



(12)发明专利申请

(10)申请公布号 CN 111476375 A

(43)申请公布日 2020.07.31

(21)申请号 202010237571.2

G06K 9/62(2006.01)

(22)申请日 2020.03.30

G06Q 50/30(2012.01)

H04W 8/18(2009.01)

(71)申请人 中移动信息技术有限公司

地址 101407 北京市昌平区未来科学城英才北三街16号院16号楼1006室

申请人 中移(杭州)信息技术有限公司  
中国移动通信集团有限公司

(72)发明人 王伟杰 张文 许鑫伶 朱君瑀  
李文倩 鲁银冰

(74)专利代理机构 上海晨皓知识产权代理事务所(普通合伙) 31260

代理人 成丽杰

(51)Int.Cl.

G06N 20/20(2019.01)

G06N 5/00(2006.01)

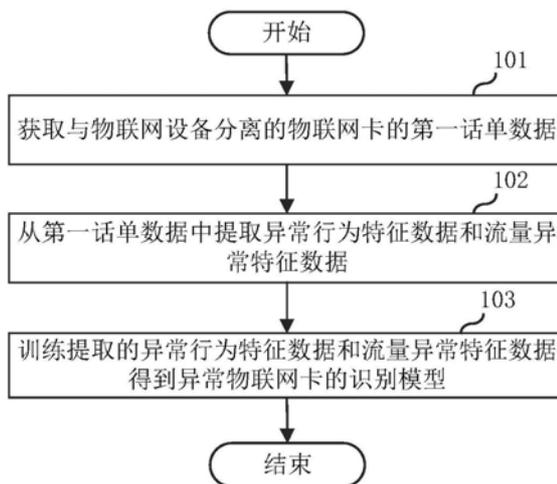
权利要求书2页 说明书8页 附图4页

(54)发明名称

一种确定识别模型的方法、装置、电子设备及存储介质

(57)摘要

本发明实施例涉及计算机技术领域,公开了一种确定识别模型的方法、装置、电子设备及存储介质。本发明中确定识别模型的方法,包括:获取与物联网设备分离的物联网卡的第一话单数据;从第一话单数据中提取异常行为特征数据和流量异常特征数据,异常行为特征数据为违规物联网卡的基础行为的特征数据,流量异常特征数据为违规物联网卡的流量变化的特征数据;训练提取的异常行为特征数据和流量异常特征数据,得到异常物联网卡的识别模型。通过采用本实施方式,使得确定的识别模型的准确率高,通过该识别模型可以快速、准确地识别异常物联网卡。



1. 一种确定识别模型的方法,其特征在于,包括:

获取与物联网设备分离的物联网卡的第一话单数据;

从所述第一话单数据中提取异常行为特征数据和流量异常特征数据,所述异常行为特征数据为违规物联网卡的基础行为的特征数据,所述流量异常特征数据为所述违规物联网卡的流量变化的特征数据;

训练提取的所述异常行为特征数据和所述流量异常特征数据,得到异常物联网卡的识别模型。

2. 根据权利要求1所述的确定识别模型的方法,其特征在于,所述异常物联网卡的识别模型用于识别被违规转售的物联网卡;

所述从所述第一话单数据中提取异常行为特征数据和流量异常特征数据,包括:

通过聚类算法将所述第一话单数据划分为N个数据簇,每个数据簇包括相同类别的特征数据;

从N个所述数据簇中删除属于正常行为类别的数据簇,得到异常行为类别的数据簇,将所有所述异常行为类别的数据簇中的特征数据作为异常行为特征数据;

从N个所述数据簇中提取流量异常的特征数据作为流量异常特征数据。

3. 根据权利要求2所述的确定识别模型的方法,其特征在于,所述流量异常特征数据包括:流量回归误差,和/或,流量异常产生日的个数;

所述从N个所述数据簇中提取流量异常的特征数据作为流量异常特征数据,包括:

针对每个数据簇中每个物联网卡的特征数据进行如下处理:

提取所述物联网卡在所述第一预设时间段内每天产生的流量,构建所述物联网卡产生流量的时间序列;

根据所述时间序列,确定所述物联网卡在所述第一预设时长的流量回归误差,所述流量回归误差为指示所述物联网卡的流量波动的特征数据;和/或;根据所述时间序列,确定所述物联网卡在所述第一预设时长内出现流量异常产生日的个数。

4. 根据权利要求3所述的确定识别模型的方法,其特征在于,所述根据所述时间序列,确定所述物联网卡在所述第一预设时长的流量回归误差,包括:

拟合所述时间序列,得到所述物联网卡在所述第一预设时间段内的流量回归误差。

5. 根据权利要求3所述的确定识别模型的方法,其特征在于,所述根据所述时间序列,确定所述物联网卡在所述第一预设时长内出现流量异常产生日的个数,包括:

计算所述时间序列对应的加权移动平均线;

根据所述加权移动平均线以及预设倍数,确定所述时间序列对应的振幅范围;

将超出所述振幅范围的日期个数作为所述物联网卡在所述第一预设时长内出现流量异常产生日的个数。

6. 根据权利要求1至5中任一项所述的确定识别模型的方法,其特征在于,所述获取与物联网设备分离的物联网卡的第一话单数据,包括:

