



(12) 发明专利申请

(10) 申请公布号 CN 119341994 A

(43) 申请公布日 2025. 01. 21

(21) 申请号 202411375667.X

(22) 申请日 2024.09.29

(71) 申请人 中国联合网络通信集团有限公司

地址 100033 北京市西城区金融大街21号

(72) 发明人 张兴伟 高洁 王海鹏 王剑

张扬

(74) 专利代理机构 北京天昊联合知识产权代理

有限公司 11112

专利代理师 苏赛军 邓伯英

(51) Int. Cl.

H04L 47/24 (2022.01)

H04L 41/16 (2022.01)

G06F 18/2415 (2023.01)

G06F 18/214 (2023.01)

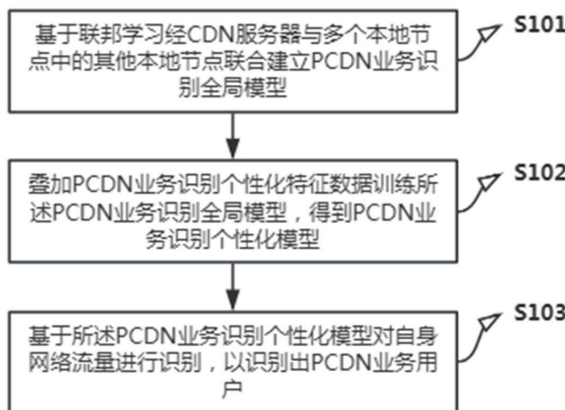
权利要求书3页 说明书16页 附图3页

(54) 发明名称

PCDN业务识别方法、本地节点、CDN服务器、设备及存储介质

(57) 摘要

本发明提供一种PCDN业务的识别方法、本地节点、CDN服务器、电子设备及计算机可读存储介质,涉及网络检测技术领域。识别方法包括:基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型;叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型;基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。以至少解决相关技术中存在的PCDN业务识别准确度不高的问题。适应于网络检测以及业务检测场景。



1. 一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的多个本地节点中的任意一个本地节点,其特征在于,PCDN业务的识别方法包括:

基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型;

叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型;

基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

2. 根据权利要求1所述的PCDN业务的识别方法,其特征在于,在所述基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型之前,还包括:

将自身所采集的本地业务数据进行加密后发送至CDN服务器;

接收CDN服务器发送的PCDN业务识别特征数据,其中,所述PCDN业务识别特征数据由所述CDN服务器将加密后的本地业务数据与BSS节点加密后的BSS业务数据基于用户ID匹配对齐得到,所述PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据;

将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据。

3. 根据权利要求1所述的PCDN业务的识别方法,其特征在于,所述基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,具体包括:

基于PCDN业务识别共性特征数据训练PCDN业务识别模型,得到本地PCDN业务识别模型;

将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,N为正整数;

接收CDN服务器发送的PCDN业务识别全局模型。

4. 根据权利要求3所述的PCDN业务的识别方法,其特征在于,所述将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,具体包括:

将第i轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器对第i轮次接收到的多个本地节点的模型参数进行加权求和,生成第i轮次的优化模型参数,其中, $i=1,2,\dots,N$,N表示轮次阈值;

接收所述CDN服务器发送的第i轮次的优化模型参数,调整第i轮次训练后的本地PCDN业务识别模型,得到第i+1轮次训练后的本地PCDN业务识别模型;

判断第i+1轮次训练后的本地PCDN业务识别模型是否满足收敛状态;

响应于第i+1轮次训练后的本地PCDN业务识别模型满足收敛状态,将第i+1轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器聚合多个本地节点的模型参数生成PCDN业务识别全局模型。

5. 根据权利要求2所述的PCDN业务的识别方法,其特征在于,BSS业务数据包括用户ID以及以下至少之一:用户实名登记信息、用户的位置数据、套餐类型、消费金额信息、归属地

信息,

在所述将自身所采集的本地业务数据进行加密后发送至CDN服务器之前,还包括:

采集本地业务数据,其中,所述本地业务数据包括用户ID以及以下至少之一:用户业务数据、用户信令数据、用户业务频次、业务日均上下行流量、业务流向、路由信息。

6.一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的CDN服务器,其特征在于,PCDN业务的识别方法包括:

基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型;

将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

7.根据权利要求6所述的PCDN业务的识别方法,其特征在于,在所述基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型之前,还包括:

接收多个本地节点发送的加密后的本地业务数据,其中,本地业务数据用于表征本地节点采集的业务数据;

获取BSS节点加密后的BSS业务数据;

将加密后的本地业务数据与加密后的BSS业务数据基于用户ID匹配对齐,得到PCDN业务识别特征数据,其中,PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据;

向各本地节点发送PCDN业务识别特征数据,以使各本地节点将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据训练本地PCDN业务识别模型。

8.一种本地节点,应用于目标区域CDN,其特征在于,包括第一建立模块、训练模块和识别模块,

第一建立模块,用于基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,

训练模块,与第一建立模块连接,用于叠加业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,

识别模块,与训练模块连接,用于基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

9.一种CDN服务器,应用于目标区域CDN,其特征在于,包括第二建立模块和第二发送模块,

第二建立模块,用于基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型,

第二发送模块,与第二建立模块连接,用于将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

10.一种电子设备,其特征在于,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器被设置为运行所述计算机程序以实现如权利要求1至5任一项所述的一种PCDN业务的识别方法或权利要求6至7任一项所述的一种PCDN业务的识别方法。

11.一种计算机可读存储介质,其上存储有计算机程序,其特征在于,所述计算机程序

被处理器执行时,实现如权利要求1至5任一项所述的一种PCDN业务的识别方法或权利要求6至7任一项所述的一种PCDN业务的识别方法。

PCDN业务识别方法、本地节点、CDN服务器、设备及存储介质

技术领域

[0001] 本发明涉及网络检测技术领域,尤其涉及一种PCDN业务的识别方法、本地节点、CDN服务器、电子设备及计算机可读存储介质。

背景技术

[0002] 由于PCDN(Peer to Peer Content Delivery Network,点对点内容分发网络,也称为基于P2P技术的CDN)技术直接把内容放在用户的手机终端上,或者路由器上,手机终端或路由器变成规模更大的“内容源”,给运营商以及用户的权益带来了巨大的风险和损失,具体如下:

[0003] (1) 终端变成PCDN节点,与其它设备存在数据交互,可能导致隐私数据的泄露,或产生安全漏洞,遭到恶意攻击。

[0004] (2) 海量的PCDN流量,对运营商骨干传输网络,形成巨大压力,会影响整体用户的业务感知和体验。

[0005] 运营商传统的PCDN业务甄别手段是通过采用DPI (deep

[0006] packet inspection,深度数据包检测) 流量识别以及流量异常检测方式来甄别PCDN业务,其中,DPI流量识别以及流量异常检测方式包括规则识别和AI识别两种方法。规则识别算法通常是结合宽带用户的历史行为表现,利用特定定义的规则进行识别,而AI识别方法是通过对宽带用户的历史行为数据进行建模,从而通过模型对用户实时行为进行判断。

[0007] 然而,运营商传统的PCDN业务甄别手段需要集中地采集和分析处理数据,算力要求高,但PCDN业务甄别准确度却不高,还需要数据频繁地进行传输和交互,存在数据安全风险。

发明内容

[0008] 本发明所要解决的技术问题是针对现有技术的上述不足,提供一种PCDN业务的识别方法、本地节点、CDN服务器、电子设备及计算机可读存储介质,该方法可以实现高效且精准地PCDN业务识别,提升PCDN业务识别模型的数据分析能力,提升共性特征数据和个性化特征数据分析处理的准确性和效率,减少维护和更新成本的同时,实现更加精准和个性化的PCDN业务识别服务。

