

# 移动边缘计算下基于联邦学习的动态 QoS 优化

张鹏程<sup>1)</sup> 魏芯淼<sup>1)</sup> 金惠颖<sup>1)</sup>

<sup>1)</sup>(河海大学计算机与信息学院,南京 211000)

**摘要** 在 5G 边缘网络飞速发展的过程中,边缘用户对高带宽、低时延的网络服务的质量要求也显著提高.从移动边缘网络的角度,网络内的整体服务质量与边缘用户的分配息息相关,用户移动的复杂性为边缘用户分配带来困难,边缘用户分配过程中还存在隐私泄露问题.本文提出一种移动边缘环境下基于联邦学习的动态 QoS(Quality of Service)优化方法 MECFLD\_QoS,基于联邦学习的思想,优化边缘区域的服务缓存,在动态移动场景下根据用户位置分配边缘服务器,有效保护用户隐私,实现区域服务质量优化,对动态用户移动场景有更好的适应性.MECFLD\_QoS 主要做了以下几个方面的优化工作:(1)优化了传统 QoS 数据集,将数据集映射到边缘网络环境中,充分考虑边缘计算的移动、分布式、实时性、复杂场景等特点,形成边缘 QoS 特征数据集;(2)优化了边缘服务器缓存,在用户终端训练用户偏好模型,与区域公有模型交互时只传输参数,将用户的隐私数据封装在用户终端中,避免数据的传输可以有效地保护用户特征隐私;(3)优化了用户移动场景,在动态移动场景中收集用户移动信息,利用用户接入基站的地理位置拟合用户的移动轨迹进行预测,有效地模糊了用户的真实位置,在轨迹预测的同时有效地保护了用户的位置隐私;(4)优化了用户分配方法,提出改进的基于二维解的人工蜂群算法对边缘网络中的用户分配问题进行优化,事实证明改进的人工蜂群算法针对其多变量多峰值的特点有效地优化了用户分配,达到了较优的分配效果.通过边缘 QoS 特征数据集实验表明,本方法在多变量多峰值的用户分配问题中能产生全局最优的分配.

**关键词** 移动边缘; 联邦学习; 移动感知; 边缘用户分配; 服务质量

中图法分类号 TP18

## Dynamic QoS Optimization Method based on Federal Learning in Mobile Edge Computing

ZHANG Peng-Cheng<sup>1)</sup> WEI Xin-Miao<sup>1)</sup> JIN Hui-Ying<sup>1)</sup>

<sup>1)</sup>(School of Computer and Information, Hohai University, Nanjing, China)

**Abstract** The development of 5G networks has expanded the Internet of Things, promoted the operation of cellular networks, and further pushed the Internet to the edge of the network. In the process of rapid development of 5G edge networks, mobile edge users have significantly improved the quality of service (QoS) requirements for high-bandwidth and low-latency Web services. Traditional QoS calculation and optimization are from the perspective of service, and the calculation of quality of experience (QoE) is from the perspective of edge users. From the perspective of mobile edge networks, the overall service quality of the edge area is closely related to the edge users allocation (EUA). As a result, EUA needs to be taken into consideration in the optimization work. The allocation of edge users to achieve the best overall service quality in the area is a hot topic. The complexity of user movement behavior and scenes brings difficulties to the edge user allocation. There are also privacy leakage problems in the edge user allocation process. This paper proposes a dynamic QoS optimization method in mobile edge environment based on federal learning (MECFLD\_QoS). Based on the essence of federal learning, the proposed method uses gradient descent to solve the logistic regression model, trains the regional public

model to optimize the Web service cache of the edge area, and allocates edge servers according to the user location in the dynamic mobile scenario, effectively protecting user privacy. The regional service quality is optimized, and it has better adaptability to dynamic user mobility scenarios. MECFLD\_QoS mainly optimized the following aspects: (1) Optimizes the traditional QoS data set, maps the data set to the edge network environment, and fully considers the mobile, distributed, real-time, and complex scenarios features of edge computing, and forms an edge QoS feature data set; (2) Optimizes the edge server cache, trains the user preference model on the user terminals, only transmits model parameters when interacting with the regional public model, and encapsulates the user's private data in the user terminal to avoid data transmission, which can effectively protect user characteristic privacy; (3) Optimizes user mobile scenarios, collects user mobile information in dynamic mobile scenarios, and uses the geographic location of the user to access the base station to fit the user's movement trajectory for prediction, which effectively blurs the user's reality location while predicting the trajectory, effectively protects the users' location privacy; (4) Optimizes the user allocation method, and proposed an improved two-dimensional-solutionbased artificial bee colony algorithm (ABC) to optimize the user allocation problem in the edge network. Facts have proved that the improved artificial bee colony algorithm effectively optimizes user allocation according to its multi-variable and multi-peak characteristics, and achieves a better allocation effect. To sum up, the MECFLD\_QoS method proposed in this paper protects the user's characteristic privacy and location privacy while ensuring the optimization effect. So it has a good preference on achieving good results in privacy protection. The experiments on edge QoS feature data set show that the method proposed can produce the global optimal allocation in the multi-variable and multi-peak user allocation problem.

**Key words** Mobile Edge; Federal Learning; Mobility Aware; Edge User Allocation; Quality of Service

如何优化 QoS 成为边缘计算的研究热点。

## 1 引言

在面向服务体系架构<sup>[1]</sup>(Service-Oriented Architecture, SOA)中, Web 服务是独立于平台的低耦合、自包含、可编程的应用程序. 在用户选择、调用 Web 服务的过程中, 可感知 Web 服务的非功能属性, 包括响应时间、吞吐量等, 称为 Web 服务的服务质量(Quality of Service, QoS).

