振动故障总是对应着一定的频率成分。同时,不同的振动故障,总会在水电机组的某个部位检测到超过允许值的振动或摆度。本文以水电机组振动信号的频谱特征中  $(0.4 \sim 0.5)x$ 、 $1x$ 、 $2x$ 、 $3x$ 、 $>3x$  五个频段上的幅值分量作为故障征兆属性( $x$ 为转频),按照文献[23]所提供的训练样本完成诊断模型的学习,并用测试样本对诊断模型的有效性进行评估,进行不平衡、不对中及涡带偏心三种典型故障的诊断。表3和表4分别示出训练样本和测试样本,所有样本都经过归一化处理。

表3 振动频谱征兆域训练样本

Table 3 Samples for training					
故障类型	故障特征				
	$(0.4 \sim 0.5)x$	$1x$	$2x$	$3x$	$>3x$
F1 涡带偏心	0.88	0.22	0.02	0.04	0.06
	0.90	0.20	0.05	0.02	0.02
	0.92	0.21	0.03	0.01	0.02
	0.85	0.25	0.06	0.02	0.01
	0.04	0.98	0.10	0.02	0.02
F2 不平衡	0.02	1.00	0.08	0.03	0.01
	0.05	0.90	0.11	0.05	0.02
	0.03	0.96	0.12	0.04	0.03
	0.02	0.41	0.43	0.34	0.15
F3 不对中	0.01	0.52	0.40	0.32	0.10
	0.01	0.40	0.47	0.35	0.18
	0.02	0.45	0.42	0.28	0.29

表 4 振动频谱征兆域测试样本  
Table 4 Samples for testing

标号	故障特征					现场诊断结论
	(0.4~0.5)x	1x	2x	3x	>3x	
1	0.82	0.28	0.05	0.04	0.03	涡带偏心
2	0.02	0.91	0.08	0.01	0.02	不平衡
3	0.01	0.48	0.48	0.36	0.20	不对中

采用所提出的聚类算法对表 3 的训练样本进行聚类，得到涡带偏心故障的聚类中心为：[0.8875, 0.2200, 0.0400, 0.0225, 0.0275]；不平衡故障的聚类中心为：[0.0350, 0.9600, 0.1025, 0.0350, 0.0200]；不对中故障的聚类中心为：[0.0150, 0.4450, 0.4300, 0.3225, 0.1800]。经过计算，表 4 中的样本 1、2、3 分别与涡带偏心故障类、不平衡故障类及不对中故障类聚类中心最近。因此，分别被诊断为涡带偏心、不平衡、不对中故障，与现场诊断结论一致。将测试样本分别划归到各自的类别中，重新计算各类的聚类中心矢量为：涡带偏心故障[0.8740, 0.2320, 0.0420, 0.0260, 0.0280]，不平衡故障[0.0320, 0.9500, 0.0980, 0.0300, 0.0200]和不对中故障[0.0140, 0.4520, 0.4400, 0.3300, 0.1840]。

2.3.2 实例 2

某水电站轴流式机组发生了以下三种典型振动故障：尾水管低频涡带(F1)、转轮叶片开口不均匀(F2)和发电机极频振动故障(F3)。从该机组采集到若干振动信号，依据文献[14]利用小波分析对这些振动信号进行分解，得到的每层信号的能量按照重构信号的顺序构成小波能量特征向量并归一化得到： $T=[E_1, E_2, \dots, E_7]$ 。每类典型故障整理出 6 个故障样本，如表 5 所示。

每类选取前 5 个样本作为训练样本，最后一个样本作为测试样本。采用本文所述的方法聚类得到故障 F1、F2、F3 的聚类中心分别为：[0.7654, 0.0686, 0.0010, 0.0138, 0.0588, 0.0740, 0.0184]，[0.0852, 0.0925, 0.0052, 0.2063, 0.4060, 0.0723, 0.0264]和 [0.0547, 0.0707, 0.0010, 0.0140, 0.0585, 0.0735, 0.7275]。最后，三个测试样本都被正确诊断出来。将测试样本添加进其相应的故障类之后得到新的 F1、F2、F3 的聚类中心分别为：[0.7072, 0.0772, 0.0013, 0.0215, 0.0792, 0.0868, 0.0270]，[0.0976, 0.0979, 0.0050, 0.1854, 0.3983, 0.0924, 0.0325]和 [0.0698, 0.0812, 0.0014, 0.0232, 0.0828, 0.0890, 0.6528]。

由上述两个工程实例可见本文所提出的诊断模型运算简单，具有较高的诊断率，在样本较少的情况下能够高效诊断出水电机组的典型故障。

表 5 小波能量特征表达的某水电机组故障样本  
Table 5 Fault samples of one HGU denoted  
by wavelet energy features

