(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 110247910 B
(45)授权公告日 2022.08.09

(21)申请号 201910511621.9

(22)申请日 2019.06.13

(65)同一申请的已公布的文献号
申请公布号 CN 110247910 A

(43)申请公布日 2019.09.17

(73)专利权人 深信服科技股份有限公司
地址 518055 广东省深圳市南山区学苑大道1001号南山智园A1栋一层

(72)发明人 陈晓帆 吴东磊

(74)专利代理机构 深圳市深佳知识产权代理事务所(普通合伙) 44285
专利代理师 王仲凯

(51)Int.Cl.
H04L 9/40(2022.01)
G06N 20/00(2019.01)

(56)对比文件
US 2018097822 A1,2018.04.05
CN 108234500 A,2018.06.29
CN 102263790 A,2011.11.30
CN 109714324 A,2019.05.03
CN 109347872 A,2019.02.15
CN 107846392 A,2018.03.27
CN 109829543 A,2019.05.31
CN 108093406 A,2018.05.29
CN 108023876 A,2018.05.11
CN 102291392 A,2011.12.21
CN 108959566 A,2018.12.07
CN 109861988 A,2019.06.07
CN 107766418 A,2018.03.06

审查员 吴超

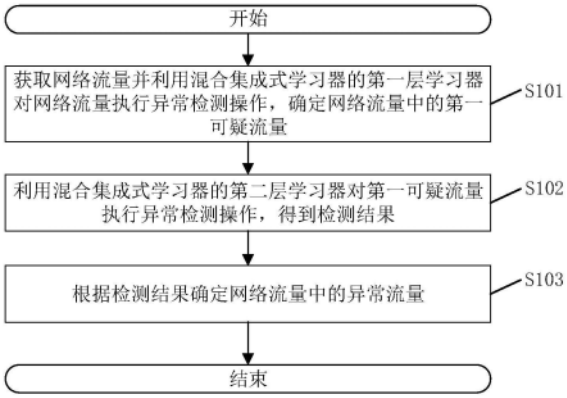
权利要求书2页 说明书10页 附图6页

(54)发明名称

一种异常流量的检测方法、系统及相关组件

(57)摘要

本申请公开了一种异常流量的检测方法,所述检测方法包括获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作,确定网络流量中的第一可疑流量;其中,第一层学习器为Stacking集成学习器;利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作,得到检测结果;根据检测结果确定网络流量中的异常流量。本申请能够提高异常流量检测的准确度,避免出现误判、漏检的状况。本申请还公开了一种异常流量的检测系统、一种计算机可读存储介质及一种电子设备,具有以上有益效果。



CN 110247910 B

1.一种异常流量的检测方法,其特征在于,包括:

获取网络流量并利用混合集成式学习器的第一层学习器对所述网络流量执行异常检测操作,确定所述网络流量中的第一可疑流量;其中,所述第一层学习器为Stacking集成学习器;所述Stacking集成学习器的元学习器为基于无监督学习算法的学习器;

将所述第一可疑流量输入级联集成学习器,并利用所述级联集成学习器的每一层子学习器依次对所述第一可疑流量执行异常监测操作,得到检测结果;其中,所述级联集成学习器由多层子学习器级联得到,每一层子学习器的输出结果为下一层子学习器的输入数据;所述级联集成学习器为所述混合集成式学习器的第二层学习器;

根据所述检测结果确定所述网络流量中的异常流量。

2.根据权利要求1所述检测方法,其特征在于,当所述级联集成学习器为2层子学习器级联得到的学习器时,利用所述级联集成学习器的每一层子学习器依次对所述第一可疑流量执行异常检测操作包括:

利用所述级联集成学习器的第一层子学习器对所有所述第一可疑流量执行异常检测操作,得到第二可疑流量;

利用所述级联集成学习器的第二层子学习器对所有所述第二可疑流量执行异常检测操作;

其中,所述第一层子学习器和所述第二层子学习器为不同类型的子学习器。

3.根据权利要求2所述检测方法,其特征在于,所述第一层子学习器为Local Outlier Factor学习器,所述第二层子学习器为K-means学习器。

4.根据权利要求1至3任一项所述检测方法,其特征在于,所述利用混合集成式学习器的第一层学习器对所述网络流量执行异常检测操作包括:

将所述网络流量输入至所述Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果;

拼接所有所述预训练结果得到特征矩阵;

将所述特征矩阵输入所述Stacking集成学习器的元学习器,以便所述元学习器对所述特征矩阵进行异常流量检测操作。

5.一种异常流量的检测系统,其特征在于,包括:

第一检测模块,用于获取网络流量并利用混合集成式学习器的第一层学习器对所述网络流量执行异常检测操作,确定所述网络流量中的第一可疑流量;其中,所述第一层学习器为Stacking集成学习器;所述Stacking集成学习器的元学习器为基于无监督学习算法的学习器;

第二检测模块,用于将所述第一可疑流量输入级联集成学习器,并利用所述级联集成学习器的每一层子学习器依次对所述第一可疑流量执行异常监测操作,得到检测结果;其中,所述级联集成学习器由多层子学习器级联得到,每一层子学习器的输出结果为下一层子学习器的输入数据;所述级联集成学习器为所述混合集成式学习器的第二层学习器;

异常流量确定模块,用于根据所述检测结果确定所述网络流量中的异常流量。

6.根据权利要求5所述检测系统,其特征在于,当所述级联集成学习器为2层子学习器级联得到的学习器时,所述第二检测模块包括:

第一子处理单元,用于利用所述级联集成学习器的第一层子学习器对所有所述第一可

疑流量执行异常检测操作,得到第二可疑流量;

第二子处理单元,用于利用所述级联集成学习器的第二层子学习器对所有所述第二可疑流量执行异常检测操作;

其中,所述第一层子学习器和所述第二层子学习器为不同类型的子学习器。

7.根据权利要求6所述检测系统,其特征在于,所述第一层子学习器为Local Outlier Factor学习器,所述第二层子学习器为K-means学习器。

8.根据权利要求5至7任一项所述检测系统,其特征在于,所述第一检测模块包括:

基学习器执行单元,用于获取网络流量并将所述网络流量输入至所述Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果;

结果拼接单元,用于拼接所有所述预训练结果得到特征矩阵;

元学习器执行单元,用于将所述特征矩阵输入所述Stacking集成学习器的元学习器,以便所述元学习器对所述特征矩阵进行异常流量检测操作。

9.一种计算机可读存储介质,其特征在于,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权利要求1至4任一项所述异常流量的检测方法的步骤。

