



(12) 发明专利

(10) 授权公告号 CN 113642618 B

(45) 授权公告日 2024.03.01

(21) 申请号 202110851829.2

G06F 18/241 (2023.01)

(22) 申请日 2021.07.27

G06F 18/2433 (2023.01)

(65) 同一申请的已公布的文献号

G06Q 10/04 (2023.01)

申请公布号 CN 113642618 A

G06Q 10/20 (2023.01)

G06Q 50/04 (2012.01)

(43) 申请公布日 2021.11.12

(56) 对比文件

(73) 专利权人 上海展湾信息科技有限公司

CN 103064287 A, 2013.04.24

地址 200090 上海市杨浦区黄兴路217号A

CN 110222782 A, 2019.09.10

座14层A1401、A1403号

CN 111275546 A, 2020.06.12

(72) 发明人 叶思汐 唐栋

US 2018281256 A1, 2018.10.04

(74) 专利代理机构 上海百一领御专利代理事务

CN 112861422 A, 2021.05.28

所(普通合伙) 31243

CN 113139053 A, 2021.07.20

专利代理师 王奎宇 朱永海

CN 111618655 A, 2020.09.04

CN 110232405 A, 2019.09.13

CN 112101220 A, 2020.12.18

(51) Int. Cl.

G06F 18/214 (2023.01)

G06F 18/21 (2023.01)

审查员 王佩思

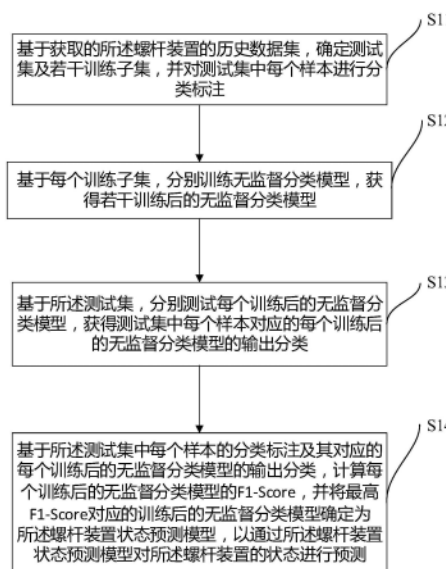
权利要求书2页 说明书12页 附图2页

(54) 发明名称

一种用于螺杆装置状态预测模型训练的方法及设备

(57) 摘要

本申请提供一种用于螺杆装置状态预测模型训练的方法及设备,该方法基于获取的螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;接着基于每个训练子集,分别训练无监督分类模型;然后基于测试集,分别测试每个训练后的无监督分类模型,获得相应的输出分类;最后基于所述测试集中每个样本的分类标注及其对应的每个训练后的模型的输出分类,计算每个训练后的模型的F1-Score,将最高F1-Score对应的模型确定为螺杆装置状态预测模型。通过该方法得到的无监督分类预测模型,可以用来对螺杆装置的健康状况进行预测评估,实现对螺杆装置的故障预测和健康管理,减少停机风险,较好地延长螺杆装置的剩余使用寿命,并可降低维护成本。



1. 一种用于螺杆装置状态预测模型训练的方法,其特征在于,所述方法包括:

获取所述螺杆装置的原始历史数据集,并对所述原始历史数据集做数据清洗,以得到所述历史数据集,基于所述历史数据集,确定其中每日历史数据对应的特征数据,并基于所有特征数据,构建样本集,将所述样本集划分为测试集及训练集,并对测试集中每个样本进行标注,基于所述训练集,确定若干训练子集,并将每个训练子集中的样本进行标准化处理,以得到若干标准化处理后的训练子集,基于所述测试集及每个训练子集,将测试集中的样本进行标准化处理,以得到每个标准化处理后的训练子集对应的标准化处理后的测试子集,其中,所述特征数据包括:所述历史数据集中每日历史数据对应的统计特征数据,以及以下至少一项:所述历史数据集中每日历史数据对应的滑动窗口特征数据,所述历史数据集中每日历史数据对应的时间序列特征数据,所述历史数据集中每日历史数据对应的频域特征数据,其中,每类特征数据还包含多个特征参数;

基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;

基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;

基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型,以通过所述螺杆装置状态预测模型对所述螺杆装置的状态进行预测。

2. 根据权利要求1所述的方法,其特征在于,所述方法还包括:

基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数据,并将所述当日特征数据进行标准化处理;

将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。

3. 根据权利要求2所述的方法,其特征在于,所述方法还包括:

若所述螺杆装置当日状态为异常,将所述当日数据与阈值进行比较;

若比较结果满足预设条件,则确认所述螺杆装置需要检修,否则,判断所述螺杆装置当日之前预设连续天数的状态是否异常,以确认所述螺杆装置是否需要检修。

4. 根据权利要求3所述的方法,其特征在于,所述阈值的确定包括:

确定用于训练所述螺杆装置状态预测模型的训练子集中,所有输出分类为正常所对应的标准化处理后的特征数据;

将所有所述标准化处理后的特征数据对应的历史数据的均值的预设倍数确定为阈值。

5. 一种用于螺杆装置状态预测模型训练的设备,其特征在于,所述设备包括:

第一装置,用于获取所述螺杆装置的原始历史数据集,并对所述原始历史数据集做数据清洗,以得到所述历史数据集,基于所述历史数据集,确定其中每日历史数据对应的特征数据,并基于所有特征数据,构建样本集,将所述样本集划分为测试集及训练集,并对测试集中每个样本进行标注,基于所述训练集,确定若干训练子集,并将每个训练子集中的样本进行标准化处理,以得到若干标准化处理后的训练子集,基于所述测试集及每个训练子集,将测试集中的样本进行标准化处理,以得到每个标准化处理后的训练子集对应的标准化处理后的测试子集,其中,所述特征数据包括:所述历史数据集中每日历史数据对应的统计特

征数据,以及以下至少一项:所述历史数据集中每日历史数据对应的滑动窗口特征数据,所述历史数据集中每日历史数据对应的时间序列特征数据,所述历史数据集中每日历史数据对应的频域特征数据,其中,每类特征数据还包含多个特征参数;

第二装置,用于基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;

第三装置,用于基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;

