计算机研究与发展
Journal of Computer Research and Development
DOI:10.7544/issn1000-1239.20200668
59(2):478-487, 2022

基于联邦学习的多源异构数据融合算法

莫慧凌　郑海峰　高　敏　冯心欣
(福州大学物理与信息工程学院　福州　350108)
(1298948502@qq.com)

Multi-Source Heterogeneous Data Fusion Based on Federated Learning

Mo Huiling, Zheng Haifeng, Gao Min, and Feng Xinxin
(*College of Physics and Information Engineering, Fuzhou University, Fuzhou* 350108)

Abstract　With the rapid development of technology, the number of network edge devices with the capability of computation and memory is increasing, and the volume of the generated data is growing exponentially, which makes it difficult for a centralized processing model with cloud computing as the core to efficiently process data generated by edge devices. Not only will the network delay increase, but the data is likely to be leaked on the upload link, and data security cannot be guaranteed. In addition, due to the diversity of edge devices and the continuous enrichment of data representation methods, multi-modal data exists widely. The processing of multi-source heterogeneous data collected by different edge devices has become an urgent problem in big data research. In order to make full use of heterogeneous data on edge devices and solve the problem of "data communication barriers" caused by data privacy in edge computing, in this paper we propose a novel fusion algorithm for multi-source heterogeneous data based on Tucker decomposition in federated learning. For the fusion problem of heterogeneous data without interaction in federated learning, the proposed algorithm introduces Tucker decomposition theory to capture the multi-dimensional characteristics of heterogeneous data by constructing a high-order tensor. Finally, the effectiveness of this algorithm is verified on the MOSI dataset.

Key words　edge computing; federated learning; deep learning; tensor theory; heterogeneous data fusion

摘　要　随着科技的迅猛发展,具有计算和存储能力的边缘设备数量不断增加,产生的数据流量更是呈指数式增长,这使得以云计算为核心的集中式处理模式难以高效处理边缘设备产生的数据.另外,由于边缘网络设备的多样性以及数据表示手段的不断丰富,多模态数据广泛存在.为充分利用边缘设备上的异构数据,解决边缘计算中由于数据隐私引起的"数据通信壁垒"问题,提出了一种联邦学习中基于Tucker分解的多源异构数据融合算法.该算法针对异构数据在无交互条件下的融合问题,引入张量Tucker分解理论,通过构建一个具有异构空间维度特性的高阶张量以捕捉异构数据的高维特征,从而实现联邦学习中多源异构数据的融合.最后,在MOSI数据集上验证了算法的有效性.

关键词　边缘计算;联邦学习;深度学习;张量理论;异构数据融合

中图法分类号　TP391

收稿日期:2020-08-31;**修回日期**:2021-02-25

基金项目:国家自然科学基金项目(61971139)
This work was supported by the National Natural Science Foundation of China (61971139).

通信作者:郑海峰(zhenghf@fzu.edu.cn)

在信息化时代,海量的边缘数据将给以云计算模型为核心的数据集中化处理模式带来许多问题,一是将数据全部上传至云端的处理方式,不仅效率低下,而且造成额外的带宽开销,同时网络延迟也会增加.二是由于用户隐私意识的提高,边缘设备的数据很有可能在上传链路时泄密,个人隐私的安全问题无法得到保障.而分布式数据处理模式可以有效地解决传统云计算存在的时延和效率问题[1].同时,针对"数据孤岛"问题,谷歌公司首次提出了"联邦学习"这个概念[2].通过在多个边缘设备上利用各自的训练样本对模型进行单独的训练,并通过模型参数聚合实现在不透露用户隐私的前提下多源信息的共享.

此外,边缘设备的多样性使得设备采集到的数据在标注、语义和存在形式等方面都呈现多样性.多模态数据广泛存在.不同的模态数据可以从多个方面描述目标对象,通过消除冗余数据和融合各种数据源进行关联补充分析,数据可以涌现出更多有价值的新信息,从而实现 $1+1>2$ 的效果[3].从互联网和移动设备收集的多媒体数据是典型的非结构化数据[4],与传统的易于存储的结构化数据格式存在显著差异.因此,对不同边缘设备采集到的多源异构数据处理成为大数据研究中亟需解决的问题.

在传统的多源异构数据融合算法中,数据集中化处理在实际应用中存在数据隐私泄露的风险.因此,对于不透露用户隐私前提下的多源异构数据处理还存在许多难题:首先,由于企业竞争和用户的隐私保护意识,使得数据互通长期处于闭塞状态,无法实现信息共享,从而无法充分发挥异构数据的价值.其次,利用神经网络对数据进行处理,根据数据设计的模型一旦确定后就无法更改.然而在边缘计算中,边缘设备所采集的数据结构和种类数目存在差异.若针对各个网络边缘设备上的数据设计适用于各自数据特征的神经网络,工作量极大,同时该模型只能适用于单一节点或者和该节点数据特征相同的边缘设备,普适性不高,也无法充分发挥物联网中其他的异构数据的价值.

为解决边缘计算中,在不泄露用户隐私的前提下实现多源异构数据的融合问题,本文提出了一种基于联邦学习的多源异构数据融合算法.从边缘设备采集到的数据结构特点入手,结合张量 Tucker 分解理论,研究能够在各异的边缘设备上自适应处理多源异构数据模型,解决联邦学习中由于处理异构数据的模型不统一带来的单一适应性问题.

1 相关工作

目前,针对联邦学习以及异构数据融合的研究已经取得了众多的成果.

1.1 联邦学习

由于移动设备和边缘设备的广泛使用,Yang 等人[2]提出的一种人工智能技术——联邦学习,是由一个中央服务器协调多个客户端在不公开数据的前提下,协同完成一个学习任务.

联邦学习有很多优点.首先,相比于云计算模型,联邦学习只发送更新的模型参数进行聚合,这极大降低了数据通信的成本,提高了网络带宽的利用率.其次,用户的原始数据不需要发送至云端,这避免了数据在上传链路时泄露用户隐私的可能.再者,联邦学习的模型训练可以在边缘节点或终端设备上进行训练和实时决策,时间延迟会比在云端进行决策时得到极大地改善.