10.一种电子设备,其特征在于,包括:

存储器,用于存储计算机程序;

处理器,用于执行所述计算机程序时实现如权利要求1至4任一项所述异常流量的检测方法的步骤。

一种异常流量的检测方法、系统及相关组件

技术领域

[0001] 本发明涉及网络安全技术领域，特别涉及一种异常流量的检测方法、系统、一种计算机可读存储介质及一种电子设备。

背景技术

[0002] 目前主流的流量异常检测算法是基于机器学习和深度学习的理论实现的，常见的流量异常检测算法如SVM，BP神经网络，循环神经网络对于带有标注的数据的异常检测任务已经取得了令人满意的效果。然而，很多时候运维人员所能获取的原始流量数据并未被人工标注，特别是对于连续性强，数据量大的网络流量数据，若要进行人工标注需要花费大量的人力物力，故当待检测的数据为无标注的流量数据时，因此这类监督学习算法将不再适用。

[0003] 相关技术中，往往通过单一的无监督异常流量检测算法实现，例如使用One Class SVM算法、Isolation Forest算法、One Class SVM算法、Elliptic Envelope算法等。但是，如上述相关技术中仅单独的使用某一种特定算法对已有的无标签数据进行异常检测往往存在着难以避免的误判、漏检等状况。

[0004] 因此，如何提高异常流量检测的准确度，避免出现误判、漏检的状况是本领域技术人员目前需要解决的技术问题。

发明内容

[0005] 本申请的目的是提供一种异常流量的检测方法、系统、一种计算机可读存储介质及一种电子设备，能够提高异常流量检测的准确度，避免出现误判、漏检的状况。

[0006] 为解决上述技术问题，本申请提供一种异常流量的检测方法，该检测方法包括：

[0007] 获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作，确定网络流量中的第一可疑流量；其中，第一层学习器为Stacking集成学习器；

[0008] 利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作，得到检测结果；

[0009] 根据检测结果确定网络流量中的异常流量。

[0010] 可选的，当第二层学习器为级联集成学习器时，利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作包括：

[0011] 将第一可疑流量输入级联集成学习器，并利用级联集成学习器的每一层子学习器依次对第一可疑流量执行异常监测操作；

[0012] 其中，级联集成学习器由多层子学习器级联得到，每一层子学习器的输出结果为下一层子学习器的输入数据。

[0013] 可选的，当级联集成学习器为2层子学习器级联得到的学习器时，利用级联集成学习器的每一层子学习器依次对第一可疑流量执行异常检测操作包括：

[0014] 利用级联集成学习器的第一层子学习器对所有第一可疑流量执行异常检测操作，

得到第二可疑流量；

[0015] 利用级联集成学习器的第二层子学习器对所有第二可疑流量执行异常检测操作；

[0016] 其中，第一层子学习器和第二层子学习器为不同类型的子学习器。

[0017] 可选的，第一层子学习器为Local Outlier Factor学习器，第二层子学习器为K-means学习器。

[0018] 可选的，利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作包括：

[0019] 将网络流量输入至Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果；

[0020] 拼接所有预训练结果得到特征矩阵；

[0021] 将特征矩阵输入Stacking集成学习器的元学习器，以便元学习器对特征矩阵进行异常流量检测操作。

[0022] 可选的，元学习器为基于无监督学习算法的学习器。

[0023] 本申请还提供了一种异常流量的检测系统，该检测系统包括：

[0024] 第一检测模块，用于获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作，确定网络流量中的第一可疑流量；其中，第一层学习器为Stacking集成学习器；

[0025] 第二检测模块，用于利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作，得到检测结果；

[0026] 异常流量确定模块，用于根据检测结果确定网络流量中的异常流量。

[0027] 可选的，当第二层学习器为级联集成学习器时，第二检测模块具体为用于将第一可疑流量输入级联集成学习器，并利用级联集成学习器的每一层子学习器依次对第一可疑流量执行异常监测操作的模块；

[0028] 其中，级联集成学习器由多层子学习器级联得到，每一层子学习器的输出结果为下一层子学习器的输入数据。

[0029] 可选的，当级联集成学习器为2层子学习器级联得到的学习器时，第二检测模块包括：

[0030] 第一子处理单元，用于利用级联集成学习器的第一层子学习器对所有第一可疑流量执行异常检测操作，得到第二可疑流量；

[0031] 第二子处理单元，用于利用级联集成学习器的第二层子学习器对所有第二可疑流量执行异常检测操作；

[0032] 其中，第一层子学习器和第二层子学习器为不同类型的子学习器。

[0033] 可选的，第一层子学习器为Local Outlier Factor学习器，第二层子学习器为K-means学习器。

[0034] 可选的，第一检测模块包括：

[0035] 基学习器执行单元，用于获取网络流量并将网络流量输入至Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果；

[0036] 结果拼接单元，用于拼接所有预训练结果得到特征矩阵；

[0037] 元学习器执行单元，用于将特征矩阵输入Stacking集成学习器的元学习器，以便

元学习器对特征矩阵进行异常流量检测操作。

[0038] 可选的，元学习器为基于无监督学习算法的学习器。

[0039] 本申请还提供了一种计算机可读存储介质，其上存储有计算机程序，计算机程序执行时实现上述异常流量的检测方法执行的步骤。

[0040] 本申请还提供了一种电子设备，包括存储器和处理器，存储器中存储有计算机程序，处理器调用存储器中的计算机程序时实现上述异常流量的检测方法执行的步骤。

[0041] 本申请提供了一种异常流量的检测方法，包括获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作，确定网络流量中的第一可疑流量；其中，第一层学习器为Stacking集成学习器；利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作，得到检测结果；根据检测结果确定网络流量中的异常流量。

[0042] 本申请利用Stacking集成学习器和第二层学习器构建混合集成式学习器，先通过混合集成式学习器的第一层学习器，即Stacking集成学习器，对异常流量进行初步筛选得到第一可疑流量，再利用第二层学习器以第一可疑流量作为输入数据再次执行异常检测操作得到检测结果，进而确定异常流量。由于Stacking集成学习器中可以包括多个基学习器和一个元学习器，对模型的表达能力有较大提升，能够减小检测过程中欠拟合问题的发生，即降低漏检率。通过混合集成式学习器的第二层学习器对Stacking集成学习器得输出结果执行进一步的过滤操作，可以使混合集成式学习器具有较好的泛化性能减少过拟合问题的发生，即降低误判率。因此先后经过第一层学习器和第二层学习器能够明显降低异常流量的误检率和漏检率，由此可知本申请能够提高异常流量检测的准确度，避免出现误判、漏检的状况。本申请同时还提供了一种异常流量的检测系统、一种计算机可读存储介质和一种电子设备，具有上述有益效果，在此不再赘述。