随着政府报告重点支持“两新一重”建设, 加快布局 5G 网络、物联网等新型基础设施. 5G 技术也发展成为国家未来的战略发展方向, 5G 网络、物联网等新基建陆续有新政策出台支持其发展, 从而带动 5G 行业的发展. 5G 基站作为 5G 网络的核心设备, 能够实现有线通信网络和无线终端之间的信号传输, 也将实现更高密度的部署. 边缘计算作为 5G 核心技术之一, 将高带宽、低时延、本地化的业务下沉到网络边缘<sup>[2]</sup>, 解决了网络整体流量大、时延长的问题, 为 5G 网络中的实时性和带宽密集型业务提供更好的支持. 在业务下沉过程中, 由于边缘服务器的计算性能不同, 用户移动设备的位置也在不断发生变化, 边缘环境下不同的用户移动设备的分配策略会产生不同的服务质量. 因此在边缘环境下

大量学者从 Web 服务的角度, 在计算卸载和任务调度方面开展 QoS 优化的研究, 通过考虑边缘节点资源消耗约束, 服务的业务逻辑约束, 以及用户移动性等约束, 形成任务调度策略, 计算相应的响应时间和吞吐量等 QoS 属性值, 找到最优的调度策略, 主要有优化任务接受率<sup>[3]</sup>、优化策略成本<sup>[4][5]</sup>、优化计算卸载<sup>[6][7]</sup>和优化资源分配<sup>[8][9]</sup>. 尽管这些方法在移动感知的边缘计算优化方面取得一些进展, 但现有的方法不适用于复杂的用户移动场景中, 无法感知复杂移动下的相关约束, 另一方面现有的工作未能在动态边缘环境下优化边缘服务的服务质量, 优化效果无法统一比较<sup>[3][4][6][7]</sup>. 现有的移动边缘计算的优化策略存在问题归纳如下:

(1) 现有的计算卸载和任务调度方法未考虑边缘服务器的服务缓存对边缘服务的质量的影响. 考虑移动边缘云的有限计算, 通信和存储资源<sup>[9]</sup>, 需要设计能量消耗最小的服务缓存方案减少服务部署时间. 现有的服务缓存策略取决于用户对服务的偏好程度. 服务缓存和计算卸载与任务调度的优化同样能够缩短边缘服务调用的响应时间, 因此服务缓存作为优化路线, 应在进行 QoS 优化时列入考虑.

(2) 现有的移动感知方法尽管考虑用户移动性, 但未考虑移动场景的动态变化. 当前基于用户移动

<sup>1</sup> [http://www.gov.cn/premier/2020-05/29/content\\_5516072.htm](http://www.gov.cn/premier/2020-05/29/content_5516072.htm)

性的优化方法只考虑特定静态场景<sup>[7][8][10]</sup>,如病人健康护理监控、智慧警务系统以及指定路口的用户移动场景,提出的方法对于复杂多变的动态移动场景没有较好的适配度,无法随场景的变化预测用户的移动路线,因此不能反映真实 5G 场景下的用户移动和边缘服务复杂性.本文考虑了在复杂、多元的边缘计算环境下的用户移动问题,在动态移动场景下解决 QoS 优化问题.

(3)用户信息泄露现象严重.随着越来越多的用户移动设备调用边缘服务,需要对用户的行为特征进行研究以预测用户在边缘网络环境中的位置和进行用户分配,这一过程会导致用户的特征隐私和的位置隐私的泄露<sup>[3][9]</sup>.因此隐私保护在动态 QoS 优化过程中具有非常重要的意义.

针对上述三个问题,本文提出一种移动边缘环境下基于联邦学习的服务质量优化方法(Dynamic QoS Optimization base on Federal Learning in Mobile Edge Computing, MECFLD\_QoS).从移动边缘网络的角度,针对复杂多变的移动边缘环境,还原真实的用户移动和服务调用行为,收集准确的移动边缘数据.一方面通过联邦学习训练用户移动设备偏好模型和区域公有模型确定边缘服务缓存,在缩短边缘服务响应时间的同时保护用户的特征信息.另一方面,通过定位基站的位置来还原用户的移动信息,确定可分配的边缘服务器,优化边缘用户分配以保护用户的位置信息.最后利用改进的基于二维解的人工蜂群算法分配边缘用户,优化区域服务质量.总体来说,本文的贡献包括以下三个方面:

(1)针对问题一,由于边缘服务器的服务缓存对边缘服务调用产生的 QoS 有一定影响<sup>[11]</sup>,已缓存的边缘服务相比未缓存的边缘服务在调用时传输数据量更小、计算速度更快,因此通过训练区域公有模型确定边缘服务器的服务缓存,将调用频率更高的边缘服务缓存在边缘服务器中,缩短更多边缘服务调用的响应时间,以提高区域服务质量.

(2)针对问题二,由于当前的移动感知方法适用于小范围且固定场景中,未考虑动态场景下的用户移动和行为特征.因此本文提出的拉格朗日轨迹预测能够在移动边缘动态场景中考虑用户的移动性,从而扩大移动感知方法的适用范围,增强其可用性.

(3)针对问题三,针对用户的特征隐私,由于联邦学习具有共同学习、合作建模的特点<sup>[12]</sup>,在建立公有模型时只需上传训练参数,可有效避免用户隐

私信息在传输过程中泄露,因此本文利用联邦学习优化服务缓存可以有效地保护用户的特征隐私;针对用户的位置隐私,我们采用边缘基站的位置来预测用户的移动路径,结合“K 匿名”思想<sup>[13]</sup>的拉格朗日轨迹预测方法可以有效地模糊用户的具体位置信息,以达到保护用户的位置隐私的目的,同时在动态场景下预测用户的移动性.

本文的第 2 节介绍移动边缘计算的相关优化工作;第 3 节介绍本方法的背景知识以及相关方法的理论基础;第 4 节详细介绍本文提出的移动边缘环境下基于联邦学习的动态 QoS 优化方法 MECFLD\_QoS;第 5 节进行实验设计并评估本文方法;第 6 节总结本文的工作并进行展望.

