

DOI: 10.7667/PSPC151450

# 基于 PSO 优化核主元分析的海上风电机组运行工况分类

郑小霞<sup>1</sup>, 李美娜<sup>1</sup>, 王靖<sup>2</sup>, 任浩翰<sup>2</sup>, 符杨<sup>1</sup>

(1. 上海电力学院自动化工程学院, 上海 200090; 2. 上海东海风力发电有限公司, 上海 200090)

**摘要:** 海上风电机组运行环境复杂多变, 对其工况进行分类可以提高机组运行健康状态评价的准确性, 为制定合理的运行维护策略提供可靠依据。提出一种基于 PSO 优化核主元分析(KPCA)的多参数工况分类方法。针对核函数参数难以确定的问题, 综合考虑类内散度和类间散度构建优化核参数的适应度函数, 应用 PSO 算法对其进行寻优, 将优化后的 KPCA 用于数据的特征提取, 在此基础上采用模糊 C-均值聚类(FCM)建立分类模型。通过对 UCI 数据库中的三组实验数据进行分类验证了该方法的有效性。最后, 应用该方法对某海上风电场实测数据进行工况分类, 并与 PCA+FCM、KPCA+FCM 两种方法进行比较。结果表明, 提出方法的分类结果优于其他两种, 能够得到清晰准确的分类结果, 利于分工况建立准确的机组运行健康状态评价模型。

**关键词:** 海上风电机组; 工况分类; PSO; 核主元分析; 类别可分性; 模糊 C-均值聚类

## Operational conditions classification of offshore wind turbines based on kernel principal analysis optimized by PSO

ZHENG Xiaoxia<sup>1</sup>, LI Meina<sup>1</sup>, WANG Jing<sup>2</sup>, REN Haohan<sup>2</sup>, FU Yang<sup>1</sup>

(1. School of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China;

2. Shanghai Donghai Wind Power Co., Ltd., Shanghai 200090, China)

**Abstract:** As considering the complex operational conditions of offshore wind turbines, classifying the operational conditions can improve the accuracy of health condition evaluation of wind turbines and provide a reliable basis for the reasonable operation maintenance strategies. An operational conditions classification method based on kernel principal analysis (KPCA) optimized by particle swarm optimization algorithm (PSO) is proposed. The fitness function of kernel function parameter optimizations is constructed by considering within class scatter and between-class scatter of data to avoid the problem that choosing a proper kernel function parameter is difficult. The KPCA optimized by PSO is applied to data feature extraction and a classification model is built by using fuzzy C-means (FCM) clustering algorithm. The simulations on three groups data of UCI database prove the method's validity. Finally, the method is used to classify the operational conditions of offshore wind turbines, and the results show that the proposed method can get much better classification effect than PCA+FCM and KPCA+FCM. It is helpful for establishing health condition evaluation model for each condition.

This work is supported by National Natural Science Foundation of China (No. 51507098).

**Key words:** offshore wind turbines; operational conditions classification; PSO; kernel principal analysis; sort separability criterion; fuzzy C-means clustering

## 0 引言

随着风电资源的大规模开发, 海上风电因具有风能资源能量效益高, 海上风湍流强度小, 不占用

宝贵的土地资源等优势在全球范围内得到大力发展。风电机组运行状态具有复杂多变的特点, 而海上风电机组面临着强风载荷、海水腐蚀和波浪冲击等更加恶劣的运行环境<sup>[1]</sup>, 机组的运行工况更加复杂多变, 并且风机故障率较高, 一旦故障停运时间更长, 因此机组的运行维护成本更大、难度更高。通常, 运行维护成本约占海上风电场总投资的 15%~30%。对机组运行工况进行分类, 有利于针对

基金项目: 国家自然科学基金项目(51507098); 上海绿色能源并网工程技术研究中心(13DZ2251900); 上海市科委重点科技攻关项目(14DZ1200905); 上海市电站自动化技术重点实验室项目(13DZ2273800)

不同工况做出相应的客观评价, 制定合理的运行维护策略, 降低运行维护成本。目前, 关于风电机组工况分类的研究文献很少, 文献[2]考虑单一参数根据风速划分风电机组运行工况, 文献[3]克服了单一参数的缺陷, 针对机组振动监测阈值的设定进行工况划分的研究, 提出了基于数据采集与监控系统 (Supervisory control and data acquisition, SCADA) 数据和支持向量机的多运行参数运行工况划分方法, 文献[4]考虑机组运行工况提出一种基于工况辨识的机组健康状态实时评价方法, 由于文献中主要研究评价方法, 并没有过多考虑工况划分的效果, 且以上都是对陆上风电机组的研究, 对海上风电机组的研究更少。核主元分析 (Kernel principal component analysis, KPCA) 利用非线性变换将输入数据空间映射到高维特征空间, 然后在高维空间使用主元分析 (Principal component analysis, PCA) 方法提取非线性主元<sup>[5]</sup>, 该方法已在处理实际的非线性问题中取得了较好的效果。核主元分析应用中的关键问题是核参数的选择, 但目前还没有太多的理论作指导, 研究如何针对实际问题选择客观合理的核参数具有重要意义。传统的 KPCA 在选取主元时考虑的是尽可能保留原始信息, 但对于分类问题并没有充分考虑到类别信息的最大化, 数据降维后不一定能够得到较适于分类的成分。粒子群算法 (Particle swarm optimization, PSO) 是基于群体智能理论的优化算法, 目前已广泛应用于函数优化、数据挖掘等领域<sup>[6-8]</sup>。为此, 本文提出了一种基于 PSO 优化核主元分析的海上风电机组运行工况分类方法, PSO 优化核参数时, 不仅考虑保留大量原始信息, 而且尽量使所选主元的类别信息最大化, 将核主元分析提取的特征集作为模糊 C-均值聚类算法的输入, 建立海上风电机组运行工况分类模型。