附图说明

[0043] 为了更清楚地说明本申请实施例，下面将对实施例中所需要使用的附图做简单的介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

[0044] 图1为本申请实施例所提供的一种异常流量的检测方法的流程图；

[0045] 图2为Stacking集成学习器的异常数据检测结果示意图；

[0046] 图3为级联集成学习器的异常数据检测结果示意图；

[0047] 图4为本申请实施例所提供的一种二级级联集成学习器的异常流量检测方法的流程图；

[0048] 图5为本申请实施例所提供的一种Stacking集成学习器的异常流量检测方法的流程图；

[0049] 图6为本申请实施例提供的一种混合集成式异常流量检测学习器的检测算法示意图；

[0050] 图7为本申请实施例所提供的一种异常流量的检测系统的结构示意图。

具体实施方式

[0051] 为使本申请实施例的目的、技术方案和优点更加清楚，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。

[0052] 流量数据的异常检测是保证网络信息安全的重要环节，通过对流量数据的异常检测，网络维护人员能够及时发现在特定时间特定网络节点的异常现象，从而及时进行异常的分析和故障的排查。因此数据流量的异常检测是网络维护的关键，异常流量检测算法的有效性及准确率备受关注。不准确的流量检测结果将会对网络的维护带来严重的后果，例如异常误判，漏检等不准确的结果会在后续的维护过程中带来更多人力、物力方面的消耗。相关技术中通常采用无监督异常检测算法实现对异常流量的检测，例如基于统计学的异常数据检测算法、基于聚类的异常点检测算法和特定异常点检测算法等，但是上述相关技术中的检测算法只是单独的使用某一种特定算法对已有的无标签数据进行异常检测，由于在特定的任务中不同算法的表现力存在不同程度的差别，因此存在漏检、误检的情况。基于上述相关技术中的种种缺陷，本申请通过以下几个实施例提供新的异常流量检测方式，能够提高异常流量检测的准确度，避免出现误判、漏检的状况。

[0053] 下面请参见图1，图1为本申请实施例所提供的一种异常流量的检测方法的流程图。

[0054] 具体步骤可以包括：

[0055] S101：获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作，确定网络流量中的第一可疑流量；

[0056] 其中，本步骤首先获取网络流量，即原始网络流量，此处不限定网络流量是否为经过人工标注的网络流量，该网络流量可以为连续性强、数据量大的未经标注的流量数据。本步骤不限定网络流量的来源，该网络流量可以为网络系统中任意一个或几个网络设备的网络流量，该网络设备可以包括交换机、路由器、等保一体机和防火墙等。对于本实施例中获取网络流量的过程可以通过多种方式实现，可以对目标网络设备的网络流量进行实时获取实时监控，也可以按照预设周期进行获取进而对周期内的所有网络流量进行检测，还可以是对目标网络设备的网络流量进行抽样检测，本实施例不限定网络流量的具体获取策略，本领域的技术人员可以根据实际应用场景进行灵活选择。

[0057] 本实施例中提到的混合集成式学习器的第一层学习器为Stacking集成学习器，混合集成式学习器可以包括第一层学习器和第二层学习器，第二层学习器的输入数据为第一层学习器的输出结果。在获取了网络流量的基础上，本步骤利用Stacking集成学习器对网络流量进行了初步的检测操作。Stacking集成学习器指基于Stacking集成学习策略的流量检测模型，Stacking集成学习器将学习过程分为两子层，其思想是在第二子层训练一个元学习器(Meta-leaner)来对第一子层中各个基学习器(Base-leaner)的学习的结果进一步学习，从而提高检测准确率。基于Stacking集成学习器对于模型的表达能力有较大的提升，可以减小欠拟合问题的发生，该方法能有效降低漏判出现的情况发生。Stacking集成学习器的具体方法可以是使用k-fold把待学习的数据分成不相交的k个部分，首先用第一层中每个基学习器对k-1部分进行训练，对剩下的那部分进行预测，直到对所有k个部分都完成预测，预测的结果便作为这个基学习器在第一子层的输出，并且对每个基学习器都迭代的重复此步骤，直到所有学习器都把原始数据集预测完毕。然后将第一子层各基学习器的输

出拼接(Stacking)成一个特征矩阵,作为第二子层元学习器的输入,最后通过第二子层的元学习器输出的预测结果,进而根据预测结果确定网络流量中的第一可疑流量。k-fold(k折交叉验证)是一种能够有效防止训练过拟合的训练策略,该算法通过把原始数据集划分成不相交的N等分,每次取其中N-1份进行模型训练,对剩下的一部分进行预测,直到把所有N等分都预测完,作为最终数据集的学习结果。

[0058] S102:利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作,得到检测结果;

[0059] 其中,本步骤建立在已经将网络流量输入至Stacking集成学习器并经过Stacking集成学习器的检测确定第一可疑流量的基础上,由于仅利用Stacking集成学习器进行异常流量检测可能存在一定的误判情况,因此为了提升异常网络流量的检测准确率,本实施例在得到第一可疑流量后利用混合集成式学习器的第二层学习器进行了进一步的检测操作,第二层学习器包括N个子学习器,N为任意正整数。

[0060] 作为一种可行的实施方式,混合集成式学习器的第二层学习器可以为级联集成学习器,即基于Cascade集成学习策略的异常流量检测模型。通过级联的策略把前层学习到的预测结果再次使用学习器进行进一步学习,筛选出最有可能为异常的数据点作为输出。Cascade集成学习策略采用的是数据过滤的思想,先将原始数据送入到第一层子学习器进行学习,再将那些结果为负的数据取出再次采用第二层子学习器进行学习,最终得到准确度更高的学习结果。作为一种可行的实施方式,当本实施例中的网络流量为无标签的数据时,级联集成学习器中每一层子学习器均可以为基于无监督学习算法的学习器。相对于相关技术中只采用单个学习器的预测模型,基于Cascade集成学习策略有更强的非线性表述能力,能够降低泛化误差并降低误检测概率。可以理解的是,级联集成学习器可以包括由多层子学习器级联得到,上一层子学习器的输出结果为下一层子学习器的输入数据,本实施例不限定级联集成学习器中包括的子学习器的数量和种类,本领域技术人员可以根据实际应用场景选择适当数量和种类的子学习器。具体的,级联集成学习器可以由多个相同种类的子学习器级联构建得到,也可以由多个种类不同的子学习器级联构建得到。

