

基于动态阈值 AdaBoost 算法的风电机组发电机 电气故障诊断研究

彭艳来¹, 樊永¹, 杨晓峰¹, 杨宏宇², 王志新³

(1. 龙源电力集团(上海)新能源有限公司 综合管理部, 上海 200000; 2. 上海博英信息科技有限公司 软件部, 上海 200000; 3. 上海交通大学 电气工程系, 上海 200030)

摘要: 针对风场 SCADA 数据量庞大, 难以实现风电机组发电机快速故障分类的问题, 提出一种动态阈值 AdaBoost 算法(DTAdaBoost)。在该算法中引入动态阈值对样本集数据进行筛选, 剔除训练模型贡献较小的数据, 提出的算法不仅可以减小样本容量, 还能使被错误分类的数据多次训练, 最终实现发电机气隙不均、匝间短路、断条等电气故障的快速诊断。在与其他方法对比的实验结果表明: DTAdaBoost 算法在电气故障诊断中比常规 AdaBoost 算法、SWTAdaBoost 算法、GAAdaBoost 算法运算时间短、准确率高, 为风电机组的快速故障诊断提供理论基础。

关键词: 故障分类; SCADA 数据; AdaBoost 算法; 数据处理

中图分类号: TM614 **文献标识码:** A **DOI:** 10.19457/j.1001-2095.dqed24172

Research on Electrical Fault Diagnosis of Wind Turbine Generator Based on Dynamic Threshold AdaBoost Algorithm

PENG Yanlai¹, FAN Yong¹, YANG Xiaofeng¹, YANG Hongyu², WANG Zhixin³

(1. General Management Department, Longyuan Power Group(Shanghai) New Energy Co., Ltd., Shanghai 200000, China; 2. Software Department, Shanghai Proinvent Information Tech Co., Ltd., Shanghai 200000, China; 3. Department of Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200030, China)

Abstract: Aiming at the problems that the amount of data in wind farm SCADA data is huge and it is difficult to classify the fast fault of wind turbine generator, a dynamic threshold AdaBoost algorithm (DTAdaBoost) was proposed. Based on AdaBoost algorithm, the dynamic threshold was introduced to screen the sample set data, and eliminate the data that contribute little to the training model, reduce the sample size, train the incorrectly classified data for many times, at last the rapid diagnosis of electrical faults such as uneven generator air gap, inter turn short circuit and broken bar was realized. Compared with other classification methods, the experiments results show that DTAdaBoost algorithm has faster operation time and higher accuracy than conventional AdaBoost algorithm, SWTAdaBoost algorithm and GAAdaBoost algorithm in electrical fault diagnosis. It provides a theoretical basis for rapid fault diagnosis of wind turbine.

Key words: fault classification; SCADA data; AdaBoost algorithm; data processing

随着能源危机的不断加剧与对新型清洁能源的迫切需求, 我国的风力发电技术得到了迅猛的发展。发电机是风电机组核心部件之一, 但由于其长期工作在沙漠、戈壁等极端恶劣环境中, 承受着瞬态交变载荷, 所以发电机经常发生故障, 一旦其发生故障, 将严重影响风电机组的正

常运行。因此, 为保证风电机组正常运转, 对发电机故障诊断是十分必要的^[1-2]。

数据采集与监视控制系统(supervisory control and data acquisition, SCADA)可用于数据采集与监视控制, 具有实时监测, 数据易于获取的优点。在风力发电机组的故障诊断研究过程中, 国内外

基金项目: 国家重点研发计划(2018YFB1503000)

作者简介: 彭艳来(1987—), 男, 本科, 工程师, Email: 12052486@chnenergy.com.cn

通讯作者: 杨宏宇(1982—), 男, 本科, 高级工程师, Email: Red_rainsy@163.com

许多专家学者采用 SCADA 数据进行分析^[3-4]。池昊等^[5]利用 SCADA 数据中温度等参数对风电机组进行诊断,对风电机组的安全维护提供参考意义;JIN X 等^[6]利用功率、转速等数据对风机异常状态进行诊断,保证风机正常运转。

