(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

(43) 国际公布日
2019 年 7 月 18 日 (18.07.2019)

WIPO | PCT

(10) 国际公布号
WO 2019/136990 A1

(51) 国际专利分类号:
H04L 29/06 (2006.01)

(21) 国际申请号: PCT/CN2018/099948

(22) 国际申请日: 2018 年 8 月 10 日 (10.08.2018)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810032278.5 2018年1月12日 (12.01.2018) CN

(71) 申请人: 深圳壹账通智能科技有限公司(ONE CONNECT SMART TECHNOLOGY CO., LTD. (SHENZHEN)) [CN/CN]; 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201 室, Guangdong 518052 (CN)。

(72) 发明人: 羊志敏 (YANG, Zhimin); 中国广东省深圳市前海深港合作区前湾一路 1 号 A 栋 201室, Guangdong 518052 (CN)。

(74) 代理人: 广州华进联合专利商标代理有限公司 (ADVANCE CHINA IP LAW OFFICE); 中国广东省广州市天河区珠江东路 6 号 4501 房（部位：自编01-03和08-12单元）（仅限办公用途）, Guangdong 510623 (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, 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:** NETWORK DATA DETECTION METHOD, APPARATUS, COMPUTER DEVICE AND STORAGE MEDIUM

(54) **发明名称**: 网络数据检测方法、装置、计算机设备和存储介质

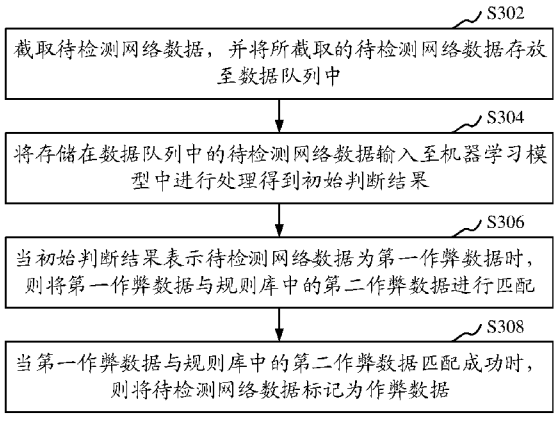


图 3

S302 Intercept network data to be detected, and store said intercepted network data within a data queue
S304 Input said intercepted network data within the data queue into a machine learning model for processing to obtain an initial determination result
S306 When the initial determination result indicates that the intercepted network data to be detected is first cheating data, then match the first cheating data with second cheating data within a rule library
S308 When the matching of the first cheating data within the second cheating data within the rule library is successful, then mark the network data to be detected as cheating data

(57) **Abstract:** A network data detection method, comprising: intercepting network data to be detected, and storing said intercepted network data within a data queue; inputting said intercepted network data within the data queue into a machine learning model for processing to obtain an initial determination result; when the initial determination result indicates that the intercepted network data to be detected is first cheating data, then matching the first cheating data with second cheating data within a rule library; when the matching of the first cheating data within the second cheating data within the rule library is successful, then marking the network data to be detected as cheating data.

(57) **摘要**: 一种网络数据检测方法，包括：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

WO 2019/136990 A1

[见续页]

SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

网络数据检测方法、装置、计算机设备和存储介质

相关申请的交叉引用

本申请要求于 2018 年 1 月 12 日提交中国专利局，申请号为 2018100322785，申请名称为“网络数据检测方法、装置、计算机设备和存储介质”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

本申请涉及一种网络数据检测方法、装置、计算机设备和存储介质。

背景技术

随着计算机技术的发展，出现了各种各样的营销活动，但是为了获取到该些营销活动中的奖励，出现了羊毛党，，羊毛党是指活跃在各互联网金融平台上，搜集各大电子商城、银行、实体店等各渠道的优惠促销活动、免费业务之类的信息，以此赚取小额奖励的投资者。

为了识别出该些羊毛党，发明人意识到，传统上是基于事前规则进行的，即通过事前规则判断对应的网络数据是否为作弊数据来判断该网络数据对应的用户是否为羊毛党。然而基于规则的反作弊系统需要事先制定规则，并根据该制定的规则对网络数据进行判断，并根据羊毛党的特征手动调整规则，但是这种判断方式中，由于手动调整的规则极易出现错误，因此通过手动调整规则极易造成羊毛党的判断错误。

发明内容

根据本申请公开的各种实施例，提供一种网络数据检测方法、装置、计算机设备和存储介质。

一种网络数据检测方法，包括:

截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中;

将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果;

当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及

当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

一种网络数据检测装置，包括:

截取模块，用于截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列

中；

机器学习模型判断模块，用于将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；

规则判断模块，用于当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及

标记模块，用于当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

一种计算机设备，包括存储器和一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述处理器执行时，使得所述一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例中的技术方案，下面将对实施例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它的附图。

图 1 为根据一个或多个实施例中网络数据检测方法的应用场景图。

图 2 为根据一个或多个实施例中的检测平台的架构图。

图 3 为根据一个或多个实施例中网络数据检测方法的流程示意图。

图 4 为根据一个或多个实施例中网络数据检测装置的框图。

图 5 为根据一个或多个实施例中计算机设备的框图。

具体实施方式

为了使本申请的技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请提供的网络数据检测方法，可以应用于如图 1 所示的应用环境中。终端通过网络与服务器通过网络进行通信。检测平台可以截取到终端与服务器通信的网络数据，并对该网络数据进行检测。例如，在检测平台截取到待检测网络数据后，则将所截取的待检测网路数据存放在数据队列中，然后依次通过机器学习模型和规则库中的规则对数据队列中的待检测网络数据进行检测，在机器学习模型和规则库中的规则均判定该待检测网络数据为作弊数据时，则将该待检测网络数据标记为作弊数据。终端可以但不限于是各种个人计算机、笔记本电脑、智能手机、平板电脑和便携式可穿戴设备，服务器即为网络服务器，其可以用独立的服务器或者是多个服务器组成的服务器集群来实现。检测平台也可以是以服务器或者多个服务器组成的服务器集群的形式实现。

参见图 2，图 2 为一实施例中检测平台的架构图，包括 H5 侦测引擎、Android 侦测引擎、IOS 侦测引擎、数据队列模块、机器学习模型模块、规则管理模块、风控引擎、活动管理模块、决策引擎模块、日志管理模块、数据监控平台、黑名单库和白名单库。H5 侦测引擎、Android 侦测引擎、IOS 侦测引擎用于截取待检测网络数据，并将所截取到的待检测网络数据放在数据队列中，机器学习模型模块用于对所截取到的待检测网络数据进行判断得到初始判断结果，当初始判断结果为作弊数据时，则继续通过规则管理模块调用风控引擎中的规则对该作弊数据进行进一步的判断。风控引擎中存储有对应的规则，活动管理模块中存储有对应的活动，且建立活动与规则之间的关系，决策引擎模块即将判断结果输出至数据监控平台，数据监控平台可以查看系统中的数据并对活动等进行管理。日志管理模块中存储有每次的判断记录，黑名单库中存储有作弊数据，白名单库中存储有正常数据等。

