

DOI: 10.19650/j.cnki.cjsi.J1905106

# 基于集成 NSET 和模糊软聚类的风电机组齿轮箱状态监测\*

王梓齐<sup>1</sup>, 刘长良<sup>1,2</sup>, 刘 帅<sup>1</sup>

(1. 华北电力大学 控制与计算机工程学院 保定 071003;  
2. 新能源电力系统国家重点实验室(华北电力大学) 北京 102206)

**摘要:**风电机组齿轮箱的故障频率和维修成本较高,有必要对其运行状态进行实时监测。非线性状态估计(NSET)算法有着对记忆矩阵依赖大、无法有效利用数据资源改善精度、实时性差等不足。为此,提出一种基于模糊软聚类和集成 NSET 的状态监测方法;使用模糊软聚类将历史数据分为边界有重叠的不同类别,实现工况的软划分并构造多个不同工况的 NSET 模型作为个体学习器;以参数回归方法作为结合器,可在不影响实时性的同时,使用大量数据训练参数以改善精度。用某 2 MW 风电机组的齿轮箱故障数据进行验证,结果表明,相比常规方法,提出方法的精度和实时性均更优;通过预测残差均值和基于残差构造的健康指数,能够灵敏、准确的反映齿轮箱的早期故障及其发展趋势。

**关键词:** 集成学习;非线性状态估计;模糊软聚类;风电机组齿轮箱;状态监测

**中图分类号:** TM315      **文献标识码:** A      **国家标准学科分类代码:** 470.4047

## Condition monitoring of wind turbine gearbox based on ensemble nonlinear state estimation technique and soft fuzzy clustering

Wang Ziqi<sup>1</sup>, Liu Changliang<sup>1,2</sup>, Liu Shuai<sup>1</sup>

(1. School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, China;  
2. State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources  
(North China Electric Power University), Beijing 102206, China)

**Abstract:** The failure frequency and maintenance cost of wind turbine gearbox are relatively high. It is necessary to monitor its operation condition in real-time. Nonlinear state estimation technique (NSET) has the problems of high dependence on memory matrix, low accuracy caused by insufficient data utilization, bad real-time performance, etc. Therefore, a condition monitoring method based on soft fuzzy C-means clustering (SFCM) and ensemble NSET is proposed. SFCM is adopted to divide the historical data into different classes with overlapping boundaries to achieve the soft condition division. NSET models under different conditions are constructed as individual learner. The parametric regression method is used as the combiner. Lots of data can be used to train the parameters without affecting the real-time performance and the accuracy can be improved accordingly. The gearbox fault data of a 2 MW wind turbine are taken to evaluate the method. Compared with NSET, experimental results show that the proposed method has better accuracy and real-time performance. Through the means of predicted residual and the health index based on residual, it can reflect the early fault and its development trend of gearbox sensitively and accurately.

**Keywords:** ensemble learning; nonlinear state estimate technique; soft fuzzy clustering; wind turbine gearbox; condition monitoring

## 0 引言

据 2016 年对中国风电机组故障情况的统计<sup>[1]</sup>, 齿轮箱的故障频率仅次于变频器、发电机和变桨系统, 在主要设备中位列第四。齿轮箱的维护成本也相对较高, 维修时间和费用均位于各主要设备的前列<sup>[2]</sup>。

齿轮箱的状态监测方法主要有油液分析、振动分析和运行数据分析 3 大类。油液分析是对润滑油进行理化性能分析, 以金属磨粒等指标监测齿轮箱的运行状态<sup>[3]</sup>, 能够直观反映其机械劣化情况, 可靠性较高; 但金属磨粒传感器的成本较高且实时性较差, 有必要进一步研究低成本解决方案。振动分析是从齿轮箱的故障特征敏感部位采集振动信号, 通过时、频域信号处理方法提取特征因子并进行分析<sup>[4]</sup>, 监测信息的敏感度和实时性较好, 能够对全寿命周期进行监测和故障诊断; 但振动分析的应用多以事后诊断为主, 很难做到天以上提前预警齿轮箱的运行异常, 有必要与其他监测方法配合使用。

近年来, 基于监控与数据采集 (supervisory control and data acquisition, SCADA) 系统运行数据的状态监测方法受到了广泛关注, 主要是使用历史数据建立设备正常运行时的状态变量模型, 通过状态变量预测值和测量值间的残差进行状态监测。

以是否显式包含参数分类, 可将状态变量建模方法分为参数和非参数 2 类。常规的参数方法有 BP 神经网络<sup>[5]</sup>、混合高斯模型<sup>[6]</sup>等, 学习能力强但泛化能力较差、建模精度较低。深度学习<sup>[7]</sup>是如今的研究热点, 能够有效利用海量、多类型的数据进行学习, 建模精度和泛化能力相比常规参数方法有较大幅度的提升。但现阶段深度学习模型的可解释性较差、需要掌握较好的调参技巧才能保证算法的性能, 有必要针对实际工程应用进一步深入研究。

非参数方法有非线性状态估计 (nonlinear state estimate technique, NSET)<sup>[8]</sup>、核回归<sup>[9]</sup>等, 有着无需进行训练、建模精度高、泛化能力强等优势, 但实时计算速度相对较慢。其中, NSET 是根据设备正常工作时的运行数据构成记忆矩阵, 进而对状态变量进行预测, 具有易于实现、建模精度高等特点, 受到了较多关注。如何高效利用现有数据资源, 提高计算速度和建模精度是 NSET 算法的研究重点。

文献[10]基于 Relief 算法进行特征选择, 降低了记忆矩阵的维度、提高了计算速度。文献[11]基于马氏距离对数据进行等距抽样构造记忆矩阵, 提高了建模精度和计算速度。文献[12]基于相似度原理对记忆矩阵进行约简, 在对精度影响不大的同时减少了计算时间。文献[13]在基于欧式距离的基础上引入极限学习机算法

计算权重, 提高了建模精度。文献[14]以 2 范数概率密度为指标抽取样本构造记忆矩阵, 提高了计算速度。文献[15]基于模糊 C 均值 (fuzzy c-means, FCM) 聚类进行工况划分并构造多个 NSET 模型, 根据当前运行工况调用对应的模型进行计算, 改善了计算速度。