## 1 核主元分析的基本原理

核主元分析是主元分析的非线性推广, 其基本思想是通过一个非线性变换  $\Phi$ , 将输入数据空间中的  $M$  个样本  $\mathbf{x}_k (k=1, 2, \dots, M)$ , 映射到高维特征空间  $F$  中, 得到高维空间的样本点  $\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_M)$ , 样本中心化得到  $\Phi(\tilde{\mathbf{x}}_k)$ , 再利用主元分析提取主元, 实现非线性数据降维。

中心化后数据的协方差矩阵为

$$\mathbf{C}^F = \frac{1}{M} \sum_{j=1}^M \Phi(\tilde{\mathbf{x}}_j) \Phi(\tilde{\mathbf{x}}_j)^T \quad (1)$$

对应的特征方程为

$$\lambda \mathbf{v} = \mathbf{C}^F \mathbf{v} \quad (2)$$

求解式(2)中的特征值  $\lambda (\lambda \geq 0)$  和特征向量  $\mathbf{v}$ , 每个样本与式(2)做内积, 进而有

$$\lambda [\Phi(\tilde{\mathbf{x}}_k) \cdot \mathbf{v}] = \Phi(\tilde{\mathbf{x}}_k) \cdot \mathbf{C}^F \mathbf{v}, \quad k=1, 2, \dots, M \quad (3)$$

上式中  $\mathbf{v}$  可由  $\Phi(\tilde{\mathbf{x}}_k)$  线性表出, 即

$$\mathbf{v} = \sum_{i=1}^M \alpha_i \Phi(\tilde{\mathbf{x}}_i) \quad (4)$$

其中,  $\alpha_i$  为相关系数。从而可得:

$$\frac{1}{M} \sum_{i=1}^M \alpha_i \cdot \left[ \Phi(\tilde{\mathbf{x}}_k) \cdot \sum_{i=1}^M \Phi(\tilde{\mathbf{x}}_j) \right] \cdot \left[ \Phi(\tilde{\mathbf{x}}_j) \cdot \Phi(\tilde{\mathbf{x}}_i) \right] = \lambda \sum_{i=1}^M \alpha_i [\Phi(\tilde{\mathbf{x}}_k) \cdot \Phi(\tilde{\mathbf{x}}_i)] \quad (5)$$

定义  $M \times M$  核矩阵  $\mathbf{K}$ :

$$\mathbf{K}_{ij} = \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\tilde{\mathbf{x}}_i) \cdot \Phi(\tilde{\mathbf{x}}_j) \quad (6)$$

由于  $\mathbf{K}$  为对称阵, 则上式可简化为

$$M \lambda \mathbf{K} \alpha = \mathbf{K}^2 \alpha \quad (7)$$

$$M \lambda \alpha = \mathbf{K} \alpha \quad (8)$$

通过对上式的求解, 即可得到要求的特征值和特征向量, 归一化特征向量后, 可得样本在  $F$  空间中第  $k$  个主成分, 即在向量  $\mathbf{V}_k$  上的投影。

$$\mathbf{V}^k \cdot \Phi(\tilde{\mathbf{x}}) = \sum_{i=1}^M \alpha_i [\Phi(\tilde{\mathbf{x}}_i) \cdot \Phi(\tilde{\mathbf{x}})] \quad (9)$$

上述过程假设数据已中心化, 但一般情况下中心化是不成立的, 因此  $\mathbf{K}$  用  $\tilde{\mathbf{K}}$  代替:

$$\tilde{\mathbf{K}} = \mathbf{K} - \mathbf{I} \mathbf{K} - \mathbf{K} \mathbf{I} + \mathbf{I} \mathbf{K} \mathbf{I} \quad (10)$$

其中,  $\mathbf{I} = \begin{bmatrix} 1 & & \\ & \ddots & \\ & & 1 \end{bmatrix}_{l \times l}$ 。

核主元分析是利用核函数内积运算实现数据的非线性映射, 而无需关注具体映射形式。常用核函数如下。

多项式函数:

$$\mathbf{K}(\mathbf{x}, \mathbf{z}) = (\langle \mathbf{x}, \mathbf{z} \rangle + l)^p, \quad l > 0 \quad (11)$$

高斯径向基核函数:

$$\mathbf{K}(\mathbf{x}, \mathbf{z}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{z}\|^2}{2\sigma^2}\right), \quad \sigma > 0 \quad (12)$$