[0009] 第一方面,本发明提供一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的多个本地节点中的任意一个本地节点,PCDN业务的识别方法包括:基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型;叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型;基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0010] 优选地,在所述基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联

合建立PCDN业务识别全局模型之前,PCDN业务的识别方法还包括:将自身所采集的本地业务数据进行加密后发送至CDN服务器;接收CDN服务器发送的PCDN业务识别特征数据,其中,所述PCDN业务识别特征数据由所述CDN服务器将加密后的本地业务数据与BSS节点加密后的BSS业务数据基于用户ID匹配对齐得到,所述PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据;将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据。

[0011] 优选地,所述基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,具体包括:基于PCDN业务识别共性特征数据训练PCDN业务识别模型,得到本地PCDN业务识别模型;将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,N为正整数;接收CDN服务器发送的PCDN业务识别全局模型。

[0012] 优选地,所述将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,具体包括:将第i轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器对第i轮次接收到的多个本地节点的模型参数进行加权求和,生成第i轮次的优化模型参数,其中, $i=1,2,\dots,N$,N表示轮次阈值;接收所述CDN服务器发送的第i轮次的优化模型参数,调整第i轮次训练后的本地PCDN业务识别模型,得到第i+1轮次训练后的本地PCDN业务识别模型;判断第i+1轮次训练后的本地PCDN业务识别模型是否满足收敛状态;响应于第i+1轮次训练后的本地PCDN业务识别模型满足收敛状态,将第i+1轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器聚合多个本地节点的模型参数生成PCDN业务识别全局模型。

[0013] 优选地,BSS业务数据包括用户ID以及以下至少之一:用户实名登记信息、用户的位置数据、套餐类型、消费金额信息、归属地信息,在所述将自身所采集的本地业务数据进行加密后发送至CDN服务器之前,PCDN业务的识别方法还包括:采集本地业务数据,其中,所述本地业务数据包括用户ID以及以下至少之一:用户业务数据、用户信令数据、用户业务频次、业务日均上下行流量、业务流向、路由信息。

[0014] 第二方面,本发明还提供一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的CDN服务器,PCDN业务的识别方法包括:基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型;将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0015] 优选地,在所述基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型之前,PCDN业务的识别方法还包括:接收多个本地节点发送的加密后的本地业务数据,其中,本地业务数据用于表征本地节点采集的业务数据;获取BSS节点加密后的BSS业务数据;将加密后的本地业务数据与加密后的BSS业务数据基于用户ID匹配对齐,得到PCDN业务识别特征数据,其中,PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据;向各本地节点发送PCDN业务识别特征数据,以使各本地节点将PCDN业务

识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据训练本地PCDN业务识别模型。

[0016] 第三方面,本发明还提供一种本地节点,应用于目标区域CDN,包括第一建立模块、训练模块和识别模块,第一建立模块,用于基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,训练模块,与第一建立模块连接,用于叠加业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,识别模块,与训练模块连接,用于基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0017] 第四方面,本发明还提供一种CDN服务器,应用于目标区域CDN,包括第二建立模块和第二发送模块,第二建立模块,用于基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型,第二发送模块,与第二建立模块连接,用于将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0018] 第五方面,本发明还提供一种电子设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器被设置为运行所述计算机程序以实现上述第一方面或第二方面提供的PCDN业务的识别方法。

[0019] 第六方面,本发明还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时,实现上述第一方面或第二方面提供的PCDN业务的识别方法。

[0020] 本发明提供了一种PCDN业务的识别方法、本地节点、CDN服务器、电子设备及计算机可读存储介质,通过CDN服务器联合目标区域CDN的多个本地节点的PCDN业务识别共性特征数据,快速建立泛化能力强的PCDN业务识别全局模型,以分发给目标区域CDN的多个本地节点;再通过目标区域CDN的各个本地节点的PCDN业务识别个性化特征数据,训练PCDN业务识别全局模型,生成满足本地节点自身网络流量识别需求的PCDN业务识别个性化模型,其中,PCDN业务识别个性化模型不仅能快速分析处理PCDN业务识别共性特征数据,也能快速分析处理PCDN业务识别个性化特征数据,进而能快速且精准地识别出本地节点自身的PCDN业务用户。因此,本发明能够实现高效且精准地PCDN业务识别,提升PCDN业务识别模型的数据分析能力,提升共性特征数据和个性化特征数据分析处理的准确性和效率,减少维护和更新成本的同时,实现更加精准和个性化的PCDN业务识别服务。

附图说明

- [0021] 图1为本发明实施例1的一种PCDN业务的识别方法的流程图;
- [0022] 图2为本发明实施例1中本地业务数据与BSS业务数据匹配对齐的示例图;
- [0023] 图3为本发明实施例2的一种PCDN业务的识别方法的流程图;
- [0024] 图4为本发明实施例3的一种PCDN业务的识别方法的流程图;
- [0025] 图5为本发明实施例4的一种PCDN业务的识别装置的结构示意图;
- [0026] 图6为本发明实施例5的一种PCDN业务的识别装置的结构示意图。

具体实施方式

[0027] 为使本领域技术人员更好地理解本发明的技术方案,下面将结合附图对本发明实施方式作进一步地详细描述。

[0028] 可以理解的是,此处描述的具体实施例和附图仅仅用于解释本发明,而非对本发明的限定。

[0029] 可以理解的是,在不冲突的情况下,本发明中的各实施例及实施例中的各特征可相互组合。

[0030] 可以理解的是,为便于描述,本发明的附图中仅示出了与本发明相关的部分,而与本发明无关的部分未在附图中示出。

[0031] 可以理解的是,本发明的实施例中所涉及的每个单元、模块可仅对应一个实体结构,也可由多个实体结构组成,或者,多个单元、模块也可集成为一个实体结构。

[0032] 可以理解的是,在不冲突的情况下,本发明的流程图和框图中标注的功能、步骤可按照不同于附图中所标注的顺序发生。

[0033] 可以理解的是,本发明的流程图和框图中,示出了按照本发明各实施例的系统、装置、设备、方法的可能实现的体系架构、功能和操作。其中,流程图或框图中的每个方框可代表一个单元、模块、程序段、代码,其包含用于实现规定的功能的可执行指令。而且,框图和流程图中的每个方框或方框的组合,可用实现规定的功能的基于硬件的系统实现,也可用硬件与计算机指令的组合来实现。

[0034] 可以理解的是,本发明实施例中所涉及的单元、模块可通过软件的方式实现,也可通过硬件的方式来实现,例如单元、模块可位于处理器中。

[0035] 实施例1:

[0036] 如图1所示,本实施例提供一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的多个本地节点中的任意一个本地节点。

[0037] 本实施例中,内容分发网络CDN(Content Delivery Network,内容分发网络)包括多个区域,每个区域内存在若干个客户端,即本地节点。CDN是运营商正规的边缘计算网络,是构建在现有互联网基础之上的一层智能虚拟网络,主要通过CDN各个区域对应部署一个CDN服务器作为自身区域的中心服务器,CDN服务器负责自身区域内本地节点的数据分发和采集。CDN服务器采集和分析自身区域内所有本地节点的DPI数据以及流量监控数据,生成CDN服务器的源站内容,并将CDN服务器的源站内容分发至自身区域内所有的本地节点,使本地节点的用户可以就近获得所需的内容,源站内容包括但不限于:DPI数据以及流量监控数据的分析结果信息和存储信息。