第四装置,用于基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型。

6. 根据权利要求5所述的设备,其特征在于,所述设备还包括:

第五装置,用于基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数据,并将所述当日特征数据进行标准化处理;

第六装置,用于将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。

7. 一种计算机可读介质,其特征在于,

其上存储有可读的计算机程序指令,所述可读的计算机程序指令被处理器执行以实现如权利要求1至4中任一项所述的方法。

8. 一种用于螺杆装置状态预测模型训练的设备,其特征在于,所述设备包括:

一个或多个处理器;以及

存储有计算机可读指令的存储器,所述计算机可读指令在被执行时使所述处理器执行如权利要求1至4中任一项所述方法的操作。

一种用于螺杆装置状态预测模型训练的方法及设备

技术领域

[0001] 本申请涉及机械设备维护领域,尤其涉及一种用于螺杆装置状态预测的技术。

背景技术

[0002] 在工业生产中,具有螺杆装置的设备在使用寿命周期内,除了正常的疲劳失效,在螺杆装置工作时经常出现的粉尘堆积、润滑不及时、应力变大等因素都会对螺杆装置的健康状况产生不利影响,因此,对具有螺杆装置的设备的日常清洗、润滑等日常性维护是必不可少的。

[0003] 但是,日常性维护属于被动维护,并不能很好地延长具有螺杆装置的设备的剩余使用寿命,可考虑采取预测性维护来更好地延长具有螺杆装置的设备的剩余使用寿命。而直接针对设备的预测性维护存在数据难收集且收集周期长、数据与剩余使用寿命关联度不高甚至无关联等困难。

[0004] 传动轴作为具有螺杆装置的设备实现作业的纽带,其本身的工作状态直接影响螺杆装置的工作状态,从而影响到设备的剩余使用寿命。而传动轴的扭矩数据可直接反映传动轴的工作状态,因此,可考虑通过监测螺杆装置的扭矩数据的变化来对螺杆装置的健康状态进行预测,从而实现对具有螺杆装置的设备进行预测性维护。

发明内容

[0005] 本申请的目的是提供一种用于螺杆装置状态预测模型的方法及设备,用以解决具有螺杆装置的设备的预测性维护难的技术问题。

[0006] 根据本申请的一个方面,提供了一种用于螺杆装置状态预测模型训练的方法,其中,所述方法包括:

[0007] 基于获取的所述螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;

[0008] 基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;

[0009] 基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;

[0010] 基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型,以通过所述螺杆装置状态预测模型对所述螺杆装置的状态进行预测。

[0011] 可选地,其中,所述基于获取的所述螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注包括:

[0012] 获取所述螺杆装置的原始历史数据集,并对所述原始历史数据集做数据清洗,以得到所述历史数据集;

- [0013] 基于所述历史数据集,确定其中每日历史数据对应的特征数据,并基于所有特征数据,构建样本集;
- [0014] 将所述样本集划分为测试集及训练集,并对测试集中每个样本进行标注;
- [0015] 基于所述训练集,确定若干训练子集,并将每个训练子集中的样本进行标准化处理,以得到若干标准化处理后的训练子集;
- [0016] 基于所述测试集及每个训练子集,将测试集中的样本进行标准化处理,以得到每个标准化处理后的训练子集对应的标准化处理后的测试子集。
- [0017] 可选地,其中,所述特征数据包括:
- [0018] 所述历史数据集中每日历史数据对应的统计特征数据,以及以下至少一项:
- [0019] 所述历史数据集中每日历史数据对应的滑动窗口特征数据;
- [0020] 所述历史数据集中每日历史数据对应的时间序列特征数据;
- [0021] 所述历史数据集中每日历史数据对应的频域特征数据。
- [0022] 可选地,所述一种用于螺杆装置状态预测模型训练的方法还包括:
- [0023] 基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数据,并将所述当日特征数据进行标准化处理;
- [0024] 将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。
- [0025] 可选地,所述方法还包括:
- [0026] 若所述螺杆装置当日状态为异常,将所述当日数据与阈值进行比较;
- [0027] 若比较结果满足预设条件,则确认所述螺杆装置需要检修,否则,判断所述螺杆装置当日之前预设连续天数的状态是否异常,以确认所述螺杆装置是否需要检修。
- [0028] 可选地,其中,所述阈值的确定包括:
- [0029] 确定用于训练所述螺杆装置状态预测模型的训练子集中,所有输出分类为正常所对应的标准化处理后的特征数据;
- [0030] 将所有所述标准化处理后的特征数据对应的历史数据的均值的预设倍数确定为阈值。
- [0031] 根据本申请的另一方面,还提供了一种用于螺杆装置状态预测模型训练的设备,其中,所述设备包括:
- [0032] 第一装置,用于基于获取的所述螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;
- [0033] 第二装置,用于基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;
- [0034] 第三装置,用于基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;
- [0035] 第四装置,用于基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型。
- [0036] 可选地,所述设备还包括:
- [0037] 第五装置,用于基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数

据,并将所述当日特征数据进行标准化处理;

[0038] 第六装置,用于将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。

[0039] 与现有技术相比,本申请提供一种用于螺杆装置状态预测模型训练的方法及设备,首先基于获取的螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;接着基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;然后基于测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;最后基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型,以通过所述螺杆装置状态预测模型对所述螺杆装置的状态进行预测。进一步地,可根据预测结果对螺杆装置状态进行预测性维护。通过该方法,采用无标注样本训练集对无监督分类模型进行训练,得到预测模型,可以用来对螺杆装置的健康状况进行预测性维护,实现对螺杆装置的故障预测和健康管理,减少停机风险,较好地延长螺杆装置的剩余使用寿命,并可降低维护成本。

附图说明

[0040] 通过阅读参照以下附图所作的对非限制性实施例所作的详细描述,本发明的其它特征、目的和优点将会变得更明显:

[0041] 图1示出根据本申请一个方面的一种用于螺杆装置状态预测模型训练的方法流程图;

[0042] 图2示出一个实施例的一个螺杆装置的特征数据分布可视化示意图;

[0043] 图3示出根据本申请另一个方面的一种用于螺杆装置状态预测模型训练的设备示意图;