在其中一个实施例中，如图 3 所示，提供了一种网络数据检测方法，以该方法应用于图 1 中的检测平台为例进行说明，包括以下步骤：

S302：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中。

具体地，该检测平台可以通过设置代理服务器的方式截取待检测网络数据，例如设置代理服务器，并在终端预先设置调用程序，当终端与服务器发生数据交互时，则代理服务器可以通过该调用程序获取到终端所发出的待检测网络数据，并将该待检测网络数据发送给检测平台，从而检测平台在接收到该待检测网络数据后将该待检测网络数据存储至数据队列中，从而保证后续检测平台可以从数据队列中获取数据进行处理。之所以设置数据队列预先存储是为了在大量数据的情况出现时，可以防止检测平台故障，检测平台依然可以根据当前处理情况获取到对应数量的待检测网络数据进行处理，从而保证检测平台有条不紊地处理待检测网络数据。

S304：将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到

初始判断结果。

具体地，机器学习模型是预先设置的模型，该模型可以对待检测网络数据进行处理以判断该待检测网络数据是否为作弊数据，该机器学习模型可以是根据历史网络数据学习得到的。

机器学习模型可以根据当前在处理的待检测网络数据的数量从数据队列中获取待检测网络数据，还可以是分批从数据队列中获取到待检测网络数据，即在当前在处理的待检测网络数据处理完成后，才会从数据队列中获取到下一次要处理的待检测网络数据。机器学习模型可以同时对多个待检测网络数据进行处理，例如通过多线程的方式实现等，以提高数据处理的速度。

S306：当初始判断结果表示待检测网络数据为第一作弊数据时，则将第一作弊数据与规则库中的第二作弊数据进行匹配。

具体地，当初始判断结果，即经过机器学习模型所得到的待检测网络数据为作弊数据时，则可以通过规则库进行进一步的判断，以保证机器学习模型的判断结果正确。当初始判断结果，即经过机器学习模型所得到的待检测网络数据为正常数据时，则不会对该待检测网络数据进行拦截等，以保证终端与服务器的正常通信，或者也可以将正常数据发送至规则库中进行二次判断，以防止机器学习模型将作弊数据判断为正常数据的情况的发生。

，规则库中存储有大量的用于判断待检测网络数据是否为作弊数据的数据规则，可以第一作弊数据与规则库中的规则中的数据进行匹配。

S308：当第一作弊数据与规则库中的第二作弊数据匹配成功时，则将待检测网络数据标记为作弊数据。

具体地，当第一作弊数据与规则库中的第二作弊数据匹配成功时，则说明该待检测网络数据为作弊数据，从而通过两者相互查漏补缺，可以提高判断的准确率，防止机器学习模型产生误判的情况。且当第一作弊数据与规则库中的第二作弊数据未匹配成功时，还可以将该第一作弊数据输出，以便于用户进行人工干预，对规则库进行调整等，例如当人工判断出该第一作弊数据确实为作弊数据时，则可以对规则库进行调整，增加相应的规则，实现规则库的扩充；当人工判断出该第一作弊数据不为作弊数据时，则可以通过该第一作弊数据对机器学习模型进行修正，以进一步提高机器学习模型的准确度。

，当检测平台截取到待检测网络数据后，需要判断该网络数据是否为羊毛党的终端发出的网络数据，即作弊数据，且为了在接收大量数据情况下检测平台能够正常运行，检测平台首先将待检测网络数据存放在数据队列中，然后通过机器学习模型对该待检测数据进行预判断得到初始判断结果，即先根据大数据智能模型（AI 学习模型，即上述的机器学习模型）对该待检测数据进行初始判断，当该待检测网络数据为作弊数据时，则将该作弊数据再输入至规则库中进行检测，从而确保机器学习模型所作出的判断正确，通过机器学习模型首先对待检测网络数据进行判断，而不是首先通过规则库对待检测网络数据进行

判断的原因是由于机器学习模型的判断速度更加快，从而可以提高待检测网络数据的检测速度，其由于大多数网络数据均是正常网络数据，而非羊毛党的终端发出的网络数据，因此通过机器学习模型可以大量地减少输入至规则库的待检测网络数据，从而相比于先通过规则库进行判断的方式效率更高。

通过机器学习模型的判断，可以是首先提取到待检测网络数据的特征数据，该特征数据的提取方式可以是根据下文中提取历史网络数据的特征数据的方式相同，即首先对该待检测网络数据进行分词处理，然后计算当前分词在当前待检测网络数据中的词频，并根据包含当前分词的待检测网络数据的条数以及待检测网络数据的总条数获取到当前分词的逆文档频率；根据当前分词的词频和逆文档频率计算当前分词的词权重；获取词权重最大的当前分词作为当前待检测网络数据的特征数据；并通过所提取到的特征数据与机器模型中的特征数据进行比对，从而可以判断出该待检测网络数据是否为作弊数据，即是否为羊毛党对应的终端发出的数据，如果是，则继续将该作弊数据输入至规则库，并通过规则库对该作弊数据进行进一步的判断，由于机器学习模型的处理速度比较快，可以预先过滤掉大部分无效数据，再传输给规则库，可以提高判断的效率。

上述网络数据检测方法，首先通过机器学习模型对待检测网络数据进行初步判断，然后再通过规则库进行判断，两者相互查漏补缺，可以提高判断的准确率，且机器学习模型的处理速度比较快，可以预先过滤掉大部分无效数据，再传输给规则库，可以提高判断的效率。且机器学习模型的可以根据历史网络数据产生大量的新的规则，不需要人工添加，减少人力投入。

在其中一个实施例中，机器学习模型的生成主要包括三个步骤：定义目标问题，收集数据和特征，训练模型。因此该方法还包括一机器学习模型生成步骤，该机器学习模型生成步骤可以包括：获取已经标记的历史网络数据，并提取历史网络数据的特征数据；将标记为作弊数据的历史网络数据添加至黑名单库，将标记为正常数据的历史网络数据添加至白名单库；根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型。

，特征数据可以是可以代表该网络数据的数据，即如果某个特征数据在某一网络数据中出现的频率大，并且在其他网络数据中很少出现，则认为此特征数据具有很好的类别区分能力，因此该特征数据可以代表该网络数据，可选地，该特征数据可以为某一个词或某一个短语。

具体地，检测平台可以获取到以前标记完成的历史网络数据，该些历史网络数据已经进行了二分类，即已经根据标记分为正常数据和作弊数据，可以将正常数据存放在白名单库中，将作弊数据存放在黑名单库中；并提取该些历史网络数据的特征数据，根据特征数据和历史网络数据的类型即可以通过贝叶斯算法来进行建模生成机器学习模型。

，在生成机器学习模型后，还可以包括一个交叉训练步骤，即可以获取到其他的历史网络数据作为测试集用来对该机器学习模型进行优化。例如，通过 GridSearchCV 模型

