



(12) 发明专利

(10) 授权公告号 CN 118658303 B

(45) 授权公告日 2024. 11. 05

(21) 申请号 202411132664.3

(22) 申请日 2024.08.19

(65) 同一申请的已公布的文献号

申请公布号 CN 118658303 A

(43) 申请公布日 2024.09.17

(73) 专利权人 四川万物纵横科技股份有限公司

地址 610041 四川省成都市高新区天府大道中段1388号12栋7层1号

(72) 发明人 谭利明 毛开福 杨帆

(74) 专利代理机构 成都贞元会专知识产权代理

有限公司 51390

专利代理师 黄娟

(51) Int. Cl.

G08G 1/01 (2006.01)

H04L 67/12 (2022.01)

(56) 对比文件

CN 113112088 A, 2021.07.13

CN 113837222 A, 2021.12.24

审查员 方涵

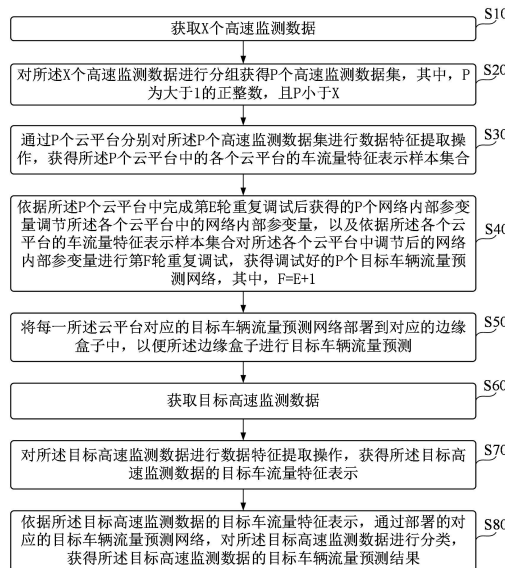
权利要求书3页 说明书25页 附图3页

(54) 发明名称

一种低延迟边缘协同数据处理方法及系统

(57) 摘要

本申请提供一种低延迟边缘协同数据处理方法及系统,在网络调试过程中,通过P个云平台分别对P个高速监测数据集进行数据特征提取操作以及通过各个云平台对各个云平台的车流量特征表示样本集合进行重复调试来调试目标车辆流量预测网络,也即基于各个云平台对网络调试所依赖的算力消耗进行分散,这样可以对各云平台的空置算力进行有效使用,使得P个目标车辆流量预测网络的调试的算力成本得到降低,此外,减少每次重复调试对算力和存储的要求。在第F轮重复调试中,基于同享P个云平台中完成第E轮重复调试后获得的P个网络内部参变量,完成对P个目标车辆流量预测网络的联邦学习,确保P个目标车辆流量预测网络的调校质量。



1. 一种低延迟边缘协同数据处理方法,其特征在于,包括:

获取X个高速监测数据;

对所述X个高速监测数据进行分组获得P个高速监测数据集,其中,P为大于1的正整数,且P小于X;

通过P个云平台分别对所述P个高速监测数据集进行数据特征提取操作,获得所述P个云平台中的各个云平台的车流量特征表示样本集合;

依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,其中, $F=E+1$;

将所述各个云平台对应的目标车辆流量预测网络部署到对应的边缘盒子中,以便所述边缘盒子进行目标车辆流量预测;

其中,所述各个云平台与至少一个边缘盒子通信连接,针对每一个所述边缘盒子,在进行目标车辆流量预测时,包括:

获取目标高速监测数据;

对所述目标高速监测数据进行数据特征提取操作,获得所述目标高速监测数据的目标车流量特征表示;

依据所述目标高速监测数据的目标车流量特征表示,通过部署的对应的目标车辆流量预测网络,对所述目标高速监测数据进行分类,获得所述目标高速监测数据的目标车辆流量预测结果;

其中,所述依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,包括:

通过所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量的聚合参变量,调节所述各个云平台中的网络内部参变量;

依据所述各个云平台的车流量特征表示样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络;

所述各个云平台中的初始网络内部参变量一致,所述P个高速监测数据集中的各个高速监测数据集包括X/P个高速监测数据;

或者,所述依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,包括:

对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合;

通过所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量的聚合参变

量,调节所述各个云平台中的网络内部参变量;

依据所述各个云平台的学习样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,获得所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量;

依据所述各个云平台的验证样本集合,对所述各个云平台中的网络内部参变量的误差进行评估,获得用以指示所述P个云平台中的网络内部参变量的误差的目标误差值;

判断所述目标误差值是否符合评估条件,如果所述目标误差值符合所述评估条件,获得所述P个目标车辆流量预测网络;如果所述目标误差值不符合评估条件,则通过所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量的聚合参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的学习样本集合,对所述各个云平台中的网络内部参变量进行第G次重复调试,直到所述目标误差值符合所述评估条件时,获得所述P个目标车辆流量预测网络,其中, $G=F+1$ 。

2.根据权利要求1所述的方法,其特征在于,所述依据所述各个云平台的车流量特征表示样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络,包括:

将所述各个云平台中的网络内部参变量与所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示进行作积,将作积结果与所述各个云平台的车流量特征表示样本集合中的先验标记进行作差,获得作差结果,将所述作差结果与所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示进行作积,获得所述各个云平台中的网络内部参变量的偏导向量;

将所述各个云平台中的网络内部参变量的调谐参数和所述各个云平台中的网络内部参变量的偏导向量进行作积,获得所述各个云平台中的网络内部参变量的调节结果;

通过所述各个云平台中的网络内部参变量减去所述各个云平台中的网络内部参变量的调节结果,获得所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量;

如果F等于所述事先确定的最大重复轮次或所述各个云平台中的网络内部参变量的偏导向量等于0时,获得所述P个目标车辆流量预测网络;

如果F小于所述事先确定的最大重复轮次或所述各个云平台中的网络内部参变量的偏导向量等于0时,则依据所述各个云平台的车流量特征表示样本集合和所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量,对所述各个云平台中的网络内部参变量进行第G次重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络,其中, $G=F+1$ 。

3.根据权利要求1所述的方法,其特征在于,所述对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合,包括:

依据所述各个云平台的车流量特征表示样本集合的类别,基于设定的百分比通过对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合。

4.根据权利要求3所述的方法,其特征在于,所述依据所述各个云平台的车流量特征表

示样本集合的类别,基于设定的百分比通过对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合,包括:

如果所述各个云平台的车流量特征表示样本集合的类别为符合特征空间一致但样本空间各异的分布式数据类别,则基于所述设定的百分比通过对所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示和先验标记进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合;

如果所述各个云平台的车流量特征表示样本集合的类别为符合特征空间不同但样本空间相同的分布式数据类别,则基于所述设定的百分比通过对所述P个云平台中的第一云平台的车流量特征表示样本集合中的目标车流量特征表示和先验标记进行稀疏提取,并基于所述设定的百分比通过对所述P个云平台中所述第一云平台以外的云平台的车流量特征表示样本集合中的目标车流量特征表示进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合。

5.根据权利要求1所述的方法,其特征在于,所述依据所述各个云平台的验证样本集合,对所述各个云平台中的网络内部参变量的误差进行评估,获得用以指示所述P个云平台中的网络内部参变量的误差的目标误差值,包括:

通过将所述各个云平台中的网络内部参变量与所述各个云平台的验证样本集合中的目标车流量特征表示进行作积,获得所述各个云平台的阶段性预估值;

通过预设分类器对所述各个云平台的阶段性预估值进行分类,获得所述各个云平台得到的目标车辆流量预测结果;

依据所述各个云平台得到的目标车辆流量预测结果和所述各个云平台的验证样本集合中的先验标记,确定所述目标误差值。

6.根据权利要求5所述的方法,其特征在于,所述依据所述各个云平台得到的目标车辆流量预测结果和所述各个云平台的验证样本集合中的先验标记,确定所述目标误差值,包括:

通过所述各个云平台的验证样本集合中的先验标记减去所述各个云平台得到的目标车辆流量预测结果,获得所述各个云平台中的网络内部参变量的误差目标车流量特征表示;

对所述各个云平台中的网络内部参变量的误差目标车流量特征表示进行翻转,将翻转后的特征表示与所述各个云平台中的网络内部参变量的误差目标车流量特征表示进行作积,获得所述各个云平台中的网络内部参变量的误差值;

对所述各个云平台中的网络内部参变量的误差值进行均值计算,得到均值误差,将所述均值误差作为所述目标误差值。

7.一种边缘协同数据处理系统,其特征在于,包括多个边缘盒子和多个云平台,多个所述边缘盒子和多个所述云平台通信连接,其中,所述云平台包括:一个或多个处理器;存储器;一个或多个计算机程序;其中一个或多个所述计算机程序被存储在所述存储器中并被配置为由一个或多个所述处理器执行,一个或多个所述计算机程序被所述处理器执行时,实现如权利要求1~6任一项所述的方法。

一种低延迟边缘协同数据处理方法及系统

技术领域

[0001] 本申请涉及边缘计算、机器学习领域,具体而言,涉及一种低延迟边缘协同数据处理方法及系统。

背景技术

[0002] 随着高速公路交通流量的不断增加,对实时、准确的车辆流量预测需求日益迫切。传统的车辆流量预测方法往往依赖于中心化的数据处理系统,然而,这种中心化的处理方式面临着数据传输延迟高、实时性差等问题,难以满足现代交通管理对低延迟、高效率的要求。近年来,随着云计算和边缘计算技术的快速发展,为车辆流量预测提供了新的解决方案。云计算通过强大的计算能力和海量的存储资源,能够处理大规模的数据集,但其延迟问题限制了其在实时性要求较高的场景中的应用。而边缘计算则将计算任务推向网络边缘,利用更靠近数据源的计算资源进行数据处理和分析,从而降低了数据传输延迟,提高了实时性。

[0003] 然而,单纯的边缘计算或云计算解决方案在车辆流量预测领域仍存在局限性。边缘计算虽然能够实现低延迟处理,但其计算资源相对有限,难以处理复杂的数据分析任务;而云计算则面临着数据传输延迟和实时性不足的问题。因此,如何将云计算和边缘计算的优势相结合,实现高效、低延迟的车辆流量预测,成为了当前研究的热点。目前,已有一些研究尝试通过云平台协同工作来提高车辆流量预测的性能。然而,这些研究大多集中在单一云平台内的数据处理和模型训练,缺乏对多个云平台之间协同工作的深入探索。此外,如何在保护数据隐私的同时,充分利用各个云平台的数据和计算资源,也是一个亟待解决的问题。

发明内容

[0004] 本发明的目的在于提供一种低延迟边缘协同数据处理方法及系统。本发明是这样实现的:

[0005] 第一方面,本申请实施例提供了一种低延迟边缘协同数据处理方法,包括:获取X个高速监测数据;对所述X个高速监测数据进行分组获得P个高速监测数据集,其中,P为大于1的正整数,且P小于X;通过P个云平台分别对所述P个高速监测数据集进行数据特征提取操作,获得所述P个云平台中的各个云平台的车流量特征表示样本集合;依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,其中, $F=E+1$;将所述各个云平台对应的目标车辆流量预测网络部署到对应的边缘盒子中,以便所述边缘盒子进行目标车辆流量预测;其中,所述各个云平台与至少一个边缘盒子通信连接,针对每一个所述边缘盒子,在进行目标车辆流量预测时,包括:获取目标高速监测数据;对所述目标高速监测数据进行数据特征提取操作,获得所述目标高速监测数据的目标车流量特

征表示;依据所述目标高速监测数据的目标车流量特征表示,通过部署的对应的目标车辆流量预测网络,对所述目标高速监测数据进行分类,获得所述目标高速监测数据的目标车辆流量预测结果。

[0006] 第二方面,本申请提供一种低延迟边缘协同数据处理方法,应用于边缘盒子,所述边缘盒子与目标云平台通信连接,所述方法包括:获取目标高速监测数据;对所述目标高速监测数据进行数据特征提取操作,获得所述目标高速监测数据的目标车流量特征表示;依据所述目标高速监测数据的目标车流量特征表示,通过所述目标云平台中的目标车辆流量预测网络,对所述目标高速监测数据进行分类,获得所述目标高速监测数据的目标车辆流量预测结果,其中,所述目标车辆流量预测网络基于以上第一方面提供的所述方法调试得到。

