



(12) 发明专利

(10) 授权公告号 CN 113918552 B

(45) 授权公告日 2024.09.24

(21) 申请号 202111187610.3

G06F 17/16 (2006.01)

(22) 申请日 2021.10.12

(56) 对比文件

(65) 同一申请的已公布的文献号

CN 108412710 A, 2018.08.17

申请公布号 CN 113918552 A

CN 111044176 A, 2020.04.21

(43) 申请公布日 2022.01.11

审查员 娄贝贝

(73) 专利权人 西安热工研究院有限公司

地址 710048 陕西省西安市碑林区兴庆路
136号

(72) 发明人 高晨 童博 赵勇 程方 韩毅

宋子琛 张宝锋

(74) 专利代理机构 西安通大专利代理有限责任

公司 61200

专利代理师 白文佳

(51) Int. Cl.

G06F 16/215 (2019.01)

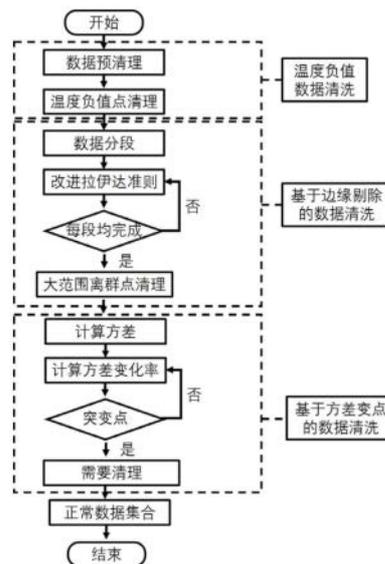
权利要求书2页 说明书7页 附图3页

(54) 发明名称

一种风电机组变频器温度数据清洗方法及系统

(57) 摘要

本发明提供了一种风电机组变频器温度数据清洗方法及系统,包括以下步骤:步骤1,将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵A;步骤2,对步骤1中获取到的矩阵A中的温度超限点数据进行清洗,得到矩阵A';步骤3,对步骤2中获取到的矩阵A'中的大范围温度离群点数据进行清洗,得到矩阵A'';步骤4,对步骤3中获取到的矩阵A''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据;本发明所提数据清洗方法可实现上述三类异常数据的清洗,有效数据删除率低,且无需根据历史数据进行预训练,具有较强的通用性。



1. 一种风电机组变频器温度数据清洗方法,其特征在于,包括以下步骤:

步骤1,将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵 A ;

步骤2,对步骤1中获取得到的矩阵 A 中的温度超限点数据进行清洗,得到矩阵 A' ;

步骤3,对步骤2中获取得到的矩阵 A' 中的大范围温度离群点数据进行清洗,得到矩阵 A'' ;

步骤4,对步骤3中获取得到的矩阵 A'' 中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据;

步骤1中,将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵 A ,具体方法是:

将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据记为矩阵 A ,其中,矩阵 A 为 n 行 2 列,第一列数据为机组运行功率,按大小顺序排列,记为 $P=[P_1, P_2, \dots, P_i, \dots, P_n]$;第二列数据为第一列功率数据中不同功率下对应的温度监测数值,记为 $T=[T_1, T_2, \dots, T_i, \dots, T_n]$;

步骤2中,对步骤1中获取得到的矩阵 A 中的温度超限点数据进行清洗,得到矩阵 A' ,具体方法是:

将矩阵 A 中小于0或大于机组最大允许功率值的第 i 行机组运行功率对应的温度监测值记为0,之后将数值为0的对应行进行剔除,得到矩阵 A' ;

步骤3中,对步骤2中获取得到的矩阵 A' 中的大范围温度离群点数据进行清洗,得到矩阵 A'' ,具体方法是:

对步骤2中获取得到的矩阵 A' 进行分段,得到 L 个子矩阵;

利用改进型拉伊达准则对每个子矩阵中的大范围温度离群点数据进行识别剔除,得到矩阵 A'' ;

利用改进型拉伊达准则对每个子矩阵中的大范围温度离群点数据进行识别剔除,得到矩阵 A'' ,具体方法是:

计算每个子矩阵对应温度向量的期望;并统计每个子矩阵对应温度向量的中位数和众数;

根据得到的期望、中位数和众数计算对应子矩阵的数据分布特性参数;

计算每个子矩阵对应温度向量的标准差;

利用每个子矩阵的数据分布特性参数和标准差,结合改进型拉伊达准则对温度向量中的大范围温度离群点数据进行识别;得到矩阵 A'' ;

将矩阵 A'' 中的大范围温度离群点数据进行剔除,得到矩阵 A'' ;

步骤4,对步骤3中获取得到的矩阵 A'' 中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据,具体方法是:

