

DOI: 10.19783/j.cnki.pspc.191560

# 基于 AdaBoost-SAMME 的风力发电机组变桨异常识别系统

樊帅<sup>1</sup>, 唐群先<sup>2</sup>

(1. 北京金风科创风电设备有限公司, 北京 100176; 2. 对外经济贸易大学, 北京 100029)

**摘要:** 为了实现对风力发电机组变桨系统的预见性维护, 提出了在机组健康管理平台上开发变桨异常识别系统的设计思路。在模型算法构建过程中, 提出将变桨速度的频域特征作为分类特征, 实现了对异常征兆的精确刻画。利用 AdaBoost-SAMME 算法将变桨速度分为高频、低频以及正常三类, 与人工神经网络、支持向量机、随机森林等五种算法对比发现, 即使各类样本数量严重不均衡, AdaBoost-SAMME 算法的准确率、查准率、查全率及 G-mean 等评价指标也都优于其他算法。为了提高系统的自学习能力, 提出了一种基于欧式距离的新样本判断方法, 以此自动扩增训练样本规模。应用实例验证表明, 基于 AdaBoost-SAMME 算法的变桨异常识别系统具有显著的分类效果和良好的稳定性, 解决了不同机型的变桨速度频域特征普适性规律不明显, 常规逻辑判断方法无法识别变桨系统异常的技术难题。该系统实现了在故障前识别出异常的预警功能, 能够指导现场人员开展预见性维护, 提高机组的可靠性和可利用率。

**关键词:** 风力发电机组; 变桨系统; 异常识别; AdaBoost-SAMME; 特征选择

## Wind turbine pitch anomaly recognition system based on AdaBoost-SAMME

FAN Shuai<sup>1</sup>, TANG Qunxian<sup>2</sup>

(1. Beijing Goldwind Science & Creation Windpower Equipment Co., Ltd., Beijing 100176, China;  
2. University of International Business and Economics, Beijing 100029, China)

**Abstract:** To realize predictive maintenance of the pitch system in wind turbines, a design idea is presented whereby a pitch anomaly recognition system can be developed on a wind turbines health management platform. In constructing the model algorithm, the frequency domain characteristics of pitch velocity are suggested as classification features, and precise characterization of abnormal symptoms is achieved. The pitch velocity is divided into three categories: high, low and normal frequency with AdaBoost-SAMME as a classification algorithm. Comparing with five algorithms including an artificial neural network, a support vector machine and a random forest, it is found that the AdaBoost-SAMME algorithm is superior in accuracy, precision, recall and G-mean index although a serious disproportion exists in the number of samples. In order to improve the self-learning ability of the recognition system, a new sample judgment method based on Euclidean distance is proposed to automatically expand the training sample size. The application example demonstrates that the pitch anomaly recognition system using the AdaBoost-SAMME algorithm possesses noticeable classifying quality and stability. It solves the technical problem that for the general adaptive law of pitch velocity frequency domain characteristics in different models is implicit, and a conventional logic judgment method cannot be used to identify a pitch system abnormality. The system achieves the early warning function of identifying anomalies before faults. This can guide field personnel to practice predictive maintenance and improve the reliability and utilization rate of wind turbines.

This work is supported by National Key Research and Development Program of China (No. 2016YFB1000705).

**Key words:** wind turbines; pitch system; anomaly recognition; AdaBoost-SAMME; feature selection

## 0 引言

近年来, 随着全球范围内新能源产业的快速发

展, 风电装机容量正在迅猛增长。但是风力发电机组却通常位于位置偏远、环境恶劣、工况复杂的地方, 对风力发电机组实施常规计划性维护, 将在很大程度上增加人力和财力投入<sup>[1-2]</sup>。而分析机组运行数据, 在故障之前发现异常状态, 实施预见性维护,

基金项目: 国家重点研发计划项目资助(2016YFB1000705)

提高机组的可靠性与可利用率<sup>[3-4]</sup>, 将是风力发电行业亟待解决的一个新问题。

风力发电机组的桨叶作为整机的重要部件之一, 一方面可以改变机组的转速和功率输出; 另一方面还可以改变风机受力。变桨距机组的桨叶角度<sup>[5]</sup>(即桨距角)是可变的, 旋转角度通常在 0~90°。桨距角是通过编码器组件测量的, 变桨速度<sup>[6-7]</sup>是桨距角在单位时间内的变化速率, 是一个矢量信号, 具有方向性。该变量反映了桨距角的变化趋势, 是分析变桨系统运行情况的一项重要数据。

目前已经应用的针对风力发电机组变桨系统的异常判断大都是故障级别的, 即变桨系统运行参数超过事先设置的故障阈值时, 执行故障停机。这种异常判断方法虽然能够对变桨设备起到一定的预警保护作用, 但是无法从早期识别出变桨系统的异常状态。而且如果机组长期运行在亚健康状态, 会导致机组出力下降, 甚至变桨电机等关键部件出现损坏<sup>[8-9]</sup>。

在阈值判断之外, 基于 SCADA(Supervisory Control and Data Acquisition)数据, 自适应模糊神经网络、非线性状态估计以及核主成分分析等机器学习算法在风机变桨系统异常识别与故障诊断方面的应用也在逐渐地探索与推进。文献[10]将 Fisher 判别分析法应用到变桨系统故障诊断中, 构建一种基于 FDA(Fisher Discriminant Analysis, FDA)贡献图的故障源分离方法。文献[11]利用非线性状态评估方法, 建立起风机变桨系统所有正常状态的健康模型, 当变桨系统故障时, 模型预测值与正常状态会出现偏差, 依据事先选定的每一个运行参数对偏差的影响确定故障根因。文献[12]提出采用自适应神经模糊推理方法来拟合风速等运行参数与机组运行状态之间的规律, 通过对比实际输出与期望输出的差异进行故障诊断。针对 SCADA 系统提供的故障信息较多, 不易迅速定位故障类别的实际问题, 文献[13]利用非线性状态估计技术作为数据挖掘方法, 在多个故障信息中识别出主、次故障, 指导检修人员确定检修顺序。文献[14]建立了基于风速作为输入的变桨系统特征参量的支持向量机预测模型, 并提出了基于多特征参量距离的变桨系统异常识别方法。文献[15]提出了基于核主成分分析的变桨系统故障诊断方法。