获取第二话单数据,所述第二话单数据为在第二预设时间段内运行于预设运营商所属网段的所有物联网卡的话单数据;

从第二话单数据中分离出与物联网设备分离的物联网卡的话单数据,作为第一话单数据。

7. 根据权利要求6所述的确定识别模型的方法,其特征在于,所述从第二话单数据中分离出与物联网设备分离的物联网卡的话单数据,作为第一话单数据,包括:

将对应多个国际移动设备识别号IMEI的物联网卡的话单数据,存入第一话单数据库,将所述第一话单数据库中的话单数据作为所述第一话单数据;

和/或;

针对每个物联网卡进行如下处理:获取所述物联网卡当前对应的IMEI,判断获取的所述IMEI与所述物联网卡的激活IMEI是否相同,若不同,则将所述物联网卡的话单数据,存入第一话单数据库,将所述第一话单数据库中存储的话单数据作为所述第一话单数据。

8. 一种确定识别模型的装置,其特征在于,包括:获取模块、特征提取模块和训练模块;所述获取模块用于获取与物联网设备分离的物联网卡的第一话单数据;

所述特征提取模块用于从所述第一话单数据中提取异常行为特征数据和流量异常特征数据,所述异常行为特征数据为违规物联网卡的基础行为的特征数据,所述流量异常特征数据为所述违规物联网卡的流量变化的特征数据;

所述训练模块用于训练提取的所述异常行为特征数据和所述流量异常特征数据,得到异常物联网卡的识别模型。

9. 一种电子设备,其特征在于,包括:

至少一个处理器;以及,

与所述至少一个处理器通信连接的存储器;其中,

所述存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行如权利要求1-7任一所述的确定识别模型的方法。

10. 一种计算机可读存储介质,存储有计算机程序,其特征在于,所述计算机程序被处理器执行时实现权利要求1至7中任一项所述的确定识别模型的方法。

## 一种确定识别模型的方法、装置、电子设备及存储介质

### 技术领域

[0001] 本发明实施例涉及计算机技术领域,特别涉及一种确定识别模型的方法、装置、电子设备及存储介质。

### 背景技术

[0002] 随着物联网卡行业的迅速发展,物联网卡的安全问题也突显出来。工信部也明确要求移动通信转售企业进一步做好物联网行业卡安全管理,切实管好物联网行业卡。但是,仍旧有部分企业对物联网卡的管控不规范,存在违规销售的问题。

[0003] 发明人发现现有技术中至少存在如下问题:物联网卡行业的每日数据是数亿级别,极大的提高了管控的难度。目前通常通过识别模型识别出异常物联网卡,如被违规转售的物联网卡,但是目前的识别模型的识别准确率低。

### 发明内容

[0004] 本发明实施方式的目的在于提供一种确定识别模型的方法、装置、电子设备及存储介质,使得确定的识别模型的准确率高,通过该识别模型可以快速、准确地识别异常物联网卡。

[0005] 为解决上述技术问题,本发明的实施方式提供了一种确定识别模型的方法,包括:获取与物联网设备分离的物联网卡的第一话单数据;从第一话单数据中提取异常行为特征数据和流量异常特征数据,异常行为特征数据为违规物联网卡的基础行为的特征数据,流量异常特征数据为违规物联网卡的流量变化的特征数据;训练提取的异常行为特征数据和流量异常特征数据,得到异常物联网卡的识别模型。

[0006] 本发明的实施方式还提供了一种确定识别模型的装置,包括:获取模块、特征提取模块和训练模块;获取模块用于获取与物联网设备分离的物联网卡的第一话单数据;特征提取模块用于从第一话单数据中提取异常行为特征数据和流量异常特征数据,异常行为特征数据为违规物联网卡的基础行为的特征数据,流量异常特征数据为违规物联网卡的流量变化的特征数据;训练模块用于训练提取的异常行为特征数据和流量异常特征数据,得到异常物联网卡的识别模型。

[0007] 本发明的实施方式还提供了一种电子设备,包括:至少一个处理器;以及,与至少一个处理器通信连接的存储器;其中,存储器存储有可被所述至少一个处理器执行的指令,所述指令被所述至少一个处理器执行,以使所述至少一个处理器能够执行上述的确定识别模型的方法。

[0008] 本发明的实施方式还提供了一种计算机可读存储介质,存储有计算机程序,计算机程序被处理器执行时实现上述的确定识别模型的方法。

[0009] 本发明实施方式相对于现有技术而言,通过获取与物联网卡设备分离的物联网卡的第一话单数据,而不是采集所有的物联网卡的话单数据,缩小了用于训练的数据的数据量,同时与对应的物联网卡设备发生分离的物联网卡中出现被违规转售的物联网卡的概率

大,缩小了训练数据的范围,缩小了从第一话单数据中提取特征数据的难度,提高了提取异常行为特征数据和流量异常特征数据的速度;提取第一话单数据中的异常行为特征数据和流量异常特征数据,由于物联网卡通常消耗流量,被违规转售的物联网卡会出现流量异常,因而通过提取流量异常特征数据,丰富了训练识别模型的训练数据的特征,增加了用于训练的数据的准确性;从而提高了该异常物联网卡的识别模型的准确性。