对机器学习模型中的参数进行优化，即遍历机器学习模型中多种参数组合，通过交叉验证确定最佳效果参数。，参数可以包括测试集中的网络数据与机器学习模型中的特征数据的相似度等，选择相似度合适的参数作为机器学习模型的参数，例如当大于90%的相似度则认为待检测网络数据与机器学习模型中的特征数据相同，此时机器学习模型的判断准确率最好，则将该参数最为机器学习模型的参数。在对机器学习模型进行交叉训练后，将模型进行存储，以作为标准模型进行使用。

上述实施例中，通过对历史网络数据进行二分类和提取特征，并根据贝叶斯算来生成机器学习模型，使得该机器学习模型可靠，且在生成模型后对该机器学习模型进行交叉训练调参，进一步提高了机器学习模型的可靠性。

在其中一个实施例中，提取历史网络数据的特征数据的步骤可以包括：对历史网络数据进行分词处理；根据当前分词在当前历史网络数据中的词频，并根据包含当前分词的历史网络数据的条数以及历史网络数据的总条数获取到当前分词的逆文档频率；根据当前分词的词频和逆文档频率计算当前分词的词权重；获取词权重最大的当前分词作为当前历史网络数据的特征数据。

具体地，可以对历史网络数据进行分词处理，该分词处理可以是根据预设的分词逻辑进行处理或者是对历史网络数据进行随机分词，在此不做限制。如果根据预设的分词逻辑进行处理时，可以设置分词逻辑库，在分词处理之前，首先获取到该分词逻辑库，再将该分词逻辑库中的分词与历史网络数据进行匹配，匹配成功，则在该词前后加入分词符以进行分词。如果是对历史网络数据进行随机分词，则可以预先设定分词步长，例如可以将预设步长设定为3个字符、5个字符等等。具体的预设步长可以根据实际经验进行设置。检测平台根据预设步长按照预设切分顺序将历史网络数据依次切分成多个分词，预设切分顺序可以为由右至左的顺序，也可以为由左至右的顺序，也可以为其他顺序。

，当前分词在当前历史网络数据中的词频=当前分词在当前历史网络数据中出现的次数/当前历史网络数据的总分词数据；当前分词的逆文档频率=log（包含当前分词的历史网络数据的条数）/（历史网络数据的总条数+1）；词权重=词频*逆文档频率。

例如，当某次活动的历史网络数据包含1000条，第一条历史网络数据共计10个单词，单引号有3个，from也有3个，1000条历史网络数据中有10条语句包含单引号，100条语句包含from，则当前分词“单引号”的词频=3/10=0.3，逆文档频率=log（10/（1000+1））=1.958，因此词权重=0.3*1.958=0.5874。当前分词“from”的词频=3/10=0.3，逆文档频率=log（100/（1000+1））=0.995，因此词权重=0.3*0.995=0.3318。具体可以参见下表：

	包含该词的网络数据的条数	词频	逆文档频率	词权重
单引号	10	0.3	1.958	0.5874
From	100	0.3	0.995	0.3318

，单引号的词权重=0.587 > from 的词权重=0.3318，因此单引号的词权重对比之下更大，比 from 更能代表一条历史网络数据是否是攻击语句，即作弊数据。

上述实施例中，特征数据可以是可以代表该网络数据的数据，即如果某个特征数据在某一网络数据中出现的频率大，并且在其他网络数据中很少出现，则认为此特征数据具有很好的类别区分能力，因此该特征数据可以代表该网络数据，因此通过计算分词的词权重可以准确地判断出该分词是否可以作为特征数据，为提高后续机器学习模型的准确性奠定基础。

在其中一个实施例中，由于羊毛党可能根据机器学习对数据进行一定的修改，因此使得机器学习模型不再适应当前羊毛党的判断，为了避免这种情况的发生，可以对生成机器学习模型的黑名单库和白名单库中的历史网络数据进行优化，以提高样本的准确性，从而可以设置时间段重新生成该机器学习模型，提高机器学习模型的适应性。对于黑名单库和白名单库中的历史网络数据的优化可以包括历史网络数据的增加以及历史网络数据的删除，对于历史网络数据的增加可以包括黑名单库中的历史网络数据的增加和白名单库中的实例网络数据的增加。对于历史网络数据的删除主要是针对数据去重和加密数据的删除。

具体地，对于黑名单库中历史网络数据的增加可以是在截取待检测网络数据的步骤之后，即在截取待检测网络数据的步骤之后还可以包括获取待检测网络数据对应的终端的网络地址；统计截取到的网络地址对应的终端发送的待检测网络数据的频率；当频率大于预设频率时，则将待检测网络数据添加到黑名单库中。即可以通过日志管理模块搜集动态网络地址黑名单，将动态网络地址所发送的网络数据添加到黑名单中，一般会认为检测平台持续截取到某一网络地址对应的网络数据，则认为该网络地址所对应的网络数据为作弊数据，因此可以统计截取到的网络地址对应的终端发送的待检测网络数据的频率，即在预设时间段内截取到该网络地址所发送的网络数据的次数，如果大于预设次数，则认为该网络数据为作弊数据，直接加入到黑名单库中即可。

具体地，对于黑名单库中历史网络数据的增加可以包括当初始判断结果表示待检测网络数据不为第一作弊数据时，则将待检测网络数据标记为正常数据；将正常数据添加到白名单库中。即可以定期获取到正常数据的原始网络数据添加到白名单库中，以更新白名单库，保证白名单库的时效性，白名单数据可以包括通过机器学习模型判断得到的正常数据，还可以包括人工干预后确定的为正常数据的第一作弊数据，这样可以增加正常数据的多样性，提高正常数据的覆盖范围，从而可以保证白名单库中的网络数据的覆盖面更全。在该实施例中，可以定期将标记为正常数据的网络数据添加到白名单库中，例如检测平台可以获取到日志管理模块中的日志数据，从选取标记为正常数据的新的网络数据，例如可以根据上次添加的日期，获取上次添加日期后的新的网络数据，将该新的网络数据添加至对应的白名单库中。

具体地，对于历史网络数据的删除包括去重和加密数据的删除，去重包括按照黑名单库中的历史网络数据的顺序获取黑名单库中相同的历史网络数据；按照白名单库中的历史网络数据的顺序获取白名单库中相同的历史网络数据；对相同的历史网络数据进行去重处理。例如可以通过黑名单库中的第一条历史网络数据依次去匹配黑名单库中的其他历史网络数据，当两条历史网络数据的相似度大于预设相似度时，则认为该两条历史网络数据相同，因此删除掉一条，仅保留一条历史网络数据即可。同样地对于白名单库，可以通过白名单库中的第一条历史网络数据依次去匹配白名单库中的其他历史网络数据，当两条历史网络数据的相似度大于预设相似度时，则认为该两条历史网络数据相同，因此删除掉一条，仅保留一条历史网络数据即可，这样可以尽可能地将样本的数量，提高数据处理的效率。对于加密数据的去重，可以是根据预设参数名称从白名单库和黑名单库中去重，例如可以将加密的参数的名称预先进行存储，当该历史网络数据中存在该参数的名称时，则直接从白名单库和黑名单库中删除该参数对应的网络数据。