[0007] 第三方面,本申请提供一种边缘协同数据处理系统,包括多个边缘盒子和多个云平台,多个所述边缘盒子和多个所述云平台通信连接,其中,所述云平台包括:一个或多个处理器;存储器;一个或多个计算机程序;其中一个或多个所述计算机程序被存储在所述存储器中并被配置为由一个或多个所述处理器执行,一个或多个所述计算机程序被所述处理器执行时,实现以上第一方面提供的所述方法。

[0008] 本申请至少包含的有益效果:本申请提供一种低延迟边缘协同数据处理方法及系统,在网络调试过程中,通过P个云平台分别对P个高速监测数据集进行数据特征提取操作以及通过各个云平台对各个云平台的车流量特征表示样本集合进行重复调试来调试目标车辆流量预测网络,也即基于各个云平台对网络调试所依赖的算力消耗进行分散,这样可以对各云平台的空置算力进行有效使用,使得P个目标车辆流量预测网络的调试的算力成本得到降低,此外,减少每次重复调试对算力和存储的要求。依据P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节各个云平台中的网络内部参变量,以及依据各个云平台的车流量特征表示样本集合对各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,也即在第F轮重复调试中,基于同享P个云平台中完成第E轮重复调试后获得的P个网络内部参变量,完成对P个目标车辆流量预测网络的联邦学习,确保P个目标车辆流量预测网络的调校质量。

附图说明

[0009] 为了更清楚地说明本申请实施例中的技术方案,下面将对本申请实施例描述中所需要使用的附图作简单地介绍。

[0010] 图1是本申请实施例提供的边缘协同数据处理系统的架构示意图。

[0011] 图2是本申请实施例提供的一种低延迟边缘协同数据处理方法的流程图。

[0012] 图3是本申请实施例提供的一种云平台的组成示意图。

[0013] 附图标记:101-处理器;102-总线;103-存储器;104-收发器;105-边缘盒子;106-云平台。

具体实施方式

[0014] 下面结合本申请实施例中的附图对本申请实施例进行描述。本申请实施例的实施方式部分使用的术语仅用于对本申请的具体实施例进行解释,而非旨在限定本申请。

[0015] 本申请实施例提供的低延迟边缘协同数据处理方法,可以应用于如图1所示的边缘协同数据处理系统的云平台106和边缘盒子105中。其中,边缘盒子105通过网络与云平台106进行通信。数据存储系统可以存储云平台106在训练神经网络时需要处理的数据。数据存储系统可以集成在云平台106上。其中,在神经网络推理阶段,高速监测数据可存储在边缘盒子105的本地存储中,也可存储至数据存储系统或者与云平台106关联的云端存储中。可以理解,边缘盒子105是安装在道路沿线或关键节点上的智能设备,作为智慧交通系统的一部分。这些边缘盒子具备数据处理、通信和一定的智能分析能力,能够实时收集和处理来自各类传感器的交通数据,如车流量、车速、道路占用率等,并将处理后的数据上传至云端或发送给其他相关系统。具体来说,边缘盒子可以集成多种传感器接口、数据处理芯片、通信模块等硬件组件,以及相应的软件和算法。它们通常被部署在靠近数据源的位置,以便能够更快地响应和处理数据,减少数据传输的延迟。云平台106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。

[0016] 本申请实施例提供了一种低延迟边缘协同数据处理方法,该方法应用于云平台,如图2所示,该方法包括:

[0017] 网络训练阶段,包括步骤S10~S40:

[0018] 步骤S10:获取X个高速监测数据。

[0019] 在步骤S10中,云平台从高速公路监测系统中获取X个高速监测数据。这些高速监测数据是实时收集的,用于反映高速公路的交通状况。具体地,这些数据通常来自于安装在高速公路上的各种传感器和设备,如交通流量传感器、车速检测器、摄像头等。以一个典型的高速公路边缘协同数据处理场景为例,假设某条高速公路上有多个监测点,每个监测点都配备了一系列传感器来收集数据。这些传感器例如包括用于测量车辆通过数量的红外传感器、用于记录车速的雷达测速仪以及用于捕捉实时交通画面的高清摄像头。这些设备会不断生成大量的原始数据,这些数据包含了丰富的交通信息。在步骤S10中,云平台连接到这些传感器和设备,通过数据采集接口实时获取这些高速监测数据。这些数据可能包括每分钟的车辆通过数量、平均车速、车辆类型分布、道路占用率等。此外,如果摄像头拍摄了实时交通画面,这些图像数据也会被收集起来,以供后续的数据分析和处理使用。

[0020] 假设X为10000,那么云平台获取这10000个高速监测数据样本。这些数据样本可能是以时间序列的形式记录的,每个样本都包含了一段时间内的交通数据。通过获取这些数据,云平台可以构建一个庞大的数据集,用于后续的模式训练和车辆流量预测。

[0021] 步骤S20:对所述X个高速监测数据进行分组获得P个高速监测数据集,其中,P为大于1的正整数,且P小于X。

[0022] 在步骤S20中,云平台对获取到的X个高速监测数据进行分组,以形成P个高速监测数据集。这一步的目的是将庞大的数据集拆分成多个较小的子集,以便在多个云平台上并行处理,从而提高数据处理效率。

[0023] 以高速公路边缘协同数据处理为例,假设X为10000,这表示云平台已经收集了10000个高速监测数据样本。这些数据可能来自不同的高速公路路段、不同的时间段或不同的监测设备。由于数据量较大,直接在一个云平台上进行处理可能会面临资源瓶颈和效率问题。因此,在步骤S20中,云平台根据一定的策略将这10000个数据样本分成P个较小的数据集。P是一个大于1的正整数,且小于X,具体取值可以根据实际情况确定。例如,如果云平

台拥有5个可用的云平台,那么可以将数据分成5个数据集(即 $P=5$),每个数据集包含2000个数据样本。

[0024] 分组策略可以根据具体的应用需求和数据特性来制定。一种常见的策略是按照高速公路路段进行分组,即将同一路段的数据样本划分到同一个数据集中。这样可以确保每个数据集都反映了相同路段的交通状况,有利于后续的数据分析和模型训练。另外,也可以考虑按照时间段进行分组,比如将一天的数据按照小时进行划分,每个小时的数据形成一个数据集。这样做的好处是可以捕捉到不同时间段的交通流量变化,有助于分析交通流量的周期性规律。

[0025] 一旦确定了分组策略,云平台就会根据策略对 X 个高速监测数据进行分组,生成 P 个高速监测数据集。每个数据集都包含了一部分原始数据样本,并且保持了原始数据的完整性和准确性。接下来,这些数据集将被分别发送到不同的云平台上进行后续的数据处理和模型训练。需要注意的是,分组过程中需要确保每个数据集都具有一定的代表性,即能够反映整体数据的某些特性。同时,还需要考虑数据集的平衡性,避免某个数据集过大或过小,导致后续处理过程中的资源分配不均。通过合理的分组策略,可以最大限度地发挥多个云平台并行处理的优势,提高整个数据处理流程的效率。

[0026] 步骤S30:通过 P 个云平台分别对所述 P 个高速监测数据集进行数据特征提取操作,获得所述 P 个云平台中的各个云平台的车流量特征表示样本集合。

[0027] 在步骤S30中,云平台将利用 P 个云平台对 P 个高速监测数据集进行并行的数据特征提取操作,以获得每个云平台对应的车流量特征表示样本集合。这一步骤是构建车辆流量预测模型的关键步骤之一,因为它涉及到从原始的高速监测数据中提取出有意义的信息,这些信息将用于后续的模型训练。

[0028] 具体地,假设 P 为5,意味着有5个云平台将参与数据特征提取过程。每个云平台都将接收一个包含2000个高速监测数据样本的数据集(假设原始数据集 X 被均匀地分成 P 个数据集)。这些高速监测数据样本可能包括车流量、车速、道路占用率、天气条件等多种类型的实时信息。

[0029] 在每个云平台上,数据特征提取操作将通过应用特定的机器学习算法或模型来实现。例如,可以使用卷积神经网络(CNN)或循环神经网络(RNN)来从图像数据中提取车辆流量相关的视觉特征,或者使用决策树、随机森林或梯度提升机等算法从数值数据中提取统计特征。这些特征可能包括车辆数量的时间序列变化、平均车速的分布、道路拥堵指数等。

[0030] 以一个具体的例子来说,如果某个云平台接收到了包含摄像头拍摄图像的数据集,它可以利用一个预训练的CNN模型来提取图像中的车辆数量和车辆密度等视觉特征。这些特征将被编码为数值向量,每个向量代表一个时间点的车辆流量情况。类似地,如果数据集包含车速和车流量的数值数据,云平台可以计算这些数据的统计特征,如平均值、中位数、标准差等,以及时间序列上的变化特征,如增长率、周期性等。通过这个过程,每个云平台都将生成一个车流量特征表示样本集合。这个集合包含了一系列特征向量,每个向量对应一个高速监测数据样本,并且包含了从原始数据中提取出的有意义的信息。这些特征向量将作为后续模型训练的输入,用于学习从特征到车辆流量的映射关系。

[0031] 通过并行地在 P 个云平台上进行数据特征提取操作,云平台能够充分利用各个云平台的计算资源,加速整个数据处理流程。同时,由于每个云平台都处理一个独立的数据

集,因此还可以在在一定程度上提高数据的多样性和模型的泛化能力。

[0032] 步骤S40:依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,其中, $F=E+1$ 。

[0033] 在步骤S40中,云平台将执行一个迭代的过程,用于调试和优化P个目标车辆流量预测网络。这个过程涉及到依据P个云平台在第E轮重复调试后获得的网络内部参变量来调节各云平台中的网络内部参变量,并基于各云平台的车流量特征表示样本集合对调节后的网络内部参变量进行第F轮($F=E+1$)重复调试。以下是对该步骤的详细解释,结合高速公路边缘协同数据处理的场景进行举例。

[0034] 首先,假设P为5,表示有5个云平台参与车辆流量预测网络的训练和优化。每个云平台都维护了一个车辆流量预测网络,这些网络在初始时可能具有相同的网络结构和初始参数,但随着训练的进行,它们将逐渐学习到各自数据集上的特定模式。在第E轮重复调试后,每个云平台都获得了一组网络内部参变量(例如神经网络的权重和偏置)。这些参变量代表了网络在当前状态下的知识,即如何从输入的车流量特征表示中预测车辆流量。然而,由于每个云平台处理的数据集可能不同,因此它们获得的网络内部参变量也可能存在差异。

[0035] 为了充分利用这些差异并提升整体模型的性能,云平台将执行一个聚合操作。具体来说,它将收集所有云平台在第E轮调试后获得的网络内部参变量,并计算它们的聚合参变量。聚合操作的具体方式可以根据实际情况选择,例如可以是取平均值、中位数或者通过某种优化算法来找到一个全局最优的参变量集合。

[0036] 接下来,云平台将使用这些聚合参变量来调节各个云平台中的网络内部参变量。这一步骤的目的是将各个云平台的知识进行融合,使得每个云平台都能够从其他云平台的学习经验中受益。通过调节网络内部参变量,每个云平台上的车辆流量预测网络都将更新其状态,以更好地适应整个数据集的分布。

[0037] 在调节完成后,云平台将依据各个云平台的车流量特征表示样本集合对调节后的网络内部参变量进行第F轮($F=E+1$)重复调试。这一步骤是为了进一步微调网络参数,以提高其在特定数据集上的预测性能。调试过程通常涉及到前向传播(计算预测值)、计算损失(预测值与真实值之间的差异)、反向传播(计算梯度并更新网络参数)等步骤。具体来说,对于每个云平台,云平台将使用其车流量特征表示样本集合中的一部分作为训练数据,另一部分作为验证数据。在训练过程中,云平台将前向传播输入特征通过车辆流量预测网络,计算预测的车辆流量值,并与真实值进行比较以计算损失。然后,云平台将通过反向传播算法计算梯度,并使用优化器(如梯度下降算法)来更新网络参数。在验证过程中,云平台将使用验证数据来评估网络的性能,并根据需要调整超参数(如学习率、批次大小等)以优化训练效果。

[0038] 这个过程将重复进行多轮(即F轮),直到满足某个停止准则(如达到预设的最大轮数、验证损失不再下降等)。在每一轮调试中,各个云平台都将根据其在当前轮次中获得的网络内部参变量和车流量特征表示样本集合来更新其车辆流量预测网络。最终,当调试过程收敛或达到预设的停止准则时,云平台将获得P个调试好的目标车辆流量预测网络,这些