以上研究在一定程度上增强了 NSET 的实用性, 但原理上算法对记忆矩阵的依赖较大, 无法有效利用记忆矩阵以外的数据。要想进一步提高建模精度, 往往需要对记忆矩阵进行扩展, 会对实时性造成较大负面影响, 甚至出现病态问题导致算法失效<sup>[16]</sup>。

针对上述问题, 本文基于集成学习理论, 将聚类算法、参数回归方法与 NSET 算法进行结合, 提出了一种基于集成 NSET 的风电机组齿轮箱状态监测方法: 基于模糊软聚类算法, 对少量运行数据进行工况软划分, 进而构造多个不同工况的 NSET 子模型作为个体学习器; 以参数回归方法作为结合器, 将 NSET 子模型的输出进行综合, 可使用大量运行数据训练结合器参数, 在保留 NSET 泛化能力强、建模精度高等优势的同时, 有效利用大量运行数据资源进行训练, 从而改善 NSET 对记忆矩阵依赖大的不足。

以某 2 MW 风电机组 SCADA 系统的实际运行数据为算例, 针对齿轮箱的状态监测问题, 将本文提出的集成 NSET 方法与几种参数方法及常规 NSET 方法进行了对比, 并基于该机组齿轮箱的实际故障验证了算法的故障预警能力。

## 1 算法基础

### 1.1 非线性状态估计

NSET 是由 Gross 等<sup>[17]</sup>提出的一种非线性、非参数的状态监测方法, 原理如下。

设共有  $n$  个运行参数参与状态变量建模, 对应的测量值记为观测向量  $X = [x_1, x_2, \dots, x_n]^T$ 。设备正常运行时, 采样得到  $m$  个观测向量构造记忆矩阵  $D$  如式(1)所示。

$$D = [X_1, X_2, \dots, X_m] = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{m1} \\ x_{12} & x_{22} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{mn} \end{bmatrix}_{n \times m} \quad (1)$$

$D$  的构造是对设备正常运行状态的记忆和学习。

NSET 的输入为某一时刻的观测向量  $X_e$ , 输出为  $X_e$  的估计值  $\hat{X}_e$ 。对任意  $X_e$ , 均计算一个权值向量  $W = [w_1, w_2, \dots, w_m]^T$  令输出  $\hat{X}_e$  等于记忆矩阵  $D$  和  $W$  的乘积:  $\hat{X}_e = D \cdot W = w_1 \cdot X_1 + w_2 \cdot X_2 + \dots + w_m \cdot X_m$  (2) 权值向量  $W$  通过最小化  $X_e$  与  $\hat{X}_e$  间的残差得到<sup>[18]</sup>:

$$W = (D^T \otimes D)^{-1} \cdot (D^T \otimes X_0) \quad (3)$$

式中： $\otimes$ 为非线性运算符，常选取为2向量间的欧氏距离。与参数方法不同，NSET计算输出时需要重新计算权值向量。

将式(3)代入式(2)， $X_e$ 的计算公式如式(4)所示。

$$X_e = D \cdot (D^T \otimes D)^{-1} \cdot (D^T \otimes X_0) \quad (4)$$

若设备正常运行，状态变量能被记忆矩阵中的观测向量精确重构、估计精度较高，反之则精度较低。

输出 $X_e$ 受记忆矩阵 $D$ 的影响较大，故加入新的观测向量可能会改善建模精度，但会导致以下问题。

1) 由于权值向量 $W$ 需要重复计算，若 $D$ 的规模较大，式(4)中 $D^T \otimes X_0$ 项的计算时间会大幅增长，严重影响算法在线运算的实时性。

2) 加入的观测向量可能与 $D$ 中已有数据的相似度较高、对改善精度的作用有限，同时会使矩阵各列的相似性增大，导致式(4)中 $D^T \otimes D$ 的条件数变大，使 $(D^T \otimes D)^{-1}$ 转化为病态问题无法求解<sup>[16]</sup>。

3) 加入的观测向量可能有异常数据或较大的噪声，会对建模精度造成负面影响。

## 1.2 模糊C均值软聚类

按照划分结果的不同，可将聚类算法分为硬聚类和软聚类2种<sup>[19]</sup>；硬聚类的划分结果是类别标签，样本属于某一类的概率是0或1且只能属于单一类别；软聚类的划分结果是隶属度，样本属于某类别的概率在 $[0, 1]$ 区间内且可属于多个类别。软聚类划分结果如图1所示。

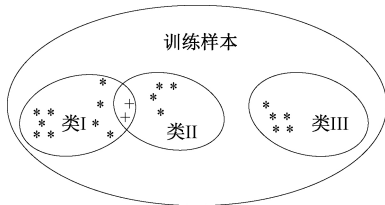


图1 软聚类的划分结果

Fig.1 Partition results of soft clustering

模糊聚类算法的聚类结果是隶属度矩阵，但实际应用中多基于最大隶属度原则，将隶属度矩阵转化为类别标签，得到硬聚类的结果。但是，位于类别边界处的样本可能不完全属于某单一类别，使得基于模糊聚类结果建立的回归模型，对处于类别边界处的数据预测精度相对较低<sup>[20]</sup>。

Lyu等<sup>[21]</sup>在FCM的基础上，提出了一种模糊C均值软聚类(soft fuzzy c-means, SFCM)算法，用软划分规则替代最大隶属度原则，使样本可被划分给多个类别，得到在类别边界处有部分重叠的软聚类结果。近年来，SFCM算法在燃煤电站脱硝系统的建模<sup>[22-23]</sup>中得到了一定应

用，有效改善了模型的预测精度。SFCM算法的实现流程如下。

首先，基于常规FCM算法将样本进行聚类，得到相应的隶属度矩阵 $U = [u_{ij}]_{m \times T}$ ，其中 $m$ 为样本个数， $T$ 为聚类个数。之后，基于如下软划分规则进行划分。