[0061] S103:根据检测结果确定网络流量中的异常流量。

[0062] 其中,利用混合集成式学习器的第一层学习器和第二层学习器执行异常检测操作相当于对网络流量进行聚类,根据得到的检测结果(即聚类结果)可以将网络流量划分为异常流量和正常流量。在确定网络流量中的异常流量之后,本实施例还可以上报该异常流量,以便对异常流量进行相关的分析,进而维护网络信息安全。

[0063] 请参见图2,图2为Stacking集成学习器的异常数据检测结果示意图,利用上述实际应用中的实施例提到的Stacking集成学习器可以对漏检情况得到了很好的抑制,但是仍存在着少量的误判,即将那些非异常点判断成了异常点的情况(如图2中线框所示)。请参见图3,图3为级联集成学习器的异常数据检测结果示意图,基于Cascade的异常检测算法可以看出,误判情况已经完全消除,但存在着少量的漏检问题发生(如图3中线框所示)。由于实际应用中数据量往往十分巨大,因此误判和漏检的数量也会相应大幅度增加。根据以上分析,可见基于Stacking集成学习器的集成策略对于模型的表达能力有较大的提升,它从某种角度上等同于减小了欠拟合问题的发生。而基于Cascade的集成策略对模型的泛化性能,它从某种角度上等同于减小了过拟合问题的发生。因此本实施例将两种集成策略相结合,

设计了一种混合集成式异常检测算法来进一步优化异常检测的结果。将Stacking集成学习器的模型输出的结果作为混合集成式模型的前层输入，将第一层的输出中那些判为异常的数据点(负样本)提取出来作为第二层聚类算法的输入。使用K-means聚类算法作为第二层模型并再次将输入数据聚为两类，选取其中簇较大的类别中的样本点作为最终的异常数据点。通过本实施例的方案，最终的异常检测结果无论是从误判还是漏检方面都有了进一步的提升。

[0064] 本实施例利用Stacking集成学习器和第二层学习器构建混合集成式学习器，先通过混合集成式学习器的第一层学习器，即Stacking集成学习器，对异常流量进行初步筛选得到第一可疑流量，再利用第二层学习器以第一可疑流量作为输入数据再次执行异常检测操作得到检测结果，进而确定异常流量。由于Stacking集成学习器中可以包括多个基学习器和一个元学习器，对模型的表达能力有较大提升，能够减小检测过程中欠拟合问题的发生，即降低漏检率。通过混合集成式学习器的第二层学习器对Stacking集成学习器得输出结果执行进一步的过滤操作，可以使混合集成式学习器具有较好的泛化性能减少过拟合问题的发生，即降低误判率。因此先后经过第一层学习器和第二层学习器能够明显降低异常流量的误检率和漏检率，由此可知本实施例能够提高异常流量检测的准确度，避免出现误判、漏检的状况。

[0065] 作为对于图1对应的实施例的进一步补充，S102中利用第二层学习器对第一可疑流量执行异常检测操作的过程可以具体为：将第一可疑流量输入级联集成学习器，并利用级联集成学习器的每一层子学习器依次对第一可疑流量执行异常监测操作；其中，级联集成学习器由多层子学习器级联得到，每一层子学习器的输出结果为下一层子学习器的输入数据。

[0066] 图1对应的实施例可以不对第二层学习器中的子学习器数量进行限定，但是级联集成学习器中级联的层级越多模型复杂度越高，并且层级过多时又会存在另一个问题：即最终剩下判定为异常数据点的个数会越来越少。因此作为一种优选的实施方式，可以根据实际应用中的具体问题进行实验，然后根据实验效果来选择算法的层数以及每一层的具体的算法。通过综合考量漏检率、误判率以及模型复杂度多种影响因素的基础上，采用二层级联为较好的选择。下面请参见图4，图4为本申请实施例所提供的一种二级级联集成学习器的异常流量检测方法的流程图；本实施例是对图1对应实施例中S102的进一步描述，可以将本实施例与图1对应的实施例相结合得到更为优选的实施方式，本实施例的具体步骤可以包括：

[0067] S201：利用级联集成学习器的第一层子学习器对所有第一可疑流量执行异常检测操作，得到第二可疑流量；

[0068] S202：利用级联集成学习器的第二层子学习器对所有第二可疑流量执行异常检测操作；

[0069] 由于异常检测算法本质上是先对输入数据进行聚类，再通过聚类的结果来分析哪些点属于正常点，哪些点属于异常点。但是向学习器中输入的原始数据(即图1对应的实施例中的网络流量)复杂多变，例如这些原始数据中往往存在着一些与正常数据点的值差异较大的异常数据点，或者原始数据点中包含着多个数值大小不同，但都属于正常数据的簇。若使用K-means学习器这类对异常值(噪声)比较敏感且受到簇分布差别影响较大的聚类算

法作为第一层的检测算法，那么第一层中输出的结果中很可能就是含有较多误判数据点，如果再将这些点作为下一层的输入将会对后层算法的结果造成不良影响。因此本实施例可以使用对于包含不同密度簇以及噪声的数据鲁棒性更高的LOF(Local Outlier Factor局部异常因子)算法用于对原始的数据进行第一层的检测，这从某种意义上相当于先使用一层不那么敏感的过滤器先对原始数据进行过滤，然后再对过滤后的数据采用类似K-means这种速度快，且能够有较好的聚类性能的算法进行进一步校准。

[0070] 总之，对于二级级联集成学习器来说，第一层子学习器应选择对输入数据的分布不均匀、噪声等问题有较强的鲁棒性的学习器。第二层子学习器可以为计算复杂度低且聚类性能良好的学习器。第一层子学习器和第二层子学习器可参照以上原则根据具体情况来选择实验结果较好的算法。

