第43卷第5期 2015年3月1日

电力系统保护与控制 PowerSystemProtectionandControl

Vol.43 No.5 Mar.1, 2015

# 基于最小最大核K均值聚类算法的水电机组

张孝远，张新萍²，苏保平1

(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

ZHANG Xiaoyuan', ZHANG Xinping2, SU Baoping' (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 cancluster thefault samples intoa 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, afault 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 NationalNatural ScienceFoundation of China (No.51409095). Key words: hydro-turbine generating unit; vibration; fault diagnosis; minmax K-means clustering algorithm;kernel function

# 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振动信号的内部特征也将发生变化。据报道，水电 安全还直接关系到其互联电网的稳定性[-2]。因此，

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)

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 水电机组的运行健康状况不仅关系到水电厂自身的振动信号的内部特征也将发生变化。据报道，水电 安全还直接关系到其互联电网的稳定性[-2]。因此， 机组约有80%的故障或事故都在振动信号中有所反 映3。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47。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

-28

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[8-11]和支持向量机[6,12-14]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[15-17]。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。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-一最小最大 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

# 1.1K均值聚类

将一个数据集x={x;,x,∈R"无监督地划分 到M个互无交集的簇{c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。

Esum= ∑∑o llx, mel [ =1 if x,∈ Ck =0otherwise

电力系统保护与控制

为簇内样本个数。m，=（>x）/>6）为第k个簇 i=1 的“质心”即聚类中心。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， 使得总体上看来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。

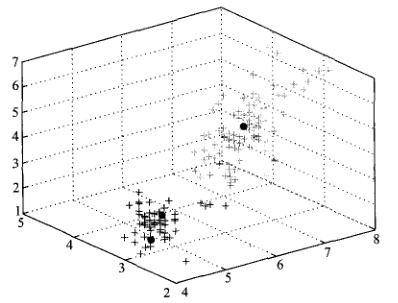
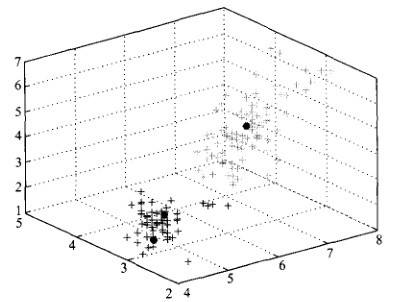


图1iris数据集一次差的随机初始化的聚类中心 Fig.1 Bad initialization clustering centers foriris data



(1)

(2)

本文将簇内方差定义为v= Zox llx-ml, 其与一般意义上的方差v之间的关系为v=Nv，N

图2iris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 poorK-means solution for iris data set underbadinitialization

张孝远，等 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

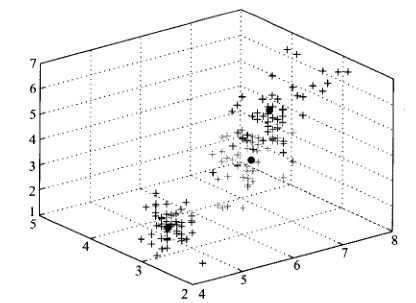


图3iris数据集实际的类别中心 Fig.3 Actual class centers of iris data set

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ε实现聚类，被称为最小 最大K均值聚类方法(TheMinMaxK-means clustering algorithm)。

max=max v=max Zox; --m 1SkSM=]

29-

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的不断增加，只有非常接近聚类中心m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。

1 k=argminsesm w|x, -ml 1 S 0otherwise m=

(7) (8)

由于0≤p<1，1/(1-p)>0，所以簇内方差v越 大则其所对应的权重w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 个算法的选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。 Wk' 1/(1-p) ）p/(1-p) ，由于0≤p<1，随着 (Vx) 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。

(3)

直接最小化式(3)的ε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)。

M M. 8= ∑wV=∑w∑ollx;-m1² k=1 M wx ≥0, ∑w =1, 0≤p<1 =

(4)

式中：V、O的意义同式(1)和式(2)；w为各个簇 的簇内方差的权重；P是w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)。

min maxE c}{w M s.t. w ≥0,∑w =1,0≤ p<1 k=1

(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。

1/(1-p Wk=v(1-p) v =∑ox1x, -mk|2

6)

w=βw(-)+(1-β) 0≤β≤1

(9)

30

这里β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，β=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。

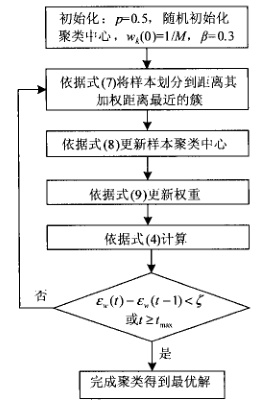


图4最小最大K均值聚类的算法流程 Fig.4 Flowchart of minmax K-means clusteringalgorithm

算法开始时，随机初始化聚类中心。参数设置 p=0.5，β=0.3，初始权重w(0)=1/M。将样本划 分到距离其最近的，更新聚类中心，更新权重直 到算法满足终止条件。终止条件中是一个事先给 定的很小的正数，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x={x},x,∈R在非线性映射的作 用下可以被映射到特征空间H中，对应样本变为： {x,)》。特征空间的点积可以用核函数来表示， 即K(x,x;)=(x)(x;）。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。