上述实施例中，对白名单库和黑名单库中的历史网络数据进行清洗，可以通过机器学习大量的样本并产生新的样本，大大释放原有人工的维护量。

在其中一个实施例中，还包括一对机器学习模型进行评估的步骤，该步骤可以是在生成机器学习模型的步骤之后进行，即在根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型的步骤之后，还可以包括：将预设正样本和预设负样本输入至机器学习模型中得到测试结果；将测试结果与预期结果进行比较得到比较结果；根据比较结果计算机器学习模型的评分；当评分小于预设分数时，则通过预设正样本和预设负样本对机器学习模型进行修正；当评分大于等于预设分数时，则继续将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

具体地，预设的正样本和负样本是已经确定其为正常数据还是作弊数据的网络数据，将该网络数据输入至机器学习模型中可以得到测试结果，将该测试结果与预期结果进行比较得到比较结果，并可以根据比较结果得到机器学习模型的评分。例如参见下表所示：

真实情况（预期结果）	预测结果（测试结果）	
	正常数据	作弊数据
正常数据	TP	FN
作弊数据	FP	TN

将该测试结果与预期结果进行比较得到比较结果：TP：正确肯定——实际是正常数据，识别为正常数据；FN：错误否定（漏报）——实际是正常数据，却识别成了作弊数据；FP：错误肯定（误报）——实际是作弊数据，却识别成了正常数据；TN：正确否定——实际是作弊数据，识别为作弊数据。并根据比较结果计算机器学习模型的评分：评分的参数包括：准确率、误分率、召回率、查准率、TPR、FAR、FRR 以及 f1-score 等，该

些参数的具体计算方式如下:

AccuracyRate（准确率）:（TP+TN）/（TP+TN+FN+FP）;

ErrorRate（误分率）:（FN+FP）/（TP+TN+FN+FP）;

Recall（召回率、查全率、击中概率）: TP/（TP+FN），在所有正常数据中有多少被识别为正常数据;

Precision（查准率）: TP/（TP+FP），在所有识别成正常数据中有多少是真正的正常数据;

TPR（TruePositive Rate）: TP/（TP+FN），实际就是 Recall;

FAR（FalseAcceptance Rate）或 FPR（False Positive Rate）: FP/（FP+TN），错误接收率，误报率，在所有作弊数据中有多少被识别为正常数据;

FRR（FalseRejection Rate）: FN/（TP+FN），错误拒绝率，拒真率，在所有正常数据中有多少被识别为作弊数据，它等于 1-Recall;

f1-score 是召回率和准确率的调和平均数，并假设两者一样重要,计算公式: f1-score=(2*Recall*Accuracy) / (Recall+Accuracy)。

根据所计算的参数可以对机器学习模型进行打分或分级，只有分数或等级合适的机器学习模型才会投入使用。即当评分大于等于预设分数时，则继续将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。当评分小于预设分数时，则通过预设正样本和预设负样本对机器学习模型进行修正，例如通过预设正常本和预设负样本对在机器学习模型的基础上再次进行训练等。

上述实施例中，在生成机器学习模型之后，并非直接将机器学习模型投入使用，而是对机器学习模型进行打分，只有分数合适的机器学习模型才会投入使用，进一步提高了待检测网络数据的检测准确性。

应该理解的是，虽然图 3 的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，图 3 中的至少一部分步骤可以包括多个子步骤或者多个阶段，这些子步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些子步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤的子步骤或者阶段的至少一部分轮流或者交替地执行。

在一个实施例中，如图 4 所示，提供了一种网络数据检测装置，包括：截取模块 100、机器学习模型判断模块 200、规则判断模块 300、标记模块 400，其中:

截取模块 100，用于截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中。

机器学习模型判断模块 200，用于将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

规则判断模块300，用于当初始判断结果表示待检测网络数据为第一作弊数据时，则将第一作弊数据与规则库中的第二作弊数据进行匹配。

标记模块400，用于当第一作弊数据与规则库中的第二作弊数据匹配成功时，则将待检测网络数据标记为作弊数据。

在其中一个实施例中，装置还包括:

特征数据提取模块，用于获取已经标记的历史网络数据，并提取历史网络数据的特征数据。

库生成模块，用于将标记为作弊数据的历史网络数据添加至黑名单库，将标记为正常数据的历史网络数据添加至白名单库。

机器学习模型生成模块，用于根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型。

在其中一个实施例中，特征数据提取模块包括:

分词单元，用于对历史网络数据进行分词处理。

词特征获取单元，用于根据当前分词在当前历史网络数据中的词频，并根据包含当前分词的历史网络数据的条数以及历史网络数据的总条数获取到当前分词的逆文档频率。

词权重计算单元，用于根据当前分词的词频和逆文档频率计算当前分词的词权重。

特征数据提取单元，用于获取词权重最大的当前分词作为当前历史网络数据的特征数据。

在其中一个实施例中，装置还包括:

网络地址获取模块，用于在截取待检测网络数据后，获取待检测网络数据对应的终端的网络地址。

频率统计模块，用于统计截取到的网络地址对应的终端发送的待检测网络数据的频率。

黑名单库添加模块，用于当频率大于预设频率时，则将待检测网络数据添加到黑名单库中。

在其中一个实施例中，标记模块400还用于当初始判断结果表示待检测网络数据不为第一作弊数据时，则将待检测网络数据标记为正常数据。

装置还包括：白名单库添加模块，用于将正常数据添加到白名单库中。

在其中一个实施例中，装置还包括:

相同数据获取模块，用于按照黑名单库中的历史网络数据的顺序获取黑名单库中相同的历史网络数据；按照白名单库中的历史网络数据的顺序获取白名单库中相同的历史网络数据。

去重模块，用于对相同的历史网络数据进行去重处理。

在其中一个实施例中，装置还包括:

测试模块，用于将预设正样本和预设负样本输入至机器学习模型中得到测试结果。

比较模块，用于将测试结果与预期结果进行比较得到比较结果。

评分模块，用于根据比较结果计算机器学习模型的评分。

修正模块，用于当评分小于预设分数时，则通过预设正样本和预设负样本对机器学习模型进行修正。

机器学习模型判断模块还用于当评分大于等于预设分数时，则继续将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

关于网络数据检测装置的具体限定可以参见上文中对于网络数据检测方法的限定，在此不再赘述。上述网络数据检测装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 5 所示。该计算机设备包括通过系统总线连接的处理器、存储器、网络接口和数据库。，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质、内存储器。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存储器为非易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储黑名单库、白名单库以及日志数据等。该计算机设备的网络接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种网络数据检测方法。

本领域技术人员可以理解，图 5 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

一种计算机设备，包括存储器和一个或多个处理器，存储器中储存有计算机可读指令，计算机可读指令被处理器执行时，使得一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当初始判断结果表示待检测网络数据为第一作弊数据时，则将第一作弊数据与规则库中的第二作弊数据进行匹配；当第一作弊数据与规则库中的第二作弊数据匹配成功时，则将待检测网络数据标记为作弊数据。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：获取已经标记的历史网络数据，并提取历史网络数据的特征数据；将标记为作弊数据的历史网络数据添加至黑名单库，将标记为正常数据的历史网络数据添加至白名单库；根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型。

