(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115277555 B
(45)授权公告日 2024.01.16

(21)申请号 202210663072.9
(22)申请日 2022.06.13
(65)同一申请的已公布的文献号
申请公布号 CN 115277555 A
(43)申请公布日 2022.11.01
(73)专利权人 香港理工大学深圳研究院
地址 518057 广东省深圳市南山区粤海街道高新技术产业园南区粤兴一道18号香港理工大学产学研大楼205室
(72)发明人 王丹
(74)专利代理机构 深圳市君胜知识产权代理事务所(普通合伙) 44268
专利代理师 李可 王永文
(51)Int.Cl.
H04L 47/10(2022.01)
H04L 47/2441(2022.01)
(56)对比文件
CN 113434859 A,2021.09.24
CN 114528304 A,2022.05.24
CN 113033712 A,2021.06.25
CN 113112027 A,2021.07.13
WO 2021259090 A1,2021.12.30
Shulai Zhang.Dubhe: Towards Data Unbiasedness with Homomorphic Encryption in Federated Learning Client Selection.ACM.2021,全文.
Zibo Wang.FedACS: Federated Skewness Analytics in Heterogeneous Decentralized Data Environments.IEEE.2021,第I-III部分.
审查员 曹彦

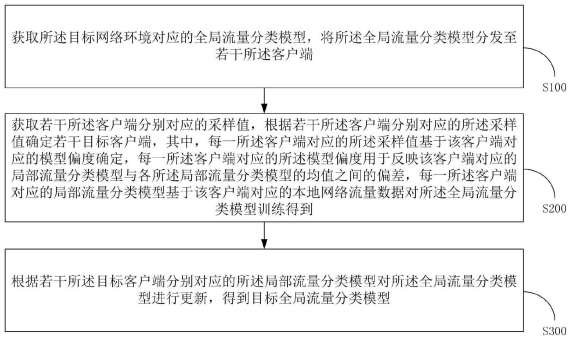
权利要求书3页 说明书10页 附图3页

(54)发明名称

异构环境的网络流量分类方法、装置、终端及存储介质

(57)摘要

本发明公开了异构环境的网络流量分类方法、装置、终端及存储介质,通过将目标网络环境的全局流量分类模分发至各客户端;获取各客户端的采样值,其中,每一客户端的采样值基于该客户端的模型偏度确定;根据各客户端的采样值确定目标客户端;根据各目标客户端对全局流量分类模型进行更新,得到目标全局流量分类模型。本发明通过模型偏度选取目标客户端,再通过各目标客户端的局部流量分类模型更新全局流量分类模型,可以解决现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,导致流量分类精度下降的问题。



CN 115277555 B

1.一种异构环境的网络流量分类方法,其特征在于,所述方法应用于目标网络环境,所述目标网络环境包括若干客户端,若干所述客户端分别对应不同的网络流量数据分布,所述方法包括:

获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端;客户端进行本地训练和模型偏度计算:使用客户端本地的网络流量数据训练接收到的全局流量分类模型,得到客户端的局部流量分类模型,其中,针对每一客户端,计算该客户端当前的局部流量分类模型与各局部流量分类模型的均值的偏差,得到该客户端的模型偏度,保存训练得到的局部流量分类模型;客户端将计算得到的模型偏度上传给服务器;服务器根据上传客户端的模型偏度挑选参与联合计算的客户端;

获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;

根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型;

所述获取若干所述客户端分别对应的采样值,包括:

获取若干所述客户端分别对应的概率分布数据,其中,每一所述客户端对应的所述概率分布数据基于该客户端对应的所述模型偏度确定;

根据若干所述客户端分别对应的所述概率分布数据,确定若干所述客户端分别对应的所述采样值,其中,每一所述客户端对应的所述采样值基于该客户端对应的所述概率分布数据抽样得到;

所述根据若干所述客户端分别对应的所述采样值确定若干目标客户端,包括:

根据所述采样值最大的所述客户端确定候选客户端,判断所述候选客户端对应的总量是否达到目标总量;

当所述候选客户端对应的总量未达到所述目标总量时,针对若干所述客户端中除所述候选客户端之外的客户端,继续执行所述获取若干所述客户端分别对应的采样值,根据所述采样值最大的所述客户端确定候选客户端的步骤,直至所述候选客户端对应的总量达到所述目标总量,得到若干所述候选客户端;

根据若干所述候选客户端,确定预设数量的若干所述目标客户端;

所述根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型,包括:

根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型;

判断所述更新全局流量分类模型是否收敛至目标值;

当所述更新全局流量分类模型未收敛至所述目标值时,根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据;

将所述更新全局流量分类模型作为所述全局流量分类模型,继续执行所述将所述全局流量分类模型分发至若干所述客户端的步骤,直至所述更新全局流量分类模型收敛至所述

目标值,得到所述目标全局流量分类模型;

每一所述概率分布数据为Beta分布函数,每一所述Beta分布函数基于第一分布参数和第二分布参数确定,所述第一分布参数用于反映该Beta分布函数对应的所述客户端未成为所述目标客户端的次数,所述第二分布参数用于反映该Beta分布函数对应的所述客户端成为所述目标客户端的次数;

所述根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据,包括:

对若干所述目标客户端中任意两个客户端,根据预设数值增加所述模型偏度最高的客户端对应的所述第一分布参数,根据所述预设数值增加所述模型偏度最低的客户端所对应的所述第二分布参数。