[0010] 另外,异常物联网卡的识别模型用于识别被违规转售的物联网卡;从第一话单数据中提取异常行为特征数据和流量异常特征数据,包括:通过聚类算法将第一话单数据划分为N个数据簇,每个数据簇包括相同类别的特征数据;从N个数据簇中删除属于正常行为类别的数据簇,得到异常行为类别的数据簇,将所有异常行为类别的数据簇中的特征数据作为异常行为特征数据;从N个数据簇中提取流量异常的特征数据作为流量异常特征数据。通过聚类算法将第一话单数据划分为N个数据簇,由于可以分别从每个数据簇中提取流量异常特征数据,使得无需一次性处理大量的数据,从而提高了提取流量异常特征数据的速度,另外通过删除正常的行为的类别的数据簇,可以快速得到异常行为特征数据。

[0011] 另外,流量异常特征数据包括:流量回归误差,和/或,流量异常产生日的个数;从N个数据簇中提取流量异常的特征数据作为流量异常特征数据,包括:针对每个数据簇中每个物联网卡的特征数据进行如下处理:提取物联网卡在第一预设时间段内每天产生的流量,构建物联网卡产生流量的时间序列;根据时间序列,确定物联网卡在第一预设时长的流量回归误差,流量回归误差为指示物联网卡的流量波动的特征数据;和/或;根据时间序列,确定物联网卡在第一预设时长内出现流量异常产生日的个数。被违规转售的物联网卡,通常流量数据会发生异常,通过确定流量回归误差和/或流量异常产生日的个数,可以有效的确定出异常物联网卡,从而使得训练数据更加准确。

[0012] 另外,根据时间序列,确定物联网卡在第一预设时长的流量回归误差,包括:拟合时间序列,得到物联网卡在第一预设时间段内的流量回归误差。通过拟合时间序列的方式,可以快速且准确的得到流量回归误差。

[0013] 另外,根据时间序列,确定物联网卡在第一预设时长内出现流量异常产生日的个数,包括:计算时间序列对应的加权移动平均线;根据加权移动平均线以及预设倍数,确定时间序列对应的振幅范围;将超出振幅范围的日期个数作为物联网卡在第一预设时长内出现流量异常产生日的个数。通过加权移动平均线,可以快速确定出流量异常产生日的个数。

[0014] 另外,获取与物联网设备分离的物联网卡的第一话单数据,包括:获取第二话单数据,第二话单数据为在第二预设时间段内运行于预设运营商所属网段的所有物联网卡的话单数据;从第二话单数据中分离出与物联网设备分离的物联网卡的话单数据,作为第一话单数据。

[0015] 另外,从第二话单数据中分离出与物联网设备分离的物联网卡的话单数据,作为第一话单数据,包括:将对应多个国际移动设备识别号IMEI的物联网卡的话单数据,存入第一话单数据库,将第一话单数据库中的话单数据作为第一话单数据;和/或;针对每个物联网卡进行如下处理:获取物联网卡当前对应的IMEI,判断获取的IMEI与物联网卡的激活IMEI是否相同,若不同,则将物联网卡的话单数据,存入第一话单数据库,将第一话单数据库中存储的话单数据作为第一话单数据。提供多种方式确定第一话单数据,使得可以灵活提取第一话单数据。

## 附图说明

[0016] 一个或多个实施例通过与之对应的附图中的图片进行示例性说明,这些示例性说明并不构成对实施例的限定,附图中具有相同参考数字标号的元件表示为类似的元件,除非有特别申明,附图中的图不构成比例限制。

[0017] 图1是根据本发明第一实施方式提供的一种确定识别模型的方法的流程图;

[0018] 图2是根据本发明第一实施方式提供的一种提取异常行为特征数据和流量异常特征数据的示意图;

[0019] 图3是根据本发明第一实施方式提供的拟合时间序列的效果示意图;

[0020] 图4是根据本发明第一实施方式提供的一种确定物联网卡在第一预设时长内出现流量异常产生日的个数的示意图;

[0021] 图5是根据本发明第一实施方式提供的确定物联网卡在第一预设时长内出现流量异常产生日的个数的效果示意图;

[0022] 图6是根据本发明第二实施方式提供的获取与物联网设备分离的物联网卡的第一话单数据的一种实现的示意图;

[0023] 图7是根据本发明第三实施方式提供的一种确定识别模型的装置的具体结构示意图;

[0024] 图8是根据本发明第四实施方式提供的一种电子设备的具体结构示意图。

## 具体实施方式

[0025] 为使本发明实施例的目的、技术方案和优点更加清楚,下面将结合附图对本发明的各实施方式进行详细的阐述。然而,本领域的普通技术人员可以理解,在本发明各实施方式中,为了使读者更好地理解本申请而提出了许多技术细节。但是,即使没有这些技术细节和基于以下各实施方式的种种变化和修改,也可以实现本申请所要求保护的技术方案。