利用方差变化点法对矩阵 A'' 中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据;

利用方差变化点法对矩阵 A'' 中的剩余温度离群点数据及堆积温度点数据进行清洗,

得到正确的温度监测数据,具体方法是:

计算矩阵 A'' 中每个子矩阵中温度向量中的各点方差,进而得到每个子矩阵对应的方差向量;

计算每个子矩阵对应的方差向量的变化率,得到每个子矩阵对应的变化率向量;

计算每个子矩阵对应的变化率向量构建目标函数;

计算构建得到的目标函数,根据计算得到的目标函数值对变化率向量中的突变点进行剔除,最终得到正确的温度监测数据。

2. 根据权利要求1所述的一种风电机组变频器温度数据清洗方法,其特征在于,对步骤2中获取得到的矩阵 A' 进行分段,得到 L 个子矩阵,具体方法是:

将矩阵 A' 按照矩阵 A' 中的机组运行功率进行分段,得到 L 个子矩阵;其中, L 表示矩阵 A' 中最大功率监测值与最小功率监测值之差除以500并向上取整。

3. 一种风电机组变频器温度数据清洗系统,其特征在于,该系统能够运行权利要求1-2中任一项所述的方法,包括:

数据预处理单元,用于将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵 A ;

第一清洗单元,用于对获取得到的矩阵 A 中的温度超限点数据进行清洗,得到矩阵 A' ;

第二清洗单元,用于对获取得到的矩阵 A' 中的大范围温度离群点数据进行清洗,得到矩阵 A'' ;

第三清洗单元,用于对获取得到的矩阵 A'' 中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据。

一种风电机组变频器温度数据清洗方法及系统

技术领域

[0001] 本发明涉及风力发电领域,具体涉及一种风电机组变频器温度数据清洗方法及系统。

背景技术

[0002] 风力发电机组变频器是风电机组关键的能量转化设备,机组运行过程中变频器高频通断,监测温度成为能够直观反映设备状态的参量,然而通常通过SCADA系统进行监测获取的数据中存在大量异常数据,传统数据清理方法通用性不强,不仅较难实现分类识别,也有可能将正确反映设备温度信息的有效数据过滤,不利于后续分析工作的开展。因此,实现变频器温度数据的有效清洗对变频器健康评价模型的准确建立以及故障诊断、健康状态评估、寿命预测等工作均具有重要意义。

发明内容

[0003] 本发明的目的在于提供一种风电机组变频器温度数据清洗方法及系统,解决了现有技术中存在的上述不足。

[0004] 为了达到上述目的,本发明采用的技术方案是:

[0005] 本发明提供的一种风电机组变频器温度数据清洗方法,包括以下步骤:

[0006] 步骤1,将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵A;

[0007] 步骤2,对步骤1中获取得到的矩阵A中的温度超限点数据进行清洗,得到矩阵A';

[0008] 步骤3,对步骤2中获取得到的矩阵A'中的大范围温度离群点数据进行清洗,得到矩阵A'';

[0009] 步骤4,对步骤3中获取得到的矩阵A''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据。

[0010] 优选地,步骤1中,将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵A,具体方法是:

[0011] 将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据记为矩阵A,其中,矩阵A为n行2列,第一列数据为机组运行功率,按大小顺序排列,记为 $P = [P_1, P_2, \dots, P_i, \dots, P_n]$;第二列数据为第一列功率数据中不同功率下对应的温度监测数值,记为 $T = [T_1, T_2, \dots, T_i, \dots, T_n]$ 。

[0012] 优选地,步骤2中,对步骤1中获取得到的矩阵A中的温度超限点数据进行清洗,得到矩阵A',具体方法是:

[0013] 将矩阵A中小于0或大于机组最大允许功率值的第i行机组运行功率对应的温度监测值记为0,之后将数值为0的对应行进行剔除,得到矩阵A'。

[0014] 优选地,步骤3中,对步骤2中获取得到的矩阵A'中的大范围温度离群点数据进行清洗,得到矩阵A'',具体方法是:

- [0015] 对步骤2中获取得到的矩阵A'进行分段,得到L个子矩阵;
- [0016] 利用改进型拉伊达准则对每个子矩阵中的大范围温度离群点数据进行识别剔除,得到矩阵A''。
- [0017] 优选地,对步骤2中获取得到的矩阵A'进行分段,得到L个子矩阵,具体方法是:
- [0018] 将矩阵A'按照矩阵A'中的机组运行功率进行分段,得到L个子矩阵;其中,L表示矩阵A'中最大功率监测值与最小功率监测值之差除以500并向上取整。
- [0019] 优选地,利用改进型拉伊达准则对每个子矩阵中的大范围温度离群点数据进行识别剔除,得到矩阵A'',具体方法是:
- [0020] 计算每个子矩阵对应温度向量的期望;并统计每个子矩阵对应温度向量的中位数和众数;
- [0021] 根据得到的期望、中位数和众数计算对应子矩阵的数据分布特性参数;
- [0022] 计算每个子矩阵对应温度向量的标准差;
- [0023] 利用每个子矩阵的数据分布特性参数和标准差,结合改进型拉伊达准则对温度向量中的大范围温度离群点数据进行识别;得到矩阵A'';
- [0024] 将矩阵A''中的大范围温度离群点数据进行剔除,得到矩阵A'''。
- [0025] 优选地,步骤4,对步骤3中获取得到的矩阵A'''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据,具体方法是:
- [0026] 利用方差变化点法对矩阵A'''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据。
- [0027] 优选地,利用方差变化点法对矩阵A'''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据,具体方法是:
- [0028] 计算矩阵A'''中每个子矩阵中温度向量中的各点方差,进而得到每个子矩阵对应的方差向量;
- [0029] 计算每个子矩阵对应的方差向量的变化率,得到每个子矩阵对应的变化率向量;
- [0030] 计算每个子矩阵对应的变化率向量构建目标函数;
- [0031] 计算构建得到的目标函数,根据计算得到的目标函数值对变化率向量中的突变点进行剔除,最终得到正确的温度监测数据。
- [0032] 一种风电机组变频器温度数据清洗系统,该系统能够运行所述的方法,包括:
- [0033] 数据预处理单元,用于将获取得到的风电机组变频器的实际运行过程中的温度原始监测数据进行预处理,得到矩阵A;
- [0034] 第一清洗单元,用于对获取得到的矩阵A中的温度超限点数据进行清洗,得到矩阵A';
- [0035] 第二清洗单元,用于对获取得到的矩阵A'中的大范围温度离群点数据进行清洗,得到矩阵A'';
- [0036] 第三清洗单元,用于对获取得到的矩阵A''中的剩余温度离群点数据及堆积温度点数据进行清洗,得到正确的温度监测数据。
- [0037] 与现有技术相比,本发明的有益效果是:
- [0038] 本发明提供了一种风电机组变频器温度数据清洗方法及系统,基于风电机组变频器温度异常数据的分布特征,将异常数据分为温度超限点、大范围离群型异常数据和局部

堆砌型异常数据,根据异常数据分布特征提出基于边缘剔除与方差变点的变频器温度数据清洗方法,通过三个步骤清洗异常数据,实现正常变频器温度数据的提取;通过算例验证结果表明,本发明所提数据清洗方法可实现上述三类异常数据的清洗,有效数据删除率低,且无需根据历史数据进行预训练,具有较强的通用性。

附图说明

- [0039] 图1是本发明的流程示意图;
 [0040] 图2是本发明的流程框图;
 [0041] 图3是变频器温度原始监测数据图;
 [0042] 图4是剔除超限(灰色点)数据后变频器温度监测数据(黑色点)图;
 [0043] 图5是初步剔除大范围温度离群点数据的温度监测数据;
 [0044] 图6是剔除剩余离群点与堆积点的温度监测数据。

具体实施方式

- [0045] 下面结合附图,对本发明进一步详细说明。
 [0046] 实施例具体流程如附图1所示,具体步骤如下:
 [0047] 步骤1,收集实际运行过程中变频器模块的功率-温度原始监测数据,该功率-温度原始监测数据涵盖不同运行功率下的温度监测数据,记为矩阵A。矩阵A为n行2列,第一列数据为机组运行功率监测数据,按大小顺序排列,记为 $P = [P_1, P_2, \dots, P_i, \dots, P_n]$;第二列数据为第一列功率数据中不同功率下对应的温度监测数值,记为 $T = [T_1, T_2, \dots, T_i, \dots, T_n]$;本方案中向量均为列向量,则有:

$$[0048] \quad A = [P \ T] = \begin{bmatrix} P_1 T_1 \\ P_2 T_2 \\ \vdots \\ P_i T_i \\ \vdots \\ P_n T_n \end{bmatrix}$$

- [0049] 其中, T_1 为 P_1 功率条件时的温度监测数值。
 [0050] 根据矩阵A两列数据作出二维分布图,如附图2所示,从附图2中可以看出,矩阵A中不同功率条件下得温度监测数据可以分为正常记录数据、温度超限点数据、大范围温度离群点数据以及堆积温度点数据。

[0051] 步骤2,首先对步骤1得到的矩阵A进行温度超限点数据清除,可根据公式(1)实现此类异常数据的识别,即:将预处理后的温度监测数据中小于0或大于机组最大允许功率值的第i行机组运行功率对应的温度监测值记为0,之后利用常规方法将含有0数据的对应行的数据进行剔除,只保留非0行,完成超限数据剔除后矩阵记为 A' ,结果如图3所示:

- [0052] 若 $P_i < 0$ 或 $P_i > P_m$,则 $T'_i = 0$;
 [0053] $T'_t = T_t$,若 $0 \leq P_i \leq P_m$,则 $T'_i = T_i$ (1)
 [0054] $1 \leq i \leq n$

- [0055] 其中, P_i 为矩阵A中第i行对应的功率监测数据值; P_m 为机组允许的最大功率;
 [0056] T'_i 为矩阵A中第i行第二列的温度监测数值。

[0057] 经过温度超限点数据剔除得到矩阵A'如下所示:

$$[0058] \quad A' = [P' \ T'] = \begin{bmatrix} P_j & T_j \\ P_{j+1} & T_{j+1} \\ \vdots & \vdots \\ P_l & T_l \\ \vdots & \vdots \\ P_q & T_q \end{bmatrix}$$

[0059] 矩阵A'表明将原矩阵前j-1行、末尾n-q行对应的温度超限点数据剔除,则对于矩阵A'有:

$$[0060] \quad P_j = \min(P');$$

$$[0061] \quad P_q = \max(P')$$

[0062] 步骤3,对步骤2中得到的矩阵A'按照第一列即P'中功率监测数据的取值范围进行分段,得到L个子矩阵,分别记为A1、A2、...AU、...AL,其中AU也可以表示[PU TU],设置分段数为L,L表示矩阵A'中最大功率监测值与最小功率监测值之差除以500并向上取整,即:

$$[0063] \quad L = \left(\frac{P_q - P_j}{500} \right)$$

[0064] 则分段完成后矩阵A'可以表示为:

$$[0065] \quad A' = [P' \ T'] = \begin{bmatrix} A1 \\ A2 \\ \vdots \\ AU \\ \vdots \\ AL \end{bmatrix} = \begin{bmatrix} P1 \ T1 \\ P2 \ T2 \\ \vdots \\ PU \ TU \\ \vdots \\ PL \ TL \end{bmatrix}, 1 \leq U \leq L$$

[0066] 其中,P1的取值范围为:

$$[0067] \quad P1_{max} = \left(\frac{P_q - P_j}{L} \right)$$

$$[0068] \quad P1_{min} = P_j$$

[0069] T1表示P1取值范围内对应的温度数据,t表示数据个数:

$$[0070] \quad T1 = [T_{11}, T_{12}, \dots, T_{1t}]$$

[0071] 同理,可以得到PU的取值范围为:

$$[0072] \quad PU_{max} = \left(\frac{P_q - P_j}{L} \right) \times U$$

$$[0073] \quad PU_{min} = \left(\frac{P_q - P_j}{L} \right) \times (U - 1)$$

$$[0074] \quad 1 < U \leq L$$

[0075] 步骤4,根据步骤3中分段结果对大范围温度离群点数据进行识别与剔除,以剔除矩阵A'的子矩阵A1大范围温度离群点数据为例,首先计算子矩阵A1对应的温度向量T1的期望 E_{P1} ,统计此组数据的中位数 Mid_{P1} 和众数 $Ma_{j_{P1}}$,并根据期望 E_{P1} 、中位数 Mid_{P1} 与众数 $Ma_{j_{P1}}$,计算反映此组的数据分布特性参数 λ_{P1} ,计算方式如下:

$$[0076] \quad \lambda_{P1} = Mid_{P1}, \frac{1}{2} \times \frac{|E_{P1} + Maj_{P1}|}{Mid_{P1}} \times 100\% \leq 5\%;$$

$$[0077] \quad \lambda_{P1} = \frac{1}{2} \times |Maj_{P1} + Mid_{P1}|, \text{其他};$$

[0078] 步骤5, 计算子矩阵A1对应温度向量T1的标准差 σ_{P1} , 并根据改进型拉伊达准则对温度向量T1的每个数据进行大范围离群数据点数据识别, 每个数据识别结果记为 $Iden_x$, 具体计算方法如下:

$$[0079] \quad Iden_x = 1, T_x < \lambda_{P1} \pm 3\sigma_{P1};$$

$$[0080] \quad Iden_x = 0, T_x \geq \lambda_{P1} \pm 3\sigma_{P1};$$

[0081] 其中, T_x 表示温度向量T1的第x个数据, 且有 $1 \leq x \leq t$, 则经过计算可以得到一组由0、1组成的识别结果向量, 记为 $Iden(T1)$, 有:

$$[0082] \quad Iden(T1) = [Iden_1, Iden_2, Iden_3, \dots, Iden_x, \dots, Iden_t]$$