在联邦学习中,数据安全是一个主要研究方面.Mcmahan 等人[5]提出了用户级的差分隐私训练算法,通过将隐私保护添加到聚合算法中,有效地降低了从传输模型中恢复个人信息的可能.另一方面,Beimel 等人[6]提出了差分隐私混合模型,通过用户的信任偏好对用户进行分区,从而减少所需用户基数的大小.Dong 等人[7]将梯度选择和秘密分享的算法结合起来,在保证用户隐私和数据安全的情况下大幅提升了通信效率.

在联邦学习中的资源优化方面,Tran 等人[8]考虑通过无线网络进行联邦学习,提出了优化能源消耗和全球联邦学习时间的问题.Wang 等人[9]提出了一种控制算法,可以在全局参数聚合和局部模型更新之间进行权衡,以在资源预算约束下将损失函数降至最低.Wang 等人[10]提出了一种联邦学习框架 In-Edge-AI,以实现边缘计算中的智能资源管理.

1.2 多源异构数据融合

数据融合系统中的数据逐渐多元化且数量巨大,这迫使人们对系统效率的提高有了更高的要求.Microsoft 研究院的 Zheng[11]将异构数据融合方法分为 3 种类型:1)基于阶段的数据融合方法;2)基于特征的数据融合方法;3)基于语义的数据融合方法.

基于阶段的数据融合方法是指在数据挖掘的过程中,不同阶段利用不同的数据进行分析.Pan 等人[12]首先使用 GPS 轨迹数据和道路网络数据检测交通异常,通过检索与交通异常位置相关的社交媒体

信息(例如 Twitter),最后分析交通异常的特定事件内容.这种融合方法的异构数据之间没有交互作用,失去了异构数据之间互补的优势,很难实现真正的内在数据融合.

基于特征的融合方法通过提取每个异构数据的特征,然后对特征进行分析和处理.因此,提取的特征质量以及融合方法都将对融合效果具有决定性的影响.Liu 等人[13]整合不同视图的面部信息,将不同维度的深度学习特征向量融合以实现基于深度异构性特征的面部识别.Ouyang 等人[14]将人类异构信息特征的 3 个来源进行非线性融合,可以更准确地估计身体姿势.Wang 等人[15]设计了张量深度学习计算模型,利用张量对多源异构数据的复杂性进行建模,将向量空间数据扩展到张量空间,并在张量空间中进行特征提取.Zadeh 等人[16]提出了张量融合网络用于解决多模态的情感分析,通过笛卡儿积的方式将多种模态进行融合,实现对情感的分类分析.

基于语义的融合方法了解每个数据集以及跨数据集的特征之间的关系,认为提取到的异构数据的特征是可解释的.Zheng 等人[17]提出了一种基于协同训练的模型来预测整个城市空气质量,利用空气质量具有时间以及空间的依赖性的特点,分别针对时空数据设计了 2 个分类器,通过将不同的时空特征输入到不同的分类器,从而在不同标签上生成 2 组概率,最大化地选择标签.

上述工作主要考虑了单一节点上的多源异构数据融合问题.而针对联邦学习中的多源异构数据融合问题,目前尚未有相关工作报道.

2 基于联邦学习的多源异构数据融合

本文主要针对网络边缘设备由于数据隐私性,无法进行数据通信情况下实现多源异构数据的融合进行研究.通过引入张量分解理论,构建一个具有异构数据空间维度特性的高阶记忆单元,在不透露用户隐私的前提下,利用记忆单元对多源异构数据进行有效地融合.同时,能够在不额外增加模型规模的条件下实现对多源异构数据的自适应学习.

2.1 联邦学习系统模型

本文针对异构数据在不进行数据互通前提下的融合问题,考虑在边缘计算中引入联邦学习,在不暴露用户自身隐私的前提下实现对多用户潜在特征的学习,其系统基本架构如图 1 所示.在该框架中,系统由边缘节点、物联网和云端服务器组成,其中边缘节点通过物联网(如网关和路由器)与云端服务器互联.

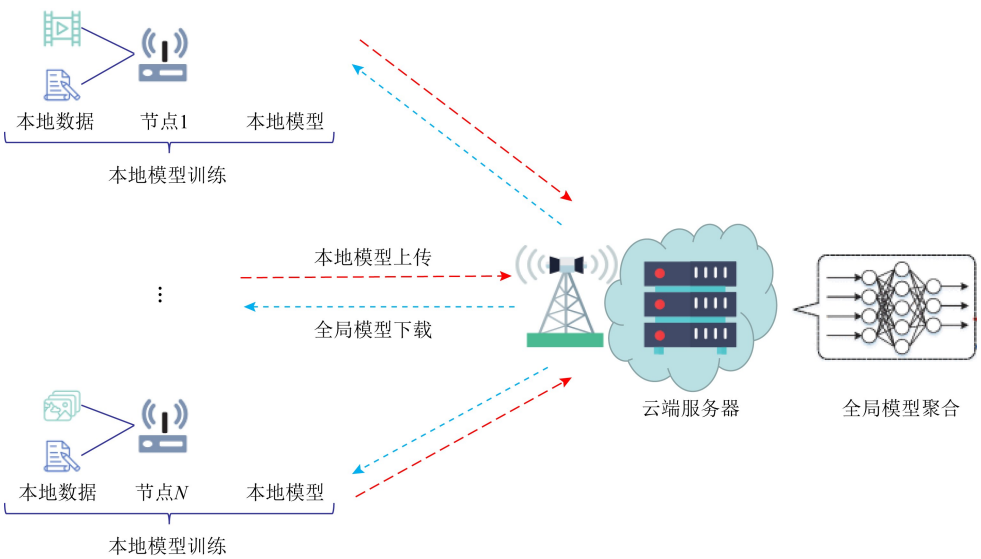

Fig. 1 The system model for federated learning

图 1 联邦学习系统模型

联邦学习是一种分布式学习框架,其中原始数据被收集并存储在多个边缘节点上,并在节点处执行模型训练,然后将模型通过节点与云端服务器的交互逐步优化学习模型.

基于以上框架,联邦学习可以从多个独立的边缘节点上使用本地数据协同训练一个泛化的共享模型,通过模型传输替代数据传输,规避了用户隐私泄露的风险.

2.2 算法的总体设计

如图 2 所示,本文所提出的基于联邦学习的多源异构数据融合算法主要分成特征提取模块、特征融合模块和特征决策模块.其中特征提取模块由各种异构数据对应的特征提取子网络构成.