[0071] 图4对应的实施例中说提到的第一层子学习器和第二层子学习器可以为不同类型的子学习器，不同类型的子学习的检测特性能够进行优势互补，提高检测的准确率。作为一种可行的实施方式，第一层子学习器可以为Local Outlier Factor学习器，第二层子学习器可以为K-means学习器。具体的，首先将原始数据(相当于图1对应的实施例中的第一可疑流量)输入第一层Local Outlier Factor学习器。其次，将第一层学习器所分出来为可能存在异常的数据点(负样本)输入到第二层K-means学习器中再次学习。最后使用K-means学习器将输入聚为两类，其中数值较大的那一类作为异常检测结果输出。

[0072] 下面请参见图5，图5为本申请实施例所提供的一种Stacking集成学习器的异常流量检测方法的流程图；本实施例是对图1对应的实施例中S101的进一步描述，可以将本实施例与图1对应的实施例相结合得到更为优选的实施方式，本实施例的步骤可以包括：

[0073] S301：将网络流量输入至Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果；

[0074] S302：拼接所有预训练结果得到特征矩阵；

[0075] S303：将特征矩阵输入Stacking集成学习器的元学习器，以便元学习器对特征矩阵进行异常流量检测操作。

[0076] 其中，Stacking集成学习器中可以包括两类学习器，即基学习器和元学习器，相关技术中Stacking集成学习策略只应用于监督学习算法中，并不涉及无监督学习算法。当本实施例中的Stacking集成学习器的元学习器为基于无监督学习算法的学习器时，Stacking集成学习策略即可以应用于无监督学习中。具体的，本领域的相关技术中Stacking集成学习器的元学习器(Meta

[0077] Learner)通常采用逻辑回归或者KNN来实现投票机制，使用线性回归实现平均机制。但是上述提到的相关技术中Stacking集成学习策略采用的算法依然是有监督学习算法，同样不适用于无标签数据的场景。本实施例可以将无监督学习算法(如Isolation Forest)应用于元学习器中并取得了较好效果。

[0078] 下面通过实际应用中的实施例说明Stacking集成学习器的算法框架，

[0079] Stacking集成学习器的基学习器可以分别设置为Isolation Forest，Elliptic Envelop，DBSCAN，Gaussian，Local Outlier Factor，K-means这六种不同的算法，Stacking集成学习器的元学习器设置为Isolation Forest。可以将上述Stacking集成学习器与K-means学习器构建集成式异常流量检测学习器请参见图6，图6为本申请实施例提供的一种

混合集成式异常流量检测学习器的检测算法示意图。

[0080] 孤立森林(Isolation Forest)是一种基于树的异常检测算法,它是适用于连续数据的无监督异常检测方法,常用语挖掘异常数据,如网络安全中的流量异常检测和攻击检测等。

[0081] 椭圆包络(Elliptic Envelope)是一种基于统计分布的异常检测算法,该算法的策略是假设正常的数据是来自一个已知的高斯分布。根据这个假设,会尝试定义一个数据的“形状”,然后那些距离这个形状足够远的数据点就可以认为是离群点。

[0082] DBSCAN(Density-Based Spatial Clustering of Applications with Noise)是一种基于密度的空间聚类算法。该算法将具有足够密度的区域划分为簇,并在具有噪声的空间数据库中发现任意形状的簇,它通过将簇定义为密度相连的点的最大集合来对数据进行聚类,从而进行异常分析。

[0083] 高斯分布检测(Gaussian)是一种基于统计学概念的异常检测算法,在假设正常数据服从高斯分布的前提下,对数据建立高斯分布模型,从而用该模型估计待测样本属于非异常样本的可能性。

[0084] 局部离群因子检测方法(Local Outlier Factor):Local Outlier Factor是基于密度的离群点检测方法中的经典算法。该算法会给数据集中的每个点计算一个离群因子,通过判断该点的利群因子是否接近于1来判定是否是异常点。若离远大于1,则认为是异常点,接近于1,则是正常点。

[0085] K均值聚类(K-means)是一种经典的基于划分的聚类算法,该算法以空间中k个点为形心进行聚类,对最靠近他们的对象归类。通过迭代的方法,逐次更新各簇的形心的值,直至得到最好的聚类结果。

[0086] 请参见图7,图7为本申请实施例所提供的一种异常流量的检测系统的结构示意图;

[0087] 该系统可以包括:

[0088] 第一检测模块100,用于获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作,确定网络流量中的第一可疑流量;其中,第一层学习器为Stacking集成学习器;

[0089] 第二检测模块200,用于利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作,得到检测结果;

[0090] 异常流量确定模块300,用于根据检测结果确定网络流量中的异常流量。

[0091] 本实施例利用Stacking集成学习器和第二层学习器构建混合集成式学习器,先通过混合集成式学习器的第一层学习器,即Stacking集成学习器,对异常流量进行初步筛选得到第一可疑流量,再利用第二层学习器以第一可疑流量作为输入数据再次执行异常检测操作得到检测结果,进而确定异常流量。由于Stacking集成学习器中可以包括多个基学习器和一个元学习器,对模型的表达能力有较大提升,能够减小检测过程中欠拟合问题的发生,即降低漏检率。通过混合集成式学习器的第二层学习器对Stacking集成学习器得输出结果执行进一步的过滤操作,可以使混合集成式学习器具有较好的泛化性能减少过拟合问题的发生,即降低误判率。因此先后经过第一层学习器和第二层学习器能够明显降低异常流量的误检率和漏检率,由此可知本实施例能够提高异常流量检测的准确度,避免出现误

判、漏检的状况。

[0092] 进一步的，当第二层学习器为级联集成学习器时，第二检测模块200具体为用于将第一可疑流量输入级联集成学习器，并利用级联集成学习器的每一层子学习器依次对第一可疑流量执行异常监测操作的模块；

[0093] 其中，级联集成学习器由多层子学习器级联得到，每一层子学习器的输出结果为下一层子学习器的输入数据。

[0094] 进一步的，当级联集成学习器为2层子学习器级联得到的学习器时，第二检测模块200包括：

[0095] 第一子处理单元，用于利用级联集成学习器的第一层子学习器对所有第一可疑流量执行异常检测操作，得到第二可疑流量；

[0096] 第二子处理单元，用于利用级联集成学习器的第二层子学习器对所有第二可疑流量执行异常检测操作；

[0097] 其中，第一层子学习器和第二层子学习器为不同类型的子学习器。

[0098] 进一步的，第一层子学习器为Local Outlier Factor学习器，第二层子学习器为K-means学习器。

[0099] 进一步的，第一检测模块100包括：

[0100] 基学习器执行单元，用于获取网络流量并将网络流量输入至Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果；