[0083] 步骤6, 根据步骤5中识别结果剔除子矩阵A1对应的温度向量T1中大范围温度离群点数据, 将识别结果向量 $Iden(T1)$ 的每一个元素与温度向量T1向量对应元素相乘, 得到 $T1''$, 同理得到 $P1''$, 具体计算方法如下:

$$[0084] \quad T1'' = [Iden_1 T_{11}, Iden_2 T_{12}, Iden_3 T_{13}, \dots, Iden_x T_{1x}, \dots, Iden_t T_{1t}];$$

$$[0085] \quad P1'' = [Iden_1 P_{11}, Iden_2 P_{12}, Iden_3 P_{13}, \dots, Iden_x P_{1x}, \dots, Iden_t P_{1t}]$$

$$[0086] \quad A1'' = [P1'' T1'']$$

[0087] 其中, $A1''$ 为经过识别后的矩阵, 包含若干全为0元素的行。

[0088] 利用常规方法将带有0元素的行剔除, 只保留非0元素的行, 完成大范围温度离群点数据的剔除, 记为:

$$[0089] \quad A1''' = [P1''' T1''']$$

[0090] 步骤7, 重复步骤4至步骤6, 完成 $A1, A2, \dots, AU, \dots, AL$ 对应温度向量 $T1, T2, \dots, TU, \dots, TL$ 中大范围离群点数据的识别与剔除, 则经过大范围温度离群点数据的识别与剔除的数据记为 A''' , 有:

$$[0091] \quad A''' = [P''' T'''] = \begin{bmatrix} A1''' \\ A2''' \\ \vdots \\ AU''' \\ \vdots \\ AL''' \end{bmatrix} = \begin{bmatrix} P1''' T1''' \\ P2''' T2''' \\ \vdots \\ PU''' TU''' \\ \vdots \\ PL''' TL''' \end{bmatrix}, 1 \leq U \leq L$$

[0092] 结果如附图4所示。

[0093] 步骤8, 对步骤7获得的数据矩阵 A''' , 采用方差变化点法识别剩余温度离群点数据及堆积温度点数据, 以识别 $A1'''$ 中的异常数据为例, 首先求出 $A1'''$ 中的温度向量 $T1'''$ 的各点方差, 方差计算公式为:

$$[0094] \quad s_h = \frac{\sum_{h=1}^k (T'''_h - \lambda_g)^2}{h}$$

[0095] 其中, T'''_h 为温度向量 $T1'''$ 中第h个元素; λ_g 为温度向量 $T1'''$ 中前h个元素的分布特性参数, 计算方式与步骤4中分布特性的参数计算方法相同; s_h 为第h个点的方差, 用以表征

第h点数据与前h-1个数据之间的离散程度,k表示T1''的数据个数,则通过计算可以获得一组方差向量记为S1,有:

$$[0096] \quad S1 = [s_1, s_2, s_3, \dots, s_h, \dots, s_k]$$

[0097] 步骤9,计算步骤8获得的方差向量S1的变化率,计算方法如下:

$$[0098] \quad v_h = \frac{|s_{h-1} - s_h|}{s_{h-1}}, 1 < h \leq k$$

[0099] 其中, s_{h-1} 与 s_h 分别为前h项数据方差与前h-1项数据方差, v_h 表示前h项数据方差与前h-1项数据方差之间的变化率,逐项计算后可以得到变化率向量V1,有:

$$[0100] \quad V1 = [v_2, v_3, \dots, v_h, \dots, v_k]$$

[0101] 步骤10,针对步骤9获得的变化率向量V1,识别向量中变化率突变的数据点,构建目标函数Targ_h,以识别 v_h 是否为突变点为例:

[0102] 首先定义 v_h 的一个邻域作为目标计算向量,记为 $v_{h\mu}$,则有:

$$[0103] \quad v_{h\mu} = [v_{h-\mu}, v_{h-\mu+1}, \dots, v_h, \dots, v_{h+\mu-1}, v_{h+\mu}]$$

[0104] 其中, $\mu \ll k$,则根据统计规律可以得出,若 v_h 对应的数据点为变化率突变点,则在 v_h 处,必然同时满足两个线性方程,即:

$$[0105] \quad v_h = \beta_1 y + \varphi_1$$

$$[0106] \quad v_h = \beta_2 y + \varphi_2$$

[0107] 其中 β_1 、 β_2 、 φ_1 、 φ_2 为模型参数,y为模型中自变量。

[0108] 则可构建目标函数Targ_h:

$$[0109] \quad Targ_h = \sum_{\alpha=h-\mu}^h [v_\alpha - \beta_1 y + \varphi_1]^2 + \sum_{\alpha=h}^{h+\mu} [v_\alpha - \beta_2 y + \varphi_2]^2$$