在初始化阶段,中心控制节点对模型中的特征提取模块、特征融合模块和特征决策模块进行网络参数随机初始化,并下发至边缘节点.

在模型训练阶段,边缘节点接收到中心控制节点下发的模型后,根据本地节点上的数据集结构选择对应的特征提取模块,并利用本地数据集对特征提取模块、特征融合模块和特征决策模块进行训练.边缘节点新一轮训练的终止条件是本地节点训练轮数超过给定的训练轮数.待训练完成后将各自的训练模型返回至中心控制节点进行模型聚合.

在模型聚合阶段,对于特征融合模块和特征决策模块采用平均聚合算法,对于特征提取模块,则是根据得到的对应特征提取子模块进行平均聚合,以确保同一模态的数据提取的特征具有相似性.最后,将更新后的模型重新下发至边缘节点进行新一轮的训练.

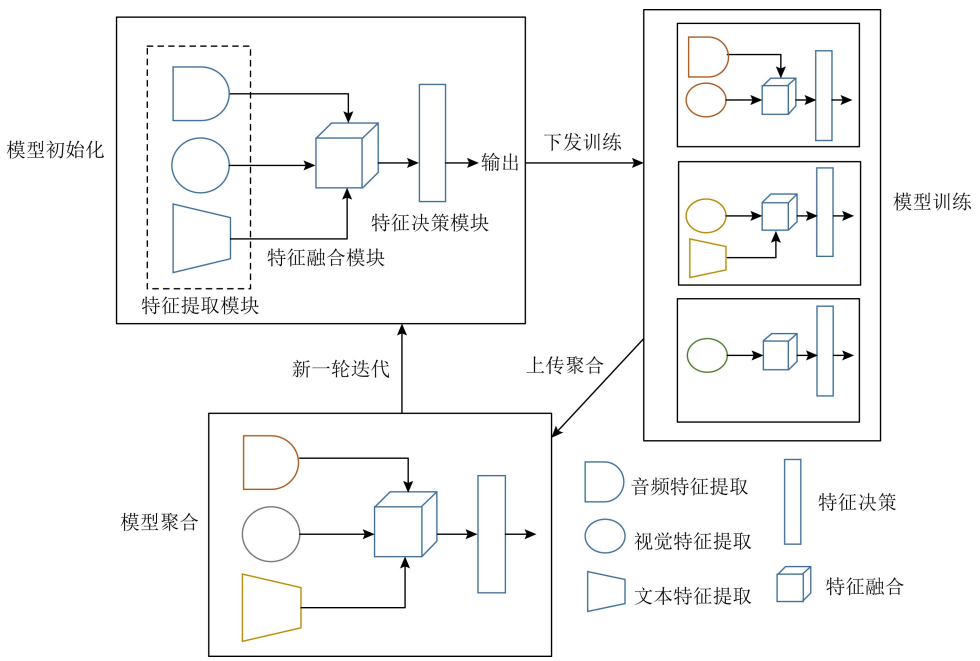

Fig. 2 The overview of the proposed algorithm

图 2 算法总体框架

2.3 子模块设计

2.3.1 特征提取模块

本文假设待处理的异构数据分别为音频、视觉和文本数据.在特征提取模块,本节根据不同模态的特征,采用了不同的特征提取子网络分别对音频、视觉及文本信息进行特征提取.

1) 音频、视觉特征子网络.针对音频信息和视觉信息,分别采用了 COVAREP[18] 声学分析框架和 FACET[19] 面部表情分析框架对 MOSI 数据集进行特征采样提取(采样频率分别为 100 Hz 和 30 Hz).

2) 文本特征子网络.口语文本在语法及表达上不同于书面文本,例如"我觉得挺好的……,不过,我觉得这个方法还有待改善"这种口语在书面语言中很少出现.处理口语这种具有多变性语言的关键在于建立能够在不可靠的情况下运行的模型,以及通过关注重要的词语来表现特殊的言语特征.如图 3 所示,本文提出的文本特征提取网络在编码部分先采用全局词向量对口语词进行预处理,同时使用知短时记忆(long short-term memory, LSTM)网络

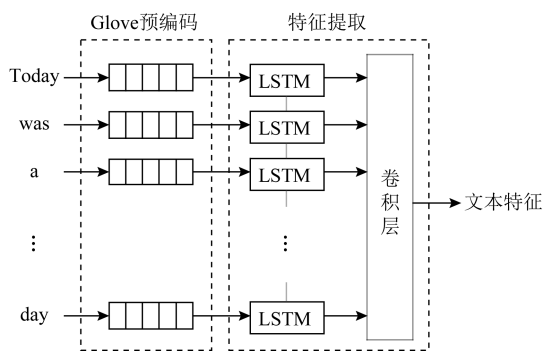

Fig. 3 Text feature sub-network

图 3 文本特征子网络

学习与时间相关的语言表示,并将其作为 CNN 网络的输入.在卷积层中,通过卷积核对文本信息实现细粒度更小的局部特征提取.

2.3.2 特征融合模块

在本节中,假设待处理的异构数据特征分别为音频数据特征 $\boldsymbol{z}_a=(\boldsymbol{z}_a^1,\boldsymbol{z}_a^2,\cdots,\boldsymbol{z}_a^q)$,视觉数据特征 $\boldsymbol{z}_v=(\boldsymbol{z}_v^1,\boldsymbol{z}_v^2,\cdots,\boldsymbol{z}_v^p)$,文本数据特征 $\boldsymbol{z}_t=(\boldsymbol{z}_t^1,\boldsymbol{z}_t^2,\cdots,\boldsymbol{z}_t^m)$,经过特征融合模块后的特征输出为 $\boldsymbol{Z}$.接下来,将以上述假设作为基本条件,阐述提出的基于 Tucker 分解的异构数据融合算法的基本原理.如图 4 所示,该模块通过引入一个具有异构数据特征空间的高阶张量 $\boldsymbol{W}$,该张量的每一模态对应于一种异构数据特征的空间映射.因此,在对每一种异构数据特征进行融合的时候,高阶张量 $\boldsymbol{W}$ 不仅能够引入其余异构数据模态的特征进行修正,同时也会将正在进行的异构数据模态特征进行记忆.

