



(12) 发明专利

(10) 授权公告号 CN 110162888 B

(45) 授权公告日 2022. 12. 09

(21) 申请号 201910439351.5

G06K 9/62 (2022.01)

(22) 申请日 2019.05.24

G06N 20/00 (2019.01)

(65) 同一申请的已公布的文献号

G06Q 50/06 (2012.01)

申请公布号 CN 110162888 A

G06F 119/08 (2020.01)

(43) 申请公布日 2019.08.23

(56) 对比文件

(73) 专利权人 北京天泽智云科技有限公司

CN 109522627 A, 2019.03.26

地址 100191 北京市海淀区知春路17-7号

CN 109376801 A, 2019.02.22

致真大厦C座4层北区

US 2012226485 A1, 2012.09.06

(72) 发明人 王旻轩 鲍亭文 金超 晋文静

韩彦岭等. 主动学习与半监督技术相结合的海冰图像分类.《遥感信息》.2019, (第02期),

史喆 李杰

叶春霖等. 基于数据挖掘的风电机组叶片结冰故障诊断.《噪声与振动控制》.2018,

(74) 专利代理机构 北京煦润律师事务所 11522

审查员 张媛媛

专利代理师 惠磊

(51) Int. Cl.

G06F 30/17 (2020.01)

G06F 30/27 (2020.01)

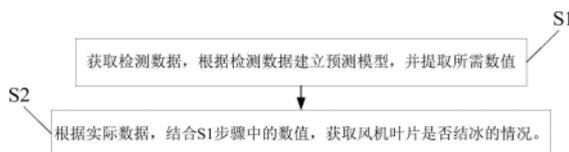
权利要求书2页 说明书7页 附图2页

(54) 发明名称

一种基于半监督式学习的风机叶片结冰检测的方法

(57) 摘要

本发明提出了一种不依赖样本质量和故障标签的,物理规则和机器学习模型相结合的基于半监督式学习的风机叶片结冰检测的方法;包括S1、获取监测数据,根据监测数据建立预测模型,并提取所需数值;S2、根据实际数据,结合S1步骤中的数值,获取风机叶片是否结冰的情况;本发明提出的基于半监督式学习的风机叶片结冰检测的方法利用SCADA数据进行检测,能够覆盖风机运行的各个工况性能参数;同时基于物理模型和机器学习模型,极大地提高了算法的准确度和稳健性;且不依赖积累的故障样本,可以在无标签的情况下对叶片结冰进行分类判别。



1. 一种基于半监督式学习的风机叶片结冰检测的方法,其特征在于,包括:
- S1、获取监测数据,根据监测数据建立预测模型,并提取所需数值;
- S2、根据实际数据,结合S1步骤中的数值,获取风机叶片是否结冰的情况;
- 其中,所述S1步骤包括:
- S11、获取风机运行时的监测数据,并对监测数据进行选取;
- S12、根据监测数据构建风机正常运行的风速-功率预测模型;
- S13、在监测数据中提取风机原始通道振动序列和机舱内温度与环境温度的差值序列,并确定风机原始通道振动的基线阈值和机舱内外温差的基线阈值;
- 其中,所述S2步骤包括:
- S21、输入实际数据,并将实际数据按照时间变量进行排序;
- S22、对S21排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工况,并获得新增特征风速平方、机舱内外温差和扭矩,并根据新增特征和S1步骤中的数值获取偏移参数;
- S23、根据S22步骤中新的风机运行工况,使用S1步骤中的预测模型设置阈值的方式对数据标记结冰标签;
- S24、将S23步骤中的结冰标签作为正样本,并针对正样本进行分析,得到预测概率;
- S25、根据S24步骤中的结果,将预测概率大于判别阈值的样本点确定为结冰样本,并得出最终结果;
- 其中,所述S22步骤中:
- 风速平方设定为 vs_square ;机舱内外温差设定为 $temp_d$;扭矩设定为 $torque$;
- $$torque = \frac{9550P}{n}$$
- ,其中,P为风机功率,n为风轮转速;
- 风能利用系数 C_p :
- $$C_p = \frac{P_{wt}}{\frac{\rho S V^3}{2}}$$
- ,其中 ρ 为当地空气密度,单位为 kg/m^3 ;S为风轮扫风面积,单位为 m^2 ;V为上游风速,单位为 m/s ;且 ρ 和S可以视作常数项;
- 矩阵系数 C_t :
- $$C_t = \frac{4M}{\rho D A v^3}$$
- ,其中M为扭矩,D为风轮直径;且D可以视作常数项;
- 叶尖速比TSR:
- $$TSR = \frac{R\omega}{V} = \frac{2\pi R n}{60V}$$
- ,其中n为风轮转速,单位为 r/min ;R为叶尖半径,单位为 m ;V为上游风速,单位为 m/s ; ω 为风轮角速度,单位为 rad/s ;且R和 ω 可以视作常数项;
- 功率偏移 $power_d$,根据风机历史数据构造的理论风速-功率模型,使用理论功率减去实际功率作为功率偏移参数。
2. 如权利要求1所述的基于半监督式学习的风机叶片结冰检测的方法,其特征在于,所述S11步骤为:获取风机运行时SCADA的监测数据,并对监测数据进行选取。
3. 如权利要求1所述的基于半监督式学习的风机叶片结冰检测的方法,其特征在于,步骤23包括:

