



(12) 发明专利

(10) 授权公告号 CN 115146881 B

(45) 授权公告日 2023.01.13

(21) 申请号 202211081062.0

(22) 申请日 2022.09.05

(65) 同一申请的已公布的文献号
申请公布号 CN 115146881 A

(43) 申请公布日 2022.10.04

(73) 专利权人 广东电网有限责任公司江门供电局
地址 529000 广东省江门市蓬江区建设二路152号

(72) 发明人 陈钢 李波 武建平 邓瑞麒
郑广勇 晏梦璇 郭亮 柳军停
黄伟杰 陈锦洪 郑文杰 卢伟钿
易晋 黄定威 张锦添

(74) 专利代理机构 北京集佳知识产权代理有限公司 11227
专利代理师 任文生

(51) Int.Cl.

G06Q 10/04 (2012.01)

G06Q 50/06 (2012.01)

H02J 3/00 (2006.01)

(56) 对比文件

CN 106022528 A, 2016.10.12

CN 104123682 A, 2014.10.29

CN 108053149 A, 2018.05.18

EP 3640869 A1, 2020.04.22

US 2017018038 A1, 2017.01.19

CN 108846527 A, 2018.11.20

郭辉. 基于电站群聚类 and 相似气象日的光伏电站不良数据修复.《优秀硕士学位论文工程科技II辑》.2019,

审查员 杨黎鹏

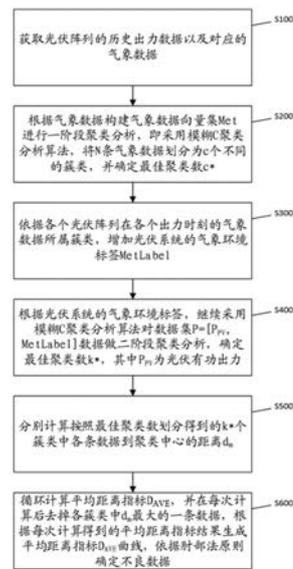
权利要求书2页 说明书6页 附图2页

(54) 发明名称

一种低压分布式光伏系统不良数据检测方法和系统

(57) 摘要

本发明提供了一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法和系统,通过采集光伏出力数据对应的气象数据,构建气象数据向量集,并通过一阶段模糊C聚类分析,按照各时刻光伏出力的气象数据所属簇类,标记气象环境标签。在二阶段模糊C聚类过程中,将光伏出力数据与气象环境标签组合构建数据集,依据偏离指标选择最佳聚类数,让二阶段聚类划分与一阶段的聚类划分尽可能贴合。根据最终光伏有功出力的划分簇类,依次剔除最外围数据,并绘制平均距离指标DAVE的下降曲线,依据肘部法原则检测不良数据。本发明可以准确检测到传输到数据存储端的部分不良数据,降低其对光伏系统的特性分析以及出力预测的影响。



1. 一种低压分布式光伏系统不良数据检测方法,其特征在于,包括:

获取光伏阵列的历史出力数据以及对应的气象数据;

根据所述气象数据构建气象数据向量集 Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将 N 条气象数据划分为 c 个不同的簇类,并确定最佳聚类数 c^* ;

依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气象环境标签 $MetLabel$;

根据所述光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{PV}, MetLabel]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{PV} 为光伏有功出力;

分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m ;

循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

2. 根据权利要求1所述的低压分布式光伏系统不良数据检测方法,其特征在于,所述气象数据具体包括:气温 $Temp$ 、光照强度 Lux 、空气湿度 RH 、云量 TCC 。

3. 根据权利要求2所述的低压分布式光伏系统不良数据检测方法,其特征在于,所述一阶段聚类分析具体包括:

令初始聚类数 $c=2$,采用模糊C聚类算法对所述气象数据向量集 $Met = [Temp, Lux, RH, TCC]$ 做聚类分析;

根据XB聚类有效性指标确定最佳聚类数 c^* ,构建气象环境指标 $1, 2, \dots, c^*$;

依据各个采样时刻光伏出力对应的气象数据所属簇类,对光伏处理数据标记气象环境标签 $MetLabel$ 。

4. 根据权利要求1所述的低压分布式光伏系统不良数据检测方法,其特征在于,所述二阶段聚类分析具体包括:

令初始聚类数 $k = 2$,采用模糊C算法对 N 条数据集 $P = [P_{PV}, MetLabel]$ 做聚类分析,根据偏离指标最小选择最佳聚类数 k^* ;

根据所述最佳聚类数结果,分别计算各个簇类的数据到聚类中心的距离 d_m ;

根据计算所得的各个簇类数据的距离值,计算平均距离指标 D_{AVE} ;

去掉各簇类中 d_m 最大的一条数据,更新数据集并重新计算平均距离指标 D_{AVE} ,按照每次计算一次 D_{AVE} 的同时去掉一条 d_m 最大的数据的规则,生成平均距离指标 D_{AVE} 曲线,依据肘部法原则筛选不良数据。

5. 根据权利要求1或4所述的低压分布式光伏系统不良数据检测方法,其特征在于,依据肘部法原则确定不良数据具体为:

确定所述平均距离指标 D_{AVE} 曲线中的拐点,将所述拐点对应的数据点记为不良数据点,做剔除处理。

6. 一种低压分布式光伏系统不良数据检测系统,其特征在于,包括:

数据获取单元,用于获取光伏阵列的历史出力数据以及对应的气象数据;

第一聚类分析单元,用于根据所述气象数据构建气象数据向量集 Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将 N 条气象数据划分为 c 个不同的簇类,并确定最佳聚类数 c^* ;还用于依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气

象环境标签MetLabel;

第二聚类分析单元,用于根据所述光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{pv}, \text{MetLabel}]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{pv} 为光伏有功出力;还用于分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m ;还用于循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

7. 根据权利要求6所述的低压分布式光伏系统不良数据检测系统,其特征在于,所述气象数据具体包括:气温Temp、光照强度Lux、空气湿度RH、云量TCC。

8. 根据权利要求7所述的低压分布式光伏系统不良数据检测系统,其特征在于,在所述第一聚类分析单元中,所述一阶段聚类分析具体包括:

令初始聚类数 $c=2$,采用模糊C聚类算法对所述气象数据向量集 $\text{Met} = [\text{Temp}, \text{Lux}, \text{RH}, \text{TCC}]$ 做聚类分析;

根据XB聚类有效性指标确定最佳聚类数 c^* ,构建气象环境指标 $1, 2, \dots, c^*$;

依据各个采样时刻光伏出力对应的气象数据所属簇类,对光伏处理数据标记气象环境标签MetLabel。

9. 根据权利要求6所述的低压分布式光伏系统不良数据检测系统,其特征在于,在所述第二聚类分析单元中,所述二阶段聚类分析具体包括:

令初始聚类数 $k = 2$,采用模糊C算法对 N 条数据集 $P = [P_{pv}, \text{MetLabel}]$ 做聚类分析,根据偏离指标最小选择最佳聚类数 k^* ;

根据所述最佳聚类数结果,分别计算各个簇类的数据到聚类中心的距离 d_m ;

根据计算所得的各个簇类数据的距离值,计算平均距离指标 D_{AVE} ;

去掉各簇类中 d_m 最大的一条数据,更新数据集并重新计算平均距离指标 D_{AVE} ,按照每次计算一次 D_{AVE} 的同时去掉一条 d_m 最大的数据的规则,生成平均距离指标 D_{AVE} 曲线,依据肘部法原则剔除不良数据。

10. 根据权利要求6或9所述的低压分布式光伏系统不良数据检测系统,其特征在于,在所述第二聚类分析单元中,依据肘部法原则确定不良数据具体为:

确定所述平均距离指标 D_{AVE} 曲线中的拐点,将所述拐点对应的数据点记为不良数据点,做剔除处理。

一种低压分布式光伏系统不良数据检测方法和系统

技术领域

[0001] 本发明属于光伏出力预测技术领域,具体涉及一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法和系统。