[0101] 结果拼接单元，用于拼接所有预训练结果得到特征矩阵；

[0102] 元学习器执行单元，用于将特征矩阵输入Stacking集成学习器的元学习器，以便元学习器对特征矩阵进行异常流量检测操作。

[0103] 进一步的，元学习器为基于无监督学习算法的学习器。

[0104] 由于系统部分的实施例与方法部分的实施例相互对应，因此系统部分的实施例请参见方法部分的实施例的描述，这里暂不赘述。

[0105] 本申请还提供了一种计算机可读存储介质，其上存有计算机程序，该计算机程序被执行时可以实现上述实施例所提供的步骤。该存储介质可以包括：U盘、移动硬盘、只读存储器(Read-Only Memory，ROM)、随机存取存储器(Random Access Memory，RAM)、磁碟或者光盘等各种可以存储程序代码的介质。

[0106] 本申请还提供了一种电子设备，可以包括存储器和处理器，所述存储器中存有计算机程序，所述处理器调用所述存储器中的计算机程序时，可以实现上述实施例所提供的步骤。当然所述电子设备还可以包括各种网络接口，电源等组件。

[0107] 说明书中各个实施例采用递进的方式描述，每个实施例重点说明的都是与其他实施例的不同之处，各个实施例之间相同相似部分互相参见即可。对于实施例公开的系统而言，由于其与实施例公开的方法相对应，所以描述的比较简单，相关之处参见方法部分说明即可。应当指出，对于本技术领域的普通技术人员来说，在不脱离本申请原理的前提下，还可以对本申请进行若干改进和修饰，这些改进和修饰也落入本申请权利要求的保护范围内。

[0108] 还需要说明的是，在本说明书中，诸如第一和第二等之类的关系术语仅仅用来将一个实体或者操作与另一个实体或操作区分开来，而不一定要求或者暗示这些实体或操作

之间存在任何这种实际的关系或者顺序。而且，术语“包括”、“包含”或者其任何其他变体意在涵盖非排他性的包含，从而使得包括一系列要素的过程、方法、物品或者设备不仅包括那些要素，而且还包括没有明确列出的其他要素，或者是还包括为这种过程、方法、物品或者设备所固有的要素。在没有更多限制的状况下，由语句“包括一个……”限定的要素，并不排除在包括所述要素的过程、方法、物品或者设备中还存在另外的相同要素。

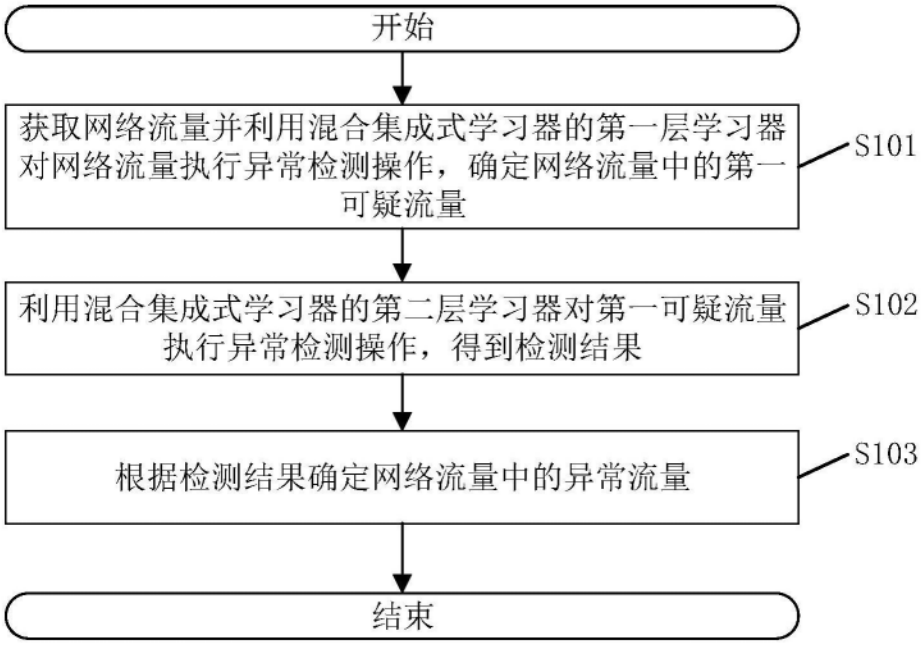

开始
获取网络流量并利用混合集成式学习器的第一层学习器对网络流量执行异常检测操作，确定网络流量中的第一可疑流量
S101
利用混合集成式学习器的第二层学习器对第一可疑流量执行异常检测操作，得到检测结果
S102
根据检测结果确定网络流量中的异常流量
S103
结束