故障	小波能量特征						
	$E_1$	$E_2$	$E_3$	$E_4$	$E_5$	$E_6$	$E_7$
F1	0.740	0.087	0.001	0.014	0.061	0.077	0.019
	0.753	0.077	0.001	0.014	0.060	0.076	0.019
	0.766	0.068	0.001	0.014	0.059	0.074	0.018
	0.778	0.059	0.001	0.014	0.058	0.072	0.018
	0.790	0.052	0.001	0.013	0.056	0.071	0.018
	0.416	0.120	0.003	0.060	0.181	0.151	0.070
F2	0.101	0.113	0.006	0.233	0.455	0.075	0.018
	0.095	0.103	0.006	0.239	0.466	0.073	0.018
	0.089	0.094	0.006	0.245	0.476	0.072	0.018
	0.084	0.086	0.006	0.251	0.485	0.070	0.017
	0.080	0.078	0.006	0.256	0.494	0.068	0.017
	0.172	0.130	0.004	0.060	0.352	0.213	0.069
F3	0.068	0.091	0.001	0.015	0.061	0.077	0.686
	0.062	0.081	0.001	0.014	0.060	0.076	0.0706
	0.056	0.072	0.001	0.014	0.059	0.074	0.724
	0.050	0.064	0.001	0.014	0.058	0.072	0.742
	0.045	0.056	0.001	0.013	0.056	0.071	0.758
	0.130	0.123	0.003	0.060	0.180	0.151	0.354

3 结论

考虑到标准 K 均值聚类方法受初始聚类中心影响而易于陷入局部最优解的问题，本文提出一种基于最小最大核 K 均值聚类方法的水电机组故障诊断模型。相比标准 K 均值聚类方法，最小最大核 K 均值聚类方法考虑了各个簇之间的簇内方差的大小关系，通过为各个簇的簇内方差赋以与之成正比的权重，绕开了常规 K 均值聚类算法的局部最优解问题。同时引入核函数技术以提高对线性不可分的样本的聚类效果。在标准测试数据集上将所提聚类方法与标准 k-means 和 k-means++方法比较，验证了所提方法的有效性与优越性。基于所提方法构建了水电机组振动故障诊断模型，通过诊断实例的验证，证明该模型能够诊断出水电机组的典型故障。

参考文献

[1] 汪旋, 刘辉, 陈武晖, 等. 水轮发电机组励磁与水门协调控制设计[J]. 电工技术学报, 2013, 28(7): 262-268.  
WANG Ni, LIU Hui, CHEN Wuhui, et al. Coordinated control for the excitation and governing of hydroturbine generator set[J]. Transactions of China Electrotechnical Society, 2013, 28(7): 262-268.