2.根据权利要求1所述的异构环境的网络流量分类方法,其特征在于,所述根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型,包括:

根据若干所述目标客户端分别对应的所述局部流量分类模型的均值,确定目标模型均值;

根据目标模型均值对所述全局流量分类模型进行更新,得到所述更新全局流量分类模型。

3.一种异构环境的网络流量分类装置,其特征在于,所述装置包括:

分发模块,用于获取目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干客户端;客户端进行本地训练和模型偏度计算:使用客户端本地的网络流量数据训练接收到的全局流量分类模型,得到客户端的局部流量分类模型,其中,针对每一客户端,计算该客户端当前的局部流量分类模型与各局部流量分类模型的均值的偏差,得到该客户端的模型偏度,保存训练得到的局部流量分类模型;客户端将计算得到的模型偏度上传给服务器;服务器根据上传客户端的模型偏度挑选参与联合计算的客户端;

筛选模块,用于获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;

更新模块,用于根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型;

所述获取若干所述客户端分别对应的采样值,包括:

获取若干所述客户端分别对应的概率分布数据,其中,每一所述客户端对应的所述概率分布数据基于该客户端对应的所述模型偏度确定;

根据若干所述客户端分别对应的所述概率分布数据,确定若干所述客户端分别对应的所述采样值,其中,每一所述客户端对应的所述采样值基于该客户端对应的所述概率分布数据抽样得到;

所述根据若干所述客户端分别对应的所述采样值确定若干目标客户端,包括:

根据所述采样值最大的所述客户端确定候选客户端,判断所述候选客户端对应的总量是否达到目标总量;

当所述候选客户端对应的总量未达到所述目标总量时,针对若干所述客户端中除所述候选客户端之外的客户端,继续执行所述获取若干所述客户端分别对应的采样值,根据所述采样值最大的所述客户端确定候选客户端的步骤,直至所述候选客户端对应的总量达到所述目标总量,得到若干所述候选客户端;

根据若干所述候选客户端,确定预设数量的若干所述目标客户端;

所述根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型,包括:

根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型;

判断所述更新全局流量分类模型是否收敛至目标值;

当所述更新全局流量分类模型未收敛至所述目标值时,根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据;

将所述更新全局流量分类模型作为所述全局流量分类模型,继续执行所述将所述全局流量分类模型分发至若干所述客户端的步骤,直至所述更新全局流量分类模型收敛至所述目标值,得到所述目标全局流量分类模型;

每一所述概率分布数据为Beta分布函数,每一所述Beta分布函数基于第一分布参数和第二分布参数确定,所述第一分布参数用于反映该Beta分布函数对应的所述客户端未成为所述目标客户端的次数,所述第二分布参数用于反映该Beta分布函数对应的所述客户端成为所述目标客户端的次数;

所述根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据,包括:

对若干所述目标客户端中任意两个客户端,根据预设数值增加所述模型偏度最高的客户端对应的所述第一分布参数,根据所述预设数值增加所述模型偏度最低的客户端所对应的所述第二分布参数。

4.一种终端,其特征在于,所述终端包括有存储器和一个或者一个以上处理器;所述存储器存储有一个或者一个以上的程序;所述程序包含用于执行如权利要求1-2中任一所述的异构环境的网络流量分类方法的指令;所述处理器用于执行所述程序。

5.一种计算机可读存储介质,其上存储有多条指令,其特征在于,所述指令适用于由处理器加载并执行,以实现上述权利要求1-2任一所述的异构环境的网络流量分类方法的步骤。

异构环境的网络流量分类方法、装置、终端及存储介质

技术领域

[0001] 本发明涉及互联网技术领域,尤其涉及的是一种异构环境的网络流量分类方法、装置、终端及存储介质。

背景技术

[0002] 移动网络流量分类是将边缘设备产生的网络流量数据分为不同的类别,在流量工程(Traffic Engineering)、网络入侵检测、网络监控和服务质量(QoS)等网络安全和管理应用中发挥着重要作用。传统的网络流量分类方法主要可分为三类:基于端口的方法、基于有效负载的方法和基于机器学习的方法。随着人工智能(Artificial Intelligence)技术的快速发展,深度学习(Deep Learning)等机器学习方法越来越受欢迎。但是所有传统的方法都是从不同的边缘设备收集移动流量数据,并集中处理这些流量数据,这种处理方式存在隐私泄露和威胁数据安全等问题。因此,基于隐私保护的网络流量分类研究已引起全世界的关注。

[0003] 基于联邦学习的移动网络流量分类是利用本地客户端中的原始流量数据来解决了流量分类中的隐私和安全问题,在保护用户隐私方面显示出了巨大的潜力。然而现有的联邦学习流量分类解决方案,主要是对局部流量分类模型进行平均,以获得全局模型。然而,由于应用服务和用户偏好的不同,在不同的客户端中的流量类别分布是不同的,即存在流量数据异构性(或非独立相同分布(非IID)数据)。因此现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,最终导致流量分类精度下降。

[0004] 因此,现有技术还有待改进和发展。

发明内容

[0005] 本发明要解决的技术问题在于,针对现有技术的上述缺陷,提供一种异构环境的网络流量分类方法、装置、终端及存储介质,旨在解决现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,导致流量分类精度下降的问题。

[0006] 本发明解决问题所采用的技术方案如下:

[0007] 第一方面,本发明实施例提供一种异构环境的网络流量分类方法,其中,所述方法应用于目标网络环境,所述目标网络环境包括若干客户端,若干所述客户端分别对应不同的网络流量数据分布,所述方法包括:

[0008] 获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端;

[0009] 获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分

类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;

[0010] 根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。

[0011] 在一种实施方式中,所述获取若干所述客户端分别对应的采样值,包括:

[0012] 获取若干所述客户端分别对应的概率分布数据,其中,每一所述客户端对应的所述概率分布数据基于该客户端对应的所述模型偏度确定;

[0013] 根据若干所述客户端分别对应的所述概率分布数据,确定若干所述客户端分别对应的所述采样值,其中,每一所述客户端对应的所述采样值基于该客户端对应的所述概率分布数据抽样得到。

[0014] 在一种实施方式中,所述根据若干所述客户端分别对应的所述采样值确定若干目标客户端,包括:

[0015] 根据所述采样值最大的所述客户端确定候选客户端,判断所述候选客户端对应的总量是否达到目标总量;

[0016] 当所述候选客户端对应的总量未达到所述目标总量时,针对若干所述客户端中除所述候选客户端之外的客户端,继续执行所述获取各所述客户端分别对应的采样值,根据所述采样值最大的所述客户端确定候选客户端的步骤,直至所述候选客户端对应的总量达到所述目标总量,得到若干所述候选客户端;

[0017] 根据若干所述候选客户端,确定预设数量的若干所述目标客户端。

[0018] 在一种实施方式中,所述根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型,包括:

[0019] 根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型;

[0020] 判断所述更新全局流量分类模型是否收敛至目标值;

[0021] 当所述更新全局流量分类模型未收敛至所述目标值时,根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据;

[0022] 将所述更新全局流量分类模型作为所述全局流量分类模型,继续执行所述将所述全局流量分类模型分发至若干所述客户端的步骤,直至所述更新全局流量分类模型收敛至所述目标值,得到所述目标全局流量分类模型。

[0023] 在一种实施方式中,所述根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型,包括:

[0024] 根据若干所述目标客户端分别对应的所述局部流量分类模型的均值,确定目标模型均值;

[0025] 根据目标模型均值对所述全局流量分类模型进行更新,得到所述更新全局流量分类模型。

[0026] 在一种实施方式中,每一所述概率分布数据为Beta分布函数,每一所述Beta分布函数基于第一分布参数和第二分布参数确定,所述第一分布参数用于反映该Beta分布函数对应的所述客户端未成为所述目标客户端的次数,所述第二分布参数用于反映该Beta分布函数对应的所述客户端成为所述目标客户端的次数。

[0027] 在一种实施方式中,所述根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据,包括:

[0028] 对若干所述目标客户端中任意两个客户端,根据预设数值增加所述模型偏度最高的客户端对应的所述第一分布参数,根据所述预设数值增加所述模型偏度最低的客户端所对应的所述第二分布参数。

[0029] 第二方面,本发明实施例还提供一种异构环境的网络流量分类装置,其中,所述装置包括:

[0030] 分发模块,用于获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端;

[0031] 筛选模块,用于获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;

[0032] 更新模块,用于根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。

[0033] 第三方面,本发明实施例还提供一种终端,其中,所述终端包括有存储器和一个或者一个以上处理器;所述存储器存储有一个或者一个以上的程序;所述程序包含用于执行如上述任一所述的异构环境的网络流量分类方法的指令;所述处理器用于执行所述程序。

[0034] 第四方面,本发明实施例还提供一种计算机可读存储介质,其上存储有多条指令,其中,所述指令适用于由处理器加载并执行,以实现上述任一所述的异构环境的网络流量分类方法的步骤。

[0035] 本发明的有益效果:本发明实施例通过将目标网络环境的全局流量分类模分发至各客户端;获取各客户端的采样值,其中,每一客户端的采样值基于该客户端的模型偏度确定;根据各客户端的采样值确定目标客户端;根据各目标客户端对全局流量分类模型进行更新,得到目标全局流量分类模型。本发明通过模型偏度选取目标客户端,再通过各目标客户端的局部流量分类模型更新全局流量分类模型,可以解决现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,导致流量分类精度下降的问题。

附图说明

[0036] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述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发明中记载的一些实施例,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。

[0037] 图1是本发明实施例提供的异构环境的网络流量分类方法的流程示意图。

[0038] 图2是本发明实施例提供的FEAT算法的流程示意图。

[0039] 图3是本发明实施例提供的异构环境的网络流量分类装置的模块示意图。

[0040] 图4是本发明实施例提供的终端的原理框图。

具体实施方式

[0041] 本发明公开了异构环境的网络流量分类方法、装置、终端及存储介质，为使本发明的目的、技术方案及效果更加清楚、明确，以下参照附图并举实施例对本发明进一步详细说明。应当理解，此处所描述的具体实施例仅用以解释本发明，并不用于限定本发明。

[0042] 本技术领域技术人员可以理解，除非特意声明，这里使用的单数形式“一”、“一个”、“所述”和“该”也可包括复数形式。应该进一步理解的是，本发明的说明书中使用的措辞“包括”是指存在所述特征、整数、步骤、操作、元件和/或组件，但是并不排除存在或添加一个或多个其他特征、整数、步骤、操作、元件、组件和/或它们的组。应该理解，当我们称元件被“连接”或“耦接”到另一元件时，它可以直接连接或耦接到其他元件，或者也可以存在中间元件。此外，这里使用的“连接”或“耦接”可以包括无线连接或无线耦接。这里使用的措辞“和/或”包括一个或更多个相关联的列出项的全部或任一单元和全部组合。