## 2 相关工作

现有的移动边缘网络 QoS 优化工作主要包括用户分配优化和任务调度优化,基于计算资源和服务质量对移动边缘网络进行优化.

### 2.1 边缘用户分配

在移动边缘环境中,用户移动设备的存储和计算能力有限,因此移动边缘网络支持它们将密集的计算任务填充到邻近的边缘服务器.边缘服务器在满足覆盖范围和任务量时可为接入的用户移动设备提供服务.传统的 QoS 计算和优化从服务的角度计算服务在调用过程中的各项属性;服务体验(Quality of Experience, QoE)是从用户的角度计算用户在调用过程中获得的不同级别 QoS 的满意程度<sup>[14]</sup>;从移动边缘网络的角度看,区域整体 QoS 的计算与边缘用户的分配息息相关,因此需要考虑边缘用户的分配以实现区域整体服务质量的最优.服务提供商在部署服务时,也需考虑边缘计算中应用的按需付费定价与边缘服务器数量之间的经济效益.

He<sup>[14][14][16]</sup>将此类问题定义为边缘用户分配问题(Edge User Allocation, EUA),对此类问题进行建模,分析边缘环境的资源、距离等约束,提出启发式方法解决 EUA 问题;在此基础上,He 进一步考虑了边缘服务用户的动态 QoS 级别,以找到一种可最大程度提高应用程序用户总体 QoE 的解决方案,以及基于距离感知和对抗感知的边缘用户分配. Peng<sup>[17]</sup>也在边缘用户分配问题上进行了研究,基于用户的移动性和时变性提供在线决策.上述的边缘用户分配问题没有针对区域的整体 QoS 进行优化.

## 2.2 计算卸载和任务调度

现有的大量边缘计算优化工作集中在用户卸载决策问题上,这类工作先估算研究任务的延迟和能耗,再制定将任务部署到服务器或在本地执行的策略,最后观察执行调度策略后的能耗和时延是否达到优化效果.

Deng<sup>[8]</sup>提出了在资源和性能要求受限时在 MEC 环境中通过优化应用程序部署减小部署成本;Miao<sup>[6]</sup>提出基于任务预测的计算分流和任务迁移算法,在数据和服务增加的同时有效地减小总任务的延时.在边缘环境中用户的移动性也是不容忽视的一个因素,Xu<sup>[7]</sup>在医疗监控场景下,通过调整移动感知的工作流卸载和调度策略,降低能耗和工作量以实现最佳的适应性,Wang<sup>[4]</sup>同时考虑了任务属性、用户移动性和网络约束,以减少任务调度对任务执行的延迟.

## 2.3 服务缓存

由于将服务放置在边缘服务器上可以减少用户获得服务的等待时间,因此许多研究人员研究服务放置,目的在于探索最佳服务放置方案,以实现最大程度地提高移动设备的服务质量.近年来,服务放置吸引了许多学者的目光,专家学者提出了一些服务放置方案,如 Liang<sup>[18]</sup>提出在总放置成本不超过指定的预算阈值的约束下,决定将服务实体放置在交互式应用程序用户之间以实现低延迟成对交互的位置,You<sup>[19]</sup>等提出权衡成本与延迟的服务放置模型,Qian<sup>[9]</sup>提出基于联邦学习确定服务放置策略,结合用户偏好放置服务,同时保护了用户移动设备的隐私.

上述 QoS 优化方法中,基于计算卸载和任务调度的优化问题在考虑用户的移动性时,研究场景不具有动态性和不确定性,无法在动态场景中应用,现有工作只单一考虑用户、服务其中之一进行 QoS 优化,没有进行综合优化,对用户特征隐私和位置隐私的泄露问题也没有提出解决方案.因此目前还不存在面向隐私保护的移动边缘环境下基于联邦学习的动态 QoS 优化方法.

# 3 预备知识

## 3.1 移动边缘计算

5G 网络推动了物联网的发展,推进了蜂窝网络的运营,将互联网推向网络边缘.近几年,响应国家

“两新一重”建设,国家加强新型基础设施建设,发展新一代信息网络,拓展 5G 应用,这意味着 5G 成为国家未来的战略发展方向,国内通信运营商也加大 5G 基站的建设规模.在去中心化的 5G 网络中,诸多高带宽、低时延的任务需要在网络边缘部署小规模数据中心,实现请求的本地化处理<sup>[20]</sup>.在移动边缘网络中,基站是网络的通信结点,而边缘服务器是网络的计算结点,由于在实际部署中,基站和边缘服务器部署在同一位置,因此本文默认基站和边缘服务器地处相同位置.

## 3.2 联邦学习

联邦学习作为一种新型的人工智能技术,2016年由谷歌最先提出,用于处理安卓移动设备的本地化模型训练,联邦学习在保证本地数据的隐私的同时,在多参与方或多计算结点之间开展高效率的机器学习<sup>[21]</sup>.联邦学习使用共同学习、合作建模的思想,其参与者地位对等,可以在彼此保持独立性的同时交换信息参数,同时能够保证各参与者之间数据隔离,在模型质量无损的同时满足用户隐私保护和数据安全的要求.针对不同的参与各方数据源分布情况,可以将联邦学习分为三类:联邦迁移学习、纵向联邦学习和横向联邦学习.

联邦迁移学习在数据集用户和用户特征都重合较少的情况下,解决数据和标签稀疏的问题;纵向联邦学习在数据集用户重叠较多但用户特征重叠较少的情况下,训练特征维度不同的数据;横向联邦学习在数据集用户特征重叠较多但用户重叠较少的情况下训练用户维度不同的数据.由于本文针对的是同一边缘区域中的不同用户,因此训练该边缘区域的不同用户的用户偏好模型,在区域内进行合作建模,该问题中各用户的特征维度相同,在不同用户移动设备上训练用户偏好模型,因此选用横向联邦学习来构建区域公有模型.