与传统分类方法相比,自适应增强(adaptive boosting, AdaBoost)算法具有适应数据类型多、克服采样频率低的优点,因此处理 SCADA 数据常采用 AdaBoost 算法^[7-9]。叶春霖等^[10]利用 SCADA 数据,对总体样本进行随机采样,使 AdaBoost 算法样本数据量减少,但容易丢失关键信息;Seyed-hosseini 等^[11]提出新型加权选取结合 AdaBoost (weighted novelty selection-AdaBoost, WNS-AdaBoost)算法,在原有样本中选出新样本集,但样本集选择耗时长,导致运算速度下降。熊霖等^[12]提出一种动静态权重结合的 AdaBoost 算法,在静态权重基础上加入可适应待测样本的动态权重,该算法虽然提高了准确率,但降低了计算效率。

虽然 AdaBoost 算法具有较强的适应能力,但由于 SCADA 数据量庞大,这会导致 AdaBoost 算法在数据处理上会产生计算效率低、无法实时监测的问题。因此,本文提出一种动态阈值 AdaBoost 算法(dynamic threshold AdaBoost, DTAdaBoost),来实现风力机故障的快速诊断,并满足实时监测的要求。

1 DTAdaBoost 算法原理

1.1 AdaBoost 算法

在实际故障诊断分类问题中,直接构建强分类器将带来巨大困难。AdaBoost^[13-14]是一种将弱分类器集成为强分类器的运算方法。

AdaBoost 算法基本流程如下:

1) 初始化训练数据的权值分布,使每个数据组的权值为 $w_{1,i} = 1/m, i=1, 2, \dots, m, m$ 为数据组中数据的总个数。

2) 使用具有权值分布的训练数据集学习,得到弱分类器 $G_l(x), l=1, 2, \dots, L, L$ 为弱分类器的总个数。

3) 更新训练数据集的权值分布:

$$D_{l+1} = (w_{l+1,1}, w_{l+1,2}, \dots, w_{l+1,m}) \quad (1)$$

4) 构建强分类器 $G(x)$:

$$G(x) = \text{sign}[f(x)] \quad (2)$$

式中: $G(x)$ 为强分类器; $\text{sign}(\cdot)$ 为符号函数; $f(x)$ 为强分类器的输入。

AdaBoost 算法在每轮迭代后都会降低正确分类样本的权值。导致此类数据在训练模型时作用小、运算效率低。

1.2 DTAdaBoost 算法

针对 AdaBoost 算法处理 SCADA 数据时运算效率低的问题,本文采取在迭代过程中设置动态阈值的方式实现快速分类,该动态阈值选取权值较大的数据来训练弱分类器,其选取数据方式如图 1 所示。

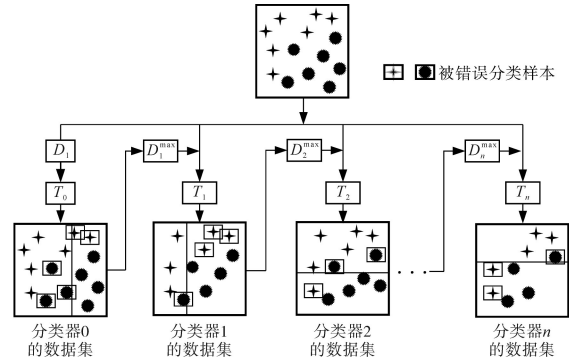


图1 动态阈值选取数据集示意图

Fig.1 Schematic diagram of dynamic threshold selection data set

DTAdaBoost 算法流程如下:

输入: 训练集 $T = \{(x_1, y_1), \dots, (x_m, y_m)\}$ 。其中 $x_i \in X, X$ 为训练集, m 为训练集中输入的个数; $y_i = \{-1, 1\}$ 。

输出: 强分类器 $G(x)$ 。

1) 初始化训练数据的权值分布,每个数据组的权值为 $w_{1,i} = 1/m, i=1, 2, \dots, m$ 。

2) 引入权重阈值 T_n , 每次迭代时,选取样本中权重最大值与训练集系数计算权重阈值 T_n :

$$T_n = \frac{K}{m} \cdot \max(w_{n,i}) \quad (3)$$

式中: K 为训练系数; n 为迭代次数。

3) 选取权值 $w_{n,i} > T_n$ 的数据作为新数据集 $D_n^{\max}$ 。在迭代初期样本权值差异较小,权重阈值小,选取数据量多。随着迭代进行,样本间权值差异变大,权重阈值随之变大,选取数据量变少。

$$w_{n,i} \geq T_n = \begin{cases} x_i \text{ 加入新数据集 } D_n^{\max} & \text{是} \\ \text{舍弃} & \text{否} \end{cases} \quad (4)$$

4) 新样本集 $D_n^{\max}$ 中数据,包含步骤 3) 中权值较高的数据进行训练。利用新样本集 $D_n^{\max}$ 中数据得到分类器 $G_n^{\max}$ 。