[0044] 附图中相同或相似的附图标记代表相同或相似的部件。

具体实施方式

[0045] 下面结合附图对本发明作进一步详细描述。

[0046] 在本申请一个典型的配置中,所述设备、装置或模块均包括一个或多个处理器(CPU)、输入/输出接口、网络接口和内存。

[0047] 内存可能包括计算机可读介质中的非永久性存储器,随机存取存储器(RAM)和/或非易失性内存等形式,如只读存储器(ROM)或闪存(flash RAM)。内存是计算机可读介质的示例。

[0048] 计算机可读介质包括永久性和非永久性、可移动和非可移动媒体可以由任何方法或技术来实现信息存储。信息可以是计算机可读指令、数据结构、程序的模块或其他数据。计算机的存储介质的例子包括,但不限于相变内存(PRAM)、静态随机存取存储器(SRAM)、动态随机存取存储器(DRAM)、其他类型的随机存取存储器(RAM)、只读存储器(ROM)、电可擦除可编程只读存储器(EEPROM)、快闪记忆体或其他内存技术、只读光盘只读存储器(CD-ROM)、数字多功能光盘(DVD)或其他光学存储、磁盒式磁带,磁带磁盘存储或其他磁性存储设备或

任何其他非传输介质,可用于存储可以被计算设备访问的信息。按照本文中的界定,计算机可读介质不包括非暂存电脑可读媒体(transitory media),如调制的数据信号和载波。

[0049] 为更进一步阐述本申请所采取的技术手段及取得的效果,下面结合附图及优选实施例,对本申请的技术方案,进行清楚和完整的描述。

[0050] 图1示出本申请一个方面的一种用于螺杆装置状态预测模型训练的方法流程图,其中,一个实施例的方法包括:

[0051] S11基于获取的所述螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;

[0052] S12基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;

[0053] S13基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;

[0054] S14基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型,以通过所述螺杆装置状态预测模型对所述螺杆装置的状态进行预测。

[0055] 在本申请中,所述方法通过设备1执行,所述设备1为可获取螺杆装置传动轴扭矩数据的计算机设备和/或云。其中,所述计算机设备包括但不限于个人计算机、笔记本电脑、工业计算机、网络主机、单个网络服务器、多个网络服务器集;所述云由基于云计算(Cloud Computing)的大量计算机或网络服务器构成,其中,云计算是分布式计算的一种,由一群松散耦合的计算机集组成的一个虚拟超级计算机。

[0056] 在此,所述计算机设备和/或云仅为举例,其他现有的或者今后可能出现的设备和/或资源共享平台如适用于本申请也应包含在本申请的保护范围内,在此,以引用的方式包含于此。

[0057] 在该实施例中,在所述步骤S11中,设备1获取到螺杆装置的历史数据集,比如与螺杆装置每日生产中的有效扭矩数据对应的历史数据组成的历史数据集,划分测试集及若干训练子集,并对测试集中的每个样本进行分类标注。

[0058] 可选地,其中,所述步骤S11包括:

[0059] S111获取所述螺杆装置的原始历史数据集,并对所述原始历史数据集做数据清洗,以得到所述历史数据集;

[0060] S112基于所述历史数据集,确定其中每日历史数据对应的特征数据,并基于所有特征数据,构建样本集;

[0061] S113将所述样本集划分为测试集及训练集,并对测试集中每个样本进行标注;

[0062] S114基于所述训练集,确定若干训练子集,并将每个训练子集中的样本进行标准化处理,以得到若干标准化处理后的训练子集;

[0063] S115基于所述测试集及每个训练子集,将测试集中的样本进行标准化处理,以得到每个标准化处理后的训练子集对应的标准化处理后的测试子集。

[0064] 在该实施例的步骤S111中,设备1获取螺杆装置的原始历史数据集,比如设备1与PLC传感器电连接,直接获取PLC传感器采集到的螺杆装置的传动轴的扭矩数据并保存,以

得到原始的历史扭矩数据集。或者是上位机与PLC传感器电连接,上位机获取PLC传感器采集到的螺杆装置的传动轴的扭矩数据,并保存,以得到原始的历史扭矩数据集,而设备1获取到上位机上存储的原始的历史扭矩数据集。其中,原始的历史扭矩数据集中的每条历史扭矩数据通常包括数据采集的日期(DAY)、时间(time)及一个无量纲、单位的百分比数值(value),即螺杆装置日常生产中传动轴的实际扭矩与其额定扭矩的比值。一个实施例的一个螺杆装置的原始的历史扭矩数据集示例如下表1所示。

[0065] 表1

[0066] 由于各种客观原因,原始的历史扭矩数据集中的数据存在异常数据、重

	time	value
0	2020-07-01 08:00:00.031	0.000000
1	2020-07-01 08:00:01.060	0.000000
2	2020-07-01 08:00:02.090	0.000000
3	2020-07-01 08:00:03.120	0.000000
4	2020-07-01 08:00:04.149	0.000000

[0067]

...
5759656	2020-12-31 16:33:42.401	46.799999
5759657	2020-12-31 16:33:43.431	46.799999
5759658	2020-12-31 16:33:44.460	46.799999
5759659	2020-12-31 16:33:45.490	46.799999
5759660	2020-12-31 16:33:46.519	46.799999

5759661 rows × 2 columns

[0068] 复数据和无效数据等,比如,螺杆装置的加工过程是非连续过程,在此过程中PLC采集到的历史扭矩数据属于异常数据;当日螺杆装置开机后的预热阶段(例如2小时,此时间段内采集到的扭矩数据波动较大)以内的PLC采集到的历史扭矩数据也被认为是异常数据;当日螺杆装置总计加工时长小于若干小时(例如1小时)所对应的PLC采集到的历史扭矩数据也被认为是无效数据;还有螺杆装置传动轴一个加工周期(比如3秒)内的PLC采集到的多个扭矩数据(若PLC的数据采集周期约为1秒,一个加工周期内采集到3个扭矩数据)可被认为是重复数据。