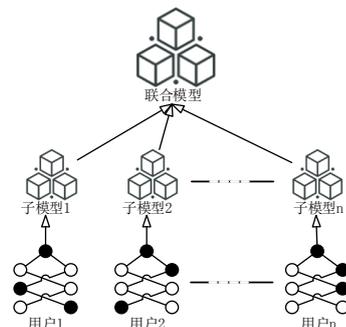


图 1 横向联邦学习过程

横向联邦学习的学习过程如图 1 所示,各参与用户特征维度相同但个体不同,公有模型需要对其

进行样本融合. 因此各参与用户下载最新模型, 各参与用户利用本地的数据和特征信息训练模型参数, 上传到公有模型中, 公有模型聚合各用户的梯度更新模型参数, 再将更新参数传回参与方用于更新本地的训练模型用于下一次训练.

### 3.3 拉格朗日插值法

在数值分析中, 许多实际问题都用函数来表示某种内在联系或规律<sup>[22][23]</sup>. 拉格朗日插值法对于实践中的物理量, 可以恰好找到一个多项式, 使其在观测点取到物理量的观测值. 设函数  $y = f(x)$  在区间  $[a, b]$  上有定义, 且给出一系列观测点上的观测值  $y_i = f(x_i) (i = 0, 1, 2, \dots, n)$ , 则可以找到一个  $n$  次多项式  $p_n(x)$ , 使得

$$p_n(x_i) = y_i (i = 0, 1, 2, \dots, n) \quad (1)$$

函数  $p_n(x)$  为  $f(x)$  的插值函数,  $x_0, x_1, \dots, x_n$  称为插值结点, 所在区间  $[a, b]$  为插值区间,  $p_n(x_i) = y_i$  称为插值条件. 构造的  $n$  次多项式可表示为:

$$P_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (2)$$

当  $n = 1$  时, 问题即简化为要求通过  $(x_0, y_0), (x_1, y_1)$  两点的一次多项式  $P_1(x)$ , 由此可以得出直线方程的两点式:

$$L_1(x) = \frac{x-x_1}{x_0-x_1}y_0 + \frac{x-x_0}{x_1-x_0}y_1 \quad (3)$$

其中令 1 次多项式  $l_0(x) = \frac{x-x_1}{x_0-x_1}, l_1(x) = \frac{x-x_0}{x_1-x_0}$ ,

可以得到拉式基函数<sup>[23]</sup>:

$$L_1(x) = \sum_{i=0}^1 l_i(x)y_i \quad (4)$$

插值法是数学建模中的一种常用方法, 一般适合数据准确或数据量小的情形, 鉴于数据的稀疏性, 本文通过线性插值拟合用户的移动轨迹.

### 3.4 人工蜂群算法

人工蜂群算法(Artificial Bee Colony Algorithm, ABC)通过模拟蜜蜂群体的觅食行为解决多变量函数优化问题<sup>[24]</sup>, 蜜蜂群体模型与 QoS 优化问题对应关系如表 1 所示. 由表 1 可知, 在人工蜂群算法中, 一个蜜源位置对应 QoS 优化问题的一个可行的用户-基站分配, 而蜜源的花蜜含量对应 QoS 优化问题中该可行解的整体 QoS. 蜂群的采蜜系统由蜜源、

雇佣蜂与非雇佣蜂三个部分组成. 雇佣蜂只包含采蜜蜂一个工种, 负责与某个蜜源产生联系并将蜜源信息在蜂巢的舞蹈区传递; 非雇佣蜂包含观察蜂和侦察蜂两个工种, 主要负责开采新蜜源, 侦察蜂以一定的领域搜索规则搜索新蜜源, 同时在舞蹈区传递信息, 观察蜂观察蜜源信息并对蜜源质量进行评价, 向高质量的蜜源聚集. 在人工蜂群算法中, 所有蜜蜂共分为 3 个工种: 采蜜蜂、观察蜂和侦察蜂. 首先, 人工蜂群算法随机产生  $N$  个初始解, 即  $N$  个采蜜蜂和初始蜜源, 每个解是一个  $D$  维向量,  $D$  为优化参数的数量. 在初始化蜜源后, 采蜜蜂对所有初始解进行循环搜索. 侦察蜂会以一定概率改变记忆中的食物源的位置从而进行领域搜索, 找到新蜜源, 并计算新解的适应度值, 若产生的新解适应度值更高, 则用新蜜源位置替换原来的蜜源位置. 侦察蜂完成领域搜索后与观察蜂共享蜜源信息, 观察蜂根据采蜜蜂搜集到的信息, 以一定的概率选择与现有蜜源位置相关的蜜源位置, 计算新适应度值, 若高于原蜜源的适应度, 则替换蜜源.

表 1 QoS 优化问题与蜜蜂群体模型对应关系表

蜜蜂群体模型	QoS 优化问题
采蜜蜂	初始化用户-基站分配进程
侦察蜂	领域搜索可行分配进程
观察蜂	选择可行分配进程
蜜源位置	用户-基站可行分配
蜜源花蜜量	可行分配的 QoS

## 4 MAFLD\_QoS 优化方法

传统的 QoS 计算和优化从服务的角度出发优化资源卸载和任务调度; 而 QoE 的计算是从用户的角度优化用户对边缘服务质量的满意程度; 在移动边缘环境中, 移动边缘网络(Mobile Edge Network, MEN)在靠近用户的地方提供计算和存储服务, 以满足高工作量和低延迟的需求<sup>[4]</sup>, MEN 由移动边缘计算中具有计算和存储功能的小型蜂窝基站构成, 边缘用户也是网络中的行为主体, 在边缘用户和基

