



(12) 发明专利

(10) 授权公告号 CN 111237134 B

(45) 授权公告日 2022. 04. 01

(21) 申请号 202010038529.8

(22) 申请日 2020.01.14

(65) 同一申请的已公布的文献号
申请公布号 CN 111237134 A

(43) 申请公布日 2020.06.05

(73) 专利权人 上海电力大学
地址 200090 上海市杨浦区平凉路2103号

(72) 发明人 魏书荣 张鑫 符杨 任子旭
缪舒馨 闫梦飞

(74) 专利代理机构 上海科盛知识产权代理有限公司 31225

代理人 赵继明

(51) Int. Cl.

F03D 17/00 (2016.01)

F03D 80/00 (2016.01)

F03D 9/25 (2016.01)

G06Q 50/06 (2012.01)

(56) 对比文件

CN 110414154 A, 2019.11.05

CN 103308855 A, 2013.09.18

CN 110659770 A, 2020.01.07

CN 108204341 A, 2018.06.26

CN 104937263 A, 2015.09.23

CN 108376298 A, 2018.08.07

CN 109146137 A, 2019.01.04

CN 110285403 A, 2019.09.27

CN 110659779 A, 2020.01.07

CN 106844826 A, 2017.06.13

CN 110210648 A, 2019.09.06

曹淦昆等. 基于LSTM神经网络的风机齿轮带断裂故障预测.《测量与检测技术》.2019,第41卷(第4期),第92-95页.

史佳琪等. 基于多模型融合Stacking集成学习方式的负荷预测方法.《中国电机工程学报》.2019,第39卷(第14期),第4032-4041页.

审查员 陈友庆

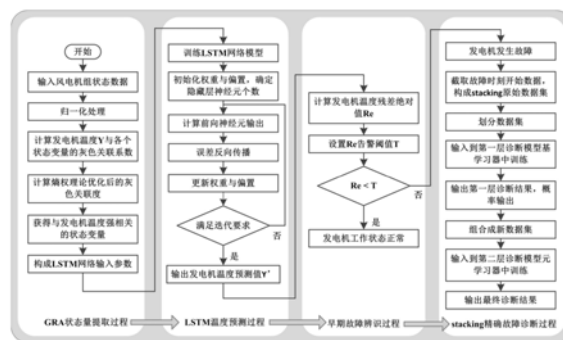
权利要求书2页 说明书10页 附图8页

(54) 发明名称

一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法

(57) 摘要

本发明涉及一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,该方法包括以下步骤:步骤1:通过GRA对SCADA状态变量进行分析,筛选出与发电机温度相关的状态变量作为LSTM网络的输入;步骤2:通过LSTM对正常状态下的发电机温度进行预测,得到预测值;步骤3:计算实际值与预测值的残差绝对值,使用统计学方法设置告警阈值,据此对发电机早期故障进行辨识并提取出故障样本;步骤4:通过stacking融合算法对提取的故障样本进行数据处理后输出最终的精确诊断结果。与现有技术相比,本发明具有故障诊断精确、通用性强,故障样本获取简单等优点。



CN 111237134 B

1. 一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,该方法包括以下步骤:

步骤1:通过GRA对SCADA状态变量进行分析,筛选出与发电机温度相关的状态变量作为LSTM网络的输入;

步骤2:通过LSTM对正常状态下的发电机温度进行预测,得到预测值;

步骤3:计算实际值与预测值的残差绝对值,使用统计学方法设置告警阈值,据此对发电机早期故障进行辨识并提取出故障样本;

步骤4:通过stacking融合算法对提取的故障样本进行数据处理后输出最终的精确诊断结果;

所述的步骤4具体包括:对步骤3提取的故障样本进行数据集划分,采用K折交叉验证对stacking融合算法第一层基学习器进行训练,输出作为诊断结果的概率输出,组合第一层基学习器输出的概率作为新数据集,并作为第二层元学习器的输入进行训练,输出最终的精确诊断结果;

所述第一层基学习器中选择的相关算法模型包括:支持向量机、K最邻近分类算法、采用bagging方式集成的随机森林、梯度提升决策树以及采用boosting方式集成的极端梯度提升树,所述第一层基学习器中的各个基学习器经过单独诊断后,还对诊断结果进行spearman关联性分析,所述spearman关联性分析的计算公式为:

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

式中, ρ_{xy} 为spearman关联性分析对应结果量, \bar{x} 和 \bar{y} 分别为各向量元素的平均值, x_i 和 y_i 分别为各向量元素的实际诊断值;

所述第二层元学习器中选择的相关算法模型包括GBDT。

2. 根据权利要求1所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤1包括以下分步骤:

步骤11:提取SCADA系统采集的风机正常运行的状态数据,并对其进行归一化处理;

步骤12:计算发电机温度与各个状态变量的灰色关联系数;

步骤13:使用熵权理论优化权重大小,计算各个状态变量与发电机温度的灰色关联度;

步骤14:对灰色关联度进行排序,并选择相关的状态变量作为LSTM网络的输入。

3. 根据权利要求2所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤12中的灰色关联系数,其计算公式为:

$$\xi_{ij} = \frac{\min_i \min_j |\Delta x| + \rho \max_i \max_j |\Delta x|}{|\Delta x| + \rho \max_i \max_j |\Delta x|}$$

式中, ξ_{ij} 为灰色关联系数, ρ 为分辨率, Δx 为作为参考序列的状态变量与作为比较序列的发电机温度的变化量, $i=1,2,\dots,m$, $j=1,2,\dots,n$, m 为采样样本数目, n 为状态变量数目。

4. 根据权利要求2所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤13中的灰色关联度,其计算公式为:

$$r_j = \frac{1}{m} \sum_{i=1}^m \omega_i \xi_{ij}$$

式中, r_j 为灰色关联度, $\sum_{i=1}^m \omega_i$ 由于使用熵权理论进行优化,即取值为1。

5. 根据权利要求1所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤2具体包括:将GRA模型提取的状态变量作为输入,发电机温度作为输出,接着对LSTM网络进行训练,通过迭代更新权值和偏置使误差最小,获得发电机温度预测模型,输出发电机温度预测值。

6. 根据权利要求1所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤3包括以下分步骤:

步骤31:获取发电机温度实际值与预测值的残差绝对值;

步骤32:采用统计学方法设置残差绝对值告警阈值;

步骤33:据此对发电机早期故障进行辨识,一旦发电机温度残差绝对值超过告警阈值提取出告警之后的状态数据作为下一步精确故障诊断的故障样本。

7. 根据权利要求6所述的一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,其特征在于,所述的步骤32中的告警阈值,其计算公式为:

$$\int_T^{+\infty} f(\text{Re}) d_{\text{Re}} = \alpha$$

式中, $f(\text{Re})$ 为概率密度函数, Re 为残差绝对值, T 为告警阈值, α 为显著性水平。

一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法

技术领域

[0001] 本发明涉及海上风电机组故障诊断领域,尤其是涉及一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法。