在一个实施例中，处理器执行计算机可读指令时所实现的提取历史网络数据的特征数据的步骤，可以包括：对历史网络数据进行分词处理；根据当前分词在当前历史网络数

据中的词频，并根据包含当前分词的历史网络数据的条数以及历史网络数据的总条数获取到当前分词的逆文档频率；根据当前分词的词频和逆文档频率计算当前分词的词权重；获取词权重最大的当前分词作为当前历史网络数据的特征数据。

在一个实施例中，处理器执行计算机可读指令时所实现的截取待检测网络数据的步骤之后，还可以包括：获取待检测网络数据对应的终端的网络地址；统计截取到的网络地址对应的终端发送的待检测网络数据的频率；当频率大于预设频率时，则将待检测网络数据添加到黑名单库中。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：当初始判断结果表示待检测网络数据不为第一作弊数据时，则将待检测网络数据标记为正常数据；将正常数据添加到白名单库中。

在一个实施例中，处理器执行计算机可读指令时还实现以下步骤：按照黑名单库中的历史网络数据的顺序获取黑名单库中相同的历史网络数据；按照白名单库中的历史网络数据的顺序获取白名单库中相同的历史网络数据；对相同的历史网络数据进行去重处理。

在一个实施例中，处理器执行计算机可读指令时所实现的根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型的步骤之后，还可以包括：将预设正样本和预设负样本输入至机器学习模型中得到测试结果；将测试结果与预期结果进行比较得到比较结果；根据比较结果计算机器学习模型的评分；当评分小于预设分数时，则通过预设正样本和预设负样本对机器学习模型进行修正；当评分大于等于预设分数时，则继续将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，计算机可读指令被一个或多个处理器执行时，使得一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当初始判断结果表示待检测网络数据为第一作弊数据时，则将第一作弊数据与规则库中的第二作弊数据进行匹配；当第一作弊数据与规则库中的第二作弊数据匹配成功时，则将待检测网络数据标记为作弊数据。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：获取已经标记的历史网络数据，并提取历史网络数据的特征数据；将标记为作弊数据的历史网络数据添加至黑名单库，将标记为正常数据的历史网络数据添加至白名单库；根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型。

在一个实施例中，计算机可读指令被处理器执行时所实现的提取历史网络数据的特征数据的步骤，可以包括：对历史网络数据进行分词处理；根据当前分词在当前历史网络数据中的词频，并根据包含当前分词的历史网络数据的条数以及历史网络数据的总条数获取到当前分词的逆文档频率；根据当前分词的词频和逆文档频率计算当前分词的词权重；获取词权重最大的当前分词作为当前历史网络数据的特征数据。

在一个实施例中，计算机可读指令被处理器执行时所实现的截取待检测网络数据的步骤之后，还可以包括：获取待检测网络数据对应的终端的网络地址；统计截取到的网络地址对应的终端发送的待检测网络数据的频率；当频率大于预设频率时，则将待检测网络数据添加到黑名单库中。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：当初始判断结果表示待检测网络数据不为第一作弊数据时，则将待检测网络数据标记为正常数据；将正常数据添加到白名单库中。

在一个实施例中，计算机可读指令被处理器执行时还实现以下步骤：按照黑名单库中的历史网络数据的顺序获取黑名单库中相同的历史网络数据；按照白名单库中的历史网络数据的顺序获取白名单库中相同的历史网络数据；对相同的历史网络数据进行去重处理。

在一个实施例中，计算机可读指令被处理器执行时所实现的根据黑名单库中的历史网络数据的特征数据以及白名单库中的历史网络数据的特征数据建立机器学习模型的步骤之后，还可以包括：将预设正样本和预设负样本输入至机器学习模型中得到测试结果；将测试结果与预期结果进行比较得到比较结果；根据比较结果计算机器学习模型的评分；当评分小于预设分数时，则通过预设正样本和预设负样本对机器学习模型进行修正；当评分大于等于预设分数时，则继续将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读取存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。，本申请所提供的各实施例中所使用的对存储器、存储、数据库或其它介质的任何引用，均可包括非易失性和/或易失性存储器。非易失性存储器可包括只读存储器（ROM）、可编程 ROM（PROM）、电可编程 ROM（EPROM）、电可擦除可编程 ROM（EEPROM）或闪存。易失性存储器可包括随机存取存储器（RAM）或者外部高速缓冲存储器。作为说明而非局限，RAM 以多种形式可得，诸如静态 RAM（SRAM）、动态 RAM（DRAM）、同步 DRAM（SDRAM）、双数据率 SDRAM（DDRSDRAM）、增强型 SDRAM（ESDRAM）、同步链路（Synchlink） DRAM（SLDRAM）、存储器总线（Rambus）直接 RAM（RDRAM）、直接存储器总线动态 RAM（DRDRAM）、以及存储器总线动态 RAM（RDRAM）等。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对发明专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范

围。因此，本申请专利的保护范围应以所附权利要求为准。

权利要求书

1、一种网络数据检测方法，包括:

截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中;

将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果;

当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及

当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

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:

获取已经标记的历史网络数据，并提取所述历史网络数据的特征数据;

将标记为作弊数据的所述历史网络数据添加至黑名单库，将标记为正常数据的所述历史网络数据添加至白名单库；及

根据所述黑名单库中的历史网络数据的特征数据以及所述白名单库中的历史网络数据的特征数据建立机器学习模型。

3、根据权利要求 2 所述的方法，其特征在于，所述提取所述历史网络数据的特征数据，包括:

对所述历史网络数据进行分词处理;

根据当前分词在当前历史网络数据中的词频，并根据包含所述当前分词的历史网络数据的条数以及所述历史网络数据的总条数获取到所述当前分词的逆文档频率;

根据所述当前分词的词频和逆文档频率计算所述当前分词的词权重；及

获取词权重最大的当前分词作为所述当前历史网络数据的特征数据。

4、根据权利要求 2 或 3 所述的方法，其特征在于，所述截取待检测网络数据之后，还包括:

获取所述待检测网络数据对应的终端的网络地址;

统计截取到的所述网络地址对应的终端发送的所述待检测网络数据的频率；及

当所述频率大于预设频率时，则将所述待检测网络数据添加到所述黑名单库中。

5、根据权利要求 2 或 3 所述的方法，其特征在于，所述方法还包括:

当所述初始判断结果表示所述待检测网络数据不为第一作弊数据时，则将所述待检测网络数据标记为正常数据；及

将所述正常数据添加到所述白名单库中。

6、根据权利要求 2 或 3 所述的方法，其特征在于，所述方法还包括:

按照所述黑名单库中的历史网络数据的顺序获取所述黑名单库中相同的历史网络数据;

按照所述白名单库中的历史网络数据的顺序获取所述白名单库中相同的历史网络数

据；及

对所述相同的历史网络数据进行去重处理。

7、根据权利要求 2 或 3 所述的方法，其特征在于，所述根据所述黑名单库中的历史网络数据的特征数据以及所述白名单库中的历史网络数据的特征数据建立机器学习模型之后，还包括：