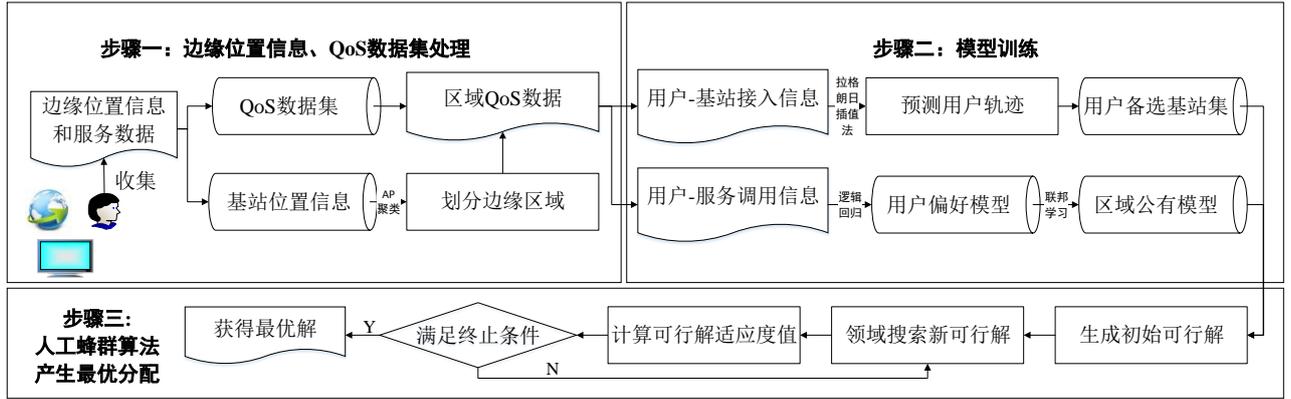


图 2 MECFLD\_QoS 方法图

站的交互过程中调用边缘服务产生服务质量。从 MEN 的角度优化区域 QoS, 则与边缘用户的分配息息相关, 因此需要考虑边缘用户的分配以实现优化区域整体服务质量的目标。此外, 本文考虑隐私保护和用户移动场景的动态性, 提出了一种移动边缘环境下基于联邦学习的动态服务质量优化方法 MECFLD\_QoS, 流程如图 2 所示。由第三方收集数据集、划分区域, 本方法在用户层面提取用户的移动信息、预测用户移动方向从而实现场景动态优化, 同时在边缘节点层面训练用户偏好模型, 实现公有模型优化, 以优化边缘服务响应时间, 最后根据优化的场景和 QoS 属性进行区域 QoS 优化。主要分为以下三个步骤:

(1) 边缘位置信息、QoS 数据集收集和处理: 收集边缘地理位置信息和 QoS 属性值, 分别对两个数据集进行处理。首先将基站的经纬度值映射到平面直角坐标系, 以反映基站之间的实际距离, 从而获取边缘服务器的分布情况和移动用户的接入情况; 再将传统的 QoS 数据集中的用户根据用户接入情况分配到边缘服务器中, 形成边缘环境下边缘 QoS 特征数据集。

(2) 模型训练: 在得到边缘 QoS 特征数据集后, 根据其用户-服务调用信息, 基于联邦学习的思想, 利用梯度下降的逻辑回归模型训练区域公有模型, 确定边缘服务器服务缓存的同时有效保护了用户的特征信息; 根据其用户-基站接入信息, 利用拉格朗日插值法还原用户的运行轨迹, 并预测其移动方向以确定用户备选基站集。

(3) 区域 QoS 优化: 根据下一时刻用户调用的服务信息和模型训练用户备选基站集信息为用户分配边缘服务器, 每一种可行分配为该问题的一个可行解, 根据历史信息和模型训练结果计算出每个

可行解的适应度值, 即整体 QoS 值。最后利用人工蜂群算法更新、迭代, 找出最优解。

#### 4.1 数据收集和预处理

参考文献<sup>[25]</sup>的方法, 收集完整的服务质量数据和边缘位置信息, 分别对基站位置点和服务质量数据进行预处理, 具体预处理过程由以下步骤组成:

##### (1) 边缘区域位置划分

根据边缘位置结点的位置关系划分区域, 分为以下三个步骤, 步骤一: 筛选研究区域内的有效点, 选取不重复的位置作为边缘位置点, 每个边缘位置点为一个边缘服务器和基站; 步骤二: 利用通用横轴墨卡托投影 (Universal Transverse Mercator Projection, UTM)<sup>[26]</sup>将数据集中的每个边缘节点的经纬度投影到平面坐标中, 减少投影形变, 以准确反映边缘结点之间的实际距离; 步骤三: 利用 AP 聚类算法<sup>[27]</sup>划分区域,  $N = \{n_i\}_{i=1}^K$  为待聚类的各个位置点集, 其中  $K$  表示位置点的个数,  $S$  为一个刻画位置点之间相似度的矩阵, 对于集合  $N$  中的任何位置点  $n_k$ , 可以用吸引信息  $r(i, k)$  和归属信息  $\alpha(i, k)$  刻画该位置点的特征。

吸引信息  $r(i, k)$  描述位置点  $k$  作为聚类中心对位置点  $i$  的吸引程度, 吸引信息  $r_{t+1}(i, k)$  迭代如下:

$$r_{t+1}(i, k) = s(i, k) - \max_{k \neq k'} \{\alpha_t(i, k') + s(i, k')\} \quad (5)$$

其中  $s(i, k)$  代表矩阵  $S$  中位置点  $i$  和位置点  $k$  之间的相似度。归属信息  $\alpha(i, k)$  描述位置点  $k$  被位置点  $i$  选为聚类中心的适应程度, 归属信息  $\alpha(i, k)$  按如下方式迭代:

$$\alpha_{t+1}(i, k) = \min\{0, r_t(k, k) + \sum_{i' \in \{i, k\}} \max\{0, r_t(i', k)\}, i \neq k\} \quad (6)$$

$$\alpha_{t+1}(k, k) = \sum_{i' \neq k} \max\{0, r_t(i', k)\} \quad (7)$$

初始化后,若 $r(i,k)$ 变大,则 $\alpha(i,k)$ 变小; $\alpha(i,k)$ 变小反馈信息促使 $r(i,k)$ 继续变大,从而达到一个稳定的值.根据 AP 聚类算法对以上步骤进行迭代,若这些决策经过若干次迭代之后保持不变或者算法执行超过设定的迭代次数,或一个小区域内的关于样本点的决策经过数次迭代后保持不变,则算法结束.由此得到聚类结果.