[0026] 以下各个实施例的划分是为了描述方便,不应对本发明的具体实现方式构成任何限定,各个实施例在不矛盾的前提下可以相互结合相互引用。

[0027] 发明人发现物联网卡的流量资费低,出现了许多违规转售物联网卡的问题,而目前通常采用识别模型识别出被违规转售的物联网卡。识别模型的准确性将决定是否准确识别出被违规转售的物联网卡,而识别模型的准确度依赖于模型训练,由于物联网卡的数据量庞大,从庞大的物联网卡数据中提取出与违规转售相关的特征数据非常困难,使得识别模型的训练困难;另外提取的特征数据与违规转售弱相关,也导致识别模型非常不准确。

[0028] 本发明的第一实施方式涉及一种确定识别模型的方法。该确定识别模型的方法可以应用于电子设备,如服务器、电脑等。该识别模型用于识别被违规转售的物联网卡,该确定识别模型的方法的具体流程如图1所示:

[0029] 步骤101:获取与物联网设备分离的物联网卡的第一话单数据。

[0030] 具体的说,通常正规售卖的物联网卡与指定类型的物联网设备对应,例如,正常销售的情况,物联网卡A与物联网设备B绑定,或者物联网卡A与C类别的物联设备关联。因此,在发生了机卡分离的物联网卡中更容易出现被违规转售的物联网卡,为了减小对庞大数据量的处理,可以获取与物联网设备分离的物联网的话单数据作为第一话单数据。

[0031] 需要说明的是,物联网卡通常用于上网,为了减少冗余数据,可以将与物联网设备分离的物联网卡的GPRS话单数据,作为第一话单数据。

[0032] 步骤102:从第一话单数据中提取异常行为特征数据和流量异常特征数据,异常行为特征数据为违规物联网卡的基础行为的特征数据,流量异常特征数据为违规物联网卡的流量变化的特征数据。

[0033] 在一个例子中,从第一话单数据中提取异常行为特征数据和流量异常特征数据可以采用如图2所示的子步骤。

[0034] 子步骤S11:通过聚类算法将第一话单数据划分为N个数据簇,每个数据簇包括相同类别的特征数据。

[0035] 具体的说,基于GPRS话单行为的话单数据中包括多种类别的特征数据,表1列举了基于该GPRS话单行为的数据中的14种类别的特征数据。

数学符号	特征	说明
F1	day_nums	当月发生流量使用的天数
F2	notiot_max_day_flow	非行业终端一天使用的最大流量(单位:G)
F3	iot_max_day_flow	行业终端一天使用的最大流量(单位:G)
F4	imei_change_th_avg	当月imei平均变化次数
F5	total_iot_terminal_avg	当月使用的行业终端平均个数
F6	total_notiot_terminal_avg	当月使用的非行业终端平均个数
F7	terminal_change_avg	当月终端类别平均变化次数
F8	max_flow_month	当月使用最大总流量(单位:G)
F9	avg_flow_month	当月使用平均流量(单位:G)
F10	cv	当月流量变异系数(标准差/均值)
F11	using_nums	当月流量使用总次数/60
F12	imei_nums	当月涉及IMEI个数
F13	province_nums	当月拜访省份个数
F14	time_nums	当月使用总时间(小时)

[0037] 表1

[0038] 由于第一话单数据种包含多个不同类别的特征数据,通过聚类算法可以将第一话单数据划分为不同类别的数据簇,聚类算法可以采用KMeans算法。下面具体介绍划分N个数据簇的过程:

[0039] 按照聚类算法,确定聚类类别N,N为大于1的整数;N的值可以是第一话单数据中的类别数量;例如,如表1所示,特征类别有14个,故N可以设置为14。从第一话单数据中随机选取N个点作为初始聚集的簇心(或称为聚类中心);计算第一话单数据种每个数据到各簇心的距离,确定对应的最小聚类中心;归并最近聚类中心相同的数据集,重新确定新的聚类中心;若聚类中心发生变化,重新计算第一话单数据中每个数据到新的簇心的距离,重新确定对应的最小聚类中心,归并最近与聚类中心相同的数据;若聚类中心趋近稳定不变,则结束聚类。

[0040] 值得一提的是,通过聚类算法,可以快速将第一话单数据划分为N个数据簇,便于后续步骤中快速根据数据簇的类别,选出与违规转售行为强相关的特征数据,以减少对庞大的第一话单数据,提高处理数据的效率。

[0041] 子步骤S12:从N个数据簇中删除属于正常行为类别的数据簇,得到异常行为类别的数据簇,将所有异常行为类别的数据簇中的特征数据作为异常行为特征数据。

[0042] 具体的说,针对每个数据簇进行如下处理:检测该数据簇的簇心的类别,若簇心的类别为正常行为类别,则确定该数据簇属于正常行为类别的数据簇;若簇心的类别为异常行为类别的数据簇,则确定该数据簇属于异常行为类别的数据簇。可以理解的是,可以预先设置正常行为类别的类型和异常行为类别的类型。

