基于多接入边缘计算的任务卸载和资源分配策略研究

Research on Task Offloading and Resource Allocation Strategy Based on Multi-access Edge Computing

许　波[1]，陈礼波[1]，张　鹏[2]，曹童杰[1]（1. 中讯邮电咨询设计院有限公司，北京 100048；2. 中国联合网络通信集团有限公司，北京 100033）

Xu Bo[1], Chen Libo[1], Zhang Peng[2], Cao Tongjie[1]（1. China Information Technology Designing & Consulting Institute Co., Ltd., Beijing 100048, China; 2. China United Network Communications Group Co., Ltd., Beijing 100033, China）

摘　要：

随着移动通信技术和工业互联网的飞速发展，移动设备端日渐庞大的数量和复杂的应用对大量计算密集和低时延提出了要求，也因此引出了基于多接入边缘计算的任务卸载概念。这种任务卸载方式能够有效地利用边缘云服务器资源，将复杂的计算任务卸载至邻近的低消耗边缘服务器，提高任务计算效率和更高的服务质量。提出了一种基于拓扑结构的任务卸载策略和边缘资源分配策略，旨在解决边缘计算场景中，任务卸载效率低、资源利用率不足等问题。

关键词：

多接入边缘计算；任务卸载；资源分配

doi：10.12045/j.issn.1007-3043.2022.05.014

文章编号：1007-3043（2022）05-0066-05

中图分类号：TN919

文献标识码：A

开放科学（资源服务）标识码（OSID）：

Abstract：

With the rapid development of mobile communication technology and industrial Internet, the increasingly large number and complex applications of mobile devices put forward requirements for a large number of computationally intensive and low latency, which leeds to the concept of task offloading based on multi-access edge computing. This task offloading method can effectively use edge cloud server resources, offload complex computing tasks to neighboring low-consumption edge servers, which improves task computing efficiency and higher service quality. It proposes a topology-based task offloading strategy and edge resource allocation strategy, which aims to solve the problems of low task offloading efficiency and insufficient resource utilization in edge computing scenarios.

Keywords：

Multi-access edge computing; Task offloading; Resource allocation

引用格式：许波，陈礼波，张鹏，等. 基于多接入边缘计算的任务卸载和资源分配策略研究[J]. 邮电设计技术，2022(5)：66-70.

0 引言

近年来，商用5G已初具规模[1]。与前几代移动通信技术相比，5G的高带宽、低时延、广连接特性使其在在线游戏、智能识别、增强现实和虚拟现实（AR&VR）[2]、自动驾驶[3]等领域得到广泛应用。然而，由于终端的计算能力和电池容量等方面的局限性，终端本身已无法满足日益增长的各种计算需求，云服务器应运而生。传统的云服务器采用集中式的计算方法，由中心云完成计算任务并将结果返回至设备端，这就导致设备端接收计算结果的时间取决于中心云与设备端之间的距离和核心网络的流量，在多数情况下很难满足低时延甚至“无感知“的需求。为了解决这一难点，科学家提出了一种分布式的计算策略——多接入边缘计算（Multi-access Edge Computing，MEC）[4]。MEC将中心云分解为多个“微云”，即云服务节点，并将其分散在设备附近，以完成设备卸载的计算任务。由于“微云”和设备的距离很近，流量较少，能够满足各类计算请求对时延的需求；其缺点是计算能力较弱，一般只接收附近终端的计算任务。任务卸载和资源分配是移动边缘计算的关键问题[5]。合理的任务卸载和资源分配策略对提升MEC的服务性能和用户服

收稿日期：2022-03-16

务质量（Quality of Service，QoS）[6]具有重要的意义。

1 相关研究

近几年，有关MEC的研究比较热。MEC将单一的云功能分解并下沉至多个边缘服务器[7]，利用物理位置的变化减少任务执行处理的时延[8]。部分研究将MEC简化为在小型基站上部署一部分额外的资源[9-10]，但没有考虑到基站位置的不可移动性，使得边缘服务器的部署有了很大限制；现有MEC多为云-边协同架构[11]，既有云端高速处理的优势，又有边缘侧低时延的优势[12]；对于MEC中的任务卸载问题，参考文献[13]认为边缘服务器具有统一的计算能力，并提出了一种伪在线任务调度算法。Jia等人在参考文献[14]中利用排队理论设计了用户和服务器模型，并提出了一种启发式策略来解决调度问题。本文在这些理论研究的基础上，结合任务之间的拓扑结构，提出了一种细粒度的任务卸载和资源分配方式。

