

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局



(10) 国际公布号

WO 2021/184727 A1

(43) 国际公布日

2021 年 9 月 23 日 (23.09.2021)

(51) 国际专利分类号:

G06F 11/30 (2006.01)

G06N 3/04 (2006.01)

G06K 9/62 (2006.01)

G06N 3/08 (2006.01)

(21) 国际申请号:

PCT/CN2020/119060

(22) 国际申请日:

2020 年 9 月 29 日 (29.09.2020)

(25) 申请语言:

中文

(26) 公布语言:

中文

(30) 优先权:

202010195664.3

2020 年 3 月 19 日

(19.03.2020) CN

(71) 申请人: 平安科技(深圳)有限公司(PING AN TECHNOLOGY (SHENZHEN) CO.,LTD.) [CN/CN];

中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

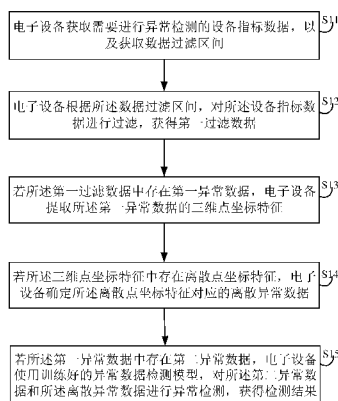
(72) 发明人: 徐锐杰(XU, Ruijie); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼, Guangdong 518000 (CN)。

(74) 代理人: 深圳市沃德知识产权代理有限公司(普通合伙)(SHENZHEN WORLD INTELLECTUAL PROPERTY AGENCY (GENERAL PARTNERSHIP)); 中国广东省深圳市福田区梅林街道中康路 128 号卓越城 1 期 2 栋 301B 于志光, Guangdong 518000 (CN)。

(81) 指定国(除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL,

(54) Title: DATA ABNORMALITY DETECTION METHOD AND APPARATUS, ELECTRONIC DEVICE AND STORAGE MEDIUM

(54) 发明名称: 数据异常检测方法、装置、电子设备及存储介质



S11 An electronic device acquires device index data that needs to be detected for abnormalities, and acquires a data filtering interval
S12 The electronic device filters the device index data according to the data filtering interval to obtain first filtered data
S13 If first abnormal data is present in the first filtered data, then the electronic device extracts three-dimensional point coordinate features of the first abnormal data
S14 If discrete point coordinate features are present in the three-dimensional point coordinate features, then the electronic device determines discrete abnormal data corresponding to the discrete point coordinate features
S15 If second abnormal data is present in the first abnormal data, then the electronic device performs abnormality detection on the second abnormal data and the discrete abnormal data by using a trained abnormal data detection model to obtain a detection result

图 1

(57) Abstract: A data abnormality detection method and apparatus, an electronic device and a storage medium, which may improve the accuracy of abnormality detection and reduce the consumption of computer resources. The method comprises: acquiring device index data that needs to be detected for abnormalities, and acquiring a data filtering interval (S11); filtering the device index data according to the data filtering interval to obtain first filtered data (S12); if first abnormal data is present in the first filtered data, then extracting three-dimensional point coordinate features of the first abnormal data (S13); if discrete point coordinate features are present in the three-dimensional point coordinate features, then determining discrete abnormal data corresponding to the discrete point coordinate features (S14); and if second abnormal data is present in the first abnormal data, then performing abnormality detection on the second abnormal data and the discrete abnormal data by using a trained abnormal data detection model to obtain a detection result (S15).

ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US,
UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区
保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ,
NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM,
AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG,
CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU,
IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT,
RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI,
CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:
— 包括国际检索报告 (条约第21条(3))。

(57) 摘要: 一种数据异常检测方法、装置、电子设备及存储介质, 能提高异常检测的准确率以及减少计算机资源的消耗。所述方法包括: 获取需要进行异常检测的设备指标数据, 以及获取数据过滤区间 (S11); 根据数据过滤区间, 对设备指标数据进行过滤, 获得第一过滤数据 (S12); 若第一过滤数据中存在第一异常数据, 提取第一异常数据的三维点坐标特征 (S13); 若三维点坐标特征中存在离散点坐标特征, 确定离散点坐标特征对应的离散异常数据 (S14); 若第一异常数据中存在第二异常数据, 使用训练好的异常数据检测模型, 对第二异常数据和离散异常数据进行异常检测, 获得检测结果 (S15)。

数据异常检测方法、装置、电子设备及存储介质

本申请要求于 2020 年 3 月 19 日提交中国专利局、申请号为 CN202010195664.3、名称为“数据异常检测方法、装置、电子设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请涉及异常检测技术领域，尤其涉及一种数据异常检测方法、装置、电子设备及存储介质。

背景技术

目前，随着计算机技术的发展，由大量主机以及网络交换机等设备组成的云平台得到大规模的应用。为了确保云平台可以稳定地提供服务，需要对云平台的设备的一些指标数据进行实时监控，发明人意识到，使用无监督的学习算法进行异常检测，容易出现误告警问题，异常检测的准确率不高，而使用基于神经网络算法的模型进行异常检测，需要消耗的计算机资源太多。

因此，如何提高异常检测的准确率以及减少计算机资源的消耗是一个亟需解决的技术问题。

发明内容

本申请的第一方面提供一种数据异常检测方法，所述方法包括：

获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

本申请的第二方面提供一种数据异常检测装置，所述装置包括：

获取模块，用于获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

过滤模块，用于根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

判断模块，用于根据拉依达准则、高斯模型以及指数加权移动平均法，判断所述第一过滤数据中是否存在第一异常数据；

提取模块，用于若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

添加模块，用于将所述第一异常数据的三维点坐标特征添加至坐标特征集合中；

所述判断模块，还用于使用基于密度的聚类算法，对所述坐标特征集合中的特征进行聚类分析，获得每个所述三维点坐标特征对应的至少一个类别，并根据所述每个所述三维点坐标特征对应的至少一个类别，判断所述三维点坐标特征中是否存在离散点坐标特征；

确定模块，用于若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

所述判断模块，还用于使用主成分分析算法，判断所述第一异常数据中是否存在第二异常数据；

检测模块，用于若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检

测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

本申请的第三方面提供一种电子设备，所述电子设备包括处理器和存储器，所述处理器用于执行所述存储器中存储的计算机程序时实现如下步骤：

获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

本申请的第四方面提供一种计算机可读存储介质，所述计算机可读存储介质上存储有计算机程序，所述计算机程序被处理器执行时实现如下步骤：

获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

附图说明

图1是本申请公开的一种数据异常检测方法的较佳实施例的流程图。

图2是本申请公开的一种数据异常检测装置的较佳实施例的功能模块图。

图3是本申请实现数据异常检测方法的较佳实施例的电子设备的结构示意图。

具体实施方式

下面将结合本申请实施例中的附图，对本申请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申请一部分实施例，而不是全部的实施例。基于本申请中的实施例，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他实施例，都属于本申请保护的范围。