[0038] 需要说明的是,随着光通信技术的不断发展,运营商的宽带业务不仅是数字经济发展的先决条件,更是驱动数字经济快速成长的关键引擎之一,运营商的宽带业务有力地支撑了数字经济从消费互联网向产业互联网的延伸和融合,促进了经济社会的高质量发展。具体体现在以下几个方面:①运营商通过持续投资和升级宽带网络基础设施,实现了全国范围内的高覆盖率和高速率接入,为数字经济提供了坚实的物理层支撑。②宽带业务的快速发展显著提高了网络速度和稳定性,降低了单位带宽资费,使得企业和个人用户能以更低的成本享受到更优质的网络服务,从而促进了在线商务、远程办公、在线教育、视频流媒体等数字经济业态的繁荣。③高速宽带网络能够支持大数据、云计算、物联网、5G(5th

Generation Mobile Communication Technology, 第五代移动通信技术)、人工智能的新一代信息技术的应用,促使传统行业实现数字化转型,形成新的经济增长点。④数字经济的核心在于数据流动与处理,运营商提供的宽带服务保障了数据的高效传输与存储,有利于企业挖掘数据价值、创新商业模式,同时也能满足消费者对于实时、高清、互动的数字化服务的需求。⑤通过扩大城乡宽带覆盖范围,降低接入门槛,运营商有助于消除地域间的信息不对称和数字鸿沟,共享数字化带来的经济和社会效益。⑥高质量的宽带服务也有力地支持了电子政务、智慧城市、远程医疗、应急响应公共服务领域的数字化进程,提升了社会治理效能和公共服务水平。

[0039] 然而,PCDN业务也给运营商和用户在宽带业务上的收益带来了巨大的风险和损失。因此,迫切需要运营商对PCDN业务进行严格排查。

[0040] PCDN业务的识别方法,包括:

[0041] S101,基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型。

[0042] 本实施例中,通过中心化的联邦学习框架,让目标区域CDN网络的各个本地节点的本地业务数据和集中采集数据都保留在目标区域CDN网络中,通过多个参与者联合数据建立虚拟的共有模型,来实现PCDN业务的识别算法。

[0043] 需要说明的是,联邦学习算法包括但不限于:联邦平均算法(Federated Averaging, FedAvg)、联邦随机梯度下降(Federated stochastic gradient descent, FedSGD)、联邦近端优化(Federalized Proximal Algorithm, FedProx)、联邦仿射模型(Federated Recurrent Neural Network, Federated RNN)、分层联邦学习(Hierarchical Federated Learning)、安全多方计算(Secure Multi-Party Computation, SMPC)、差分隐私(Differential Privacy)和联邦迁移学习(Federated Transfer Learning)。本实施例以联邦学习算法为联邦平均算法为例。

[0044] 需要说明的是,联邦平均算法通过在每个设备上局部训练模型并将模型权重发送到中心服务器,然后进行平均更新来聚合模型。联邦随机梯度下降通过在每一轮中只使用单个设备进行梯度更新,并将更新结果发送给中心服务器。联邦近端优化通过在FedAvg的基础上加入一个正则项,以解决设备间数据分布不均的问题。联邦仿射模型指用于处理序列数据的联邦学习模型,通常涉及循环神经网络(Recurrent Neural Network, RNN)。分层联邦学习适用于多层次网络架构,如设备与边缘服务器之间的联邦学习。安全多方计算通过密码学方法确保数据隐私,允许各方共同计算模型而无需暴露数据。差分隐私在模型训练过程中加入噪声,保护个体数据的隐私。联邦迁移学习结合迁移学习的思想,允许在不同的数据分布上进行模型训练。

[0045] 可选地,在S101:基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型之前,PCDN业务的识别方法还包括步骤S104-步骤S106:

[0046] S104,将自身所采集的本地业务数据进行加密后发送至CDN服务器。

[0047] 可选地,在S104:将自身所采集的本地业务数据进行加密后发送至CDN服务器之前,PCDN业务的识别方法还包括:

[0048] S107,采集本地业务数据,其中,本地业务数据包括用户ID以及以下至少之一:用户业务数据、用户信令数据、用户业务频次、业务日均上下行流量、业务流向、路由信息。

[0049] 本实施例中,参与联邦学习建模的数据包括各个本地节点自身所采集的本地业务数据。各个本地节点在积累自身的本地业务数据时,不仅会标识业务数据的用户ID,本实施例以用户ID为用户电话号码为例,形成以用户电话号码区分的本地业务数据集合。而对于本地节点采集自身的本地业务数据时,还会给自身的本地业务数据标注自身的数据来源标识,数据来源标识用于表征本地业务数据的来源节点。

[0050] 本实施例各个本地节点所采集的本地业务数据通常为各类别(企业、商圈、住宅)的用户宽带业务数据,即用户O域信息,用户O域信息包括但不限于:用户业务数据、用户信令数据、用户业务频次、业务日均上下行流量、业务流向、路由信息。

[0051] 具体地,BSS业务数据包括用户ID以及以下至少之一:用户实名登记信息、用户的位置数据、套餐类型、消费金额信息、归属地信息。

[0052] 本实施例中,参与联邦学习建模的数据还包括运营商BSS(Business Support System,业务支撑系统)节点所采集的BSS业务数据。同理,运营商BSS节点所采集的BSS业务数据也形成以用户电话号码区分的BSS业务数据集合。

[0053] 本实施例BSS节点所采集的BSS业务数据通常为用户B域信息,包括但不限于:用户实名登记信息、用户的位置数据、套餐类型、消费金额信息、归属地信息。

[0054] S105,接收CDN服务器发送的PCDN业务识别特征数据,其中,PCDN业务识别特征数据由CDN服务器将加密后的本地业务数据与BSS节点加密后的BSS业务数据基于用户ID匹配对齐得到,PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据。

[0055] 本实施例中,在采集各个本地节点的本地业务数据和BSS节点的BSS业务数据之后,为了避免泄露数据隐私,各个本地节点和BSS节点分别对自身的本地业务数据和BSS业务数据进行加密。另外,现有技术通常集中收集各个本地节点的本地业务数据,并通过DPI流量识别和流量异常检测方法,如:规则识别和AI识别方法,识别出各个本地节点的PCDN业务用户的业务数据。因此,本实施例在CDN服务器上也可预设DPI流量识别和流量异常检测方法,在各个本地节点将自身的本地业务数据分别发送给CDN服务器,以及BSS节点将BSS业务数据发送给CDN服务器后,CDN服务器可通过预设DPI流量识别和流量异常检测方法,初步识别出各个本地节点的本地业务数据以及BSS节点的BSS业务数据中,各个本地节点以及BSS节点的PCDN业务用户的相应业务数据,并对各个本地节点以及BSS节点的PCDN业务用户的相应业务数据进行标识。

[0056] 进而CDN服务器能在BSS节点和所有本地节点不暴露各自业务数据的前提下确认双方的共有用户,并且不暴露、不互相重叠的用户的业务数据,以便联合共有用户的业务数据,实现所有本地节点与BSS节点公用用户下本地业务数据和BSS业务数据的互相补充,以得到所有本地节点的PCDN业务识别特征数据,具体为:CDN服务器采取Paillier同态加密算法,对齐并融合所有本地节点与BSS节点同一用户ID下的加密后的本地业务数据和BSS业务数据,得到所有本地节点的PCDN业务识别特征数据,如图2所示。本实施例通过CDN服务器匹配和对齐本地业务数据和BSS业务数据,得到PCDN业务识别特征数据以用于后续建立PCDN业务识别全局模型和PCDN业务识别个性化模型,实现多域数据的融合建模和全量业务数据的判别,提高PCDN业务用户群体识别的精准性。通过采取Paillier同态加密算法实现业务数据和BSS业务数据的匹配和对齐,Paillier同态加密算法允许在加密数据上直接进行计