[0069] 因此,设备1在获取到原始的历史扭矩数据集后,通常需要对原始的历史扭矩数据集做数据清洗,得到数据清洗后的历史扭矩数据集。比如,以时间为扭矩数据划分单位,剔除当前前2个小时的扭矩数据;若当日总计加工时长小于1小时,则剔除该日的扭矩数据;剔除非连续加工过程的扭矩数据;剔除一个加工周期内的重复扭矩数据;还可通过箱线图,将小于箱线图下边缘的扭矩数据(比如数据值为0和/或异常数据值)剔除。一个实施例的一个螺杆装置的数据清洗后的历史扭矩数据集示例如下表2所示。

[0070] 表2

	time	DAY	value
0	2020-07-01 08:16:17.685	2020-07-01	46.099998
1	2020-07-01 08:16:18.714	2020-07-01	42.400002
2	2020-07-01 08:16:22.833	2020-07-01	45.500000
3	2020-07-01 08:16:23.862	2020-07-01	42.500000
4	2020-07-01 08:16:28.464	2020-07-01	44.700001

[0071]

...
1087805	2020-12-31 12:01:44.336	2020-12-31	46.400002
1087806	2020-12-31 12:01:46.910	2020-12-31	45.500000
1087807	2020-12-31 12:01:49.484	2020-12-31	46.400002
1087808	2020-12-31 12:01:52.572	2020-12-31	47.299999
1087809	2020-12-31 12:01:55.661	2020-12-31	45.500000

1087810 rows × 3 columns

[0072] 在无法直接通过对历史数据标注来区分螺杆装置的健康加工状态或异常加工状态的情况下,可通过提取历史数据的特征来预测螺杆装置的健康状况。

[0073] 继续在该实施例中,在步骤S112中,设备1将数据清洗后的历史扭矩数据集按时间日为单位,确定其中每日的历史扭矩数据对应的特征数据,并将每日的特征数据组织起来,构建样本集。

[0074] 可选地,其中,所述特征数据包括:

[0075] 所述历史数据集中每日历史数据对应的统计特征数据,以及以下至少一项:

[0076] 所述历史数据集中每日历史数据对应的滑动窗口特征数据;

[0077] 所述历史数据集中每日历史数据对应的时间序列特征数据;

[0078] 所述历史数据集中每日历史数据对应的频域特征数据。

[0079] 例如,设备1经过步骤S111后得到历史扭矩数据集,以时间日为单位,可获取每日的历史扭矩数据的统计特征数据,其中,该统计特征数据具体可以包括如下特征参数:每日的历史扭矩数据的均值(mean),每日的历史扭矩数据的均方差(std),每日的历史扭矩数据的最大值(max)及最小值(min),每日的历史扭矩数据的分布特性等。

[0080] 为了避免影响训练效果,提升将特征数据作为样本来预测螺杆装置的健康状况的准确率,及适应不同的应用场景,还可以获取每日的历史扭矩数据的滑动窗口特征数据和/或时间序列特征数据、频域特征数据,选择不同特征数据进行拟合,可将每日的历史扭矩数据的滑动窗口特征数据和/或时间序列特征数据、频域特征数据也纳入样本,其中,每类特征数据还可包含多个特征参数。比如,针对每日的历史扭矩数据,设置窗口大小为100条历史扭矩数据,则可通过滑动窗口来获取该日的历史扭矩数据中,每100条历史扭矩数据的均值、均方差,最大值、最小值、变化率等滑动窗口特征参数;获取每日的历史扭矩数据的自回归系数、近似熵、趋势特征、窗口差异值特征等时间序列特征参数;获取每日的历史扭矩数据的傅里叶系数、幅度、峰度、偏度等频域特征参数。上述实施例的一个螺杆装置的原始的历史扭矩数据集对应的特征数据集示例如下表3所示。

[0081] 表3

	mean	std	max	min	mean_change_abs	hmean	gmean	kstat_1	moment_1	kstat_2	...
0	45.903295	0.922042	49.000000	43.200001	-0.000366	45.884853	45.894061	45.903295	0.0	0.850161	...
1	43.641790	1.311897	50.799999	40.599998	-0.000975	43.602593	43.622152	43.641790	0.0	1.721073	...
2	44.213768	1.204700	47.799999	40.799999	-0.000289	44.181060	44.197395	44.213768	0.0	1.451303	...
3	45.541192	0.892818	48.500000	42.099998	-0.000267	45.523636	45.532424	45.541192	0.0	0.797125	...
4	45.055933	0.923263	49.200001	42.700001	0.000315	45.037164	45.046524	45.055933	0.0	0.852414	...
5	47.764613	1.135170	51.500000	44.500000	-0.000654	47.737659	47.751133	47.764613	0.0	1.288612	...
6	46.966654	1.092021	51.200001	43.400002	-0.000431	46.941372	46.953996	46.966654	0.0	1.192509	...
7	48.515564	0.958773	51.599998	45.299999	-0.000452	48.496604	48.506087	48.515564	0.0	0.919245	...
8	41.684471	0.602092	43.900002	39.599998	-0.000070	41.675769	41.680121	41.684471	0.0	0.362515	...
9	42.175262	0.614344	44.400002	40.200001	-0.000094	42.166292	42.170781	42.175262	0.0	0.377419	...
10	42.571203	0.621142	44.599998	40.099998	0.000079	42.562129	42.566669	42.571203	0.0	0.385818	...
11	44.993233	0.911244	48.000000	42.099998	-0.000120	44.974763	44.984001	44.993233	0.0	0.830365	...
12	42.901722	1.533143	48.400002	40.400002	-0.000998	42.848638	42.874899	42.901722	0.0	2.350527	...
13	41.946646	0.504722	43.900002	40.400002	-0.000342	41.940591	41.943616	41.946646	0.0	0.254745	...
14	42.898635	1.341370	51.000000	40.200001	-0.000385	42.845967	42.866094	42.898635	0.0	1.799273	...
15	44.909106	0.840856	48.000000	42.299999	-0.000148	44.893409	44.901250	44.909106	0.0	0.707039	...
16	47.078330	1.054765	50.099998	44.299999	-0.000792	47.054730	47.066526	47.078330	0.0	1.112530	...