除非另有定义，本文所使用的所有的技术和科学术语与属于本申请的技术领域的技术人员通常理解的含义相同。本文中在本申请的说明书中所使用的术语只是为了描述具体的实施例的目的，不是旨在限制本申请。

本申请实施例的数据异常检测方法应用在电子设备中，也可以应用在电子设备和通过网络与所述电子设备进行连接的服务器所构成的硬件环境中，由服务器和电子设备共同执行。网络包括但不限于：广域网、城域网或局域网。

所述电子设备是一种能够按照事先设定或存储的指令，自动进行数值计算和/或信息处理的设备，其硬件包括但不限于微处理器、专用集成电路(ASIC)、现场可编程门阵列(FPGA)、数字处理器(DSP)、嵌入式设备等。所述电子设备还可包括网络设备和/或用户设备。其中，所述网络设备包括但不限于单个网络设备、多个网络设备组成的服务器组或基于云计算(Cloud Computing)的由大量主机或网络设备构成的云，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超级虚拟计算机。所述用户设备包括但不限于任何一种可与用户通过键盘、鼠标、遥控器、触摸板或声控设备等方式进行人机交互的电子产

品，例如，个人计算机、平板电脑、智能手机、个人数字助理 PDA 等。

请参见图 1，图 1 是本申请公开的一种数据异常检测方法的较佳实施例的流程图。其中，根据不同的需求，该流程图中步骤的顺序可以改变，某些步骤可以省略。

S11、电子设备获取需要进行异常检测的设备指标数据，以及获取数据过滤区间。

其中，所述设备指标数据可以为 CPU 使用量、内存使用量、硬盘存储空间、网络交换机请求数量以及网络出口访问量等。

具体的，所述获取数据过滤区间包括：

获取用户设置的数据过滤区间；或

获取历史统计数据，并根据所述历史统计数据，确定数据过滤区间。

在该可选的实施方式中，用户可以根据实际情况预先设置一个数据过滤区间，比如设置内存使用率过滤区间为 $[10\%, 100\%]$ ，如果内存的实时使用率没有达到 10% ，则无需对这时的数据进行异常检测，节省计算机资源。若用户没有预先设置数据过滤区间，则可以统计一个月的历史数据，获得历史统计数据，然后获取历史统计数据的百分位数 0.01 与百分位数 0.99，将百分位数 0.01 与百分位数 0.99 确定数据过滤区间，比如，内存使用率的历史统计数据的百分位数 0.01 是 10% ，内存使用率的历史统计数据的百分位数 0.99 为 88% ，则将内存使用率的数据过滤区间确定为 $[10\%, 88\%]$ 。

作为一种可选的实施方式，在步骤 S11 之前，所述方法还包括：

获取用于训练的第一样本数据；

使用无监督的学习算法对所述第一样本数据进行筛选，获得第二样本数据；

使用所述第二样本数据进行训练，获得训练好的异常检测模型。

其中，所述无监督的学习算法可以为基于拉依达准则、高斯模型、指数加权移动平均法、基于密度的聚类算法以及主成分分析算法等。

在该可选的实施方式中，因为设备在大多数时间都是处于正常运行状态，所以获取的大部分的设备指标数据都为正常数据，而用于训练的第一样本数据也是通过统计设备指标数据来的，因此，第一样本数据中的异常数据所占的比例极低，从而导致使用第一样本数据训练出来的异常检测模型的准确度较低。可以通过无监督的学习算法，对所述第一样本数据进行筛选，即过滤掉大部分正常数据，以提高异常数据所占的比例。通过筛选获得第二样本数据，然后使用第二样本数据进行训练，可以获得准确度较高的异常检测模型。

可选的，可以获取当前点的前三小时的数据，昨天同一时刻前后三小时的数据，前天同一时刻前后三个小时的数据，7 天前同一时刻的前后三个小时的数据，以及 14 天前前后三个小时的数据共五个片段。分别提取上述五个片段的最大值、最小值、均值、中值、偏度、峰度、差分等统计特征，使用 DNN (Deep Neural Networks, 深度神经网络) 神经网络算法进行分类训练，假设每个片段提取上述 7 个特征，则 5 个片段共提取 35 个特征点 $(x_1, x_2, x_3, \dots, x_{35})$ ，因此输入的向量 X 为 $[x_1, x_2, x_3, \dots, x_{35}]$ ，假设输出的检测结果 Y 的值如果为 0 则代表正常、为 1 则代表异常。可以通过梯度下降算法训练出模型的参数 W 、 b ，假设 σ 为一个激活函数，则模型公式如下：

$$Y = \sigma(WX + b);$$

假设 L 为损失函数， y 为异常检测模型的检测结果， $\hat{y}$ 为人工为样本标记的结果，则损失函数的公式如下：

$$L(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}).$$

S12、电子设备根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据。

本申请实施例中，通过确定数据过滤区间，可以过滤不需要进行监控的数据，节省了计算机资源。

作为一种可选的实施方式,所述步骤 S12 之后,所述方法还包括:

针对每个所述第一过滤数据,根据拉依达准则,判断所述第一过滤数据是否属于第一预设数据区间;

若所述第一过滤数据属于第一预设数据区间,根据高斯分布模型,判断所述第一过滤数据对应的密度估计值是否小于预设密度估计阈值;

若所述第一过滤数据对应的密度估计值大于或等于预设密度估计阈值,使用指数加权移动平均法,判断所述第一过滤数据对应的指数加权平均值是否属于第二预设数据区间;

若所述第一过滤数据对应的指数加权平均值不属于第二预设数据区间,确定所述第一过滤数据中存在第一异常数据;或

若所述第一过滤数据对应的指数加权平均值属于第二预设数据区间,确定所述第一过滤数据未存在第一异常数据。

其中,所述拉依达准则(3sigma)指先假设一组检测数据只含有随机误差,对其进行计算处理得到标准偏差,按一定概率确定一个区间,认为凡超过这个区间的误差,就不属于随机误差而是粗大误差,含有该误差的数据应予以剔除。

其中,所述高斯模型是用高斯概率密度函数(正态分布曲线)精确地量化的事物,将一个事物分解为若干的基于高斯概率密度函数(正态分布曲线)形成的模型。

其中,所述指数加权移动平均法(EWMA, Exponentially Weighted Moving-Average)是是对观察值分别给予不同的权数,按不同权数求得移动平均值,并以最后的移动平均值为基础,确定预测值的方法,各数值的加权系数随时间呈指数式递减,越靠近当前时刻的数值加权系数就越大。