本文提出了在机组健康管理平台上开发变桨异常识别系统的设计思路, 以此实现将占用计算资源的复杂分析作业部署在基于大数据集群的机组健康管理平台。在模型算法构建过程中, 提出将变桨速度的频域特征作为分类特征。利用 AdaBoost-

SAMME 算法将变桨速度分为高频、低频以及正常三类, 为了验证该算法在解决变桨系统异常识别问题上的可行性和有效性, 选取正确率、查准率、查全率等六项评价指标, 对 AdaBoost-SAMME 算法与人工神经网络等五种算法进行了对比分析。此外, 为了提高系统的自学习能力, 还提出了一种基于欧式距离的新样本判断方法。本文提出的基于 AdaBoost-SAMME 算法的风机变桨异常识别系统可以解决不同机型的变桨速度频域特征普适性规律不明显, 常规逻辑判断方法无法识别变桨系统异常的问题, 并可实现在故障前识别出异常的预警功能。

## 1 风机变桨异常信号识别系统

### 1.1 变桨速度异常信号分析

通过分析异常案例, 发现变桨系统出现异常时, 通常异常征兆在变桨速度信号中有直接体现。正常的变桨速度时域曲线呈现一种随机状态, 不含有明显的高频或者低频成分, 如图 1 所示, x 轴表示采样点数, 采样周期为 0.02 s; y 轴表示变桨速度, 单位是度/秒。异常的变桨速度可以分为高频与低频两类, 前者在时域中最明显的特征是变桨速度小幅、密集波动, 如图 2 所示; 后者则是变桨速度波动幅值较大, 且周期性特征显著, 如图 3 所示。

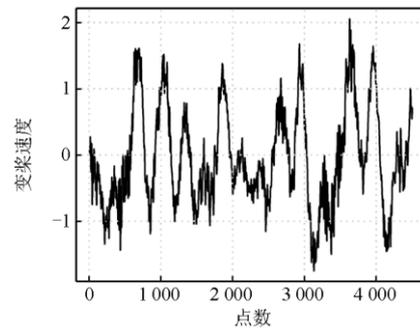


图 1 变桨速度正常信号

Fig. 1 Normal signal of pitch velocity

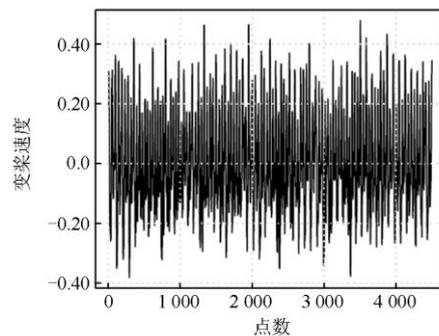


图 2 变桨速度高频信号

Fig. 2 High frequency signal of pitch velocity

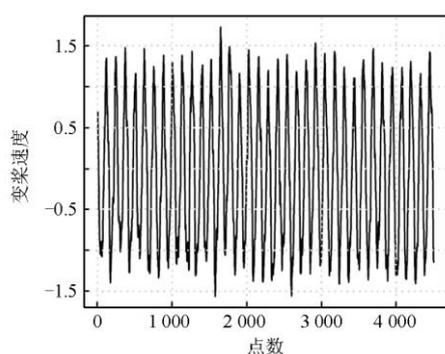


图 3 变桨速度低频信号

Fig. 3 Low frequency signal of pitch velocity

考虑到变桨系统出现异常时变桨速度信号中呈现的上述特征与变桨速度信号的频域特征强相关, 因此提出将变桨速度的频域特征, 即变桨速度频谱

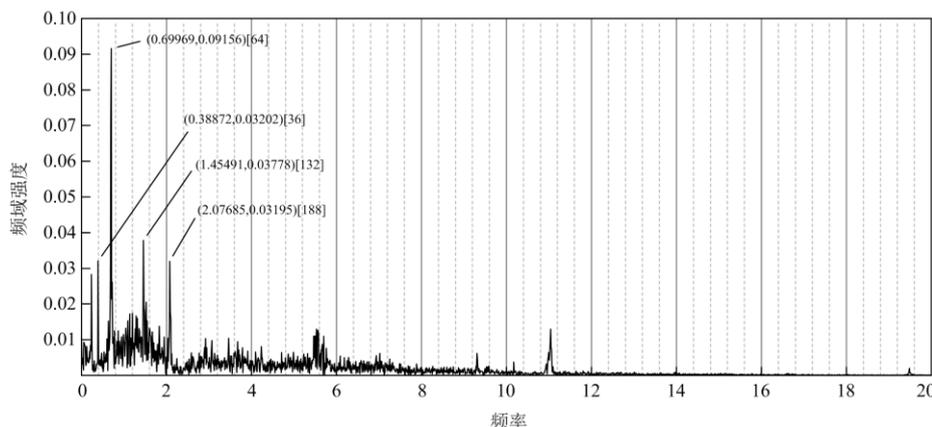


图 4 变桨速度频谱分析图

Fig. 4 Spectrum analysis chart of pitch velocity

此外, 不同机型的风力发电机组的变桨系统出现异常时, 变桨速度的频域特征没有普适性规律, 即变桨速度频谱分析后的主要频率点及其对应的频域强度不尽相同, 用常规的逻辑判断方法很难实现对变桨系统异常的识别与预警。鉴于此, 利用基于 SAMME 分类器权值更新策略的 AdaBoost 算法作为分类算法, 实现对变桨异常中高频、低频类的准确识别。

## 1.2 变桨异常识别系统原理

变桨异常识别系统作为风力发电机组健康管理平台的一个应用系统, 能够从机组的运行数据中, 尽早判断机组的变桨信号中是否含有明显的异常频率成分, 进而发出预警信息, 指导现场运维人员开展预见性维护。图 5 为风力发电机组健康管理平台的结构示意图, 分为物理层、数据层以及应用层。