S231、判断风机外部风速是否处于风机运行范围内,如果风速达到了启机风速,进行后续操作;

S232、判断风机外部环境温度是否低于设定值,设定值一般为 0°C ,如果环境温度低于设定温度,进行后续操作;

S233、判断当前风速下理论输出功率与风机实时输出功率差值与历史正常情况的偏移是否超出阈值,如果偏移超过根据正常运行状态得到的阈值,进行后续操作;

S234、判断风机原始通道振动值与历史正常情况的偏移是否超出阈值,若偏移超过根据正常运行状态得到的阈值,表示风机有其他故障,若偏移低于阈值,表示风机可能结冰并继续;

S235、判断风机机舱内外温差与历史正常情况的偏移是否超出阈值,若偏移超过历史正常状态得到的阈值,表示风机已经结冰,为到这一步骤筛选到所有满足条件的数据打上结冰标签。

4. 如权利要求1所述的基于半监督式学习的风机叶片结冰检测的方法,其特征在于,所述S24步骤为:

将S23步骤中的结冰标签作为正样本,并以人工标记结冰标签的正样本为基础,对剩余样本基于bagging袋装法构造机器学习模型,并迭代进行二分类,辅助并扩大、修正结冰标签,以迭代均值作为最终的结冰预测概率。

5. 如权利要求3所述的基于半监督式学习的风机叶片结冰检测的方法,其特征在于,所述S24步骤包括:

S241、将人工分为两部分的实际数据结冰样本和未标记样本;标注出具有标签的正样本个数,并确定一个迭代次数;

S242、根据正样本的个数,随机可重复地从未标记样本中选取和数量相同的样本作为负样本,并打上相应的未结冰标签;

S243、根据具有相同数量正样本和负样本的数据集,构造二分类器,并对在S1步骤获取的数据之外的未标记样本进行预测,并输出每个未标记样本点属于结冰样本的概率值;

S244、随机抽样,不断地重新训练分类模型,并每次都对未标记样本进行预测;

S245、使用迭代的平均值作为最终预测的概率输出,若S241步骤中人工标记正样本数大于剩余样本数,则使用正样本所占比例作为预测概率输出;

S246、将结冰预测概率作为风险值,将风险值大于判别阈值的样本点确定为结冰样本,对风机整体结冰情况做出评估。

一种基于半监督式学习的风机叶片结冰检测的方法

技术领域

[0001] 本发明涉及风机叶片监测领域,特别是指一种基于半监督式学习的风机叶片结冰检测的方法。

背景技术

[0002] 数据表明,风力发电机叶片结冰会改变叶片叶形,破坏叶片气动特性,从而导致风机效率下降和运行不稳定,进而对电网的稳定运行产生影响。因此,实现早期叶片结冰预测可以有效提高风机运行效率和电网运行安全;

[0003] 现有技术中,通过如下几种方式进行监测:

[0004] 1、物理模型阈值判定法:

[0005] 中国专利申请公布号CN 108167140A,名称:风力发电机组叶片结冰的监测方法和装置;中国专利申请公布号CN 103899485A,名称:一种监测风机运行时叶片结冰的方法;上述两个专利中采用了一些基于物理规则人为设定阈值的方法,即通过SCADA系统中得到的监测测点数据设定阈值进行判断,以诊断风机叶片是否结冰;而此种方法在风机主控制系统中存在广泛应用,优点是比较直观、简单,但是通常传感器的测点的影响因素诸多,基于物理模型的方法通常只利用了SCADA系统中的一小部分测点变量,并且人工设置阈值需要大量的实验基础,以及对风机在各种不同复杂工况下运行机理的深入了解,单独的阈值监测方法容易在正常的情况下引起误报,也很难达到良好的普适性、稳健性;

[0006] 2、基于机器学习、深度学习的分类模型判定法:

[0007] 中国专利申请公布号CN109522627A,名称:基于SCADA数据的风机叶片结冰预测方法;中国专利申请公布号CN 109026563 A,名称:一种基于特征选择和XGBoost的风机叶片结冰预测方法;中国专利申请公布号CN 109209790 A,名称:一种基于数据建模的风电叶片结冰推测方法;上述专利中采用了基于特征选择和机器学习分类模型的风机叶片结冰预测方法;

[0008] 中国专利申请公布号CN109376801 A,名称:基于集成深度神经网络的风力发电机叶片使用了集成深度学习模型来预测风机结冰;这一类方法不依赖于物理机理,省去了人工筛选特征和设定阈值的工作,但是其缺陷在于对于训练数据量和标签质量依赖较强,而在实际的生产环境中很难积累到满足训练要求的数据和故障标签;

[0009] 3、基于声音和图像的诊断方法:

[0010] 中国专利申请公布号CN 207538980 U,名称一种风机叶片结冰监测装置;中国专利申请公布号CN 107642464 A,名称一种风机叶片结冰监测装置及监测方法;上述专利中使用了音频转换电流信号,通过比较噪音信号的异同来诊断结冰;

[0011] 中国专利申请公布号CN 109117858 A,名称监测风力发电机叶片结冰的方法及装置,根据摄像头拍摄风机照片,对图像处理后根据与训练好的模型对图像进行识别,来判断是否结冰。

[0012] 这一类方法直接从声光的角度出发,比较直接,但是处理起来比较复杂,也依赖大

量的结冰和非结冰样本来构建样本库,而且由于没有使用SCADA的数据,会丧失一些风机运行工况的参数。

[0013] 急需一种方法解决上述技术问题。

发明内容

[0014] 本发明提出一种不依赖样本质量和故障标签的,物理规则和机器学习模型相结合的基于半监督式学习的风机叶片结冰检测的方法。

[0015] 本发明的技术方案是这样实现的:

[0016] 一种基于半监督式学习的风机叶片结冰检测的方法,包括:

[0017] S1、获取监测数据,根据监测数据建立预测模型,并提取所需数值;

[0018] S2、根据实际数据,结合S1步骤中的数值,获取风机叶片是否结冰的情况。

[0019] 作为进一步的技术方案,所述S1步骤包括:

[0020] S11、获取风机运行时的监测数据,并对监测数据进行选取;

[0021] S12、根据监测数据构建风机正常运行的风速-功率预测模型;

[0022] S13、在监测数据中提取风机原始通道振动序列和机舱内温度与环境温度的差值序列,并确定风机原始通道振动的基线阈值和机舱内外温差的基线阈值。

[0023] 作为进一步的技术方案,所述S11步骤为:获取风机运行时SCADA的监测数据,并对监测数据进行选取。

[0024] 作为进一步的技术方案,所述S2步骤包括:

[0025] S21、输入实际数据,并将实际数据按照时间变量进行排序;

[0026] S22、对S21排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工况,并根据S1步骤中的数值获取偏移参数;

[0027] S23、根据S22步骤中新的风机运行工况,使用S1步骤中的预测模型设置阈值的方式对数据标记结冰标签;

[0028] S24、将S23步骤中的结冰标签作为正样本,并以人工标记结冰标签的正样本为基础,对剩余样本基于bagging袋装法构造机器学习模型,并迭代进行二分类,辅助并扩大、修正结冰标签,以迭代均值作为最终的结冰预测概率;

[0029] S25、根据S24步骤中的结果,以结冰预测概率作为风险值(若人工构造标签的样本数量超过一半,则直接使用结冰标签样本比例作为风险值),将风险值大于判别阈值的样本点确定为结冰样本,并得出最终结果。