在该可选的实施方式中,假设均值为 μ ,标准差为 σ ,基于拉依达准则,则在 $\mu \pm 3\sigma$ 范围内的数据为正常数据,基于高斯模型,可以设定一个密度估计阈值,若数据的密度估计值小于这个密度估计阈值,则该数据为异常数据,其中,计算所述密度估计值 $p(x)$ 的公式为: $p(x) = \prod p(x; u; \sigma^2)$ 。基于指数加权移动平均法,假设权重为 α ,指数加权移动平均值 $m_i = \alpha m_{i-1} + (1 - \alpha)x_i$ 。若指数加权移动平均值在 $\mu \pm 3\sigma$ 的范围内,则认为数据正常。

S13、若所述第一过滤数据中存在第一异常数据,电子设备提取所述第一异常数据的三维点坐标特征。

具体的,所述提取所述第一异常数据的三维特点坐标征包括:

获取所述第一异常数据的差分值、当前值以及指数加权移动平均值;

计算所述当前值与所述指数加权移动平均值的目标差值;

将所述差分值、所述当前值以及所述目标差值确定为所述第一异常数据的三维点坐标特征。

其中,所述差分值可以指当前时间点的数据与上一个时间点的数据的差值。

在该可选的实施方式中,可以将所述第一异常数据的差分值、当前值以及指数加权移动平均值作为所述第一异常数据的X轴、Y轴以及Z轴的坐标值(即三维点坐标)。一个三维点坐标特征可以表示一个第一异常数在三维图中的一个点。

作为一种可选的实施方式,所述提取所述第一异常数据的三维点坐标特征之后,所述方法还包括:

将所述第一异常数据的三维点坐标特征添加至坐标特征集合中;

使用基于密度的聚类算法,对所述坐标特征集合中的特征进行聚类分析,获得每个所述三维点坐标特征对应的至少一个类别;

根据每个所述三维点坐标特征对应的至少一个类别,统计每个所述三维点坐标特征对应的类别数量;

判断所有所述三维点坐标特征中是否存在类别数量小于预设类别数量阈值的三维点坐标特征；

若所述三维点坐标特征中存在类别数量小于预设类别数量阈值的三维点坐标特征，将类别数量小于预设类别数量阈值的三维点坐标特征确定为离散点坐标特征；或

若所述三维点坐标特征中未存在类别数量大于或等于预设类别数量阈值的三维点坐标特征，确定所述三维点坐标特征中未存在离散点坐标特征。

其中，所述坐标特征集合保存着在以往时间提取的异常数据的三维点坐标特征。

其中，所述基于密度的聚类算法（Density-Based Spatial Clustering of Applications with Noise, DBSCAN）能够把具有足够高密度的区域划分为簇，并可在噪声的空间数据库中发现任意形状的聚类。所述基于密度的聚类算法可以将三维图（空间）中的点进行聚类，如果一个点所处区域的密度越高，则该点属于的类别越多，通过所述基于密度的聚类算法，可以确定每个点对应的类别数量，找出对应的类别数量少的点，即离散的点（该离散的点的坐标就是离散点坐标特征）。

S14、若所述三维点坐标特征中存在离散点坐标特征，电子设备确定所述离散点坐标特征对应的离散异常数据。

其中，所述离散点坐标特征所述表示的点是离散的点，离散的点很大概率是异常的点，即所述离散点坐标特征对应的离散异常数据很可能是真的异常数据。

作为一种可选的实施方式，所述确定所述离散点坐标特征对应的离散异常数据之后，所述方法还包括：

使用主成分分析算法，对所述第一异常数据进行降维，获得降维数据；

对所述降维数据进行逆向计算，获得目标数据；

判断所述目标数据与所述第一异常数据的差值是否大于预设差值阈值；

若所述还原数据与所述第一异常数据的差值大于预设差值阈值，确定所述第一异常数据中存在第二异常数据；或

若所述还原数据与所述第一异常数据的差值小于或等于预设差值阈值，确定所述第一异常数据中未存在第二异常数据。

其中，所述主成分分析算法（Principle Component Analysis, PCA）可以将数据映射到低维特征空间，然后由低维特征空间重新映射回原空间，尝试用低维特征重构原始数据，看重构误差的大小。如果一个数据不容易被重构出来，即重构（逆向计算）获得的所述目标数据和对应第一异常数据的差值大于预设差值阈值，表示该第一异常数据的特征跟整体数据的特征不一致，该第一异常数据被认为是真的异常数据。

S15、若所述第一异常数据中存在第二异常数据，电子设备使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

其中，所述检测结果可以是正常，可以是异常，可以输出数字 0 表示所述检测结果为正常，可以输出数字 1 表示所述检测结果为异常。

本申请实施例中，因为，所述第二异常数据以及所述离散异常数据中还存在部分误报的正常数据，可以用异常检测模型对所述第二异常数据以及所述离散异常数据进行异常检测，因为所述第二异常数据以及所述离散异常数据是经过了两层无监督的学习算法筛选的（一层无监督的学习算法是指拉依达准则、高斯模型以及指数加权移动平均法，另一层无监督的学习算法是指基于密度的聚类算法以及主成分分析算法），所述第二异常数据以及所述离散异常数据相较于所述设备指标数据，数据量少很多，这时再用异常检测模型对少量的数据进行异常检测，可以节省计算机资源，同时，可以提高异常检测的准确率。

在图 1 所描述的方法流程中，可以通过拉依达准则、指数加权移动平均法以及高斯模型这些简单的无监督的学习算法对设备指标数据进行初步的判定，获得第一异常数据，其中，第一异常数据中存在很多误报的正常数据，可以通过基于密度的聚类算法以及主成分分析算法这些相对高级的无监督的学习算法对第一异常数据进行进一步的判定，获得第二异常数据，其中，第二异常数据中依然存在误报的正常数据，通过两次的判定获得的第二异常数据的数量已经是很少了的，此外，整个过程中采用的无监督的学习算法（如拉依达准则、高斯模型以及指数加权移动平均法、密度的聚类算法等）比较简单，所消耗的计算机资源较少，而且异常检测模型只需要对少量的数据进行检测，也节约了计算机资源，另外，异常检测模型的训练样本是通过无监督的学习算法筛选的样本，在训练样本中不会出现异常数据样本的数量过少的问题，从而提高了异常检测模型的准确率。

以上所述，仅是本申请的具体实施方式，但本申请的保护范围并不局限于此，对于本领域的普通技术人员来说，在不脱离本申请创造构思的前提下，还可以做出改进，但这些均属于本申请的保护范围。

请参见图 2，图 2 是本申请公开的一种数据异常检测装置的较佳实施例的功能模块图。

在一些实施例中，所述数据异常检测装置运行于电子设备中。所述数据异常检测装置可以包括多个由程序代码段所组成的功能模块。所述数据异常检测装置中的各个程序段的程序代码可以存储于存储器中，并由至少一个处理器所执行，以执行图 1 所描述的数据异常检测方法中的部分或全部步骤。

本实施例中，所述数据异常检测装置根据其所执行的功能，可以被划分为多个功能模块。所述功能模块可以包括：获取模块 201、过滤模块 202、提取模块 203、确定模块 204 及检测模块 205。本申请所称的模块是指一种能够被至少一个处理器所执行并且能够完成固定功能的一系列计算机程序段，其存储在存储器中。