[0043] 值得一提的是,由于第一话单数据中的数据量大、特征维度多、数据不完备(存在大量的无效数据)的特点,通过删除正常行为类别的数据簇,可以快速得到异常行为类别的特征数据,缩小用于训练的数据量,同时由于减少了特征维度,也提高了后续训练识别模型的速度。

[0044] 子步骤S13:从N个数据簇中提取流量异常的特征数据作为流量异常特征数据。

[0045] 针对每个数据簇中每个物联网卡的特征数据进行如下处理:提取物联网卡在第一预设时间段内每天产生的流量,构建物联网卡产生流量的时间序列;根据时间序列,确定物联网卡在第一预设时长的流量回归误差,流量回归误差为指示物联网卡的流量波动的特征数据;和/或;根据时间序列,确定物联网卡在第一预设时长内出现流量异常产生日的个数。

[0046] 具体的说,第一预设时间段可以根据需要设置,例如,本实施方式中第一预设时间段可以为一个月的时长。时间序列的提取有多种方式,例如,可以提取每个物联网卡一个月内每天消耗的总流量,按照当月的日期构建出一组基于每日的日期的时间序列;也可以按照每月构建时间序列,本实施方式中,为了提高训练的准确性,基于每日的日期构建一组时间序列。

[0047] 在一个例子中,确定物联网卡在第一预设时长的流量回归误差的过程可以为:拟合时间序列,得到物联网卡在第一预设时间段内的流量回归误差。

[0048] 具体的说,拟合时间序列的方式可以采用最小二乘法,通过最小二乘法拟合时间序列后,得到该物联网卡当月对应的回归误差,记作Ed\_1r;将Ed\_1r作为该物联网卡在当月的流量回归误差。例如,对物联网卡A在一个月内的时间序列进行最小二乘法的拟合,拟合效果如图3所示,实线表示为流量回归误差,小点表示该物联网卡的一日的流量数据,横坐标为日期,纵坐标为流量数。。

[0049] 在一个例子中,确定物联网卡在第一预设时长内出现流量异常产生日的个数的方式可以采用如图4所示的子步骤。

[0050] 子步骤S21:计算时间序列对应的加权移动平均线。

[0051] 具体的说,本实施方式中基于加权移动平均(weighted moving average,简称“WMA”)算法统计出现流量异常产生日的个数。该WMA的表达式如公式(1)所示:

$$[0052] \quad WMA_m = \frac{n * p_m + (n-1) * p_{m-1} + \dots + 2 * p_{m-n+2} + 1 * p_{m-n+1}}{n + (n-1) + \dots + 1} \quad \text{公式(1);}$$

[0053] 其中,WMA<sub>m</sub>表示加权移动平均值,n日加权移动平均的最近一个数乘以n,次近期的数乘以n-1;以此类推,直至n为0;P<sub>m</sub>可以表示第n日的流量。例如,本实施方式中,可以按照公式(1)依次计算当天的前4个日期的平均值,并乘以不同的权重,构建滑动曲线,该滑动曲线即为该加权移动平均线。

[0054] 子步骤S22:根据加权移动平均线以及预设倍数,确定时间序列对应的振幅范围。

[0055] 具体的说,预设倍数可以根据实际需要进行设置,本实施方式中预设倍数可以设置为1倍。由于可以通过均值计算出对应的方差,因而根据该加权移动平均线,可以确定出与该加权移动平均线对应的均值线,将对应的均值线乘以预设倍数,得到该时间序列所对应的振幅范围。

[0056] 子步骤S23:将超出振幅范围的日期个数作为物联网卡在第一预设时长内出现流量异常产生日的个数。

[0057] 具体的说,统计超出振幅范围的日期个数;将该超出振幅范围的日期个数作为物联网卡在第一预设时长内出现流量异常产生日的个数。如图5所示,图5中虚线表示物联网卡产生流量的时间序列,实线表示振幅范围,从图5中可以看出,物联网卡在第一预设时长内出现流量异常产生日的个数为2,图5中横坐标为日期,纵坐标为流量值。

[0058] 步骤103:训练提取的异常行为特征数据和流量异常特征数据,得到异常物联网卡的识别模型。

[0059] 具体的说,在训练识别模型过程中,将提取的异常行为特征数据和流量异常特征数据作为训练数据,采用机器学习算法进行训练,得到异常物联网卡的识别模型。该机器学习方法可以采用极限梯度提升迭代决策树(Extreme Gradient Boosting Decision Tree,简称“XGBoost”)算法;得到的识别模型即可用于识别被违规转售的物联网卡。

[0060] 表2列举了采用本实施方式训练得到的识别模型识别被违规转售的物联网卡的准确率。

[0061]

模型\效果	准确率	时间(s)
Xgboost的识别模型	87.6%	3670
聚类后训练得到的识别模型	90.4%	296
本实施方式得到的识别模型	98.5%	313