网络将用于在实际应用中预测高速公路的车辆流量。

[0039] 在网络的应用阶段,包括步骤S50~S80:

[0040] 步骤S50:将所述各个云平台对应的目标车辆流量预测网络部署到对应的边缘盒子中,以便所述边缘盒子进行目标车辆流量预测。

[0041] 在步骤S50中,将每个云平台训练好的目标车辆流量预测网络部署到对应的边缘盒子中。这一步骤是边缘协同数据处理流程的关键环节,它使得车辆流量预测的能力能够下沉到网络边缘,实现低延迟的实时预测。本申请实施例中,边缘盒子通常被部署在接近数据源的位置,即高速公路的监测点附近。这些边缘盒子具备了一定的计算能力和存储能力,可以实时接收和处理来自监测点的数据。通过将车辆流量预测网络部署到边缘盒子中,可以实现数据的就地处理,大大减少数据传输的延迟,提高预测的实时性。接下来详细解释步骤S50的执行过程。假设有5个云平台,每个云平台都训练出了一个目标车辆流量预测网络。这些网络经过多轮的调试和优化,已经具备了较好的预测性能。此时,云平台将开始将这些网络部署到对应的边缘盒子中。

[0042] 具体来说,对于每个云平台,云平台将首先将其训练好的目标车辆流量预测网络进行序列化操作,将网络的结构和参数转换为一种可以在边缘盒子上运行的格式。这个格式通常是轻量级的,以适应边缘盒子的计算资源限制。然后,云平台将通过网络连接将序列化后的网络模型传输到对应的边缘盒子。

[0043] 在边缘盒子接收到网络模型后,它将进行反序列化操作,将模型还原为可执行的状态。同时,边缘盒子还会根据需要进行一些配置工作,如设置输入数据的格式、指定输出的结果存储位置等。完成这些配置后,边缘盒子就具备了进行车辆流量预测的能力。

[0044] 以一个具体的例子来说,假设某个边缘盒子负责处理来自某个高速公路监测点的数据。当监测点采集到新的高速监测数据时,这些数据将被实时传输到边缘盒子。边缘盒子接收到数据后,将对其进行必要的的数据预处理操作,如数据清洗、特征提取等。然后,它将使用已经部署在其上的目标车辆流量预测网络对处理后的数据进行预测,得到预测结果。这个预测结果将被边缘盒子实时存储并用于后续的应用场景,如交通调度、路况发布等。通过步骤S50的执行,云平台成功地将车辆流量预测的能力下沉到了网络边缘,实现了低延迟的实时预测。这不仅能够提高预测的准确性和实时性,还能够减轻中心服务器的负载压力,提高整个系统的处理能力和效率。

[0045] 其中,所述各个云平台与至少一个边缘盒子通信连接,针对每一个所述边缘盒子,在进行目标车辆流量预测时,包括:

[0046] 步骤S60:获取目标高速监测数据。

[0047] 在步骤S60中,边缘盒子针对每一个边缘盒子进行目标车辆流量预测时,获取目标高速监测数据。这一步骤是边缘盒子进行实时车辆流量预测的基础,确保后续的数据分析和预测能够基于最新的、准确的交通数据。

[0048] 本申请实施例中,边缘盒子通常被部署在高速公路的监测点附近,这些监测点配备了各种传感器和设备,用于实时采集交通数据。这些传感器可能包括交通流量传感器、车速检测器、摄像头等,它们能够实时地监测并记录高速公路上的车辆流量、车速、道路占用率等信息。当边缘盒子进行车辆流量预测时,通过与其连接的通信接口(如网络接口、串口等)从监测点的传感器和设备中获取目标高速监测数据。这些数据可能包括特定时间段内

的车辆通过数量、平均车速、道路拥堵情况等实时信息。

[0049] 例如,假设一个边缘盒子负责处理来自某个高速公路收费站的监测数据。当收费站的车道上有车辆通过时,交通流量传感器实时记录车辆的通过时间和车道号等信息,并将这些数据传输到边缘盒子。边缘盒子在接收到这些数据后,会将其作为目标高速监测数据,用于后续的车辆流量预测。需要注意的是,由于高速公路的交通状况是实时变化的,因此边缘盒子需要不断地从监测点获取最新的高速监测数据,以确保预测结果的准确性和实时性。同时,由于不同监测点的数据类型和格式可能存在差异,边缘盒子还需要对获取到的数据进行必要的预处理和格式化操作,以便能够输入到车辆流量预测网络中进行处理。

[0050] 通过执行步骤S60,云平台能够确保边缘盒子在进行车辆流量预测时能够基于最新的、准确的交通数据,从而提高预测结果的准确性和实时性。

[0051] 步骤S70:对所述目标高速监测数据进行数据特征提取操作,获得所述目标高速监测数据的目标车流量特征表示。

[0052] 在步骤S70中,云平台将对获取到的目标高速监测数据进行数据特征提取操作,以生成目标车流量特征表示。这一步骤是车辆流量预测过程中的关键一环,它通过对原始数据进行处理和分析,提取出与车辆流量预测密切相关的特征信息,为后续的预测模型提供输入。

[0053] 本申请实施例中,目标高速监测数据通常来自于部署在高速公路沿线的各种传感器和设备,如交通流量传感器、车速检测器、摄像头等。这些数据包含了实时的交通信息,如车辆通过数量、车速、道路占用率等。然而,原始数据往往包含大量的冗余信息和噪声,直接用于预测模型可能会导致预测结果的不准确。因此,需要通过数据特征提取操作来提取出与目标车辆流量预测密切相关的特征。

[0054] 在步骤S70中,边缘盒子将应用一系列的数据预处理和特征提取技术来处理目标高速监测数据。首先,边缘盒子对原始数据进行清洗和整理,去除重复、异常或无效的数据,确保数据的准确性和一致性。然后,边缘盒子根据车辆流量预测的需求,选择适当的特征提取方法。

[0055] 例如,对于交通流量传感器采集到的车辆通过数量数据,边缘盒子可以直接将其作为一个特征。对于车速检测器采集到的车速数据,边缘盒子可以计算其平均值、最大值、最小值等统计特征,以反映不同时间段内的车速分布情况。对于摄像头采集到的图像数据,边缘盒子可以应用图像处理技术来提取车辆数量、密度、速度等视觉特征。具体地,假设目标高速监测数据包括某一时段内某一路段的车辆通过数量和车速数据。边缘盒子首先会对这些数据进行清洗和整理,去除无效或异常的数据点。然后,边缘盒子可以提取以下特征:车辆通过数量特征:直接使用传感器记录的车辆通过数量作为特征值。平均车速特征:计算该时段内所有车速数据的平均值,反映该路段的平均行驶速度。车速标准差特征:计算车速数据的标准差,反映车速的波动情况。这些提取出的特征将形成目标车流量特征表示,用于后续的车辆流量预测模型。通过步骤S70的数据特征提取操作,边缘盒子能够从原始的高速监测数据中提取出与目标车辆流量预测密切相关的特征信息,为后续的预测提供有力的支持。

[0056] 步骤S80:依据所述目标高速监测数据的目标车流量特征表示,通过部署的对应的目标车辆流量预测网络,对所述目标高速监测数据进行分类,获得所述目标高速监测数据

的目标车辆流量预测结果。

[0057] 在步骤S80中,边缘盒子将根据目标高速监测数据的目标车流量特征表示,利用已经部署在边缘盒子上的目标车辆流量预测网络,对目标高速监测数据进行分类预测,从而得到目标车辆流量预测结果。这一步骤是边缘协同数据处理流程中的关键一环,直接决定了最终预测结果的准确性和实用性。

[0058] 具体来说,当边缘盒子获取到目标高速监测数据后,对其进行数据特征提取操作(步骤S70),得到目标车流量特征表示。这个特征表示包含了与目标车辆流量密切相关的关键信息,如车辆通过数量、平均车速、道路占用率等。接下来,边缘盒子将把这些特征表示作为输入,传递给部署在边缘盒子上的目标车辆流量预测网络。这个预测网络是在多个云平台上通过分布式训练得到的,已经学习到了从特征表示到车辆流量的映射关系。因此,当接收到新的特征表示时,预测网络能够迅速计算出对应的车辆流量预测结果。

[0059] 在预测过程中,预测网络会对输入的特征表示进行一系列的计算和推理。这个过程可能涉及到多层神经网络的前向传播,每一层都会根据上一层的输出和自身的参数计算出新的输出。最终,网络会输出一个或多个预测值,这些预测值代表了目标高速监测数据对应的车辆流量情况。

[0060] 以一个具体的例子来说,假设边缘盒子获取到了一组目标高速监测数据,经过数据特征提取后得到了一个包含车辆通过数量、平均车速和道路占用率等特征的特征向量。这个特征向量被输入到目标车辆流量预测网络中,网络经过一系列的计算后输出了一个预测值,比如“中流量”。这个预测值就是目标车辆流量预测结果,它表示在未来一段时间内,该路段的车辆流量将处于中等水平。

[0061] 需要注意的是,由于高速公路的交通状况是实时变化的,因此边缘盒子需要不断地获取新的目标高速监测数据,并实时地进行车辆流量预测。这样,交通管理部门就能够根据预测结果及时调整交通调度策略,优化道路资源的使用效率,提高交通运行的效率和安全性。

[0062] 作为一种实施方式,在上述的网络训练过程中,所述各个云平台中的初始网络内部参变量一致,所述P个高速监测数据集中的各个高速监测数据集包括X/P个高速监测数据。

[0063] 在一种实施方式中,云平台确保各个云平台在训练车辆流量预测网络时,其初始网络内部参变量(例如神经网络的权重和偏置)保持一致。同时,将获取到的X个高速监测数据均匀分成P个高速监测数据集,每个数据集包含X/P个高速监测数据。这种实施方式旨在实现云平台间的协同学习,并确保各云平台能够从整个数据集的不同部分中学习到相似的车辆流量预测模式。

[0064] 在机器学习中,模型的初始参数对于训练结果有着重要影响。当多个云平台使用相同的初始网络内部参变量开始训练时,它们能够在后续的训练过程中保持一定的同步性。这意味着,尽管各云平台处理的数据子集可能不同,但由于初始参数相同,它们学习到的特征表示和预测模式会具有更高的相似性。这对于后续的模型聚合和参数更新是非常有利的。

[0065] 以高速公路边缘协同数据处理为例,假设X为10000个高速监测数据样本,P为5个云平台。云平台首先会将这10000个样本均匀分成5个数据集,每个数据集包含2000个样本。

然后,为这5个云平台分别初始化车辆流量预测网络,并确保它们的初始网络内部参变量完全一致。

[0066] 接下来,每个云平台将使用各自的数据集进行网络训练。由于初始参变量相同,各云平台上的网络在训练初期会表现出相似的性能。随着训练的进行,每个云平台上的网络将逐渐学习到其数据集上特有的车辆流量模式。但由于初始参数的一致性,这些网络在整体结构和参数更新方向上仍会保持一定的同步性。这种实施方式的好处在于,它充分利用了多个云平台的计算能力,实现了对大规模数据集的并行处理。同时,通过保持初始网络内部参变量的一致性,确保了各云平台训练得到的网络模型在结构和性能上具有较高的相似性。这为后续的模型聚合和参数更新提供了便利,有助于提升整个车辆流量预测系统的性能和稳定性。

[0067] 在实际应用中,这种实施方式可以帮助交通管理部门更加准确地预测高速公路的车辆流量情况。通过实时获取和处理高速监测数据,各云平台能够迅速响应交通状况的变化,并提供准确的预测结果。这对于优化交通调度、缓解交通拥堵、提高道路使用效率等方面都具有重要意义。

[0068] 作为一种实施方式,所述步骤S40,依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,包括:

[0069] 步骤S41:通过所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量的聚合参变量,调节所述各个云平台中的网络内部参变量。

[0070] 在步骤S41中,云平台通过P个云平台在完成第E轮重复调试后获得的网络内部参变量的聚合参变量来调节各个云平台中的网络内部参变量。这一步骤是联邦学习(Federated Learning)中的一个典型应用,旨在通过多个云平台协同工作,共同优化车辆流量预测网络,同时保护数据隐私。