[x-y K(x,y)= exp(- 2o²

电力系统保护与控制

o(x;)-0(m)² = [Φ(x;)-Φ(m)][Φ(x;)-Φ(m)]= [K(x,x;)-2K(x,mk)+K(mk,m）]= 2[1-K(x,m）]

(11)

将式(11)替换1.2节中相应的样本与簇中心的距离 －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20相比较以验证算法的有效 性。 k-means++算法的基本思想是在i1个聚类中心 (i=l,",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均值算法的大大改 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iris18winel18、segment和 pendigits18] 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Scorect表示类别i的实际的样本个数， Scomoute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 Table1Basic 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 | 2310 | 19 | 7 | 是 |
| pendigits | 7 494 | 16 | 10 | 香 |

Scoreet MS Re(i)= Scoreet

(12)

(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

Scoreet 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。

张孝远，等 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

表2不同方法的聚类全局召回率比较 Table2Comparisonofglobalrecall ratefor 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 |

# 基于最小最大K均值聚类算法的水电机组振

# 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飘曲、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”，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

31-

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：（C1,C2，,CM)。 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 2(1- K(X,C))

(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 (0.4~0.5)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 3Samples for training

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 故障类型 | 故障特征 | | | | |
| (0.4~0.5)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 |
| F2不平衡 | 0.85 0.04 | 0.25 0.98 | 0.06 0.10 | 0.02 0.02 | 0.01 0.02 |
| 0.02 | 1.00 | 0.08 | 0.03 | 0.01 |
| 0.05 | 0.90 | 0.11 | 0.05 | 0.02 |
| F3不对中 | 0.03 0.02 | 0.96 | 0.12 | 0.04 | 0.03 0.15 |
|  | 0.41 | 0.43 | 0.34 | 0.10 |
| 0.01 | 0.52 | 0.40 | 0.32 0.35 | 0.18 |
| 0.01 | 0.40 | 0.47 |  |  |
| 0.02 | 0.45 | 0.42 | 0.28 | 0.29 |

- 32 -

表4振动频谱征兆域测试样本 Table4Samplesfor 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,E2,,E,]。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。

电力系统保护与控制

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 故 障 | 小波能量特征 | | | | | | |
| E | E2 | E3 | EA | Es | E6 | E |
| 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 |  | 0.006 | 0.233 | 0.455 | 0.075 | 0.018 |
| 0.095 | 0.113 |  | 0.239 | 0.466 | 0.073 | 0.018 |
| 0.089 | 0.103 0.094 | 0.006 0.006 | 0.245 | 0.476 | 0.072 | 0.018 |
|  | 0.086 | 0.006 | 0.251 | 0.485 | 0.070 | 0.017 |
| 0.084 |  | 0.006 | 0.256 | 0.494 | 0.068 | 0.017 |
| 0.080 | 0.078 0.130 | 0.004 | 0.060 | 0.352 | 0.213 |  |
| 0.172 0.068 | 0.091 | 0.001 | 0.015 | 0.061 | 0.077 | 0.069 0.686 |
| 0.062 | 0.081 | 0.001 | 0.014 | 0.060 | 0.076 | 0.0706 |
| F3 |  |  |  |  |  |  |  |
| 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

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 controlfor the excitation and governing of hydroturbine generator set[J].Transactions of China Electrotechnical Society,2013,28(7):262-268.

张孝远，等 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

[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].PowerSystem 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 entropyweighttheoryandinformationfusion 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[].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]瞿翠，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刀].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[].Transactions of China Electrotechnical Society, 2005, 20(3): 94-99. [7]彭文季，罗兴.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,LUOXingqi.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，韩风琴，桂中华.基于神经网络的水电机组

33-

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 hydroelectricgeneratingsets[J]. PowerSystem Technology,2006,30(1): 40-43,98. [9]彭文季，罗兴.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，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[刀.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 spectrumanalysisandRBFnetworkmethod[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，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[].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-SVMandinformationfusiontechnology[J]. Proceedings ofthe 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 unitfaultdiagnosis[J].ChinaRuralWaterand 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.

- 34 -

CHEN Tiehua, CHEN Qijuan.Fuzzy clustering analysis based vibrationfault diagnosisof 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]l. 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/mlearn/MI.Repository.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

电力系统保护与控制

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. FANShiying.Analysis of stabilityfor 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 networkmethodforvibrationfauitdiagnosisof 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）