获取模块 201，用于获取需要进行异常检测的设备指标数据，以及获取数据过滤区间。

过滤模块 202，用于根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据。

提取模块 203，用于若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征。

确定模块 204，用于若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据。

检测模块 205，用于若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

作为一种可选的实施方式，所述获取模块 201 获取数据过滤区间的方式具体为：获取用户设置的数据过滤区间；或

获取历史统计数据，并根据所述历史统计数据，确定数据过滤区间。

作为一种可选的实施方式，所述数据异常检测还可以包括

第一判断模块，用于所述过滤模块 202 根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据之后，针对每个所述第一过滤数据，根据拉依达准则，判断所述第一过滤数据是否属于第一预设数据区间；

所述第一判断模块，还用于若所述第一过滤数据属于第一预设数据区间，根据高斯分布模型，判断所述第一过滤数据对应的密度估计值是否小于预设密度估计阈值；

所述第一判断模块，还用于若所述第一过滤数据对应的密度估计值大于或等于预

设密度估计阈值，使用指数加权移动平均法，判断所述第一过滤数据对应的指数加权平均值是否属于第二预设数据区间；

所述确定模块 204，还用于若所述第一过滤数据对应的指数加权平均值不属于第二预设数据区间，确定所述第一过滤数据中存在第一异常数据；或

所述确定模块 204，还用于若所述第一过滤数据对应的指数加权平均值属于第二预设数据区间，确定所述第一过滤数据未存在第一异常数据。

作为一种可选的实施方式，所述数据异常检测装置还可以包括：

添加模块，用于所述提取模块 203 提取所述第一异常数据的三维点坐标特征之后，将所述第一异常数据的三维点坐标特征添加至坐标特征集合中；

分析模块，用于使用基于密度的聚类算法，对所述坐标特征集合中的特征进行聚类分析，获得每个所述三维点坐标特征对应的至少一个类别；

统计模块，用于根据每个所述三维点坐标特征对应的至少一个类别，统计每个所述三维点坐标特征对应的类别数量；

第二判断模块，用于判断所有所述三维点坐标特征中是否存在类别数量小于预设类别数量阈值的三维点坐标特征；

所述确定模块 204，还用于若所述三维点坐标特征中存在类别数量大于或等于预设类别数量阈值的三维点坐标特征，将类别数量小于预设类别数量阈值的三维点坐标特征确定为离散点坐标特征；或

所述确定模块 204，还用于若所述三维点坐标特征中未存在类别数量小于预设类别数量阈值的三维点坐标特征，确定所述三维点坐标特征中未存在离散点坐标特征。

作为一种可选的实施方式，所述数据异常检测装置还可以包括：

降维模块，用于所述确定模块 204 确定所述离散点坐标特征对应的离散异常数据之后，使用主成分分析算法，对所述第一异常数据进行降维，获得降维数据；

计算模块，用于对所述降维数据进行逆向计算，获得目标数据；

第三判断模块，用于判断所述目标数据与所述第一异常数据的差值是否大于预设差值阈值；

所述确定模块 204，还用于若所述还原数据与所述第一异常数据的差值大于预设差值阈值，确定所述第一异常数据中存在第二异常数据；或

所述确定模块 204，还用于若所述还原数据与所述第一异常数据的差值小于或等于预设差值阈值，确定所述第一异常数据中未存在第二异常数据。

作为一种可选的实施方式，所述提取模块 203 提取所述第一异常数据的三维特征的方式具体为：

获取所述第一异常数据的差分值、当前值以及指数加权移动平均值；

计算所述当前值与所述指数加权移动平均值的目标差值；

将所述差分值、所述当前值以及所述目标差值确定为所述第一异常数据的三维点坐标特征。

作为一种可选的实施方式，所述获取模块 201，还用于获取用于训练的第一样本数据；

所述数据异常检测装置还可以包括：

筛选模块，用于使用无监督的学习算法对所述第一样本数据进行筛选，获得第二样本数据；

训练模块，用于使用所述第二样本数据进行训练，获得训练好的异常检测模型。

在图 2 所描述的数据异常检测装置中，可以通过拉依达准则、指数加权移动平均法以及高斯模型这些简单的无监督的学习算法对设备指标数据进行初步的判定，获得第一异常数据，其中，第一异常数据中存在很多误报的正常数据，可以通过基于密度

的聚类算法以及主成分分析算法这些相对高级的无监督的学习算法对第一异常数据进行进一步的判定,获得第二异常数据,其中,第二异常数据中依然存在误报的正常数据,通过两次的判定获得的第二异常数据的数量已经是很少了的,此外,整个过程中采用的无监督的学习算法(如拉依达准则、高斯模型以及指数加权移动平均法、密度的聚类算法等)比较简单,所消耗的计算机资源较少,而且异常检测模型只需要对少量的数据进行检测,也节约了计算机资源,另外,异常检测模型的训练样本是通过无监督的学习算法筛选的样本,在训练样本中不会出现异常数据样本的数量过少的问题,从而提高了异常检测模型的准确率。

如图3所示,图3是本申请实现数据异常检测方法的较佳实施例的电子设备的结构示意图。所述电子设备3包括存储器31、至少一个处理器32、存储在所述存储器31中并可在所述至少一个处理器32上运行的计算机程序33及至少一条通讯总线34。

本领域技术人员可以理解,图3所示的示意图仅仅是所述电子设备3的示例,并不构成对所述电子设备3的限定,可以包括比图示更多或更少的部件,或者组合某些部件,或者不同的部件,例如所述电子设备3还可以包括输入输出设备、网络接入设备等。

所述至少一个处理器32可以是中央处理单元(Central Processing Unit, CPU),还可以是其他通用处理器、数字信号处理器(Digital Signal Processor, DSP)、专用集成电路(Application Specific Integrated Circuit, ASIC)、现场可编程门阵列(Field-Programmable Gate Array, FPGA)或者其他可编程逻辑器件、晶体管逻辑器件、分立硬件组件等。该处理器32可以是微处理器或者该处理器32也可以是任何常规的处理器等,所述处理器32是所述电子设备3的控制中心,利用各种接口和线路连接整个电子设备3的各个部分。

所述存储器31可用于存储所述计算机程序33和/或模块/单元,所述处理器32通过运行或执行存储在所述存储器31内的计算机程序和/或模块/单元,以及调用存储在存储器31内的数据,实现所述电子设备3的各种功能。所述存储器31可主要包括存储程序区和存储数据区,其中,存储程序区可存储操作系统、至少一个功能所需的应用程序(比如声音播放功能、图像播放功能等)等;存储数据区可存储根据电子设备3的使用所创建的数据(比如音频数据等)等。此外,存储器31可以包括非易失性存储器,例如硬盘、内存、插接式硬盘,智能存储卡(Smart Media Card, SMC),安全数字(Secure Digital, SD)卡,闪存卡(Flash Card)、至少一个磁盘存储器件、闪存器件等。