背景技术

[0002] 近年来随着国家政策的大力支持,光伏技术的日益成熟,光伏组件的安装成本也呈指数级下降。因此,在我国各个地区出现了集中式或分布式光伏系统。然而,光伏电站的大规模建设也给此类可再生能源的准确监控带来技术难题。目前,用以监控光伏系统数据的设备主要包括:SCADA系统监控下光伏接入点可获取的有功功率以及无功功率;PMU系统监控下光伏接入点可检测的高频相量数据。随着光伏容量的快速增加,各式数据采集设备也相应投入使用,导致系统中的监控数据爆炸式增长。由于数据采集装置的量测误差,以及数据传输过程中的通信故障,通常容易导致传输到数据存储端的数据存在部分不良数据。这些不良数据的存在对后续光伏系统的特性分析以及出力预测都带来间接影响。

发明内容

[0003] 有鉴于此,本发明提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法和系统,旨在准确检测传输到数据存储端的部分不良数据,降低其对光伏系统的特性分析以及出力预测的影响。

[0004] 为了解决上述技术问题,本发明提供以下技术方案:

[0005] 第一方面,本发明提供了一种低压分布式光伏系统不良数据检测方法,包括:

[0006] 获取光伏阵列的历史出力数据以及对应的气象数据;

[0007] 根据气象数据构建气象数据向量集 Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将 N 条气象数据划分为 c 个不同的簇类,并确定最佳聚类数 c^* ;

[0008] 依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气象环境标签 $MetLabel$;

[0009] 根据光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{PV}, MetLabel]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{PV} 为光伏有功出力;

[0010] 分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m ;

[0011] 循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

[0012] 进一步的,气象数据具体包括:气温 $Temp$ 、光照强度 Lux 、空气湿度 RH 、云量 TCC 。

[0013] 进一步的,一阶段聚类分析具体包括:

[0014] 令初始聚类数 $c=2$,采用模糊C聚类算法对气象数据向量集 $Met = [Temp, Lux, RH, TCC]$ 做聚类分析;

[0015] 根据XB聚类有效性指标确定最佳聚类数 c^* ,构建气象环境指标 $1, 2, \dots, c^*$;

[0016] 依据各个采样时刻光伏出力对应的气象数据所属簇类,对光伏处理数据标记气象环境标签MetLabel。

[0017] 进一步的,二阶段聚类分析具体包括:

[0018] 令初始聚类数 $k = 2$,采用模糊C算法对N条数据集 $P = [P_{PV}, \text{MetLabel}]$ 做聚类分析,根据偏离指标最小选择最佳聚类数 k^* ;

[0019] 根据最佳聚类数结果,分别计算各个簇类的数据到聚类中心的距离 d_m ;

[0020] 根据计算所得的各个簇类数据的距离值,计算平均距离指标 D_{AVE} ;

[0021] 去掉各簇类中 d_m 最大的一条数据,更新数据集并重新计算平均距离指标 D_{AVE} ,按照每计算一次 D_{AVE} 的同时去掉一条 d_m 最大的数据的规则,生成平均距离指标 D_{AVE} 曲线,依据肘部法原则剔除不良数据。

[0022] 进一步的,依据肘部法原则确定不良数据具体为:

[0023] 确定平均距离指标 D_{AVE} 曲线中的拐点,将拐点对应的数据点记为不良数据点,做剔除处理。

[0024] 第二方面,本发明提供了一种低压分布式光伏系统不良数据检测系统,包括:

[0025] 数据获取单元,用于获取光伏阵列的历史出力数据以及对应的气象数据;

[0026] 第一聚类分析单元,用于根据气象数据构建气象数据向量集Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将N条气象数据划分为c个不同的簇类,并确定最佳聚类数 c^* ;还用于依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气象环境标签MetLabel;

[0027] 第二聚类分析单元,用于根据光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{PV}, \text{MetLabel}]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{PV} 为光伏有功出力;还用于分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m ;还用于循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

[0028] 进一步的,气象数据具体包括:气温Temp、光照强度Lux、空气湿度RH、云量TCC。

[0029] 进一步的,在第一聚类分析单元中,一阶段聚类分析具体包括:

[0030] 令初始聚类数 $c=2$,采用模糊C聚类算法对气象数据向量集 $\text{Met} = [\text{Temp}, \text{Lux}, \text{RH}, \text{TCC}]$ 做聚类分析;