分析后的 top3 频率及其对应的频域强度作为识别依据, 实现对变桨系统异常征兆的精确刻画。

研究中发现, 在一个频域强度较大的频率点附近, 通常会存在另一个较强的频率点, 而这两个异常频率的信号源一致。为了避免邻域信号的干扰, 在选定了 top1 信号的频率和频域强度后, top1 信号左右 0.5 Hz 范围内的信号不计入 top2 排序; top3 的排序沿用同样的方法。如图 4 所示, 图中标注的小括号内有两个数字, 第一个是频率值, 第二个是频域强度, 中括号里的数字表示点序号。由于第 36 点的频率值 0.388 72 Hz 与第 64 点的频率值 0.699 69 Hz 之间的差值为 0.310 97 Hz, 小于 0.5 Hz, 所以第 36 点所对应的频率信号是第 64 点次生而来。在按照频域强度由大到小排序时, 可以将第 36 点剔除, 频域强度 top3 排序依次为: 第 64 点 > 第 132 点 > 第 188 点。

物理层负责将风机运行数据通过交换机和网关等传回服务器。存储在服务器中的数据被大数据平台调用后, 分发给相应的分析模型; 服务推送引擎能够将应用层中模型输出的预警、诊断及评估等结果推送给用户。在应用层, 模型指的是具有预警、诊断及评估等功能的算法程序, 它的输入是大数据平台传送来的数据, 输出形式多样, 可以是分析报告, 也可以是可视化交互界面, 还可以是预警、诊断及评估的数据结果。本文中的变桨异常识别系统位于图 5 应用层的模型位置。

变桨异常识别系统的识别流程如图 6 所示。

首先, 初始化模型参数, 利用训练样本训练模型, 得到一个当前最优的识别模型。其次, 从风力发电机组健康管理平台获取数据文件, 提取数据文件中的变桨速度信号, 对变桨速度 FFT 变换求取其

频域特征量,即变桨速度频谱分析后的 top3 频率及其对应的频域强度,提取时需剔除主频率左右 0.5 Hz 范围内的干扰信号。然后,将获取的特征量送入已经训练并保存好的模型中进行分类。最后,经模型计算后得到识别结果。

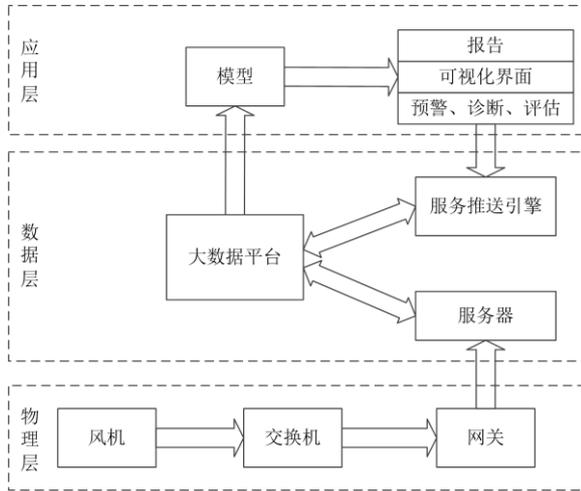


图 5 风力发电机组健康管理平台结构示意图  
Fig. 5 Schematic diagram of health management platform for wind turbines

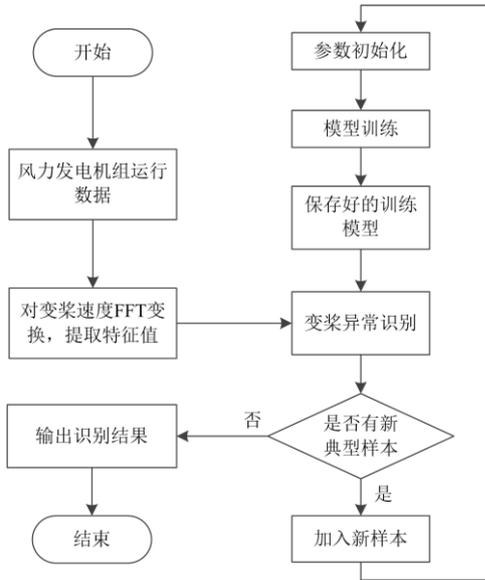


图 6 变桨异常识别流程图  
Fig. 6 Flow chart of pitch abnormal signal recognition

在上述过程中,如果有新的典型样本进入系统,则该样本将会被作为新的训练样本,系统也会重新训练模型,以此提高模型的自学习能力,式(1)给出了基于欧式距离的新样本判断方法。

$$d = \sqrt{\sum_{i=1}^3 (x_i - \varepsilon_i)^2} \quad (1)$$

式中:  $(x_1, x_2, x_3)$  表示样本的 top3 频率值;  $(\varepsilon_1, \varepsilon_2, \varepsilon_3)$  表示类别中心,类别中心通过计算各类别训练样本的 top3 频率均值获得。在本实例中,样本分为 3 个类别,对应有 3 个  $\varepsilon$  向量,计算一个样本与各类别中心的距离时结果也将是 3 个  $d$ 。对于每一个距离数值  $d$ ,当  $d > \sigma_1$  &  $d > \sigma_2$  &  $d > \sigma_3$  时,表示当前样本与各类别中心距离均超出上限,当前样本为新典型样本,  $\sigma_i (i=1,2,3)$  表示设定阈值,其设置的原则是能够覆盖当前样本集中的所有数据,即由当前样本集中的数据代入式(1),取最大值得到各类别中心的设定阈值。

## 2 AdaBoost算法原理

AdaBoost 作为一类集成学习算法,通过将多个基学习进行组合,进而构建一个集成的分类模型,其组合机制大大降低了识别误差<sup>[16-17]</sup>。AdaBoost 基本思想是:首先,构造一个弱分类器,将训练集送入该弱分类器,并依据分类结果给每个样本赋予一个权值,原则是如果该样本被正确分类则减小其权值,而如果未被正确分类则增大其权值。其次,根据每个样本的权值进行抽样,构建第二个弱分类器,最大程度地使权值较大的样本能够被正确分类。然后,根据前两个分类器的结果重新调整各样本权值,构建第三个分类器。重复上述操作,可以构建多个弱分类器。最后,采用投票的方式找出该样本多数情况下的分类结果,并作为最终分类结果<sup>[18-19]</sup>。