5) 计算 $G_n^{\max}$ 在训练集上的分类误差 $e_n^{\max}$:

$$e_n^{\max} = P[G_n^{\max}(x_i) \neq y_i] = \sum_{i=1}^m w_{n,i} I[G_n^{\max}(x_i) \neq y_i] \quad (5)$$

式中: $P(\cdot)$ 为由强分类器 $G_n^{\max}$ 计算出的分类概率。

当 $e_n^{\max} \geq 0.5$ 且 $D_n^{\max} = D$ 时, 令 $N = n - 1$, 停止迭代; 当 $e_n^{\max} \geq 0.5$ 且 $D_n^{\max} \neq D$ 时, 则令 $T_n = 0$, 转至步骤 3)。

6) 计算 $G_n^{\max}$ 的系数:

$$\alpha_n^{\max} = \frac{1}{2} \ln \frac{1 - e_n^{\max}}{e_n^{\max}} \quad (6)$$

7) 更新训练数据集的权值分布:

$$D_{n+1}^{\max} = (w_{n+1,1}, \dots, w_{n+1,i}, \dots, w_{n+1,m}) \quad (7)$$

$$w_{n+1,i} = \frac{w_{n,i}}{Z_n} \exp[-\alpha_n^{\max} y_i G_n^{\max}(x_i)] \quad (8)$$

式中: Z_n 为规范化因子, 可保证 $D_{n+1}^{\max}$ 是一个概率分布。

8) 构建强分类器 $G(x)$:

$$G(x) = \text{sign} \left[\sum_{n=1}^N \alpha_n^{\max} G_n^{\max}(x) \right] \quad (9)$$

从流程步骤 2)~步骤 5) 中可以看出, 本文提出的 DTAdaBoost 算法能够对全体 SCADA 数据进行筛选, 该算法既可以减小样本容量, 又能够减小正确分类样本的训练次数, 最终提高运算效率与准确率。该算法的具体流程如图 2 所示。

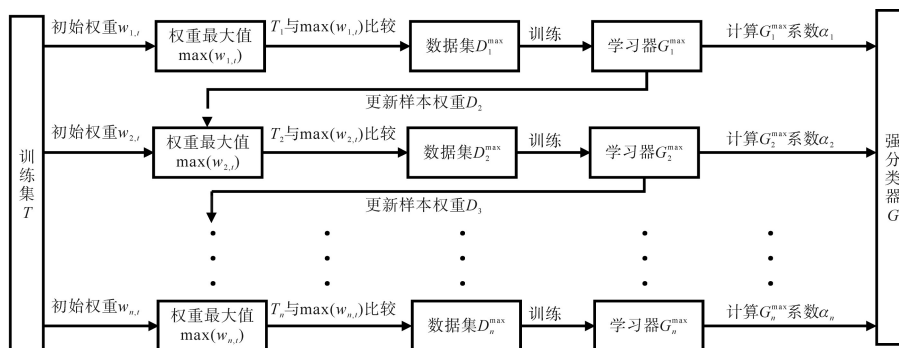


图 2 DTAdaBoost 算法流程图

Fig.2 Flow chart of DTAdaBoost algorithm

利用 SCADA 数据对风电机组发电机进行快速诊断的 DTAdaBoost 算法的实时监控流程图如图 3 所示。

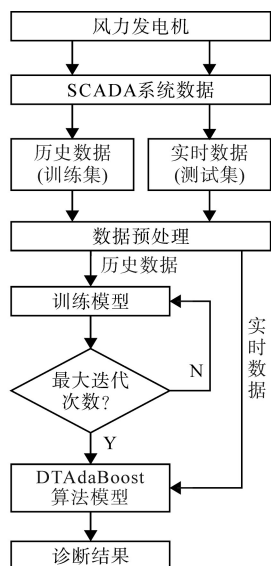


图 3 实时监控流程图

Fig.3 Flow chart of real time monitoring

2 实验研究与结果分析

2.1 实验基本信息

采集上海某风场 2019 年—2021 年 SCADA 数

据, 数据总量为 157 680 组, 每组数据含有风速、转速、功率、发电机温度、电流、电压、有功功率、无功功率、叶片螺旋角、叶片温度等特征, 数据采样频率为 10 min/次。该数据集由正常状态、转子断条、转子气隙不均、转子匝间短路 4 种状态组成, 他们的分类标签如表 1 所示。