[0043] 本技术领域技术人员可以理解，除非另外定义，这里使用的所有术语(包括技术术语和科学术语)，具有与本发明所属领域中的普通技术人员的一般理解相同的意义。还应该理解的是，诸如通用字典中定义的那些术语，应该被理解为具有与现有技术的上下文中的意义一致的意义，并且除非像这里一样被特定定义，否则不会用理想化或过于正式的含义来解释。

[0044] 移动网络流量分类是将边缘设备产生的网络流量数据分为不同的类别，在流量工程(Traffic Engineering)、网络入侵检测、网络监控和服务质量(QoS)等网络安全和管理应用中发挥着重要作用。传统的网络流量分类方法主要可分为三类:基于端口的方法、基于有效负载的方法和基于机器学习的方法。随着人工智能(Artificial Intelligence)技术的快速发展，深度学习(Deep Learning)等机器学习方法越来越受欢迎。但是所有传统的方法都是从不同的边缘设备收集移动流量数据，并集中处理这些流量数据，这种处理方式存在隐私泄露和威胁数据安全等问题。因此，基于隐私保护的网络流量分类研究已引起全世界的关注。

[0045] 基于联邦学习的移动网络流量分类是利用本地客户端中的原始流量数据来解决了流量分类中的隐私和安全问题，在保护用户隐私方面显示出了巨大的潜力。然而现有的联邦学习流量分类解决方案，主要是对局部流量分类模型进行平均，以获得全局模型。然而，由于应用服务和用户偏好的不同，在不同的客户端中的流量类别分布是不同的，即存在流量数据异构性(或非独立相同分布(非IID)数据)。因此现有的基于联邦学习的移动网络流量分类在异构环境下，无法消除模型聚合中数据分布差异大的客户端参与平均的影响，最终导致流量分类精度下降。

[0046] 针对现有技术的上述缺陷，本发明提供一种异构环境的网络流量分类方法，其中，所述方法应用于目标网络环境，所述目标网络环境包括若干客户端，若干所述客户端分别对应不同的网络流量数据分布，所述方法包括:获取所述目标网络环境对应的全局流量分类模型，将所述全局流量分类模型分发至若干所述客户端;获取若干所述客户端分别对应的采样值，其中，每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定，每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述

局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;根据若干所述客户端分别对应的所述采样值确定若干目标客户端;根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。本发明通过模型偏度选取目标客户端,再通过各目标客户端的局部流量分类模型更新全局流量分类模型,可以解决现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,导致流量分类精度下降的问题。

[0047] 如图1所示,所述方法包括:

[0048] 步骤S100、获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端。

[0049] 具体地,本实施例中目标网络环境属于异构环境,其包括多个客户端,且各客户端的网络流量数据分布是不同的,即各客户端手机的数据的统计特征不同。为了保护各客户端的用户隐私和数据安全,本实施例将目标网络环境的全局流量分类模型下发至各客户端,使得各客户端收集的数据可以仅在本地环境进行处理,不需要上传包含用户敏感信息的隐私数据。

[0050] 如图1所示,所述方法还包括如下步骤:

[0051] 步骤S200、获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到。

[0052] 简单来说,为了消除数据分布差异大的客户端后续参与更新全局流量分类模型的影响,本实施例需要获取各客户端的采样值,由于每一客户端的采样值是基于该客户端的模型偏差确定的,因此通过各客户端的采样值可以筛选出数据分布差异较小的目标客户端。具体地,针对每一客户端,该客户端接收到下发的全局流量分类模型后,会根据本地网络流量数据对全局流量分类模型进行训练,得到该客户端的局部流量分类模型。通过确定该客户端的局部流量分类模型与各客户端的局部流量分类模型的均值的偏差,可以得到该客户端的模型偏度。然后在基于该客户端的模型偏度确定其对应的采样值。最终根据各客户端的采样值筛选出用于对全局流量分类模型进行更新的多个目标客户端。

[0053] 在一种实现方式中,所述获取若干所述客户端分别对应的采样值,包括:

[0054] 步骤S201、获取若干所述客户端分别对应的概率分布数据,其中,每一所述客户端对应的所述概率分布数据基于该客户端对应的所述模型偏度确定;

[0055] 步骤S202、根据若干所述客户端分别对应的所述概率分布数据,确定若干所述客户端分别对应的所述采样值,其中,每一所述客户端对应的所述采样值基于该客户端对应的所述概率分布数据抽样得到。

[0056] 简单来说,针对每一客户端,该客户端都具有一个概率分布数据,该概率分布数据受该客户端的模型偏度影响,可以反映该客户端被成功选为目标客户端的概率分布情况。通过对该客户端的概率分布数据进行抽样,例如随机抽样,即可得到该客户端的采样值。为

了消除数据分布差异大的客户端后续参与更新全局流量分类模型的影响,本实施例会根据各客户端的采样值筛选出目标客户端。

[0057] 在一种实现方式中,所述根据若干所述客户端分别对应的所述采样值确定若干目标客户端,包括:

[0058] 步骤S203、根据所述采样值最大的所述客户端确定候选客户端,判断所述候选客户端对应的总量是否达到目标总量;

[0059] 步骤S204、当所述候选客户端对应的总量未达到所述目标总量时,针对若干所述客户端中除所述候选客户端之外的客户端,继续执行所述获取各所述客户端分别对应的采样值,根据所述采样值最大的所述客户端确定候选客户端的步骤,直至所述候选客户端对应的总量达到所述目标总量,得到若干所述候选客户端;