若样本 $x_i$ 的最大隶属度 $u_{ik} = \max_{1 \leq j \leq T} \{u_{ij}\}$ 满足式(5)条件，则 $x_i$ 被唯一地划分为类别 $k$ ：

$$u_{ik} > 0.5 + 0.5T^{-1} \quad (5)$$

否则，若隶属度 $u_{ij}$ 满足式(6)条件，则 $x_i$ 被划分为类别 $j$ ，此时 $x_i$ 可属于多个类别：

$$u_{ij} > (T + \delta)^{-1} \quad (6)$$

式中： $\delta$ 为重叠度； $\delta$ 越大，边界上的重叠部分越大。

当样本距离某类别中心较近且距离其他类别较远时，SFCM以式(5)为原则进行硬划分，与最大隶属度原则一致；当样本距离某几个类别中心均较为接近时，SFCM以式(6)为原则将该样本划分给多个类别，得到软聚类的结果。

对于机械设备特别是齿轮箱等旋转机械，运行工况的变化是较为连续的，不同工况间存在过渡状态。对于过渡状态下的运行数据，相比硬聚类，使用软聚类可能会将其同时分配给邻近的多个工况，更符合工况连续变化的实际情况。此外，SFCM算法仅是在类别划分时采用软划分规则替代最大隶属度原则，无需改变常规FCM的聚类过程，较易于工程实现。

## 2 集成NSET状态监测方法

### 2.1 集成学习

集成学习是机器学习领域的研究热点之一，核心思想是针对同一问题训练多个相异的个体学习器，并使用结合器将结果综合后输出<sup>[24]</sup>，可应用于分类、聚类、回归等任务中。多数回归算法与集成学习进行结合后，往往能够在不加入新的训练样本、不改变参数训练方法的同时，有效提高精度和泛化能力。

以非参数方法作为个体学习器的集成学习模型，主要有集成支持向量机、集成 $K$ 最近邻、随机森林等，但多以研究分类问题为主。对于状态监测这一回归问题，目前仍以集成学习与参数方法的结合为主。

### 2.2 集成NSET状态变量建模方法

本文的集成NSET状态变量建模方法如图2所示。图2中，近期历史数据的含义为训练样本中时间标签相对较晚、数量相对较少的部分数据，用来构造NSET的记忆矩阵；远期历史数据的含义为除近期历史数据以外、时间标签较早的训练样本，与近期历史数据一起用来训练结合器的参数。

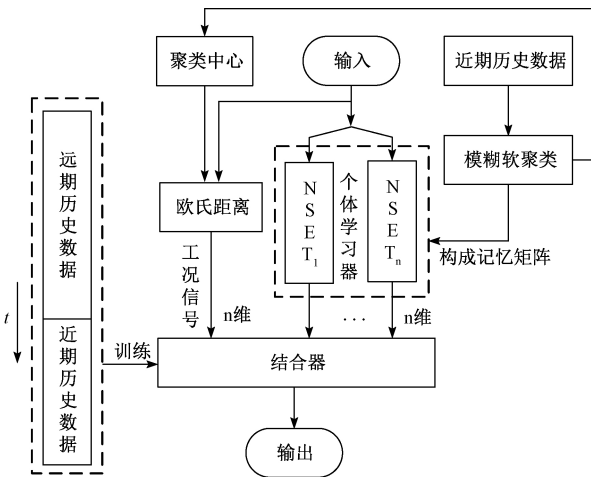


图 2 集成 NSET 状态变量建模方法

Fig.2 Ensemble NSET condition variable modeling method

如图 2 所示,集成 NSET 模型主要由个体学习器和结合器 2 部分构成,模型输入为设备当前的观测向量  $X_0$ , 输出为状态变量的预测值  $\hat{y}$ 。

个体学习器是基于近期历史数据和 SFCM 算法构造的 NSET 模型:基于 SFCM 算法将近期历史数据进行聚类,得到不同工况对应的聚类中心和各样本的类别标签;将属于同一工况的样本作为观测向量构成 NSET 的记忆矩阵,模型个数  $n$  和聚类个数  $T$  相等。

结合器是一个输入为  $2n$  维、输出为  $1$  维的参数回归模型,结构如下。

$$\hat{y} = f(\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n, d_1, d_2, \dots, d_n) \quad (7)$$

式中:  $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n$  为  $n$  个 NSET 模型的输出  $X_1^e, X_2^e, \dots, X_n^e$  中状态变量  $y$  的预测值;  $d_1, d_2, \dots, d_n$  为输入  $X_0$  与工况对应聚类中心的欧氏距离,具体如式(8)所示。

$$d_i = \sqrt{(C_i - X_0) \cdot (X_0 - C_i)^T} \quad (8)$$

式中:  $C_i$  为第  $i$  个工况对应的聚类中心。

如式(7)所示,结合器的输入由 2 部分构成:1)个体学习器的输出,与一般的集成学习模型相同;2)输入向量与聚类中心的欧式距离,是表征当前运行工况的辅助信号。结合器的具体算法应选择能充分拟合非线性函数关系的参数回归方法,如神经网络、支持向量机等。

集成 NSET 建模方法主要有如下 3 个优点。

1)对于集成学习模型的整体结构,以非参数方法作为个体学习器,继承了 NSET 算法泛化能力强、建模精度高的优势。以参数回归方法作为结合器,克服了 NSET 对记忆矩阵的依赖,能够充分利用记忆矩阵外大量的运行数据进行训练,在不影响实时性的同时进一步改善建模精度。

2)对于工况划分的聚类算法,SFCM 算法使属于不同工况过渡阶段的数据,能够被多个 NSET 个体学习器

利用,能够改善工况过渡阶段的建模精度,另一方面也能提高近期历史数据的利用率。

3)对于结合器的模型结构,额外构造了工况信号作为输入,使结合器能够直接感知设备的运行工况,从而更好的学习不同个体学习器在结合器内部的权重、进一步提高集成模型整体的精度。