将预设正样本和预设负样本输入至所述机器学习模型中得到测试结果；

将所述测试结果与预期结果进行比较得到比较结果；

根据所述比较结果计算所述机器学习模型的评分；

当所述评分小于预设分数时，则通过所述预设正样本和预设负样本对所述机器学习模型进行修正；及

当所述评分大于等于所述预设分数时，则继续将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

8、一种网络数据检测装置，包括：

截取模块，用于截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；

机器学习模型判断模块，用于将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；

规则判断模块，用于当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及

标记模块，用于当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

9、一种计算机设备，包括存储器及一个或多个处理器，所述存储器中储存有计算机可读指令，所述计算机可读指令被所述一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

10、根据权利要 9 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：获取已经标记的历史网络数据，并提取所述历史网络数据的特征数据；将标记为作弊数据的所述历史网络数据添加至黑名单库，将标记为正常数据的所述历史网络数据添加至白名单库；及根据所述黑名单库中的历史网络数据的特征数据以及所述白名单库中的历史网络数据的特征数据建立机器学习模型。

11、根据权利要 10 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时所实现的所述提取所述历史网络数据的特征数据，包括：对所述历史网络数据进

行分词处理；根据当前分词在当前历史网络数据中的词频，并根据包含所述当前分词的历史网络数据的条数以及所述历史网络数据的总条数获取到所述当前分词的逆文档频率；根据所述当前分词的词频和逆文档频率计算所述当前分词的词权重；及获取词权重最大的当前分词作为所述当前历史网络数据的特征数据。

12、根据权利要 10 或 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时所实现的所述截取待检测网络数据之后，还包括：获取所述待检测网络数据对应的终端的网络地址；统计截取到的所述网络地址对应的终端发送的所述待检测网络数据的频率；及当所述频率大于预设频率时，则将所述待检测网络数据添加到所述黑名单库中。

13、根据权利要 10 或 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：当所述初始判断结果表示所述待检测网络数据不为第一作弊数据时，则将所述待检测网络数据标记为正常数据；及将所述正常数据添加到所述白名单库中。

14、根据权利要 10 或 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时还执行以下步骤：按照所述黑名单库中的历史网络数据的顺序获取所述黑名单库中相同的历史网络数据；按照所述白名单库中的历史网络数据的顺序获取所述白名单库中相同的历史网络数据；及对所述相同的历史网络数据进行去重处理。

15、根据权利要 10 或 11 所述的计算机设备，其特征在于，所述处理器执行所述计算机可读指令时所实现的所述根据所述黑名单库中的历史网络数据的特征数据以及所述白名单库中的历史网络数据的特征数据建立机器学习模型之后，还包括：将预设正样本和预设负样本输入至所述机器学习模型中得到测试结果；将所述测试结果与预期结果进行比较得到比较结果；根据所述比较结果计算所述机器学习模型的评分；当所述评分小于预设分数时，则通过所述预设正样本和预设负样本对所述机器学习模型进行修正；及当所述评分大于等于所述预设分数时，则继续将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果。

16、一个或多个存储有计算机可读指令的非易失性计算机可读存储介质，所述计算机可读指令被一个或多个处理器执行时，使得所述一个或多个处理器执行以下步骤：截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中；将存储在所述数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果；当所述初始判断结果表示所述待检测网络数据为第一作弊数据时，则将所述第一作弊数据与规则库中的第二作弊数据进行匹配；及当所述第一作弊数据与所述规则库中的第二作弊数据匹配成功时，则将所述待检测网络数据标记为作弊数据。

17、根据权利要 16 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：获取已经标记的历史网络数据，并提取所述历史网络数据的特征数据；将标记为作弊数据的所述历史网络数据添加至黑名单库，将标记为正常数据的所

述历史网络数据添加至白名单库；及根据所述黑名单库中的历史网络数据的特征数据以及所述白名单库中的历史网络数据的特征数据建立机器学习模型。

18、根据权利要 17 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时所实现的所述提取所述历史网络数据的特征数据，包括：对所述历史网络数据进行分词处理；根据当前分词在当前历史网络数据中的词频，并根据包含所述当前分词的历史网络数据的条数以及所述历史网络数据的总条数获取到所述当前分词的逆文档频率；根据所述当前分词的词频和逆文档频率计算所述当前分词的词权重；及获取词权重最大的当前分词作为所述当前历史网络数据的特征数据。

19、根据权利要 17 或 18 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时所实现的所述截取待检测网络数据之后，还包括：获取所述待检测网络数据对应的终端的网络地址；统计截取到的所述网络地址对应的终端发送的所述待检测网络数据的频率；及当所述频率大于预设频率时，则将所述待检测网络数据添加到所述黑名单库中。

20、根据权利要 17 或 18 所述的存储介质，其特征在于，所述计算机可读指令被所述处理器执行时还执行以下步骤：当所述初始判断结果表示所述待检测网络数据不为第一作弊数据时，则将所述待检测网络数据标记为正常数据；及将所述正常数据添加到所述白名单库中。