[0071] 具体来说,假设P为5,即有5个云平台参与车辆流量预测网络的训练。在每个云平台上,都运行着一个车辆流量预测网络,该网络可能是基于深度学习(如卷积神经网络CNN或循环神经网络RNN)构建的。这些网络在训练过程中,会根据各自处理的高速监测数据集,不断地更新其内部参变量(即神经网络的权重和偏置)。在第E轮重复调试完成后,每个云平台都获得了一组更新后的网络内部参变量。这些参变量代表了云平台在当前数据集上的学习成果。然而,由于不同云平台处理的数据集可能存在差异,因此它们获得的参变量也可能有所不同。

[0072] 为了整合这些差异并得到一个全局的更新方向,云平台采用了聚合策略。聚合策略的选择可以根据具体应用场景和需求来确定,常见的聚合策略包括取平均值、取中位数、联邦平均(Federated Averaging)等。在本例中,假设采用取平均值作为聚合策略。云平台首先会收集所有云平台在第E轮调试后获得的网络内部参变量,并将它们对应位置上的值进行平均。例如,假设每个云平台都有一个权重矩阵W,其中W1、W2、W3、W4和W5分别表示5个云平台在第E轮调试后的权重矩阵。云平台计算这些矩阵对应位置上的元素平均值,得到一个聚合权重矩阵W_avg。

[0073] 其中, $W_avg = (W1 + W2 + W3 + W4 + W5) / 5$ 。

[0074] 这个聚合权重矩阵 W_{avg} 就代表了所有云平台在第E轮调试中的共同学习成果,是一个全局的更新方向。接下来,云平台将这个聚合权重矩阵 W_{avg} 作为新的网络内部参变量,去调节各个云平台中的网络内部参变量。具体地,每个云平台都会将其当前的权重矩阵更新为聚合权重矩阵 W_{avg} ,从而使其网络参数与全局更新方向保持一致。

[0075] 通过这一步骤,云平台实现了多个云平台之间的协同学习。每个云平台都贡献了自己的学习成果,通过聚合策略整合为一个全局的更新方向,并据此调节各自的网络参数。这种方式不仅能够充分利用各个云平台的计算能力,加速网络训练过程,还能够提高网络模型的泛化能力,使其能够更好地适应实际高速公路交通状况的变化。同时,由于数据不需要离开云平台,因此也保护了数据隐私。

[0076] 步骤S42:依据各个云平台的车流量特征表示样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络。

[0077] 在步骤S42中,云平台依据各个云平台的车流量特征表示样本集合,对各个云平台中的网络内部参变量进行第F轮($F=E+1$)重复调试。这一步骤是车辆流量预测网络训练过程中的关键一环,旨在通过迭代优化来提高网络的预测性能。

[0078] 一、调试过程的初始化。

[0079] 在开始第F轮重复调试之前,云平台已经从步骤S41中获得了经过聚合参变量调节后的网络内部参变量。这些参变量作为本轮调试的初始值,将被用于指导网络的参数更新。

[0080] 二、基于车流量特征表示样本集合的调试。

[0081] 每个云平台都拥有其独特的车流量特征表示样本集合,这些样本集合包含了从高速监测数据中提取出的关键特征,如车辆通过数量、平均车速、道路占用率等。这些特征对于车辆流量的预测至关重要。在第F轮调试中,云平台将使用每个云平台的车流量特征表示样本集合作为训练数据,输入到对应的车辆流量预测网络中。网络根据当前的网络内部参变量进行前向传播计算,得到预测的车辆流量值。然后,云平台将预测值与真实值进行比较,计算预测误差(损失函数值)。

[0082] 三、反向传播与参数更新。

[0083] 在得到预测误差后,云平台将进行反向传播计算,通过链式法则逐层计算网络中每个参数的梯度。梯度代表了参数对损失函数的影响程度,指导了参数更新的方向。接着,云平台使用优化算法(如梯度下降算法)根据计算得到的梯度来更新网络内部参变量。优化算法通过调整参数的大小来减小预测误差,使网络逐渐逼近最优解。

[0084] 四、迭代调试与收敛判断。

[0085] 上述的前向传播、计算误差、反向传播和参数更新过程将不断重复进行,构成了一个迭代优化的循环。在每一轮调试中,网络都会根据新的训练数据更新其内部参变量,以提高预测性能。然而,迭代过程并不是无限进行的。当满足某个停止条件时,迭代过程将终止。停止条件可以是达到事先确定的最大重复轮次(例如设定为100轮),也可以是网络内部参变量收敛(即参数变化量小于某个预设阈值)。当满足这些条件之一时,云平台认为网络已经得到了充分的训练,并获得了调试好的目标车辆流量预测网络。

[0086] 例如,假设有3个云平台参与车辆流量预测网络的训练,每个云平台都拥有其独特

的车流量特征表示样本集合。在第F轮调试中,每个云平台都将其样本集合输入到对应的网络中,进行前向传播计算得到预测值。然后,每个云平台分别计算预测误差,并进行反向传播和参数更新。经过多轮迭代调试后,当所有云平台的网络内部参变量变化量均小于预设阈值时,迭代过程终止,获得了3个调试好的目标车辆流量预测网络。这些网络将被部署到边缘盒子中,用于实时的车辆流量预测。

[0087] 作为一种实施方式,所述步骤S42,依据所述各个云平台的车流量特征表示样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络,包括:

[0088] 步骤S421:将所述各个云平台中的网络内部参变量与所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示进行作积,将作积结果与所述各个云平台的车流量特征表示样本集合中的先验标记进行作差,获得作差结果,将所述作差结果与所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示进行作积,获得所述各个云平台中的网络内部参变量的偏导向量。

[0089] 在步骤S421中,假设有P个云平台,每个云平台都维护一个车辆流量预测网络,并且每个网络都包含一组网络内部参变量,这些参变量通常包括神经网络的权重和偏置。同时,每个云平台都拥有其独特的车流量特征表示样本集合,这些样本集合包含了从高速监测数据中提取的关键特征,如车辆数量、车速、道路占用率等。步骤S421的执行过程分为几个子步骤:

[0090] 特征表示与网络参变量的作积:

[0091] 对于每个云平台,云平台首先会将其车流量特征表示样本集合中的目标车流量特征表示与网络内部参变量进行作积。这里的目标车流量特征表示通常是一个特征向量,包含了多个与车辆流量相关的特征值。而网络内部参变量则代表了网络当前的状态,即如何从输入特征映射到输出预测值。作积操作实际上是网络的前向传播过程,它模拟了网络根据当前参数对输入特征的响应。

[0092] 举例来说,假设某个云平台上的车辆流量预测网络是一个简单的全连接神经网络,其权重矩阵为 W ,偏置向量为 b ,而某个目标车流量特征表示为 x 。那么前向传播的过程就是将 x 与 W 进行矩阵乘法,再加上 b ,得到一个输出预测值 y 。

[0093] 预测值与先验标记的作差:

[0094] 接下来,云平台将前向传播得到的预测值与车流量特征表示样本集合中的先验标记进行作差。先验标记是真实的车辆流量数据,用于与网络预测值进行比较,从而评估网络的预测性能。作差操作得到的结果是一个误差向量,它反映了网络预测值与真实值之间的差异。

[0095] 继续上面的例子,假设先验标记为 y_{true} ,那么误差向量 $e=y_{true}-y$ 。

[0096] 误差向量与目标车流量特征表示的作积:

[0097] 最后,云平台将误差向量与目标车流量特征表示进行作积,得到的结果就是网络内部参变量的偏导向量(梯度)。这个偏导向量指示了在当前参数下,网络应该如何调整其内部参变量以减小预测误差。具体来说,偏导向量的每个元素都对应了网络内部参变量中的一个参数(如权重或偏置),其值的大小和方向表示了该参数应该调整的量 and 方向。

[0098] 在上面的例子中,偏导向量 g 可以通过误差向量 e 与目标车流量特征表示 x 的作积得到,即 $g = e * x$ (这里的“ $*$ ”表示按元素乘法)。需要注意的是,在实际应用中,由于网络内部参变量可能包含大量的权重和偏置,因此作积操作通常是基于矩阵和向量的,并且可能涉及到更复杂的计算过程。通过步骤S421的计算,云平台得到了每个云平台中车辆流量预测网络的偏导向量,这些偏导向量将用于后续的参数更新过程,以逐步优化网络的预测性能。

[0099] 步骤S422:将所述各个云平台中的网络内部参变量的调谐参数和所述各个云平台中的网络内部参变量的偏导向量进行作积,获得所述各个云平台中的网络内部参变量的调节结果。

[0100] 在步骤S422中,云平台利用梯度下降算法的核心思想,通过计算网络内部参变量的调节结果来指导网络参数的更新。这一步是优化车辆流量预测网络性能的关键环节,它确保网络能够朝着减小预测误差的方向逐步调整其内部参数。网络内部参变量的调谐参数,通常被称为学习率(learning rate),它决定了网络在每次迭代中参数更新的步长。一个合适的学习率能够确保网络稳定且高效地收敛到最优解。偏导向量(即梯度)则是在步骤S421中计算得出的,它指示了在当前参数下,网络应该如何调整其内部参变量以减小预测误差。

[0101] 本申请实施例中,假设每个云平台上的车辆流量预测网络都经过了前向传播和误差计算,得到了各自的偏导向量。这个偏导向量是一个与网络内部参变量维度相同的向量,它的每个元素都对应了网络参数中的一个权重或偏置项,并指示了这些参数应该调整的方向和大小。接下来,云平台将执行步骤S422,即将每个云平台中的网络内部参变量的调谐参数(学习率)和偏导向量进行作积。这个作积操作实际上是对偏向导量的每个元素进行缩放,缩放因子就是学习率。学习率的大小决定了参数更新的幅度,较大的学习率意味着每次迭代中参数更新的步长较大,可能加快收敛速度但也可能导致网络在最优解附近震荡;较小的学习率则可能导致收敛速度较慢但更加稳定。

[0102] 举例来说,假设某个云平台上的车辆流量预测网络有一个权重矩阵 W ,其对应的偏导向量为 g ,调谐参数为 η 。那么,该云平台上的网络内部参变量的调节结果 ΔW 可以通过以下公式计算得出: $\Delta W = \eta * g$ 。

[0103] 这里的 ΔW 是一个与 W 维度相同的矩阵,它的每个元素都表示在下一轮迭代中应该调整的量。通过计算 ΔW ,云平台得到了网络参数更新的指导信息。最后,云平台将使用调节结果 ΔW 来更新网络内部参变量 W ,即执行 $W = W - \Delta W$ 的操作。这样,网络就根据最新的训练数据调整了其内部参数,以减小预测误差并提高预测性能。通过多轮迭代和参数更新,车辆流量预测网络能够逐渐学习到从输入特征到输出预测值的准确映射关系。

[0104] 步骤S423:通过所述各个云平台中的网络内部参变量减去所述各个云平台中的网络内部参变量的调节结果,获得所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量。

[0105] 在步骤S423中,云平台通过从各个云平台中的网络内部参变量中减去对应的调节结果,来获得更新后的网络内部参变量。这一步骤是梯度下降算法中的核心环节,用于指导网络模型向减小预测误差的方向进行迭代优化。具体来说,本申请实施例中,每个云平台都维护着一个车辆流量预测网络,该网络包含一组网络内部参变量,如神经网络的权重和偏

置。在步骤S422中,云平台已经计算出了每个云平台上的网络内部参变量的调节结果,这个结果是根据网络的偏导向量和调谐参数计算得出的,它指示了在当前参数下网络应该如何调整其内部参变量以减小预测误差。

[0106] 在步骤S423中,对于每个云平台上的车辆流量预测网络,云平台会从当前的网络内部参变量中减去对应的调节结果。这个减法操作实际上是在网络参数空间中沿着梯度下降的方向进行了一次小步移动,从而使网络参数逐渐逼近最优解。举例来说,假设某个云平台上的车辆流量预测网络有一个权重矩阵 W ,其对应的调节结果为 ΔW 。在步骤S423中,云平台将执行 $W = W - \Delta W$ 的操作,即从原始的权重矩阵 W 中减去调节结果 ΔW ,得到更新后的权重矩阵 W' 。这个 W' 就是该云平台在第 F 轮重复调试后获得的新的网络内部参变量,它包含了网络根据最新训练数据调整后的参数信息。