图1

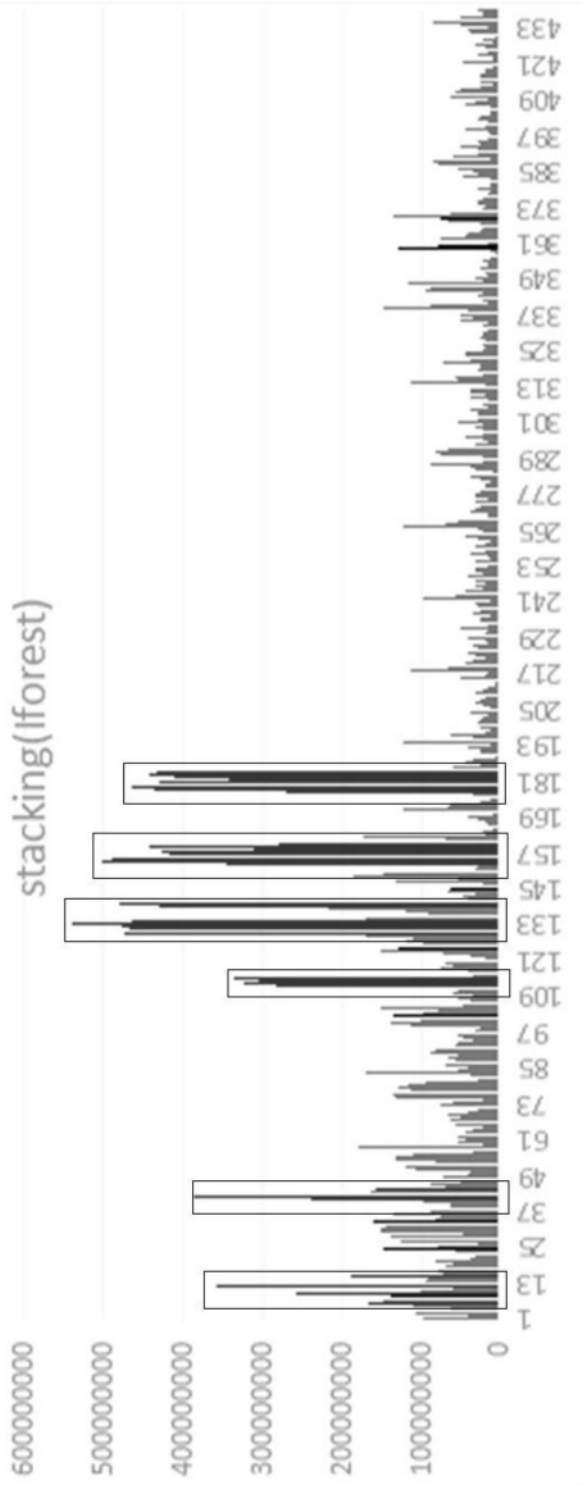

stacking(Iforest)
600000000
500000000
400000000
300000000
200000000
100000000
0
1
13
25
37
49
61
73
85
97
109
121
133
145
157
169
181
193
205
217
229
241
253
265
277
289
301
313
325
337
349
361
373
385
397
409
421
433

图2

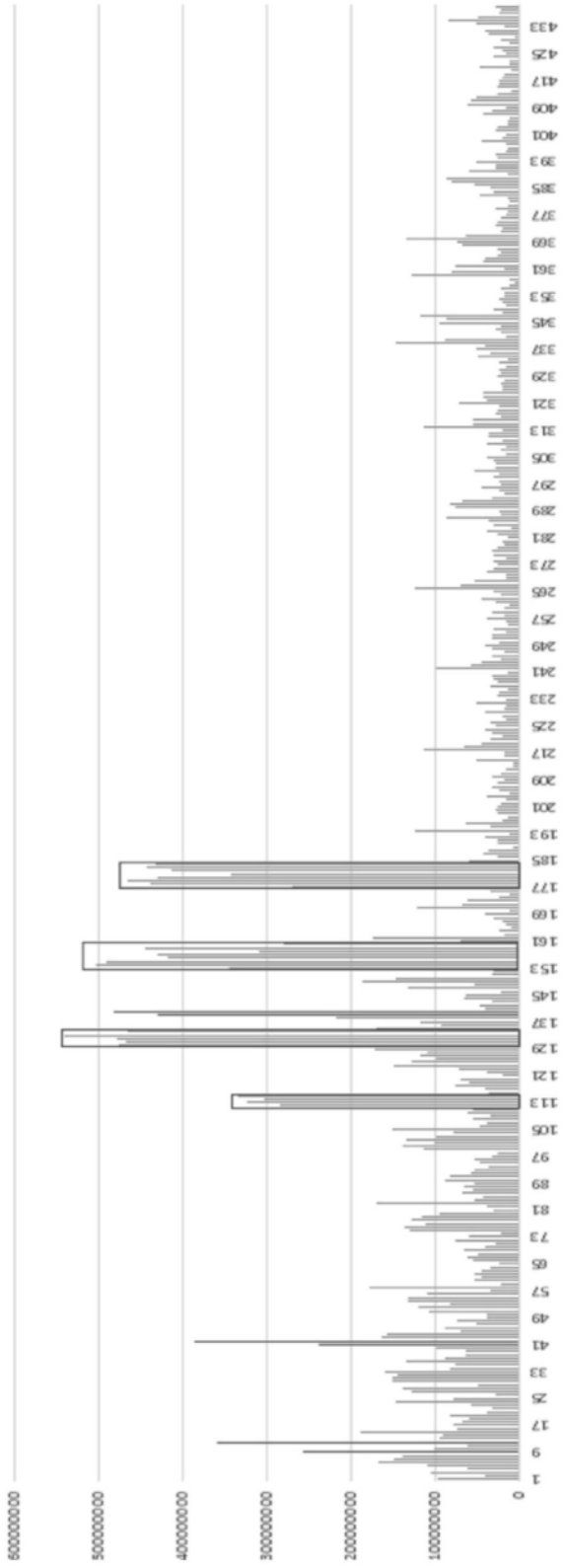

600000000
500000000
400000000
300000000
200000000
100000000
0

图3

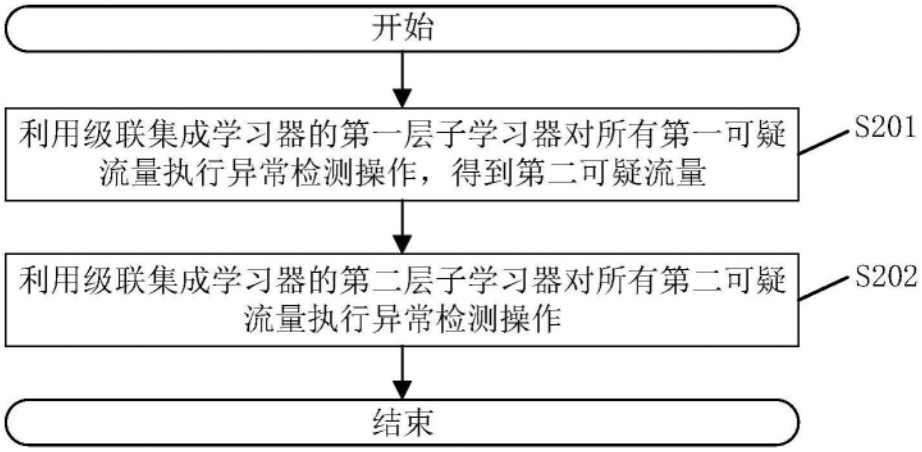

开始
利用级联集成学习器的第一层子学习器对所有第一可疑流量执行异常检测操作，得到第二可疑流量
S201
利用级联集成学习器的第二层子学习器对所有第二可疑流量执行异常检测操作
S202
结束