### (2) QoS 数据集处理

现有的 QoS 数据集无法体现边缘特征,亦无法满足 5G 网络中高带宽、低时延的需求,其表示通过以下方式将传统的服务质量数据集中的记录根据用户接入基站情况分配到边缘服务器中形成的边缘环境下的 QoS 特征数据集.在  $T$  时间段内,用户集  $U$  基于边缘服务器集  $ES$  调用了边缘服务集  $WS$ .首先,提取基站位置信息中一定时间内的数据,根据传统服务质量数据集的时间间隔为其时间进行编号,并提取各段内用户接入边缘节点的信息,得到用户接入边缘服务器的信息表;其次,整理传统 QoS 数据集,提取各时间段内用户调用边缘服务的信息以及相应的服务质量属性,得到用户调用边缘服务信息表.

### (3) 边缘服务质量特征数据集

完成边缘区域划分后,根据时间和用户编号将 QoS 数据集中相应的服务调用及其产生的服务质量属性分配到各个区域的各个基站中,形成边缘环境下边缘 QoS 特征数据集.具体分配过程如图 3 所示.

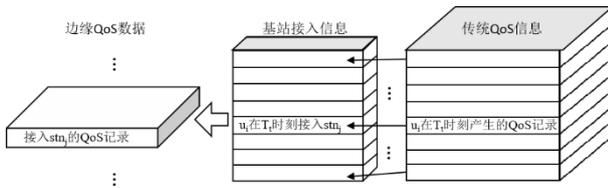


图 3 边缘节点用户分配过程

所得边缘 QoS 特征数据集在一个实验区域内,每组不重复的经纬度坐标对应一组边缘服务器和网络基站,在一个时间段内接入该坐标的用户在 QoS 数据集中对应产生的服务调用记录和相关的服务质量属性分配到该位置点,形成边缘 QoS 记录,当用户发生移动、接入的位置点发生改变时,其服务调用记录也随之分配到新的位置点中.记一个实验区域包含的边缘节点集  $ES = \{es_1, es_2, \dots, es_n\}$ , 用户集  $U = \{u_1, u_2, \dots, u_m\}$ , 所调用的边缘服务集  $WS = \{ws_1, ws_2, \dots, ws_k\}$ , 其中  $n, m, k$  均为正整数.本文通过优化服务缓存以及预测用户移动方向进行区域 QoS 优化,因此对每个实验区域,需提取区

域内用户-服务调用信息和用户-基站接入信息以训练服务缓存模型和用户移动模型.

## 4.2 基于联邦学习的梯度下降逻辑回归模型算法

联邦学习的概念是为了解决数据所有者之间的隐私泄露问题而提出的.联邦学习通过允许用户协作训练共享模型,将个人隐私数据保留在用户端,从而减轻其隐私泄露问题.联邦学习作为一种分布式学习框架,在隐私保护上有出色的表现.在移动边缘计算环境中,边缘服务器和物联网设备可能面临安全攻击.在这个问题上,本方法将用户私有模型封装在移动设备中,在联合训练公有模型时,只传递模型参数,并不发生用户敏感数据的传递,攻击者无法在数据传输于计算结点间时获取用户敏感数据.若攻击者攻击各计算结点内的敏感数据,则需由各计算结点的防火墙对此类攻击进行防御,本方法不涉及防火墙隐私保护的相关内容.因此,联邦学习可以用在移动边缘网络上机器学习模型训练中,在网络中保护用户隐私<sup>[9]</sup>.在本文提出的问题中,用户在移动边缘环境中移动,调用边缘服务从而产生了服务质量.在用户偏好模型的训练过程中,用户在移动端训练用户模型,再将参数上传,训练区域公有模型.各用户的维度特征基本相同,各子模型的用户样本几乎不重合,故采用横向联邦学习进行训练.联邦学习系统中主要有两个实体,即数据所有者和模型所有者.本文基于一个实验区域内研究服务缓存问题,数据所有者即移动用户,模型所有者即区域内边缘服务器集群.每个数据所有者  $u_i$  都有一个私有数据集  $D_i \in U$ , 包括用户的隐私特征数据和各个边缘服务的历史调用数据,每个数据所有者  $u_i$  使用其数据集  $D_i$  来训练局部模型,并且仅将局部模型参数发送到联邦学习模型所有者.