[0107] 通过执行步骤S423,云平台完成了对各个云平台上的车辆流量预测网络的参数更新。这些更新后的网络内部参变量将用于下一轮的前向传播和误差计算,以进一步迭代优化网络模型。通过多轮重复调试和参数更新,车辆流量预测网络能够逐渐学习到从输入特征到输出预测值的准确映射关系,从而提高预测的准确性和鲁棒性。

[0108] 步骤S424:如果 F 等于所述事先确定的最大重复轮次或所述各个云平台中的网络内部参变量的偏导向量等于0时,获得所述 P 个目标车辆流量预测网络。

[0109] 在步骤S424中,云平台根据迭代调试的轮次 F 以及各个云平台中网络内部参变量的偏导向量来判断是否达到了停止条件,从而确定是否获得了目标车辆流量预测网络。这一步骤是车辆流量预测网络训练过程中的重要环节,确保网络在充分学习和优化后停止迭代,避免过拟合或其他不必要的计算资源消耗。

[0110] 具体来说,本申请实施例中,每个云平台都运行着一个车辆流量预测网络,该网络通过多轮迭代调试来优化其内部参变量(如神经网络的权重和偏置)。在每轮迭代中,云平台都会计算网络的偏导向量,并根据偏导向量和学习率来更新网络参数。这个过程会重复进行,直到满足某个停止条件。

[0111] 步骤S424中定义了两个停止条件:

[0112] 迭代轮次达到最大重复轮次:在训练开始前,通常会设定一个最大重复轮次 F_{\max} ,作为迭代过程的上限。这个值可以根据具体的应用场景和计算资源来确定。当迭代轮次 F 达到或超过 F_{\max} 时,云平台认为网络已经经过了充分的训练,可以停止迭代并输出当前的网络作为目标车辆流量预测网络。

[0113] 例如,假设设定最大重复轮次 F_{\max} 为100轮。如果系统在第100轮迭代后仍然没有达到其他停止条件(如偏导向量接近于0),则云平台会在第100轮迭代后停止训练,并将此时的网络作为目标车辆流量预测网络。

[0114] 偏导向量接近于0:偏导向量反映了网络参数在特征空间上应该调整的方向和大小。当偏导向量接近于0时,意味着网络参数在当前数据上的梯度已经非常小,即网络参数已经接近或达到了最优解。此时,继续迭代可能不会对网络的性能产生显著提升,反而会浪费计算资源。因此,当偏导向量小于某个预设的阈值 ϵ 时(例如 $\epsilon=1e-6$),云平台也会停止迭代并输出当前的网络作为目标车辆流量预测网络。举例来说,假设在某个迭代轮次中,云平台计算得到的偏导向量的所有元素都小于 $1e-6$ 。这意味着网络参数在当前数据上的梯度已经非常小,云平台可以认为网络已经收敛到了最优解附近。因此,云平台会停止迭代,并将

此时的网络作为目标车辆流量预测网络。

[0115] 通过步骤S424的判断,云平台能够确保在车辆流量预测网络的训练过程中既充分利用了计算资源进行优化,又避免了不必要的资源浪费。这有助于提高网络的预测性能,同时降低训练成本。

[0116] 步骤S425:如果F小于所述事先确定的最大重复轮次或所述各个云平台中的网络内部参变量的偏导向量等于0时,则依据所述各个云平台的车流量特征表示样本集合和所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量,对所述各个云平台中的网络内部参变量进行第G次重复调试,直到所述各个云平台中的网络内部参变量的重复的轮次等于事先确定的最大重复轮次或所述各个云平台中的网络内部参变量收敛时,获得所述P个目标车辆流量预测网络,其中, $G=F+1$ 。

[0117] 在步骤S425中,云平台根据当前的迭代轮次F和各个云平台中网络内部参变量的偏导向量来判断是否需要进行下一轮的重复调试。这是确保车辆流量预测网络能够达到最优性能的关键步骤,通过不断迭代调试,网络系统能够逐步逼近最优解。

[0118] 具体来说,本申请实施例中,每个云平台都运行着一个车辆流量预测网络,并且这些网络都经历了多轮的迭代调试。在每次迭代中,云平台都会根据当前的网络内部参变量和车流量特征表示样本集合来计算网络的偏导向量,并根据偏导向量和调谐参数来更新网络参数。这个过程会重复进行,直到满足某个停止条件。

[0119] 然而,在步骤S424中,如果当前迭代轮次F小于事先确定的最大重复轮次 F_{\max} ,或者虽然F达到了 F_{\max} 但各个云平台中网络内部参变量的偏导向量仍然不等于0(即网络尚未收敛),则云平台需要继续执行步骤S425,进行下一轮的迭代调试。在步骤S425中,云平台首先会收集各个云平台在第F轮迭代后获得的网络内部参变量和对应的车流量特征表示样本集合。然后,云平台会将这些参变量和样本集合作为输入,开始第G轮($G=F+1$)的迭代调试。这一轮调试的过程与之前的迭代类似,包括前向传播计算预测值、计算预测误差、反向传播计算偏导向量以及更新网络参数等步骤。

[0120] 通过不断重复步骤S425中的迭代调试过程,车辆流量预测网络能够逐步学习到从输入特征到输出预测值的准确映射关系。随着迭代轮次的增加,网络的预测性能也会逐渐提高。当满足停止条件(即迭代轮次达到 F_{\max} 或网络内部参变量收敛)时,云平台认为网络已经得到了充分的训练,并将此时的网络作为目标车辆流量预测网络输出。

[0121] 举例来说,假设最大重复轮次 F_{\max} 设定为100轮,当前迭代轮次F为95轮。在第95轮迭代后,云平台计算得到的偏导向量仍然不等于0,这意味着网络尚未收敛到最优解。因此,云平台会执行步骤S425,开始第96轮($G=95+1$)的迭代调试。这一轮调试过程中,云平台会根据第95轮迭代后的网络内部参变量和车流量特征表示样本集合来更新网络参数,并继续提高网络的预测性能。当迭代轮次达到 F_{\max} 或网络内部参变量收敛时,云平台停止迭代并输出目标车辆流量预测网络。

[0122] 作为另一种实施方式,所述步骤S40,依据所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的车流量特征表示样本集合对所述各个云平台中调节后的网络内部参变量进行第F轮重复调试,获得调试好的P个目标车辆流量预测网络,包括:

[0123] 步骤S4A:对各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个

云平台的学习样本集合和所述各个云平台的验证样本集合。

[0124] 在步骤S4A中,云平台对各个云平台上的车流量特征表示样本集合进行稀疏提取,也就是通过采样技术,将原始的样本集合划分为学习样本集合(Training Set)和验证样本集合(Test Set)。这一步骤是机器学习中常见的数据预处理步骤,特别是在使用监督学习算法训练模型时,将数据集划分为训练集和测试集是确保模型泛化性能的关键。本申请实施例中,每个云平台都拥有大量的车流量特征表示样本,这些样本是从高速监测数据中提取出来的,包含了与车辆流量预测相关的关键特征。然而,直接使用全部样本进行模型训练往往不是最高效的,因为过多的样本可能导致训练时间过长,而且可能包含冗余信息或噪声。因此,通过稀疏提取将样本划分为训练集和测试集,可以更有效地利用样本数据。

[0125] 具体来说,云平台首先确定一个采样策略,这个策略可以是基于样本数量的随机采样,也可以是基于样本特征分布的分层采样等。然后,根据这个策略,从原始的车流量特征表示样本集合中随机选择一部分样本作为学习样本集合,用于训练车辆流量预测网络。剩余的样本则作为验证样本集合,用于在训练过程中评估网络的性能。例如,假设某个云平台拥有10000条车流量特征表示样本。云平台可以采用随机采样的策略,从中选择80%的样本(即8000条)作为学习样本集合,用于训练车辆流量预测网络;剩余的20%的样本(即2000条)作为验证样本集合,用于在训练过程中评估网络的预测误差。通过步骤S4A的稀疏提取操作,云平台不仅将原始的样本集合划分为训练集和测试集,还确保了训练集和测试集之间的独立性。这使得在训练过程中能够更准确地评估模型的性能,并有助于防止过拟合现象的发生。同时,通过合理地划分训练集和测试集,云平台能够更有效地利用有限的计算资源,加速模型的训练和优化过程。

[0126] 步骤S4B:通过所述P个云平台中完成第E轮重复调试后获得的P个网络内部参变量的聚合参变量,调节所述各个云平台中的网络内部参变量。

[0127] 在步骤S4B中,云平台通过P个云平台在第E轮重复调试后获得的网络内部参变量的聚合参变量来调节各个云平台中的网络内部参变量。这一步骤体现了联邦学习的核心思想,即在保护数据隐私的同时,通过多个云平台协同工作来优化全局模型。本申请实施例中,每个云平台都运行着一个车辆流量预测网络,这些网络通过训练学习从输入特征到车辆流量的映射关系。然而,由于各个云平台处理的数据集可能不同,因此它们训练得到的网络内部参变量(如神经网络的权重和偏置)也可能存在差异。为了充分利用这些差异并得到一个全局最优的模型,云平台采用了聚合参变量的策略。

[0128] 具体来说,在第E轮重复调试完成后,每个云平台都获得了一组更新后的网络内部参变量。这些参变量代表了云平台在当前数据集上的学习成果。为了整合这些成果,云平台采用了一种聚合策略(如取最小参数、平均值、取中位数等),将P个云平台上的网络内部参变量进行聚合,得到一个聚合参变量。这个聚合参变量代表了所有云平台在第E轮调试中的共同学习成果,是一个全局的更新方向。接下来,云平台使用这个聚合参变量来调节各个云平台中的网络内部参变量。这个过程实际上是将每个云平台上的网络参数更新为聚合参变量的值。通过这一步骤,每个云平台上的车辆流量预测网络都融入了其他云平台的学习成果,使得整个系统能够在全局范围内优化模型性能。

[0129] 举例来说,假设有3个云平台参与车辆流量预测网络的训练,每个云平台在第E轮调试后都得到了更新后的网络内部参变量,分别为W1、W2和W3(这里W1、W2、W3表示神经网络

的权重矩阵的简化表示)。云平台采用取平均值的聚合策略,计算得到聚合参变量 $W_{avg} = (W1+W2+W3)/3$ 。然后,云平台将这个聚合参变量 W_{avg} 作为新的网络内部参变量,分别更新到3个云平台上的车辆流量预测网络中,即 $W1'=W_{avg}$, $W2'=W_{avg}$, $W3'=W_{avg}$ 。这样,3个云平台上的网络参数就得到了统一的更新,反映了所有云平台在第E轮调试中的共同学习成果。通过步骤S4B的执行,云平台实现了多个云平台之间的协同学习,充分利用了各个云平台的数据和计算资源,提高了车辆流量预测网络的性能和泛化能力。同时,由于数据不需要离开云平台,因此也保护了数据隐私和安全。

[0130] 步骤S4C:依据所述各个云平台的学习样本集合,对所述各个云平台中的网络内部参变量进行第F轮重复调试,获得所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量。

[0131] 在步骤S4C中,云平台根据各个云平台的学习样本集合,对各个云平台中的车辆流量预测网络进行第F轮重复调试。这一步骤是机器学习模型训练过程中的核心环节,通过迭代优化网络内部参变量(如神经网络的权重和偏置),使网络能够更准确地预测高速公路上的车辆流量。

[0132] 本申请实施例中,每个云平台都维护着一个车辆流量预测网络,并持有一个学习样本集合,该集合是通过稀疏提取从原始车流量特征表示样本集中获得的。学习样本集合包含了用于训练网络模型的关键特征数据,如车辆通过数量、平均车速、道路占用率等,以及对应的真实车辆流量标签。

[0133] 在步骤S4C中,云平台首先会将每个云平台上的学习样本集合输入到对应的车辆流量预测网络中。网络根据当前的内部参变量(即神经网络的权重和偏置)对输入样本进行前向传播计算,生成预测的车辆流量值。然后,云平台会比较预测值与真实标签之间的差异,计算损失函数值(如均方误差MSE),以量化网络在当前参数下的预测性能。接下来,云平台使用反向传播算法计算损失函数关于网络内部参变量的梯度(即偏导向量)。这个梯度指明了在当前参数下,网络应该如何调整其内部参变量以减小预测误差。基于这个梯度信息,云平台会使用优化算法(如梯度下降算法)来更新网络内部参变量,使得网络的预测性能在下一轮迭代中得到提升。