AdaBoost.M1 是 AdaBoost 算法的一个典型例子,以下将具体说明 AdaBoost.M1 的计算步骤<sup>[20-21]</sup>。

为训练集中每个样本的权值设置初始值,如式(2)所示。

$$\beta_i = \frac{1}{n}, i=1, 2, \dots, n \quad (2)$$

式中,  $n$  为样本个数。

以下计算需要持续  $m$  轮(式(3)一式(6)),在每一轮计算中,依据每个样本的权值  $\beta_i$  构建该轮的分类器  $h_j(x)$ ,之后,计算该分类器加权的错误率,如式(3)所示。

$$err_j = \frac{\sum_{i=1}^n \beta_i I(y_i \neq h_j(x_i))}{\sum_{i=1}^n \beta_i} \quad (3)$$

式中,  $x$  表示输入,  $x_i \in R^d$ 。

函数  $I(x)$  定义为

$$I(x) = \begin{cases} 1, & \text{如果 } x = \text{TRUE} \\ 0, & \text{如果 } x = \text{FALSE} \end{cases} \quad (4)$$

该分类器的最优权值可表示为

$$\omega_j = \frac{1}{2} \ln \frac{1 - err_j}{err_j} \quad (5)$$

每个样本的权值按照式(6)更新。

$$\beta_i = \beta_i \exp(2\omega_j I(y_i \neq h_j(\mathbf{x}_i))), i = 1, 2, \dots, n \quad (6)$$

经过  $m$  轮计算后, 顺次得到  $m$  个分类器。最终的分分类器聚合为

$$f(x) = \text{sign} \left( \sum_{j=1}^m \omega_j h_j(\mathbf{x}) \right) \quad (7)$$

AdaBoost-SAMME 算法与 AdaBoost.M1 算法的结构相同, 分类器的权值计算方法不同, AdaBoost-SAMME 算法的分类器权值更新方法为<sup>[22]</sup>

$$\omega_j = \frac{1}{2} \ln \frac{1 - err_j}{err_j} + \ln(K - 1) \quad (8)$$

式中,  $K$  为样本类别数量, 当  $K=2$  时, AdaBoost-SAMME 与 AdaBoost.M1 的分类器权值更新方法相同。分类器权值  $\omega_j$  应满足  $\omega_j > 0$ , 将此条件代入式(5), 得出  $err_j < 0.5$ ; 而将此条件代入式(8), 得出  $err_j < 1 - 1/K$ 。由此可知, AdaBoost-SAMME 较 AdaBoost.M1 降低了对各分类器的分类正确率的要求, 以期获得更多符合要求的分类器。

表 2 训练集样本实例

Table 2 Example of training set samples

序号	F1	A1	F2	A2	F3	A3	类别
1	2.490 538	0.037 760	0.042 730	0.025 842	1.214 748	0.025 598	1
2	1.550 482	0.156 791	3.082 652	0.033 152	0.549 383	0.029 444	1
3	0.689 781	0.034 638	0.036 626	0.025 374	1.342 937	0.009 560	1
4	0.708 094	0.067 542	0.195 336	0.013 333	2.545 477	0.011 812	1
N	N	N	N	N	N	N	N
41	0.396 777	0.347 939	0.958 369	0.024 802	1.843 487	0.017 811	2
42	0.396 777	0.169 527	1.184 227	0.033 556	3.107 069	0.025 090	2
43	0.396 777	0.215 220	1.190 331	0.049 128	3.393 969	0.010 917	2
44	0.042 730	0.073 820	1.794 653	0.002 776	0.573 801	0.002 430	2
N	N	N	N	N	N	N	N
235	0.006 104	0.169 855	0.842 388	0.024 497	1.489 440	0.008 520	3
236	0	0	0.500 549	0	1.001 099	0	3
237	0.006 104	0.003 626	0.531 071	0.002 958	1.031 620	0.002 930	3
238	0.006 104	0.680 321	0.555 488	0.021 295	1.086 558	0.010 066	3
N	N	N	N	N	N	N	N
1515	0.140 398	0.333 914	0.714 199	0.074 055	1.263 582	0.019 120	3

### 3.2 评价指标

在结果分析之前, 需要结合实际问题选择合适的评价指标。该部分将以二分类问题为例说明, 三分类问题与之类似, 二分类问题中, 假设少数类为正例, 多数类为负例, 混淆矩阵如表 3 所示。

## 3 算法实现与结果分析

### 3.1 数据源

本文所用到的样本取自于全国各地 9 159 台 1.5 MW 机组的运行数据, 采样周期为 0.02 s, 每一个数据文件长度为 2 min。对每一个文件的变桨速度信号进行 FFT 变换, 按照频域强度由大到小的顺序进行排序, 选取 top3 的三个频率及其频域强度作为分类特征量, 提取时需剔除主频率左右 0.5 Hz 范围内的干扰信号。

本文选取的训练样本和测试样本的比例为 4:1, 具体各类别及其数量如表 1 所示。

表 1 训练样本和测试样本基本情况

Table 1 Basic situation of training samples and test samples

	高频	低频	正常	总计
训练样本/个	40	194	1 281	1 515
测试样本/个	10	48	320	378
总计/个	50	242	1 601	1 893

每个样本包含变桨速度信号 FFT 变换后的 top3 频率 F1、F2、F3 及其 FFT 频域强度 A1、A2、A3。类别中“1”表示高频, “2”表示低频, “3”表示正常, 部分训练集样本信息如表 2 所示。

表 3 混淆矩阵

Table 3 Confusion matrix

	正例(预测)	负例(预测)
正例(实际)	TP	FN
负例(实际)	FP	TN

正确率(Accuracy)表示分类正确的样本数与样本总数的比值,反映了分类器对所有样本的分类能力,是最常用的分类性能评价指标,计算方法为

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (9)$$

但是,当遇到样本数量不均衡问题时,即全部样本中某些类别的样本数量明显小于其他类别的样本数量,这时如果仍然采用分类正确率作为分类性能评价指标,就会忽略了分类器对样本数量较少类别的识别准确率,正确率并不能全面地反映分类器的性能。鉴于此,在分类正确率的基础上,又引入查准率(Precision)、查全率(Recall)及 G-mean 作为整体的分类性能评价指标<sup>[23]</sup>。