2 多接入边缘计算

MEC系统的核心就是分布式的边缘服务器，即在源头提供具有网络、计算、存储等功能的服务器节点[15]，计算任务被分别卸载到边缘服务器上，经过资源分配和计算后再将结果返回至设备端。这样，能同时满足多设备的计算需求，有效减少资源浪费，提高计算效率。无论是在空间距离上还是网络拓扑上，边缘节点都更靠近设备侧，与传统的云中心计算模式相比，MEC时延更低、安全性更高，对时延敏感型和带宽密集型任务更加友好[16]。MEC的系统架构如图1所示。

边缘服务器分布在网络的边缘侧，一般在基站或接入点附近，更靠近设备侧，图1示出了2种任务卸载的路径：卸载至边缘服务器或直接卸载至云端。其连接性、约束性和分布性是5G网络协议提供可靠服务的保证。MEC可以显著缩短通信距离，被广泛应用于在线游戏、智能制造、智慧交通、智慧城市、物联网、车联网等领域，这些领域的需求主要体现在时延、带宽与安全3个方面。然而，使用MEC系统架构还存在一定的挑战性，其关键在于任务卸载的决策和计算资源的分配。

3 任务卸载

MEC的实际应用中往往催生大量的算力需求。

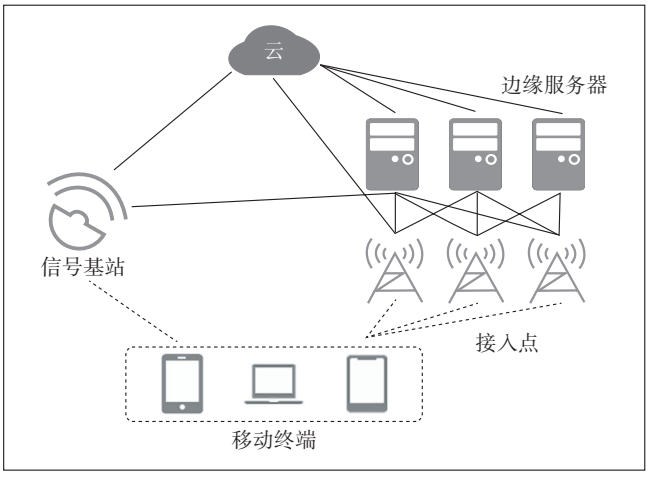

图1　MEC系统架构

当移动终端自身的算力无法满足时，计算任务会被就近分配至边缘服务器，待计算完成后将结果返回至终端，这一过程就是任务卸载[17]。任务卸载可以帮助终端分担复杂的计算任务，降低能耗的同时实现更快的任务处理速度，提供更好的QoS。

任务卸载一般有2个方式：水平卸载和垂直卸载[18]。水平卸载一般指在相同层的设备和服务器之间进行通信和协同计算；垂直卸载是根据计算能力将系统划分为云中心、边缘服务器、移动终端3层，卸载方向为低算力层到高算力层，如图2所示。

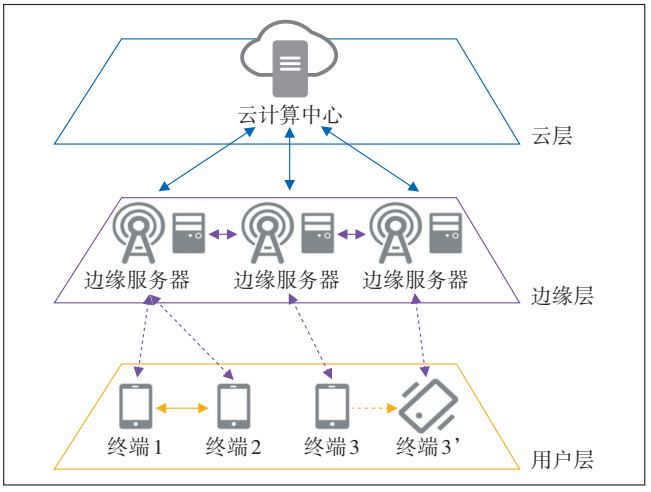

图2　垂直任务卸载示意图

现阶段对任务卸载的研究集中在判断是否要进行卸载，把任务当成一个整体，而忽略了部分卸载的可能性。判断任务卸载的结果至少应该包括不卸载（进行本地计算）、全部卸载和部分卸载3种情况，如图3所示。详细卸载过程如图4所示。