[0030] 作为进一步的技术方案,所述S22步骤为:

[0031] 对S21排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工况,并获得新增特征风速平方、机舱内外温差和扭矩,并根据新增特征和 S1步骤中的数值获取偏移参数。

[0032] 作为进一步的技术方案,所述S22步骤中:

[0033] 风速平方设定为 vs_square ;机舱内外温差设定为 $temp_d$;扭矩设定为 $torque$;

[0034] $torque = \frac{9550P}{n}$,其中,P为风机功率,n为风轮转速;

[0035] 风能利用系数 C_p :

[0036] $C_p = \frac{P_{wf}}{\rho S V^3}$, 其中 ρ 为当地空气密度, 单位为 kg/m^3 ; S 为风轮扫风面积, 单位为 m^2 ; V

为上游风速, 单位为 m/s ; 且 ρ 和 S 可以视作常数项;

[0037] 矩阵系数 C_t :

[0038] $C_t = \frac{4M}{\rho D A v^3}$, 其中 M 为扭矩, D 为风轮直径; 且 D 可以视作常数项;

[0039] 叶尖速比TSR:

[0040] $TSR = \frac{R\omega}{V} = \frac{2\pi Rn}{60V}$, 其中 n 为风轮转速, 单位为 r/min ; R 为叶尖半径, 单位为 m ; V 为

上游风速, 单位为 m/s ; ω 为风轮角速度, 单位为 rad/s ; 且 R 和 ω 可以视作常数项;

[0041] 功率偏移 power_d , 根据风机历史数据构造的理论风速-功率模型, 使用理论功率减去实际功率作为功率偏移参数。

[0042] 作为进一步的技术方案, 步骤23包括:

[0043] S231、判断风机外部风速是否处于风机运行范围内, 如果风速达到了启机风速, 进行后续操作;

[0044] S232、判断风机外部环境温度是否低于设定值, 设定值一般为 0°C , 如果环境温度低于设定温度, 进行后续操作;

[0045] S233、判断当前风速下理论输出功率与风机实时输出功率差值与历史正常情况的偏移是否超出阈值, 如果偏移超过根据正常运行状态得到的阈值, 进行后续操作;

[0046] S234、判断风机原始通道振动值与历史正常情况的偏移是否超出阈值, 若偏移超过根据正常运行状态得到的阈值, 表示风机有其他故障, 若偏移低于阈值, 表示风机可能结冰并继续;

[0047] S235、判断风机机舱内外温差与历史正常情况的偏移是否超出阈值, 若偏超过历史正常状态得到的阈值, 表示风机已经结冰, 为到这一步骤筛选到所有满足条件的数据打上结冰标签。

[0048] 作为进一步的技术方案, 所述S24步骤为:

[0049] 将S23步骤中的结冰标签作为正样本, 并以人工标记结冰标签的正样本为基础, 对剩余样本基于bagging袋装法构造机器学习模型, 并迭代进行二分类, 辅助并扩大、修正结冰标签, 以迭代均值作为最终的结冰预测概率。作为进一步的技术方案, 所述S24步骤包括:

[0050] S241、将人工分为两部分的实际数据结冰样本和未标记样本; 标注出具有标签的正样本个数, 并确定一个迭代次数;

[0051] S242、根据正样本的个数, 随机可重复地从未标记样本中选取和数量相同的样本作为负样本, 并打上相应的未结冰标签;

[0052] S243、根据具有相同数量正样本和负样本的数据集, 构造二分类器, 并对在S1步骤获取的数据之外的未标记样本进行预测, 并输出每个未标记样本点属于结冰样本的概率值;

[0053] S244、随机抽样, 不断地重新训练分类模型, 并每次都对未标记样本进行预测;

[0054] S245、使用迭代的平均值作为最终预测的概率输出, 若S241步骤中人工标记正样本数大于剩余样本数, 则使用正样本所占比例作为预测概率输出;

[0055] S246、将结冰预测概率作为风险值,将风险值大于判别阈值的样本点确定为结冰样本,对风机整体结冰情况做出评估。

[0056] 本发明技术方案利用SCADA数据进行检测,能够覆盖风机运行的各个工况性能参数;同时基于物理模型和机器学习模型,极大地提高了算法的准确度和稳健性;且不依赖积累的故障样本,可以在无标签的情况下对叶片结冰进行分类判别。