[0060] 步骤S205、根据若干所述候选客户端,确定预设数量的若干所述目标客户端。

[0061] 具体地,本实施例预先设定了候选客户端的目标总量,每一次将采样值最大的客户端作为候选客户端,每选出一个候选客户端即需要判断一次候选客户端的总量是否满足目标总量,未满足时,已成为候选客户端不再参数下一轮筛选。继续获取剩余的各客户端的采样值,将采样值最大的客户端作为候选客户端,直至选出满足目标总量的多个候选客户端。最后从各候选客户端中选取一定数量的目标客户端,例如可以采用随机选取的方式。

[0062] 如图1所示,所述方法还包括如下步骤:

[0063] 步骤S300、根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。

[0064] 具体地,本实施例筛选出的目标客户端都是数据分布差异较小的客户端,因此采用这些目标客户端的局部流量分类模型对全局流量分类模型进行更新,最后得到的目标全局流量分类模型的流量分类精度会大幅提高。

[0065] 在一种实现方式中,所述步骤S300具体包括:

[0066] 步骤S301、根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到更新全局流量分类模型;

[0067] 步骤S302、判断所述更新全局流量分类模型是否收敛至目标值;

[0068] 步骤S303、当所述更新全局流量分类模型未收敛至所述目标值时,根据若干所述目标客户端分别对应的所述模型偏度,更新若干所述目标客户端分别对应的所述概率分布数据;

[0069] 步骤S304、将所述更新全局流量分类模型作为所述全局流量分类模型,继续执行所述将所述全局流量分类模型分发至若干所述客户端的步骤,直至所述更新全局流量分类模型收敛至所述目标值,得到所述目标全局流量分类模型。

[0070] 具体地,本实施例中对全局流量分类模型的更新是迭代更新。针对每一轮更新,首先通过当前轮筛选出来的各目标客户端的局部流量分类模型对当前的全局流量分类模型进行更新,得到更新全局流量分类模型。然后判断更新全局流量分类模型是否收敛至目标值,若未收敛至目标值时,针对各目标客户端的模型偏度对各目标客户端的概率分布数据进行更新。然后再将更新全局流量分类模型下发给各客户端,进行下一轮的目标客户端的筛选和全局流量分类模型的更新,直至得到的更新全局流量分类模型收敛至目标值,将此时的更新全局流量分类模型作为目标环境最终的目标全局流量分类模型。简言之,本实施

例会通过分析计算每一轮中每个客户端上传的模型偏度,来挑选参与该轮次模型聚合的目标客户端,通过不断的选取合适的目标客户端对全局流量分类模型进行更新,以此提高全局流量分类模型的分类准确度。

[0071] 在一种实现方式中,所述步骤S301具体包括如下步骤:

[0072] 步骤S3011、根据若干所述目标客户端分别对应的所述局部流量分类模型的均值,确定目标模型均值;

[0073] 步骤S3012、根据目标模型均值对所述全局流量分类模型进行更新,得到所述更新全局流量分类模型。

[0074] 简单来说,首先本实施需要将各目标客户端的局部流量分类模型进行聚合,其中,聚合过程采用的是取各目标客户端的局部流量分类模型的模型参数的平均值的方式。然后根据聚合后得到的目标模型均值对全局流量分类模型进行更新。

[0075] 在一种实现方式中,每一所述概率分布数据为Beta分布函数,每一所述Beta分布函数基于第一分布参数和第二分布参数确定,所述第一分布参数用于反映该Beta分布函数对应的所述客户端未成为所述目标客户端的次数,所述第二分布参数用于反映该Beta分布函数对应的所述客户端成为所述目标客户端的次数。

[0076] 举例说明,对于每一个客户端$c \in C$,其概率分布数据均为Beta分布参数A_c,B_c,其中,C表示客户端集合,A_c,B_c分别表示客户端c未成为所述目标客户端的次数和成为所述目标客户端的次数。

[0077] 在一种实现方式中,所述步骤S303具体包括如下步骤:

[0078] 步骤S3031、对若干所述目标客户端中任意两个客户端,根据预设数值增加所述模型偏度最高的客户端对应的所述第一分布参数,根据所述预设数值增加所述模型偏度最低的客户端所对应的所述第二分布参数。

[0079] 为了实现对各目标客户端的概率分布数据进行更新,本实施例首先需要比较各目标客户端的模型偏度。具体地,由于本实施例的目的是争取采用模型偏度交底的局部流量分类模型参与聚合、更新,因此针对目标客户端中任意两个客户端,增加模型偏度较高的客户端的第一分布参数,同时增加模型偏度交底的客户端的第二分布参数。概率分布数据更新以后,模型偏度较低的客户端更容易被抽选为目标客户端。

[0080] 举例说明,对于参与聚合的客户端集合S的任意两个客户端s,s',比较其对应的R_s和$R_{s'}$,如果R_s大于$R_{s'}$,则更新$A_s = A_s + 1$,$B_{s'} = B_{s'} + 1$。

[0081] 为了便于理解,如图2所示,示例本方法的一种算法流程(定义为FEAT算法):

[0082] 1.将全局网络流量分类模型参数分发到每一个客户端上。

[0083] 2.客户端进行本地训练和模型偏度(Skewness)计算:

[0084] (1)客户端接收服务器分发的全局网络流量分类模型;