表 1 故障标签

Tab.1 Fault tag

故障类型	故障缩写	标签
正常状态	N	0
转子断条	BR	1
转子气隙不均	UAg	2
转子匝间短路	SC	3

标签 0, 1, 2 和 3 分别代表正常状态、转子断条故障、转子气隙不均故障、转子匝间短路故障。

2.2 数据处理

本文利用皮尔逊相关系数法选取 SCADA 数据中相关性大于中度相关的指标作为故障特征指标。皮尔逊相关系数表达式为

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (10)$$

式中:上标“-”为相应变量的平均值。

选取 36 802 组 SCADA 数据进行相关性分析,转速、风速、有功功率、无功功率、电压、电流、温度 7 个特征的相关系数如表 2 所示。

表 2 指标间相关系数

Tab.2 Correlation coefficient between indicators		
故障类型	指标	相关系数
正常	电流	0.905
	电压	0.481
	有功功率	0.817
	无功功率	0.792
	转速	0.651
	温度	0.746
	风速	0.586
转子断条	电流	0.845
	电压	0.443
	有功功率	0.845
	无功功率	0.734
	转速	0.819
	温度	0.694
	风速	0.612
转子气隙不均	电流	0.821
	电压	0.501
	有功功率	0.821
	无功功率	0.733
	转速	0.779
	温度	0.648
	风速	0.517
转子匝间短路	电流	0.833
	电压	0.471
	有功功率	0.833
	无功功率	0.706
	转速	0.784
	温度	0.661
	风速	0.547

选择指标区间相关性 $r_{xy} \geq 0.5$ 的指标作为故障特征指标。因此选择转速、风速、有功功率、无功功率、电流、温度 6 个指标作为故障特征指标。

AdaBoost 算法会着重对分类困难的数据进行训练,导致算法对异常数据比较敏感。在风电机组运行时,由于风机停机维护、风力冲击等因素的影响,会产生风机的停机数据、风机限制功率数据、“噪声”数据^[15]等异常数据,这类异常数据在迭代过程中,多次训练并获得较高权重,这将

直接影响模型训练的准确率与运算效率。故需要在数据预处理时剔除这些异常数据。

剔除异常数据后得到的风速—功率曲线如图 4 所示。

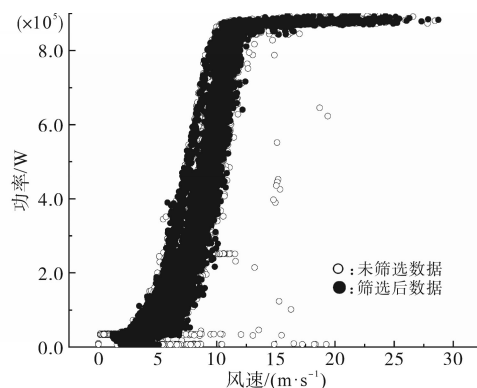


图 4 风速—功率数据处理图

Fig.4 Wind speed—power data processing diagram

2.3 结果分析

为了进一步凸显本文算法的运算速度,将传统的 AdaBoost 算法、遗传 AdaBoost 算法(GAAdaBoost)、静态权重剪切 AdaBoost 算法(SWTAdaBoost)三种方法与 DTAdaBoost 算法进行对比。处理结果选取 2019 年—2021 年 SCADA 数据中的 36 000 组数据作为训练集,得到的 4 种算法在不同迭代次数下的准确率曲线如图 5 所示。

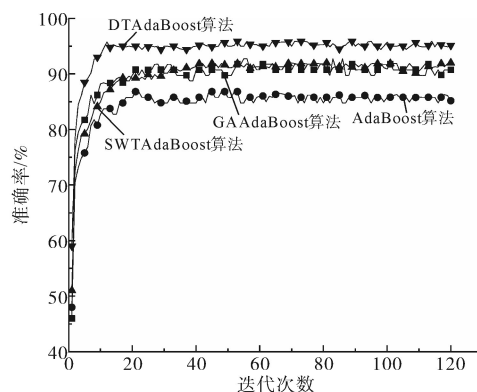


图 5 算法准确率

Fig.5 Accuracy of algorithms

由图 5 可以看出,DTAdaBoost 算法准确率高达 94%,均优于其他算法。SWTAdaBoost 算法与 GAAdaBoost 算法的准确率均高于 AdaBoost 算法,但低于 DTAdaBoost 算法。在图 5 中,迭代 10 次与迭代 120 次准确率基本相同。所以若要进行快速诊断,只需将迭代次数设置为 10。