[0110] 步骤11,计算步骤10获得的目标函数Targ_h,若Targ_h为 v_h 邻域内最小,则识别 v_h 为向量中变化率突变的点,反之,则表明 v_h 不为向量中变化率突变的点,即:

[0111] T1''向量中第h个数据需要剔除,即 $T1h''' = 0$,Targ_h为邻域内最小;

[0112] T1''向量中第h个不数据需要剔除,即 $T1h''' = T1h''$,Targ_h不为邻域内最小

[0113] 步骤12,重复步骤11,计算h不同取值条件下的目标函数Targ_h,识别出所有需要剔除的点,对于矩阵A1'',利用传统方法将含有0元素的行剔除,生成矩阵A1''',即代表剔除完堆积数据与小范围离群数据的结果。

[0114] 步骤13,重复步骤8至步骤12,完成整个A''矩阵堆积数据与小范围离群数据清理,最终结果记为A''',如图5所示。

[0115] 本发明的工作原理:

[0116] 本发明基于风电机组变频器温度异常数据的分布特征,将异常数据分为温度超限点数据、大范围温度离群点数据以及堆积温度点数据。根据异常数据分布特征提出基于边缘剔除与方差变点的变频器温度数据清洗方法,通过三个步骤清洗异常数据,实现正常变频器温度数据的提取。通过算例验证结果表明,本文所提数据清洗方法可实现上述三类异常数据的清洗,有效数据删除率低,且无需根据历史数据进行预训练,具有较强的通用性。

[0117] 本发明所采用的改进型拉伊达准则对温度向量中的每个数据进行大范围离群数

据点数据识别,传统的拉伊达准则利用中位数加减三倍标准差进行识别,而本申请利用了统计中的众数、中位数和期望计算的参数加减三倍标准差进行识别,拓宽了应用范围,拉伊达准则效果最好的就是正态分布和近似正态分布,但是采集的数据有可能不符合正态分布。