附图说明

[0057] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动的前提下,还可以根据这些附图获得其他的附图。

[0058] 图1为本发明一种基于半监督式学习的风机叶片结冰检测的方法的流程图;

[0059] 图2为本发明中实施例的流程图;

[0060] 图3为图2的详细流程图。

具体实施方式

[0061] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0062] 如图1所示,本发明提出的一种基于半监督式学习的风机叶片结冰检测的方法,包括如下步骤:

[0063] 获取监测数据,根据监测数据建立预测模型,并提取所需数值;具体的,获取风机运行时的监测数据,并对监测数据进行选取;根据监测数据构建风机正常运行的风速-功率预测模型;在监测数据中提取风机原始通道振动序列和机舱内温度与环境温度的差值序列,并确定风机原始通道振动的基线阈值和机舱内外温差的基线阈值;本发明中,优选的获取风机运行时SCADA(风机数据采集与监视控制)的监测数据,并对监测数据进行选取;具体的,获取风机运行时SCADA系统获取的监测数据,并对监测数据进行初步的清理和筛选,具体包括:通过数据中的风机运行状态标识筛选非停机、限电等非正常运行状况的数据,并按照时间排序;判断是否出现缺失值、重复值等过多数据质量不佳情况出现,对数据质量进行预处理(修正/剔除)得到预处理后的风机监测数据;

[0064] 根据实际数据,结合S1步骤中的数值,获取风机叶片是否结冰的情况;具体的,通过如下步骤获取是否结冰的情况:

[0065] 输入实际数据,并将实际数据按照时间变量进行排序;对排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工况,并根据S1步骤中的数值获取偏移参数;根据新的风机运行工况,使用预测模型设置阈值的方式对数据标记结冰标签;将结冰标签作为正样本,并针对正样本进行分析,得到预测概率;根据结果,将预测概率大于判别阈值的样本点确定为结冰样本,并得出最终结果;

[0066] 其中,对排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工

况,并根据数值获取偏移参数,具体为对排序后的实时数据进行筛选,根据机理和预测模型计算新的风机运行工况,并获得新增特征风速平方、机舱内外温差和扭矩,并根据新增特征和数值获取偏移参数。

[0067] 且,新增特征:

[0068] 风速平方设定为vs_square;机舱内外温差设定为temp_d;扭矩设定为 torque;

[0069] $torque = \frac{9550P}{n}$, 其中,P为风机功率,n为风轮转速;

[0070] 风能利用系数 C_p :

[0071] $C_p = \frac{P_{wt}}{\frac{\rho S V^3}{2}}$, 其中 ρ 为当地空气密度,单位为 kg/m^3 ;S为风轮扫风面积,单位为 m^2 ;V

为上游风速,单位为 m/s ;且 ρ 和S可以视作常数项;

[0072] 矩阵系数 C_t :

[0073] $C_t = \frac{4M}{\rho D A v^3}$, 其中M为扭矩,D为风轮直径;且D可以视作常数项;

[0074] 叶尖速比TSR:

[0075] $TSR = \frac{R\omega}{V} = \frac{2\pi Rn}{60V}$, 其中n为风轮转速,单位为 r/min ;R为叶尖半径,单位为 m ;V为

上游风速,单位为 m/s ; ω 为风轮角速度,单位为 rad/s ;且R和 ω 可以视作常数项;

[0076] 功率偏移power_d,根据风机历史数据构造的理论风速-功率模型,使用理论功率减去实际功率作为功率偏移参数。

[0077] 另外,根据新的风机运行工况,使用步骤中的预测模型设置阈值的方式对数据标记结冰标签;包括如下步骤:

[0078] 判断风机外部风速是否处于风机运行范围内,如果风速达到了启机风速,进行后续操作;判断风机外部环境温度是否低于设定值,设定值一般为 $0^\circ C$,如果环境温度低于设定温度,进行后续操作;判断当前风速下理论输出功率与风机实时输出功率差值与历史正常情况的偏移是否超出阈值,如果偏移超过根据正常运行状态得到的阈值,进行后续操作;判断风机原始通道振动值与历史正常情况的偏移是否超出阈值,若偏移超过根据正常运行状态得到的阈值,表示风机有其他故障,若偏移低于阈值,表示风机可能结冰并继续;判断风机机舱内外温差与历史正常情况的偏移是否超出阈值,若偏超过历史正常状态得到的阈值,表示风机已经结冰,为到这一步骤筛选到所有满足条件的数据打上结冰标签。