以图 4 为例,当待处理的异构数据特征分别为 $\boldsymbol{z}_a,\boldsymbol{z}_v,\boldsymbol{z}_t$ 时,记忆单元 $\boldsymbol{W}$ 为一个三阶张量,且此张量的 3 个维度分别对应于 3 种异构数据特征 $\boldsymbol{z}_a,\boldsymbol{z}_v,\boldsymbol{z}_t$ 的特征空间.在本节提出的异构数据特征融合中,通过将异构数据特征与记忆单元对应的特征空间进行模乘,可得到具有该异构数据特征的记忆单元,并以此进行进一步的特征融合操作.融合操作主要分成 3 个阶段:首先,记忆单元 $\boldsymbol{W}$ 沿着一阶与异构数据特征 $\boldsymbol{z}_a$ 进行模乘,得到具有 $\boldsymbol{z}_a$ 特征的新记忆单元 $\boldsymbol{W}^{(1)}$.其次,记忆单元 $\boldsymbol{W}^{(1)}$ 沿着二阶与异构数据特征 $\boldsymbol{z}_v$ 进行模乘,得到具有 $\boldsymbol{z}_a$ 和 $\boldsymbol{z}_v$ 特征的记忆单元 $\boldsymbol{W}^{(2)}$.最后,记忆单元 $\boldsymbol{W}^{(2)}$ 沿着三阶与异构数据特征 $\boldsymbol{z}_t$ 进行模乘,最终得到具有三者特征的融合张量 $\boldsymbol{Z}$.其具体过程可以表示为

$$\boldsymbol{Z}=((\boldsymbol{W}\times_1\boldsymbol{z}_a)\times_2\boldsymbol{z}_v)\times_3\boldsymbol{z}_t, \tag{1}$$

其中 $\boldsymbol{W}\in\mathbb{R}^{R_1\times R_2\times R_3}$,$\boldsymbol{z}_a\in\mathbb{R}^{P\times R_1}$,$\boldsymbol{z}_v\in\mathbb{R}^{J\times R_2}$,$\boldsymbol{z}_t\in\mathbb{R}^{K\times R_3}$.

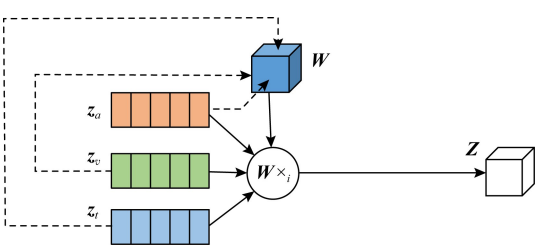

Fig. 4 Heterogeneous data fusion based on Tucker decomposition

图 4 基于 Tucker 分解的异构数据融合

2.3.3 特征决策模块

针对融合后的数据,本节采用了传统的全连接层在全局特征的基础上进行决策,包括回归模型的预测和分类模型的概率预测.在该模块中,采用了 $L1$ 范数损失函数 $L1Loss$ 对目标值和预测值之间的误差进行了衡量.其具体表达式为

$$L1Loss(\boldsymbol{Y},\boldsymbol{Y}^p)=\boldsymbol{L}=(l_1,l_2,\cdots,l_n)^{\mathrm{T}}=\begin{cases}sum(\boldsymbol{L}), reduction=\text{'sum'},\\ mean(\boldsymbol{L}), reduction=\text{'mean'},\end{cases} \tag{2}$$

其中 l_n 的表达式为

$$l_n=|y_n-y_n^p|, \tag{3}$$

y_n 和 y_n^p 分别为数据集的真实值 $\boldsymbol{Y}$ 和模型预测值 $\boldsymbol{Y}^p$ 中的样本.

2.4 基于联邦学习的多源异构数据融合

假设有 N 个边缘节点 $\{E_1,E_2,\cdots,E_N\}$ 参与共享模型的训练,且所有边缘节点共采集到 M 种异构数据.

在初始化阶段,云端根据采集到的 M 种异构数据,设计对应的特征提取模块 $\boldsymbol{F}$,特征融合模块 $\boldsymbol{I}$,特征决策模块 $\boldsymbol{C}$.则共享模型 $\boldsymbol{G}$ 可表示为

$$\boldsymbol{G}=\langle\boldsymbol{F},\boldsymbol{I},\boldsymbol{C}\rangle, \tag{4}$$

其中 $\langle\cdot\rangle$ 表示模型拼接操作.具体来说,在特征提取模块 $\boldsymbol{F}$ 中,分别针对 M 种异构数据设计对应的特征提取子网络 $\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_M$,可表示为

$$\boldsymbol{F}=\langle\boldsymbol{F}_1,\boldsymbol{F}_2,\cdots,\boldsymbol{F}_M\rangle, \tag{5}$$

其中,$\boldsymbol{F}_i$ 表示第 i 种异构数据的特征提取子网络.

在特征融合模块 $\boldsymbol{I}$ 中,构建一个具有异构数据空间维度特性的高阶张量 $\boldsymbol{W}\in\mathbb{R}^{R_1\times R_2\times\cdots\times R_M}$.经训练后,该张量沿第 i 模展开后的参数能够反映第 i 种异构数据的空间维度特性.在特征决策模块 $\boldsymbol{C}$ 中,通过对融合后的异构数据特征进行训练,对异构数据之间的潜在联系进行更深层次的挖掘,提高模型在多源异构数据上的特征表达力.

通过式(1)可知,当 $R_1=P$,$R_2=J$,$R_3=K$ 时,融合后的张量大小与记忆单元 $\boldsymbol{W}$ 的大小一致,均为 $\boldsymbol{Z}\in\mathbb{R}^{R_1\times R_2\times R_3}$.因此,当因子矩阵满足正方阵约束时,核心张量与原始张量在空间维度上存在恒等关系.利用该特性,在初始化阶段,进一步对全局进行设置,定义特征提取子网络 $\boldsymbol{F}_i$ 的特征图 $\boldsymbol{f}_i\in\mathbb{R}^{R_i\times R_i}$,$i\in\{1,2,\cdots,M\}$,从而解决了边缘计算中由于异构数据的不确定性引起的异构数据融合难问题.