Sigmoid 函数:

$$\mathbf{K}(\mathbf{x}, \mathbf{z}) = \tanh(\langle \mathbf{x}, \mathbf{z} \rangle + c) \quad (13)$$

研究表明, Sigmoid 核函数在特定参数下与径向基核函数具有类似的特性, 且径向基核函数预测速度和精度在处理大多实际问题中优于其他函数, 故本文采用高斯径向基核函数。对于高斯径向基核函数, 影响其应用结果的重要因素是核参数  $\sigma$ 。

## 2 基于 PSO 优化的核主元分析

粒子群算法是一种有效的全局寻优算法, 由 Kennedy 和 Eberhart 于 1995 年提出<sup>[9-10]</sup>。本文采用带收缩因子和惯性权重线性递减的 PSO 算法进行核参数优化, 加快收敛速度并防止陷入局部最优, 其数学表达式为

$$v_{id}^{t+1} = K[(\omega v_{id}^t + c_1(p_{id} - x_{id}^t) + c_2 r_2 (g_{id} + x_{id}^t))] \quad (14)$$

$$x_{id}^{t+1} = x_{id}^t + v_{id}^{t+1} \quad (15)$$

$$\omega = \omega_{\max} - \frac{(\omega_{\max} - \omega_{\min}) \times t}{T_{\max}} \quad (16)$$

式中:  $K$  为收缩因子, 一般取 0.729;  $\omega_{\max}$ 、 $\omega_{\min}$  分别是惯性权重的最大值和最小值;  $t$  是迭代次数;  $T_{\max}$  为最大迭代次数。

本文将核主元分析特征提取的结果作为工况分类模型的输入, 因此在优化核参数  $\sigma$  时, 不仅考虑提取的特征向量集保留的原始信息量, 同时考虑该向量集所具有的类别信息, 选取既能够反映原信息又有利于分类的核参数和特征向量集。

### 2.1 类别可分性判据

传统的 KPCA 进行特征提取时, 通常选取累积贡献率大于 85% 的前  $m$  个主元构成特征向量集, 认为能够与原数据有较高的一致性, 即只考虑提取的主元是否保留了大量的原始信息, 但是保留的信息量多并不一定是最有利于实际问题分析, 对于模式分类数据还应尽可能的提取出类别信息大的特征向量集。类别可分性判据可以衡量提取的特征向量对分类的贡献大小, 即具有的类别信息的大小。基于类内类间距离的可分性判据是一种常用的类别可分性判据。对于分类结果, 我们希望类内距离越小越好, 类间距离越大越好, 可以用类内离散度和类间离散度用来度量类内距离和类间距离, 为此构造能反映类内距离和类间距离的准则函数。

对于  $c$  类样本,  $\mathbf{X}^i(x_1, x_2, \dots, x_{n_i}), i=1, 2, \dots, c$ , 第  $i$  类的类内离散度矩阵为

$$S_w^i = \frac{1}{n_i} \sum_{j=1}^{n_i} (x_j^i - m_i)(x_j^i - m_i)^T \quad i=1, 2, \dots, c \quad (17)$$

总的类内离散度矩阵为

$$S_w = \sum_{i=1}^c \frac{n_i}{N} S_w^i \quad (18)$$

类间离散度矩阵为

$$S_b = \sum_{j=1}^{n_i} \frac{n_i}{N} (x_j^i - m)(x_j^i - m)^T \quad i=1, 2, \dots, c \quad (19)$$

式中:  $N$  是样本总数;  $n_i$  是  $i$  类的样本个数;  $m_i$  是  $i$

类的样本集均值向量;  $m$  是所有样本集均值向量。

本文采用一种用离散度矩阵给出的基于类内类间距离的可分性判据:

$$J_d(x) = \text{tr}(S_w^{-1} S_b) \quad (20)$$

$J_d$  越大表示特征向量的可分性越好, 具有的类别信息越大。

### 2.2 PSO 优化核参数模型

本文采用模糊 C-均值聚类实现工况的分类。模糊聚类引入模糊隶属度, 具有描述样本中间性的优点, 能够客观的反映实际。其中应用最广泛的是基于目标函数的模糊 C-均值聚类算法<sup>[11-12]</sup>。设  $c$  个聚类中心为  $V=\{v_1, v_2, \dots, v_c\}$ , 全部  $n$  个样本的聚类情况由一个隶属度矩阵  $U$  描述,

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1n} \\ \vdots & \vdots & \vdots \\ u_{c1} & \cdots & u_{cn} \end{bmatrix} \quad (21)$$

式中,  $u_{ij}$  为第  $j$  个样本归属于第  $i$  类的隶属度。

模糊 C-均值聚类按照如下准则和约束条件聚类, 得到最优的模糊分类<sup>[12]</sup>:

$$\min J(U, V; X) = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|v_i - x_j\|^2 \quad (22)$$

$$\begin{cases} 0 \leq u_{ij} \leq 1, 1 \leq i \leq c, 1 \leq j \leq n \\ 0 < \sum_{j=1}^n u_{ij} < n, 1 \leq i \leq c \\ \sum_{i=1}^c u_{ij} = 1, 1 \leq j \leq n \end{cases} \quad (23)$$