### 2.3 集成 NSET 状态监测方法的实现流程

针对风电机组齿轮箱,集成 NSET 状态监测方法的具体实现流程如图 3 所示。

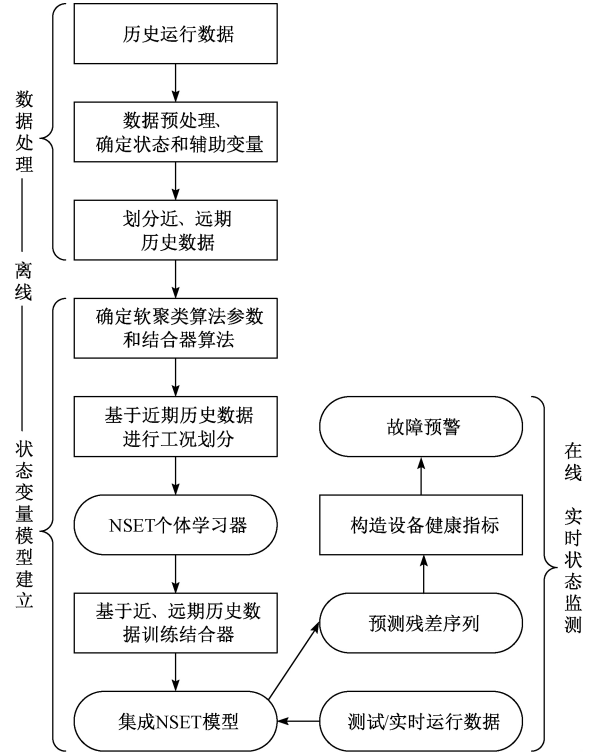


图 3 集成 NSET 状态监测方法的流程

Fig.3 Process of ensemble NSET condition monitoring method

如图 3 所示,集成 NSET 状态监测方法可分为数据处理、状态变量模型建立和实时状态监测 3 部分。

1)数据处理:从 SCADA 系统中,导出一段机组处于正常运行状态的历史数据;对运行数据进行异常值检测等数据预处理,并从运行参数中确定状态变量模型的输出(状态)变量和输入(辅助)变量;将处理后的训练数据分为近期和远期历史数据 2 部分。

2)状态变量模型建立:基于聚类有效性指标,设置 SFCM 算法的聚类个数  $T$  和重叠度  $\delta$ ,并选择结合器具体的参数回归算法;依照图 2 所示流程,基于近、远期历史数据,逐一进行工况划分、个体学习器构造、结合器训练等,建立集成 NSET 状态变量模型。

3)实时状态监测:基于 SCADA 系统的实时数据或测试样本,计算状态变量预测值和实际测量值间的预测残

差;将残差序列进行处理和变换,构造为有一定实际意义的、能直观反映设备运行状态的健康指标;最后,基于健康指标和阈值等,实现对齿轮箱运行状态的实时监测和故障预警。

### 3 算例及分析

#### 3.1 研究对象及数据处理

本文的研究对象为福建省某风场一台额定功率为2 MW的双馈式风电机组,型号为 Vestas V90,切入风速4 m/s,切出风速25 m/s,额定风速12 m/s,齿轮箱的结构为二级螺旋齿轮和一级行星齿轮,SCADA系统的采样周期为10 min。该机组于2016年7月13日10:20发现齿轮箱故障导致停运,有效运行数据的最后记录时间为2016年6月23日9:50。

从SCADA系统中导出2016年2月23日9:50~6月23日9:50,共16 193组运行数据。数据中可用的运行参数有风速、发电机转速、叶轮转速、风向角、环境温度、无功功率、有功功率和齿轮箱轴承温度8个。剔除属性缺失、有功功率小于或等于0、风速小于切入或大于切出风速的数据点,并基于拉依达准则去噪后,剩余10 250组有效数据参与试验。

齿轮箱轴承是齿轮箱主轴的载体,据统计约有50%的齿轮箱故障与轴承的选型、润滑和使用有关<sup>[25]</sup>。在可用运行参数中,齿轮箱轴承温度可以直接反映齿轮箱整体的运行状态,故将其作为模型输出即状态变量。参与建模数据中,8个运行参数的变化范围及其与齿轮箱轴承温度的皮尔逊相关系数如表1所示。

表1 运行参数的变化范围和相关分析

Table 1 Range and relevant analysis of the operation parameters

运行参数	变化范围	相关系数
风速/m·s <sup>-1</sup>	[4.00, 14.14]	0.846
发电机转速/r·min <sup>-1</sup>	[1 001.3, 1 693.0]	0.928
叶轮转速/r·min <sup>-1</sup>	[9.00, 15.02]	0.927
风向角/°	[13.5, 338.0]	-0.151
环境温度/°C	[3.0, 30.0]	-0.214
无功功率/var	[-1.54, 3.92]	-0.093
有功功率/kW	[26.1, 2 083.2]	0.829
齿轮箱轴承温度/°C	[43.63, 63.01]	1

由表1可知,轴承温度与风速、发电机转速、有功功率等存在着较强的正相关关系,符合风速增大导致叶轮、发电机等设备的转速和载荷增大,进而使齿轮箱轴承温

度升高的正常特性。风向角与变桨和偏航系统、无功功率与并网过程的关系密切,两者对齿轮箱的影响较小,故相关性较差。尽管轴承温度与环境温度的相关性一般,但当功率和风速相同时,环境温度的不同也会导致机组的运行状态有较大差异,在工况划分和状态监测时需要进行参照。

结合以上分析,确定辅助变量为风速、发电机转速、叶轮转速、环境温度和有功功率共5个。将运行数据进行min-max归一化以消除量纲影响后,公式如式(9)所示。

$$x'_{ij} = \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} \quad (9)$$

式中: $x_{ij}$ 为变量 $j$ 对应的第 $i$ 个数据; $\max(x_j)$ 和 $\min(x_j)$ 分别为变量 $j$ 的最大值和最小值; $x'_{ij}$ 为归一化后的数据。

以1~6 000号样本(时间跨度为2016年2月23日9:50~5月8日21:20)作为训练样本,用于建立状态变量模型;6 001~7 000号(2016年5月8日21:30~5月17日7:10)为第1批测试样本,用来评价状态变量模型的性能;6 001~10 250号(2016年5月8日21:30~6月23日9:50)为第2批测试样本,用来测试状态监测方法的故障预警能力。

#### 3.2 模糊软聚类算法的参数设置

确定聚类个数的常用方法是使用聚类有效性指标<sup>[26]</sup>。误差平方和 $I_{SSE}$ 是常用的聚类有效性指标,计算公式如式(10)所示。

$$I_{SSE} = \sum_{i=1}^T \sum_{x \in X_i} d^2(C_i, x) \quad (10)$$

式中: $T$ 为聚类个数; $X_i$ 为子类别; $C_i$ 为子类别对应的聚类中心; $x$ 为子类别对应的样本。

确定聚类个数的具体过程为:给定某一确定的聚类个数范围,绘制相应的 $I_{SSE}$ 曲线;曲线的拐点表示再增加类别时,误差平方和的减小幅度相对较小,故将拐点对应的个数作为最佳聚类个数<sup>[27]</sup>。确定聚类个数 $T$ 后,根据文献<sup>[21]</sup>可取重叠度 $\delta=0.1T$ 。