[0031] 根据XB聚类有效性指标确定最佳聚类数 c^* ,构建气象环境指标 $1, 2, \dots, c^*$;

[0032] 依据各个采样时刻光伏出力对应的气象数据所属簇类,对光伏处理数据标记气象环境标签MetLabel。

[0033] 进一步的,在第二聚类分析单元中,二阶段聚类分析具体包括:

[0034] 令初始聚类数 $k = 2$,采用模糊C算法对N条数据集 $P = [P_{PV}, \text{MetLabel}]$ 做聚类分析,根据偏离指标最小选择最佳聚类数 k^* ;

[0035] 根据最佳聚类数结果,分别计算各个簇类的数据到聚类中心的距离 d_m ;

[0036] 根据计算所得的各个簇类数据的距离值,计算平均距离指标 D_{AVE} ;

[0037] 去掉各簇类中 d_m 最大的一条数据,更新数据集并重新计算平均距离指标 D_{AVE} ,按照每计算一次 D_{AVE} 的同时去掉一条 d_m 最大的数据的规则,生成平均距离指标 D_{AVE} 曲线,依据肘

部法原则筛除不良数据。

[0038] 进一步的,在第二聚类分析单元中,依据肘部法原则确定不良数据具体为:

[0039] 确定平均距离指标 D_{AVE} 曲线中的拐点,将拐点对应的数据点记为不良数据点,做剔除处理。

[0040] 综上,本发明提供了一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法和系统,通过采集光伏出力数据对应的气象数据,构建气象数据向量集,并通过一阶段模糊C聚类分析,将气象环境的划分结果构建气象环境标签,按照各时刻光伏出力的气象数据所属簇类,标记气象环境标签。在二阶段模糊C聚类过程中,将光伏出力数据与气象环境标签组合构建数据集合,依据偏离指标选择最佳聚类数,让二阶段聚类划分与一阶段的聚类划分尽可能贴合。根据最终光伏有功出力的划分簇类,依次剔除最外围数据,并绘制平均距离指标 D_{AVE} 的下降曲线,依据肘部法原则检测不良数据。本发明可以准确检测出传输到数据存储端的部分不良数据,降低其对光伏系统的特性分析以及出力预测的影响。

附图说明

[0041] 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性的前提下,还可以根据这些附图获得其它的附图。

[0042] 图1为本发明实施例提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法的流程图;

[0043] 图2为本发明实施例提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法的流程框图。

具体实施方式

[0044] 为使得本发明的目的、特征、优点能够更加的明显和易懂,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,下面所描述的实施例仅仅是本发明一部分实施例,而非全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其它实施例,都属于本发明保护的范围。

[0045] 近年来随着国家政策的大力支持,光伏技术的日益成熟,光伏组件的安装成本也呈指数级下降。因此,在我国各个地区出现了集中式或分布式光伏系统。然而,光伏电站的大规模建设也给此类可再生能源的准确监控带来技术难题。目前,用以监控光伏系统数据的设备主要包括:SCADA系统监控下光伏接入点可获取的有功功率以及无功功率;PMU系统监控下光伏接入点可检测的高频相量数据。随着光伏容量的快速增加,各式数据采集设备也相应投入使用,导致系统中的监控数据爆炸式增长。由于数据采集装置的量测误差,以及数据传输过程中的通信故障,通常容易导致传输到数据存储端的数据存在部分不良数据。这些不良数据的存在对后续光伏系统的特性分析以及出力预测都带来间接影响。

[0046] 基于此,本发明提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法和系统。

[0047] 以下对本发明的一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法的实施例进行详细的介绍。

[0048] 请参阅图1,本实施例提供了一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法,包括:

[0049] S100:获取光伏阵列的历史出力数据以及对应的气象数据。

[0050] 需要说明的是,光伏出力有功出力记为 P_{PV} ,气象数据具体包括气温Temp、光照强度Lux、空气湿度RH、云量TCC。

[0051] S200:根据气象数据构建气象数据向量集Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将N条气象数据划分为c个不同的簇类,并确定最佳聚类数 c^* 。

[0052] 可以理解的是,每一个簇类物理意义即为光伏电站工作的一个气象环境。

[0053] S300:依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气象环境标签MetLabel。