[0079] 本发明中,将结冰标签作为正样本,并针对正样本进行分析,得到预测概率;具体为将结冰标签作为正样本,并针对正样本进行分析,对为被分析的样本进行二分类,以迭代均值作为预测概率,具体包括:

[0080] 将人工分为两部分的实际数据结冰样本和未标记样本;标注出具有标签的正样本个数,并确定一个迭代次数;根据正样本的个数,随机可重复地从未标记样本中选取和数量相同的样本作为负样本,并打上相应的未结冰标签;根据具有相同数量正样本和负样本的数据集,构造二分类器(本发明中通过XGBoost算法构造二分类器),并对获取的数据之外的未标记样本进行预测,并输出每个未标记样本点属于结冰样本的概率值;随机抽样,不断地重新训练分类模型,并每次都对未标记样本进行预测;使用迭代的平均值作为最终预测的

概率输出若正样本数大于剩余样本数,使用正样本所占比例作为预测概率输出。

[0081] 其中,

[0082] 本半监督学习算法具体步骤:

[0083] 输入: P = 通过人工标记得到的结冰样本, U = 未被标记的剩余样本, K = bootstrap自助法有放回随机抽样的样本个数, T = bootstrap自助法迭代次数

[0084] 输出:结冰分值 s : U 中个样本属于结冰样本的概率值, $s: U \rightarrow \mathbb{R}$

[0085] 初始化对所有未标记样本中的样本 x ,初始化 x 属于结冰样本的概率 $f(x)$ 为0, x 被抽取到的次数 $u(x)$ 为0, $\forall x \in U, n(x) \leftarrow 0, f(x) \leftarrow 0$

[0086] 对循环 $t=1$ 到 T :

[0087] 从 U 中有放回随机抽样 K 个样本得到bootstrap自助法样本 U_t 训练分类器 f_t 区分出在 U_t 中可能存在的结冰样本 P 对任意属于未标记样本 U 抽样出的样本 $x, x \in U \setminus U_t$,更新:

[0088] $f(x) \leftarrow f(x) + f_t(x)$

[0089] $n(x) \leftarrow n(x) + 1$

[0090] 循环结束

[0091] 对所有的未标记样本中的样本 $x, x \in U$,输出分值 $s(x) = \frac{f(x)}{n(x)}$

[0092] 结冰样本(正样本)为设定为 P ,未标记样本设定为 U ,正样本个数设定为 K ,确定的一个迭代次数设定为 T ;负样本(为结冰样本)设定为 N ;

[0093] 则,将人工分为两部分的实际数据结冰样本 P 和未标记样本 U ;标注出具有标签的正样本个数 K ,并确定一个迭代次数 T ;根据正样本 P 的个数,每次随机可重复地从未标记样本 U 中选取和 P 数量相同的样本作为负样本 N ,并打上相应的未结冰标签;根据具有相同数量正样本和负样本的数据集,构造二分类器(本发明中通过XGBoost算法构造二分类器),并对获取的数据之外的未标记样本进行预测,并输出每个未标记样本点属于结冰样本的概率值;随机抽样进行至 T 次,不断地重新训练分类模型,并每次都对未标记样本进行预测;使用 T 次迭代的平均值作为最终预测的概率输出;若正样本数大于剩余样本数,使用正样本所占比例作为预测概率输出。

[0094] 如图2和图3所示,为更好的理解本发明的技术方案,特举例说明:

[0095] 使用的风机数据供19000多条,其中为结冰样本比例为88.68%,结冰样本比例为5.58%;

[0096] 1、获得风机设备传感器的监测数据;

[0097] 2、对数据进行预处理,根据数据中的风机状态标识删除设备在故障、维护、限电等非正常运行状态时的数据;

[0098] 3、根据风机历史正常运行时的运行数据进行风功率模型训练得到风速-功率关系的拟合模型,并确定温度阈值、切入风速阈值、机舱内外温差阈值和原始通道振动阈值;

[0099] 4、判断风功率模型的拟合残差是否服从正态分布,若服从则对于输入数据首先使用训练好的风功率模型进行拟合;