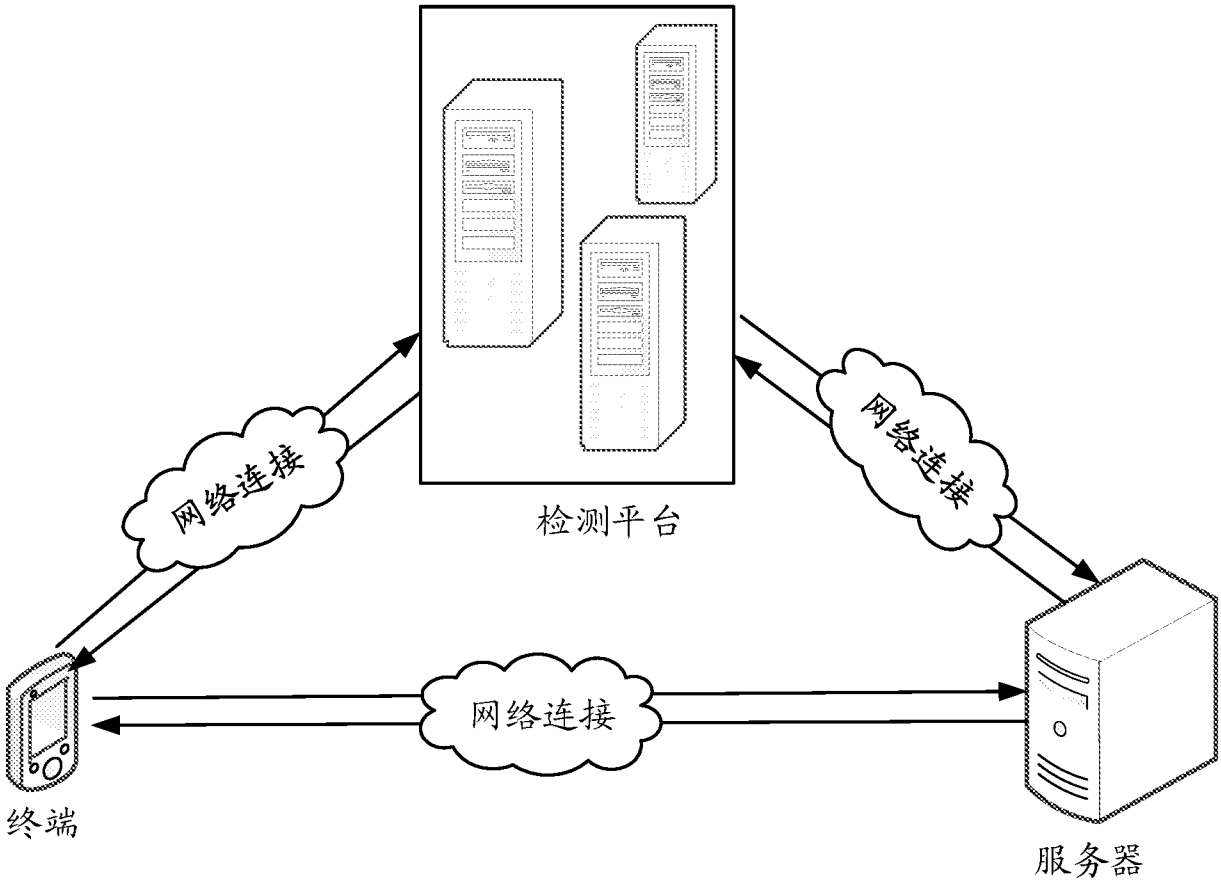


图 1

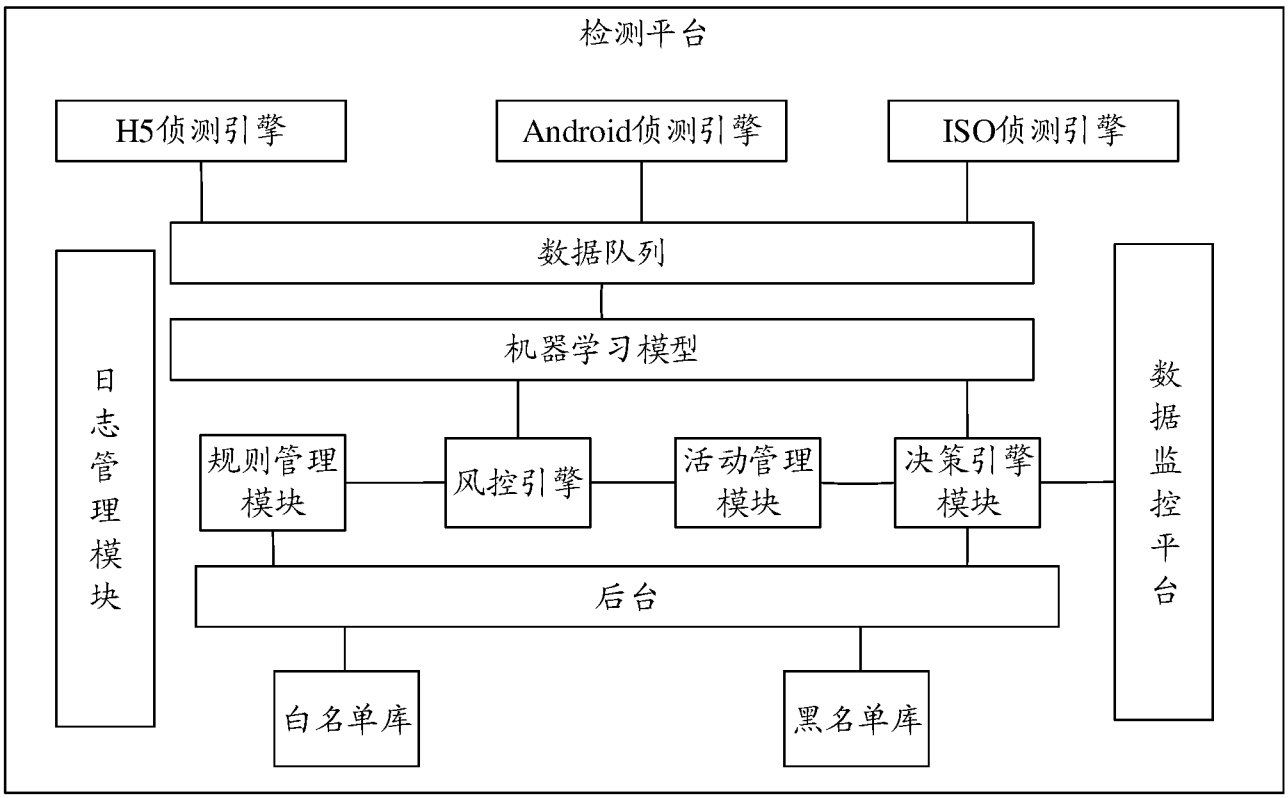

检测平台
H5侦测引擎
Android侦测引擎
ISO侦测引擎
数据队列
机器学习模型
日志管理模块
规则管理模块
风控引擎
活动管理模块
决策引擎模块
数据监控平台
后台
白名单库
黑名单库

图 2

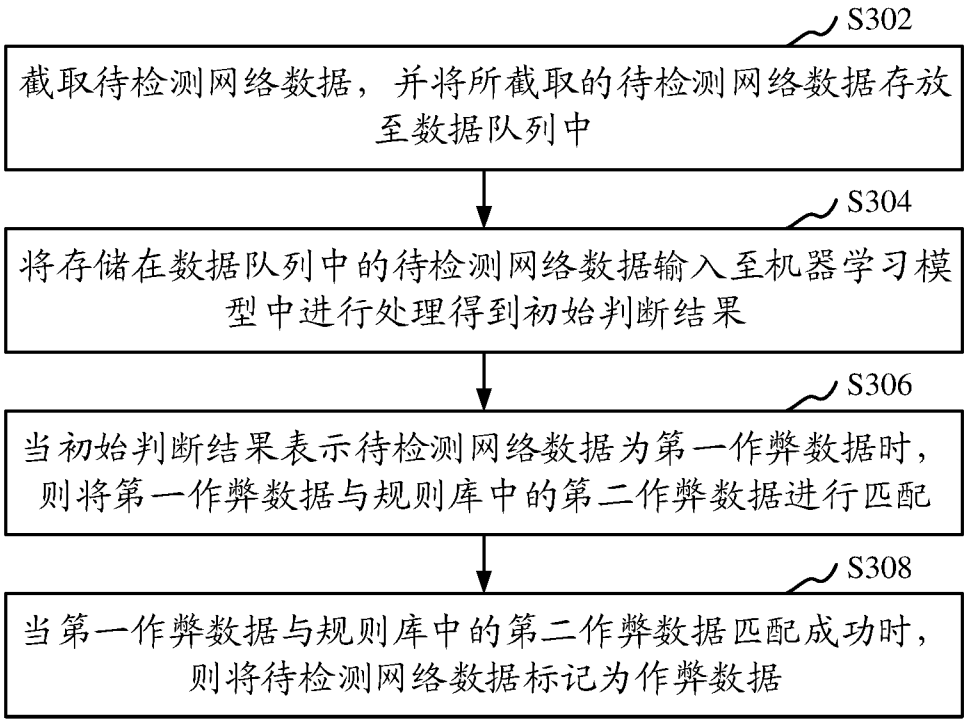

S302
截取待检测网络数据，并将所截取的待检测网络数据存放至数据队列中
S304
将存储在数据队列中的待检测网络数据输入至机器学习模型中进行处理得到初始判断结果
S306
当初始判断结果表示待检测网络数据为第一作弊数据时，则将第一作弊数据与规则库中的第二作弊数据进行匹配
S308
当第一作弊数据与规则库中的第二作弊数据匹配成功时，则将待检测网络数据标记为作弊数据