[0054] 需要说明的是,步骤S200-S300为一阶段聚类分析过程。具体的,本阶段的聚类分析过程如下:

[0055] 令初始聚类数 $c = 2$,采用模糊C算法对气象数据集 $Met = [Temp, Lux, RH, TCC]$ 做聚类分析。修正聚类数 c 的值,并计算XB聚类有效性指标,循环采用模糊C算法划分得出 c 个簇类,直到XB指标达到谷值时,确定最佳聚类数 c^* 个气象环境。其中,XB聚类有效性指标的计算方法表示如下:

$$[0056] \quad V_{XB} = \frac{\frac{1}{N} \sum_j^c \sum_i^N \mu_{ij}^2 \|x_i - v_j\|^2}{\min_{j \neq i} \|x_i - v_j\|^2}$$

[0057] c 表示聚类数, N 表示所有待聚类数据样本数目, μ_{ij} 为样本 x_i 隶属于 i 类隶属度值, v_j 是第 j 个簇类的聚类中心。XB聚类有效性指标越小,说明聚类结果中类内各点距离越小,类间各点之间距离越大,聚类划分越明显。

[0058] 根据一阶段的气象数据集聚类结果,构建气象环境指标 $MetLabel=1, 2, \dots, c^*$ 。并按照各个采样时刻光伏出力PPV对应气象数据所属簇类,对光伏出力数据PPV标记气象环境指标 $MetLabel$ 。气象环境标签 $MetLabel$,可以有效依据气象环境数据,划分光伏阵列在各个时刻所属的工作气象环境。

[0059] S400:根据光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{PV}, MetLabel]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{PV} 为光伏有功出力。

[0060] S500:分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m 。

[0061] S600:循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

[0062] 需要说明的是,步骤S400-S600为二阶段聚类分析过程。具体的,本阶段的聚类分

析过程如下：

[0063] 令初始聚类数 $k = 2$ ，采用模糊C算法对N条由光伏出力和对应气象环境指标构建的数据集 $P = [P_{pv}, \text{MetLabel}]$ 做聚类分析。修正聚类数 k 的值，根据偏离指标(Dev)最小选择最佳聚类数 k^* 。其中，偏离指标(Dev)计算方法如下：

$$[0064] \quad \text{Dev} = \frac{1}{k} \sum_{j=1}^k (n_j - n_{j,\text{mode}})$$

[0065] k 表示聚类数， n_j 表示第 j 个簇类中的数据数目， $n_{j,\text{mode}}$ 表示第 j 个簇类所有数据的MetLabel构成数据集的众数。因此，偏离指标(Dev)主要用以衡量两阶段聚类划分的差异性。循环采用模糊C聚类算法划分得出 k 个簇类，直到Dev指标达到谷值时，取偏离指标(Dev)最小时的 k 值为最佳二阶段聚类数 k^* 。

[0066] 得出最佳二阶段聚类结果后，分别计算各个簇类的数据到聚类中心的距离 d_m 。其中， d_m 计算方法如下 $y_{m,j}$ 表示第 j 个簇类的第 m 个样本， e_j 表示二阶段聚类中心：

$$[0067] \quad d_{m,j} = \|y_{m,j} - e_j\| \quad (m = 1, 2, \dots, n_j; j = 1, 2, \dots, k^*)$$

[0068] 根据计算所得的各个簇类数据的距离值，计算平均距离指标 D_{AVE} 如下：

$$[0069] \quad D_{\text{AVE}} = \frac{1}{k} \sum_{m=1}^{n_j} d_m$$

[0070] 去掉各簇类中 d_m 最大的一条数据，更新数据集并重新计算平均距离指标 D_{AVE} 。以此方法，依次再次去掉各簇类剩余数据中 d_m 最大的一个，并计算平均距离指标 D_{AVE} 。观察平均距离指标 D_{AVE} 曲线下降趋势，当 D_{AVE} 曲线一直保持平稳下降，说明无不良数据；当 D_{AVE} 曲线先急速下降，后转为平稳下降，依据肘部法原则，认为使得 D_{AVE} 曲线急速下降的数据点为不良数据，做剔除处理。