[0062] 表2

[0063] XGboost的识别模型是直接对采集的话单数据进行训练后得到,聚类后训练得到的识别模型是对采集的话单数据进行聚类后进行训练得到的模型。

[0064] 本发明实施方式相对于现有技术而言,通过获取与物联网卡设备分离的物联网卡的第一话单数据,而不是采集所有的物联网卡的话单数据,缩小了用于训练的数据的数据量,同时与对应的物联网卡设备发生分离的物联网卡中出现被违规转售的物联网卡的概率大,缩小了训练数据的范围,缩小了从第一话单数据中提取特征数据的难度,提高了提取异常行为特征数据和流量异常特征数据的速度;提取第一话单数据中的异常行为特征数据和流量异常特征数据,由于物联网卡通常消耗流量,被违规转售的物联网卡会出现流量异常,因而通过提取流量异常特征数据,丰富了训练识别模型的训练数据的特征,增加了用于训练的数据的准确性;从而提高了该异常物联网卡的识别模型的准确性。

[0065] 本发明的第二实施方式涉及一种确定识别模型的方法。第二实施方式是对第一实施方式中的步骤101的详细说明,步骤101获取与物联网设备分离的物联网卡的第一话单数据可以采用如图6所示的步骤:

[0066] 步骤201:获取第二话单数据,第二话单数据为在第二预设时间段内运行于预设运营商所属网段的所有物联网卡的话单数据。

[0067] 具体的说,为了保证训练数据的准确性,获取运行于同一个运行商所属网段内所

有物联网卡的话单数据。第二预设时间段可以根据需要设置,例如,本实施方式中可以设置为3个月。物联网卡通常用于上网,为了减少冗余数据,可以仅采集物联网卡的GPRS话单数据。

[0068] 步骤202:从第二话单数据中分离出与物联网设备分离的物联网卡的话单数据,作为第一话单数据。

[0069] 在一个例子中,将对应多个国际移动设备识别号IMEI的物联网卡的话单数据,存入第一话单数据库,将第一话单数据库中的话单数据作为第一话单数据。

[0070] 具体的说,移动设备都有唯一的IMEI;每一个物联网卡应当与一个IMEI对应,若检测出一个物联网卡与多个IMEI对应,将与多个IMEI对应的物联网卡的话单数据存储第一话单数据库内;将第一话单数据库中的话单数据作为第一话单数据。

[0071] 在一个例子中,针对每个物联网卡进行如下处理:获取物联网卡当前对应的IMEI,判断获取的IMEI与物联网卡的激活IMEI是否相同,若不同,则将物联网卡的话单数据,存入第一话单数据库,将第一话单数据库中存储的话单数据作为所述第一话单数据。

[0072] 在另一个例子中,可以将对应多个国际移动设备识别号IMEI的物联网卡的话单数据,存入第一话单数据库。并针对每个物联网卡进行如下处理:以及将获取所述物联网卡当前对应的IMEI,判断获取的IMEI与物联网卡的激活IMEI是否相同,若不同,则将物联网卡的话单数据,存入第一话单数据库;将第一话单数据库中存储的话单数据作为第一话单数据。

[0073] 本实施方式提供的确定识别模型的方法,通过提供多种方式确定出第一话单数据,使得可以灵活提取第一话单数据。

[0074] 上面各种方法的步骤划分,只是为了描述清楚,实现时可以合并为一个步骤或者对某些步骤进行拆分,分解为多个步骤,只要包括相同的逻辑关系,都在本专利的保护范围内;对算法中或者流程中添加无关紧要的修改或者引入无关紧要的设计,但不改变其算法和流程的核心设计都在该专利的保护范围内。

[0075] 本发明第三实施方式涉及一种确定识别模型的装置,该确定识别模型的装置结构如图7所示,包括:获取模块301、特征提取模块302和训练模块302;获取模块301用于获取与物联网设备分离的物联网卡的第一话单数据;特征提取模块302用于从第一话单数据中提取异常行为特征数据和流量异常特征数据,异常行为特征数据为违规物联网卡的基础行为的特征数据,流量异常特征数据为违规物联网卡的流量变化的特征数据;训练模块301用于训练提取的异常行为特征数据和流量异常特征数据,得到异常物联网卡的识别模型。

[0076] 不难发现,本实施方式为与第一实施方式相对应的虚拟装置的实施例,本实施方式可与第一实施方式互相配合实施。第一实施方式中提到的相关技术细节在本实施方式中依然有效,为了减少重复,这里不再赘述。相应地,本实施方式中提到的相关技术细节也可应用在第一实施方式中。

[0077] 值得一提的是,本实施方式中所涉及到的各模块均为逻辑模块,在实际应用中,一个逻辑单元可以是一个物理单元,也可以是一个物理单元的一部分,还可以以多个物理单元的组合实现。此外,为了突出本发明的创新部分,本实施方式中并没有将与解决本发明所提出的技术问题关系不太密切的单元引入,但这并不表明本实施方式中不存在其它的单元。

[0078] 本发明第四实施方式涉及一种电子设备,电子设备40如图8所示,包括:至少一个

处理器401;以及,与至少一个处理器401通信连接的存储器402;其中,存储器402存储有可被至少一个处理器401执行的指令,指令被至少一个处理器401执行,以使至少一个处理器401能够执行第一实施方式或第二实施方式中的确定识别模型的方法。