[0083] 对于同一螺杆装置,因各种原因在生产中可能存在很多种异常状态,但其健康状态应该只有一种,在健康状态下的历史数据对应的特征数据应该是相同或相近的,因此,可将不同日期的历史数据对应的特征数据提取后做可视化。例如,以日为单位,将每日的一组特征数据作为整体采用PCA(Principal Component Analysis,主成分分析)法降维,直至降至2维后,进行可视化展示,加入某日的数据清洗后额历史扭矩数据经特征提取后得到一组100维的特征数据,对该组特征数据采用PCA法进行降维处理,直至降至2维,得到一组2维数据,对应2维坐标平面的X轴及Y轴坐标,在2维坐标平面上可视化展示为一个数据点。上述实施例的一个螺杆装置的数据清洗后的历史扭矩数据集对应的特征数据集可视化示例如图2所示,健康状态下的历史扭矩数据对应的特征数据分布高度聚集,而异常状态下的历史扭矩数据对应的特征数据分布较分散。

[0084] 继续在该实施例中,在步骤S113中,设备1可将得到的样本集划分为测试集及训练集,比如,随机将样本集中样本(即每日和/或一段时间历史数据对应的特征数据)的30%划分为测试集,将样本集中剩余的70%样本划分为训练集,在此,划分比例仅为举例,可结合具体应用场景采取不同划分比例,在此不做限定。

[0085] 设备1还结合对测试集中样本的可视化结果,对测试集中每个样本进行标注,其中,将在可视化结果中处于聚集区对应的样本标注为正常,将在可视化结果中处于非聚集区对应的样本标注为异常。

[0086] 由于样本集中不同特征参数数值的纲量差异较大,造成不同特征参数对模型的影响权重不同,可将样本进行标准化处理,以消除不同纲量对模型的影响,使得标准化处理后样本中各特征参数对模型的影响权重相同。

[0087] 为了提升训练集的多样性和模型的泛化能力,继续在该实施例中,在步骤S114中,设备1继续将训练集划分成若干训练子集,进一步地,鉴于训练样本中不同特征参数数值的纲量差异较大,可将训练样本进行标准化处理,以消除不同纲量对模型的影响,使得标准化处理后样本中各特征参数对模型的影响权重相同。比如,可将训练集划分为若干包含不同

样本数量和/或不同样本的训练子集,然后计算每个训练子集中对应各特征参数的所有样本的相应参数值的均值(mean)及均方差(std),然后将各特征参数下各样本的相应参数值x经过如下标准化公式处理,

[0088] $(x-\text{mean})/\text{std}$

[0089] 遍历训练子集中全部特征参数下全部样本,得到标准化处理后的训练子集。

[0090] 其中,每个训练子集中对应各特征参数的mean及std只与该训练子集中的样本有关,对应各特征参数的mean及std是该训练子集中全部样本对应特征参数下所有样本值的mean及std。比如,若训练子集包含50天的统计特征数据(包括每天历史数据均值、最大值、最小值等统计特征参数),则每个统计特征参数的mean/std是该统计特征参数对应的50个特征参数值的mean/std;若训练子集包含100天统计特征数据,则每个统计特征参数的mean/std是该统计特征参数对应的100个特征参数值的mean/std。

[0091] 对如表3所示的上述实施例的一个螺杆装置的原始的历史扭矩数据集对应的特征数据集进行如上标准化处理后得到的训练子集如下表4所示。

[0092] 表4

	mean	std	max	min	mean_change_abs	hmean	gmean	kstat_1	moment_1	kstat_2	...
0	-0.193926	-0.462947	-0.405877	2.330195e-01	0.202417	-0.181998	-0.187998	-0.193926	0.0	-0.344106	...
1	-0.625599	-0.265274	-0.274185	-6.324811e-01	-0.488082	-0.636993	-0.631338	-0.625599	0.0	-0.297331	...
2	-0.516421	-0.319627	-0.493672	-5.659039e-01	0.289707	-0.521669	-0.519085	-0.516421	0.0	-0.311820	...
3	-0.263044	-0.477765	-0.442458	-1.331543e-01	0.314452	-0.254011	-0.258568	-0.263044	0.0	-0.346954	...
4	-0.355669	-0.462328	-0.391244	6.657719e-02	0.975373	-0.350995	-0.353386	-0.355669	0.0	-0.343965	...
5	0.161360	-0.354882	-0.222971	6.657691e-01	-0.124179	0.187379	0.174390	0.161360	0.0	-0.320558	...
6	0.009047	-0.376760	-0.244920	2.995966e-01	0.128404	0.028631	0.018837	0.009047	0.0	-0.325719	...
7	0.304701	-0.444323	-0.215655	9.320765e-01	0.105622	0.338684	0.321711	0.304701	0.0	-0.340396	...
8	-0.999210	-0.625176	-0.779005	-9.653656e-01	0.537997	-1.021127	-1.010305	-0.999210	0.0	-0.370296	...
9	-0.905529	-0.618964	-0.742424	-7.656341e-01	0.511736	-0.923335	-0.914558	-0.905529	0.0	-0.369496	...
10	-0.829952	-0.615517	-0.727791	-7.989233e-01	0.707207	-0.844421	-0.837304	-0.829952	0.0	-0.369045	...
11	-0.367638	-0.468422	-0.479039	-1.331543e-01	0.481398	-0.363435	-0.365587	-0.367638	0.0	-0.345169	...
12	-0.766863	-0.153092	-0.449774	-6.990570e-01	-0.514024	-0.787302	-0.777156	-0.766863	0.0	-0.263524	...
13	-0.949166	-0.674547	-0.779005	-6.990570e-01	0.230162	-0.968331	-0.958887	-0.949166	0.0	-0.376805	...
14	-0.769743	-0.250329	-0.259552	-7.656341e-01	0.181438	-0.787835	-0.778875	-0.769743	0.0	-0.293131	...
15	-0.383696	-0.504112	-0.479039	-6.657713e-02	0.449977	-0.379654	-0.381735	-0.383696	0.0	-0.351793	...
16	0.030363	-0.395651	-0.325398	5.991919e-01	-0.280878	0.051230	0.040796	0.030363	0.0	-0.330015	...