式中:  $m > 1$  是加权指数;  $\|v_i - x_j\|$  为样本  $x_j$  与聚类中心  $v_i$  之间的距离, 一般为欧式距离。

结合可分性判据, 综合考虑类别可分性和保留原信息量, 建立 PSO 优化核函数参数模型, 具体步骤如下:

- (1) 初始化粒子群各参数;
- (2) 在解空间内初始化粒子的位置和速度;
- (3) 计算粒子个体适应度值: 每个粒子进行一次核主元分析, 选取前 3 个主元, 如果累积贡献率大于 85% 则进行一次模糊 C-均值聚类, 按式(20)计算粒子的适应度值, 否则, 适应度值直接赋值为极小值。
- (4) 比较粒子当前适应度值  $p_{id}$  和其自身历史最优解  $p_{best}$ , 如果  $p_{best} \leq p_{id}$  更新个体最优解  $p_{best} = p_{id}$ 。
- (5) 比较粒子当前适应度值  $p_{id}$  和种群最优解  $g_{best}$ , 如果  $g_{best} \leq p_{id}$ , 更新种群最优解  $g_{best} = p_{id}$ 。
- (6) 按式(14)、式(15)更新粒子的速度和位置, 按式(16)更新惯性权重。

(7) 判断迭代次数是否达到终止条件, 如果  $t < T_{max}$ , 返回步骤(3), 否则将当前解作为最优解输出, 算法终止。

从以上步骤可以看出, PSO 在优化核参数时, 只考虑累积贡献率达到 85% 以上的粒子, 保留了大量的原始信息, 同时以聚类后可分性判据作为 PSO 优化的适应度函数, 考虑了聚类后提取的各主元对分类的贡献大小, 选取有利于分类的主元。因此, PSO 优化后可以得到既保留了大量的原信息又有利于工况分类的核参数。基于粒子群优化核参数的流程图如图 1 所示。

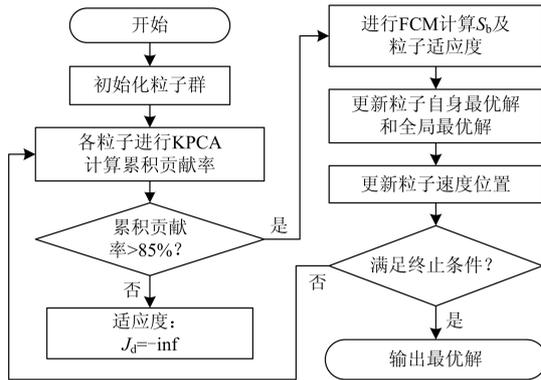


图 1 PSO 优化核参数流程图

Fig. 1 Flow chart of kernel parameter optimization based on PSO

将 PSO 优化后的核主元分析提取得到的特征向量集作为模糊 C-均值聚类的输入, 可以实现数据的分类, 为了验证方法的有效性, 本文采用 3 组实验数据进行了分类实验, 数据来源于 UCI 的 Seeds Data Set, Gesture Segmentation Data Set, Leaf Data Set, 为了便于观察将数据统一降到 3 维, 这 3 组数据都是可以用来聚类的, 降维的目标是提取有效的类别信息得到类别分明的数据, 在图中表现即为类别清晰。3 组数据的维数分别是 7 维、18 维、16 维, 得到的分类结果如图 2~图 4 所示。

由上面的分类图可以看到, PSO 优化后的核主元分析提取不同维数数据的特征向量集后, 通过模

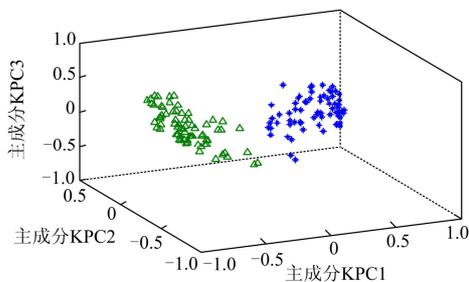


图 2 Seeds Data Set 分类图

Fig. 2 Classification result of Seeds Data Set

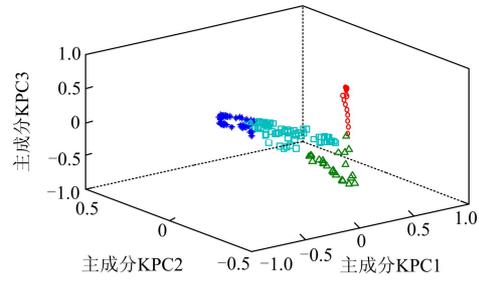


图 3 Gesture Segmentation Data Set 分类图

Fig. 3 Classification result of Gesture Segmentation Data Set

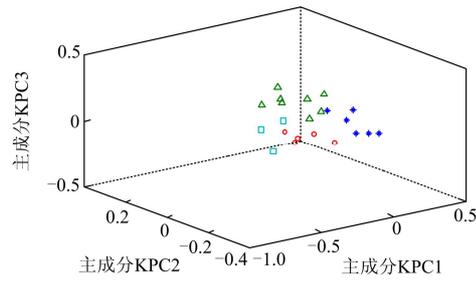


图 4 Leaf Data Set 分类图

Fig. 4 Classification result of Leaf Data Set

糊 C-均值聚类算法进行分类基本都可以得到类别清晰的分类结果, 这主要是因为 PSO 优化核参数时考虑了类别可分性, 提取出既保留大量的原始信息又同时得到具有最大类别信息的数据。