[0085] (2)使用客户端本地的网络流量数据来训练接受到的网络流量分类模型,得到客户端的局部网络流量分类模型;

[0086] (3)针对每一客户端,计算该客户端当前的局部网络流量分类模型与各局部网络流量分类模型的均值的偏差,得到该客户端的Skewness;

[0087] (4)保存此次训练得到的网络流量分类模型。

[0088] 3.客户端将计算得到的Skewness上传给服务器。

[0089] 4.服务器根据上传客户端的Skewness挑选参与联合计算的客户端:

[0090] (1)对于每一个客户端c∈C,初始化其Beta分布参数A_c,B_c等于1,其中C表示客户端集合。

[0091] (2)对于每一个客户端c∈C,计算其对应的Beta分布β(A_c,B_c),根据该分布获取采样数值,并将其加入集合P中。

[0092] (3)计算集合P中最大值对应的客户端编号,将该客户端加入到候选集合S'中,并将其从C中移除。

[0093] (4)根据设置的候选集合的大小g,重复步骤4.1-4.3,达到候选集合的最大值。其中g=γ*|C|,γ为超参数,表示对客户端的容忍程度。

[0094] (5)随机从候选集合S'中选择d个客户端加入到参与聚合的客户端集合S中。

[0095] (6)根据等式计算集合S中每个客户端的回报值R_i(相当于模型偏差),如下所示:

[0096] $$R_i = -\|\Delta\omega_i - \overline{\Delta\omega}\|_2^2$$

[0097] 其中,R_i表示客户端i的回报值;$\Delta\omega_i$表示客户端i的模型参数;$\overline{\Delta\omega}$表示客户端模型参数的均值。

[0098] (7)对于参与聚合的客户端集合S的任意两个客户端s,s',比较其对应的R_s和$R_{s'}$,如果R_s大于$R_{s'}$,则更新$A_s=A_s+1$,$B_{s'}=B_{s'}+1$。

[0099] 5.根据步骤4的客户端集合S,计算其模型权重的平均值,并更新服务器模型。

[0100] 6.重复步骤1-5,直到模型收敛到规定的阈值。

[0101] 简单来说,该算法首先自定义了一个用于衡量客户端上传模型差异程度的指标Skewness,由于客户端模型是本地数据分布的环境下训练得到的,所以其模型应该能够衡量对应本地数据分布的特征而不用暴露本地数据给外部环境,确保了用户隐私的安全性。其次通过将边缘网络设备的Skewness上传到服务器,以保护隐私的方式进行客户端Skewness估计和客户端选择;最后在服务器完成全局模型的聚合。其中,客户端选择问题是一个多强盗决斗问题,可以利用汤普森采样方法选择低偏度客户端进行模型聚合。需要说明的是,本发明在公开的真实网络数据集QUIC上进行了实验以证明其优越性。实验结果表明本发明在低异构环境下,对于网络流量的分类精度相对于随机客户端选择算法准确率提高了66.7%,而在高异构环境下提高了68.6%。此外,在低异构环境和高异构环境下,相比于传统的联邦学习算法收敛速度可以分别提高2.6倍和3.4倍。

[0102] 综上,本发明具有以下优点:

[0103] 1.本发明首先提出了在隐私保护条件下解决异构环境中网络流量分类模型精度下降的方法。

[0104] 2.本发明提出了一种能够主动、智能地选择低偏度的客户端参与联邦学习的客户端选择算法。并通过了严格的数学证明来确保其可行性。

[0105] 3.FEAT算法能够更好的提高在异构环境下模型的分类精度。和随机客户端选择算法相比,FEAT的分类准确率在低高异构环境下分别提高了66.7%和68.6%。

[0106] 4.FEAT算法可以具有更快的收敛速度。和传统的联邦学习算法相比,FEAT算法能够将收敛速度提高2.6倍和3.4倍。

[0107] 基于上述实施例,本发明还提供了一种异构环境的网络流量分类装置,如图3所示,所述装置包括:

[0108] 分发模块01,用于获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端;

[0109] 筛选模块02,用于获取若干所述客户端分别对应的采样值,根据若干所述客户端分别对应的所述采样值确定若干目标客户端,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;

[0110] 更新模块03,用于根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。

[0111] 基于上述实施例,本发明还提供了一种终端,其原理框图可以如图4所示。该终端包括通过系统总线连接的处理器、存储器、网络接口、显示屏。其中,该终端的处理器用于提供计算和控制能力。该终端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和计算机程序。该内存储器为非易失性存储介质中的操作系统和计算机程序的运行提供环境。该终端的网络接口用于与外部的终端通过网络连接通信。该计算机程序被处理器执行时以实现异构环境的网络流量分类方法。该终端的显示屏可以是液晶显示屏或者电子墨水显示屏。

[0112] 本领域技术人员可以理解,图4中示出的原理框图,仅仅是与本发明方案相关的部分结构的框图,并不构成对本发明方案所应用于其上的终端的限定,具体的终端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。

[0113] 在一种实现方式中,所述终端的存储器中存储有一个或者一个以上的程序,且经配置以由一个或者一个以上处理器执行所述一个或者一个以上程序包含用于进行异构环境的网络流量分类方法的指令。

[0114] 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非易失性计算机可读取存储介质中,该计算机程序在执行时,可包括如上述各方法的实施例的流程。其中,本发明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。