[0094] 为了确保测试集中的样本与各标准化处理后的训练子集中的样本具有相同的分布特性,继续在该实施例中,在步骤S115中,设备1对测试集中样本进行标准化处理,先获取每个训练子集中对应各特征参数的所有样本的相应参数值的均值(mean)及均方差(std),然后将测试集中各特征参数下各样本的相应参数值x经过如下标准化公式处理,

[0095] $(x-\text{mean})/\text{std}$

[0096] 得到若干个与标准化处理后的训练子集对应的标准化处理后的测试集。

[0097] 继续在该实施例中,在所述步骤S12中,设备1分别用每个训练子集来训练无监督分类模型,获得与若干个训练子集对应的训练后的无监督分类模型。其中,所述无监督分类模型可以采用k-means无监督分类算法模型,在本申请中是将训练后的模型用于对螺杆装置的状态(正常或者异常)进行预测,k取值为2。

[0098] 继续在该实施例中,在所述步骤S13中,设备1针对获得的每个训练后的无监督

分类模型,采用测试集进行测试,获得针对每个训练后的无监督分类模型的测试集中每个样本对应的输出分类,其中,所述输出分类为正常或者异常,比如,输出分类为0代表测试集中该样本对应的该日历史数据正常,代表螺杆装置该日状态正常,若输出分类为1,代表螺杆装置该日状态异常。

[0099] 继续在所述实施例中,在所述步骤S14中,设备1针对每个训练后的无监督分类模型,都可获得测试集中每个样本的输出分类,结合测试集中每个样本的分类标注,采用混淆矩阵来计算该训练后的无监督分类模型的F1-Score。遍历全部训练后的无监督分类模型,可确定每个训练后的无监督分类模型的F1-Score,然后将最高F1-Score对应的训练后的无监督分类模型确定为螺杆装置状态预测模型,还获取该最高F1-Score对应的训练子集中的正常样本对应的历史数据的均值。

[0100] 可通过该螺杆装置状态预测模型对该螺杆装置的状态进行预测。

[0101] 可选地,所述一种用于螺杆装置状态预测模型训练的方法还包括:

[0102] S15(未示出)基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数据,并将所述当日特征数据进行标准化处理;

[0103] S16(未示出)将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。

[0104] 其中,设备1在多个训练后的无监督分类模型中确定螺杆装置状态预测模型后,还可将该螺杆装置状态预测模型用于预测螺杆设备当日状态。

[0105] 继续上述实施例中,在所述步骤S15中,设备1获取该螺杆装置的当日原始扭矩数据,对原始扭矩数据进行清洗,剔除异常数据集重复数据后,获得当日有效数据,并确定与其对应的当日特征数据,其中,当日特征数据包含的特征参数应与前述训练集及测试集中样本中包含的特征参数一致。将当日特征数据进行标准化处理,其中,采用螺杆装置状态预测模型对应的训练子集中对应各特征参数的所有样本的相应参数值的均值(mean)及均方差(std),然后将当日特征数据中相应参数值x经过如下标准化公式处理,

[0106] $(x-\text{mean})/\text{std}$

[0107] 得到标准化处理后的当日特征数据。

[0108] 继续上述实施例中,在所述步骤S16中,设备1将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,获得输出分类,根据输出分类确定所述螺杆装置当日状态是否异常。比如,若输出分类为0,则可确定该螺杆装置当日状态正常;若输出分类为1,则可确定该螺杆装置当日状态异常。

[0109] 由于螺杆装置日常生产环境的复杂性,状态异常可能是螺杆装置本身因素造成的,也可能是其它客观因素造成的,而过于频繁的停机检修会造成经济损失,不可取。因此,对于螺杆装置当日状态异常的情况,还有必要对当日数据进行进一步分析处理。

[0110] 可选地,其中,所述方法还包括:

[0111] S17(未示出)若所述螺杆装置当日状态为异常,将所述当日数据与阈值进行比较;

[0112] S18(未示出)若比较结果满足预设条件,则确认所述螺杆装置需要检修,否则,则判断所述螺杆装置当日之前预设连续天数的状态是否异常,以确认所述螺杆装置是否需要检修。

[0113] 继续上述实施例中,在所述步骤S17中,若设备1获得的当日标准化处理后的特征

数据输入所述螺杆装置状态预测模型的输出分类为1,即确定所述螺杆装置当日状态为异常,则设备1获取当日扭矩数据,将所述当日扭矩数据与阈值进行比较,并对比较结果进行判断。比如,考虑到当日数据可能出现偶发性或阶段性异常,可以是将当日数据中的每一条有效数据与阈值分别进行比较,确定全部比较结果。

[0114] 继续上述实施例中,在所述步骤S18中,若比较结果满足预设条件,则设备1确认所述螺杆装置需要检修。其中,步骤S17中,若采取的是将当日数据中的每一条有效数据与阈值分别进行比较,确定全部比较结果,则预设条件是全部比较结果中有预设条有效扭矩数据大于阈值,比如,预设条件可以设定为全部比较结果中,当日扭矩数据中大于阈值的有效扭矩数据的出现个数超过35。

[0115] 在所述步骤S18中,若比较结果不满足预设条件,则设备1继续判断该螺杆装置当日之前的预设连续天数内的状态是否异常,以确认所述螺杆装置是否需要检修。比如,若预设连续天数为N天,在设备1判断比较的结果不满足预设条件后,继续判断到该螺杆装置当日之前的连续N天的状态都是异常,则确定该螺杆装置需要检修。

[0116] 可选地,其中,所述阈值的确定包括:

[0117] 确定用于训练所述螺杆装置状态预测模型的训练子集中,所有输出分类为正常所对应的标准化处理后的特征数据;

[0118] 将所有所述标准化处理后的特征数据对应的历史数据的均值的预设倍数确定为阈值。

[0119] 其中,螺杆装置健康状态下的历史扭矩数据是在一定的均值上下波动并符合正态分布的,在均值上下一定范围内的误差属于随机误差,不能体现螺杆装置状态是否异常,因此,可将用于判断螺杆装置状态的阈值设置为螺杆装置健康状态下的历史扭矩数据均值的预设倍数,即若扭矩数据偏差均值较大,不可能是随机误差造成的,可认为是螺杆装置状态异常造成的。

[0120] 在上述实施例中,假设螺杆装置状态预测模型对应的训练子集中的输出分类为正常的样本对应的历史扭矩数据值为(45.400002,45.299998,45.955550,45.000000,45.800002,45.700001),则其均值为45.525926,若预设倍数为2倍,则阈值为91.051851。