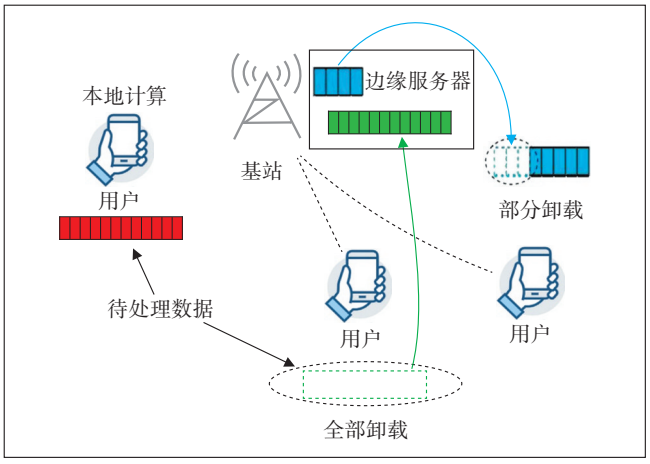

图3　任务卸载的3种方式

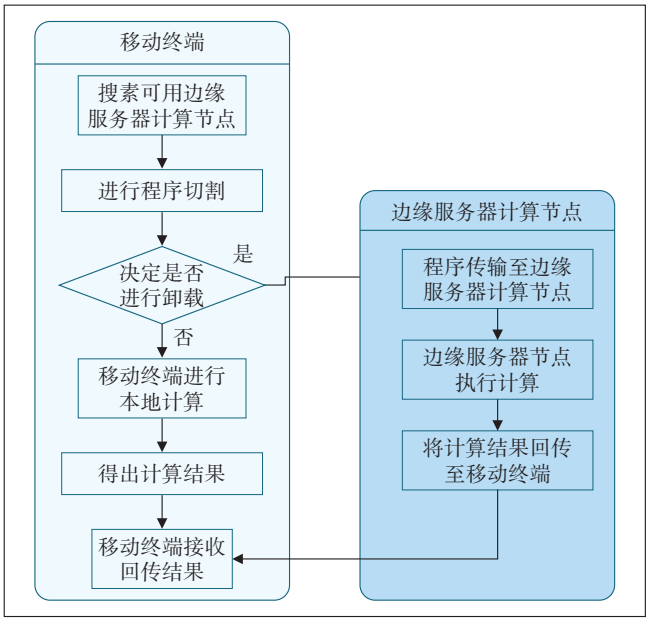

图4　计算任务卸载过程

4 资源分配

合理的任务卸载策略可以帮助MEC提高网络效能和QoS，但考虑到带宽容量和算力资源的有限性，如果同一时间涌现大量计算任务，带宽会限制数据的发送和接收，算力会限制任务的处理时长，在这种情况下，MEC就失去了其低时延、无感知的优势。因此，除了合理的任务卸载策略，MEC还需要配合资源分配策略，以保证该计算体系正常且有效的应用。资源分配面临的2个关键问题，一是如何在有限资源的情况下更加细粒度的部署网络功能，二是采用哪种部署策略能够更好地应对网络波动对服务质量的影响。针对上述2个问题，本文采取网络功能虚拟化和深度强化学习方法来制定资源分配策略，有效适应多种边缘场景，提高资源利用率和QoS。图5为该资源分配策略示意图，在此策略中，每个任务请求既有其专用通道，又有多请求复用通道，较好地权衡了时延和资源使用量之间的关系。如果进一步降低时延，需要增加至少3个虚拟机，以求每个任务请求都有专用通道；如果缩减虚拟机数量，会进一步增长时延。