[0134] 具体来说,云平台会将学习率与梯度进行相乘,得到每个内部参变量的更新量(即调节结果)。然后,云平台会从当前的内部参变量中减去这个更新量,得到更新后的内部参变量。这个过程会针对网络中的所有参数进行,从而实现对整个网络的参数更新。完成一轮参数更新后,云平台就完成了一次迭代调试。在步骤S4C中,这个过程会重复进行F轮,以确保网络能够充分学习到数据中的模式,并达到较好的预测性能。在每一轮迭代中,云平台都会使用最新的学习样本集合来评估网络的性能,并根据评估结果来指导下一轮迭代中的参数更新。

[0135] 通过步骤S4C的迭代调试过程,云平台能够逐步优化各个云平台上的车辆流量预测网络,使其能够更准确地预测高速公路上的车辆流量。这有助于交通管理部门做出更科学的决策,提高道路资源的利用效率,缓解交通拥堵等问题。

[0136] 步骤S4D:依据所述各个云平台的验证样本集合,对所述各个云平台中的网络内部参变量的误差进行评估,获得用以指示所述P个云平台中的网络内部参变量的误差的目标误差值。

[0137] 在步骤S4D中,云平台依据各个云平台的验证样本集合,对经过第F轮重复调试后各个云平台中的车辆流量预测网络的内部参变量进行误差评估。这一步骤是机器学习模型训练过程中的重要环节,用于验证网络模型在未见过的数据上的性能,并据此判断模型是否达到了预设的训练标准。

[0138] 本申请实施例中,每个云平台都持有一个验证样本集合,这是通过稀疏提取(即采样)从原始车流量特征表示样本集中获得的,并且与用于训练的学习样本集合独立。验证样本集合包含了真实的高速公路交通数据,用于模拟模型在实际应用中的表现。在步骤S4D中,云平台首先将各个云平台的验证样本集合输入到对应的车辆流量预测网络中。网络根据当前的内部参变量(即神经网络的权重和偏置)对输入样本进行前向传播计算,生成预测的车辆流量值。然后,云平台会比较预测值与验证样本中的真实车辆流量值之间的差异,计算损失函数值(如均方误差MSE),以量化网络在当前参数下的预测误差。

[0139] 损失函数值(目标误差值)是评估网络性能的关键指标。它反映了网络模型在验证样本集合上的预测准确性,即模型对未见过的数据的泛化能力。一个较小的损失函数值意味着网络模型在验证样本上的预测误差较小,性能较好。

[0140] 具体来说,假设某个云平台的验证样本集合包含1000条数据,每条数据都有一个对应的真实车辆流量值。在将验证样本输入到车辆流量预测网络后,网络会生成1000个预测的车辆流量值。然后,云平台会计算这1000个预测值与真实值之间的均方误差MSE,得到该云平台上的目标误差值。

[0141] 通过步骤S4D的误差评估过程,云平台能够了解各个云平台上的车辆流量预测网络在验证样本上的性能表现。这为后续的模型选择和调优提供了重要依据。如果目标误差值较小,说明网络已经达到了较好的预测性能;如果目标误差值较大,则可能需要继续迭代调试网络参数或调整网络结构以优化性能。

[0142] 步骤S4E:判断所述目标误差值是否符合评估条件,如果所述目标误差值符合所述评估条件,获得所述P个目标车辆流量预测网络;如果所述目标误差值不符合评估条件,则通过所述P个云平台中完成第F轮重复调试后获得的P个网络内部参变量的聚合参变量调节所述各个云平台中的网络内部参变量,以及依据所述各个云平台的学习样本集合,对所述各个云平台中的网络内部参变量进行第G次重复调试,直到所述目标误差值符合所述评估条件时,获得所述P个目标车辆流量预测网络,其中, $G=F+1$ 。

[0143] 在步骤S4E中,云平台对步骤S4D中计算得到的目标误差值进行判断,以确定当前各个云平台上的车辆流量预测网络是否已经达到了预设的训练标准。这一步骤是模型训练过程中的关键决策点,它决定了是否需要继续迭代调试网络参数以优化性能。具体来说,在步骤S4D中,云平台已经依据各个云平台的验证样本集合计算出了目标误差值,该值量化了网络模型在验证样本上的预测误差。在步骤S4E中,云平台将这个目标误差值与预设的评估条件进行比较。评估条件通常是一个具体的数值阈值,用于判断网络模型的性能是否满足要求。

[0144] 例如,假设预设的评估条件是目标误差值不大于0.01。如果某个云平台上的目标误差值小于或等于0.01,则说明该云平台上的车辆流量预测网络已经达到了预设的训练标准,可以停止迭代调试,并将当前的网络模型作为目标车辆流量预测网络输出。然而,如果目标误差值大于0.01,则说明网络模型的性能尚未满足要求,需要继续迭代调试。在这种情

况下,云平台将执行以下操作:

[0145] 聚合参变量调节:首先,云平台收集各个云平台在第F轮重复调试后获得的网络内部参变量,并计算这些参变量的聚合参变量。聚合参变量代表了所有云平台在当前迭代轮次中的共同学习成果,是一个全局的更新方向。然后,云平台使用这个聚合参变量来调节各个云平台上的网络内部参变量,使得每个云平台上的网络都能融入其他云平台的学习成果。

[0146] 重复调试:调节完网络内部参变量后,云平台将依据各个云平台的学习样本集合进行第G轮($G=F+1$)重复调试。这一轮调试的过程与之前的迭代类似,包括前向传播计算预测值、计算损失函数值、反向传播计算梯度以及更新网络参数等步骤。通过这一轮调试,网络模型将尝试进一步减小预测误差,提高性能。

[0147] 这个过程将不断重复进行,直到目标误差值满足评估条件为止。在每一轮迭代中,云平台都会根据最新的验证样本集合计算目标误差值,并根据误差值的大小来指导下一轮迭代中的参数更新方向。通过这种方式,云平台能够逐步优化各个云平台上的车辆流量预测网络,使其性能逐渐逼近最优解。最终,当所有云平台上的目标误差值都满足评估条件时,云平台将停止迭代调试,并将此时的网络模型作为目标车辆流量预测网络输出。这些目标车辆流量预测网络将具有更高的预测准确性和泛化能力,能够更好地适应实际高速公路交通状况的变化。

[0148] 作为一种实施方式,所述步骤S4A,对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合,包括:

[0149] 步骤S4A1:依据所述各个云平台的车流量特征表示样本集合的类别,基于设定的百分比通过对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合。

[0150] 在步骤S4A的实施方式中,云平台通过对各个云平台上的车流量特征表示样本集合进行稀疏提取,来划分学习样本集合和验证样本集合。这一步骤的关键在于如何根据样本集合的特性和需求来合理地进行稀疏提取。

[0151] 步骤S4A1中,云平台首先会考虑各个云平台上的车流量特征表示样本集合的类别。本申请实施例中,样本集合可能包含了多种不同类型的数据,例如不同时间段、不同路段、不同天气条件下的车流量数据。这些数据可能具有不同的统计特性和分布规律,因此在进行稀疏提取时需要考虑这些差异。接下来,基于设定的百分比,云平台对各个云平台上的车流量特征表示样本集合进行稀疏提取。这个百分比通常是根据实际需求和经验来设定的,用于确定学习样本集合和验证样本集合的大小比例。例如,可以设定80%的样本作为学习样本集合,用于训练车辆流量预测网络;剩余的20%的样本作为验证样本集合,用于在训练过程中评估网络的性能。在具体操作时,云平台根据样本集合的类别和设定的百分比,随机选择相应数量的样本作为学习样本集合和验证样本集合。这个过程通常是随机的,以确保学习样本集合和验证样本集合的独立性,并避免可能的偏差。

[0152] 举例来说,假设某个云平台上的车流量特征表示样本集合包含了10000条数据,其中涵盖了不同时间段和不同路段的车流量数据。云平台设定80%的样本作为学习样本集合,即选择8000条数据作为学习样本;剩余的20%的样本,即2000条数据作为验证样本集合。在

选择过程中,云平台会确保不同类别(如不同时间段、不同路段)的样本在学习样本集合和验证样本集合中的分布比例大致相同,以保持样本的多样性和代表性。

[0153] 通过步骤S4A1的稀疏提取操作,云平台能够有效地划分学习样本集合和验证样本集合,为后续的车辆流量预测网络训练和性能评估提供有力支持。同时,由于考虑了样本集合的类别和分布特性,这种划分方式还能够确保样本的多样性和代表性,提高模型的泛化能力和预测准确性。

[0154] 作为一种实施方式,所述步骤S4A1,依据所述各个云平台的车流量特征表示样本集合的类别,基于设定的百分比通过对所述各个云平台的车流量特征表示样本集合进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合,包括:

[0155] 步骤S4A11:如果所述各个云平台的车流量特征表示样本集合的类别为符合特征空间一致但样本空间各异的分布式数据类别,则基于所述设定的百分比通过对所述各个云平台的车流量特征表示样本集合中的目标车流量特征表示和先验标记进行稀疏提取,获得所述各个云平台的学习样本集合和所述各个云平台的验证样本集合。

[0156] 在步骤S4A11中,云平台针对的是一种特定的分布式数据类别,即特征空间一致但样本空间各异的情况,这通常出现在横向联邦学习的场景中。在高速公路边缘协同数据处理的背景下,这种情况意味着不同的云平台虽然都收集了关于车流量的特征数据(如车速、车流量、道路占用率等),但这些数据来自不同的高速公路路段或不同的时间段,因此样本空间各异。

[0157] 为了有效地利用这些数据来训练车辆流量预测网络,同时确保模型具有良好的泛化能力,云平台采用了稀疏提取的方法来划分学习样本集合和验证样本集合。具体来说,在步骤S4A11中,云平台首先确定一个设定的百分比,用于划分学习样本和验证样本的比例。这个百分比通常是基于经验或实验确定的,旨在平衡模型的训练效率和泛化能力。例如,常见的划分比例是80%的数据作为学习样本,20%的数据作为验证样本。接下来,对于每个云平台上的车流量特征表示样本集合,云平台按照设定的百分比进行随机采样。采样过程中,云平台会同时选择目标车流量特征表示和对应的先验标记(即真实的车辆流量数据)。目标车流量特征表示是一个多维度的向量,包含了描述车流量的各种特征值;而先验标记则是与这些特征值对应的真实车辆流量值,用于监督学习过程中的误差计算。

[0158] 以具体实例来说,假设某个云平台拥有10000条车流量特征表示样本,每条样本都包含了车速、车流量、道路占用率等特征值以及对应的真实车辆流量值。如果设定的百分比为80%,那么云平台将从这10000条样本中随机选择8000条作为学习样本集合,剩余的2000条作为验证样本集合。

[0159] 通过这种方式,云平台不仅确保了学习样本集合和验证样本集合的独立性,还保证了样本的多样性和代表性。学习样本集合将用于训练车辆流量预测网络,而验证样本集合则用于在训练过程中评估模型的性能,以确保模型具有良好的泛化能力。

[0160] 步骤S4A12:如果所述各个云平台的车流量特征表示样本集合的类别为符合特征空间不同但样本空间相同的分布式数据类别,则基于所述设定的百分比通过对所述P个云平台中的第一云平台的车流量特征表示样本集合中的目标车流量特征表示和先验标记进行稀疏提取,并基于所述设定的百分比通过对所述P个云平台中所述第一云平台以外的云平台的车流量特征表示样本集合中的目标车流量特征表示进行稀疏提取,获得所述各个云

平台的学习样本集合和所述各个云平台的验证样本集合。

[0161] 在步骤S4A12中,云平台处理的是另一种特定的分布式数据类别,即特征空间不同但样本空间相同的分布式数据类别,这种情况通常出现在纵向联邦学习的场景中。在高速公路边缘协同数据处理的背景下,这意味着不同的云平台收集的车流量数据对应于相同的样本(例如,相同的路段、相同的时间段),但每个云平台只拥有这些样本的部分特征信息。为了在这种场景下有效地利用数据训练车辆流量预测网络,云平台采用了一种特殊的稀疏提取策略来划分学习样本集合和验证样本集合。具体来说,云平台首先选择一个云平台作为基准云平台(假设为第一云平台),该云平台拥有完整的特征空间,即包含所有与车辆流量预测相关的特征信息。接下来,云平台根据设定的百分比,对基准云平台上的车流量特征表示样本集合进行稀疏提取。这个百分比通常基于经验或实验确定,用于划分学习样本和验证样本的比例。在提取过程中,云平台会同时选择目标车流量特征表示和对应的先验标记(即真实的车辆流量数据)。这些被提取出来的样本将构成基准云平台的学习样本集合和验证样本集合。