- [2] 郭自刚, 陈俊, 陈佳胜, 等. 大型水电机组保护若干问题探讨[J]. 电力系统保护与控制, 2011, 39(3): 148-151.  
GUO Zigang, CHEN Jun, CHEN Jiasheng, et al. Discussion on some problems of large hydroelectric unit protection[J]. Power System Protection and Control, 2011, 39(3): 148-151.
- [3] 安学利, 周建中, 刘力, 等. 基于熵权理论和信息融合技术的水电机组振动故障诊断[J]. 电力系统自动化, 2008, 32(20): 78-82.  
AN Xueli, ZHOU Jianzhong, LIU Li, et al. Vibration fault diagnosis for hydraulic generator units based on entropy weight theory and information fusion technology[J]. Automation of Electric Power Systems, 2008, 32(20): 78-82.
- [4] 张孝远, 周建中, 王常青, 等. 考虑样本交叠的水电机组振动故障诊断[J]. 电力系统保护与控制, 2012, 40(3): 8-14.  
ZHANG Xiaoyuan, ZHOU Jianzhong, WANG Changqing, et al. Vibrant fault diagnosis for hydroelectric generator unit considering overlapping fault patterns[J]. Power System Protection and Control, 2012, 40(3): 8-14.
- [5] 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 Jianzhong, LI Chaoshun, 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.
- [6] 瞿翌, 赖旭, 朱建林. 水电机组在线状态监测和故障诊断网络化系统的设计与实现[J]. 电工技术学报, 2005, 20(3): 94-99.  
QU Zhao, LAI Xu, ZHU Jianlin. Design and implementation of on-line state monitoring and fault diagnosis network system for a hydroelectric generating set[J]. Transactions of China Electrotechnical Society, 2005, 20(3): 94-99.
- [7] 彭文季, 罗兴锜. 基于粗糙集和支持向量机的水电机组振动故障诊断[J]. 电工技术学报, 2006, 21(10): 117-122.  
PENG Wenji, LUO Xingqi. Vibration fault diagnosis of hydro-turbine generating unit based on rough sets and support vector machine[J]. Transactions of China Electrotechnical Society, 2006, 21(10): 117-122.
- [8] 陈林刚, 韩凤琴, 桂中华. 基于神经网络的水电机组智能故障诊断系统[J]. 电网技术, 2006, 30(1): 40-43, 98.  
CHEN Lingang, HAN Fengqin, GUI Zhonghua. A neural network based intelligent fault diagnosis system for hydroelectric generating sets[J]. Power System Technology, 2006, 30(1): 40-43, 98.
- [9] 彭文季, 罗兴锜. 基于小波神经网络的水电机组振动故障诊断研究[J]. 水力发电学报, 2007, 26(1): 123-128.  
PENG Wenji, LUO Xingqi. Research on vibration fault diagnosis of hydro-turbine generating unit based on wavelet neural network[J]. Journal of Hydroelectric Engineering, 2007, 26(1): 123-128.
- [10] 王铁生, 张利平. 免疫小波网络在水轮机组故障诊断中的应用[J]. 水利学报, 2009, 40(6): 762-767.  
WANG Tiesheng, ZHANG Liping. Application of immune wavelet network model to fault diagnosis of hydro-turbine generating units[J]. Journal of Hydraulic Engineering, 2009, 40(6): 762-767.
- [11] 彭文季, 罗兴锜, 赵道利. 基于频谱法与径向基函数网络的水电机组振动故障诊断[J]. 中国电机工程学报, 2006, 26(9): 155-158.  
PENG Wenji, LUO Xingqi, ZHAO Daoli. Vibrant fault diagnosis of hydro-turbine generating unit base on spectrum analysis and RBF network method[J]. Proceedings of the CSEE, 2006, 26(9): 155-158.
- [12] ZHANG Xiaoyuan, ZHOU Jianzhong, GUO Jun, et al. Vibrant fault diagnosis for hydroelectric generator units with a new combination of rough sets and support vector machine[J]. Expert Systems with Applications, 2012, 39(3): 2621-2628.
- [13] 彭文季, 罗兴锜, 郭鹏程, 等. 基于最小二乘支持向量机和信息融合技术的水电机组振动故障诊断[J]. 中国电机工程学报, 2007, 27(23): 86-92.  
PENG Wenji, LUO Xingqi, GUO Pengcheng, et al. Vibration fault diagnosis of hydroelectric unit based on LS-SVM and information fusion technology[J]. Proceedings of the CSEE, 2007, 27(23): 86-92.
- [14] 邹敏, 周建中, 刘忠. 基于支持向量机的水电机组故障诊断[J]. 中国农村水利水电, 2008(1): 114-116, 119.  
ZOU Min, ZHOU Jianzhong, LIU Zhong. Support vector machines based approach for hydroelectric generating unit fault diagnosis[J]. China Rural Water and Hydropower, 2008(1): 114-116, 119.
- [15] 陈铁华, 陈启卷. 模糊聚类分析在水电机组振动故障诊断中的应用[J]. 中国电机工程学报, 2002, 22(3): 43-47.

- CHEN Tiehua, CHEN Qijuan. Fuzzy clustering analysis based vibration fault diagnosis of hydroelectric generating unit[J]. Proceedings of the CSEE, 2002, 22(3): 43-47.
- [16] 刘晓波, 黄其柏. 基于动态核聚类分析的水轮机组故障模式识别[J]. 华中科技大学学报: 自然科学版, 2005, 33(9): 47-49, 52.
- LIU Xiaobo, HUANG Qibai. Classification on the modes of hydro-generator unit fault based on dynamic kernel cluster analysis[J]. Journal of Huazhong University of Science and Technology: Nature Science, 2005, 33(9): 47-49, 52.
- [17] 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 Chaoshun, ZHOU Jianzhong, 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.
- [18] BLAKE C L, MERZ C J. UCI repository of machine learning databases[EB/OL]. [1998]. <http://www.ics.uci.edu/mllearn/MLRepository.html>.
- [19] TZORTZIS G, LIKAS A. The minmax K-means clustering algorithm[J]. Pattern Recognition, 2014, 47(7): 2505-2516.
- [20] ARTHUR D, VASSILVITSKII S. K-means++: the advantages of careful seeding[C] // Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, 2007: 1027-1035.
- [21] MICHIE D, SPIEGELHALTER D J, TAYLOR C C. Machine learning, neural and statistical classification [EB/OL]. [1994]. <ftp.ncc.up.pt/pub/statlog/>.
- [22] 樊世英. 大中型水力发电机组的安全稳定运行分析[J]. 中国电机工程学报, 2012, 32(9): 140-148.
- FAN Shiying. Analysis of stability for large and medium-sized hydro-generating unit[J]. Proceedings of the CSEE, 2012, 32(9): 140-148.
- [23] 杨晓萍, 解建宝, 孙超图. 水轮发电机组振动故障诊断的神经网络方法研究[J]. 水利学报, 1998, 29(4): 94-97.
- YANG Xiaoping, XIE Jianbao, SUN Chaotu. Neural network method for vibration fault diagnosis of hydroelectric generating set[J]. Journal of Hydraulic Engineering, 1998, 29(4): 94-97.

---

收稿日期: 2014-05-13; 修回日期: 2014-07-28

作者简介:

张孝远(1981-), 男, 通信作者, 讲师, 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

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。

(编辑 姜新丽)