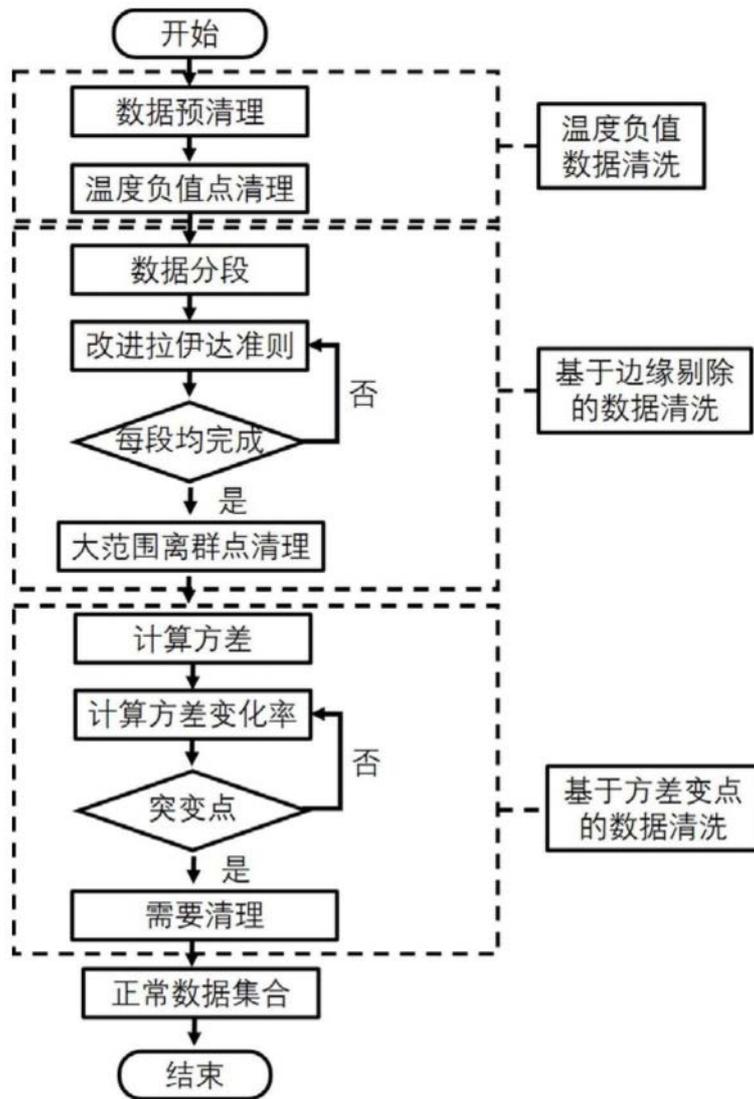


图1

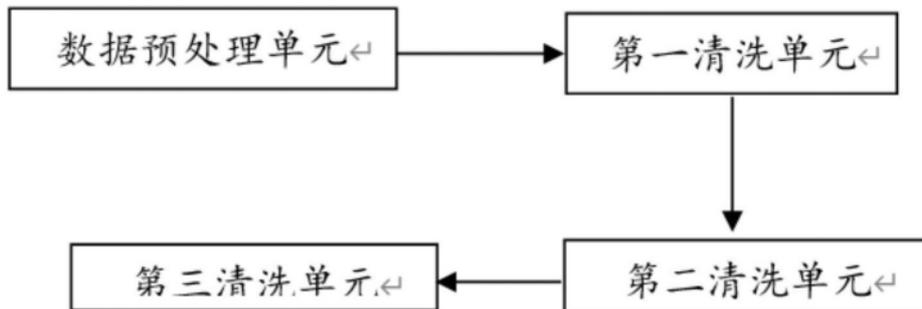


图2

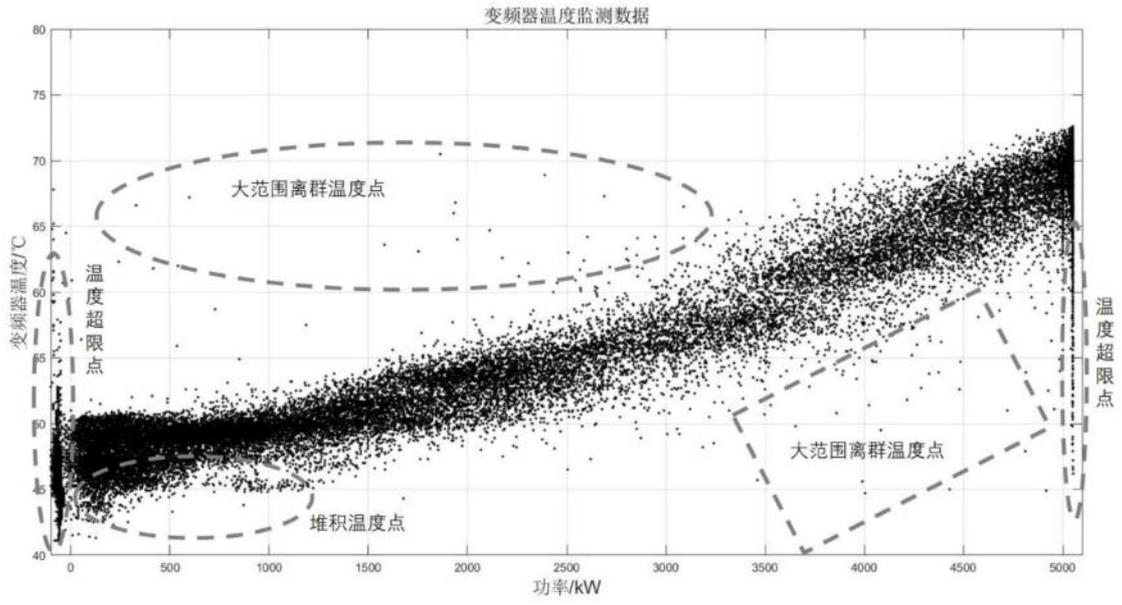


图3