在预测结果中,真正的正样本数与预测为正的样本数的比值称为查准率。

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

查全率又称为召回率,是真正的正样本数与实际为正的样本数的比值。

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

G-mean 是衡量各种算法对不平衡数据分类性能的常用评价指标,由 TPR 与 TNR 来衡量正例与负例的分类性能,数值越大分类效果越好,两者的乘积即为 G-mean。若分类器对某一类样本的分类效果欠佳,会导致 TPR 或者 TNR 偏小,进而影响 G-mean 数值较小。

$$TPR = \frac{TP}{TP + FN} \quad (12)$$

$$TNR = \frac{TN}{TN + FP} \quad (13)$$

$$G - mean = \sqrt{TPR \cdot TNR} \quad (14)$$

为了更加全面地评价分类器性能,除前述的评价指标之外,还可以引入受试者工作特征曲线 ROC 和曲线下面积 AUC、Kappa 系数。

受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC),主要作用是反映模型在不同阈值下,其敏感性(FPR)和精确性(TPR)的趋势走向。横坐标是假正率 FPR,计算方法如式(15)所示;纵坐标是真正率 TPR,计算方法如式(12)所示。

$$FPR = \frac{FP}{TN + FP} \quad (15)$$

ROC 曲线可以实现对模型的定性分析,若要定量分析,则需借助 AUC(Area under roc curve)面积,它指的是 ROC 曲线下的面积。AUC 数值越大,模型性能越好。

Kappa 系数用于衡量分类精度,其计算需借助于混淆矩阵。计算方法为

$$k = \frac{p_0 - p_e}{1 - p_e} \quad (16)$$

式中, $p_0$ 为每一类中正确分类的样本数量之和与总样本数量的比值。如果每一类的真实样本个数分别为  $a_1, a_2, \dots, a_c$ , 预测结果中每一类的样本个数分别为  $b_1, b_2, \dots, b_c$ , 总样本个数为  $n$ , 则有

$$p_e = \frac{a_1 \times b_1 + a_2 \times b_2 + \dots + a_c \times b_c}{n \times n} \quad (17)$$

### 3.3 算法实现与对比实验

按照图 6 所示流程,利用 AdaBoost-SAMME 算法建立数学模型。本实例中,寻优后迭代次数  $m_{final}$  设置为 100;将所有训练样本送入模型训练。为了验证 AdaBoost-SAMME 算法在变桨异常分类方面的性能,将它与其他五种算法进行对比分析,包括神经网络(Artificial Neural Networks, ANN)、支持向量机(Support Vector Machine, SVM)、随机森林(Random Forest)、Bagging 以及 AdaBoost.M1。

使用 ANN 时需要先对数据进行归一化,然后通过网格搜索法确定最优隐藏节点数  $size=11$ ;最优迭代次数  $maxit=200$ 。对于 SVM 算法,同样利用网格搜索法寻优,求得除去线性核外,其他核的参数  $\gamma=0.25$ ;C 分类中惩罚项  $c$  的取值  $c_{cost}=8$ ;使用 10 折交叉验证构建 SVM 模型。构建 Random Forest 模型时,寻优后确定了每次分支所选择的变量个数  $m_{try}=4$ ,决策树的数量  $n_{tree}=1000$ 。在 Bagging 模型中,决策树数量  $m_{final}=100$ 。AdaBoost.M1 模型与 AdaBoost-SAMME 模型的参数设置相同。

按照上述参数配置,利用训练数据分别对 ANN、SVM、Random Forest、Bagging、AdaBoost.M1 以及 AdaBoost-SAMME 模型进行训练,然后,利用测试数据对上述算法进行测试。

### 3.4 结果分析

通过测试,得到六种算法的正确率等分类评价指标对比情况如表 4 所示。为了更加直观地对比各种算法的分类效果,将表 4 中的测试结果以折线图的形式展示,如图 7 所示。因为风力发电机组变桨系统由异常到故障是一个缓慢的渐变过程,变桨系统的异常识别对分析时间要求不高,所以本文不对各算法的运行效率做对比分析。

由图 7 可知,对各算法按照正确率由大到小排序,依次为: AdaBoost-SAMME > AdaBoost.M1 > SVM > Bagging > RandomForest > ANN, AdaBoost-SAMME 算法的分类正确率最高,为 98.15%。

precision<sub>1</sub> 表示类别 1, 即变桨高频类别的查准率; precision<sub>2</sub> 表示类别 2, 即变桨低频类别的查准率; precision<sub>3</sub> 表示类别 3, 即变桨正常类别的查准率; recall、G-mean 与之类似。

表 4 各算法测试结果对比

Table 4 Test result comparison of each algorithm

算法	accuracy	高频			低频			正常		
		precision	recall	G-mean	precision	recall	G-mean	precision	recall	G-mean
ANN	0.870 4	0.666 7	0.4	0.516 4	0.546 7	0.854 2	0.683 4	0.956 2	0.887 5	0.921 2
SVM	0.952 4	0.428 6	0.3	0.358 6	0.975 6	0.833 3	0.833 3	0.960 6	0.990 6	0.990 6
Random Forest	0.936 5	0.333 3	0.1	0.182 6	0.807 7	0.875 0	0.840 7	0.962 8	0.971 6	0.967 3
Bagging	0.947 1	0.666 7	0.4	0.516 4	0.907 0	0.812 5	0.858 5	0.957 4	0.984 4	0.970 8
AdaBoost.M1	0.973 5	0.900 0	0.9	0.900 0	0.975 6	0.833 3	0.901 6	0.975 5	0.996 9	0.986 1
AdaBoost-SAMME	0.981 5	0.900 0	0.9	0.900 0	0.977 3	0.895 8	0.935 7	0.984 6	0.996 9	0.990 7