以后2 000组训练样本和所有6个运行参数参与聚类。设置聚类个数的范围为2~6,基于式(10)绘制对应的 $I_{SSE}$ 曲线,如图4所示。

从图4可以看出,聚类个数为3时, $I_{SSE}$ 曲线到达拐点,故确定FCM和SFCM的 $T=3, \delta=0.3$ 。

#### 3.3 齿轮箱轴承温度建模

为了横向对比集成NSET与参数回归方法和常规NSET,共建立以下3类共12个模型。

1) 参数方法:最小二乘支持向量机(least squares support vector machine, LSSVM),核函数为高斯函数。高斯过程回归(Gauss process regression, GPR),核函数为平方指数函数。回归树(regression tree, RT),结构为完全

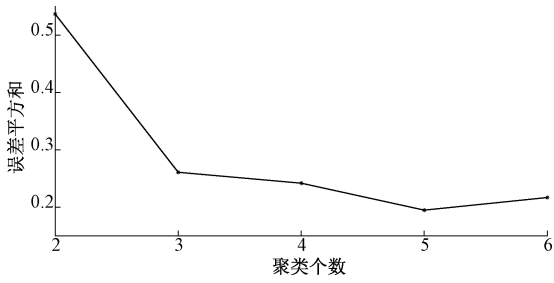


图 4 误差平方和  $I_{SSE}$  曲线

Fig.4 The curve of sum of squared error  $I_{SSE}$

二叉树。以上算法的实现,依托 MATLAB 的 Regression Learner 工具箱。

2) NSET:分别以后 1 000、2 000、3 000 组训练样本,构成规模不同的记忆矩阵,记作 NSET1、NSET2 和 NSET3,共 3 个。

3) 集成 NSET:分别以后 1 000、2 000 组训练样本为近期历史数据、其余样本为远期历史数据,基于 FCM 建立集成 NSET 模型,记作 FCM-EN1\*、FCM-EN2\* (结合器输入不含工况信号)、FCM-EN1、FCM-EN2 (含工况信号),共 4 个;基于 SFCM 建立集成 NSET 模型,记作 SFCM-EN1、SFCM-EN2 (含工况信号),共 2 个。结合器为 GPR 算法,核函数为平方指数函数。

基于训练样本和第 1 批测试样本,对各状态变量模型进行训练和测试,结果如表 2 所示。表 2 中,训练和测试误差的误差函数为均方根误差,计算时间为计算 1 000 组测试样本所需的总时间。

表 2 各状态监测方法的训练及测试结果

Table 2 Training and test results of the condition monitoring methods

建模方法	训练误差	测试误差	训练时间/s	计算时间/s
LSSVM	$7.7 \times 10^{-2}$	$8.0 \times 10^{-2}$	7.1	0.24
GPR	$7.7 \times 10^{-2}$	$7.2 \times 10^{-2}$	98.0	0.11
RT	$8.1 \times 10^{-2}$	$7.6 \times 10^{-2}$	2.5	0.07
NSET1	-	$8.5 \times 10^{-3}$	-	15.1
NSET2	-	$1.4 \times 10^{-3}$	-	75.9
NSET3	-	$1.2 \times 10^{-3}$	-	192.5
FCM-EN1* (无工况信号)	$1.7 \times 10^{-2}$	$1.6 \times 10^{-2}$	149.4	6.4
FCM-EN2* (无工况信号)	$1.3 \times 10^{-2}$	$9.4 \times 10^{-3}$	137.8	29.2
FCM-EN1	$8.4 \times 10^{-4}$	$1.4 \times 10^{-3}$	122.4	4.0
FCM-EN2	$5.3 \times 10^{-4}$	$7.7 \times 10^{-4}$	126.2	21.0
SFCM-EN1	$8.2 \times 10^{-4}$	$8.3 \times 10^{-4}$	133.2	6.2
SFCM-EN2	$6.8 \times 10^{-4}$	$6.3 \times 10^{-4}$	117.7	28.1

由表 2 可知,从建模精度的角度,参数方法的训练和测试误差相对较大,与 NSET 和集成 NSET 有数量级上的差距;NSET 的精度相对较高、通过扩展记忆矩阵能提高精度,但幅度相对有限且在相同矩阵规模下,精度低于集成 NSET;FCM-EN1、FCM-EN2 的精度要明显高于 FCM-EN1\*、FCM-EN2\*,说明加入工况信号后,集成 NSET 的精度有了较为明显的改善;SFCM-EN1 和 SFCM-EN2 的精度高于 FCM-EN1 和 FCM-EN2,说明应用 SFCM 算法后,集成 NSET 的建模精度有了进一步的提高。

从实时性的角度,由于训练后模型参数已经固化,参数方法的训练和计算时间均较短;NSET 方法的计算时间较长,且随着记忆矩阵规模的扩大计算时间大幅度增长,实时性较差;集成 NSET 方法的计算时间相比 NSET 要更短,说明将较大规模的记忆矩阵划分成多个较小的矩阵后,能够提高计算速度;引入工况信号和使用 SFCM 算法后,集成 NSET 方法的训练和计算时间并未出现较大幅度的增长,说明工况信号和聚类算法对实时性的影响不大。

因此,基于 SFCM 的集成 NSET 模型,建模精度较高且实时性较好,相比常规 NSET 有着较大的优势。

### 3.4 状态变量模型的预测残差和预警阈值

从表 2 中选择 GPR、NSET3 和 SFCM-EN2 模型(选取 GPR 的目的是对比集成 NSET 和仅使用结合器算法间的性能),使用第 2 批测试数据进行测试。基于滑动窗口法(窗口长度为 100),将残差进行均值处理,得到的残差均值序列如图 5 所示。