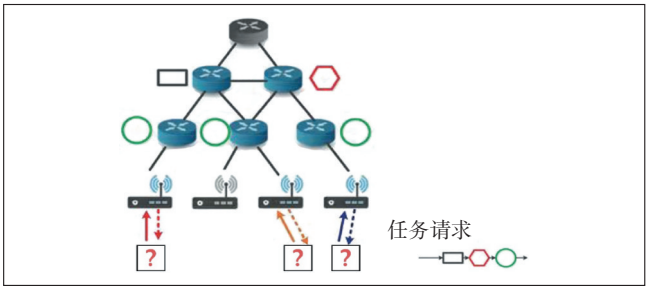

图5　一种平衡分配策略示意图

5 模型建立与算法设计

5.1 模型建立

综合考虑任务拓扑结构和计算资源的异构性，本文提出一种合适的任务卸载策略：分析任务拓扑结构，挖掘潜在的子任务卸载可能，判断哪些子任务适合卸载，哪种卸载组合可以最大限度地降低总时延；考虑传输时延波动和子任务处理时间波动，协调多用户的卸载策略以达到最小化平均延迟。任务卸载模型符号集如表1所示。

$S=[1,2,\cdots,k,\cdots,S]$表示边缘服务器集合；$D=[1,2,\cdots,i,\cdots,D]$表示终端设备集合。那么设备$i$的任务卸载策略表示为

$$\alpha_i=\begin{cases}0 & \text{任务不卸载}\\ k & \text{子任务卸载至服务器}k\end{cases}$$

香农定理是通信模型和信道建模的普适性定理，因此，依据香农定理，当确定所有卸载策略$A=(a_1,a_2,\cdots,a_N)$后，可以得到第i个设备和第k个服务器之间的无线链路传输时间为：

$$r_{i,k}(A)=B\log_2\left(1+\frac{U}{N_0+N}\right)$$

式中，B为设备和服务器之间的可用信道带宽；U表示发射信号的能量值；N_0为环境底噪；N表示与在同段时间选择同一服务器进行处理的任务对信道造成的影响，可以理解为干扰强度。

当计算任务无需进行卸载，本地计算时，子任务

表1　任务卸载模型符号集

$S=[1,2,\cdots,k,\cdots,S]$	边缘服务器集合
$D=[1,2,\cdots,i,\cdots,D]$	终端设备集合
$A=(a_1,a_2,\cdots,a_N)$	任务卸载策略
B	可用信道带宽
U	发射信号的能量值
N_0	环境底噪
N	同频干扰强度
T_{ij}	设备i中的第j个子任务
c_{ij}	处理子任务T_{ij}需要的CPU周期数
$t_{ij}^{k,Queue}$	子任务在边缘处理器中的排队时间
v_i	设备i的子任务集合

T_{ij}的任务处理时间和处理任务所需的能耗分别为：

$$t_{ij}=c_{ij}/P_i$$
$$e_{ij}=\delta_i c_{ij}$$

其中，P_i为设备i的本地处理能力；δ_i为设备i运行单个CPU周期时消耗的能量。

当计算任务需要卸载至边缘服务器上时，子任务T_{ij}卸载至服务器k所需的时间为：

$$t_{ij}^{k}=\frac{m_{ij}}{r_{i,k}(A)}+t_{ij}^{k,\mathrm{Queue}}+c_{ij}/f^k$$

其中，f^k表示服务器k的计算能力，服务器之间的异构性使得不同服务器在不同时间和任务处理情况下的计算能力各不相同；$m_{ij}/r_{i,k}(A)$表示任务上传至服务器所需的时间，m_{ij}为子任务的大小；c_{ij}/f^k表示子任务在服务器中的处理时间。

上述公式表明，任务卸载时长与任务传输、排队和计算时间息息相关，而前置任务又影响着排队时间，因此想要降低总时延就要考虑到每个子任务的卸载策略和相互之间的影响，降低竞争性并得到一个分布式的协调卸载策略。

考虑子任务在网络中的相互影响关系和任务处理的时序问题，用有向无环图（Directed Acyclic Graph，DAG）来表示子任务之间的相互关系。为了实现平均总时延的最小化，首先要对每个子任务的处理情况进行建模。当设备选择在服务器k上进行任务卸载时，任务j的开始执行时间用$\mathrm{ST}(j,k)$表示，子任务结束时间用$\mathrm{FT}(j,k)$表示。$\mathrm{ST}(j,k)$取决于所有前置任务的处理时间，因此采用递归方式从有向无环图的起始任务来定义：

$$\mathrm{ST}(j,k)=\max\{\mathrm{avail}\{0\cup[k]\},\max_{j'\in\mathrm{pred}(j)}(\mathrm{FT}(j')+C_{jj'})\}$$

$$\mathrm{FT}(j')=\min\left\{w_{ij}^{k'}+\mathrm{ST}(j',k')\right\}\quad k'\in\{0\}\cup\{k\}$$