结合图1,所述电子设备3中的所述存储器31存储多个指令以实现一种数据异常检测方法,所述处理器32可执行所述多个指令从而实现:

获取需要进行异常检测的设备指标数据,以及获取数据过滤区间;

根据所述数据过滤区间,对所述设备指标数据进行过滤,获得第一过滤数据;

若所述第一过滤数据中存在第一异常数据,提取所述第一异常数据的三维点坐标特征;

若所述三维点坐标特征中存在离散点坐标特征,确定所述离散点坐标特征对应的离散异常数据;

若所述第一异常数据中存在第二异常数据,使用训练好的异常数据检测模型,对所述第二异常数据和所述离散异常数据进行异常检测,获得检测结果,其中,所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

具体地,所述处理器32对上述指令的具体实现方法可参考图1对应实施例中相关步骤的描述,在此不赘述。

在图3所描述的电子设备3中,可以通过拉依达准则、指数加权移动平均法以及高斯模型这些简单的无监督的学习算法对设备指标数据进行初步的判定,获得第一异常数据,其中,第一异常数据中存在很多误报的正常数据,可以通过基于密度的聚类

算法以及主成分分析算法这些相对高级的无监督的学习算法对第一异常数据进行进一步的判定，获得第二异常数据，其中，第二异常数据中依然存在误报的正常数据，通过两次的判定获得的第二异常数据的数量已经是很少了的，此外，这整个过程中采用的无监督的学习算法（如拉依达准则、高斯模型以及指数加权移动平均法、密度的聚类算法等）比较简单，所消耗的计算机资源较少，而且异常检测模型只需要对少量的数据进行检测，也节约了计算机资源，另外，异常检测模型的训练样本是通过无监督的学习算法筛选的样本，在训练样本中不会出现异常数据样本的数量过少的问题，从而提高了异常检测模型的准确率。

所述电子设备 3 集成的模块/单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读取存储介质中。基于这样的理解，本申请实现上述实施例方法中的全部或部分流程，也可以通过计算机程序来指令相关的硬件来完成，所述的计算机程序可存储于一计算机可读存储介质中，所述计算机可读存储介质可以是易失性，也可以是非易失性，该计算机程序在被处理器执行时，可实现上述各个方法实施例的步骤。其中，所述计算机程序包括计算机程序代码，所述计算机程序代码可以为源代码形式、对象代码形式、可执行文件或某些中间形式等。所述计算机可读介质可以包括：能够携带所述计算机程序代码的任何实体或装置、记录介质、U 盘、移动硬盘、磁碟、光盘、计算机存储器、只读存储器（ROM，Read-Only Memory）。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，所述模块的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式。

所述作为分离部件说明的模块可以是或者也可以不是物理上分开的，作为模块显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部模块来实现本实施例方案的目的。

另外，在本申请各个实施例中的各功能模块可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用硬件加软件功能模块的形式实现。

对于本领域技术人员而言，显然本申请不限于上述示范性实施例的细节，而且在不背离本申请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申请。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，本申请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申请内。不应将权利要求中的任何附关联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。系统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。第二等词语用来表示名称，而并不表示任何特定的顺序。

最后应说明的是，以上实施例仅用以说明本申请的技术方案而非限制，尽管参照较佳实施例对本申请进行了详细说明，本领域的普通技术人员应当理解，可以对本申请的技术方案进行修改或等同替换，而不脱离本申请技术方案的精神和范围。

权利要求书

1. 一种数据异常检测方法，其中，所述方法包括：
获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；
根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；
若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；
若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；
若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。
2. 根据权利要求1所述的方法，其中，所述获取数据过滤区间包括：
获取用户设置的数据过滤区间；或
获取历史统计数据，并根据所述历史统计数据，确定数据过滤区间。
3. 根据权利要求1所述的方法，其中，所述根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据之后，所述方法还包括：
针对每个所述第一过滤数据，根据拉依达准则，判断所述第一过滤数据是否属于第一预设数据区间；
若所述第一过滤数据属于第一预设数据区间，根据高斯分布模型，判断所述第一过滤数据对应的密度估计值是否小于预设密度估计阈值；
若所述第一过滤数据对应的密度估计值大于或等于预设密度估计阈值，使用指数加权移动平均法，判断所述第一过滤数据对应的指数加权平均值是否属于第二预设数据区间；
若所述第一过滤数据对应的指数加权平均值不属于第二预设数据区间，确定所述第一过滤数据中存在第一异常数据；或
若所述第一过滤数据对应的指数加权平均值属于第二预设数据区间，确定所述第一过滤数据未存在第一异常数据。
4. 根据权利要求1所述的方法，其中，所述提取所述第一异常数据的三维点坐标特征之后，所述方法还包括：
将所述第一异常数据的三维点坐标特征添加至坐标特征集合中；
使用基于密度的聚类算法，对所述坐标特征集合中的特征进行聚类分析，获得每个所述三维点坐标特征对应的至少一个类别；
根据每个所述三维点坐标特征对应的至少一个类别，统计每个所述三维点坐标特征对应的类别数量；
判断所有所述三维点坐标特征中是否存在类别数量小于预设类别数量阈值的三维点坐标特征；
若所述三维点坐标特征中存在类别数量大于或等于预设类别数量阈值的三维点坐标特征，将类别数量小于预设类别数量阈值的三维点坐标特征确定为离散点坐标特征；或
若所述三维点坐标特征中未存在类别数量小于预设类别数量阈值的三维点坐标特征，确定所述三维点坐标特征中未存在离散点坐标特征。
5. 根据权利要求1至4中任一项所述的方法，其中，所述确定所述离散点坐标特征对应的离散异常数据之后，所述方法还包括：
使用主成分分析算法，对所述第一异常数据进行降维，获得降维数据；
对所述降维数据进行逆向计算，获得目标数据；
判断所述目标数据与所述第一异常数据的差值是否大于预设差值阈值；
若所述还原数据与所述第一异常数据的差值大于预设差值阈值，确定所述第一异常数