### 3 海上风电机组运行工况分类建模

#### 3.1 运行工况特征参数选取

合理的工况分类对机组运行健康状态评价结果有重要的影响, 而正确选择机组运行状态特征参数是得到合理工况分类的必要条件。通常情况下海上风电场均采用远程监测系统(SCADA 系统)对风电场的情况进行实时监测, SCADA 系统采集包括风速、有功功率、齿轮箱温度等几十个与风机运行状态相关的参数, 根据各因素对风电机组运行状态的影响可将这些参数分为机组性能特征参数和外界环境特征参数<sup>[13]</sup>。其中, 反映机组性能的参数主要包括齿轮箱、发电机、控制柜、机舱等相关参数, 外界环境因素参数主要包括风速、环境温度等参数。通过对历史运行数据的关联分析可以看到, 环境温度、风速和转速对机舱温度、绕组温度等参数具有明显的影响, 其相互关系曲线如图 5~图 7 所示。

图 5 中可以看到反映机组性能的机舱温度和塔筒柜温度受环境温度影响明显, 呈现近似线性关系, 均随着环境温度的增大而增大, 因此在工况分类时需要考虑环境温度的影响。

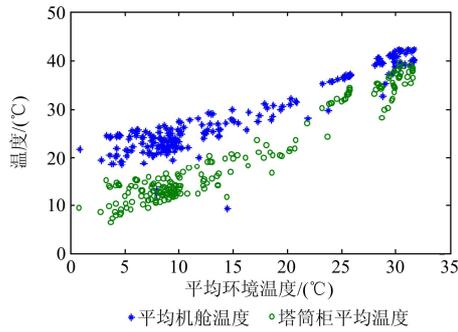


图 5 机舱温度和塔筒柜温度与环境温度的关系  
Fig. 5 Relationship among nacelle temperature, tower temperature and environment temperature

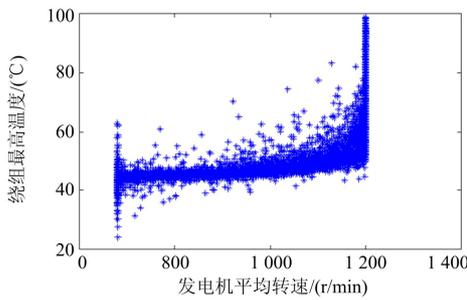


图 6 绕组最高温度与发电机转速的关系  
Fig. 6 Relationship between winding's maximum temperature and generator rotational speed

图 6 中绕组最高温度随着发电机转速的增大而有所升高，可以看出实测的绕组最高温度分布是一条宽带。

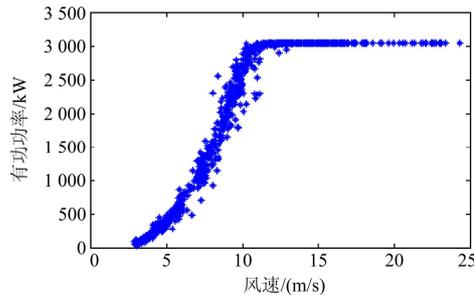


图 7 发电机有功功率与风速的关系  
Fig. 7 Relationship between active power and wind speed

图 7 是机组的功率特性曲线，由图中看到当风速低于额定风速时，风机处于功率增加阶段，有功功率随风速增大而增大；当风速在额定风速至切出风速之间时，风机保持输出功率恒定。该曲线是考核机组性能、评估机组发电能力的一项重要指标，通过比较实测的功率特性曲线与制造商提供的功率特性曲线可以评估机组的实际运行性能。

综上，环境温度、有功功率、风速和发电机转

速应该作为工况分类的特征参数。另外，齿轮箱和发电机轴承是风电机组中故障高发的部件，因此，本文选取齿轮箱油温、发电机轴承温度、环境温度、有功功率、风速和发电机转速 6 个关键参数建立机组工况分类模型。

### 3.2 建立运行工况分类模型

本文提出的基于 PSO 优化核主元析的海上风电机组运行工况分类模型的框图如图 8 所示。

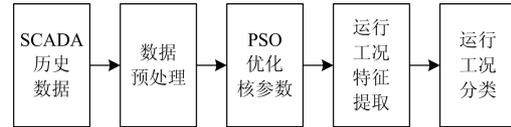


图 8 海上风电机组运行工况分类框图

Fig. 8 Framework of operational conditions classification for offshore wind turbines

具体实现步骤如下：

- (1) 数据的预处理，采集 SCADA 历史数据，合理选取数据，进行标准化处理；
- (2) 运行工况参数特征提取，利用 PSO 优化高斯核函数参数，得到最优的核参数，进行核主元分析提取特征向量，得到新的主元变量集；
- (3) 工况分类，将得到的主元变量集作为模糊 C-均值聚类的输入，得出工况分类结果。