[0079] 其中,存储器402和处理器401采用总线方式连接,总线可以包括任意数量的互联的总线和桥,总线将一个或多个处理器401和存储器402的各种电路链接在一起。总线还可以将诸如外围设备、稳压器和功率管理电路等之类的各种其他电路链接在一起,这些都是本领域所公知的,因此,本文不再对其进行进一步描述。总线接口在总线和收发机之间提供接口。收发机可以是一个元件,也可以是多个元件,比如多个接收器和发送器,提供用于在传输介质上与各种其他装置通信的单元。经处理器401处理的数据通过天线在无线介质上进行传输,进一步,天线还接收数据并将数据传送给处理器401。

[0080] 处理器401负责管理总线和通常的处理,还可以提供各种功能,包括定时,外围接口,电压调节、电源管理以及其他控制功能。而存储器可以被用于存储处理器在执行操作时所使用的数据。

[0081] 本发明第四实施方式涉及一种计算机可读存储介质,存储有计算机程序,计算机程序被处理器执行时实现第一实施方式或第二实施方式中的确定识别模型的方法。

[0082] 本领域技术人员可以理解实现上述实施例方法中的全部或部分步骤是可以通过程序来指令相关的硬件来完成,该程序存储在一个存储介质中,包括若干指令用以使得一个设备(可以是单片机,芯片等)或处理器(processor)执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储介质包括:U盘、移动硬盘、只读存储器(ROM,Read-Only Memory)、随机存取存储器(RAM,Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质。

[0083] 本领域的普通技术人员可以理解,上述各实施方式是实现本发明的具体实施例,而在实际应用中,可以在形式上和细节上对其作各种改变,而不偏离本发明的精神和范围。