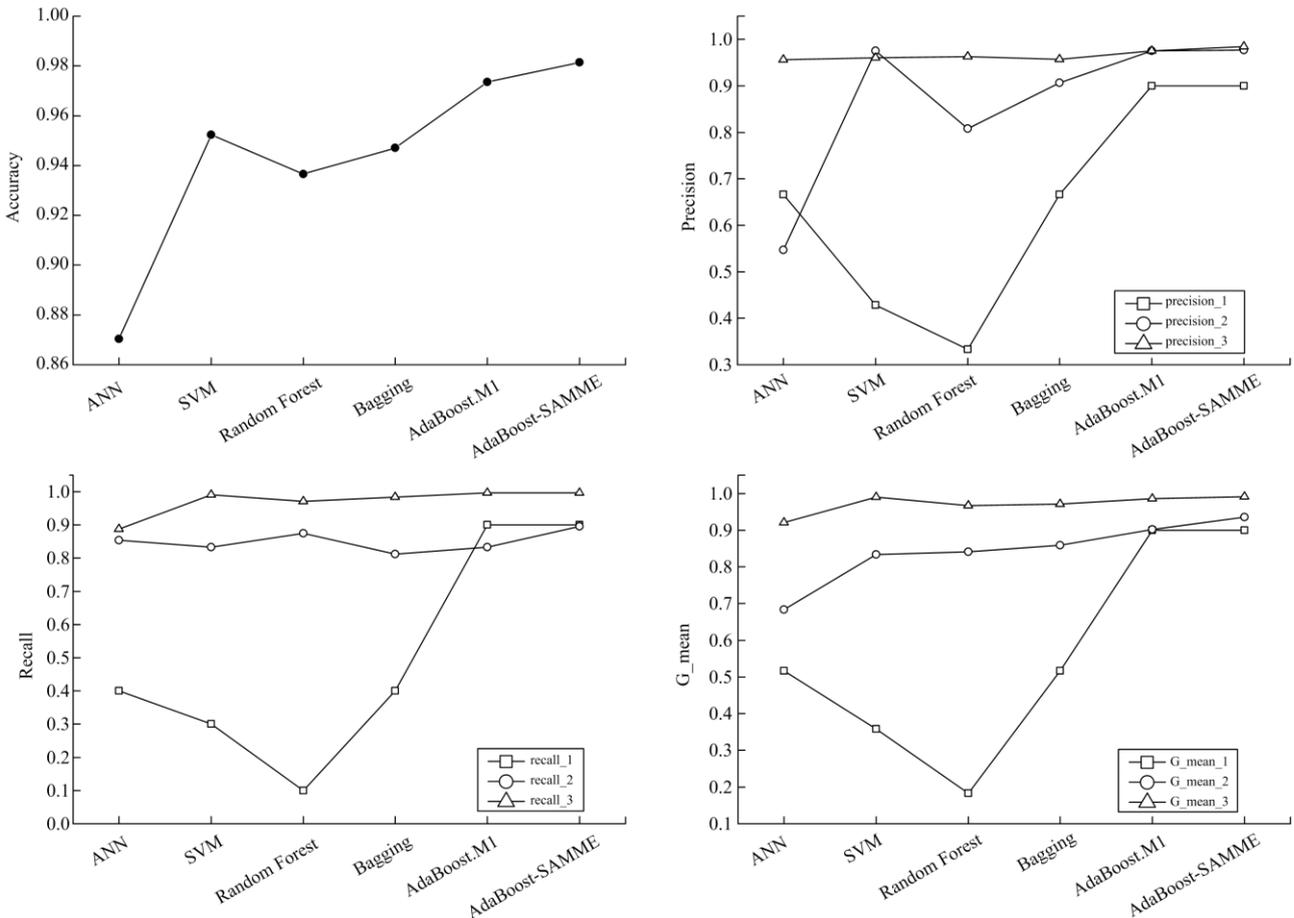


图 7 各算法分类情况对比

Fig. 7 Classification comparison of each algorithm

分析表明, AdaBoost-SAMME 在查准率和查全率方面均表现出优异性能, 即便是样本数量较少的变桨高频信号, 查准率也在 90% 以上; 三个类别的查全率均在 89% 以上。进一步分析 G-mean 指标发现, AdaBoost-SAMME 在三个类别中的 G-mean 值分别为 90%、93.57%、99.07%, 对比其他五种算法,

AdaBoost-SAMME 在每个类别下的 G-mean 均是最优的。

在上述评价指标之外, 绘制 ANN、SVM、Random Forest、Bagging、AdaBoost.M1 以及 AdaBoost-SAMME 这六种算法模型的 ROC 曲线, 并计算其 AUC 值, 结果依次为: 0.879 3, 0.873 9,

0.879 3, 0.717 4, 0.890 8, 0.900 8。其中, AdaBoost-SAMME 算法模型的 ROC 曲线如图 8 所示。与此同时, 前述六种算法模型的 Kappa 系数依次为: 0.580 9, 0.812 7, 0.754 7, 0.798 8, 0.894 6, 0.928 5。综上, AdaBoost-SAMME 算法的 AUC 值与 Kappa 系数均优于其他五种算法。

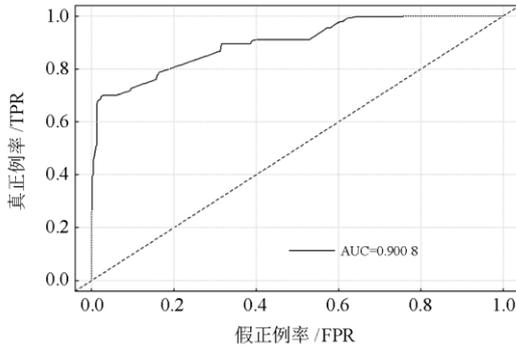


图 8 AdaBoost-SAMME 算法模型 ROC 曲线  
Fig. 8 ROC curve of AdaBoost-SAMME model

由此可知, AdaBoost-SAMME 测试结果在各类样本的正确率、查准率、查全率、Kappa 系数指标方面, 与其他五种算法相比, 在分类精确性方面显示出明显优势。在针对数据不平衡问题(变桨高频类与低频、正常类相比, 样本数量明显偏少)的 G-mean 指标方面, 即使各类样本数量不均衡的情况下, AdaBoost-SAMME 依然能够表现出显著的优越性。在模型敏感性和精确性方面, ROC 曲线走势和曲线下面积 AUC 计算表明, AdaBoost-SAMME 算法同样优于其他五种算法。综上, 从算法的精确性、敏感性以及处理数据不平衡问题能力方面考虑, AdaBoost-SAMME 算法均优于文中所述其他五种算法。由此验证了 AdaBoost-SAMME 算法应用于风机变桨异常识别的可行性和有效性。