为了进一步验证本文提出算法的优越性,当样本量依次为 20 000,36 000 和 60 000 时,4 种算法运算时间如表 3 所示。当迭代次数依次为 120,200 和 300 时,4 种算法运算时间如表 4 所示。

表 3 不同样本量迭代 120 次的运算时间

Tab.3 Operation time of 120 iterations with different sample sizes

算法	运算时间/s		
	样本量 20 000	样本量 36 000	样本量 60 000
AdaBoost	7.2	11.7	18.5
SWTAdaBoost	5.4	7.6	13.9
GAAdaBoost	4.9	7.1	12.4
DTAdaBoost	4.3	6.7	11.1

表 4 36 000 组数据不同迭代次数的运算时间

Tab.4 Operation time of 36 000 groups of data with different iteration times

算法	运算时间/s		
	迭代次数 120 次	迭代次数 200 次	迭代次数 300 次
AdaBoost	11.7	19.8	29.5
SWTAdaBoost	7.6	14.6	22.3
GAAdaBoost	7.1	12.8	19.1
DTAdaBoost	6.7	11.9	17.6

由表 3 可以看出,迭代 120 次,数据量越大,算法的运算时间越长,DTAdaBoost 算法运算效率高优势越明显。当样本量为 60 000 时,DTAdaBoost 的运算效率比 AdaBoost 高 39.8%。

由表 4 可以看出,36 000 个样本点在不同的迭代次数下,DTAdaBoost 算法运算效率最高,运算速度比传统 AdaBoost 算法运算效率高 40.8%,比 SWTAdaBoost 算法高 19.4%,比 GAAdaBoost 算法高 7.03%。

根据处理 36 000 组数据、迭代 120 次训练得出 4 种算法的模型,将风场实时采集的 60 000 组 SCADA 数据组作为测试集进行故障分类,4 种算法采用 T-SNE 算法进行可视化处理,处理结果如图 6 所示。可以看出 DTAdaBoost 算法分类效果最好,可以有效诊断风电机组发电机实时故障。

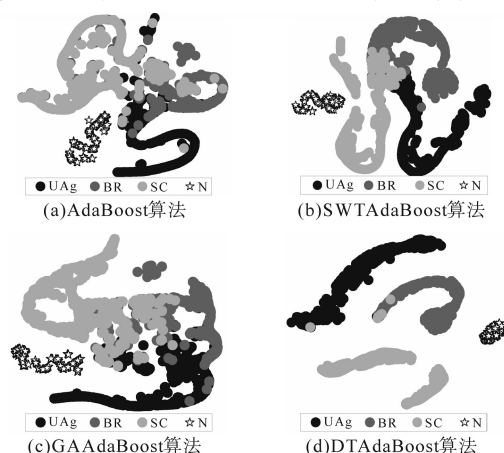


图 6 算法聚类图
Fig.6 Algorithm clustering graph

利用 36 000 组样本 120 次迭代训练得到 4 种算法的诊断模型,按照图 3 流程对 60 000 组数据进行风电机组发电机故障诊断。4 种算法均对数据进行预处理,所需时间相同,故只对比从数据预处理结束后到算法模型得出诊断结果的时间。各算法模型诊断时间如表 5 所示。

表 5 算法诊断时间

Tab.5 Algorithm diagnosis time

算法	诊断时间/s	诊断准确率/%
AdaBoost	0.312 7	83
SWTAdaBoost	0.235 1	89
GAAdaBoost	0.221 9	88
DTAdaBoost	0.188 2	93

由表 5 可以看出,采用 DTAdaBoost 算法的诊断时间最短。这是由于 DTAdaBoost 算法对数据集进行剪切,多次对相同分类的数据不再进行诊断,计算量大幅减小,可以实现风电机组发电机故障的快速诊断,且诊断准确率也比其他三种算法模型高。

3 结论

本文通过设置动态阈值的方式来优化传统的 AdaBoost 算法,提出一种高效运算速度的动态阈值 AdaBoost 算法,即 DTAdaBoost 算法。解决了传统 AdaBoost 算法在处理 SCADA 数据计算效率低的问题。通过与现有的几种改进 AdaBoost 算法进行对比,诊断结果表明,所提出的 DTAdaBoost 算法计算效率提高了 40.8%,同时,诊断精度高达 94%。因此,本文提出的算法在处理大量 SCADA 数据时,不仅提高了运算效率,还提高了准确率,可以实现风电机组发电机的快速实时诊断,为风电机组发电机的故障诊断提供重要的理论依据。