[0162] 对于其他云平台(即非基准云平台),由于它们只拥有部分特征信息,因此稀疏提取的策略略有不同。云平台根据相同的百分比,仅从这些云平台的车流量特征表示样本集合中提取目标车流量特征表示,而不提取先验标记。这是因为这些云平台缺乏完整的特征空间,无法提供与先验标记对应的完整特征向量。这些被提取出来的样本将作为非基准云平台的学习样本集合的一部分。最后,云平台将基准云平台和其他云平台的学习样本集合和验证样本集合进行整合。整合后的学习样本集合将包含来自不同云平台的样本数据,这些样本数据在样本空间上是相同的(对应于相同的路段、相同的时间段),但在特征空间上可能有所差异。整合后的验证样本集合则仅包含来自基准云平台的样本数据,因为这些样本数据具有完整的特征空间和对应的先验标记。

[0163] 以具体实例来说,假设有三个云平台参与车辆流量预测网络的训练。其中,云平台A作为基准云平台,拥有完整的特征空间,包括车速、车流量、道路占用率等特征;云平台B只拥有车速和车流量的特征;云平台C只拥有道路占用率的特征。如果设定的百分比为80%,那么云平台将从云平台A中随机选择80%的样本作为学习样本集合和验证样本集合;从云平台B和云平台C中分别随机选择80%的样本作为各自的学习样本集合,但这些样本不包含先验标记。最后,这些样本将被整合在一起,用于训练车辆流量预测网络。

[0164] 通过这种方式,云平台能够在纵向联邦学习的场景下有效地利用来自不同云平台的数据,同时确保模型训练过程的准确性和有效性。这种划分方式不仅考虑了数据的特征空间差异,还充分利用了不同云平台之间的数据互补性,有助于提高车辆流量预测网络的性能。

[0165] 作为一种实施方式,所述步骤S4D,依据所述各个云平台的验证样本集合,对所述各个云平台中的网络内部参变量的误差进行评估,获得用以指示所述P个云平台中的网络内部参变量的误差的目标误差值,包括:

[0166] 步骤S4D1:通过将所述各个云平台中的网络内部参变量与所述各个云平台的验证样本集合中的目标车流量特征表示进行作积,获得所述各个云平台的阶段性预估值。

[0167] 在步骤S4D1中,云平台执行了一个关键的操作,即通过将各个云平台中的网络内部参变量(通常是神经网络的权重和偏置)与对应云平台的验证样本集合中的目标车流量

特征表示进行作积,来获得每个云平台的阶段性预估值。这一步骤是模型评估过程中的一个重要环节,用于验证网络模型在当前参数设置下的性能。具体来说,本申请实施例中,每个云平台都维护着一个车辆流量预测网络,并且持有一个验证样本集合。验证样本集合包含了真实的高速公路交通数据,用于模拟模型在实际应用中的表现。每个样本都包含了多个与车辆流量预测相关的特征(如车速、车流量、道路占用率等),这些特征被编码为目标车流量特征表示。在步骤S4D1中,云平台从验证样本集合中取出一个样本,提取其目标车流量特征表示。然后,它将这个特征表示与当前云平台上的车辆流量预测网络的内部参变量进行作积运算。这个作积运算实际上是模拟了网络的前向传播过程,即根据当前的参数设置和输入特征来计算输出预测值。

[0168] 假设车辆流量预测网络是一个简单的全连接神经网络,其权重矩阵为 W ,偏置向量为 b ,而某个验证样本的目标车流量特征表示为 x 。那么,阶段性预估值 y 可以通过以下公式计算得出: $y=W*x+b$ 。这里的“*”表示矩阵乘法。计算得出的 y 就是一个阶段性预估值,它代表了网络在当前参数设置下对输入特征 x 的预测输出。这个预测输出是一个数值,用于表示预测的车辆流量大小或等级。

[0169] 通过重复执行步骤S4D1,云平台可以对验证样本集合中的每个样本都计算出对应的阶段性预估值。这些预估值将用于后续的步骤中,与真实的车辆流量值进行比较,从而评估网络模型的性能。通过这种方式,云平台能够了解网络模型在未见过的数据上的表现,为后续的模型选择和调优提供重要依据。

[0170] 步骤S4D2:通过预设分类器对所述各个云平台的阶段性预估值进行分类,获得所述各个云平台得到的目标车辆流量预测结果。

[0171] 在步骤S4D2中,云平台利用预设的分类器对各个云平台在前一步骤S4D1中得到的阶段性预估值进行分类,从而得到目标车辆流量预测结果。这一步骤将连续的数值型预测值转换为更具实际意义的车流量等级标签,使得预测结果更易于理解和应用。本申请实施例中,每个云平台都基于其特定的车辆流量预测网络模型和验证样本集合,通过前向传播计算得到阶段性预估值。这些预估值通常是一系列连续的数值,代表了预测的车辆流量大小。然而,对于交通管理部门而言,他们更关心的是车辆流量的等级或级别,如“少流量”、“中流量”或“多流量”,因为这些等级标签能够更直观地反映交通状况,并为交通调度和规划提供指导。

[0172] 为了满足这一需求,云平台在步骤S4D2中引入了一个预设的分类器。这个分类器可以基于多种机器学习算法实现,如逻辑回归、决策树、支持向量机等。在高速公路车辆流量预测的上下文中,一个简单的阈值分类器可能就足够了。例如,可以设定一个阈值范围,将预估值小于某个阈值的样本分类为“少流量”,预估值在两个阈值之间的样本分类为“中流量”,预估值大于较大阈值的样本分类为“多流量”。具体来说,假设预设的分类器使用了三个阈值: T_1 、 T_2 和 T_3 ($T_1 < T_2 < T_3$)。对于某个云平台的阶段性预估值 y ,如果 $y \leq T_1$,则将其分类为“少流量”;如果 $T_1 < y \leq T_2$,则分类为“中流量”;如果 $y > T_2$,则分类为“多流量”。通过这种方式,云平台可以将每个云平台的阶段性预估值转换为目标车辆流量预测结果,即车流量等级标签。

[0173] 例如,假设某个云平台的阶段性预估值为200,而预设的阈值分别为 $T_1=100$ 、 $T_2=300$ 。由于200位于 T_1 和 T_2 之间,因此该预估值将被分类为“中流量”。这个过程将对验证样本

集合中的每个预估值重复进行,从而得到每个云平台在验证样本上的目标车辆流量预测结果。这些结果将用于后续的误差计算和模型评估步骤中,为优化车辆流量预测网络提供指导。

[0174] 步骤S4D3:依据所述各个云平台得到的目标车辆流量预测结果和所述各个云平台的验证样本集合中的先验标记,确定所述目标误差值。

[0175] 在步骤S4D3中,云平台通过比较各个云平台在验证样本集合上得到的目标车辆流量预测结果与先验标记(即真实的车辆流量等级标签),来确定目标误差值。这一步骤是评估车辆流量预测网络性能的关键环节,它量化了网络预测结果与真实值之间的差异,为后续的模型调优提供了重要依据。本申请实施例中,验证样本集合包含了真实的高速公路交通数据,每条数据都有一个对应的先验标记,即真实的车辆流量等级标签(如“少流量”、“中流量”、“多流量”等)。在步骤S4D2中,云平台已经通过预设的分类器对各个云平台的阶段性预估值进行了分类,得到了目标车辆流量预测结果。

[0176] 在步骤S4D3中,云平台首先将目标车辆流量预测结果与对应的先验标记进行逐一比较。对于每个验证样本,如果预测结果与先验标记一致,则说明网络在该样本上的预测是准确的;如果不一致,则说明预测出现了误差。为了量化这种误差,云平台采用了一种误差度量方法,如准确率、召回率、F1分数等。以准确率为例,它计算了所有验证样本中预测正确的比例。具体来说,假设验证样本集合中共有N个样本,其中有M个样本的预测结果与先验标记一致,则准确率Accuracy可以通过以下公式计算得出: $Accuracy=M/N$ 。

[0177] 这个准确率就是目标误差值的一种表现形式,它直观地反映了网络模型在验证样本集合上的整体性能。准确率越高,说明网络模型的预测能力越强,误差越小。举例来说,假设某个云平台的验证样本集合中共有100个样本,其中有85个样本的预测结果与先验标记一致。那么,该云平台的目标误差值(以准确率表示)就是85%。这意味着网络模型在85%的样本上做出了准确的预测,而在剩下的15%的样本上出现了误差。根据这个误差值,云平台可以判断网络模型是否达到了预设的性能标准,进而决定是否需要进行进一步的模型调优或迭代训练。

[0178] 作为一种实施方式,所述步骤S4D3,依据所述各个云平台得到的目标车辆流量预测结果和所述各个云平台的验证样本集合中的先验标记,确定所述目标误差值,包括:

[0179] 步骤S4D31:通过所述各个云平台的验证样本集合中的先验标记减去所述各个云平台得到的目标车辆流量预测结果,获得所述各个云平台中的网络内部参变量的误差目标车流量特征表示。

[0180] 在步骤S4D31中,云平台通过比较各个云平台在验证样本集合上的先验标记(真实的车辆流量等级标签)和目标车辆流量预测结果,来计算网络内部参变量的误差目标车流量特征表示。这一步骤是评估车辆流量预测网络性能的重要一环,它帮助系统量化预测结果与实际标签之间的差异。

[0181] 本申请实施例中,每个云平台都持有一个验证样本集合,该集合包含了真实的交通数据及其对应的先验标记。先验标记通常是根据实际情况人工标注的,代表了真实的车辆流量等级(如“少流量”、“中流量”、“多流量”等)。而目标车辆流量预测结果则是车辆流量预测网络根据输入特征计算得出的预测值,同样表示为车辆流量等级。

[0182] 在步骤S4D31中,云平台首先会从验证样本集合中取出一条样本,提取其先验标记

和目标车辆流量预测结果。由于这两者都是分类标签,无法直接进行数值相减,因此需要采用一种间接的方式来量化它们之间的差异。

[0183] 一种常见的方法是为每个流量等级分配一个数值标签。例如,可以将“少流量”标签赋值为1,“中流量”标签赋值为2,“多流量”标签赋值为3。然后,云平台会比较先验标记和目标预测结果的数值标签,并计算它们之间的差值。这个差值就代表了网络在该样本上的预测误差。

[0184] 例如,假设某个验证样本的先验标记为“中流量”(数值标签为2),而车辆流量预测网络给出的目标预测结果为“少流量”(数值标签为1)。那么,云平台就会计算这两个数值标签之间的差值,即 $2-1=1$ 。这个差值1就是该样本上网络内部参变量的误差目标车流量特征表示。

[0185] 通过重复执行上述过程,云平台可以对验证样本集合中的每个样本都计算出对应的误差目标车流量特征表示。这些误差特征表示将用于后续的步骤中,进一步量化网络预测误差的大小,并评估网络的整体性能。通过这种方式,云平台能够准确地了解网络在验证样本上的表现,为后续的模型调优和迭代训练提供有力支持。

[0186] 步骤S4D32:对所述各个云平台中的网络内部参变量的误差目标车流量特征表示进行翻转,将翻转后的特征表示与所述各个云平台中的网络内部参变量的误差目标车流量特征表示进行作积,获得所述各个云平台中的网络内部参变量的误差值。在步骤S4D32中,云平台对各个云平台中的网络内部参变量的误差目标车流量特征表示进行处理,以进一步量化网络在预测车流量时的误差大小。这一步骤是评估车辆流量预测网络性能过程中的一个关键步骤,它通过对误差特征表示进行操作,得到一个能够反映网络预测误差大小的数值。

[0187] 首先,需要明确的是,在步骤S4D31中,云平台已经计算出了各个云平台在验证样本集合上的误差目标车流量特征表示。这些特征表示实际上是一组数值,每个数值对应一个验证样本的预测误差。本申请实施例中,这些误差值反映了车辆流量预测网络在特定样本上的预测准确性。