针对上述结果差异, 从 AdaBoost-SAMME 算法原理的角度分析, AdaBoost 作为一种有效的分类算法, 在处理学习难度较大的样本时, 会增大其在下一轮学习中的权重, 进而使得其在处理样本不平衡问题上能够表现出优异的性能。与此同时, AdaBoost-SAMME 算法在 AdaBoost.M1 基础上改变了各分类器权值更新的计算方法, 降低了对各分类器的分类正确率的要求, 从而让更多符合要求的分类器参与最终决策。

在未被正确分类的样本中, 有一个高频样本被识别为低频样本; 一个正常样本被识别为高频样本; 五个低频样本被识别为正常样本。从数量方面分析, 五个低频样本未被正确分类, 占全部未被正确分类

样本数量的 71.4%, 如果该部分样本被正确识别, 则本实例中的分类器性能还将得到进一步显著提升; 从对事件的影响程度分析, 五个低频样本被漏分为正常样本, 意味着风力发电机组变桨系统异常未被发现, 这样会延误现场人员对异常机组的发现与检修时机。因此, 下一步的优化工作需要将该情况纳入考虑范围。

AdaBoost-SAMME 算法计算得出的变量重要性结果如图 9 所示, 纵轴表示各变量, 横轴表示各变量的重要度, 用百分数表示。各变量的重要度是通过 AdaBoost-SAMME 每棵决策树中各变量的基尼指数与每棵决策树的权重计算得来, 该值越大, 表示当前变量对于正确分类越重要。

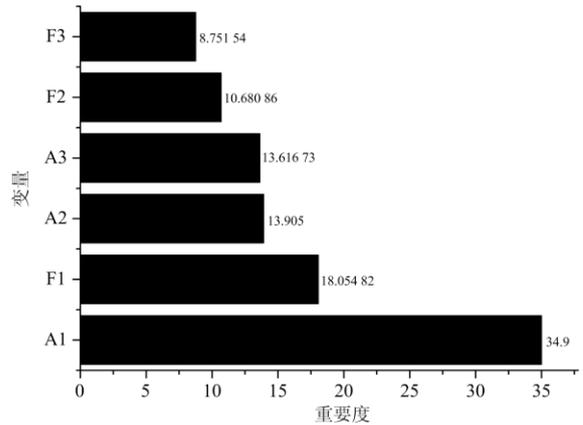


图 9 各变量重要度

Fig. 9 Importance of each variable

A1 和 F1 作为变桨速度信号的最主要频率成分及其频域强度, 它们是区分变桨信号类型(高频、低频、正常)的最主要变量。这与 AdaBoost-SAMME 算法给出的变量主要性结果相吻合, 也进一步证明了该分类算法的合理性和可行性。

#### 4 结论

本文以风力发电机组健康管理平台为基础, 建立了机组变桨异常识别系统。在该系统的分类变量选择方面, 为了实现对异常特征的精确刻画, 通过深入分析异常样本数据, 提出将变桨速度的频域特征作为异常识别的特征变量。在模型算法构建方面, 分别尝试了神经网络、支持向量机、随机森林、Bagging、AdaBoost.M1 以及 AdaBoost-SAMME。测试表明, 即使在各种类样本数量不均衡的情况下, 与文中所述其他算法相比, AdaBoost-SAMME 算法在精确性、敏感性以及处理数据不平衡问题方面也能表现出良好的性能。为了提高系统的自学习能力,

提出了一种基于欧式距离的新样本判断方法,实现自动扩大训练样本数量。

该模型算法的建立,解决了不同机型的变桨系统异常时,变桨速度频域特征普适性规律不明显,用常规逻辑判断方法无法实现变桨系统异常识别的技术难题。在结果分析部分,从 AdaBoost 算法思想和分类器权重更新方法两个方面阐述了 AdaBoost-SAMME 算法识别效果优于其他算法的原因,并从变量重要度方面简要论述了该算法的合理性和可行性。在下一步工作中,需要深入分析被错误分类样本的数据特征,从算法机理、参数寻优、样本标签合理性等方面发掘可能原因,寻找可以进一步提高识别率的技术手段。

将基于 AdaBoost-SAMME 算法的风力发电机组变桨异常识别系统应用于机组日常健康检测中,不仅能够在变桨系统出现故障前识别出异常,通过预警的方式指导现场开展预见性维护,提高机组的可靠性与可利用率。而且能够降低机组变桨电机疲劳,延长其使用寿命。此外,该变桨异常识别系统提供了一个通用型解决方案,对其他工业设备的异常识别同样具有借鉴意义。