[0115] 综上所述,本发明公开了异构环境的网络流量分类方法、装置、终端及存储介质,所述方法应用于目标网络环境,所述目标网络环境包括若干客户端,若干所述客户端分别对应不同的网络流量数据分布,所述方法包括:获取所述目标网络环境对应的全局流量分类模型,将所述全局流量分类模型分发至若干所述客户端;获取若干所述客户端分别对应的采样值,其中,每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定,每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述

局部流量分类模型的均值之间的偏差,每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到;根据若干所述客户端分别对应的所述采样值确定若干目标客户端;根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新,得到目标全局流量分类模型。本发明通过模型偏度选取目标客户端,再通过各目标客户端的局部流量分类模型更新全局流量分类模型,可以解决现有的基于联邦学习的移动网络流量分类在异构环境下,无法消除模型聚合中数据分布差异大的客户端参与平均的影响,导致流量分类精度下降的问题。

[0116] 应当理解的是,本发明的应用不限于上述的举例,对本领域普通技术人员来说,可以根据上述说明加以改进或变换,所有这些改进和变换都应属于本发明所附权利要求的保护范围。

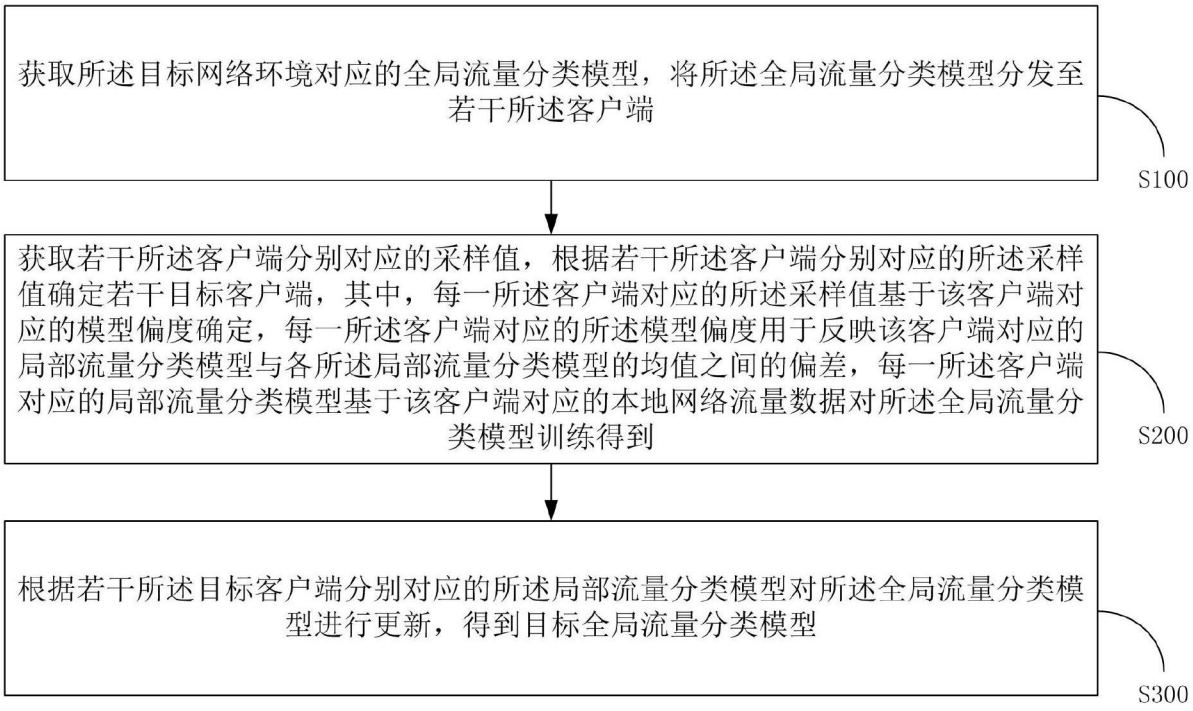

获取所述目标网络环境对应的全局流量分类模型，将所述全局流量分类模型分发至若干所述客户端
S100
获取若干所述客户端分别对应的采样值，根据若干所述客户端分别对应的所述采样值确定若干目标客户端，其中，每一所述客户端对应的所述采样值基于该客户端对应的模型偏度确定，每一所述客户端对应的所述模型偏度用于反映该客户端对应的局部流量分类模型与各所述局部流量分类模型的均值之间的偏差，每一所述客户端对应的局部流量分类模型基于该客户端对应的本地网络流量数据对所述全局流量分类模型训练得到
S200
根据若干所述目标客户端分别对应的所述局部流量分类模型对所述全局流量分类模型进行更新，得到目标全局流量分类模型
S300