算,而无需解密,从而保护敏感业务数据在传输和处理过程中的隐私,降低数据泄露风险,且不需要频繁的解密和再加密过程,从而降低了运算复杂性,提高了系统的响应速度,能够有效地支持安全和隐私要求,同时保持数据处理的效率和准确性。

[0057] 基于PCDN业务识别特征数据中本地业务数据的数据来源标识,将所有本地节点的PCDN业务识别特征数据划分为各个本地节点的PCDN业务识别特征数据,分别将各个本地节点的PCDN业务识别特征数据分发给各个本地节点。对于同一个区域CDN的不同本地节点而言,PCDN业务识别特征数据的特点为特征重叠多而用户重叠少。因此,在将所有本地节点的PCDN业务识别特征数据划分为各个本地节点的PCDN业务识别特征数据之前,本实施例还通过将同一个区域CDN下所有本地节点的PCDN业务识别特征数据进行聚合,确定同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据,并将同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据用于训练PCDN业务识别全局模型,可以最大化的提升样本数量,提升全局模型准确度。

[0058] 需要说明的是,PCDN业务识别共性特征数据指目标区域CDN下所有本地节点共同关注或普遍存在的特征数据,例如:流量模式数据、用户终端特征数据、地理位置分布数据、内容类型数据和访问频率数据。PCDN业务识别个性化特征数据指目标区域CDN下各个本地节点独特的用户需求和行为特征数据,例如:时间敏感性特征数据、内容缓存策略数据、节点内用户画像数据、用户交互行为数据和本地化内容偏好数据。

[0059] 除了Paillier同态加密算法实现业务数据和BSS业务数据的匹配和对齐之外,本实施例还可通过半同态加密(Partial Homomorphic Encryption,PHE)、层次同态加密(Liveled HE,LHE)、全同态加密(Fully Homomorphic Encryption,FHE)实现业务数据和BSS业务数据的匹配和对齐。

[0060] S106,将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据。

[0061] 本实施例中,在所有本地节点得到自身的PCDN业务识别共性特征数据之后,将同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据用于当前区域CDN下PCDN业务识别全局模型的建立和训练,以得到泛化能力强的当前区域CDN的PCDN业务识别全局模型,在当前区域CDN下本地节点多以及本地节点的网络流量大的情况下,可以捕捉到当前区域CDN下所有本地节点的PCDN业务识别特征数据的整体趋势和特征,快速建模PCDN业务识别全局模型。本实施例通过利用PCDN业务识别共性特征数据,建模PCDN业务识别全局模型,并分发给目标区域CDN下的多个本地节点,利用PCDN业务识别全局模型的泛化能力,快速实现目标区域CDN下的多个本地节点的用于识别PCDN业务用户的模型建立。

[0062] 具体地,S101:基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,包括步骤S1011-步骤S1013:

[0063] S1011,基于PCDN业务识别共性特征数据训练PCDN业务识别模型,得到本地PCDN业务识别模型。

[0064] 本实施例中,PCDN业务识别模型指CDN服务器基于分类算法初始构建的全局模型,CDN服务器将初始构建的全局模型分发给当前区域CDN的所有本地节点,以使各个本地节点分别训练初始构建的全局模型,得到本地PCDN业务识别模型。

[0065] 需要说明的是,分类算法包括但不限于:随机森林、神经网络、支持向量机。

[0066] S1012,将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,N为正整数。

[0067] 本实施例中,考虑到BSS业务数据和本地业务数据之间的差异性,以及各个区域CDN下本地节点的本地业务数据各不相同,联邦学习架构中,每个区域CDN都采取相同的训练模型结构显然模型准确度会受影响。因此,本实施例不从BSS节点全局构建一个适配各个区域CDN下所有本地节点的PCDN业务识别全局模型,而是由同一区域CDN下的所有本地节点进行PCDN业务识别全局模型的训练。本实施例通过同一区域CDN下的所有本地节点训练当前区域CDN的PCDN业务识别全局模型,并将其传给当前区域CDN的所有本地节点以进行后续个性化训练,有效提升全局模型的性能质量,进而提升PCDN业务识别全局模型的准确性。

[0068] 另外,由于各个区域CDN下所有本地节点的PCDN业务识别共性特征数据中,各个本地节点的数据量规模不同,数据特征的总体分布不同,因此,不同本地节点的PCDN业务识别共性特征对全局模型的贡献度也不同。因此,本实施例通过在本节点利用PCDN业务识别共性特征数据训练本地PCDN业务识别模型,再通过CDN服务器基于联邦学习对多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,通过在本节点进行模型训练,可以有效保护用户数据隐私,在多个本地节点并行训练可以显著提高模型训练的速度,减少CDN服务器的计算负担,通过聚合模型参数而不是直接传输数据,能够有效减少网络带宽消耗,提高系统的整体性能,通过不断地更新和优化聚合模型,能够提高PCDN业务识别的准确性和鲁棒性。

[0069] 具体地,S1012:将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,包括:将第i轮次训练后的本地PCDN业务识别模型的模型参数发送给CDN服务器,以使CDN服务器对第i轮次接收到的多个本地节点的模型参数进行加权求和,生成第i轮次的优化模型参数,其中, $i=1,2,\dots,N$,N表示轮次阈值;接收CDN服务器发送的第i轮次的优化模型参数,调整第i轮次训练后的本地PCDN业务识别模型,得到第i+1轮次训练后的本地PCDN业务识别模型;判断第i+1轮次训练后的本地PCDN业务识别模型是否满足收敛状态;响应于第i+1轮次训练后的本地PCDN业务识别模型满足收敛状态,将第i+1轮次训练后的本地PCDN业务识别模型的模型参数发送给CDN服务器,以使CDN服务器聚合多个本地节点的模型参数生成PCDN业务识别全局模型。

[0070] 本实施例中,在本节点利用PCDN业务识别共性特征数据训练本地PCDN业务识别模型的同时,本地节点会计算自身的本地PCDN业务识别模型的模型参数,并将模型参数发送给本地节点所在区域CDN的CDN服务器,其中,模型参数包括但不限于:梯度信息、统计信息、加密信息、元数据、收敛速度。本实施例以模型参数为损失函数的梯度信息 $\nabla\omega$ 为例,损失函数的梯度信息 $\nabla\omega$ 用于表征最小化损失函数时模型参数 ω 的调整需求更新量。

[0071] 本地节点所在区域CDN的CDN服务器接收本地节点所在区域CDN的所有本地节点的模型参数,以构建本地节点所在区域CDN的PCDN业务识别全局模型,具体为:①对本地节点所在区域CDN的所有本地节点的模型参数进行加权求和,计算本地节点所有区域CDN的所有本地节点的优化模型参数;②将所有本地节点的优化模型参数分发给本地节点所在区域CDN的所有本地节点,以更新所有本地节点的本地PCDN识别模型;③针对不同轮次获得的模

型参数,重复执行①和②,直至所有本地节点的本地PCDN识别模型满足收敛状态,或者达到重复执行次数阈值,聚合本地节点所在区域CDN的所有本地节点的模型参数,生成PCDN业务识别全局模型。本实施例在整个PCDN业务识别全局模型构建过程中,本地节点只与CDN服务器通信,除了共同维护的优化模型参数外,本地节点无法获取自身区域CDN的其他本地节点的任何信息,保障了多个本地节点隐私数据的机密性。