局部模型使用逻辑回归模型<sup>[28]</sup>预测服务被选中部署在边缘服务器集群的概率.私有数据集  $D_i$  包含用户特征向量  $x = (x_0, x_1, \dots, x_p)$ , 表示一个  $p$  维的用户特征,构造常数项特征  $x_0 = 1$ ,  $D_i$  使用定性变量  $y_k \in \{0,1\}$  表示对于边缘服务  $ws_k$  的历史调用情况.向量  $\omega = (\omega_0, \omega_1, \dots, \omega_p)$  为回归系数向量.逻辑回归在线性回归的基础上,套用一个逻辑函数来缩小预测范围.逻辑函数如下:

$$h_{\omega}(x) = \frac{1}{1+e^{-\omega^T x}} \quad (8)$$

函数  $h_{\omega}(x)$  表示结果  $y$  取 1 的概率,因此对于输入  $x$ , 分类结果为类别 1 和类别 0 的概率分别为:

$$P(y_k = 1|x; \omega) = h_{\omega}(x) \quad (9)$$

$$P(y_k = 0|x; \omega) = 1 - h_\omega(x) \quad (10)$$

对于单个样本, 其后验概率为:

$$P(y_k|x; \omega) = (h_\omega(x))^{y_k} (1 - h_\omega(x))^{1-y_k} \quad (11)$$

其损失函数为:

$$\begin{aligned} Cost(h_\omega(x), y_k) &= -y_k \log(h_\omega(x)) \\ &\quad - (1 - y_k) \log(1 - h_\omega(x)) \end{aligned} \quad (12)$$

对于边缘服务 $ws_k$ ,  $m$ 个用户样本的平均损失为:

$$J(\omega^k) = \frac{1}{m} \sum_{i=1}^m Cost(h_\omega(x_i), y_{ik}) \quad (13)$$

通过梯度下降方法计算梯度迭代公式:

$$\omega_i^k := \omega_i^k - \alpha \frac{1}{m} \sum_{i=1}^m (h_\omega(x_i) - y_{ik}) x_i \quad (14)$$

其中 $\alpha$ 为学习率, 迭代一定次数后停止学习.

由此可以训练用户模型获得用户 $u_i$ 调用边缘服务 $ws_k$ 的回归系数 $\omega_i^k$ , 将其发送到区域边缘服务器集群, 得到区域内调用边缘服务 $ws_k$ 的回归系数:

$$\omega^k = \frac{1}{m} \sum_{i=1}^m \omega_i^k \quad (15)$$

联邦学习模型根据区域的整体特征值 $X = (X_1, X_2, \dots, X_p)$ 预测区域缓存边缘服务 $ws_k$ 的概率:

$$P(y_k = 1) = \frac{1}{1 + e^{-\omega^k X}} \quad (16)$$

选择调用概率较高的边缘服务进行缓存. 以对用户隐私数据进行计算并将用户参数的训练封装在用户移动设备中, 从而保护了用户隐私. 整体过程的算法描述如算法 1.

**算法 1.** 基于联邦学习的梯度下降逻辑回归模型算法.

输入: 区域特征值 $X$ , 用户集 $U$ , 各用户 $u_i$ 的私有数据集 $D_i$ , 包含用户特征向量 $x$ , 和定性变量 $y_k$

输出: 区域缓存边缘服务的概率 $P(y_k = 1)$

1. 对用户集 $U$ 中的每个用户 $u_i$ :
2. 提取私有数据集 $D_i$ 的特征信息 $x$ 和定性变量 $y_k$ ;
3. 初始化回归系数 $\omega_i^k$ ;
4. 计算样本的权重分数 $\omega_i^k T x$ ;
5. 计算损失函数 $Cost(h_\omega(x), y_k)$ ;
6. 迭代:
7. 利用梯度下降方法更新回归系数 $\omega_i^k$ ;
8. 将回归系数 $\omega_i^k$ 传至区域模型;
9. 计算区域回归系数 $\omega^k$ ;
10. 根据 $X$ 计算区域缓存边缘服务的概率 $P(y_k = 1)$ ;

### 4.3 拉格朗日插值轨迹预测法

在考虑用户的移动性问题时, 一般要获取用户的实际地理位置进行分析和预测, 用户的位置隐私就会面临泄露的风险. 本文从数据集特征和保护用

户位置隐私的角度, 通过用户接入的边缘节点的位置拟合用户移动路径, 借鉴“K 匿名”的思想在还原边缘用户移动路径的同时保护了用户的位置隐私. 用户在移动边缘环境下移动时, 用户的身份信息 and 用户敏感位置信息相对应, 攻击者可以通过用户身份推测出用户敏感位置信息. 传统的处理方法有匿名、数据扰乱和数据加密, 匿名需要删除用户标识, 无法进一步处理移动信息, 数据扰乱会破坏数据的真实性和完整性, 数据加密需要过多的计算资源. 因此引入“K 匿名”方法, 通过对数据进行抽象和隐匿来发布精度较低的数据来隐藏敏感信息. 在本方法中, 通过用户接入的基站位置来抽象用户的敏感位置信息, 隐匿用户具体位置, 使攻击者无法直接获取用户敏感位置信息.

由于函数的本质为单值的映射关系, 每个输入对应唯一输出. 而在用户轨迹的预测过程中, 表示用户位置的经纬度坐标以及其映射的平面坐标, 存在一个输入对应多个输出的轨迹, 因此无法通过一个函数关系建立整段轨迹的整体模型, 故对轨迹进行分段建模以预测用户走向. 由于分段轨迹的位置数量较少且不确定, 故使用线性拉格朗日插值进行用户轨迹预测.