背景技术

[0002] 海上风电已经成为全球可再生能源发展的重点。随着海上风电技术的快速发展,欧洲各海上风电强国开始向规模化、深远海化推进。英国目前规划的Dogger Bank海上风电场离岸距离为123-289km,总装机容量达到4.8GW。德国在领海以外370.4km以内规划了13个总装机容量为21.3GW的海上风电场。中国的海上风电场亦正在向规模化、深远海化方向推进。双馈异步发电机作为远海风力发电机的主流机型之一,面临运行环境恶劣、可及性差和故障率高等问题。据统计,发电机故障是导致风电机组停机的主要因素之一,占有故障停机时间的37%。为了减少发电机故障停机带来的重大经济损失,亟需精确诊断出海上双馈风力发电机故障。

[0003] 中国海上风电开发距今已有十年的历史,大批量风电机组的数据采集与监控(Supervisory Control And Data Acquisition,SCADA)系统能够获取整个风电场海量的运行数据并进行远程或本地监控,这些海量的时序数据记录了海上风电机组实时运行工况,具有信号变化快和运行参数众多的特点。风电机组的故障特征隐含在能表征其运行状态的SCADA变量中,因此,如何充分应用SCADA数据进行风电机组状态识别、故障诊断与预警,已经成为风电领域的研究热点,并得到了国内外学者的广泛关注。现有的海上风电机组故障诊断研究已经由传统的数学方法向人工智能方向发展,如采用支持向量机、神经网络、故障树等方法对发电机、齿轮箱等风电机组重要部件进行故障诊断。以上这些诊断方法对数据的学习层次较浅、结构单一,诊断精度有待进一步提高。

[0004] 近年来,人工智能及深度学习的快速发展,为风电机组故障诊断提供了全新的思路。如采用卷积神经网络、长短记忆网络、堆叠自编码等方法对风电机组故障进行诊断,结果表明基于深度学习的故障诊断方法比浅层学习方法具有更高的准确性和泛化性。深度学习算法进行故障诊断主要是利用已知故障样本进行训练,达到故障识别的目的。然而,一旦SCADA系统预警时间不足,则会导致采集的故障样本不足,无法进行故障诊断。另外,风力发电机发生早期轻微故障时,监测的状态量通常并未超出系统阈值,此时SCADA系统无法给出有效的预警,导致故障样本无法获取。

发明内容

[0005] 本发明的目的就是为了解决针对海上风电机组特殊的运行工况以及基于SCADA数据进行故障诊断的现状和存在问题而提供一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,首先用灰色关联分析(Grey Relation Analysis,GRA)对正常运行的风电机组SCADA数据进行分析,选择与发电机温度高度相关的状态变量作为长短期记忆网络

(Long Short-Term Memory, LSTM) 温度预测模型的输入, 计算输出温度预测值与实际值的残差绝对值, 使用统计学方法设置告警阈值对发电机早期故障进行辨识。然后通过 stacking 融合算法对阈值越限的故障样本进行精确故障诊断。最后通过某海上风电场实际数据验证本发明所提方法的有效性。

[0006] 本发明的目的可以通过以下技术方案来实现:

[0007] 一种基于 GRA-LSTM-stacking 模型的海上双馈风力发电机故障诊断方法, 该方法包括以下步骤:

[0008] 步骤1: 通过 GRA 对 SCADA 状态变量进行分析, 筛选出与发电机温度相关的状态变量作为 LSTM 网络的输入;

[0009] 步骤2: 通过 LSTM 对正常状态下的发电机温度进行预测, 得到预测值;

[0010] 步骤3: 计算实际值与预测值的残差绝对值, 使用统计学方法设置告警阈值, 据此对发电机早期故障进行辨识并提取出故障样本;

[0011] 步骤4: 通过 stacking 融合算法对提取的故障样本进行数据处理后输出最终的精确诊断结果。

[0012] 进一步地, 所述的步骤1包括以下分步骤:

[0013] 步骤11: 提取 SCADA 系统采集的风机正常运行的状态数据, 并对其进行归一化处理;

[0014] 步骤12: 计算发电机温度与各个状态变量的灰色关联系数;

[0015] 步骤13: 使用熵权理论优化权重大小, 计算各个状态变量与发电机温度的灰色关联度;

[0016] 步骤14: 对灰色关联度进行排序, 并选择相关的状态变量作为 LSTM 网络的输入。

[0017] 进一步地, 所述的步骤12中的灰色关联系数, 其计算公式为:

$$[0018] \quad \xi_{ij} = \frac{\min_i \min_j |\Delta x| + \rho \max_i \max_j |\Delta x|}{|\Delta x| + \rho \max_i \max_j |\Delta x|}$$

[0019] 式中, ξ_{ij} 为灰色关联系数, ρ 为分辨率, Δx 为作为参考序列的状态变量与作为比较序列的发电机温度的变化量, $i=1, 2, \dots, m, j=1, 2, \dots, n, m$ 为采样样本数目, n 为状态变量数目。

[0020] 进一步地, 所述的步骤13中的灰色关联度, 其计算公式为:

$$[0021] \quad r_j = \frac{1}{m} \sum_{i=1}^m \omega_i \xi_{ij}$$

[0022] 式中, r_j 为灰色关联度, $\sum_{i=1}^m \omega_i$ 由于使用熵权理论进行优化, 即取值为1。

[0023] 进一步地, 所述的步骤2具体包括: 将 GRA 模型提取的状态变量作为输入, 发电机温度作为输出, 接着对 LSTM 网络进行训练, 通过迭代更新权值和偏置使误差最小, 获得发电机温度预测模型, 输出发电机温度预测值。

[0024] 进一步地, 所述的步骤3包括以下分步骤:

[0025] 步骤31:获取发电机温度实际值与预测值的残差绝对值;

[0026] 步骤32:采用统计学方法设置残差绝对值告警阈值;

[0027] 步骤33:据此对发电机早期故障进行辨识,一旦发电机温度残差绝对值超过告警阈值提取出告警之后的状态数据作为下一步精确故障诊断的故障样本。

[0028] 进一步地,所述的步骤32中的告警阈值,其计算公式为:

$$[0029] \int_T^{+\infty} f(\text{Re})d_{\text{Re}} = \alpha$$

[0030] 式中, $f(\text{Re})$ 为概率密度函数, Re 为残差绝对值, T 为告警阈值, α 为显著性水平。

[0031] 进一步地,所述的步骤4具体包括:对步骤3提取的故障样本进行数据集划分,采用K折交叉验证对stacking融合算法第一层基学习器进行训练,输出作为诊断结果的概率输出,组合第一层基学习器输出的概率作为新数据集,并作为第二层元学习器的输入进行训练,输出最终的精确诊断结果。

[0032] 进一步地,所述第一层基学习器中选择的相关算法模型包括:支持向量机、K最邻近分类算法、采用bagging方式集成的随机森林、梯度提升决策树以及采用boosting方式集成的极端梯度提升树,所述第一层基学习器中的各个基学习器经过单独诊断后,还对诊断结果进行spearman关联性分析,所述spearman关联性分析的计算公式为:

$$[0033] \rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

[0034] 式中, ρ_{xy} 为spearman关联性分析对应结果量, \bar{x} 和 \bar{y} 分别为各向量元素的平均值, x_i 和 y_i 分别为各向量元素的实际诊断值。

[0035] 进一步地,所述第二层元学习器中选择的相关算法模型包括GBDT。

[0036] 与现有技术相比,本发明具有以下优点:

[0037] (1) 本发明能有效识别发电机早期故障并精确诊断出具体故障类型,解决了SCADA系统告警时间不足以及故障样本难以获取的问题,有效的避免了故障恶化带来的巨大经济损失。

[0038] (2) 本发明首次将stacking融合算法应用于故障诊断领域,充分利用不同算法从多个角度对SCADA数据进行深入挖掘,使得各种算法能够取长补短,有效的提高了故障诊断精度。

[0039] (3) 本发明具有通用性,可推广应用至其它配备SCADA系统的风电机组中。

附图说明

[0040] 图1为本发明风力发电机故障诊断流程图;

[0041] 图2为本发明中stacking融合算法框架;

[0042] 图3为GRA-LSTM和其他模型的结果对比示意图;

[0043] 图4为GRA-LSTM和其他模型的结果片段对比示意图;

[0044] 图5为三种发电机故障案例中故障F1残差绝对值的变化示意图;

- [0045] 图6为三种发电机故障案例中故障F2残差绝对值的变化示意图；
[0046] 图7为三种发电机故障案例中故障F3残差绝对值的变化示意图；
[0047] 图8为stacking诊断模型相关性分析结果示意图。

具体实施方式

[0048] 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,所描述的实施例是本发明的一部分实施例,而不是全部实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动的前提下所获得的所有其他实施例,都应属于本发明保护的范围。

[0049] 本发明的技术方案如图1所示:

[0050] 一种基于GRA-LSTM-stacking模型的海上双馈风力发电机故障诊断方法,该方法包括如下步骤:

[0051] (1) GRA状态量提取过程:首先,提取SCADA系统采集的风机正常运行的状态数据,并对其进行归一化处理;其次,计算发电机温度与各个状态变量的灰色关联系数;然后,使用熵权理论优化权重大小,计算各个状态变量与发电机温度的灰色关联度;最后,对灰色关联度进行排序,选择高度相关的状态变量作为LSTM网络的输入。

[0052] (2) LSTM温度预测过程:将GRA模型提取的状态变量作为输入,发电机温度作为输出。接着对LSTM网络进行训练,通过迭代更新权值和偏置使误差最小,获得发电机温度预测模型,输出发电机温度预测值。

[0053] (3) 早期故障辨识过程:先计算发电机温度实际值与预测值的残差绝对值。然后,采用统计学方法设置残差绝对值告警阈值。一旦发电机温度残差绝对值超过告警阈值,说明发电机发生故障,提取告警之后的状态数据作为下一步精确故障诊断的数据集。

[0054] (4) stacking精确故障诊断过程:根据上一步截取的发电机故障时段的状态数据进行数据集划分。采用K折交叉验证对stacking融合算法第一层基学习器进行训练,输出诊断结果(概率输出)。组合第一层输出的概率为新数据集,并作为第二层元学习器的输入进行训练,输出最终的精确诊断结果。

[0055] 上述步骤4具体过程如下:

[0056] 作为本发明的创新及关键所在,设计stacking融合算法框架实现对风力发电机多种故障的精确诊断。stacking融合算法是集成学习中的一种,与融合多个相同算法的bagging算法和boosting算法不同,stacking融合算法是将不用类型的机器学习算法通过某种融合方式进行组合,以获得单一机器学习算法无法取得的优异性能。

[0057] stacking融合算法第一层诊断模型训练过程中采用K折交叉验证,具体训练方式如下:将带有标签的数据集S随机均分成K个子数据集 $\{S_1, S_2, \dots, S_k\}$ 。以基学习器1为例,把每一个子数据集分别做一次验证集,余下的K-1个子数据集作为训练集,得到K个模型下的诊断结果(概率输出),组成集合 L_1 , L_1 长度与S相同。对其它的n-1个基学习器执行同样的操作,得到集合 L_2, L_3, \dots, L_n ,组合n个基学习器诊断结果为新数据集 $L = \{L_1, L_2, \dots, L_n\}$ 。新数据集L就是第二层诊断模型元学习器的输入数据。第二层诊断算法能够及时发现并纠正第一层诊断模型中的误差,达到提升诊断模型精度的目的。

[0058] stacking融合算法集成了多样化的算法,可以充分利用不同算法从多个角度分析

数据。基于此,第一层诊断模型基学习器不仅要选择性能优异的算法,也要选择不同类型的算法。第二层诊断模型元学习器选择泛化能力较强的算法,有利于纠正第一层诊断误差,达到最优的诊断效果。本发明方法第一层基学习器初步选择支持向量机(Support Vector Machine,SVM),K最近邻分类算法(K-Nearest Neighbor,KNN),采用bagging方式集成的随机森林(Random Forest,RF)和梯度提升决策树(Gradient Boosted Decision Tree,GBDT)以及采用boosting方式集成的极端梯度提升树(eXtreme Gradient Boosting,XGBoost)。第二层选择GBDT作为元学习器。为了选择最佳的基学习器作为第一层诊断模型,本发明方法设计实验将各个基学习器进行单独诊断,并对其得到的诊断结果进行spearman关联性分析。算法之间的关联度越高,融合后的诊断精度越高。其中,spearman计算公式如下:

$$[0059] \quad \rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

[0060] 式中, ρ_{xy} 为spearman关联性分析对应结果量, \bar{x} 和 \bar{y} 分别为各向量元素的平均值, x_i 和 y_i 分别为各向量元素的实际诊断值。

[0061] 下面结合附图和具体实施例对本发明进行详细说明。

[0062] 本发明的具体实施方法可大致分为以下几步:首先,通过GRA对SCADA状态变量进行分析,筛选出与发电机温度高度相关的状态变量;然后,通过LSTM对正常状态下的发电机温度进行预测,计算实际值与预测值的残差绝对值,使用统计学方法设置告警阈值,据此对发电机早期故障进行辨识并提取出故障样本;最后,通过stacking融合算法对提取的三种发电机故障样本进行精确故障诊断。

[0063] 提取SCADA状态变量中与发电机温度高度相关的变量,利用GRA数值大小定性反映各状态变量对发电机温度影响的权重大小。GRA的思路和计算公式如下:

[0064] 1) 设定状态变量为参考序列,记作:

$$[0065] \quad X_0 = \{x_0(1), x_0(2), \dots, x_0(n)\}$$

[0066] 2) 设定发电机温度为比较序列,记作:

$$[0067] \quad X_i = \{x_i(1), x_i(2), \dots, x_i(n)\}$$

[0068] 以上两式中, $i=1, 2, \dots, m$; m 为采样样本数目; n 为状态变量的数目。

[0069] 3) 计算关联系数。参考序列与第 i 个比较序列在 j 点的关联系数为:

$$[0070] \quad \xi_{ij} = \frac{\min_i \min_j |\Delta x| + \rho \max_i \max_j |\Delta x|}{|\Delta x| + \rho \max_i \max_j |\Delta x|}$$

[0071] 式中, ξ_{ij} 为灰色关联系数, ρ 为分辨率,值介于0到1之间,为了减小极值对计算的影响,在故障诊断中一般取0.5, Δx 为作为参考序列的状态变量与作为比较序列的发电机温度的变化量, $\Delta x = x_0(j) - x_i(j)$, $i=1, 2, \dots, m$, $j=1, 2, \dots, n$, m 为采样样本数目, n 为状态变量数目。