[0121] 继续上述实施例中,若确定该螺杆装置需要检修,设备1还可以发出需对螺杆装置和/或采用螺杆装置的设备进行检修的告警信息。

[0122] 图3示出根据本申请另一个方面的一种用于螺杆装置状态预测模型训练的设备示意图,其中,一个实施例的所述设备包括:

[0123] 第一装置31,用于基于获取的所述螺杆装置的历史数据集,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注;

[0124] 第二装置32,用于基于每个训练子集,分别训练无监督分类模型,获得若干训练后的无监督分类模型;

[0125] 第三装置33,用于基于所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;

[0126] 第四装置34,用于基于所述测试集中每个样本的分类标注及其对应的每个训练后的无监督分类模型的输出分类,计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为所述螺杆装置状态预测模型。

[0127] 其中,所述设备与上述设备1相同。

[0128] 在该实施例中,采用上述方法,所述设备1的第一装置31对获取的该螺杆装置的历史数据集进行处理,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注,其中,还可以对测试集及训练子集的样本进行标准化处理;所述设备1的第二装置32将每个训练子集分别用于训练无监督分类模型,获得若干训练后的无监督分类模型;所述设备1的第三装置33采用所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;所述设备1的第四装置34根据测试集各样本的分类标注及对应每个训练后的无监督分类模型的输出分类,采用混淆矩阵计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为该螺杆装置状态预测模型。第四装置34还用于获取该最高F1-Score对应的训练子集中的正常样本对应的历史数据的均值。

[0129] 后续还可以通过该螺杆装置状态预测模型对所述螺杆装置的当日状态进行预测评估。

[0130] 可选地,其中,所述设备还包括:

[0131] 第五装置35(未示出),用于基于获取的所述螺杆装置的当日数据,确定其对应的当日特征数据,并将所述当日特征数据进行标准化处理;

[0132] 第六装置36(未示出),用于将标准化处理后的当日特征数据输入所述螺杆装置状态预测模型,以确定所述螺杆装置当日状态是否异常。

[0133] 在该实施例中,继续采用上述方法,所述设备1的第五装置35获取该螺杆装置的当日原始扭矩数据,对原始扭矩数据进行清洗,剔除异常数据集重复数据后,获得当日有效数据,并确定与其对应的当日特征数据,其中,当日特征数据包含的特征参数应与前述训练集及测试集中样本中包含的特征参数一致。将当日特征数据进行标准化处理,其中,采用螺杆装置状态预测模型对应的训练子集中对应各特征参数的所有样本的相应参数值的均值(mean)及均方差(std),然后将当日特征数据中相应参数值x经过如下标准化公式处理,

[0134] $(x - \text{mean}) / \text{std}$

[0135] 得到标准化处理后的当日特征数据。

[0136] 继续上述实施例中,所述设备1的第六装置36将标准化处理后的当日特征数据输入该螺杆装置状态预测模型,获得输出分类,根据输出分类确定所述螺杆装置当日状态是否异常。比如,若输出分类为0,则可确定该螺杆装置当日状态正常;若输出分类为1,则可确定该螺杆装置当日状态异常。

[0137] 由于螺杆装置日常生产环境的复杂性,状态异常可能是螺杆装置本身因素造成的,也可能是其它客观因素造成的,而过于频繁的停机检修会造成经济损失,不可取。因此,对于螺杆装置当日状态异常的情况,还有必要对当日数据进行进一步分析处理。

[0138] 所述设备1还可以包括第七装置37(未示出)、第八装置38(未示出)及第九装置39(未示出),若所述设备1的第六装置36确定该螺杆装置当日状态为异常,所述设备1的第七装置37则获取当日扭矩数据,将所述当日扭矩数据与阈值进行比较,对比较结果进行判断;若比较的结果满足预设条件,则所述设备1的第八装置38确认所述螺杆装置需要检修,否则,判断所述螺杆装置当日之前预设连续天数的状态是否异常,以确认所述螺杆装置是否需要检修。当确认所述螺杆装置需要检修时,所述设备1的第九装置39还可发出需对螺杆装

置和/或采用螺杆装置的设备进行检修的告警信息。

[0139] 根据本申请的还一方面,还提供了一种计算机可读介质,所述计算机可读介质存储有计算机可读指令,所述计算机可读指令可被处理器执行以实现前述方法。

[0140] 根据本申请的再一方面,还提供了一种用于螺杆装置状态预测模型训练的设备,其中,该设备包括:

[0141] 一个或多个处理器;以及

[0142] 存储有计算机可读指令的存储器,所述计算机可读指令在被执行时使所述处理器执行如前述方法的操作。

[0143] 例如,计算机可读指令在被执行时使所述一个或多个处理器:

[0144] 对获取的该螺杆装置的历史数据集进行处理,确定测试集及若干训练子集,并对测试集中每个样本进行分类标注,其中,还可以对测试集及训练子集的样本进行标准化处理;将每个训练子集分别用于训练无监督分类模型,获得若干训练后的无监督分类模型;采用所述测试集,分别测试每个训练后的无监督分类模型,获得测试集中每个样本对应的每个训练后的无监督分类模型的输出分类;根据测试集各样本的分类标注及对应每个训练后的无监督分类模型的输出分类,采用混淆矩阵计算每个训练后的无监督分类模型的F1-Score,并将最高F1-Score对应的训练后的无监督分类模型确定为该螺杆装置状态预测模型。进一步地,还可以通过该螺杆装置状态预测模型对所述螺杆装置的当日状态进行预测评估。

[0145] 对于本领域技术人员而言,显然本发明不限于上述示范性实施例的细节,而且在不背离本发明的精神或基本特征的情况下,能够以其他的具体形式实现本发明。因此,无论从哪一点来看,均应将实施例看作是示范性的,而且是非限制性的,本发明的范围由所附权利要求而不是上述说明限定,因此旨在将落在权利要求的等同要件的含义和范围内的所有变化涵括在本发明内。不应将权利要求中的任何附图标记视为限制所涉及的权利要求。此外,显然“包括”一词不排除其他单元或步骤,单数不排除复数。装置权利要求中陈述的多个单元或装置也可以由一个单元或装置通过软件或者硬件来实现。第一,第二等词语用来表示名称,而并不表示任何特定的顺序。