由图 5 可知,依据 NSET3 和 SFCM-EN2 曲线的趋势,可分为 0~1 500、1 500~3 000、3 000~最终 3 个阶段,分别表示齿轮箱正常运行(阶段 I)、发生早期故障(阶段 II)和故障恶化(阶段 III)3 种典型运行状态。

曲线在阶段 I 的变化相对较为平缓,故以阶段 I 的残差均值为基准,确定各模型的预警阈值  $E_y$  如式(11)所示。

$$E_y = \pm k \cdot \max_{0 \leq i < 1500} \{ |\bar{\varepsilon}(i)| \} \quad (11)$$

式中: $k$  是阈值系数,本文取  $k=1.1$ ;  $\bar{\varepsilon}$  为残差的滑动窗口均值。GPR、NSET3 和 SFCM-EN2 的阈值  $E_y$  分别为  $6.8 \times 10^{-2}$ ,  $1.1 \times 10^{-3}$  和  $4.9 \times 10^{-4}$ ,在图 5 中以虚线表示。

GPR 曲线的波动情况并未随时间推移发生显著改变,仅在阶段 II 短暂超过预警阈值,较难判断齿轮箱的运行状态是否发生变化、不能实现故障预警。其原因主要是建模精度相对较低,导致状态变量模型的灵敏度不够、无法有效反映设备运行状态的变化。

阶段 II 的前段,NSET3 和 SFCM-EN2 曲线均逐渐偏离横轴直至超过预警阈值,能反映出齿轮箱可能发生了早期故障,但 SFCM-EN2 超过阈值的时刻更早,说明更

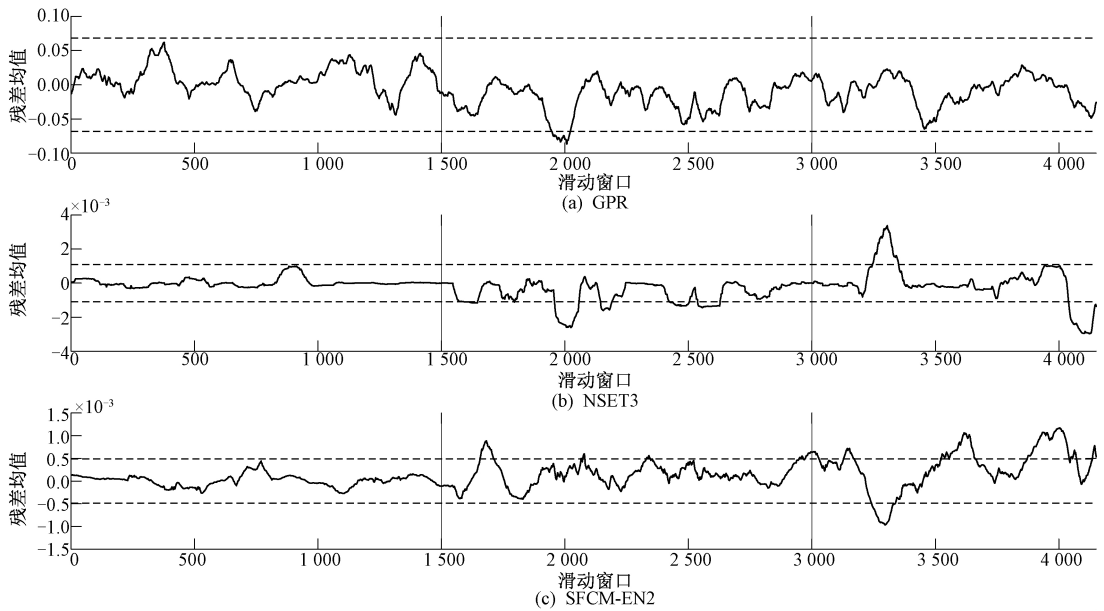


图5 3种状态变量模型的预测残差均值曲线

Fig.5 Prediction residual mean curves of the three condition variable models

早的检测出运行异常、灵敏度相对更高。阶段 II 的中后段, NSET3 曲线逐渐趋于平缓, 而 SFCM-EN2 曲线逐渐偏离横轴直至再次超过预警阈值。说明相比常规 NSET, 集成 NSET 模型残差均值的变化更能反映早期故障持续且不可逆的劣化趋势。

对于阶段 III, NSET3 曲线有显著的孤立峰值, 可能会被判断为虚警。而 SFCM-EN2 曲线继续保持大幅度波动, 多次且长时间超过预警阈值, 反映出早期故障发生了进一步的恶化, 更符合故障发展的机理。

结合以上分析, 常规 NSET 能在一定程度上反应齿轮箱故障的发展情况, 但残差均值变化的连续性较差、存在孤立的峰值; 而集成 NSET 的灵敏度相对更高, 能够更连续、更准确地反映“正常运行-发生早期故障-早期故障进一步恶化”的齿轮箱故障发展趋势, 故障预警能力相对更优。

### 3.5 基于健康指数的齿轮箱状态监测

通过图 5 的残差均值和预警阈值, 能够基本实现齿轮箱的状态监测。但残差均值的实际意义不强, 且必须配合阈值才能实现故障预警, 不利于运行人员直观感知和判断设备的运行状态。因此, 有必要结合残差和阈值, 构造能直观反映设备运行状态的健康指数。

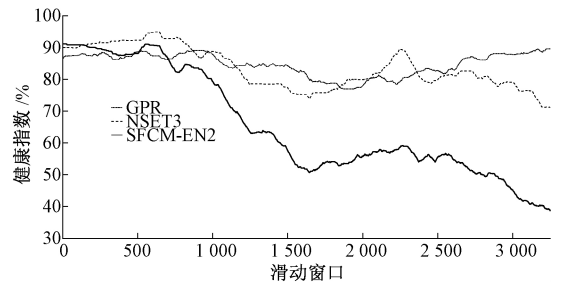
本文基于式 (12) 构造健康指数  $I$ 。

$$I = \left(1 - \frac{N}{M}\right) \times 100\% \quad (12)$$

式中:  $N$  为滑动窗口残差序列中超过预警阈值  $E_y$  的残差个数;  $M$  为残差序列的总个数即滑动窗口的长度;  $I$  的单位为百分比且值域固定在  $[0, 100\%]$ 。