在模型训练阶段,参与训练的 N 个边缘节点根据所拥有的异构数据类型 $\boldsymbol{z}=\{\boldsymbol{z}_1,\boldsymbol{z}_2,\cdots,\boldsymbol{z}_k\}$,$(k\leqslant M)$自适应地选择对应的特征提取子网络 $\boldsymbol{F}_i$

进行训练.在特征融合模块,由于在全局初始化定义阶段,将特征提取子网络 $\boldsymbol{F}_i$ 的特征图设定为 $\boldsymbol{f}_i \in \mathbb{R}^{R_i \times R_i}$,从而将异构数据融合后的张量大小约束为定值 $\boldsymbol{Z} \in \mathbb{R}^{R_1 \times R_2 \times \cdots \times R_M}$,进一步解决了多源异构数据融合自适应性问题.

如图 5 所示,本节对特征提取模块的自适应选择机制进行了更为详尽的描述.假设所有待处理的异构数据类型分别为 $\boldsymbol{z}_a$,$\boldsymbol{z}_v$,$\boldsymbol{z}_t$,且对应的记忆单元 $\boldsymbol{W}$ 的 3 个维度分别对应于 3 种异构数据 $\boldsymbol{z}_a$,$\boldsymbol{z}_v$,$\boldsymbol{z}_t$ 的特征空间.节点 1 和节点 N 采集到的异构数据类型不同.在模型训练阶段,节点 1 根据拥有的异构数据类型 $\boldsymbol{z}_a$,$\boldsymbol{z}_v$,选择对应的特征提取子网络 $\boldsymbol{F}_a$ 和 $\boldsymbol{F}_v$ 进行训练,分别得到特征图 $\boldsymbol{f}_a \in \mathbb{R}^{R_a \times R_a}$,$\boldsymbol{f}_v \in \mathbb{R}^{R_v \times R_v}$.根据节点上拥有的异构数据类型数量,将特征融合阶段分成 2 个部分:首先,记忆单元 $\boldsymbol{W}$ 沿着一阶与 $\boldsymbol{f}_a$ 特征进行模乘,得到具有 $\boldsymbol{f}_a$ 特征的新的记忆单元 $\boldsymbol{W}^{(1)}$.其次,记忆单元 $\boldsymbol{W}^{(1)}$ 沿着二阶与 $\boldsymbol{f}_v$ 特征进行模乘,得到具有以上 2 种异构数据特征的融合张量 $\boldsymbol{Z}$.该过程可表示为

$$\begin{aligned} \boldsymbol{W}^{(1)} &= I(\boldsymbol{f}_a;\boldsymbol{W}), \\ \boldsymbol{Z} &= I(\boldsymbol{f}_v;\boldsymbol{W}^{(1)}|\boldsymbol{W}), \end{aligned} \tag{6}$$

其中 $\boldsymbol{W}^{(1)}|\boldsymbol{W}$ 表示在具有 $\boldsymbol{f}_a$ 特征的基础上对 $\boldsymbol{f}_v$ 的特征进行融合.在该过程中,模型首先利用记忆单元对 $\boldsymbol{f}_a$ 特征进行记忆,得到具有 $\boldsymbol{f}_a$ 特征的模型,并将此作为 $\boldsymbol{f}_v$ 特征进行融合时的先验条件,从而在模型训练过程中,记忆单元不但能对各个异构数据的空间维度特征进行学习,还能对不同异构数据之间的潜在联系进行捕捉.

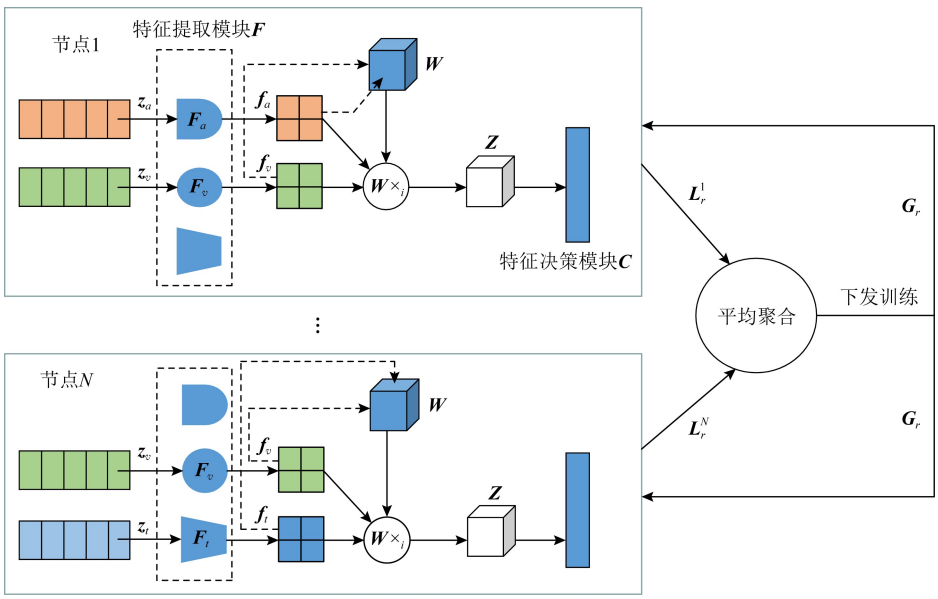

Fig. 5 Multi-source heterogeneous data fusion based on federated learning

图 5 基于联邦学习的多源异构数据融合

节点 N 上的训练机制和节点 1 类似.上述过程可表示为

$$L_r^k = L_{r-1}^k - \eta g_r^k, \tag{7}$$

其中 L_r^k 是在第 r 轮全局迭代中节点 k 上,利用本地采集到的异构数据,通过学习率为 η 的梯度下降算法得到的本地模型,g_r^k 则是对应的梯度.

在模型聚合阶段,由于各个边缘节点采用特征提取器自适应选择机制对特征提取模块进行训练,因此在模型聚合时,需要先将各个边缘节点选择训练的特征提取子网络进行归并,再采用平均聚合算法得到具有全局异构数据特征的共享模型,该过程表示为

$$\boldsymbol{F}^* = \left\langle \sum_{k=1}^{N} \frac{n_1^k}{n_1} \times \boldsymbol{F}_1^k, \sum_{k=1}^{N} \frac{n_2^k}{n_2} \times \boldsymbol{F}_2^k, \cdots, \sum_{k=1}^{N} \frac{n_M^k}{n_M} \times \boldsymbol{F}_M^k \right\rangle, \tag{8}$$