参考文献

- [1] SAWANT M,THAKARE S,RAO A P,et al. A review on state-of-the-art reviews in wind-turbine-and wind-farm-related topics [J]. Energies, 2021, 14(8): 2041–2049.
 - [2] JIN X,CHEN Y,WANG L,et al. Failure prediction, monitoring and diagnosis methods for slewing bearings of large-scale wind turbine:a review[J]. Measurement, 2021, 172(8): 126–135.
 - [3] 谷晓娇,刘天顺,李时雨. 变风速条件下风力机轴承故障特征提取[J]. 机械与电子, 2021, 39(9): 8–11.
- GU Xiaojiao, LIU Tianshun, LI Shiyu. Fault feature extraction for bearing of wind turbine under variable wind speed condition [J]. Machinery & Electronics, 2021, 39(9): 8–11.

- [4] 马宏忠,张艳,魏海增,等. 基于转子平均瞬时功率的双馈异步发电机定子绕组匝间短路故障诊断[J]. 电力自动化设备, 2018,38(4):151-156.
- MA Hongzhong, ZHANG Yan, WEI Haizeng, et al. Diagnosis of stator winding inter-turn short circuit in DFIG based on instantaneous average power in rotor side[J]. Electric Power Automation Equipment, 2018,38(4):151-156.
- [5] 池昊,彭成,贺婧,等. 基于 SCADA 数据的风机叶片结冰短期预测方法[J]. 信息与电脑(理论版), 2019,10(4):61-65.
- CHI Hao, PENG Cheng, HE Jing, et al. Short term ice prediction by using SCADA data on wind turbine blades[J]. China Computer & Communication, 2019,10(4):61-65.
- [6] JIN X, XU Z, QIAO W. Condition monitoring of wind turbine generators using SCADA data analysis[J]. IEEE Transactions on Sustainable Energy, 2020,12(1):202-210.
- [7] UPADHYAY D, MANERO J, ZAMAN M, et al. Intrusion detection in SCADA based power grids: recursive feature elimination model with majority vote ensemble algorithm[J]. IEEE Transactions on Network Science and Engineering, 2021,8(3):2559-2574.
- [8] WANG W, HARROU F, BOUYEDDOU B, et al. A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems [J]. Cluster Computing, 2021,7(6):1-18.
- [9] WU Z, WANG X, JIANG B. Fault diagnosis for wind turbines based on relief and extreme gradient boosting[J]. Applied Sciences, 2020,10(9):3258-3269.
- [10] 叶春霖,邱颖宁,冯延晖. 基于数据挖掘的风电机组叶片结冰故障诊断[J]. 噪声与振动控制, 2018,38(Z1):643-647.
- YE Chunlin, QIU Yingning, FENG Yanhui. Faults diagnosis of wind turbine blade icing based on data mining[J]. Noise and Vibration Control, 2018,38(Z1):643-647.
- [11] SEYEDHOSSEINI M, PAIVA A R C, TASDIZEN T. Fast AdaBoost training using weighted novelty selection[C]//The 2011 International Joint Conference on Neural Networks. IEEE, 2011,9(6):1245-1250.
- [12] 熊霖,唐万梅. 一种加入动态权重的 AdaBoost 算法[J]. 重庆师范大学学报(自然科学版), 2019,36(5):117-122.
- XIONG Lin, TANG Wanmei. An improved AdaBoost algorithm by add dynamic weight[J]. Journal of Chongqing Normal University (Natural Science), 2019,36(5):117-122.
- [13] 余陆斌,杜启亮,田联房. 基于自适应权值裁剪的 AdaBoost 快速训练算法[J]. 电子与信息学报, 2020,42(11):2742-2748.
- YU Lubin, DU Qiliang, TIAN Lianfang. Fast training AdaBoost algorithm based on adaptive weight trimming[J]. Journal of Electronics & Information Technology, 2020,42(11):2742-2748.
- [14] SUN Q, XIE Y, HU X, et al. Identification of generating units that abuse market power in electricity spot market based on AdaBoost-DT algorithm[J]. Mathematical Problems in Engineering, 2021,18(6):147-156.
- [15] SHI Y, LIU Y, GAO X. Study of wind turbine fault diagnosis and early warning based on SCADA data[J]. IEEE Access, 2021,9(8):124600-124615.

收稿日期:2022-01-22

修改稿日期:2022-03-02