[0071] 另外，本实施例提供的不良数据检测方法的检测对象是低压分布式光伏阵列。

[0072] 图2是根据上述步骤确定的不良数据检测方法的流程框图。

[0073] 本实施例提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法，通过采集光伏出力数据对应的气象数据，构建气象数据向量集，并通过一阶段模糊C聚类分析，将气象环境的划分结果构建气象环境标签，按照各时刻光伏出力的气象数据所属簇类，标记气象环境标签。在二阶段模糊C聚类过程中，将光伏出力数据与气象环境标签组合构建数据集，依据偏离指标选择最佳聚类数，让二阶段聚类划分与一阶段的聚类划分尽可能贴合。根据最终光伏有功出力的划分簇类，依次剔除最外围数据，并绘制平均距离指标 D_{AVE} 的下降曲线，依据肘部法原则检测不良数据。本发明可以准确检测出传输到数据存储端的部分不良数据，降低其对光伏系统的特性分析以及出力预测的影响。

[0074] 以上是对本发明的一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测方法的实施例进行的详细介绍，以下将对本发明的一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测系统的实施例进行详细的介绍。

[0075] 本实施例提供一种基于二阶段模糊C算法的低压分布式光伏系统不良数据检测系

统,包括:数据获取单元、第一聚类分析单元和第二聚类分析单元。

[0076] 在本实施例中,数据获取单元用于获取光伏阵列的历史出力数据以及对应的气象数据。

[0077] 需要说明的是,气象数据具体包括:气温Temp、光照强度Lux、空气湿度RH、云量TCC。

[0078] 在本实施例中,第一聚类分析单元用于根据气象数据构建气象数据向量集Met 进行一阶段聚类分析,即采用模糊C聚类分析算法,将N条气象数据划分为c个不同的簇类,并确定最佳聚类数 c^* ;还用于依据各个光伏阵列在各个出力时刻的气象数据所属簇类,增加光伏系统的气象环境标签MetLabel。

[0079] 在第一聚类分析单元中,一阶段聚类分析具体包括:

[0080] 令初始聚类数 $c=2$,采用模糊C聚类算法对气象数据向量集 $Met = [Temp, Lux, RH, TCC]$ 做聚类分析;

[0081] 根据XB聚类有效性指标确定最佳聚类数 c^* ,构建气象环境指标 $1, 2, \dots, c^*$;

[0082] 依据各个采样时刻光伏出力对应的气象数据所属簇类,对光伏处理数据标记气象环境标签MetLabel。

[0083] 在本实施例中,第二聚类分析单元用于根据光伏系统的气象环境标签,继续采用模糊C聚类分析算法对数据集 $P=[P_{PV}, MetLabel]$ 数据做二阶段聚类分析,确定最佳聚类数 k^* ,其中 P_{PV} 为光伏有功出力;还用于分别计算按照最佳聚类数划分得到的 k^* 个簇类中各条数据到聚类中心的距离 d_m ;还用于循环计算平均距离指标 D_{AVE} ,并在每次计算后去掉各簇类中 d_m 最大的一条数据,根据每次计算得到的平均距离指标结果生成平均距离指标 D_{AVE} 曲线,依据肘部法原则确定不良数据。

[0084] 在第二聚类分析单元中,二阶段聚类分析具体包括:

[0085] 令初始聚类数 $k = 2$,采用模糊C算法对N条数据集 $P = [P_{PV}, MetLabel]$ 做聚类分析,根据偏离指标最小选择最佳聚类数 k^* ;

[0086] 根据最佳聚类数结果,分别计算各个簇类的数据到聚类中心的距离 d_m ;

[0087] 根据计算所得的各个簇类数据的距离值,计算平均距离指标 D_{AVE} ;

[0088] 去掉各簇类中 d_m 最大的一条数据,更新数据集并重新计算平均距离指标 D_{AVE} ,按照每计算一次 D_{AVE} 的同时去掉一条 d_m 最大的数据的规则,生成平均距离指标 D_{AVE} 曲线,依据肘部法原则剔除不良数据。

[0089] 另外,依据肘部法原则确定不良数据具体为:

[0090] 确定平均距离指标 D_{AVE} 曲线中的拐点,将拐点对应的数据点记为不良数据点,做剔除处理。

[0091] 需要说明的是,本实施例提供的不良数据检测系统用于实现前述实施例提供的不良数据检测方法,各单元的具体设置均以完整实现该方法为准,在此不再赘述。

[0092] 以上实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

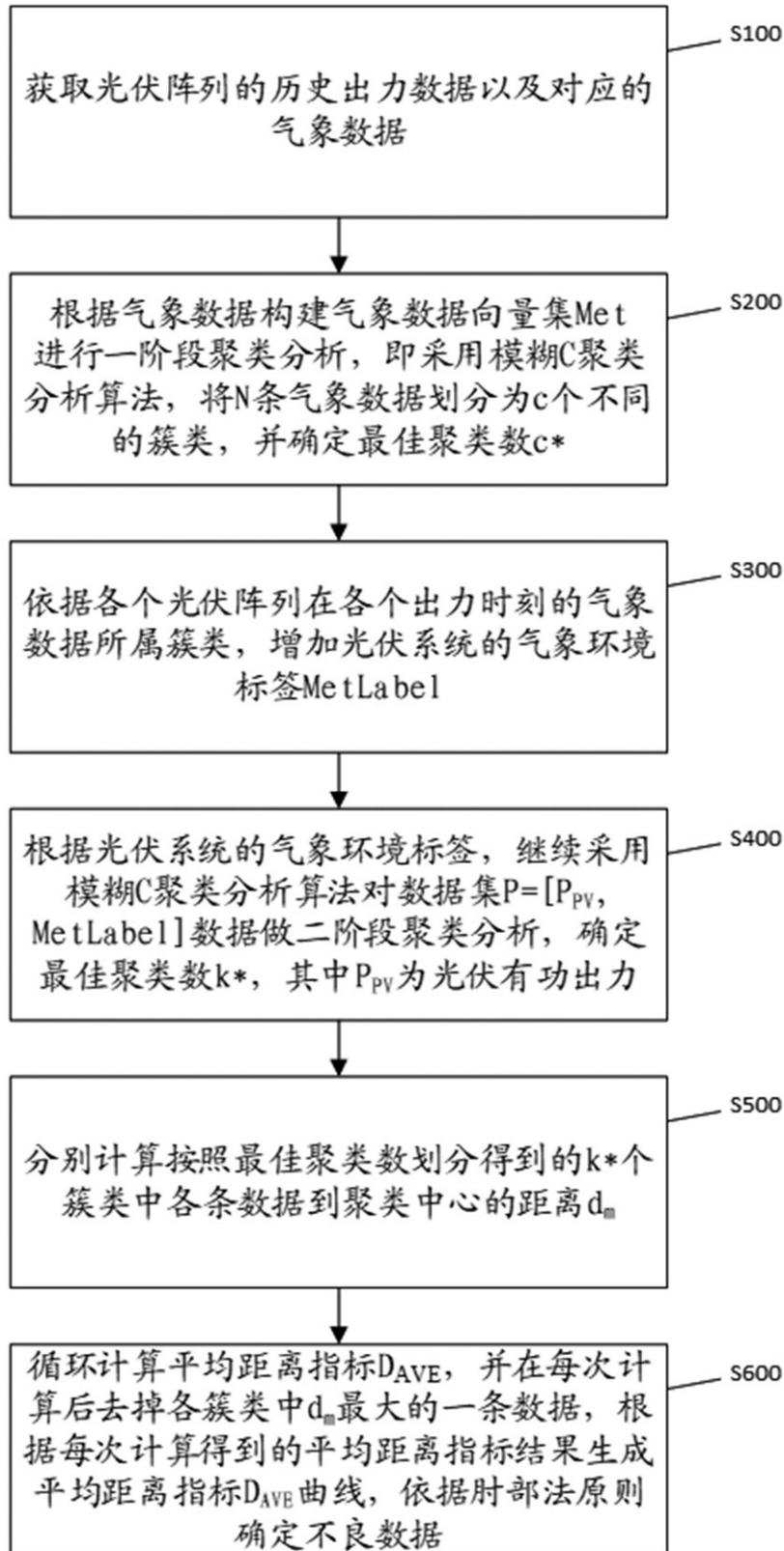


图1

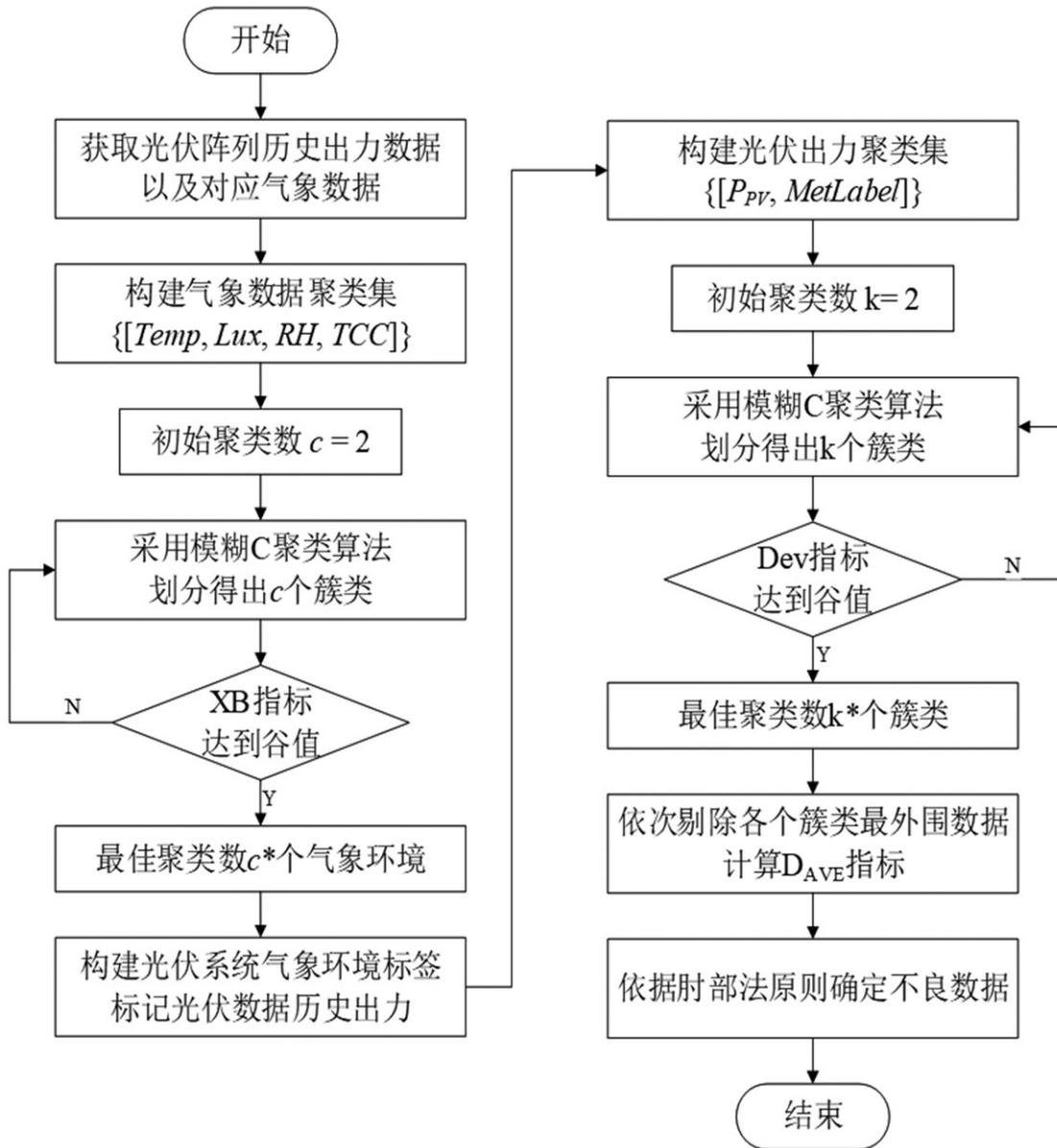


图2