[0100] 5、根据理论/实际功率偏移程度(阈值:历史未结冰数据拟合残差的均值+3 倍标准差)、温度阈值和切入风速阈值(温度阈值:0°C,风速阈值:-1)和未结冰时机舱内外温差

设定阈值(上四分位数)进行人工结冰筛选,筛选的数据中有95%为真实结冰样本;

[0101] 6、将筛选出的结冰样本作为上述半监督学习算法的正样本,通过bagging 袋装法进行迭代预测,得到每一个未被标记样本属于结冰样本的概率分值 s ;

[0102] 7、对得到的概率分值使用机器学习的评价标准AUC计算,得到最终的分类AUC分值为0.84。

[0103] 以上所述仅为本发明的较佳实施例而已,并不用以限制本发明,凡在本发明的精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

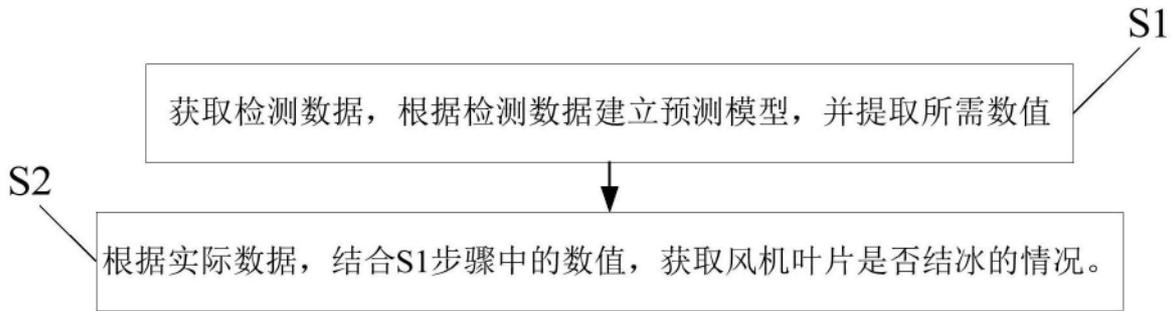


图1

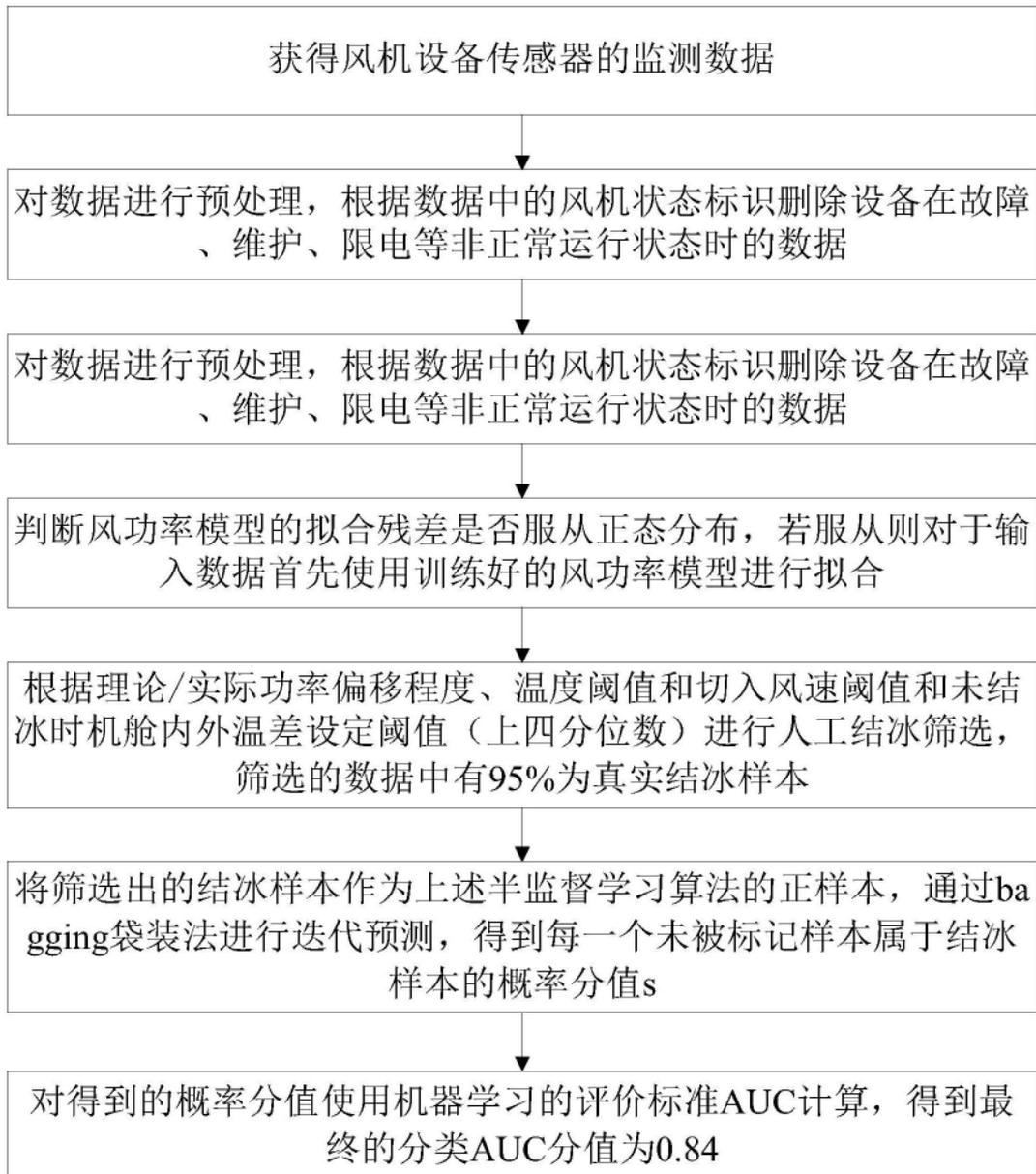


图2

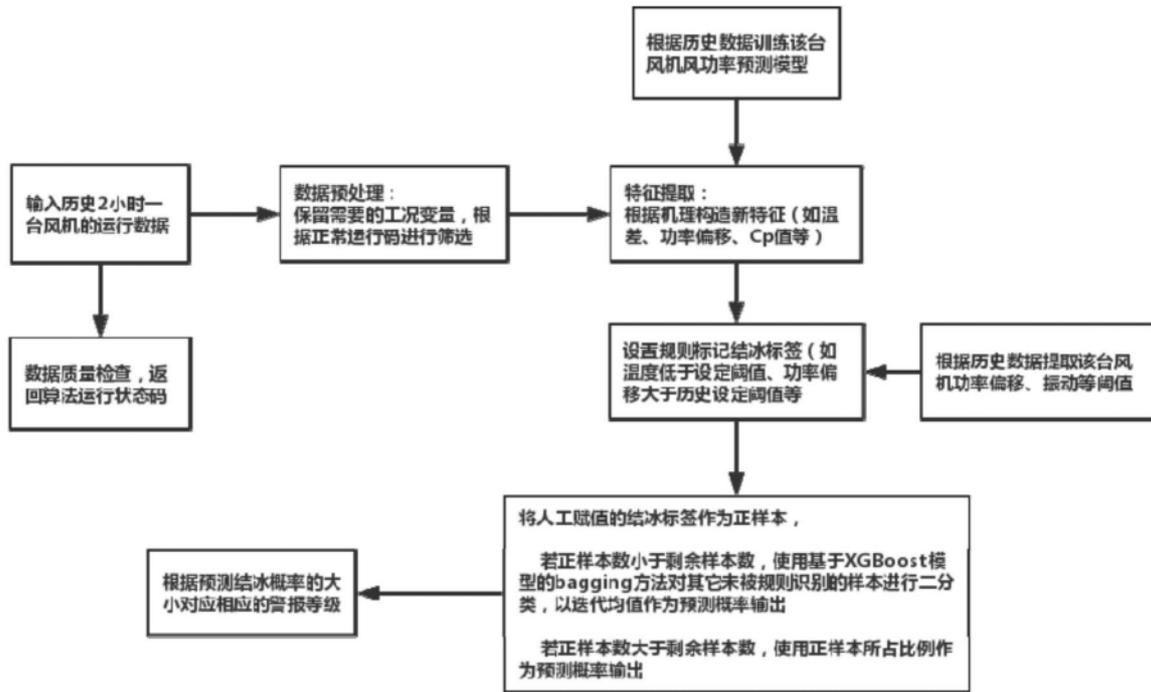


图3