其中，$\mathrm{avail}\{0\cup[k]\}$表示本地服务器或$k$服务器的可用时段；$\mathrm{pred}(j)$表示$j$的所有前置任务，$C_{jj'}$是DAG中$j\to j'$的通信消耗；$\mathrm{FT}(j')$表示$j$的处理结束时间，$w_{ij}^{k'}$表示$j$的处理时长。

$\mathrm{avail}\{0\cup[k]\}$和$\max_{j'\in\mathrm{pred}(j)}(\mathrm{FT}(j')+C_{jj'})$是子任务开始处理的2个必要条件，前者表示服务器可用，后者表示任务处理所需的数据均已处理完毕并传至服务器；二者必须同时满足，因此二者的较大值决定了任务开始的时间。

为了得到总时延，将上述2个公式进行遍历，直到得到最后一个子任务的结束时间，即为整个任务的结束时间：

$$\mathrm{FT}_i=\mathrm{ST}(\mathrm{exit})+t_{i,v_i}^{L}$$

任务卸载一般会由边缘服务器返回一个结果，本地服务器进行收集，因此可以认为最后一个子任务的执行方式为不卸载，t_{i,v_i}^{L}就用来表示最后一个子项目本地执行所需的时长。

5.2 算法设计

为了设计一个能满足低时延和合理资源配置动态平衡的算法，要考虑2个问题：子任务的优先级和边缘服务器的选择。

首先定义DAG中$j\to j'$的权重：

$$C_{jj'}=\begin{cases}0 & \text{当 } a_{ij}=a_{ij'}\\ \dfrac{\mathrm{data}_{j'j}}{r_{i,k}(A)} & \text{其他}\end{cases}$$

$\mathrm{data}_{j'j}$为2个子任务之间进行交互的数据量；$C_{jj'}$表示有关联的子任务之间的通信时延，当子任务在同一个服务器上时，该时延可认为是0；因此DAG中$j\to j'$的平均时延可以定义为：

$$\overline{C_{jj'}}=\mathrm{data}_{j'j}/(2r_{i,k}(A))$$

子任务j的平均处理时间为：

$$\overline{w_{ij}^{k}}=(w_{ij}^{k}+t_{ij}^{l})/2$$

因此，子任务j的优先级可以被定义为：

$$\mathrm{rank}(j)=\overline{w_{ij}^{k}}+\max_{j'\in\mathrm{Succ}(j)}(\overline{C_{jj'}}+\mathrm{rank}(j'))$$

其中，$\mathrm{Succ}(j)$表示直接后继任务。

根据上述模型的建立，可以得到任务卸载算法如图6所示。

该算法考虑了单服务器场景下的任务卸载策略。首先对子任务的优先级进行排序，然后按照优先级最

```
Input：G = (V,E),Data_ij
Output：a_ij
按递减顺序排列 rank(j)
while Data_ij 未被全部定义 do
选择 rank(j) 值最大，即优先级最高的子任务 i；
计算本地处理和卸载至服务器上处理所需的时间
FT_local(i) 和 FT_server(i)；
if FT_local(i) ≤ FT_server(i)
    a_ij = 0；
else
    a_ij = 1；
end if；
end while
```

图6　任务卸载算法

高到最低的顺序依次判断子任务的卸载策略，判断方法为比较不同卸载方式需要的时间（本地处理或边缘处理）并择优，最后通过每个子任务的卸载策略得到总的卸载策略。该算法利用优先级的顺序进行策略制定，避免了对每个子任务进行穷举决策，复杂度仅为O[n]。

6 结束语

本文针对传统工业互联网自动化程度低、灵活性差等问题，提出了一种基于拓扑结构的任务卸载策略和边缘资源分配策略，任务卸载策略通过前期任务处理（判定优先级）等方式，大大降低了复杂度，并有效降低端到端时延；边缘资源分配策略综合考虑了时延和资源之间的平衡，给出了一种在保证时延的条件下，尽可能减少资源浪费的方式。除此之外，本文给出了一种基于拓扑结构的任务卸载算法和流程图，解决了现有任务传输路径算法的不足，保障了工业生产的整体效率。

参考文献：

[1] FRASCOLLA V，MIATTON F，TRAN G K，et al. 5G-MiEdge：design standardization and deployment of 5G phase Ⅱ technologies：MEC and mmWaves joint development for Tokyo 2020 Olympic games[C]// 2017 IEEE Conference on Standards for Communications and Networking (CSCN). Helsinki，Finland：IEEE，2017：54-59.