## 4 实例分析

为了验证工况分类模型的有效性，本文采用某海上风电场 SCADA 系统采集的机组运行数据进行分类实验。该风电场采集的数据包含风速、环境温度、机舱温度、有功功率、发电机轴承温度、绕组温度、齿轮箱油温等 31 个运行参数，本文选取某一台 3 MW 风电机组 2012 年 1 月至 12 月的正常运行数据，该机组的额定参数如表 1 所示。

表 1 风电机组额定参数

Table 1 Rated parameters of wind turbine

风机型号	额定风速	额定功率	切入风速	切出风速
SL3000/90	13 m/s	3 000 kW	3.5 m/s	25 m/s

从中选取由齿轮箱油温、发电机轴承温度、环境温度、有功功率、风速和发电机转速 6 个参数构成的特征参数集，选取 208 组数据(去除坏点)，在选取数据时参考 IEC 标准中风电机组功率特性测试要求<sup>[14]</sup>，选取风速范围为低于切入风速 1 m/s 至切出风速，得到所选数据的风速范围为 2.64~24.36 m/s，几乎覆盖了此台机组的绝大多数运行工况。

将选取的数据按 4 中步骤实现工况分类。其中, PSO 优化核函数参数时各参数设定如表 2 所示。

表 2 粒子群算法 (PSO) 参数设定表

Table 2 Parameters set of PSO

参数名称	设定值
种群规模 $n$	40
惯性权重 $\omega_{\max}, \omega_{\min}$	0.9, 0.4
学习因子 $c1, c2$	2
收缩因子 $K$	0.729
核参数范围( $2\sigma^2$ )	[0.02, 200]
最大迭代次数 $T_{\max}$	100
粒子最大速度 $v_{\max}$	10

其中, 核参数范围是经过多次试验, 将范围由大到小逐渐缩小, 以保证能够寻得最优解。

图 9 是 PSO 优化核参数时适应度函数随迭代次数进化过程, 图 10 是 KPCA 的累积贡献率随迭代次数变化曲线, 由图中看出优化时只考虑了累积贡献率达到 85% 以上的核参数值, 得到最优的核参数 1.862 1。

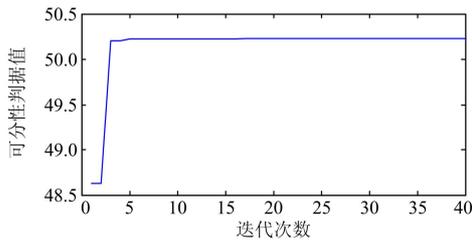


图 9 PSO 优化核参数时适应度变化曲线

Fig. 9 PSO fitness change curve during process of kernel parameter optimization

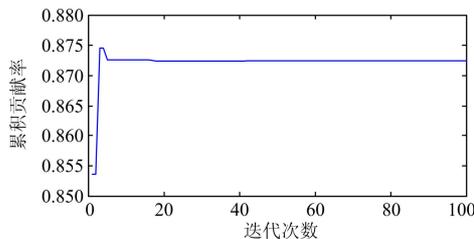


图 10 PSO 优化核参数时累积贡献率变化曲线

Fig. 10 Cumulative contribution rate change curve during process of kernel parameter optimization

表 3 给出 PSO 优化的核主元分析结果, 可以看到, 前 3 个主元的累积贡献率达到了 87.24%, 选取新主元保留了原数据的大部分信息, 可以代替原始的 31 维数据。

为体现本文所提方法的优越性, 分别采用三种

方法进行分类试验。分别是 PCA+FCM、KPCA+FCM 及 PSO-KPCA+FCM, 采用 KPCA+FCM 时核参数  $\sigma$  分别取 1、3 和 3.5, 实验得到的分类结果如图 11~图 15 所示。表 4 是各方法进行降维时得到累积贡献率。

表 3 核主元分析结果

Table 3 Kernel principal analysis result

特征值序号	特征值	累积贡献率/%
1	56.884 7	48.16
2	34.159	77.08
3	12.002 4	87.24
4	6.289 8	92.57
5	4.962 9	96.77
6	3.813 7	100