当滑动窗口中超过阈值的残差较少时, 健康指数  $I$  相对较高, 说明状态变量模型的预测精度较高、设备的健康程度相对较好; 当超过阈值的残差较多时,  $I$  会显著下降, 说明状态变量的预测值已显著偏离预测值、运行状态出现异常。

基于 3.4 节中得到的预警阈值  $E_y$ , 分别计算 GPR、NSET3 和 SFCM-EN2 在第 2 批测试数据上的健康指数  $I$ , 如图 6 所示。(为了提高曲线的平滑度, 取滑动窗口的长度  $M=1000$ )。

图6 3种状态变量模型的健康指数  $I$ Fig.6 Health index  $I$  of the three condition variable models

由图 6 可知, 在使用相同方法构造阈值和健康指数的前提下, 3 种模型健康指数曲线的变化情况有着较大的不同: GPR 曲线的变化幅度相对较小, 且最终齿轮箱发生故障时的健康指数反而高于起始点, 故障预警的效果较差; NSET3 曲线发生了较为明显的变化, 健康指数从最初的约 90% 最终降至约 70%, 但变化幅度相对较小且单调性较差; SFCM-EN2 曲线的变化幅度相对较大, 从最

初的约 90% 最终降至约 40% 且单调性较好,能够更准确的反映齿轮箱故障的发展趋势、故障预警能力相对更优。

相比残差均值,健康指数能够更为直观的反映设备运行状态的变化情况、更适宜于实际工程应用。但是,健康指数无法反映状态变量的偏离方向,即测量值相比预测值是偏高还是偏低,需要结合残差均值或故障诊断方法进一步判断具体故障类型。

## 4 结 论

本文提出了一种基于集成 NSET 和模糊软聚类的风电机组齿轮箱状态监测方法,并基于 SCADA 系统的齿轮箱实际故障数据进行了对比验证。结果表明,在由相同规模数据构成记忆矩阵的前提下,集成 NSET 方法的建模精度和实时性均优于常规 NSET,且模糊软聚类的应用进一步提高了精度;在故障预警方面,通过状态变量模型的残差均值和基于残差构造的健康指数,集成 NSET 相比常规方法,能更灵敏地响应齿轮箱发生的早期故障、更准确地反应出故障的发展趋势。

## 参考文献

- [ 1 ] LIN Y G, TU L, LIU H W, et al. Fault analysis of wind turbines in China[J]. *Renewable and Sustainable Energy Reviews*, 2016, 55: 482-490.
- [ 2 ] 金晓航, 孙毅, 单继宏, 等. 风力发电机组故障诊断与预测技术研究综述[J]. *仪器仪表学报*, 2017, 38(5): 1041-1053.  
JIN X H, SUN Y, SHAN J H, et al. Fault diagnosis and prognosis for wind turbines: An overview [J]. *Chinese Journal of Scientific Instrument*, 2017, 38 ( 5 ): 1041-1053.
- [ 3 ] 史训兵, 熊志刚, 李杨宗. 基于在线油液磨粒检测的风电机组齿轮箱磨损状态监控[J]. *机械传动*, 2014, 38(10): 74-77.  
SHI X B, XIONG ZH G, LI Y Z. Wear condition monitoring of wind turbine gearbox based on the online oil abrasion detection [ J ]. *Journal of Mechanical Transmission*, 2014, 38(10): 74-77.
- [ 4 ] 张星辉, 康建设, 高存明, 等. 基于 MoG-HMM 的齿轮箱状态识别与剩余使用寿命预测研究[J]. *振动与冲击*, 2013, 32(15): 20-25.  
ZHANG X H, KANG J SH, GAO C M, et al. Gearbox state identification and remaining useful life prediction based on MoG-HMM [ J ]. *Journal of Vibration and Shock*, 2013, 32(15): 20-25.
- [ 5 ] 杜勉, 易俊, 郭剑波, 等. 神经网络技术在风电机组 SCADA 数据分析中的应用研究[J]. *电网技术*, 2018, 42(7): 2200-2205.  
DU M, YI J, GUO J B, et al. Research on the application of neural networks on wind turbine SCADA data analysis [ J ]. *Power System Technology*, 2018, 42(7): 2200-2205.
- [ 6 ] 刘帅, 刘长良, 甄成刚, 等. 基于群体多维相似性的风机齿轮箱预警策略[J]. *仪器仪表学报*, 2018, 39(1): 180-189.  
LIU SH, LIU CH L, ZHEN CH G, et al. Fault warning strategy of wind turbines gearbox based on group multi-dimensional similarity [ J ]. *Chinese Journal of Scientific Instrument*, 2018, 39(1): 180-189.
- [ 7 ] HELBING G, RITTER M. Deep learning for fault detection in wind turbines [ J ]. *Renewable and Sustainable Energy Reviews*, 2018, 98: 189-198.
- [ 8 ] 郭鹏, INFIELD D, 杨锡运. 风电机组齿轮箱温度趋势状态监测及分析方法[J]. *中国电机工程学报*, 2011, 31(32): 129-136.  
GUO P, INFIELD D, YANG X Y. Wind turbine gearbox condition monitoring using temperature trend analysis [ J ]. *Proceedings of the CSEE*, 2011, 31(32): 129-136.
- [ 9 ] GUO P, BAI N. Wind turbine gearbox condition monitoring with AAKR and moving window statistic methods [ J ]. *Energies*, 2011, 4(11): 2077-2093.
- [ 10 ] 李大中, 毛小丽, 尹鹏娟, 等. 基于 NSET 的风力发电机后轴承温度预警方法[J]. *科学技术与工程*, 2016, 16(24): 205-209.  
LI D ZH, MAO X L, YIN P J, et al. The rear bearing temperature of early warning method in wind turbine based on NSET [ J ]. *Science Technology and Engineering*, 2016, 16(24): 205-209.
- [ 11 ] 任梦祎, 焦嵩鸣. 改进型非线性状态估计的制粉系统故障诊断[J]. *热力发电*, 2015, 44(12): 87-92.  
REN M Y, JIAO S M. Coal milling system fault diagnosis based on improved NSET [ J ]. *Thermal Power Generation*, 2015, 44(12): 87-92.
- [ 12 ] 李大中, 常城, 许炳坤. 基于样本优化的风电机组齿轮箱轴承温度预测[J]. *系统仿真学报*, 2017, 29(2): 374-380.  
LI D ZH, CHANG CH, XU B K. Wind turbine gearing temperature prediction based on sample optimization [ J ]. *Journal of System Simulation*, 2017, 29(2): 374-380.
- [ 13 ] 贾子文, 顾煜炯. 基于数据挖掘的风电机组齿轮箱运行状态分析[J]. *中国机械工程*, 2018, 29(6): 650-658.  
JIA Z W, GU Y J. Wind turbine gearbox state analysis based on data mining [ J ]. *China Mechanical Engineering*, 2018, 29(6): 650-658.
- [ 14 ] 杨婷婷, 张蓓, 吕游, 等. 基于 MSET 的电站风机故障预警技术研究[J]. *热能动力工程*, 2017, 32(9): 63-68.  
YANG T T, ZHANG B, LYU Y, et al. Research on early