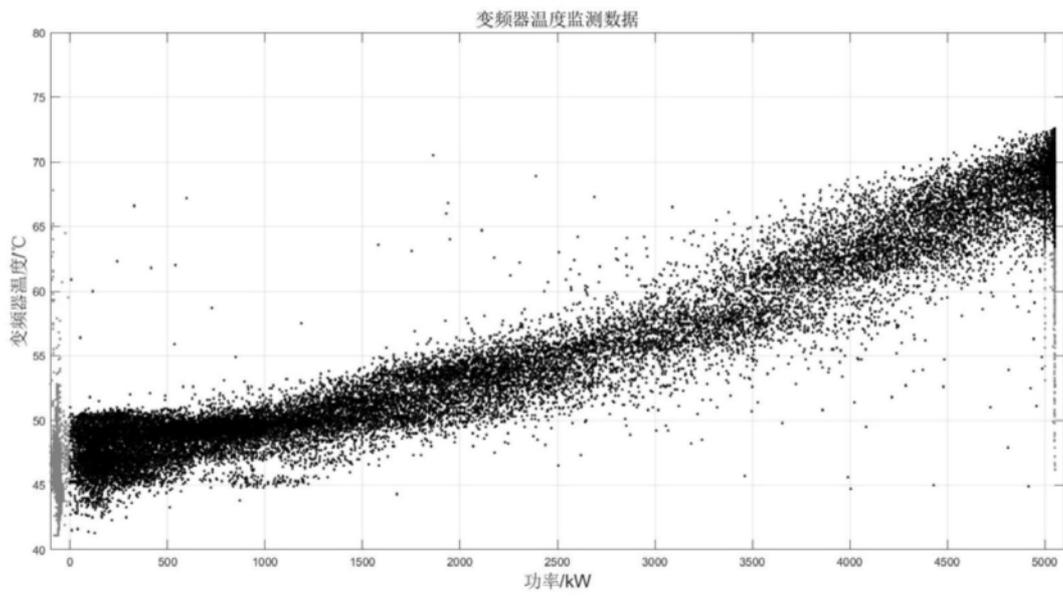


图4

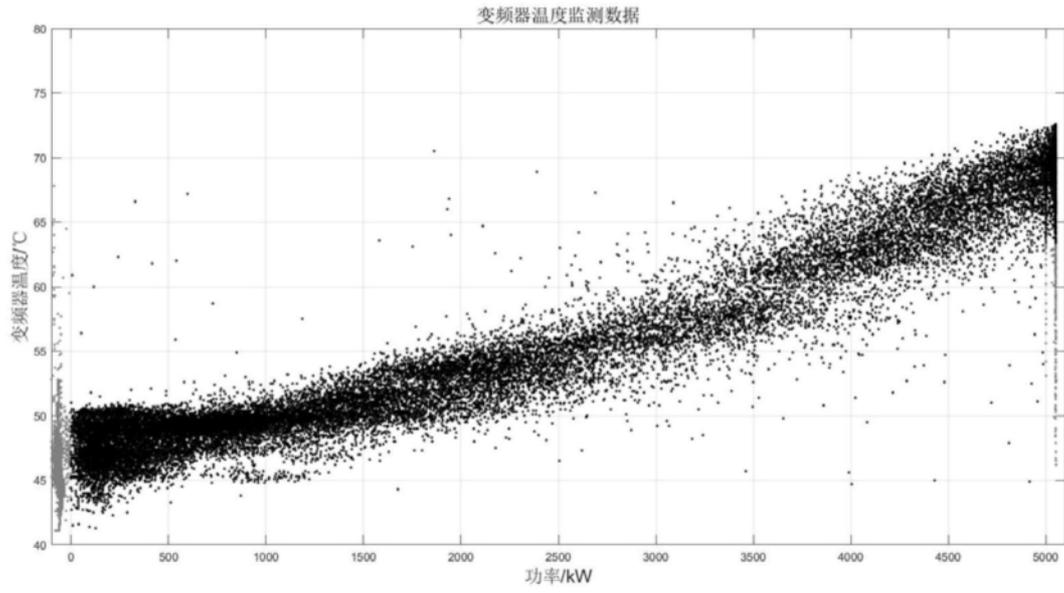


图5

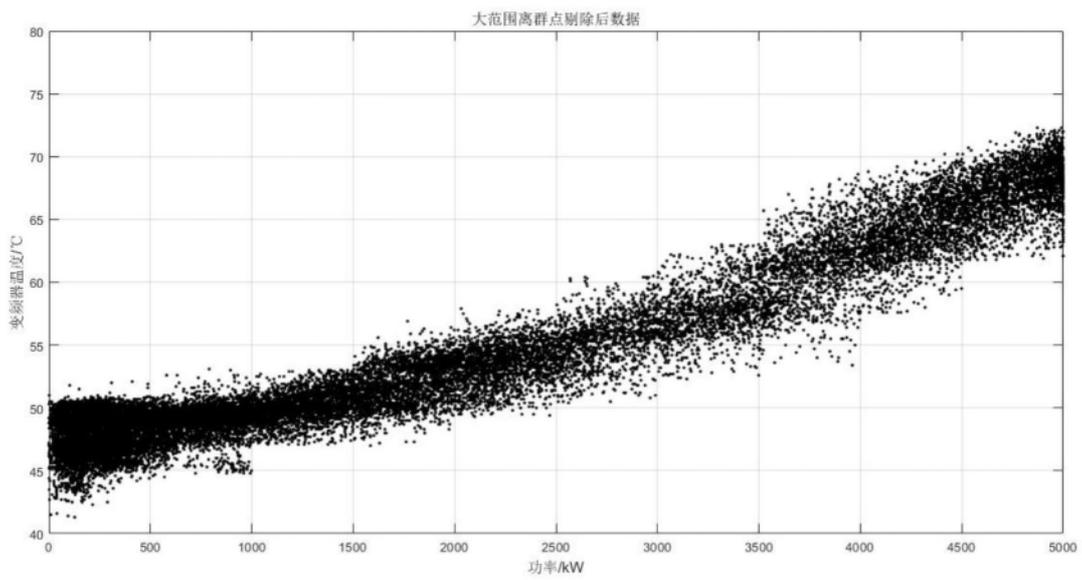


图6