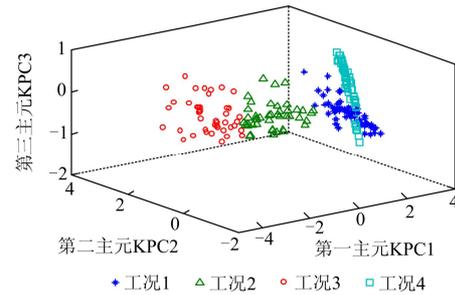


图 11 PCA+FCM 工况分类结果图

Fig. 11 Classification result of PCA+FCM

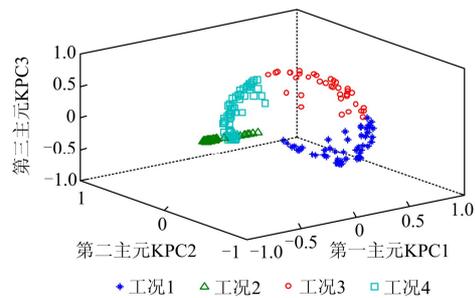


图 12  $\sigma = 1$  时 KPCA+FCM 工况分类结果图

Fig. 12 Classification result of KPCA+FCM while  $\sigma = 1$

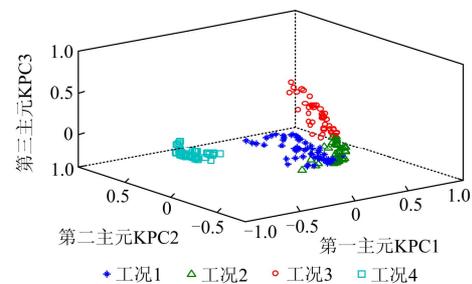


图 13  $\sigma = 3$  时 KPCA+FCM 工况分类结果图

Fig. 13 Classification result of KPCA+FCM while  $\sigma = 3$

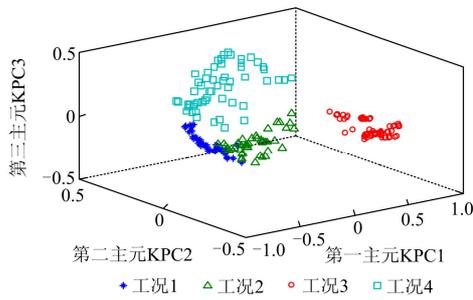


图 14  $\sigma = 3.5$  时 KPCA+FCM 工况分类结果图

Fig. 14 Classification result of KPCA+FCM while  $\sigma = 3.5$

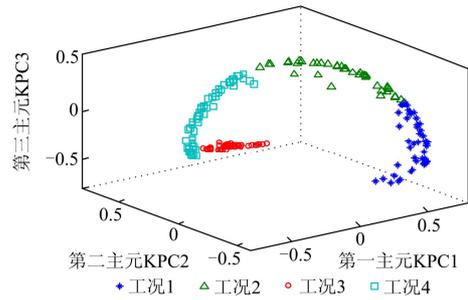


图 15 PSO-KPCA+FCM 工况分类结果图

Fig. 15 Classification result of PSO-KPCA+FCM

表 4 不同方法进行特征提取时的累积贡献率  
Table 4 Cumulative contribution rate of different feature extraction methods

特征提取方法	累积贡献率/%
PCA	97.239 7
KPCA $\sigma = 1$	75.92
KPCA $\sigma = 1.862 1$	87.24
KPCA $\sigma = 3$	92.2
KPCA $\sigma = 3.5$	93.28

从以上实验结果可以看出，PSO-KPCA+FCM 的分类效果要好于其他方法，能够清晰地划分出各个工况。对于 KPCA 方法，选择不同的核参数  $\sigma$  对分类结果影响很大，当  $\sigma$  选取最优核参数的邻近值 1、3 和 3.5 时得到的分类效果并不好。如果  $\sigma$  选择不当 KPCA+FCM 的分类效果可能没有 PCA+FCM 的效果好，如图 13 所示，当  $\sigma = 3$  时得到的分类结果明显没有图 11 所示的 PCA+FCM 的效果好。

结合表 4 与各方法分类结果图，可以看出，对于分类问题在特征提取时，不能仅考虑保留原始信息，还要尽量保留数据的分类信息，因为并不是保留信息量越多、累积贡献率越大就能够得到好的分类效果。如表 4 所示，PCA 以及 KPCA 中  $\sigma$  选取 3 和 3.5 时得到的累积贡献率都在 90% 以上，高于本文所提方法优化所得  $\sigma$  值 1.862 1 的 87.24%，但是

图 11、图 13 及图 14 表明这三种方法提取的数据分类效果并不理想，分类结果图中都有明显不同程度的工况特征点混杂在一起的情况，类别划分不清晰，边界不明显。而对于图 15，由于本文所提方法在选取 KPCA 核参数时，同时考虑了保留数据的原始信息和类别信息，在聚类后得到了清晰明确的类别划分结果。

综上，本文所提方法能够对海上机组运行工况空间进行清晰准确的划分。其中的核参数优化方法，在考虑保留原始信息量的同时兼顾数据的类别信息，能提取出更有效、更利于分类的非线性主元，很适用于解决分类问题中 KPCA 核参数的选择问题。同时，得到的机组运行工况划分结果，可以用来将 SCADA 历史数据进行相应的划分，生成相应个数的健康状态评价模型的训练样本集，通过高斯混合模型(GMM)或隐马尔科夫(HMM)等模型进行训练，建立各运行工况子空间下的健康状态评价模型，实现根据不同工况准确的建立机组的运行健康状态评价模型，并为合理制定运行维修策略提供依据。

## 5 结论

本文提出了一种基于 PSO 优化核主元分析的海上风电机组运行工况分类方法。在分类模型的特征提取部分，综合考虑了原始信息和数据的类别信息，提取出利于分类的特征参数。用本文所提的方法对某海上风电机组的 SCADA 历史数据进行工况空间划分，得到了优于其他方法的分类结果，验证了该方法的有效性和优越性，为更客观合理地实现分工况评价机组运行健康状态提供了基本保证，有利于及时制定合理的机组运行维护策略。