[0072] 4) 计算灰色关联度。

$$[0073] \quad r_j = \frac{1}{m} \sum_{i=1}^m \omega_i \xi_{ij}$$

[0074] 式中, r_j 为灰色关联度, 为了充分考虑比较序列对参考序列的影响权重大小, 本发明使用熵权理论进行优化, 即 $\sum_{i=1}^m \omega_i$ 由于使用熵权理论进行优化, 即取值为1。

[0075] 利用LSTM网络对发电机温度进行预测。LSTM网络关键在于具有记忆功能的细胞元组, 它通过控制3种门结构对细胞状态信息进行有选择的更新。LSTM网络具体计算过程如下:

[0076] 1) 遗忘门状态 f_t : 由当前层输入 x_t 和上一层输出 h_{t-1} 共同决定。

$$[0077] \quad f_t = \sigma(W_1^f g x_t + W_h^f g h_{t-1} + b_f)$$

[0078] 2) 输入门状态 i_t : 与遗忘门相同, 由当前层输入 x_t 和上一层输出 h_{t-1} 共同决定。但是, 输入门增加一步细胞状态 C_t 更新。

$$[0079] \quad i_t = \sigma(W_1^i g x_t + W_h^i g h_{t-1} + b_i)$$

$$[0080] \quad \tilde{C}_t = \tanh(W_1^C g x_t + W_h^C g h_{t-1} + b_C)$$

$$[0081] \quad C_t = i_t g \tilde{C}_t + f_t g C_{t-1}$$

[0082] 3) 输出门状态 o_t : 与输入门类似, 由当前层输入 x_t 和上一层输出 h_{t-1} 共同决定, 增加一步输出 h_t 更新。

$$[0083] \quad o_t = \sigma(W_1^o g x_t + W_h^o g h_{t-1} + b_o)$$

$$[0084] \quad h_t = o_t \times \tanh(C_t)$$

[0085] 以上各式中, W_1^f 、 W_1^i 、 W_1^C 、 W_1^o 分别为当前输入 x_t 与遗忘门、输入门、上一层细胞元组输入和输出门的连接权重; W_h^f 、 W_h^i 、 W_h^C 、 W_h^o 分别为上一层输出 h_{t-1} 与遗忘门、输入门、上一层细胞元组输入和输出门的连接权重; b_f 、 b_i 、 b_C 、 b_o 分别为遗忘门、输入门、上一层细胞元组输入和输出门的偏置; σ 为非线性sigmoid激活函数。

[0086] 风电机组正常运行时, 发电机温度实际值与预测值的残差绝对值很小。一旦发生异常情况, 残差绝对值就会偏离正常水平。随着故障程度的加深, 发电机温度残差绝对值会逐渐变大, 出现显著的抖动上升或者爬坡上升的情况。因此, 可以通过监测发电机温度残差绝对值是否超过告警阈值来辨识发电机早期故障。

[0087] 本发明方法告警阈值的设置采用统计学方法。通过拟合发电机正常运行时的温度残差绝对值, 使之服从某种统计学分布, 对应的概率密度函数为 $f(Re)$, 则告警阈值可以通过下式计算。

$$[0088] \quad \int_T^{+\infty} f(Re) d_{Re} = \alpha$$

[0089] 式中, $f(Re)$ 为概率密度函数, Re 为残差绝对值, T 为告警阈值, α 为显著性水平, 其值介于0到1之间, 为了排除异常工况的干扰, 本发明方法选取0.01。

[0090] 根据上一步截取的发电机故障时段的状态数据, 通过前述设计的stacking融合算

法框架对阈值越限的故障样本进行精确故障诊断。

[0091] 为了验证LSTM网络对发电机温度的预测效果,本发明方法选择均方根误差(root absolute error, RMSE)、平均绝对误差(mean absolute error, MAE)、平均相对误差(mean relative error, MRE)和拟合优度(goodness of fit, R^2),各评价指标的计算公式如下:

$$[0092] \quad RMSE = \sqrt{\frac{1}{n1} \sum_{i=1}^{n1} (Y(i) - Y'(i))^2}$$

$$[0093] \quad MAE = \frac{1}{n1} \sum_{i=1}^{n1} |Y(i) - Y'(i)|$$

$$[0094] \quad MRE = \frac{1}{n1} \sum_{i=1}^{n1} \frac{|Y(i) - Y'(i)|}{Y(i)}$$

$$[0095] \quad R^2 = 1 - \frac{\sum_{i=1}^{n1} (Y(i) - Y'(i))^2}{\sum_{i=1}^{n1} (Y(i) - \bar{Y})^2}$$

[0096] 以上各式中, $Y(i)$ 、 $Y'(i)$ 和 \bar{Y} 分别为第 i 个时刻发电机温度实际值、预测值和平均值; $i=1, 2, \dots, n1$; $n1$ 为测试集时间序列的长度。

[0097] 预测模型评价指标中, RMSE 表示发电机温度预测结果准确性, 预测越准确 RMSE 越小; MAE 和 MRE 表示发电机温度预测结果的一致性, 预测偏差越小 MAE 和 MRE 越小; R^2 表示预测曲线的拟合度, 拟合效果越好 R^2 越接近 1。

[0098] 同时, 本发明方法为了评价 stacking 诊断模型的性能, 采用正确率 R_T 、漏警率 R_F 和 F_1 得分进行评估。各指标计算过程如下:

$$[0099] \quad R_T = \frac{TP}{TP + FP}$$

$$[0100] \quad R_F = \frac{FN}{TP + FN}$$

$$[0101] \quad F_1 = \frac{2R_T(1 - R_F)}{R_T + (1 - R_F)}$$

[0102] 以上各式中, TP 代表正确诊断出特定故障的数目; FP 代表错将其它状态诊断为特定故障的数目; FN 代表错将特定故障诊断为其它状态的数目; F_1 为 R_T 与 R_F 的调和平均值, 代表分类器的综合性能。

[0103] 在诊断模型评价指标中, R_T 和 F_1 的值越大, R_F 的值越小, 表示故障诊断模型性能越好。

[0104] 实际实施例

[0105] 图1是本发明风力发电机故障诊断流程图, 根据流程图的过程将本发明应用到具体实例中。使用国内某海上风电场的数据进行案例研究, 以验证所提方法的可行性。该风电场共34台风机, 单机容量为3MW。选取18号风电机组2012年-2016年SCADA系统每10min采集的状态数据, 首先将采集的状态数据进行筛选, 过滤小于切入风速、大于切出风速的工况参

数。

[0106] 图2是stacking融合算法框架图,stacking融合算法第一层集成了SVM、KNN、RF、GBDT以及XGBoost五种算法,第二层选择GBDT作为元学习器。对初始数据集采用5折交叉验证。

[0107] 图3将GRA提取的9个状态变量 $X = \{T_1, v_1, T_2, I, P, T_3, T_4, v_2, T_5\}$ 作为LSTM温度预测模型的输入变量,发电机温度作为输出变量。LSTM模型的训练集为18号机组2012年1月-6月的正常运行数据,共计17066组;验证集为18号机组2012年7月-8月的正常运行数据,计5053组。设置LSTM模型的输入维数为9,输出维数为1,隐层神经元个数为15,使用随机梯度下降Adam算法优化,损失函数为交叉熵函数。经过训练集多次迭代得出最优温度预测模型,使用该模型对验证集进行预测,所得预测曲线如图3所示。