[0072] 计算本地节点所有区域CDN的所有本地节点的优化模型参数,如公式(1)所示:

$$[0073] \quad \Delta \omega = \sum_{i=1}^N \left(\frac{n_i}{\sum_{j=1}^N n_j} \right) \nabla \omega \quad (1),$$

[0074] 其中,N表示本地节点的总数量, n_j 表示第j个本地节点的样本权重, n_i 表示第i个本地节点的样本权重。

[0075] 聚合本地节点所在区域CDN的所有本地节点的模型参数,生成PCDN业务识别全局模型,如公式(2)所示:

$$[0076] \quad F(\theta)_{\text{service}} = \sum_{k=1}^K \frac{n_k}{n} F_k(\theta) \quad (2),$$

[0077] 其中, $F(\theta)_{\text{service}}$ 表示本地节点所在区域CDN的PCDN业务识别全局模型,本地节点所在区域CDN包括K个本地节点, $F_k(\theta)$ 表示第k个本地节点的本地PCDN业务识别模型, n_k 表示第k个本地节点的样本权重,n表示K个本地节点的样本权重的和。因此,PCDN业务识别全局模型的模型参数由各个本地PCDN业务识别模型的模型参数按照样本权重值进行加权求和得到。

[0078] 需要说明的是,收敛过程中的梯度信息变化具体为:在训练的初期,模型的损失可能较高,梯度信息一般较大,指向调整参数的较大方向,模型参数迅速更新,损失值大幅下降。随着训练的进行,梯度的大小可能会逐渐减小,指示损失正在减少,并且模型收敛的速度变慢,参数的更新步伐可能需要调整,以避免过度更新。当模型接近收敛时,梯度可能会非常小,说明模型已达到或接近局部最优点,收敛可能会变得非常缓慢。

[0079] S1013,接收CDN服务器发送的PCDN业务识别全局模型。

[0080] S102,叠加PCDN业务识别个性化特征数据训练PCDN业务识别全局模型,得到PCDN业务识别个性化模型。

[0081] 本实施例中,每个本地节点会基于自身的PCDN业务识别个性化特征数据叠加自身的PCDN业务识别共性特征数据,以训练接收的PCDN业务识别全局模型,即在各个本地节点的PCDN业务识别全局模型的训练数据包括自身的PCDN业务识别共性特征数据基础上,将自身的PCDN业务识别个性化特征数据也接入到PCDN业务识别全局模型的训练数据中用于训练PCDN业务识别全局模型,最终得到一个适应本地化的PCDN业务识别个性化模型。本实施例通过基于本地节点的PCDN业务识别个性化特征数据训练接收的PCDN业务识别全局模型,最终得到一个适应本地化的PCDN业务识别个性化模型,提升PCDN业务识别的准确度。

[0082] S103,基于PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0083] 本实施例提供的一种PCDN业务的识别方法,通过CDN服务器联合目标区域CDN的多个本地节点的PCDN业务识别共性特征数据,快速建立泛化能力强的PCDN业务识别全局模型,以分发给目标区域CDN的多个本地节点;再通过目标区域CDN的各个本地节点的PCDN业

务识别个性化特征数据,训练PCDN业务识别全局模型,生成满足本地节点自身网络流量识别需求的PCDN业务识别个性化模型,其中,PCDN业务识别个性化模型不仅能快速分析处理PCDN业务识别共性特征数据,也能快速分析处理PCDN业务识别个性化特征数据,进而能快速且精准地识别出自身的PCDN业务用户,实现高效且精准地PCDN业务识别,提升PCDN业务识别模型的数据分析能力,提升共性特征数据和个性化特征数据分析处理的准确性和效率,减少维护和更新成本的同时,实现更加精准和个性化的PCDN业务识别服务。另外,通过CDN服务器匹配和对齐本地业务数据和BSS业务数据,得到PCDN业务识别特征数据以用于后续建立PCDN业务识别全局模型和PCDN业务识别个性化模型,实现多域数据的融合建模和全量业务数据的判别,提高PCDN业务用户群体识别的精准性。通过采取Paillier同态加密算法实现业务数据和BSS业务数据的匹配和对齐,Paillier同态加密算法允许在加密数据上直接进行计算,而无需解密,从而保护敏感业务数据在传输和处理过程中的隐私,降低数据泄露风险,且不需要频繁的解密和再加密过程,从而降低了运算复杂性,提高了系统的响应速度,能够有效地支持安全和隐私要求,同时保持数据处理的效率和准确性。通过将同一个区域CDN下所有本地节点的PCDN业务识别特征数据进行聚合,确定同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据,并将同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据用于训练PCDN业务识别全局模型,可以最大化的提升样本数量,提升全局模型准确度。通过利用PCDN业务识别共性特征数据,建模PCDN业务识别全局模型,并分发给目标区域CDN下的多个本地节点,利用PCDN业务识别全局模型的泛化能力,快速实现目标区域CDN下的多个本地节点的用于识别PCDN业务用户的模型建立。通过同一区域CDN下的所有本地节点训练当前区域CDN的PCDN业务识别全局模型,并将其传给当前区域CDN的所有本地节点以进行后续个性化训练,有效提升全局模型的性能质量,进而提升PCDN业务识别全局模型的准确性。通过在本地节点利用PCDN业务识别共性特征数据训练本地PCDN业务识别模型,再通过CDN服务器基于联邦学习对多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,通过在本地节点进行模型训练,可以有效保护用户数据隐私,在多个本地节点并行训练可以显著提高模型训练的速度,减少CDN服务器的计算负担,通过聚合模型参数而不是直接传输数据,能够有效减少网络带宽消耗,提高系统的整体性能,通过不断地更新和优化聚合模型,能够提高PCDN业务识别的准确性和鲁棒性。在整个PCDN业务识别全局模型构建过程中,本地节点只与CDN服务器通信,除了共同维护的优化模型参数外,本地节点无法获取自身区域CDN的其他本地节点的任何信息,保障了多个本地节点隐私数据的机密性。通过基于本地节点的PCDN业务识别个性化特征数据训练接收的PCDN业务识别全局模型,最终得到一个适应本地化的PCDN业务识别个性化模型,提升PCDN业务识别的准确度。

[0084] 实施例2:

[0085] 如图3所示,本实施例提供一种点对点内容分发网络PCDN业务的识别方法,应用于目标区域内容分发网络CDN的CDN服务器。

[0086] 本实施例中,内容分发网络CDN包括多个区域,每个区域内存在若干个客户端,即本地节点。CDN是运营商正规的边缘计算网络,是构建在现有互联网基础之上的一层智能虚拟网络,主要通过CDN各个区域对应部署一个CDN服务器作为自身区域的中心服务器,CDN服务器负责自身区域内本地节点的数据分发和采集。CDN服务器采集和分析自身区域内所

有本地节点的DPI数据以及流量监控数据,生成CDN服务器的源站内容,并将CDN服务器的源站内容分发至自身区域内所有的本地节点,使本地节点的用户可以就近获得所需的内容,源站内容包括但不限于:DPI数据以及流量监控数据的分析结果信息和存储信息。

[0087] PCDN业务的识别方法包括:

[0088] S201,基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型。

[0089] 本实施例中,通过中心化的联邦学习框架,让目标区域CDN网络的各个本地节点的本地业务数据和集中采集数据都保留在目标区域CDN网络中,通过多个参与者联合数据建立虚拟的共有模型,来实现PCDN业务的识别算法。

[0090] 在S201:基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型之前,PCDN业务的识别方法还包括步骤S203-步骤S206:

[0091] S203,接收多个本地节点发送的加密后的本地业务数据,其中,本地业务数据用于表征本地节点采集的业务数据。

[0092] 本实施例中,参与联邦学习建模的数据包括各个本地节点自身所采集的本地业务数据。各个本地节点在积累自身的本地业务数据时,不仅会标识业务数据的用户ID,本实施例以用户ID为用户电话号码为例,形成以用户电话号码区分的本地业务数据集合。而对于本地节点采集自身的本地业务数据时,还会给自身的本地业务数据标注自身的数据来源标识,数据来源标识用于表征本地业务数据的来源节点。

[0093] S204,获取BSS节点加密后的BSS业务数据。

[0094] 本实施例中,参与联邦学习建模的数据还包括运营商BSS节点所采集的BSS业务数据。同理,运营商BSS节点所采集的BSS业务数据也形成以用户电话号码区分的BSS业务数据集合。

[0095] S205,将加密后的本地业务数据与加密后的BSS业务数据基于用户ID匹配对齐,得到PCDN业务识别特征数据,其中,PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据。

[0096] 本实施例中,在采集以及标识各个本地节点的本地业务数据和BSS节点的BSS业务数据之后,为了避免泄露数据隐私,各个本地节点和BSS节点分别对自身的本地业务数据和BSS业务数据进行加密。另外,现有技术通常集中收集各个本地节点的本地业务数据,并通过DPI流量识别和流量异常检测方法,如:规则识别和AI识别方法,识别出各个本地节点的PCDN业务用户的业务数据。因此,本实施例在CDN服务器上也可预设DPI流量识别和流量异常检测方法,在各个本地节点将自身的本地业务数据分别发送给CDN服务器,以及BSS节点将BSS业务数据发送给CDN服务器后,CDN服务器可通过预设DPI流量识别和流量异常检测方法,初步识别出各个本地节点的本地业务数据以及BSS节点的BSS业务数据中,各个本地节点以及BSS节点的PCDN业务用户的相应业务数据,并对各个本地节点以及BSS节点的PCDN业务用户的相应业务数据进行标识。

[0097] 进而CDN服务器能在BSS节点和所有本地节点不暴露各自业务数据的前提下确认双方的共有用户,并且不暴露、不互相重叠的用户的业务数据,以便联合共有用户的业务数据,实现所有本地节点与BSS节点公用用户下本地业务数据和BSS业务数据的互相补充,以得到所有本地节点的PCDN业务识别特征数据,具体为:CDN服务器采取Paillier同态加密算法,对齐并融合所有本地节点与BSS节点同一用户ID下的加密后的本地业务数据和BSS业务

数据,得到所有本地节点的PCDN业务识别特征数据。

[0098] 基于PCDN业务识别特征数据中本地业务数据的数据来源标识,将所有本地节点的PCDN业务识别特征数据划分为各个本地节点的PCDN业务识别特征数据,分别将各个本地节点的PCDN业务识别特征数据分发给各个本地节点。对于同一个区域CDN的不同本地节点而言,PCDN业务识别特征数据的特点为特征重叠多而用户重叠少。因此,在将所有本地节点的PCDN业务识别特征数据划分为各个本地节点的PCDN业务识别特征数据之前,本实施例还通过将同一个区域CDN下所有本地节点的PCDN业务识别特征数据进行聚合,确定同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据,并将同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据用于训练PCDN业务识别全局模型,可以最大化的提升样本数量,提升全局模型准确度。

[0099] S206,向各本地节点发送PCDN业务识别特征数据,以使各本地节点将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据训练本地PCDN业务识别模型。

[0100] 本实施例中,在所有本地节点得到各自的PCDN业务识别共性特征数据之后,将同一个区域CDN下所有本地节点的PCDN业务识别共性特征数据用于当前区域CDN下PCDN业务识别全局模型的建立和训练,以得到泛化能力强的当前区域CDN的PCDN业务识别全局模型,在当前区域CDN下本地节点多以及本地节点的网络流量大的情况下,可以捕捉到当前区域CDN下所有本地节点的PCDN业务识别特征数据的整体趋势和特征,快速建模PCDN业务识别全局模型。

[0101] PCDN业务识别模型指CDN服务器基于分类算法初始构建的全局模型,CDN服务器将初始构建的全局模型分发给当前区域CDN的所有本地节点,以使各个本地节点分别训练初始构建的全局模型,得到本地PCDN业务识别模型。

[0102] S202,将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0103] 本实施例中,考虑到BSS业务数据和本地业务数据之间的差异性,以及各个区域CDN下本地节点的本地业务数据各不相同,联邦学习架构中,每个区域CDN都采取相同的训练模型结构显然模型准确度会受影响。因此,本实施例不从BSS节点全局构建一个适配各个区域CDN下所有本地节点的PCDN业务识别全局模型,而是由同一区域CDN下的所有本地节点进行PCDN业务识别全局模型的训练。

[0104] 另外,由于各个区域CDN下所有本地节点的PCDN业务识别共性特征数据中,各个本地节点的数据量规模不同,数据特征的总体分布不同,因此,不同本地节点的PCDN业务识别共性特征对全局模型的贡献度也不同。因此,本实施例通过在本地节点利用PCDN业务识别共性特征数据训练本地PCDN业务识别模型,再通过CDN服务器基于联邦学习对多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型。

[0105] 在本地节点利用PCDN业务识别共性特征数据训练本地PCDN业务识别模型的同时,本地节点会计算自身的本地PCDN业务识别模型的模型参数,并将模型参数发送给本地节点所在区域CDN的CDN服务器,其中,模型参数包括但不限于:梯度信息、统计信息、加密信息、元数据、收敛速度。本实施例以模型参数为损失函数的梯度信息 $\nabla\omega$ 为例,损失函数的梯度

信息 $\nabla\omega$ 用于表征最小化损失函数时模型参数 ω 的调整需求更新量。

[0106] 本地节点所在区域CDN的CDN服务器接收本地节点所在区域CDN的所有本地节点的模型参数,以构建本地节点所在区域CDN的PCDN业务识别全局模型,具体为:①对本地节点所在区域CDN的所有本地节点的模型参数进行加权求和,计算本地节点所有区域CDN的所有本地节点的优化模型参数;②将所有本地节点的优化模型参数分发给本地节点所在区域CDN的所有本地节点,以更新所有本地节点的本地PCDN识别模型;③针对不同轮次获得的模型参数,重复执行①和②,直至所有本地节点的本地PCDN识别模型满足收敛状态,或者达到重复执行次数阈值,聚合本地节点所在区域CDN的所有本地节点的模型参数,生成PCDN业务识别全局模型。本实施例在整个PCDN业务识别全局模型构建过程中,本地节点只与CDN服务器通信,除了共同维护的优化模型参数外,本地节点无法获取自身区域CDN的其他本地节点的任何信息,保障了多个本地节点隐私数据的机密性。

[0107] 计算本地节点所有区域CDN的所有本地节点的优化模型参数,如公式(1)所示:

$$\Delta\omega = \sum_{i=1}^N \left(\frac{n_i}{\sum_{j=1}^N n_j} \right) \nabla\omega \quad (1), \text{其中, } N \text{ 表示本地节点的总数量, } n_j \text{ 表示第 } j \text{ 个本地}$$

节点的样本权重, n_i 表示第 i 个本地节点的样本权重。聚合本地节点所在区域CDN的所有本地节点的模型参数,生成PCDN业务识别全局模型,如公式(2)所示:

$$F(\theta)_{\text{service}} = \sum_{k=1}^K \frac{n_k}{n} F_k(\theta) \quad (2), \text{其中, } F(\theta)_{\text{service}} \text{ 表示本地节点所在区域CDN的}$$