据中存在第二异常数据；或

若所述还原数据与所述第一异常数据的差值小于或等于预设差值阈值，确定所述第一异常数据中未存在第二异常数据。

6. 根据权利要求1至4中任一项所述的方法，其中，所述提取所述第一异常数据的三维特征包括：

获取所述第一异常数据的差分值、当前值以及指数加权移动平均值；

计算所述当前值与所述指数加权移动平均值的目标差值；

将所述差分值、所述当前值以及所述目标差值确定为所述第一异常数据的三维点坐标特征。

7. 根据权利要求1至4中任一项所述的方法，其中，所述获取需要进行异常检测的设备指标数据之前，所述方法还包括：

获取用于训练的第一样本数据；

使用无监督的学习算法对所述第一样本数据进行筛选，获得第二样本数据；使用所述第二样本数据进行训练，获得训练好的异常检测模型。

8. 一种数据异常检测装置，其中，所述数据异常检测装置包括：

获取模块，用于获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

过滤模块，用于根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

提取模块，用于若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

确定模块，用于若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

检测模块，用于若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

9. 一种电子设备，其中，所述电子设备包括处理器和存储器，所述处理器用于执行存储器中存储的计算机程序以实现如下步骤：

获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散异常数据；

若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

10. 根据权利要求9所述的电子设备，其中，所述获取数据过滤区间包括：

获取用户设置的数据过滤区间；或

获取历史统计数据，并根据所述历史统计数据，确定数据过滤区间。

11. 根据权利要求9所述的电子设备，其中，所述根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据之后，所述处理器执行所述计算机程序时还实现如下步骤：

针对每个所述第一过滤数据，根据拉依达准则，判断所述第一过滤数据是否属于第一预设数据区间；

若所述第一过滤数据属于第一预设数据区间，根据高斯分布模型，判断所述第一过滤数据对应的密度估计值是否小于预设密度估计阈值；

若所述第一过滤数据对应的密度估计值大于或等于预设密度估计阈值，使用指数加权移动平均法，判断所述第一过滤数据对应的指数加权平均值是否属于第二预设数据区间；

若所述第一过滤数据对应的指数加权平均值不属于第二预设数据区间，确定所述第一过滤数据中存在第一异常数据；或

若所述第一过滤数据对应的指数加权平均值属于第二预设数据区间，确定所述第一过滤数据未存在第一异常数据。

12. 根据权利要求 9 所述的电子设备，其中，所述提取所述第一异常数据的三维点坐标特征之后，所述处理器执行所述计算机程序时还实现如下步骤：

将所述第一异常数据的三维点坐标特征添加至坐标特征集合中；

使用基于密度的聚类算法，对所述坐标特征集合中的特征进行聚类分析，获得每个所述三维点坐标特征对应的至少一个类别；

根据每个所述三维点坐标特征对应的至少一个类别，统计每个所述三维点坐标特征对应的类别数量；

判断所有所述三维点坐标特征中是否存在类别数量小于预设类别数量阈值的三维点坐标特征；

若所述三维点坐标特征中存在类别数量大于或等于预设类别数量阈值的三维点坐标特征，将类别数量小于预设类别数量阈值的三维点坐标特征确定为离散点坐标特征；或

若所述三维点坐标特征中未存在类别数量小于预设类别数量阈值的三维点坐标特征，确定所述三维点坐标特征中未存在离散点坐标特征。

13. 根据权利要求 9 至 12 中任一项所述的电子设备，其中，所述确定所述离散点坐标特征对应的离散异常数据之后，所述处理器执行所述计算机程序时还实现如下步骤：

使用主成分分析算法，对所述第一异常数据进行降维，获得降维数据；

对所述降维数据进行逆向计算，获得目标数据；

判断所述目标数据与所述第一异常数据的差值是否大于预设差值阈值；

若所述还原数据与所述第一异常数据的差值大于预设差值阈值，确定所述第一异常数据中存在第二异常数据；或

若所述还原数据与所述第一异常数据的差值小于或等于预设差值阈值，确定所述第一异常数据中未存在第二异常数据。

14. 根据权利要求 9 至 12 中任一项所述的电子设备，其中，所述提取所述第一异常数据的三维特征包括：

获取所述第一异常数据的差分值、当前值以及指数加权移动平均值；

计算所述当前值与所述指数加权移动平均值的目标差值；

将所述差分值、所述当前值以及所述目标差值确定为所述第一异常数据的三维点坐标特征。

15. 根据权利要求 9 至 12 中任一项所述的电子设备，其中，所述获取需要进行异常检测的设备指标数据之前，所述处理器执行所述计算机程序时还实现如下步骤：

获取用于训练的第一样本数据；

使用无监督的学习算法对所述第一样本数据进行筛选，获得第二样本数据；使用所述第二样本数据进行训练，获得训练好的异常检测模型。

16. 一种计算机可读存储介质，其中，所述计算机可读存储介质存储有至少一个指令，所述至少一个指令被处理器执行时实现如下步骤：

获取需要进行异常检测的设备指标数据，以及获取数据过滤区间；

根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据；

若所述第一过滤数据中存在第一异常数据，提取所述第一异常数据的三维点坐标特征；

若所述三维点坐标特征中存在离散点坐标特征，确定所述离散点坐标特征对应的离散

异常数据；

若所述第一异常数据中存在第二异常数据，使用训练好的异常数据检测模型，对所述第二异常数据和所述离散异常数据进行异常检测，获得检测结果，其中，所述异常检测模型训练时所使用的样本数据是经过无监督的学习算法筛选的数据。

17. 根据权利要求 16 所述的计算机可读存储介质，其中，所述获取数据过滤区间包括：获取用户设置的数据过滤区间；或

获取历史统计数据，并根据所述历史统计数据，确定数据过滤区间。

18. 根据权利要求 16 所述的计算机可读存储介质，其中，所述根据所述数据过滤区间，对所述设备指标数据进行过滤，获得第一过滤数据之后，所述至少一个指令被处理器执行时还实现如下步骤：

针对每个所述第一过滤数据，根据拉依达准则，判断所述第一过滤数据是否属于第一预设数据区间；

若所述第一过滤数据属于第一预设数据区间，根据高斯分布模型，判断所述第一过滤数据对应的密度估计值是否小于预设密度估计阈值；

若所述第一过滤数据对应的密度估计值大于或等于预设密度估计阈值，使用指数加权移动平均法，判断所述第一过滤数据对应的指数加权平均值是否属于第二预设数据区间；

若所述第一过滤数据对应的指数加权平均值不属于第二预设数据区间，确定所述第一过滤数据中存在第一异常数据；或

若所述第一过滤数据对应的指数加权平均值属于第二预设数据区间，确定所述第一过滤数据未存在第一异常数据。

19. 根据权利要求 16 所述的计算机可读存储介质，其中，所述提取所述第一异常数据的三维点坐标特征之后，所述至少一个指令被处理器执行时还实现如下步骤：

将所述第一异常数据的三维点坐标特征添加至坐标特征集合中；

使用基于密度的聚类算法，对所述坐标特征集合中的特征进行聚类分析，获得每个所述三维点坐标特征对应的至少一个类别；

根据每个所述三维点坐标特征对应的至少一个类别，统计每个所述三维点坐标特征对应的类别数量；

判断所有所述三维点坐标特征中是否存在类别数量小于预设类别数量阈值的三维点坐标特征；

若所述三维点坐标特征中存在类别数量大于或等于预设类别数量阈值的三维点坐标特征，将类别数量小于预设类别数量阈值的三维点坐标特征确定为离散点坐标特征；或

若所述三维点坐标特征中未存在类别数量小于预设类别数量阈值的三维点坐标特征，确定所述三维点坐标特征中未存在离散点坐标特征。

20. 根据权利要求 16 至 19 中任一项所述的计算机可读存储介质，其中，所述确定所述离散点坐标特征对应的离散异常数据之后，所述至少一个指令被处理器执行时还实现如下步骤：