[0108] 为了检验和验证所提模型在风力发电机温度预测上拥有优异的性能,分别与经典的时间序列预测模型BPNN、ARIMA以及PCA-LSTM模型进行对比分析,结果如图3所示,由于横轴时间跨度较大,对比效果不明显,提取图3中的片段(蓝色虚框),如图4所示。4种温度预测模型的误差如表1所示。

[0109] 表1预测结果的误差分析

| 预测模型 | 评价指标 | | | |
|-------------------------|----------------|----------------|----------------|----------------|
| | RMSE | MAE | MRE | R ² |
| BPNN | 2.424 9 | 1.774 7 | 0.029 5 | 0.967 8 |
| ARIMA | 1.245 8 | 0.572 6 | 0.009 5 | 0.991 5 |
| PCA-LSTM | 1.579 1 | 1.021 2 | 0.017 0 | 0.986 3 |
| GRA-LSTM (本发明方法) | 0.591 3 | 0.398 5 | 0.006 7 | 0.998 1 |

[0111] 从图4可以看出,BPNN模型预测温度的最大误差达到了3.5℃,本说明提出的GRA-LSTM模型预测的结果更接近实际值,是4种模型中预测误差最小的。从表1可知,GRA-LSTM模型的RMSE、MAE和MRE的值分别比其它模型平均降低了1.16、0.72和0.01;R²的值平均提高了0.02。因此,本发明方法所提模型的预测精度更高,为下一步早期故障辨识奠定了基础。

[0112] 图5、图6和图7是三种发电机故障案例。经过GRA-LSTM模型对验证集数据的预测得到发电机温度的预测值,接着计算预测值与实际值的残差绝对值Re。然后,使用统计学方法对Re进行分布拟合,发现Re服从对数正态分布。通过拟合发电机正常运行时的温度残差绝对值得到 $\mu = -6.04741$, $\sigma = 1.19714$,根据相关式子求得本发明方法实施例中的告警阈值T为0.0516。

[0113] 同样使用该风场18号风电机组三种发电机故障案例分析GRA-LSTM模型早期故障识别能力。

[0114] 1) 故障F1:18号机组于2012年10月4日09:55因发电机滑环碳刷磨损停机。经GRA-LSTM模型预测,发电机温度残差绝对值于2012年10月1日09:25超过告警阈值,如图5所示,比实际故障发生提前72.5小时预警,并且残差绝对值超限之前并未出现误报警的情况。

[0115] 2) 故障F2:18号机组于2012年10月6日18:10因发电机三相绕组故障停机。发电机温度残差绝对值于2012年10月8日15:50超过告警阈值,如图6所示,比实际故障发生提前51小时预警。曲线后半段出现残差绝对值减小的情况,但是仍然大于告警阈值,不影响故障辨识。

[0116] 3) 故障F3:18号机组于2016年3月11日15:30因发电机水冷循环故障停机。经GRA-

LSTM模型预测,发电机温度残差绝对值于2016年3月16日01:20超过告警阈值,如图7所示,比实际故障发生提前29.7小时预警。中间过程多次出现残差绝对值接近告警阈值的情况,但是没有触发报警。曲线后半段出现残差绝对值低于告警阈值的情况,但发电机温度残差绝对值增长的趋势很明显,可以判断故障的发生。

[0117] 图8是stacking诊断模型相关性分析。将三种发电机故障案例中温度残差绝对值超过告警阈值的故障数据提取出来组成数据集。依然选择与前述LSTM模型同样的状态变量X作为诊断模型的输入变量。输出为4种发电机状态,即故障F1、故障F2、故障F3和正常状态Nr。首先,把原始数据集拆分成训练集和测试集,将训练集分别输入RF、SVM、KNN、GBDT和XGBoost五种算法中进行模型训练,然后对测试集进行诊断,最后将得到的诊断结果进行spearman相关性分析,结果如图8所示。

[0118] 由图8可知,五种算法的spearman相关系数普遍较高,其中RF、KNN、GBDT和XGBoost的相关度最高,表示融合后的模型诊断效果最好。因此,选取RF、KNN、GBDT和XGBoost作为stacking融合算法第一层诊断模型中的4个基学习器。

[0119] 将训练集输入到构建好的stacking融合算法中进行训练,然后对测试集样本中4种发电机状态进行诊断,stacking诊断结果如表2混淆矩阵所示。混淆矩阵对角线所对应的是诊断正确的数目,非对角线所对应的行或列是诊断错误的数目。

[0120] 表2stacking融合算法诊断结果

| 真实状态 | 预测状态 | | | |
|--------------|--------------------|--------------------|--------------------|--------------------|
| | 正常 Nr ^P | 故障 F1 ^P | 故障 F2 ^P | 故障 F3 ^P |
| [0121] 正常 Nr | 73 | 0 | 0 | 0 |
| 故障 F1 | 0 | 117 | 3 | 0 |
| 故障 F2 | 0 | 3 | 97 | 0 |
| 故障 F3 | 0 | 0 | 0 | 37 |

[0122] 从表2可知,stacking融合算法完全可以诊断出故障F3。对于故障F1和F2,只出现极少次诊断错误的情况,这是因为故障F2与故障F1发生的时间很接近,导致两种故障信息重叠,stacking融合算法未能有效识别。

[0123] 在使用相同数据集的前提下,将stacking融合算法与RF、SVM、KNN、GBDT、XGBoost以及同样使用融合方式的stacking1、stacking2和stacking3诊断模型作对比,各模型的诊断结果如表3所示。

[0124] 由表3可知,采用集成方式的诊断模型(RF、GBDT和XGBoost)比单一诊断模型(SVM和KNN)具有更高的诊断精度。同样采用融合方式的4种stacking模型,诊断精度都比单一算法模型高。其中,经过spearman相关分析提取的4个基学习器融合的stacking模型诊断精度最高,说明spearman相关分析的准确性。另外,stacking融合算法与其它单一算法模型相比,诊断精度平均提高了6.12%。由此验证了stacking融合算法用于发电机故障诊断的有效性。

[0125] 表3诊断模型性能比较

| | 诊断模型 | 诊断精度/% |
|--------|--------------------------|--------------|
| [0126] | RF | 96.97 |
| | SVM | 78.79 |
| | KNN | 91.52 |
| | GBDT | 96.06 |
| | XGBoost | 96.97 |
| | stacking1 | 97.27 |
| | stacking2 | 97.58 |
| | stacking3 | 97.88 |
| | stacking(本发明方法模型) | 98.18 |

[0127] 注:stacking1融合KNN、SVM、XGBoost三种算法;stacking2融合RF、SVM、GBDT、XGBoost四种关联度最低的算法;stacking3融合RF、KNN、SVM、GBDT、XGBoost五种算法。