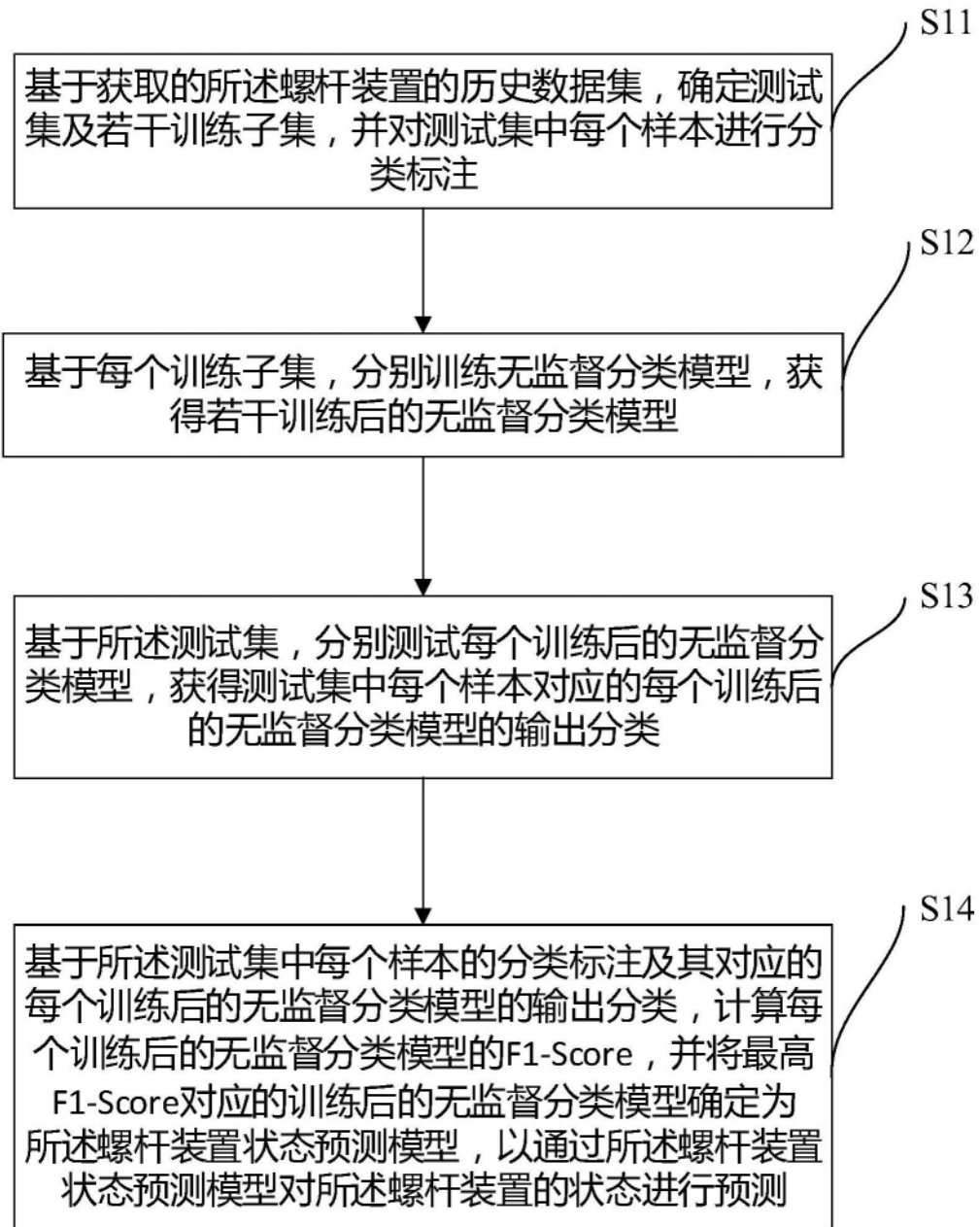


图1

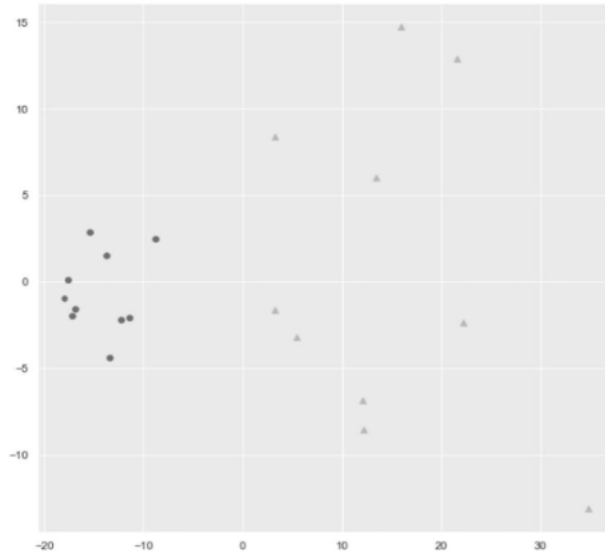


图2

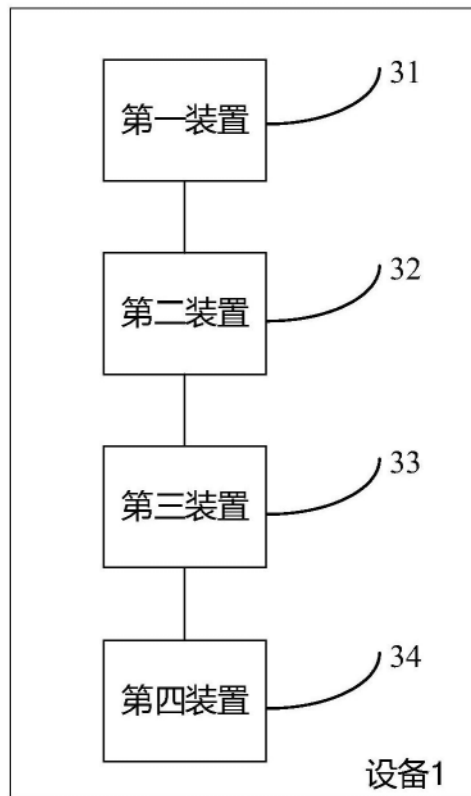


图3