使用主成分分析算法，对所述第一异常数据进行降维，获得降维数据；

对所述降维数据进行逆向计算，获得目标数据；

判断所述目标数据与所述第一异常数据的差值是否大于预设差值阈值；

若所述还原数据与所述第一异常数据的差值大于预设差值阈值，确定所述第一异常数据中存在第二异常数据；或

若所述还原数据与所述第一异常数据的差值小于或等于预设差值阈值，确定所述第一异常数据中未存在第二异常数据。

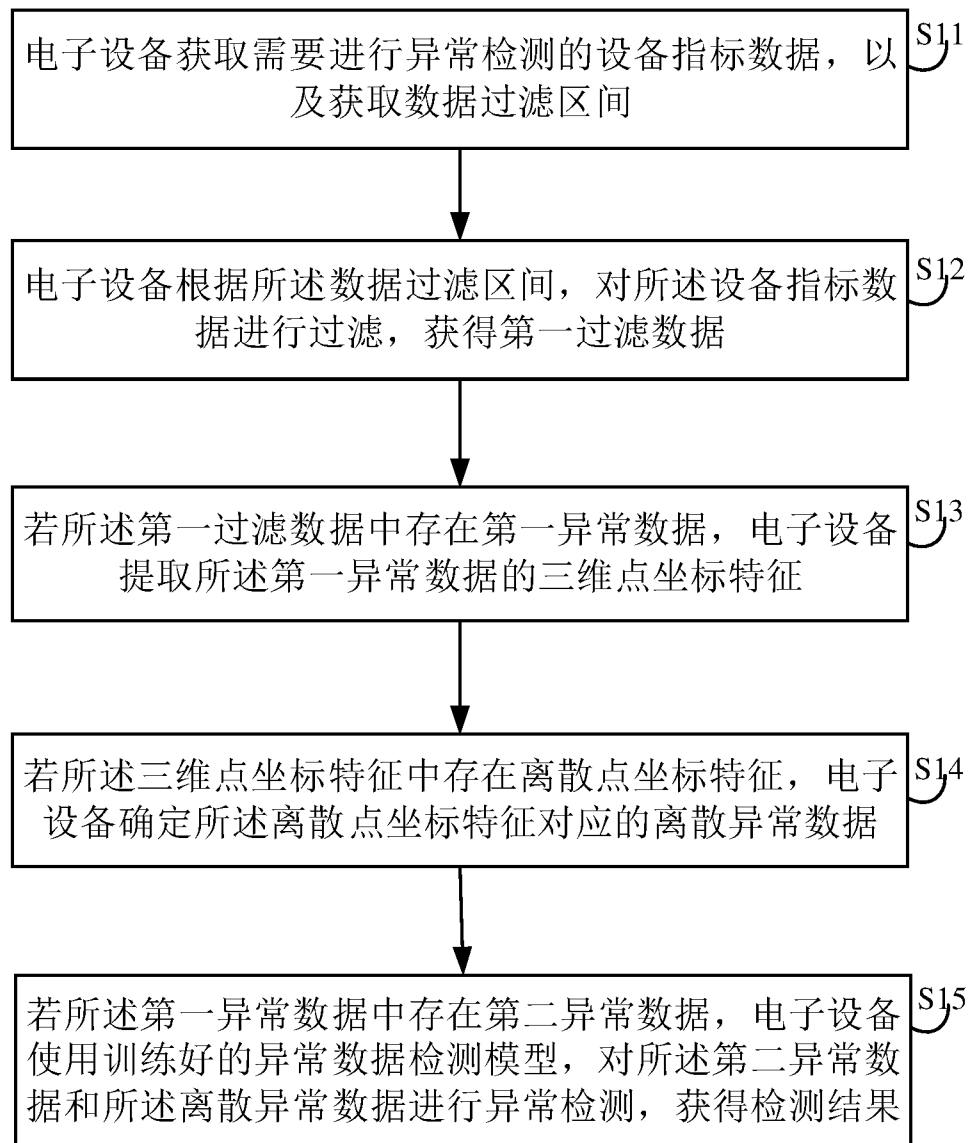


图 1

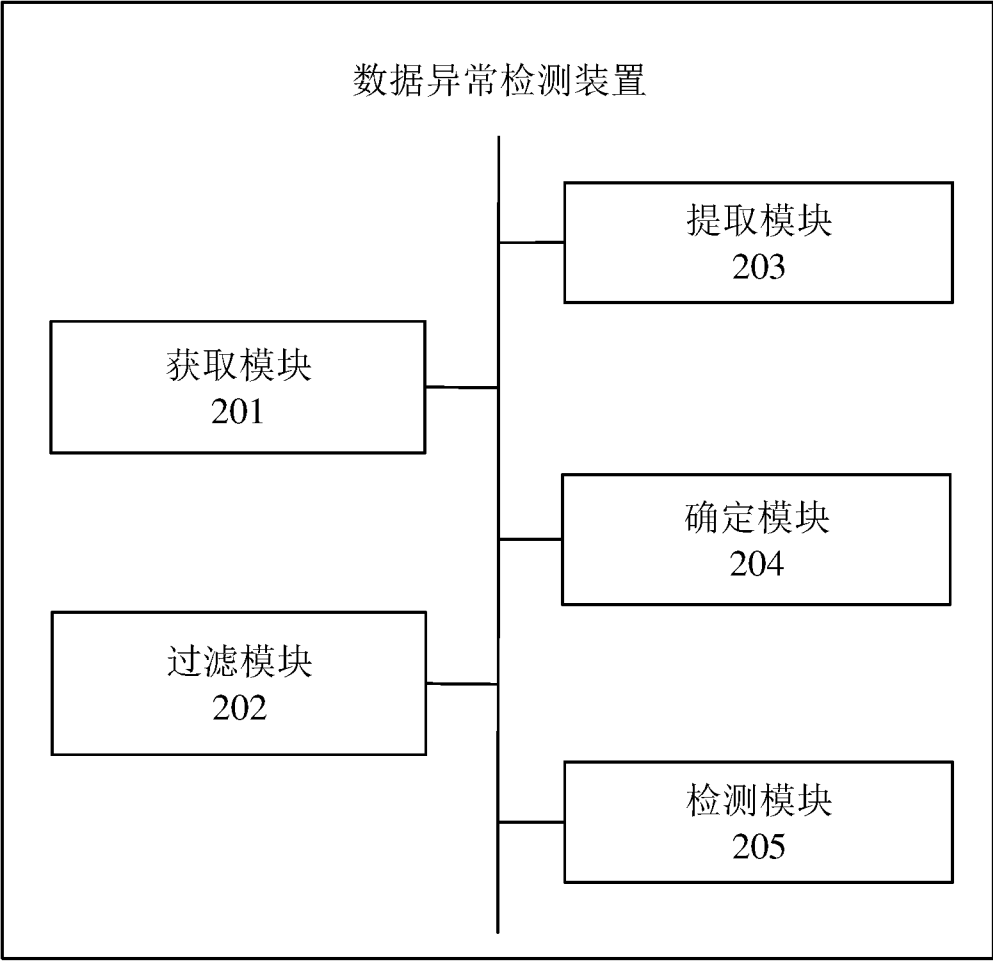


图 2

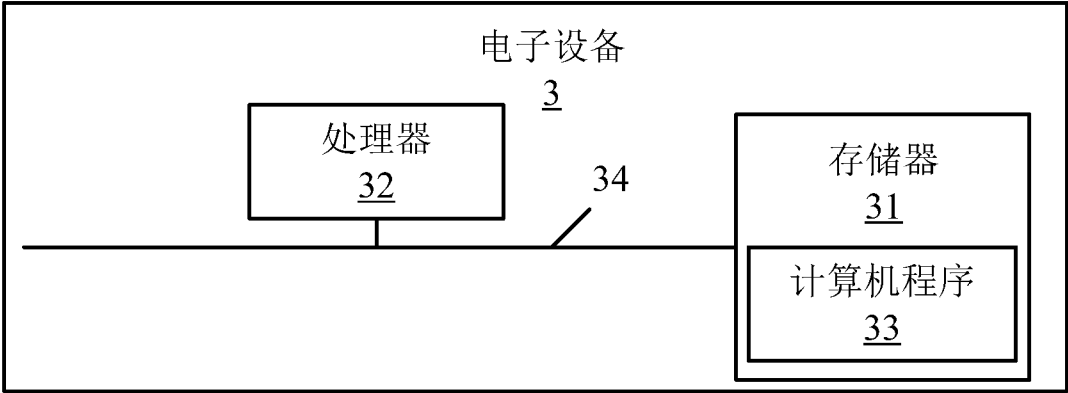


图 3

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/119060

A. CLASSIFICATION OF SUBJECT MATTER

G06F 11/30(2006.01)i; G06K 9/62(2006.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F; G06K; G06N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNPAT, EPODOC, WPI, CNKI, IEEE: 设备, 数据, 异常, 检测, 筛选, 过滤, 离散, 模型, 特征, 阈值, 类别, 降维, 二次检测, equipment, data, exception, detection, filtrate, disperse, model, character, threshold, sort, category, dimension, second

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111581046 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 25 August 2020 (2020-08-25) claims 1-10	1-20
A	CN 109032829 A (TENCENT TECHNOLOGY SHENZHEN CO., LTD.) 18 December 2018 (2018-12-18) description, paragraphs [0002]-[0048], and abstract	1-20
A	CN 107977301 A (NEUSOFT CORPORATION) 01 May 2018 (2018-05-01) entire document	1-20
A	US 2014379301 A1 (INFOSYS LIMITED) 25 December 2014 (2014-12-25) entire document	1-20
A	US 2012041575 A1 (HITACHI, LTD.) 16 February 2012 (2012-02-16) entire document	1-20



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search

17 December 2020

Date of mailing of the international search report

31 December 2020

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2020/119060

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111581046	A	25 August 2020	None			
CN	109032829	A	18 December 2018	None			
CN	107977301	A	01 May 2018	None			
US	2014379301	A1	25 December 2014	IN	201302720	I4	10 June 2016
US	2012041575	A1	16 February 2012	WO	2010095314	A1	26 August 2010
				JP	2010191556	A	02 September 2010
				CN	102282516	A	14 December 2011