[0128] 在此基础上,利用所提的诊断模型评价指标对stacking融合算法进行性能评估。各种诊断模型在三种发电机故障中的诊断性能指标如表4所示。

[0129] 表4发电机的三种不同故障识别效果

| 诊断模型 | 故障类型及评价指标 | | | | | | | | |
|----------------|--------------|-------------|--------------|-------------|-------------|-------------|---------------|-------------|--------------|
| | 故障 F1 | | | 故障 F2 | | | 故障 F3 | | |
| | $R_{T1}/\%$ | $R_{F1}/\%$ | $F_{I1}/\%$ | $R_{T2}/\%$ | $R_{F2}/\%$ | $F_{I2}/\%$ | $R_{T3}/\%$ | $R_{F3}/\%$ | $F_{I3}/\%$ |
| [0130] RF | 97.44 | 5.00 | 96.20 | 93.27 | 3.00 | 95.10 | 100.00 | 0.00 | 100.00 |
| SVM | 67.74 | 12.50 | 76.36 | 72.22 | 48.00 | 60.47 | 100.00 | 0.00 | 100.00 |
| KNN | 87.40 | 7.50 | 89.88 | 88.12 | 11.00 | 88.56 | 100.00 | 2.70 | 98.63 |
| GBDT | 94.96 | 5.83 | 94.56 | 93.20 | 4.00 | 94.58 | 100.00 | 0.00 | 100.00 |
| XGBoost | 95.83 | 4.17 | 95.83 | 95.05 | 4.00 | 95.52 | 100.00 | 0.00 | 100.00 |
| 本发明方法模型 | 97.50 | 2.50 | 97.50 | 97.00 | 3.00 | 97.00 | 100.00 | 0.00 | 100.00 |

[0131] 由表4可知,对于故障F1,stacking融合算法具有最高的正确率和最低的漏警率,同时 F_1 得分也是最高的,说明stacking融合算法诊断故障F1的综合性能最好;对于故障F2,stacking融合算法与RF都具有最低的漏警率(3%),但前者的正确率和 F_1 得分在所有诊断模型中是最高的,说明stacking融合算法诊断故障F2的性能最好;对于故障F3,除了KNN之外的诊断模型均能100%的识别,这是因为故障F3的预警时间(29.7小时)是三种故障中最短的,导致训练样本较少,使得适用于大量样本的KNN模型发生误判的情况。因此,再次验证了stacking融合算法在发电机故障诊断方面的有效性。

[0132] 以上所述,仅为本发明的具体实施方式,但本发明的保护范围并不局限于此,任何熟悉本技术领域的技术人员在本发明揭露的技术范围内,可轻易想到各种等效的修改或替换,这些修改或替换都应涵盖在本发明的保护范围之内。因此,本发明的保护范围应以权利要求要求的保护范围为准。