其中 $\boldsymbol{F}^*$ 为更新后的特征提取模块,n_i^k 和 n_i($1 \leqslant i \leqslant M$)分别是在边缘节点 k 上采集到的第 i 种异构数据样本数和第 i 种异构数据的所有样本数.因此第 r 轮更新后的全局共享模型 $\boldsymbol{G}_r = \langle \boldsymbol{F}^*, \boldsymbol{I}^*, \boldsymbol{C}^* \rangle$ 可以表示为

$$\begin{aligned} \boldsymbol{G}_r &= \langle \boldsymbol{F}^*, \boldsymbol{I}^*, \boldsymbol{C}^* \rangle = \\ &\left\langle \boldsymbol{F}^*, \sum_{k=1}^{N} \frac{m^k}{m} \times \boldsymbol{I}^k, \sum_{k=1}^{N} \frac{m^k}{m} \times \boldsymbol{C}^k \right\rangle, \end{aligned} \tag{9}$$

其中 $\boldsymbol{G}_r$ 是对 N 个边缘节点上的本地模型通过联合

平均算法进行聚合后,得到的具有全局特征的共享模型. $m^k=\sum_{i=1}^{M}n_i^k$ 和 $m=\sum_{i=1}^{M}n_i$ 分别为边缘节点上的所有异构样本数和所有异构样本总数.

3 实验结果与分析

为验证本文算法的有效性,主要从单节点异构数据融合和多节点异构数据融合 2 个方面对本文算法的性能进行评估.基于 Tucker 分解的单节点异构数据融合实验,主要是通过单节点实现多源异构数据上的回归任务和分类任务对本文算法的性能进行评估,并与目前存在的几种主流异构数据融合算法进行了对比;基于 Tucker 分解的多节点异构数据融合实验,通过多节点实现多源异构数据上的回归任务和分类任务对本文算法的性能进行评估.

3.1 数据集设置和评价指标

本实验采用 MOSI 数据集,该数据集是 YouTube 上来自于视频影评的多模态情感数据集,包含了来自 89 位评论者的 93 个视频,每个视频的长度为 2～5 min,包含了口语(字幕)、图像和语音3 种信息.对于该数据集中的各个样本,均以人工方式对情绪进行评分,其分值位于[−3,3]之间.

本实验从回归任务和分类任务 2 个方面验证所提算法在多个任务上的有效性.在分类任务中,对分值量化如表 1 所示:

Table 1 Quantification of Emotional Score

表 1 情绪分值量化表

情绪类别	情绪值
非常消极	−3
消极	−2
略微消极	−1
正常	0
略微积极	1
积极	2
非常积极	3

对于回归任务,本节实验采用平均绝对误差(*MAE*)和皮尔逊相关系数(*Corr*)对本文算法的性能进行分析,其对应表达式为

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|y_i-y_i^p|, \tag{10}$$

$$Corr(Y,Y^p)=\frac{cov(Y,Y^p)}{\sigma_Y\sigma_{Y^p}}=\frac{E[(Y-\mu_Y)(Y^p-\mu_{Y^p})]}{\sigma_Y\sigma_{Y^p}}, \tag{11}$$

其中 y_i 和 y_i^p 分别为数据集的真实值 Y 和模型预测值 Y^p 中的样本.

对于分类任务,本节实验采用精度(*Acc*)和 $F1$ 指数对本文算法性能进行评估,其中表示类别的数目,其对应的表达式为

$$Acc=\frac{TP+TN}{TP+TN+FP+FN}, \tag{12}$$

$$F1=\frac{TP}{TP+\frac{FP+FN}{2}}=\frac{2TP}{2TP+FP+FN}, \tag{13}$$

其中 TP,FP,TN,FN 分别为真阳性、假阳性、真阴性和假阴性的样本数.*Acc* 越高,模型的分类的精度越高,性能越好;$F1$ 越高,模型对各个类别的识别能力越均衡,性能越好.

3.2 单节点异构数据融合实验分析

本节实验对拥有 3 种模态的情感数据集 MOSI 分别进行了回归任务和分类任务的训练.设置 3 种模态的特征提取子模块的特征输出数为 R_k,其中 k 表示为第 k 种模态.考虑到特征提取数 R_k 对于训练效率以及实验性能起到直接的作用.因此,本实验首先对比了不同的特征提取数的组合(R_1,R_2,R_3)分别对回归任务和分类任务的性能影响.在本实验中,设置各个模态的特征提取数集合为{8,16,32}.为验证本文算法的性能,从回归任务和分类任务 2 个方面对比了 6 种算法在单节点情况下的多源异构数据融合上的表现能力.

3.2.1 回归任务

回归任务中关于各个模态特征提取数的定量实验结果如表 2 所示.观察可知,本文算法在回归任务上的 *MAE* 主要集中在 0.95～1.05 之间.

Table 2 *MAE* of MOSI Dataset Regression Task

表 2 MOSI 数据集回归任务 *MAE*

R_1	R_2	R_3		
		8	16	32
8	8	0.982 911	0.965 055	0.975 610
	16	0.973 587	1.007 763	1.012 596
	32	1.036 670	0.990 383	0.963 566
16	8	1.047 929	0.990 000	0.978 279
	16	1.048 996	1.001 936	1.007 727
	32	1.015 852	1.037 620	1.008 007
32	8	1.009 914	1.056 609	1.024 716
	16	0.993 762	1.003 780	1.015 237
	32	1.055 843	0.995 406	1.003 787

观察分析可知,对于回归任务来说,当3个模态的特征提取数的组合为(8,8,16)时,性能更为显著.

实验设置回归任务上各个模态的特征提取数分别为8,8,16.表3记录了本文算法与6种算法在回归任务上的性能对比.由表可知,本文算法与最优算法TFN[16]在回归任务上性能相当,优于大多对比算法.虽然本文提出的异构数据融合算法在单节点上的回归任务性能低于LMF[24]算法,但其旨在应用于联邦学习中对多源异构数据进行有效地融合,而现有的异构数据融合算法是在单节点上实现的,并不一定适合于联邦学习中.

Table 3 Performance Comparison of Regression Tasks on the MOSI Dataset

表3 MOSI数据集上回归任务性能对比

算法	*MAE*	*Corr*
SVM[20]	1.864	0.057
DF[21]	1.143	0.518
BC-LSTM[22]	1.079	0.581
MV-LSTM[23]	1.019	0.601
TFN[16]	0.970	0.633
LMF[24]	0.912	0.668
本文	0.965	0.624

3.2.2 分类任务

对于分类任务来说,模型对各个模态特征提取数的敏感度会低于回归任务.因此,本实验中设置分类任务上各个模态的特征提取数分别为8,8,16.表4记录了本文算法与6种算法在分类任务上的性能对比.其中Acc_2和$F1$是二分类任务评价指标,Acc_7为多分类评价指标.由表4可知,本文算法在分类任务上性能优于TFN及大多数对比算法,而与LMF算法性能相当.