[0188] 在步骤S4D32中,云平台首先会对这些误差目标车流量特征表示进行翻转(即转置)操作。然而,在这个特定的场景下,由于误差值是一维的(每个样本对应一个误差值),转置操作实际上并不改变数据的结构。这里的转置更多的是为了与后续的操作保持一致,确保数据格式的正确性。接下来,云平台将翻转后的误差特征表示(实际上与原始误差特征表示相同)与其自身进行作积。这个作积操作实际上是计算每个误差值的平方。为什么要计算平方呢?这是因为平方操作能够放大误差的影响,使得较大的误差在总误差中占据更大的比重。这对于后续的误差分析和模型调优是非常有用的。例如,假设有三个验证样本的误差目标车流量特征表示分别为1、2和3(这些数值代表了预测误差的大小)。在步骤S4D32中,计算机会对这些误差值进行平方操作,得到 $1^2=1$ 、 $2^2=4$ 和 $3^2=9$ 。这些平方后的误差值就是各个云平台中的网络内部参变量的误差值。

[0189] 通过这一步操作,云平台得到了一个包含所有样本误差平方的数组或向量。这个数组或向量直观地反映了网络在验证样本集合上的预测误差分布情况。较大的平方误差值表示网络在该样本上的预测误差较大,需要重点关注和调优。最后,这些误差值将用于后续步骤的计算,如步骤S4D33中的均值计算,以进一步量化网络的整体预测误差大小,并评估

车辆流量预测网络的性能。

[0190] 步骤S4D33:对所述各个云平台中的网络内部参变量的误差值进行均值计算,得到均值误差,将所述均值误差作为所述目标误差值。

[0191] 在步骤S4D33中,云平台对各个云平台中的网络内部参变量的误差值进行均值计算,以得到一个能够综合反映网络预测误差大小的指标,即均值误差。这个均值误差将被作为目标误差值,用于评估车辆流量预测网络在验证样本集合上的整体性能。

[0192] 本申请实施例中,每个云平台都持有一个验证样本集合,并且已经通过之前的步骤计算出了每个样本的误差值(这些误差值反映了网络在该样本上的预测误差大小)。这些误差值形成了一个误差值数组或向量,包含了网络在验证样本集合上的所有预测误差信息。在步骤S4D33中,云平台对这个误差值数组或向量进行均值计算。均值计算是一种统计方法,用于计算一组数值的平均水平。在这里,它帮助云平台将多个离散的误差值整合为一个单一的数值指标,即均值误差。

[0193] 具体来说,云平台将误差值数组或向量中的所有元素相加,然后除以元素的数量(即验证样本的数量),从而得到均值误差。这个均值误差代表了网络在验证样本集合上的平均预测误差大小,是一个综合评估网络性能的指标。

[0194] 例如,假设验证样本集合中包含100个样本,每个样本都有一个对应的误差值。这些误差值形成了一个包含100个元素的误差值数组。云平台将这100个误差值相加,得到总和,然后再除以100(样本数量),从而得到均值误差。假设总和为1000,那么均值误差就是10($1000/100=10$)。

[0195] 这个均值误差就是目标误差值,它直观地反映了车辆流量预测网络在验证样本集合上的整体预测误差大小。通过比较不同网络或不同迭代轮次下的目标误差值,云平台可以判断网络是否达到了预设的性能标准,进而决定是否需要进行进一步的模型调优或迭代训练。本申请实施例中,这有助于交通管理部门了解网络模型的预测准确性,并根据需要采取相应的措施来优化交通管理策略。

[0196] 在本申请实施例的另一方面,提供一种低延迟边缘协同数据处理方法,在该实施例中,从边缘盒子的角度进行介绍,即低延迟边缘协同数据处理方法应用于边缘盒子,所述边缘盒子与目标云平台通信连接,所述方法包括:

[0197] 步骤S100:获取目标高速监测数据;

[0198] 步骤S200:对所述目标高速监测数据进行数据特征提取操作,获得所述目标高速监测数据的目标车流量特征表示;

[0199] 步骤S300:依据所述目标高速监测数据的目标车流量特征表示,通过所述目标云平台中的目标车辆流量预测网络,对所述目标高速监测数据进行分类,获得所述目标高速监测数据的目标车辆流量预测结果,其中,所述目标车辆流量预测网络基于前述实施例的方法调试得到。此处不再对具体的原理进行赘述。

[0200] 本申请实施例提供了一种云平台,如图3所示,云平台106包括:处理器101和存储器103。其中,处理器101和存储器103相连,如通过总线102相连。可选地,云平台106还可以包括收发器104。需要说明的是,实际应用中收发器104不限于一个,该云平台106的结构并不构成对本申请实施例的限定。

[0201] 本申请实施例提供了一种云平台,本申请实施例中的云平台包括:一个或多个处

理器;存储器;一个或多个计算机程序,其中一个或多个计算机程序被存储在存储器中并被配置为由一个或多个处理器执行,一个或多个程序被处理器执行时,实现上述第一方面的方法。

[0202] 以上所述仅是本申请的部分实施方式,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,还可以做出若干改进和润饰,这些改进和润饰也应视为本申请的保护范围。

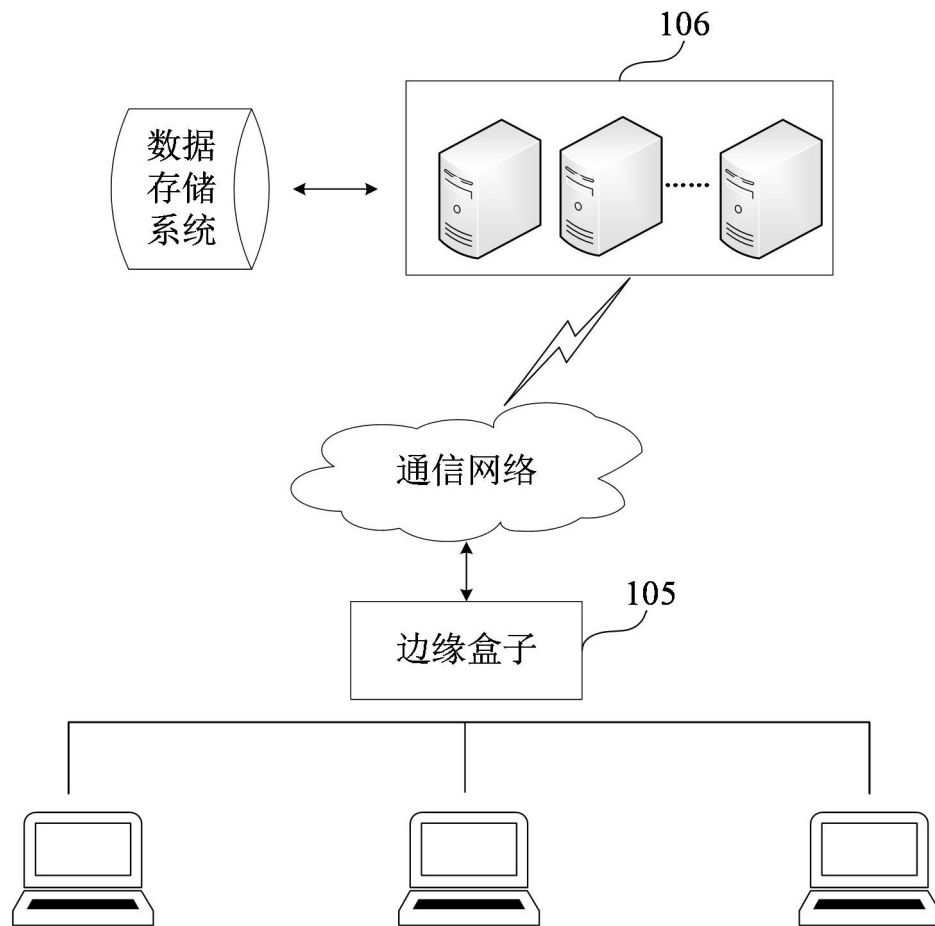


图 1

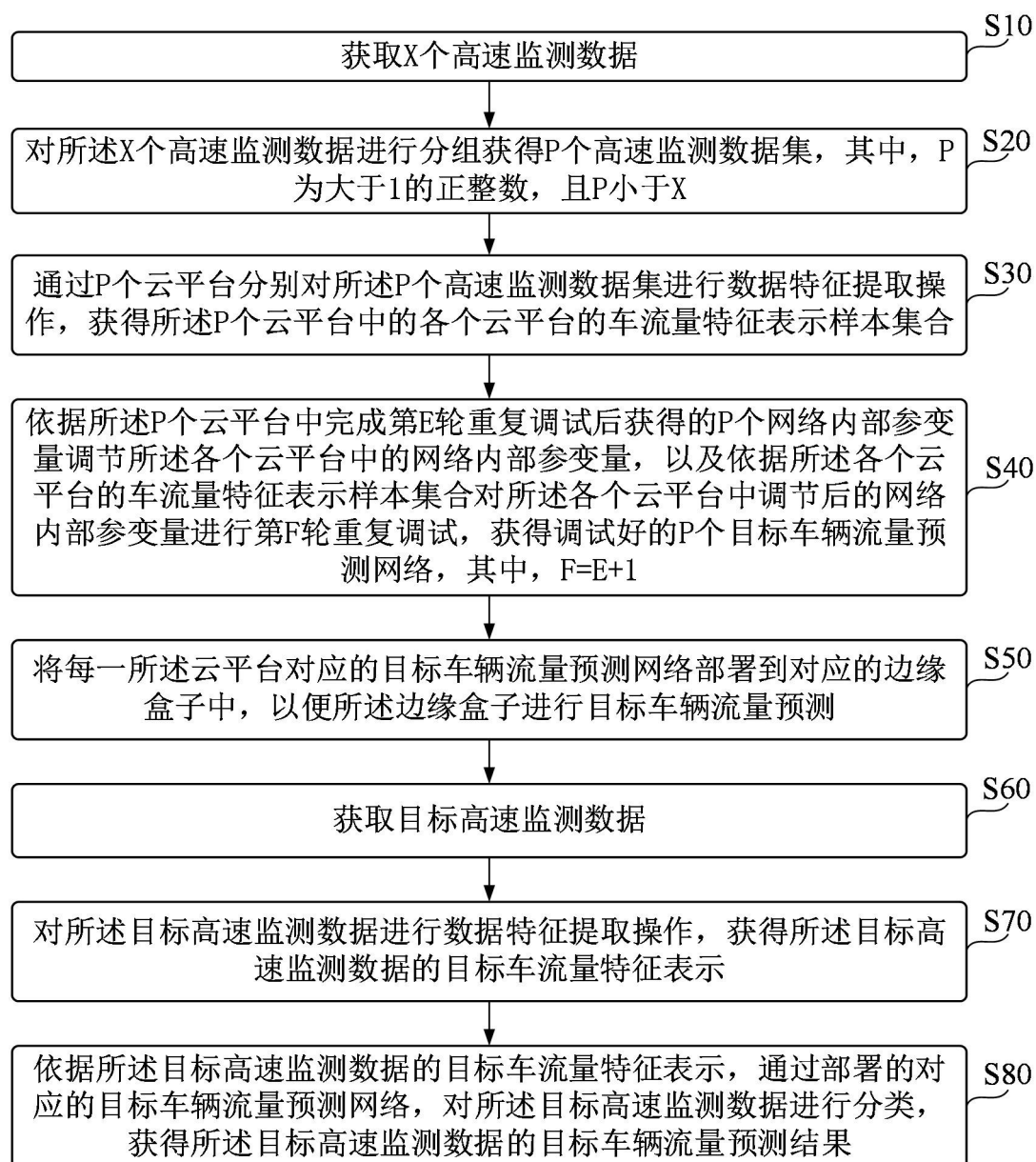


图 2

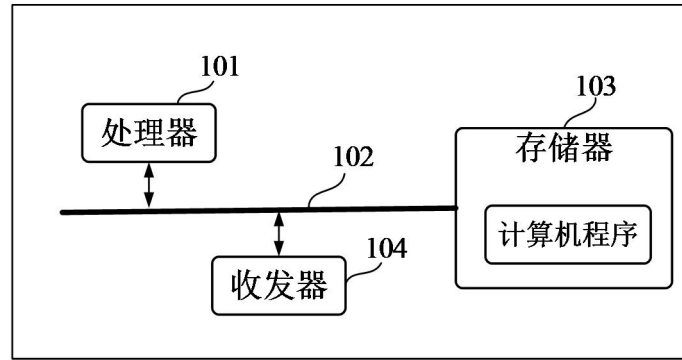
106

图 3