[2] ORLOSKY J，KIYOKAWA K，TAKEMURA H. Virtual and augmented reality on the 5G highway[J]. Journal of Information Processing，2017，25：133-141.

[3] 段惠斌，丁鹏，时晓厚，等. 基于5G边云协同的高精地图采集与应用研究[J]. 电子技术应用，2020，46(12)：32-35.

[4] 陈强，储云风，朱皆一. 面向5G的多接入边缘计算架构设计与应用[J]. 通信技术，2020，53(8)：1923-1929.

[5] LI L，ZHANG X Y，LIU K Y，et al. An energy-aware task offloading mechanism in multiuser mobile-edge cloud computing[J]. Mobile Information Systems，2018，2018：7646705.

[6] 黄冬艳，付中卫，王波. 计算资源受限的移动边缘计算服务器收益优化策略[J]. 计算机应用，2020，40(3)：765-769.

[7] QUERALTA J P，LI Q Q，ZHUO Z，et al. Enhancing autonomy with blockchain and multi-access edge computing in distributed robotic systems[C]//2020 Fifth International Conference on Fog and Mobile Edge Computing (FMEC). Paris，France：IEEE，2020：180-187.

[8] MA X T，ZHAO J H，GONG Y，et al. Key technologies of MEC towards 5G-enabled vehicular networks[M]//WANG L，QIU T，ZHAO W. Quality，Reliability，Security and Robustness in Heterogeneous Systems. Cham：Springer，2017：153-159.

[9] ANDREWS J G，CLAUSSEN H，DOHLER M. Femtocells：past，present，and future[J]. IEEE Journal on Selected Areas in Communications，2012，30(3)：497-508.

[10] DHILLON H S，GANTI R K，BACCELLI F，et al. Modeling and analysis of K-Tier downlink heterogeneous cellular networks [J]. IEEE Journal on Selected Areas in Communications，2012，30(3)：550-560.

[11] 毋涛，徐雷，刘畅. 5G MEC边缘智能架构研究[J]. 信息通信技术，2020，14(2)：46-49.

[12] 葛海波，冯安琪，王妍. 5G边缘计算环境下工作流任务的卸载策略[J]. 传感器与微系统，2020，39(8)：130-133，137.

[13] XU Z C，LIANG W F，XU W Z，et al. Efficient algorithms for capacitated cloudlet placements[J]. IEEE Transactions on Parallel and Distributed Systems，2016，27(10)：2866-2880.

[14] JIA M K，CAO J N，LIANG W F. Optimal cloudlet placement and user to cloudlet allocation in wireless metropolitan area networks[J]. IEEE Transactions on Cloud Computing，2017，5(4)：725-737.

[15] 史晨华. 基于边缘计算的5G低时延高可靠业务卸载策略研究[D]. 西安：西安电子科技大学，2018.

[16] CHU C H. Task offloading based on deep learning for blockchain in mobile edge computing[J]. Wireless Networks，2021，27(1)：117-127.

[17] 黄晓舸，崔艺凡，张东宇，等. 基于MEC的任务卸载和资源分配联合优化方案[J]. 系统工程与电子技术，2020，42(6)：1386-1394.

[18] GUO S T，LIU J D，YANG Y Y，et al. Energy-efficient dynamic computation offloading and cooperative task scheduling in mobile cloud computing[J]. IEEE Transactions on Mobile Computing，2019，18(2)：319-333.

作者简介：
许波，高级工程师，主要从事通信网络的规划设计、咨询研究及技术管理工作；陈礼波，高级工程师，硕士，主要从事政企创新业务规划、光通信网络规划咨询设计工作；张鹏，毕业于吉林大学，高级工程师，硕士，主要从事移动网络规划、移动通信新技术研究等工作；曹童杰，毕业于东华大学，工程师，硕士，主要从事矿山能源领域数字化转型工作。