Table 4 Performance Comparison of Classification Tasks on the MOSI Dataset

表4 MOSI数据集上分类任务性能对比 %

算法	Acc_2	$F1$	Acc_7
SVM[21]	50.2	50.1	17.5
DF[22]	72.3	72.1	26.8
BC-LSTM[23]	73.9	73.9	28.7
MV-LSTM[24]	77.1	74.0	33.2
TFN[16]	73.9	73.4	32.1
LMF[24]	76.4	75.7	32.8
本文	75.5	75.2	33.8

讨论:虽然本文算法在单节点上性能表现并不是最好,但在联邦学习框架下有3个优势:1)对异构数据具有更强的自适应性.与其他算法相比,本算法在训练时不需要同时输入所有类型的异构数据,因此更适合在联邦学习中的不同类型的边缘节点中应用.2)更好地保护了数据隐私.本算法避免了将多种异构数据同时发送至同一处进行训练而可能存在的隐私泄露风险.3)大大降低了传输带宽.本算法只需传输提取该边缘节点拥有的异构数据对应的特征提取子网络模型参数,而无需传输提取所有异构数据特征的模型参数.

3.3 多节点异构数据融合实验分析

为验证本文算法在联邦学习框架下对多源异构数据融合的性能,本节实验将从2个方面对本文算法的性能进行评估:首先通过对训练子节点上的异构数据的部署来验证本文算法对异构数据的自适应能力.如表5所示,对于训练子节点上数据集的具体部署策略主要分成3类:1)单模态数据,即各个训练子节点上的数据集仅为单一结构的数据;2)双模态数据,即各个训练子节点上的数据集为任意2种结构的数据;3)三模态数据,即各个训练子节点上拥有所有结构的数据.其次,通过同一种多节点异构数据部署策略下训练的模型,在不同模态的异构数据上的表现,验证本文算法的普适性和泛化性.在本实验中,采用的特征提取数的组合$(R_1,R_2,R_3)=(8,8,16)$.

Table 5 Multi-Node Heterogeneous Data Deployment Strategy

表5 多节点异构数据部署策略

数据类型	单模态数据			双模态数据			三模态数据		
	节点	节点	节点	节点	节点	节点	节点	节点	节点
口语	√			√	√		√	√	√
音频		√			√	√	√	√	√
视觉			√	√		√	√	√	√

注:"√"表示节点上存在该数据类型.

3.3.1 回归任务

对于回归任务,评价指标为*MAE*和*Corr*.实验结果如表6所示,对于回归任务来说,根据何种模态数据训练出的模型在该种模态数据上的性能表现最为显著.模型的预测值和样本的相关性随着训练模态数的增加而显著提升,且多模态训练模型对模态间潜在联系的学习具有向下兼容性.

在回归任务上,当训练数据的模态数越高,模型在与训练集模态数不同的测试集上的性能表现越

好.这是因为多模态模型在训练的过程中,除了学习各个模态本身具有的特征以外,对各个模态之间的潜在联系也进行了学习,因此,对于高模态数的训练模型来说,学习到的各个模态之间潜在联系的组合种类也就越多,回归任务的性能就越好.

Table 6 Performance of Regression Tasks on the MOSI Dataset

表 6 MOSI 数据集上回归任务性能

模态融合模型	指标	单模态	双模态	三模态
单模态数据融合	*MAE*	0.986	1.098	1.088
	Corr	0.599	0.446	0.363
双模态数据融合	*MAE*	1.160	0.838	1.098
	Corr	0.495	0.674	0.409
三模态数据融合	*MAE*	1.134 0	1.029 2	0.793 8
	Corr	0.490	0.636	0.762

3.3.2 分类任务

对于分类任务,指标为 Acc_2,$F1$,Acc_7,其中 Acc_2 和 $F1$ 为二分类任务评价指标,Acc_7 为多分类任务评价指标.实验结果如表 7 所示,根据何种模态数据训练出的模型在该种模态数据上的性能表现最为显著.单模态训练模型对于多模态数据的融合是通过云端聚合方式实现的,模型学习到的只是各个模态上各自的特征,而学习不到各个模态之间潜在的联系,得到的只是具有本地训练模态特征的局部最优解;而对于多模态训练,不论是双模态还是三模态,模型训练的过程中除了各个模态本身特征的学习,同时还对模态之间的潜在联系进行了学习,云端聚合也通过信息共享扩大了对数据样本的学习.

Table 7 Classification Task Performance on the MOSI Dataset

表 7 MOSI 数据集上分类任务性能 %

模态融合模型	指标	单模态	双模态	三模态
单模态数据融合	*Acc_2*	78.26	71.73	69.56
	F1	77.94	70.10	64.52
	Acc_7	40.57	34.05	32.60
双模态数据融合	*Acc_2*	71.01	81.15	71.73
	F1	70.64	81.35	72.11
	Acc_7	32.60	42.75	41.30
三模态数据融合	*Acc_2*	74.63	80.43	86.96
	F1	73.83	79.69	87.01
	Acc_7	31.88	33.33	47.83

4 总 结

本文提出了一种基于联邦学习的多源异构数据融合算法,该算法利用了 Tucker 分解理论,通过构建一个具有异构数据空间维度的高阶张量,实现对多模态数据的融合和记忆.相较于其他算法,该算法能够在不进行数据互通的前提下,对多源异构数据进行有效地融合,从而打破了由于隐私安全问题带来的数据通信壁垒.同时,该算法能够同时根据训练节点所拥有的异构数据结构,在不增加多余模型训练规模的前提下自适应地对不同种类的异构数据进行处理,从而在分布式训练中能够更高效地实现对通信带宽的利用率,减少不必要的传输,降低对网络边缘设备计算力和存储力的要求,同时,模型也具有更高的普适性和泛化性.

作者贡献申明:莫慧凌负责方案的实施、实验结果整理与分析以及论文撰写与修订;郑海峰指导方案设计,把握论文创新性,并指导论文撰写与修订;高敏负责方案设计与实施,以及论文撰写;冯心欣参与方案可行性讨论.