PCDN业务识别全局模型,本地节点所在区域CDN包括 K 个本地节点, $F_k(\theta)$ 表示第 k 个本地节点的本地PCDN业务识别模型, n_k 表示第 k 个本地节点的样本权重, n 表示 K 个本地节点的样本权重的和。因此,PCDN业务识别全局模型的模型参数由各个本地PCDN业务识别模型的模型参数按照样本权重值进行加权求和得到。

[0108] 每个本地节点会基于自身的PCDN业务识别个性化特征数据叠加自身的PCDN业务识别共性特征数据,以训练接收的PCDN业务识别全局模型,即在各个本地节点的PCDN业务识别全局模型的训练数据包括自身的PCDN业务识别共性特征数据基础上,将自身的PCDN业务识别个性化特征数据也接入到PCDN业务识别全局模型的训练数据中用于训练PCDN业务识别全局模型,最终得到一个适应本地化的PCDN业务识别个性化模型。

[0109] 本实施例提供的一种PCDN业务的识别方法,通过CDN服务器联合目标区域CDN的多个本地节点的PCDN业务识别共性特征数据,快速建立泛化能力强的PCDN业务识别全局模型,以分发给目标区域CDN的多个本地节点;再通过目标区域CDN的各个本地节点的PCDN业务识别个性化特征数据,训练PCDN业务识别全局模型,生成满足本地节点自身网络流量识别需求的PCDN业务识别个性化模型,其中,PCDN业务识别个性化模型不仅能快速分析处理PCDN业务识别共性特征数据,也能快速分析处理PCDN业务识别个性化特征数据,进而能快速且精准地识别出自身的PCDN业务用户,实现高效且精准地PCDN业务识别,提升PCDN业务识别模型的数据分析能力,提升共性特征数据和个性化特征数据分析处理的准确性和效率,减少维护和更新成本的同时,实现更加精准和个性化的PCDN业务识别服务。

[0110] 实施例3:

[0111] 如图4所示,本实施例还提供一种PCDN业务的识别方法,PCDN业务的识别方法包括:

[0112] S1,目标区域CDN下各个本地节点采集自身的本地业务数据,以及CDN服务器获取BSS节点的BSS业务数据。

[0113] 本实施例中,图4中的数据准备即获取各个本地节点的本地业务数据和BSS节点的BSS业务数据。

[0114] S2,CDN服务器匹配和对齐各个本地节点所采集的本地业务数据和BSS业务数据。

[0115] 本实施例中,图4中的本地数据即各个本地节点所采集的本地业务数据,图4中的运营商BSS数据即BSS业务数据。

[0116] S3,基于联邦学习经CDN服务器与多个本地节点联合建立PCDN业务识别全局模型。

[0117] 本实施例中,图4中的全局模型即PCDN业务识别全局模型。

[0118] S4,目标区域CDN下各个本地节点训练PCDN业务识别全局模型,得到PCDN业务识别个性化模型。

[0119] 本实施例中,图4中的个性化模型即PCDN业务识别个性化模型。

[0120] 本实施例提供的一种PCDN业务的识别方法,通过CDN服务器联合目标区域CDN的多个本地节点的PCDN业务识别共性特征数据,快速建立泛化能力强的PCDN业务识别全局模型,以分发给目标区域CDN的多个本地节点;再通过目标区域CDN的各个本地节点的PCDN业务识别个性化特征数据,训练PCDN业务识别全局模型,生成满足本地节点自身网络流量识别需求的PCDN业务识别个性化模型,其中,PCDN业务识别个性化模型不仅能快速分析处理PCDN业务识别共性特征数据,也能快速分析处理PCDN业务识别个性化特征数据,进而能快速且精准地识别出自身的PCDN业务用户,实现高效且精准地PCDN业务识别,提升PCDN业务识别模型的数据分析能力,提升共性特征数据和个性化特征数据分析处理的准确性和效率,减少维护和更新成本的同时,实现更加精准和个性化的PCDN业务识别服务。

[0121] 实施例4:

[0122] 如图5所示,本实施例还提供一种本地节点,应用于目标区域CDN,包括:第一建立模块41、训练模块42和识别模块43,第一建立模块41,用于基于联邦学习经CDN服务器与多个本地节点中的其他本地节点联合建立PCDN业务识别全局模型,训练模块42,与第一建立模块41连接,用于叠加业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,识别模块43,与训练模块42连接,用于基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0123] 可选地,本地节点还包括:第一发送模块44、第一接收模块45和第一确定模块46,第一发送模块44,用于将自身所采集的本地业务数据进行加密后发送至CDN服务器,第一接收模块45,用于接收CDN服务器发送的PCDN业务识别特征数据,其中,所述PCDN业务识别特征数据由所述CDN服务器将加密后的本地业务数据与BSS节点加密后的BSS业务数据基于用户ID匹配对齐得到,所述PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据,第一确定模块46,用于将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据。

[0124] 可选地,本地节点还包括:采集模块47,用于采集本地业务数据,其中,所述本地业务数据包括用户ID以及以下至少之一:用户业务数据、用户信令数据、用户业务频次、业务日均上下行流量、业务流向、路由信息。

[0125] 具体地,第一建立模块41包括:训练单元411、第一发送单元412和第一接收单元

413,训练单元411,用于基于PCDN业务识别共性特征数据训练PCDN业务识别模型,得到本地PCDN业务识别模型,第一发送单元412,用于将训练后的本地PCDN业务识别模型的模型参数发送至CDN服务器,以使CDN服务器基于联邦学习将N轮次接收到的多个本地节点的模型参数进行聚合,生成满足收敛状态的PCDN业务识别全局模型,N为正整数,第一接收单元413,用于接收CDN服务器发送的PCDN业务识别全局模型。

[0126] 具体地,训练模块42包括:第二发送单元421、第二接收单元422、判断单元423和第三发送单元424,第二发送单元421,用于将第i轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器对第i轮次接收到的多个本地节点的模型参数进行加权求和,生成第i轮次的优化模型参数,其中, $i=1,2,\dots,N$,N表示轮次阈值,第二接收单元422,用于接收所述CDN服务器发送的第i轮次的优化模型参数,调整第i轮次训练后的本地PCDN业务识别模型,得到第i+1轮次训练后的本地PCDN业务识别模型,判断单元423,用于判断第i+1轮次训练后的本地PCDN业务识别模型是否满足收敛状态,第三发送单元424,用于响应于第i+1轮次训练后的本地PCDN业务识别模型满足收敛状态,将第i+1轮次训练后的本地PCDN业务识别模型的模型参数发送给所述CDN服务器,以使所述CDN服务器聚合多个本地节点的模型参数生成PCDN业务识别全局模型。

[0127] 可以理解地,上述提供的一种本地节点执行上文所提供的实施例1对应的PCDN业务的识别方法,因此,其所能达到的有益效果可参考上文实施例1的PCDN业务的识别方法对应的方案的有益效果,此处不再赘述。

[0128] 实施例5:

[0129] 如图6所示,本实施例还提供一种CDN服务器,应用于目标区域CDN,包括第二建立模块51和第二发送模块52,第二建立模块51,用于基于联邦学习与多个本地节点联合建立PCDN业务识别全局模型,第二发送模块52,与第二建立模块51连接,用于将PCDN业务识别全局模型发送给各本地节点,以使各本地节点叠加PCDN业务识别个性化特征数据训练所述PCDN业务识别全局模型,得到PCDN业务识别个性化模型,并基于所述PCDN业务识别个性化模型对自身网络流量进行识别,以识别出PCDN业务用户。