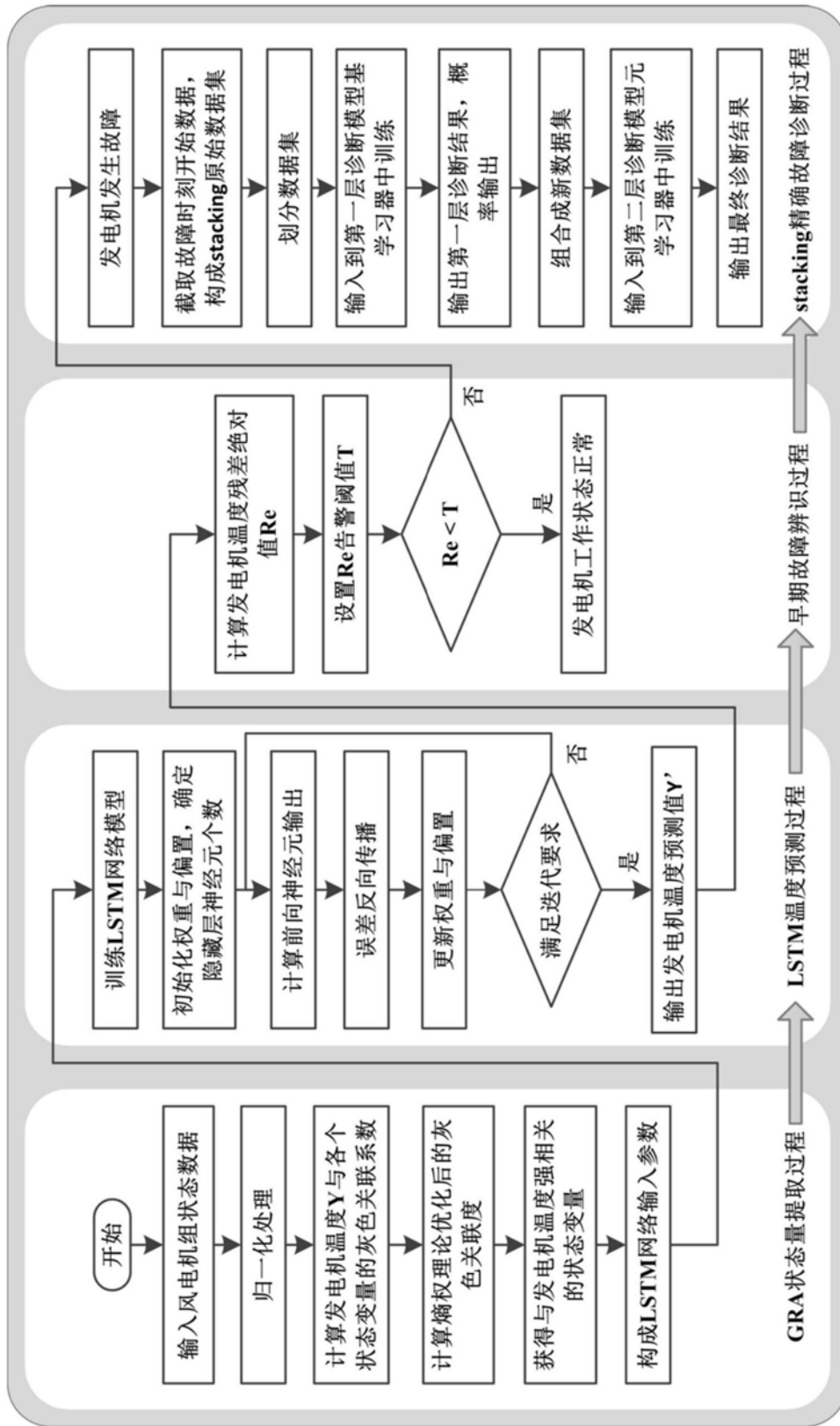


图1

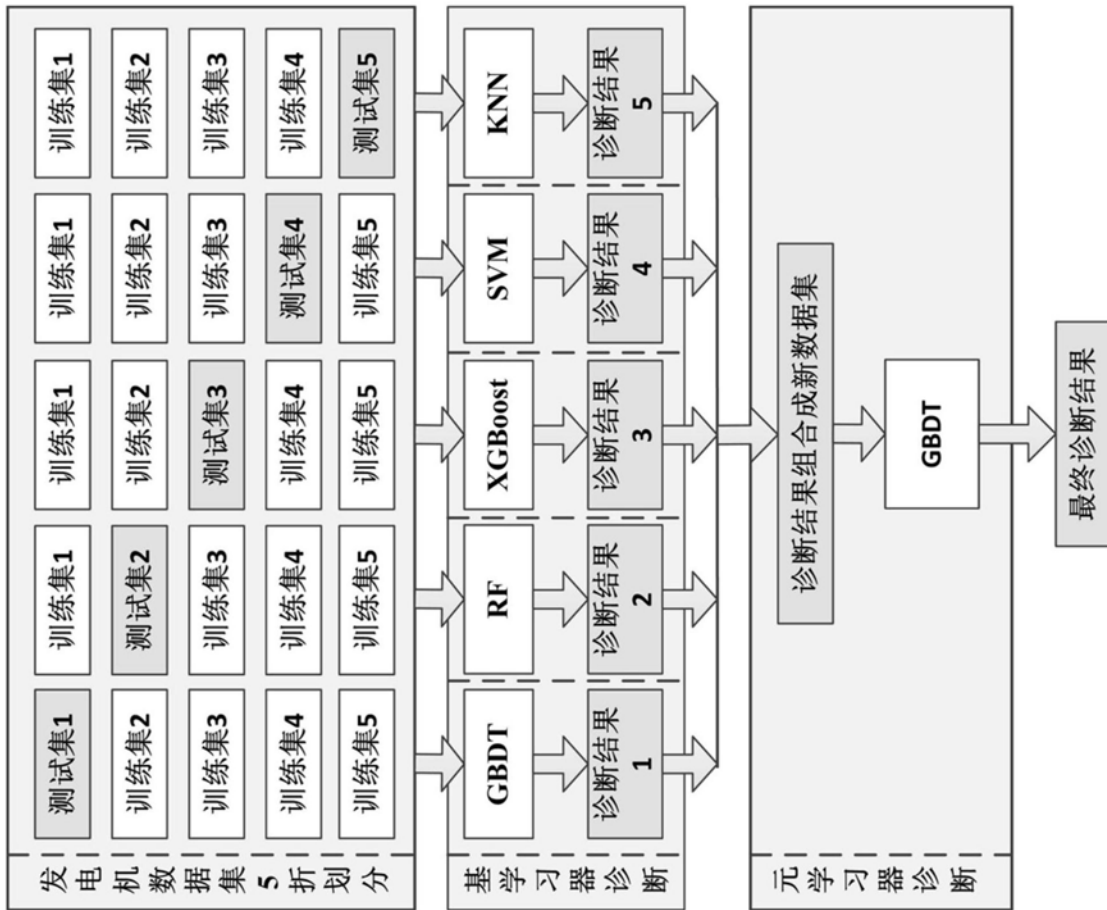


图2

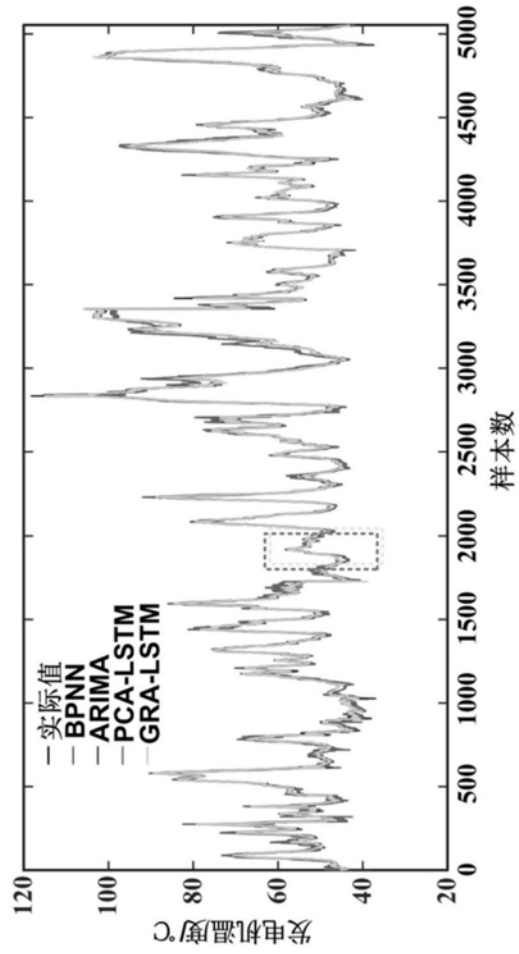


图3

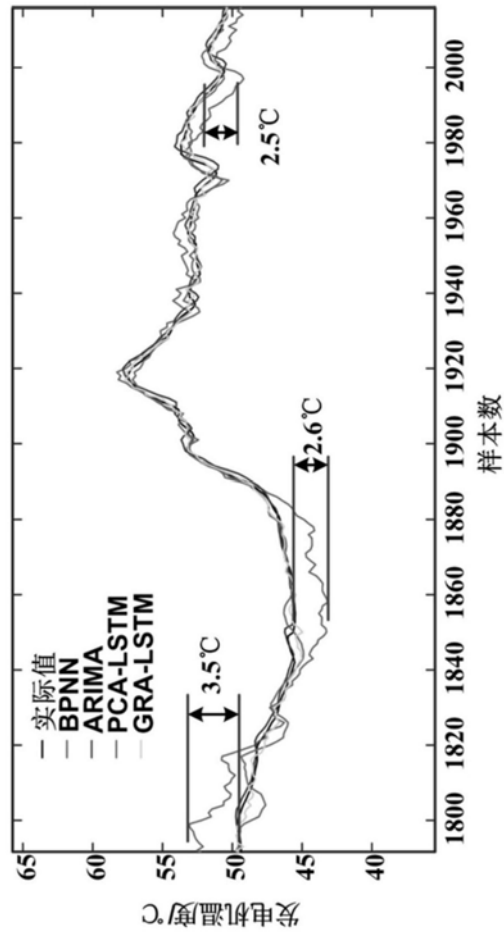


图4