### 参考文献

- [1] 王金贺, 张晓红, 曾建潮. 非完美维修模型下的风电机组最优维修决策[J]. 计算机集成制造系统, 2019, 25(5): 1151-1160.  
WANG Jinhe, ZHANG Xiaohong, ZENG Jianchao. Optimal maintenance decision for wind turbines based on imperfect maintenance model[J]. Computer Integrated Manufacturing Systems, 2019, 25(5): 1151-1160.
- [2] 赵洪山, 张兴科, 郭伟. 考虑不完全维修的风机齿轮箱优化检修策略[J]. 电力系统保护与控制, 2014, 42(10): 15-22.  
ZHAO Hongshan, ZHANG Xingke, GUO Wei. Optimized maintenance strategy with imperfect repair for the gearbox of wind turbine[J]. Power System Protection and Control, 2014, 42(10): 15-22.
- [3] 白恺, 宋鹏, 刁嘉, 等. 基于可靠性和发电性能的新能源发电设备状态评估方法[J]. 中国电力, 2016, 49(7): 140-144.  
BAI Kai, SONG Peng, DIAO Jia, et al. Condition assessment method of renewable energy generation equipment based on reliability and generation performance indicators[J]. Electric Power, 2016, 49(7): 140-144.
- [4] 李玉洁, 方瑞明. 基于改进加权灰关联分析法的风电机组可靠性研究[J]. 电力系统保护与控制, 2015, 43(14): 63-69.  
LI Yujie, FANG Ruiming. Reliability assessment for wind turbine based on weighted degree of improved grey incidence[J]. Power System Protection and Control, 2015, 43(14): 63-69.
- [5] VAN T L, NGUYUE T H, LEE D C. Advanced pitch angle control based on fuzzy logic for variable-speed wind turbine systems[J]. IEEE Transactions on Energy Conversion, 2015, 30(2): 578-587.
- [6] 贾文强, 谢双义, 何娇, 等. 独立变桨条件下的风力发电机动态载荷优化及仿真分析[J]. 中国电力, 2017, 50(1): 130-135.  
JIA Wenqiang, XIE Shuangyi, HE Jiao, et al. Dynamics load optimization and simulation analysis of wind turbine under independent pitch control[J]. Electric Power, 2017, 50(1): 130-135.
- [7] 颜湘武, 李君岩, 魏星. 直驱永磁同步风电机组在全风速范围内的控制策略研究[J]. 电力系统保护与控制, 2019, 47(23): 138-144.  
YAN Xiangwu, LI Junyan, WEI Xing. Research on control strategy of direct-drive permanent magnet synchronous wind turbine in full wind speed range[J]. Power System Protection and Control, 2019, 47(23): 138-144.
- [8] BADAL F R, DAS P, SARKER S K, et al. A survey on control issues in renewable energy integration and microgrid[J]. Protection and Control of Modern Power Systems, 2019, 4(1): 87-113. DOI: 10.1186/s41601-019-0122-8.
- [9] ALSUMIRI M, LI L, JIANG L, et al. Residue theorem based soft sliding mode control for wind power generation systems[J]. Protection and Control of Modern Power Systems, 2018, 3(3): 247-258. DOI: 10.1186/s41601-018-0097-x.
- [10] 董兴辉, 闫慧丽, 张晓亮. 基于 FDA 贡献图的风电机组变桨系统故障定位[J]. 可再生能源, 2017, 35(1): 93-100.  
DONG Xinghui, YAN Huili, ZHANG Xiaoliang. Fault location of pitch system based on FDA contribution diagram[J]. Renewable Energy, 2017, 35(1): 93-100.
- [11] 姚万业, 李新丽. 基于状态监测的风电机组变桨系统故障诊断[J]. 可再生能源, 2016, 34(3): 437-440.  
YAO Wanye, LI Xinli. Pitch system fault diagnosis for wind turbine based on condition monitoring[J]. Renewable Energy, 2016, 34(3): 437-440.
- [12] 李伟昌, 张磊. 基于风力发电系统的风电机组变桨距故障诊断[J]. 计算机仿真, 2015, 32(9): 147-151.  
LI Weichang, ZHANG Lei. Faults diagnosis of wind turbine pitch system based on ANFIS[J]. Computer Simulation, 2015, 32(9): 147-151.

- [13] 尹诗, 余忠源, 孟凯峰, 等. 基于非线性状态估计的风电机组变桨控制系统故障识别[J]. 中国电机工程学报, 2014, 34(增刊 1): 160-165.  
YIN Shi, YU Zhongyuan, MENG Kaifeng, et al. Fault identification of pitch control system of wind turbine based on nonlinear state estimation[J]. Proceedings of the CSEE, 2014, 34(S1): 160-165.
- [14] 李辉, 杨超, 李学伟, 等. 风机电动变桨系统状态特征参量挖掘及异常识别[J]. 中国电机工程学报, 2014, 34(12): 1922-1930.  
LI Hui, YANG Chao, LI Xuewei, et al. Conditions characteristic parameters mining and outlier identification for electric pitch system of wind turbine[J]. Proceedings of the CSEE, 2014, 34(12): 1922-1930.
- [15] 李贺. 基于核主元分析的风电机组变桨距系统故障诊断研究[D]. 沈阳: 沈阳工业大学, 2017.  
LI He. Fault diagnosis of variable frequency system of wind turbine based on kernel principal component analysis[D]. Shenyang: Shenyang University of Technology, 2017.
- [16] YAN Junhua, BAI Xuehan, ZHANG Wanyi, et al. No-reference image quality assessment based on AdaBoost\_BP neural network in wavelet domain[J]. Journal of Systems Engineering and Electronics, 2019, 30(2): 223-237.
- [17] 张国辉, 李志中, 王宾, 等. 基于 Adaboost 的高压直流线路连续换相失败预警方法[J]. 电力系统保护与控制, 2019, 47(19): 69-77.  
ZHANG Guohui, LI Zhizhong, WANG Bin, et al. Early warning method of HVDC line continuous commutation failure based on Adaboost algorithm[J]. Power System Protection and Control, 2019, 47(19): 69-77.
- [18] MARTINEZ W, GRAY J B. Noise peeling methods to improve boosting algorithms[J]. Computational Statistics and Data Analysis, 2016, 93: 483-497.
- [19] 曹惠玲, 高升, 薛鹏. 基于多分类 AdaBoost 的航空发动机故障诊断[J]. 北京航空航天大学学报, 2018, 44(9): 1818-1825.  
CAO Huiling, GAO Sheng, XUE Peng. Aeroengine fault diagnosis based on multi-classification AdaBoost[J]. Journal of Beijing University of Aeronautics and Astronautics, 2018, 44(9): 1818-1825.
- [20] 王桂兰, 赵洪山, 米增强. XGBoost 算法在风机主轴承故障预测中的应用[J]. 电力自动化设备, 2019, 39(1): 73-77, 83.  
WANG Guilan, ZHAO Hongshan, MI Zengqiang. Application of XGBoost algorithm in prediction of wind motor main bearing fault[J]. Electric Power Automation Equipment, 2019, 39(1): 73-77, 83.
- [21] 孙亮, 黄倩. 实用机器学习[M]. 北京: 人民邮电出版社, 2017.
- [22] ZHU J, ZOU H, ROSSET S, et al. Multi-class AdaBoost[J]. Statistics and Its Interface, 2009(2): 349-360.
- [23] 王来, 樊重俊, 杨云鹏, 等. 面向不平衡数据分类的 KFDA-Boosting 算法[J]. 计算机应用研究, 2019, 36(3): 807-811.  
WANG Lai, FAN Chongjun, YANG Yunpeng, et al. KFDA-Boosting algorithm oriented to imbalanced data classification[J]. Application Research of Computers, 2019, 36(3): 807-811.

收稿日期: 2019-12-18; 修回日期: 2020-04-25

作者简介:

樊 帅(1988—), 男, 通信作者, 工学硕士, 工程师, 从事工业设备故障诊断与机器学习应用研究。E-mail: fs1234@foxmail.com

(编辑 周金梅)