[0130] 可选地,CDN服务器还包括:第二接收模块53、获取模块54、对齐模块55和第三发送模块56,第二接收模块53,用于接收多个本地节点发送的加密后的本地业务数据,其中,本地业务数据用于表征本地节点采集的业务数据,获取模块54,用于获取BSS节点加密后的BSS业务数据,对齐模块55,用于将加密后的本地业务数据与加密后的BSS业务数据基于用户ID匹配对齐,得到PCDN业务识别特征数据,其中,PCDN业务识别特征数据包括PCDN业务识别共性特征数据和PCDN业务识别个性化特征数据,第三发送模块56,用于向各本地节点发送PCDN业务识别特征数据,以使各本地节点将PCDN业务识别共性特征数据作为用于建立PCDN业务识别全局模型时的训练数据训练本地PCDN业务识别模型。

[0131] 可以理解地,上述提供的一种CDN服务器执行上文所提供的实施例2对应的PCDN业务的识别方法,因此,其所能达到的有益效果可参考上文实施例2的PCDN业务的识别方法对应的方案的有益效果,此处不再赘述。

[0132] 实施例6:

[0133] 本实施例还提供一种电子设备,包括存储器和处理器,所述存储器中存储有计算机程序,所述处理器被设置为运行所述计算机程序以实现上述实施例1或实施例2中PCDN业

务的识别方法。

[0134] 实施例7:

[0135] 本实施例还提供一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时,实现上述实施例1或实施例2中PCDN业务的识别方法。

[0136] 可以理解的是,以上实施方式仅仅是为了说明本发明的原理而采用的示例性实施方式,然而本发明并不局限于此。对于本领域内的普通技术人员而言,在不脱离本发明的精神和实质的情况下,可以做出各种变型和改进,这些变型和改进也视为本发明的保护范围。

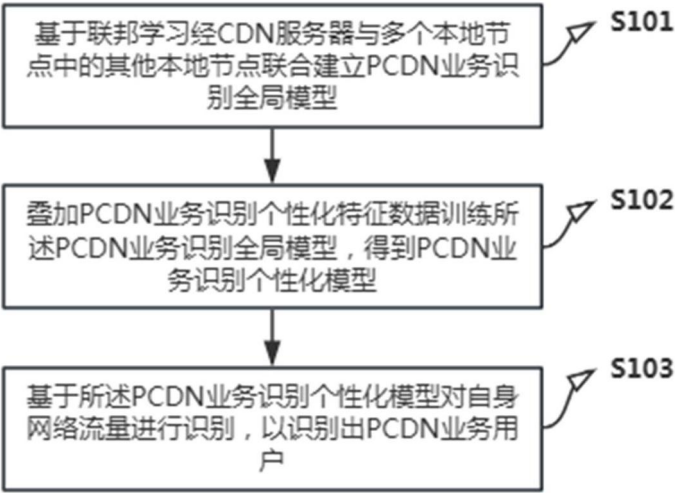


图1

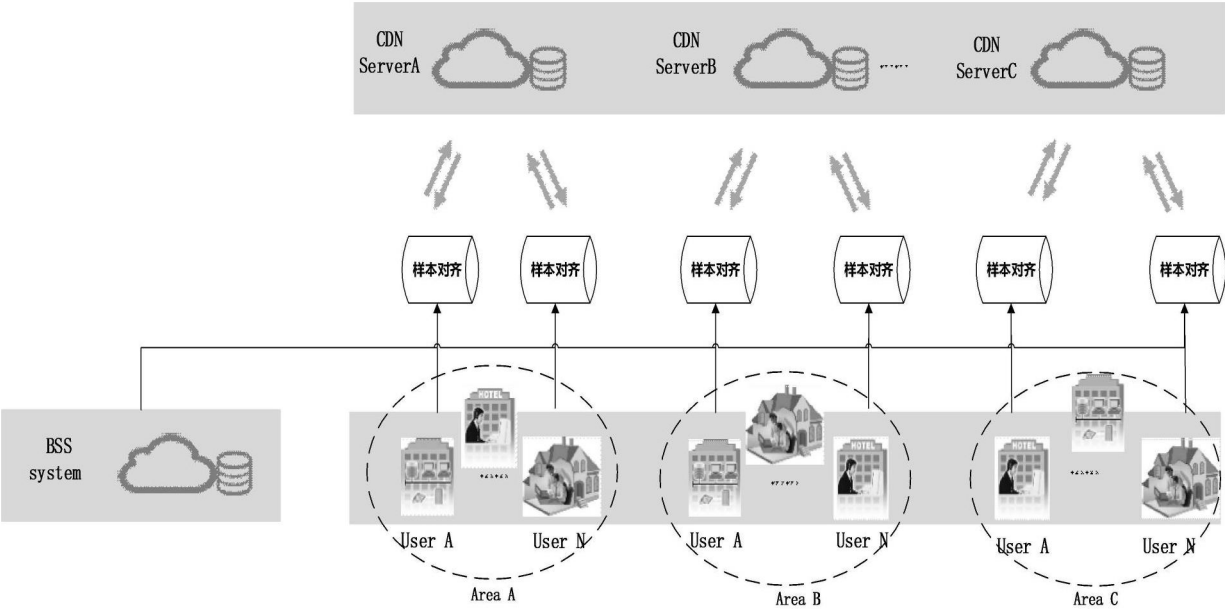


图2

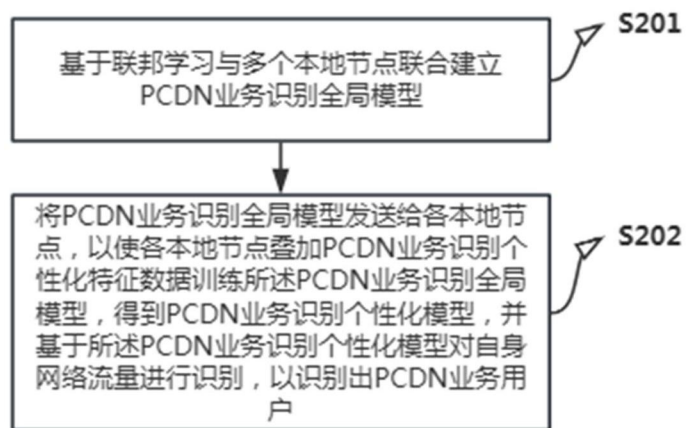


图3

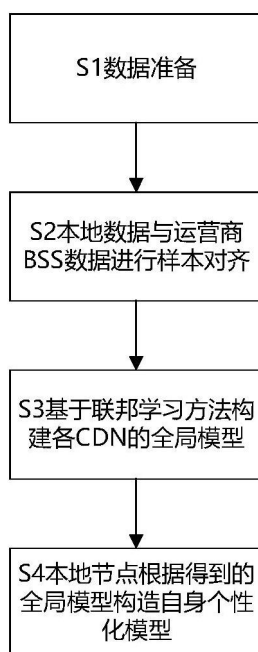


图4



图5

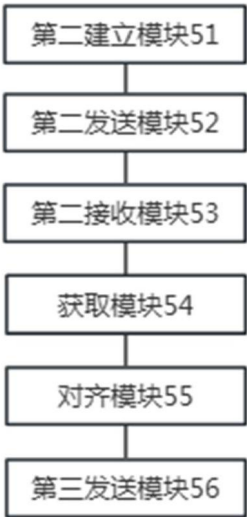


图6