图 3

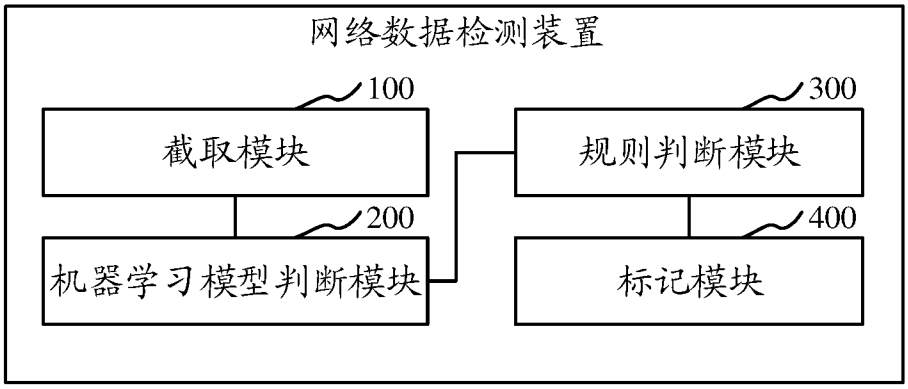

网络数据检测装置
100
截取模块
300
规则判断模块
200
机器学习模型判断模块
400
标记模块

图 4

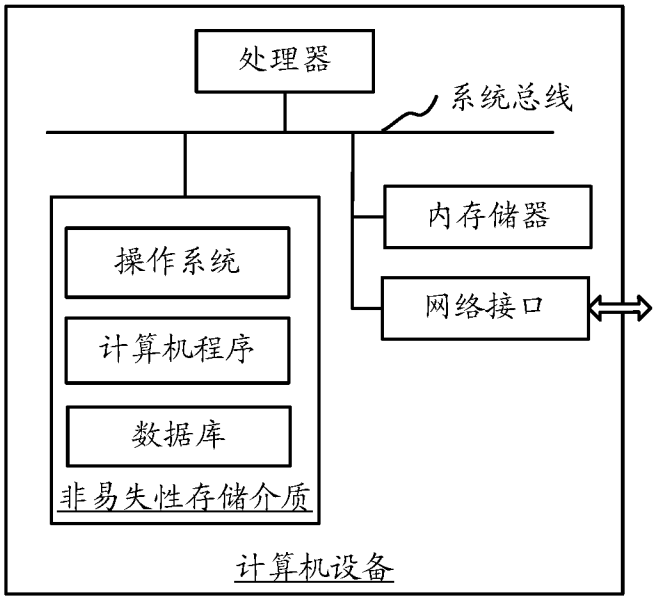

处理器
系统总线
内存储器
操作系统
网络接口
计算机程序
数据库
非易失性存储介质
计算机设备

图 5

INTERNATIONAL SEARCH REPORT

International application No.
PCT/CN2018/099948

A. CLASSIFICATION OF SUBJECT MATTER

H04L 29/06(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)

H04L; G06F

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)

CNABS; CNTXT; CNKI; VEN; WOTXT; USTXT; EPTXT: 作弊, 数据, 规则, 模型, 匹配, cheat, data, rule, model, match

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 108306864 A (ONECONNECT TECHNOLOGY CO., LTD.) 20 July 2018 (2018-07-20) claims	1-20
X	CN 107274212 A (BEIJING XIAODU INFORMATION TECHNOLOGY CO., LTD.) 20 October 2017 (2017-10-20) description, paragraphs [0031]-[0045] and [0060]-[0081], and figure 3	1-20
A	CN 103077172 A (TENCENT TECHNOLOGY (SHENZHEN) CO., LTD.) 01 May 2013 (2013-05-01) entire document	1-20
A	CN 103312703 A (SOUTHWEST UNIVERSITY) 18 September 2013 (2013-09-18) entire document	1-20
A	CN 105488091 A (HARBIN ANTIY TECHNOLOGY CO., LTD.) 13 April 2016 (2016-04-13) 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	Date of mailing of the international search report
23 October 2018	**06 November 2018**

Name and mailing address of the ISA/CN	Authorized officer
State Intellectual Property Office of the P. R. China (ISA/CN) **No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing 100088** **China**	
Facsimile No. **(86-10)62019451**	Telephone No.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2018/099948

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108306864	A	20 July 2018	None			
CN	107274212	A	20 October 2017	None			
CN	103077172	A	01 May 2013	CN	103077172	B	06 June 2017
CN	103312703	A	18 September 2013	CN	103312703	B	15 March 2017
CN	105488091	A	13 April 2016	None			

国际检索报告

国际申请号
PCT/CN2018/099948

A. 主题的分类

H04L 29/06(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)

H04L; G06F

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）)

CNABS;CNTXT;CNKI;VEN;WOTXT;USTXT;EPTXT:作弊，数据，规则，模型，匹配，cheat, data, rule, model, match

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 108306864 A (深圳壹账通智能科技有限公司) 2018年 7月 20日 (2018-07-20) 权利要求书	1-20
X	CN 107274212 A (北京小度信息科技有限公司) 2017年 10月 20日 (2017-10-20) 说明书第0031-0045、0060-0081段，附图3	1-20
A	CN 103077172 A (腾讯科技深圳有限公司) 2013年 5月 1日 (2013-05-01) 全文	1-20
A	CN 103312703 A (西南大学) 2013年 9月 18日 (2013-09-18) 全文	1-20
A	CN 105488091 A (哈尔滨安天科技股份有限公司) 2016年 4月 13日 (2016-04-13) 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	国际检索报告邮寄日期
2018年 10月 23日	2018年 11月 6日

ISA/CN的名称和邮寄地址	受权官员
中华人民共和国国家知识产权局(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	罗芳洁 电话号码 86-(010)-62411638

表 PCT/ISA/210 (第2页) (2015年1月)

国际检索报告
关于同族专利的信息

国际申请号
PCT/CN2018/099948

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	108306864	A	2018年 7月 20日	无			
CN	107274212	A	2017年 10月 20日	无			
CN	103077172	A	2013年 5月 1日	CN	103077172	B	2017年 6月 6日
CN	103312703	A	2013年 9月 18日	CN	103312703	B	2017年 3月 15日
CN	105488091	A	2016年 4月 13日	无			