参 考 文 献

[1] Zhou Jun, Shen Huajie, Lin Zhongyun, et al. Research advances on privacy preserving in edge computing [J]. Journal of Computer Research and Development, 2020, 57(10): 2027-2051 (in Chinese)
(周俊, 沈华杰, 林中允, 等. 边缘计算隐私保护研究进展[J]. 计算机研究与发展, 2020, 57(10): 2027-2051)

[2] Yang Qiang, Liu Yang, Chen Tianjian, et al. Federated machine learning: Concept and applications [J]. ACM Transactions on Intelligent Systems and Technology, 2019, 10(2): 1-19

[3] Beyer J, Heesche K, Hauptmann W, et al. Ensemble learning for multisource information fusion [C] //Proc of European Conf on Symbolic & Quantitative Approaches to Reasoning & Uncertainty. Berlin: Springer, 2009: 748-756

[4] Samuel A, Sarfraz M I, Haseeb H, et al. A framework for composition and enforcement of privacy-aware and context-driven authorization mechanism for multimedia big data [J]. IEEE Transactions on Multimedia, 2015, 17(9): 1484-1494

[5] Mcmahan H B, Ramage D, Talwar K, et al. Learning differentially private recurrent language models [J]. arXiv preprint, arXiv:1710.06963, 2018

[6] Beimel A, Korolova A, Nissim K, et al. The power of synergy in differential privacy: Combining a small curator with local randomizers [J]. arXiv preprint, arXiv: 1912.08951, 2019

[7] Dong Ye, Hou Wei, Chen Xiaojun, et al. Efficient and secure federated learning based on secret sharing and gradients selection [J]. Journal of Computer Research and Development, 2020, 57(10): 2241-2250 (in Chinese)
(董业, 侯炜, 陈小军, 等. 基于秘密分享和梯度选择的高效安全联邦学习[J]. 计算机研究与发展, 2020, 57(10): 2241-2250)

[8] Tran N H, Bao W, Zomaya A, et al. Federated learning over wireless networks: Optimization model design and analysis [C] //Proc of IEEE INFOCOM 2019-IEEE Conf on Computer Communications. Piscataway, NJ: IEEE, 2019: 1387-1395

[9] Wang Shiqiang, Tuor T, Salonidis T, et al. Adaptive federated learning in resource constrained edge computing systems [J]. IEEE Journal on Selected Areas in Communications, 2019, 37(6): 1205-1221

[10] Wang Xiaofei, Han Yiwen, Wang Chenyang, et al. In-edge AI: Intelligentizing mobile edge computing caching and communication by federated learning [J]. IEEE Network, 2019, 33(5): 156-165

[11] Zheng Yu. Methodologies for cross-domain data fusion: An averview [J]. IEEE Transactions on Big Data, 2015, 1(1): 16-34

[12] Pan Bei, Zheng Yu, Wilkie D, et al. Crowd sensing of traffic anomalies based on human mobility and social media [C] // Proc of the 21st ACM SIGSPATIAL Int Conf on Advances in Geographic Information Systems. New York: ACM, 2013: 344-353

[13] Liu Zouzhu, Zhang Wenyu, Quek T Q S, et al. Deep fusion of heterogeneous sensor data [C] //Proc of 2017 IEEE Int Conf on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ: IEEE, 2017: 5965-5969

[14] Ouyang Wanli, Chu Xiao, Wang Xiaogang. Multi-source deep learning for human pose estimation [C] //Proc of 2014 IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 2337-2344

[15] Wang Wei, Zhang Min. Tensor deep learning model for heterogeneous data fusion in Internet of things [J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2020, 4(1): 32-41

[16] Zadeh A, Chen Minghai, Poria S, et al. Tensor fusion network for multimodal sentiment analysis [J]. arXiv preprint, arXiv:1707.07250, 2017

[17] Zheng Yu, Liu Furui, Hsieh H P. U-Air: When urban air quality inference meets big data [C] //Proc of the 19th ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining. New York: ACM, 2013: 1436-1444

[18] Degottex G, Kane J, Drugman T, et al. Covarep: A collaborative voice analysis repository for speech technologies [C] //Proc of 2014 IEEE Int Conf on Acoustics, Speech and Signal Processing. Piscataway, NJ: IEEE, 2014: 960-964

[19] Paul E. An argument for basic emotions [J]. Cognition and Emotion, 1992, 6(3/4): 169-200

[20] Cortes C, Vapnik V N. Support vector networks [J]. Machine Learning, 1995, 20(3): 273-297

[21] Nojavanasghari B, Gopinath D, Koushik J, et al. Deep multimodal fusion for persuasiveness prediction [C] //Proc of the 18th ACM Int Conf on Multimodal Interaction. New York: ACM, 2016: 284-288

[22] Fukui A, Park D H, Yang D, et al. Multimodal compact bilinear pooling for visual question answering and visual grounding [J]. arXiv preprint, arXiv:1606.01847, 2016

[23] Rajagopalan S S, Morency L, Baltrusaitis T, et al. Extending long short-term memory for multi-view structured learnin[G] //LNCS 9911: Proc of European Conference on Computer Vision. Berlin: Springer, 2016: 338-353

[24] Liu Zhun, Shen Ying, Lakshminarasimhan V B, et al. Efficient low-rank multimodal fusion with modality-specific factors [J]. arXiv preprint, arXiv:1806.00064, 2018

Mo Huiling, born in 1996. Master. Her main research interests include federated learning, heterogeneous data fusion.
莫慧凌,1996 年生.硕士.主要研究方向为联邦学习、异构数据融合.

Zheng Haifeng, born in 1978. PhD, professor, PhD supervisor. Member of CCF. His main research interests include tensor theory, machine learning, Internet of things.
郑海峰,1978 年生.博士,教授,博士生导师,CCF 会员.主要研究方向为张量理论、机器学习、物联网.

Gao Min, born in 1995. Master. Her main research interests include tensor theory, deep learning and Internet of things.
高　敏,1995 年生.硕士.主要研究方向为张量理论、深度学习和物联网.

Feng Xinxin, born in 1983. PhD, associate professor. Her main research interests include tensor theory, machine learning, Internet of things.
冯心欣,1983 年生.博士,副教授.主要研究方向为张量理论、机器学习、物联网.