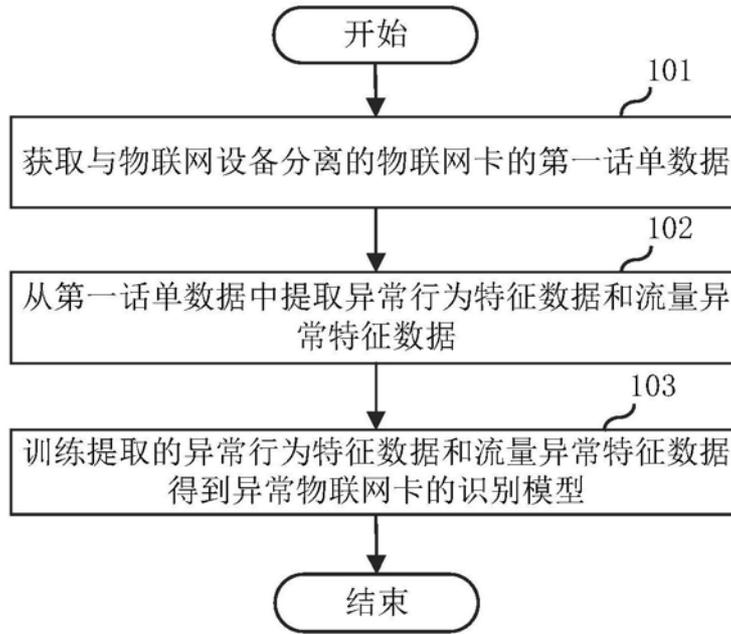


图1

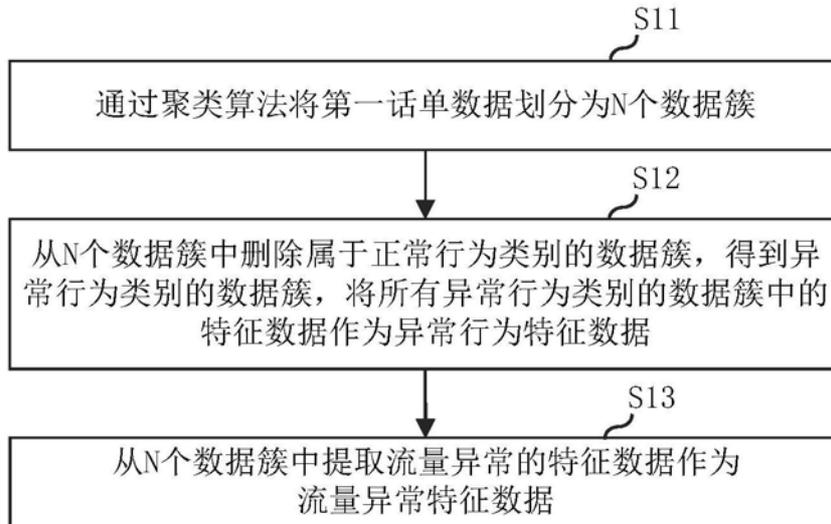


图2

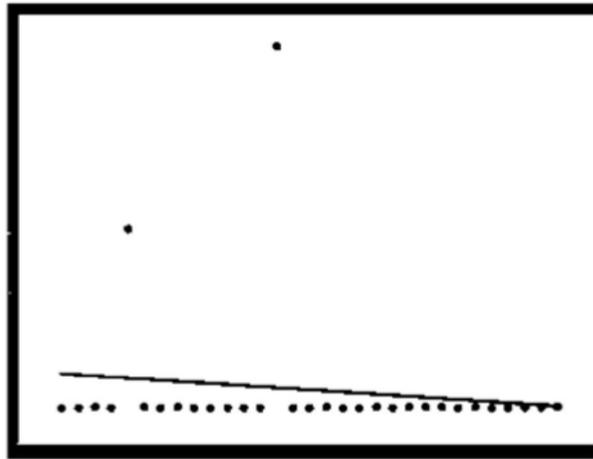


图3

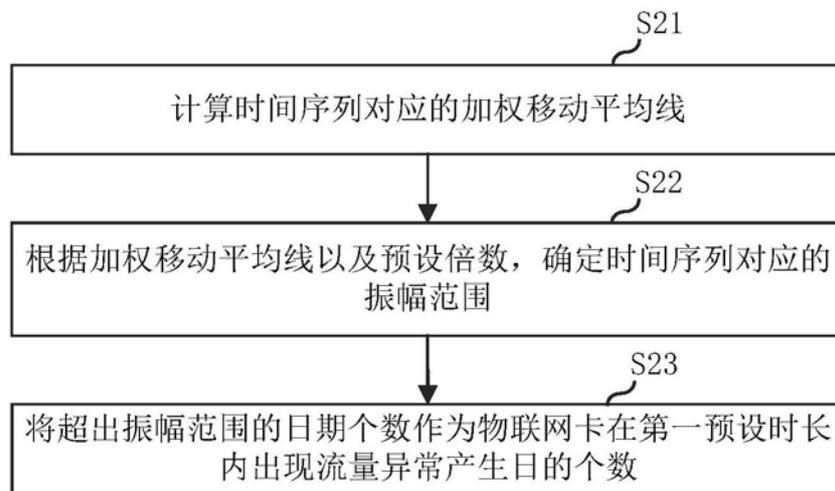


图4

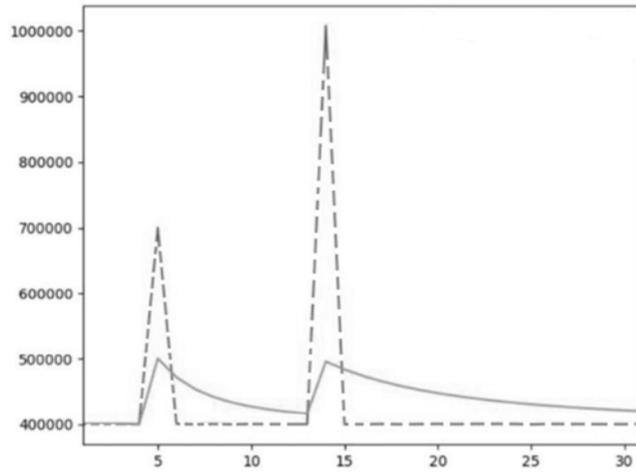


图5

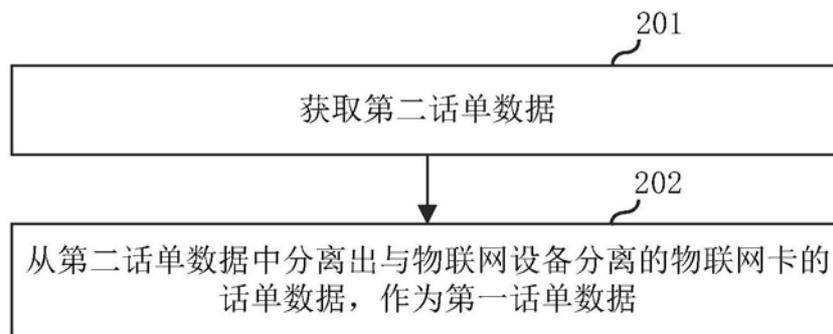


图6

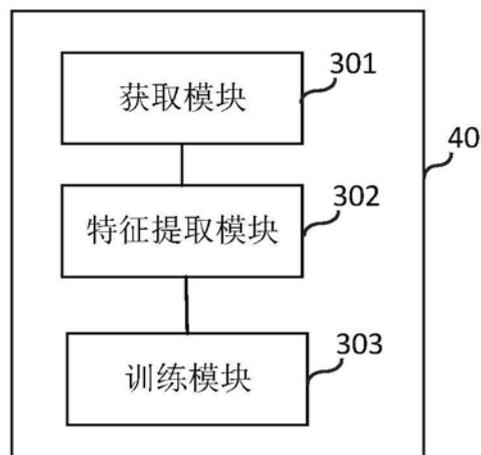


图7

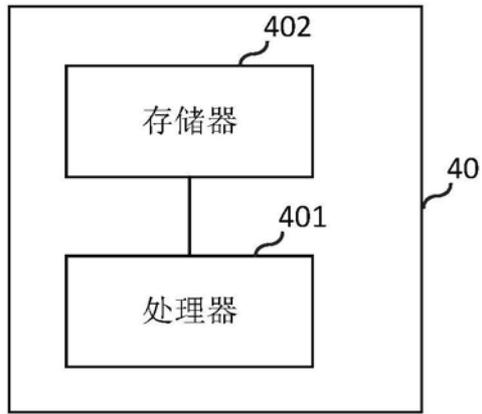


图8