图4

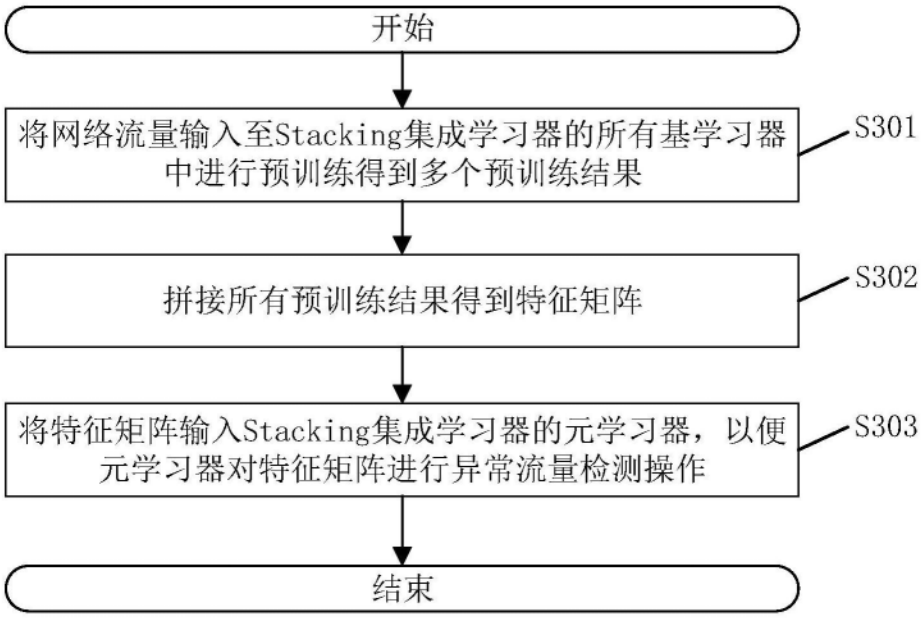

开始
将网络流量输入至Stacking集成学习器的所有基学习器中进行预训练得到多个预训练结果
S301
拼接所有预训练结果得到特征矩阵
S302
将特征矩阵输入Stacking集成学习器的元学习器，以便元学习器对特征矩阵进行异常流量检测操作
S303
结束

图5

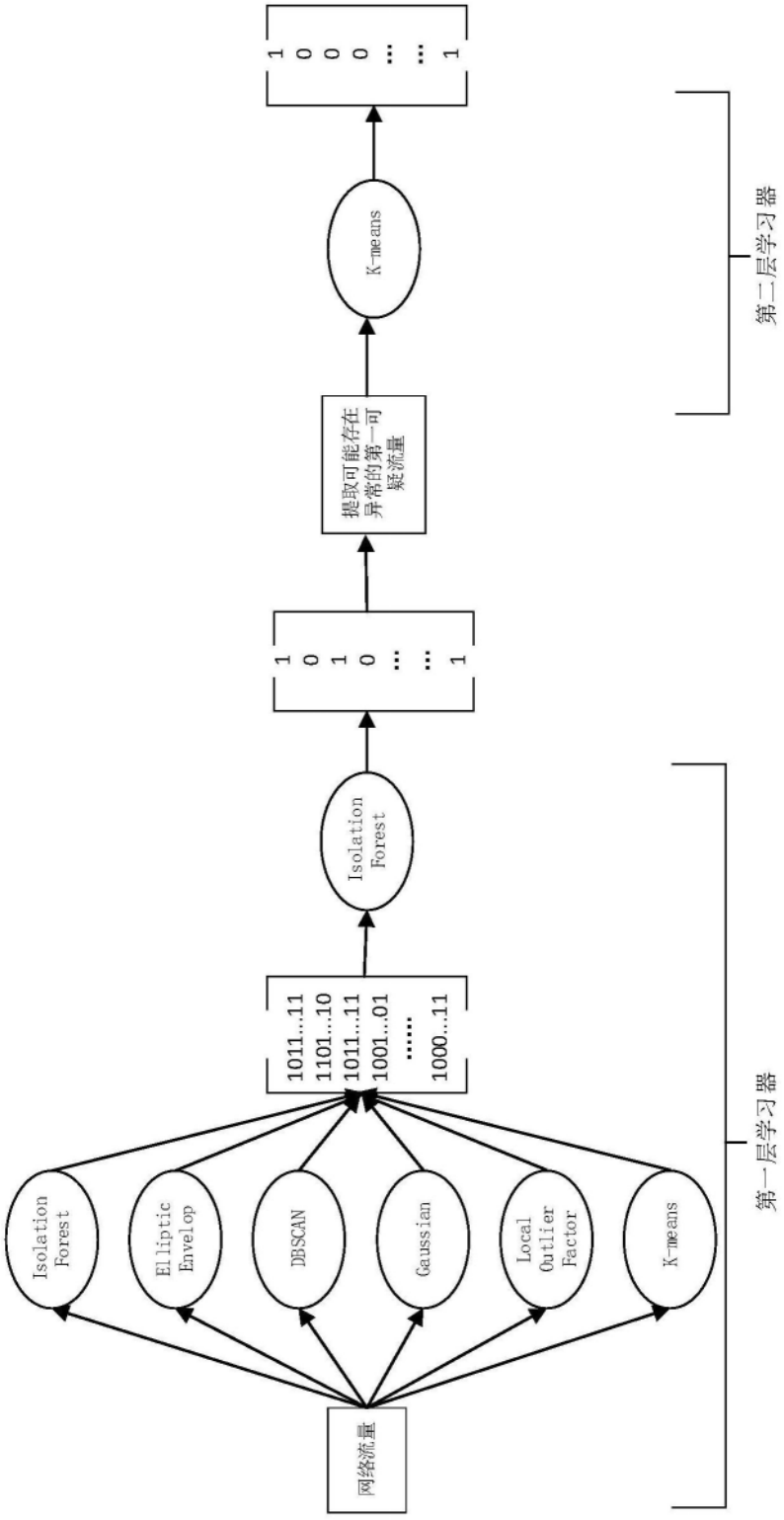

网络流量
Isolation Forest
Elliptic Envelop
DBSCAN
Gaussian
Local Outlier Factor
K-means
1011...11
1101...10
1011...11
1001...01
.....
1000...11
Isolation Forest
1 0 1 0 1
提取可能存在异常的第一可疑流量
K-means
1 0 0 0 1
第一层学习器
第二层学习器

图6

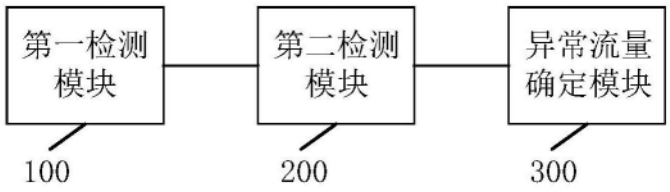


图7