- fault warning technology of fan in power plant based on MSET method[J]. Journal of Engineering for Thermal Energy and Power, 2017, 32(9): 63-68.
- [15] 刘长良, 闫萧. 基于工况辨识的风电机组齿轮箱状态监测[J]. 热能动力工程, 2016, 31(7): 41-46.  
LIU CH L, YAN X. Monitoring of the state of the gear box of a wind power generator unit based on the operating condition identification [J]. Journal of Engineering for Thermal Energy and Power, 2016, 31(7): 41-46.
- [16] GUO P, INFIELD D, YANG X. Wind turbine generator condition-monitoring using temperature trend analysis[J]. IEEE Transactions on Sustainable Energy, 2012, 3(1): 124-133.
- [17] GROSS K C, SINGER R M, WEGERICH S W, et al. Application of a model-based fault detection system to nuclear plant signals[C]. Proceedings of 9th International Conference on Intelligent Systems Application to Power System, 1997.
- [18] 郭鹏, 徐明, 白楠, 等. 基于 SCADA 运行数据的风电机组塔架振动建模与监测[J]. 中国电机工程学报, 2013, 33(5): 128-135.  
GUO P, XU M, BAI N, et al. Wind turbine tower vibration modeling and monitoring driven by SCADA data[J]. Proceedings of the CSEE, 2013, 33(5): 128-135.
- [19] BORDOGNA G, PASI G. Soft clustering for information retrieval applications[J]. Wiley Interdisciplinary Reviews Data Mining & Knowledge Discovery, 2015, 1(2): 138-146.
- [20] RAHMAN A, VERMA B. Novel layered clustering-based approach for generating ensemble of classifiers[J]. IEEE Transactions on Neural Networks, 2011, 22(5): 781-792.
- [21] LYU Y, LIU J ZH, YANG T T, et al. A novel least squares support vector machine ensemble model for NO<sub>x</sub> emission prediction of a coal-fired boiler [J]. Energy, 2013, 55(1): 319-329.
- [22] 李庆伟, 陈慧枫, 姚桂焱. 基于距离学习粒子群算法的 NO<sub>x</sub> 减排优化[J]. 动力工程学报, 2016, 36(5): 404-410.  
LI Q W, CHEN H F, YAO G H. NO<sub>x</sub> reduction optimization based on distance learning particle swarm algorithm [J]. Journal of Chinese Society of Power Engineering, 2016, 36(5): 404-410.
- [23] 赵文杰, 吕猛. 基于多 LS-SVM 集成模型的锅炉 NO<sub>x</sub> 排放量建模[J]. 电子测量与仪器学报, 2016, 30(7): 1037-1044.  
ZHAO W J, LYU M. NO<sub>x</sub> emission modeling of coal-fired boiler based on multi LS-SVM ensemble model[J]. Journal of Electronic Measurement and Instrumentation, 2016, 30(7): 1037-1044.
- [24] REN Y, ZHANG L, SUGANTHAN P N. Ensemble classification and regression-recent developments, applications and future directions [J]. IEEE Computational Intelligence Magazine, 2016, 11(1): 41-53.
- [25] 张继权, 赵洪山. 基于比例失效模型的风机齿轮箱轴承检修[J]. 电源技术, 2011, 35(5): 570-573.  
ZHANG J Q, ZHAO H SH. Maintenance decision policy of wind-power generator gearbox bearings based on proportional hazards model[J]. Chinese Journal of Power Sources, 2011, 35(5): 570-573.
- [26] 周开乐, 杨善林, 丁帅, 等. 聚类有效性研究综述[J]. 系统工程理论与实践, 2014, 34(9): 2417-2431.  
ZHOU K L, YANG SH L, DING SH, et al. On cluster validation[J]. Systems Engineering-Theory and Practice, 2014, 34(9): 2417-2431.
- [27] 张斌, 庄池杰, 胡军, 等. 结合降维技术的电力负荷曲线集成聚类算法[J]. 中国电机工程学报, 2015, 35(15): 3741-3749.  
ZHANG B, ZHUANG CH J, HU J, et al. Ensemble clustering algorithm combined with dimension reduction techniques for power load profiles[J]. Proceedings of the CSEE, 2015, 35(15): 3741-3749.

### 作者简介



王梓齐, 2017 年于华北电力大学获得学士学位, 现为华北电力大学在读博士研究生, 主要研究方向为风电机组状态监测与故障预警。

E-mail: wangziqincepu@163.com

**Wang Ziqi** received his B. Sc. degree from North China Electric Power University in 2017. He is currently a Ph. D. candidate at North China Electric Power University. His main research interests include wind turbines condition monitoring and fault warning.



刘长良 (通信作者), 分别于 1985 和 1990 年于华北电力学院获得学士和硕士学位, 2002 年于华北电力大学获得博士学位, 现为华北电力大学教授、博士生导师, 主要研究方向为风电机组故障诊断, 火电机组节能、建模与仿真等。

E-mail: 13603123513@163.com

**Liu Changliang** (Corresponding author) received his B. Sc. degree and M. Sc. degree both from North China Electric Power College in 1985 and 1990, and received his Ph. D. degree from North China Electric Power University in 2002. He is currently a professor and Ph. D. supervisor at North China Electric Power University. His main research interests include wind turbine fault diagnosis, thermal power unit modeling and simulation, etc.