图1

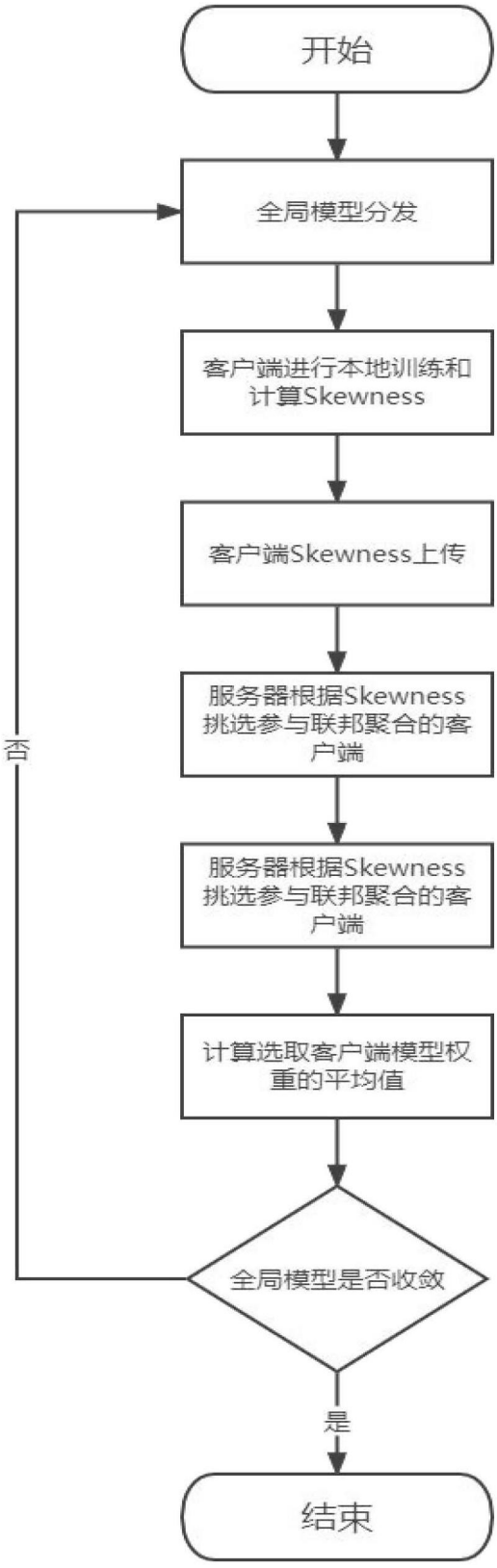

开始
全局模型分发
客户端进行本地训练和计算Skewness
客户端Skewness上传
服务器根据Skewness挑选参与联邦聚合的客户端
服务器根据Skewness挑选参与联邦聚合的客户端
计算选取客户端模型权重的平均值
全局模型是否收敛
否
是
结束

图2

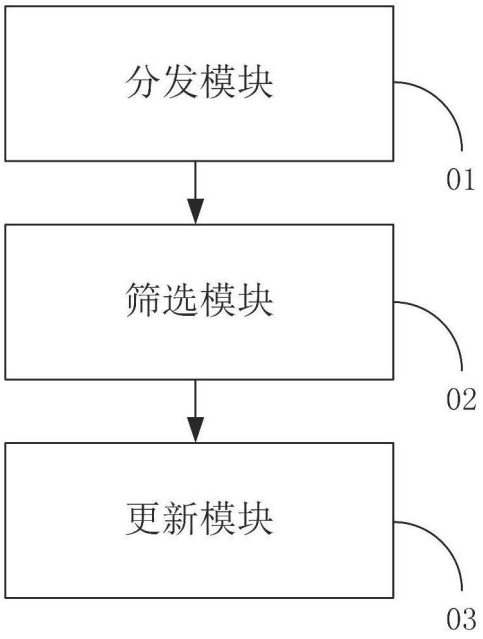

分发模块
01
筛选模块
02
更新模块
03

图3

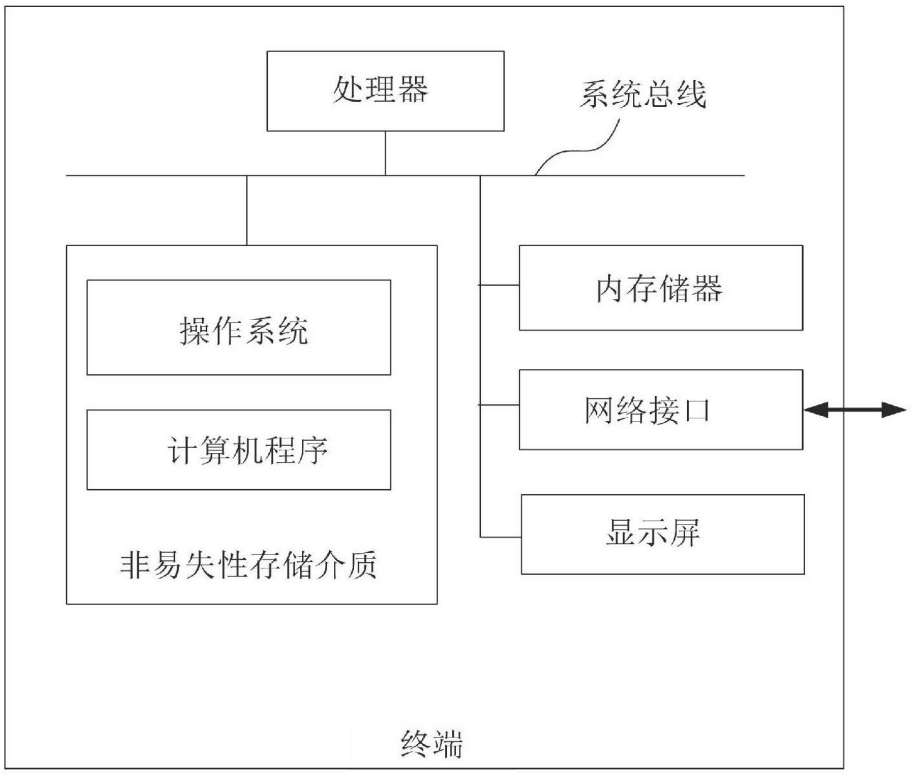

处理器
系统总线
操作系统
计算机程序
非易失性存储介质
内存储器
网络接口
显示屏
终端

图4