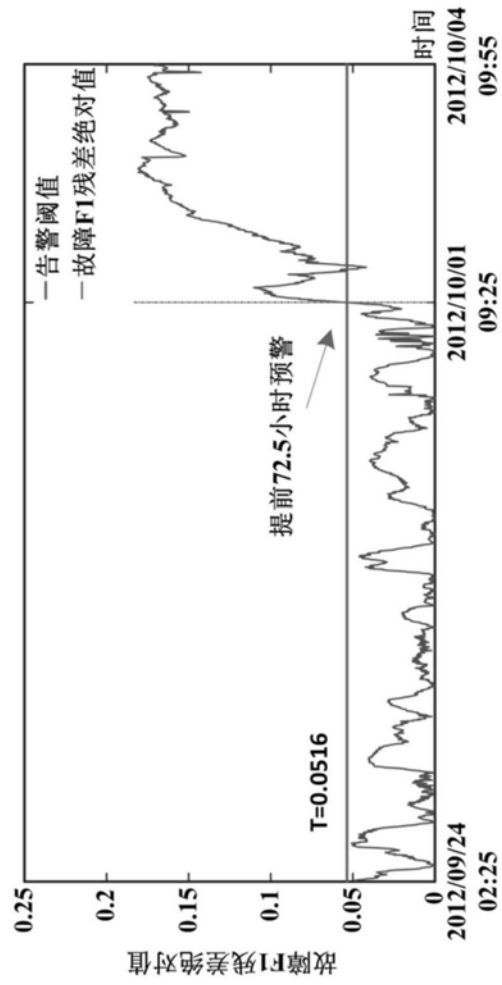


图5

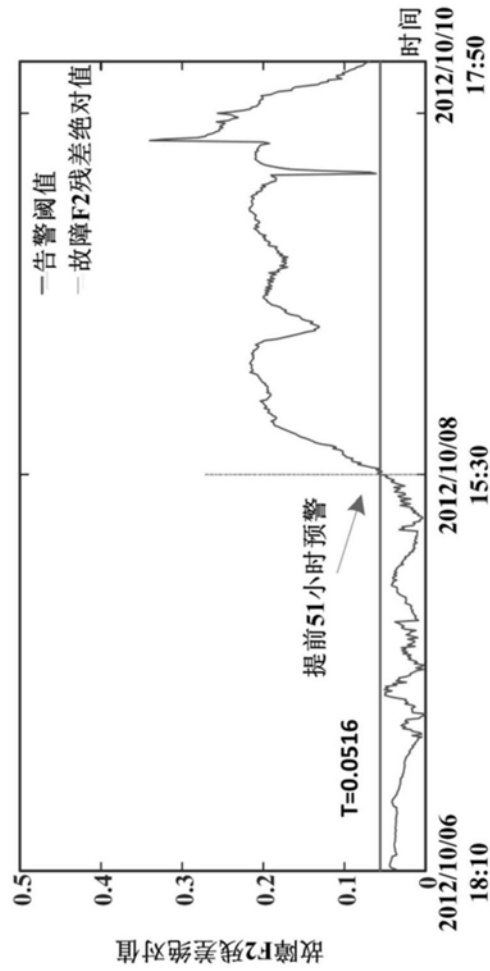


图6

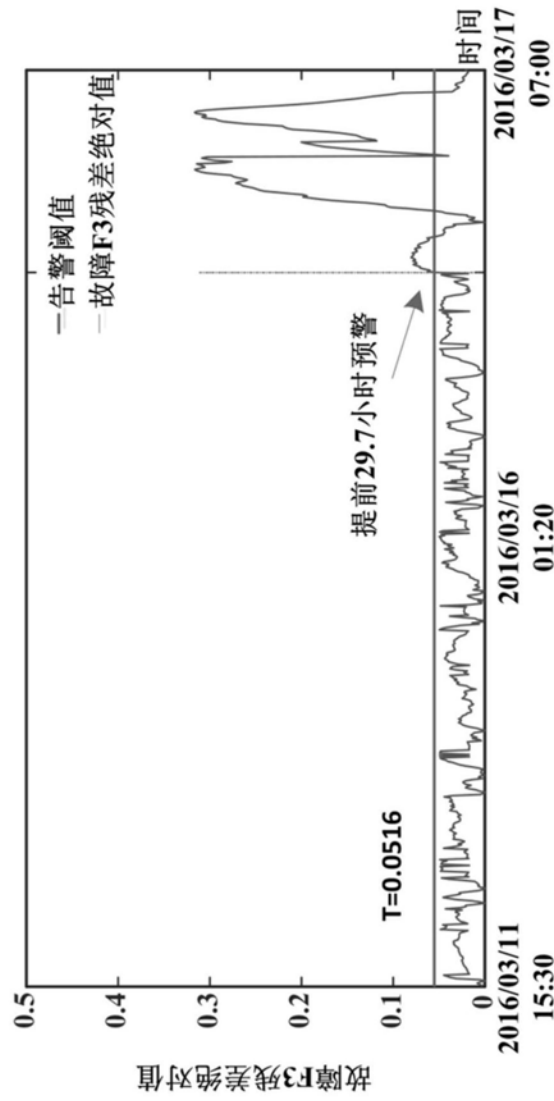


图7

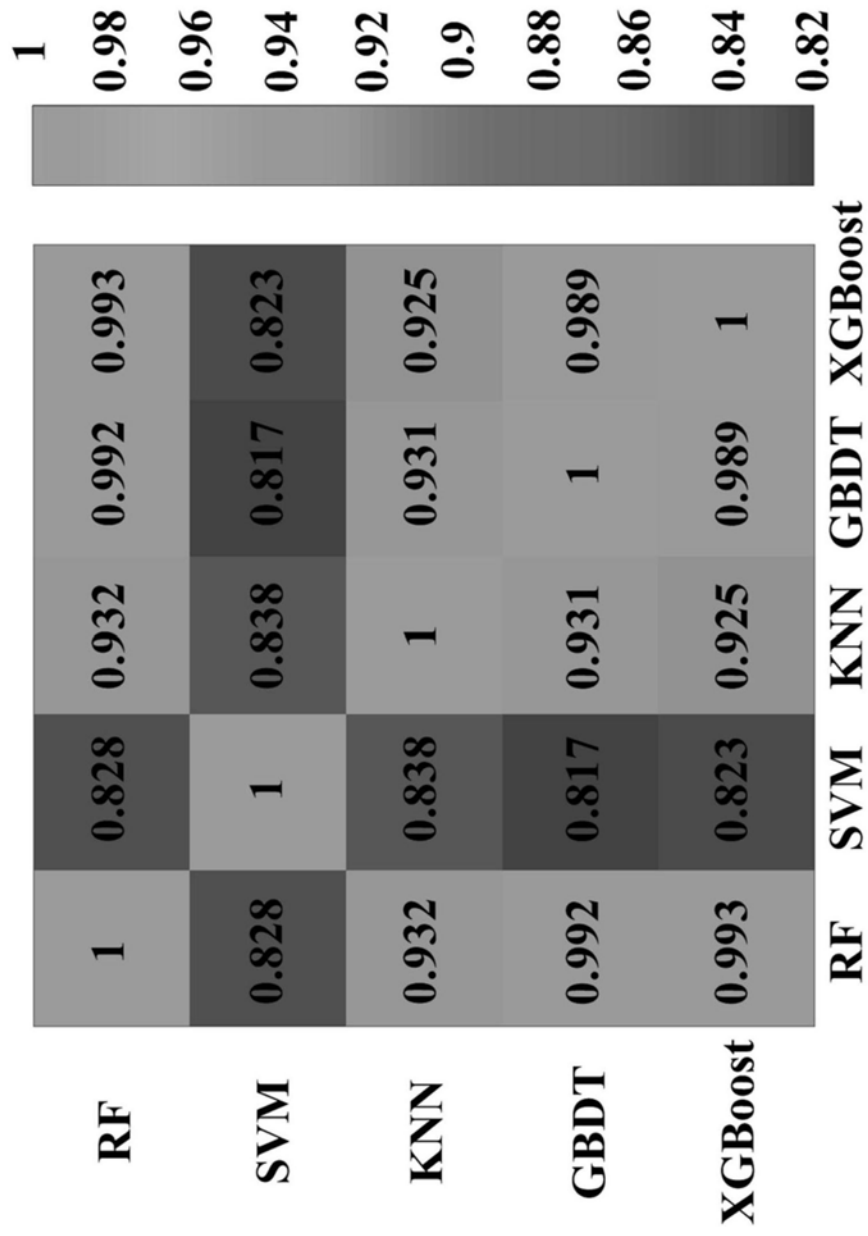


图8