				JP	2013218725	A	24 October 2013

国际检索报告

国际申请号

PCT/CN2020/119060

A. 主题的分类

G06F 11/30(2006.01)i; G06K 9/62(2006.01)i; G06N 3/04(2006.01)i; G06N 3/08(2006.01)i

按照国际专利分类(IPC)或者同时按照国家分类和IPC两种分类

B. 检索领域

检索的最低限度文献(标明分类系统和分类号)

G06F; G06K; G06N

包含在检索领域中的除最低限度文献以外的检索文献

在国际检索时查阅的电子数据库(数据库的名称, 和使用的检索词(如使用))

CNPAT, EPDOC, WPI, CNKI, IEEE: 设备, 数据, 异常, 检测, 筛选, 过滤, 离散, 模型, 特征, 阈值, 类别, 降维, 二次检测, equipment, data, exception, detection, filtrate, disperse, model, character, threshold, sort, category, dimension, second

C. 相关文件

类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 111581046 A (平安科技深圳有限公司) 2020年 8月 25日 (2020 - 08 - 25) 权利要求1-10	1-20
A	CN 109032829 A (腾讯科技深圳有限公司) 2018年 12月 18日 (2018 - 12 - 18) 说明书第[0002]-[0048]段, 摘要	1-20
A	CN 107977301 A (东软集团股份有限公司) 2018年 5月 1日 (2018 - 05 - 01) 全文	1-20
A	US 2014379301 A1 (INFOSYS LIMITED) 2014年 12月 25日 (2014 - 12 - 25) 全文	1-20
A	US 2012041575 A1 (HITACHI, LTD.) 2012年 2月 16日 (2012 - 02 - 16) 全文	1-20

☐ 其余文件在C栏的续页中列出。☒ 见同族专利附件。

* 引用文件的具体类型:

“A” 认为不特别相关的表示了现有技术一般状态的文件

“E” 在国际申请日的当天或之后公布的在先申请或专利

“L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件(如具体说明的)

“O” 涉及口头公开、使用、展览或其他方式公开的文件

“P” 公布日先于国际申请日但迟于所要求的优先权日的文件

“T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了理解发明之理论或原理的在后文件

“X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性

“Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性

“&” 同族专利的文件

国际检索实际完成的日期

2020年 12月 17日

国际检索报告邮寄日期

2020年 12月 31日

ISA/CN的名称和邮寄地址

中国国家知识产权局(ISA/CN)

中国北京市海淀区蓟门桥西土城路6号 100088

传真号 (86-10)62019451

授权官员

吴媛媛

电话号码 86-(10)-53961351

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2020/119060

检索报告引用的专利文件			公布日 (年/月/日)	同族专利			公布日 (年/月/日)
CN	111581046	A	2020年 8月 25日	无			
CN	109032829	A	2018年 12月 18日	无			
CN	107977301	A	2018年 5月 1日	无			
US	2014379301	A1	2014年 12月 25日	IN	201302720	I4	2016年 6月 10日
US	2012041575	A1	2012年 2月 16日	WO	2010095314	A1	2010年 8月 26日
				JP	2010191556	A	2010年 9月 2日
				CN	102282516	A	2011年 12月 14日
				JP	2013218725	A	2013年 10月 24日