## 参考文献

- [1] 黄玲玲, 符杨, 胡荣, 等. 基于运行维护的海上风电机组可用性评估方法[J]. 电力系统自动化, 2013, 37(16): 13-17.  
HUANG Lingling, FU Yang, HU Rong, et al. Availability evaluation of offshore wind turbine considering operation and maintenance[J]. Automation of Electric Power Systems, 2013, 37(16): 13-17.
- [2] BECKER E, POSTE PAUL, RESLER J. Keeping the blades turning: condition monitoring of wind turbine gears[J]. Wind Energy, 2006, 7(2): 26-32.
- [3] 刘永前, 王飞, 时文刚, 等. 基于支持向量机的风电机组运行工况分类方法[J]. 太阳能学报, 2010, 31(9): 1191-1197.  
LIU Yongqian, WANG Fei, SHI Wengang, et al. Operation

- condition classification method for wind turbine based on support vector machine[J]. *Acta Energeiae Solaris Sinica*, 2010, 31(9): 1191-1197.
- [4] 董玉亮, 李亚琼, 曹海斌, 等. 基于运行工况辨识的风电机组健康状态实时评价方法[J]. *中国电机工程学报*, 2013, 33(11): 88-95.  
DONG Yuliang, LI Yaqiong, CAO Haibin, et al. Real-time health condition evaluation on wind turbines based on operational condition recognition[J]. *Proceedings of the CSEE*, 2013, 33(11): 88-95.
- [5] SEHOLKOPF B, SMOLA A, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. *Neural Computation*, 1998, 10(5): 1299-1319.
- [6] 吴忠强, 贾文静, 吴昌韩, 等. 基于 PSO-BSNN 的短期风速预测[J]. *电力系统保护与控制*, 2015, 43(15): 36-41.  
WU Zhongqiang, JIA Wenjing, WU Changhan, et al. Short-term wind speed forecasting based on PSO-BSNN[J]. *Power System Protection and Control*, 2015, 43(15): 36-41.
- [7] 刘欢, 刘志刚. 基于改进粒子群算法的牵引变电所维修优化研究[J]. *电力系统保护与控制*, 2015, 43(11): 87-94.  
LIU Huan, LIU Zhigang. An improved particle swarm algorithm study on optimization model of maintenance schedules for railway traction substations[J]. *Power System Protection and Control*, 2015, 43(11): 87-94.
- [8] 卢锦玲, 苗雨阳, 张成相, 等. 基于改进多目标粒子群算法的含风电场电力系统优化调度[J]. *电力系统保护与控制*, 2013, 41(17): 25-31.  
LU Jinling, MIAO Yuyang, ZHANG Chengxiang, et al. Power system optimal dispatch considering wind farms based on improved multi-objective particle swarm algorithm[J]. *Power System Protection and Control*, 2013, 41(17): 25-31.
- [9] EBERHART R, KENNEDY J. A new optimizer using particle swarm theory[C] // *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, Nagoya Japan, 1995: 39-43.
- [10] KENNEDY J, EBERHART R. Particle swarm optimization[C] // *Proceedings IEEE International Conference on Neural Networks*, 1995: 1942-1948.
- [11] 孟安波, 卢海明, 李海亮, 等. 纵横交叉算法优化 FCM 在电力客户分类中的应用[J]. *电力系统保护与控制*, 2015, 43(20): 150-154.  
MENG Anbo, LU Hanming, LI Hailiang, et al. Electricity customer classification based on optimized FCM clustering by hybrid CSO[J]. *Power System Protection and Control*, 2015, 43(20): 150-154.
- [12] 杨茂, 熊昊, 严干贵, 等. 基于数据挖掘和模糊聚类的风电功率实时预测研究[J]. *电力系统保护与控制*, 2013, 41(1): 1-6.  
YANG Mao, XIONG Hao, YAN Gangui, et al. Real-time prediction of wind power based on data mining and fuzzy clustering[J]. *Power System Protection and Control*, 2013, 41(1): 1-6.
- [13] 胡姚刚, 唐显虎, 刘志详. 并网风电机组在线运行状态评估方法[J]. *中国电机工程学报*, 2010, 30(33): 103-109.  
HU Yaogang, TANG Xianhu, LIU Zhixiang. Method for on-line operation conditions assessment for a grid-connected wind turbine generator system[J]. *Proceedings of the CSEE*, 2010, 30(33): 103-109.
- [14] IEC 61400-12-1 Ed.1 power performance measurements of electricity producing wind turbines[S]. 2005.

收稿日期: 2015-08-17; 修回日期: 2015-11-07

作者简介:

郑小霞(1978-), 女, 博士, 副教授, 主要研究方向为风力发电故障诊断与运行维护等; E-mail: zhengxiaoxia@shiep.edu.cn

李美娜(1987-), 女, 通信作者, 硕士研究生, 主要研究方向为海上风电机组运行维护。E-mail: xiaolingna@163.com

(编辑 姜新丽)