4.1 中提取的用户-基站接入信息包含各个时间段的用户集 $U = \{u_1, u_2, \dots, u_m\}$ 和边缘基站集 $ES = \{es_1, es_2, \dots, es_n\}$ , 因此可以得到用户-基站接入矩阵 $K = \{k_{it}\} (i \leq m, t \leq T)$ , 其中 $k_{it}$ 为用户 $u_i$ 在第 $t$ 时间段接入的边缘基站 $es_j$ 的编号. 通过提取用户 $u_i$ 的接入基站序列 $G_i = \{k_{i0}, k_{i1}, \dots, k_{iT}\}$ , 根据基站的平面坐标得到一组观测值 $\{(x_{i0}, y_{i0}), (x_{i1}, y_{i1}), \dots, (x_{iT}, y_{iT})\}$ 和各个位置点的地理位置对观测值分组, 令每一组观测值的横坐标单调变化. 针对每一组观测值 $\{(x_{ihead}, y_{ihead}), \dots, (x_{ifoot}, y_{ifoot})\}$ 进行插值, 插值区间 $\hat{x}_i = (\hat{x}_{ihead}, \dots, \hat{x}_{ifoot})$ , 其中参数 $step$ 为插值步长且共有 $k$ 个插值点, 插值区间计算如下:

$$\hat{x}_{ihead} = [\min\{x_i\}/step] * step \quad (17)$$

$$\hat{x}_{ifoot} = [\max\{x_i\}/step] * step \quad (18)$$

可以计算插值预测值<sup>[22]</sup>:

$$\hat{y}_i = \sum_{p=0}^k (y_p * \prod_{q=0, q \neq p}^k \frac{\hat{x}_i - x_q}{x_p - x_q}) \quad (19)$$

拼接各组预测值得到还原的用户移动轨迹. 用户 $u_i$ 的备选边缘节点集合 $A$ 中的元素 $es_i$ 满足:

$$(\hat{x}_{ifoot}, \hat{y}_{ifoot}) \in cov(es_i) \quad (20)$$

其中 $cov(es_i)$ 表示边缘节点 $es_i$ 的覆盖范围.

#### 4.4 基于二维解的人工蜂群算法

为了优化区域的整体 QoS 值, 本文根据边缘服务器的缓存队列和用户的移动情况为移动用户动态分配边缘服务器. 由于基于缓存的动态边缘用户分配问题是一个多峰值多变量的优化问题, 事实证明, 边缘用户的分配和整体的服务质量存在多峰值的函数关系, 人工蜂群算法可以在很大程度上避免陷入局部最优解<sup>[29][30][31]</sup>, 从而找到最优解.

传统的人工蜂群算法<sup>[24]</sup>是基于一维解向量进行优化, 反映各参数的优化效果; 本文介绍针对二维解向量设计人工蜂群算法的优化过程. 本文提出的移动边缘网络优化问题的变量为一组边缘用户和基站的分配, 对这样的二维解进行迭代和进化需要对传统的人工蜂群算法进行改进, 针对二维解重新指定蜂群的蜜源搜索和迭代更新策略, 从而获得最优分配. 找到各分配方式的优化效果, 从而进行比较. 基于缓存的动态边缘用户分配问题可以形式化如下:

$$\text{maximize}(Q = \sum_{i=1}^m \sum_{j=1}^n \sum_{l=1}^k (s_{ij} * q_{ijl})) \quad (21)$$

$$\text{s. t. } (\hat{x}_{u_i}, \hat{y}_{u_i}) \in \text{cov}(es_j) \quad (22)$$

$$s_{ij} = \begin{cases} 1, & u_i \text{ 分配到 } es_j \\ 0, & \text{其它} \end{cases} \quad (23)$$

其中, 服务质量 $q_{ijl}$ 随历史数据 $q_{his_{ijl}}$ 的增大而增大, 随用户与基站距离 $\text{dist}(u_i, es_j)$ 和边缘服务器容量 $\text{capacity}(es_j)$ 的增大而减小, 因此计算方式如下:

$$q_{ijl} = q_{his_{ijl}} + \frac{\alpha}{1 + \text{dist}(u_i, es_j)} + \frac{\beta}{1 + \text{capacity}(es_j)} \quad (24)$$

为了消除用户数据的差异, 同时避免数据发散, 使用S型函数对历史的服务质量数据进行映射:

$$q_{his_{ijl}} = \frac{1}{1 + e^{-\sum_{t=0}^T q_{ijl}}} \quad (25)$$

其中T表示优化的当前时间段, 将历史服务质量进行归一化, 便于对服务质量进行比较.

利用人工蜂群算法对服务质量进行优化, 在确定参数进行初始化后, 人工蜂群算法重复迭代步骤(2)-(4)寻找最优解:

(1) 初始化蜜源. 对人工蜂群算法的参数进行初始化, 包含食物源数量 $S_N$ , 即采蜜蜂和观察蜂的数量, 优化迭代次数 $I$ 以及最大无效次数 $Invalid$ , 蜜源由如下公式产生:

$$S = (s_1^T, s_2^T, \dots, s_m^T) \quad (26)$$

$$s_i = (s_{i0}, s_{i1}, \dots, s_{in}) \quad (27)$$

$$s_{ij} = 1 \text{ iff } \text{rand}(j): es_j \in A \quad (28)$$

$$\& \exists \text{ one and only } s_{ij} = 1 \quad (29)$$

其中 $s_i$ 代表蜜源 $S$ 的第 $i$ 维度向量, 即参与分配的用户 $u_i$ 和区域内 $n$ 各边缘服务器对应的分配矩阵, 在 $u_i$ 的备选边缘服务器集合 $A$ 中随机挑选一个边缘服务器进行分配, 且仅分配到该服务器. 初始化蜜源就是对每个蜜源的所有维度通过以上公式生成一个范围内的随机向量, 从而生成 $S_N$ 各初始蜜源.

(2) 采蜜蜂阶段. 在采蜜蜂阶段更新蜜源:

$$S_i = (\dots, s_{ij}^T, \dots) S_k = (\dots, s_{kj}^T, \dots) \quad (30)$$

$$S_i' = (\dots, s_{ij}^T, \dots, \varphi(s_{kj}^T), s_{ij}^T, \dots) \quad (31)$$

其中 $S_k$ 表示邻域蜜源,  $k \in \{1, \dots, S_N\}$ 且 $k \neq i$ ,  $\varphi(s_{kj}^T)$ 表示邻域蜜源 $S_k$ 的第 $j$ 维向量的一个随机转变, 即在可行解中随机选取第 $k$ 个用户重新分配边缘服务器. 得到新蜜源后, 比较新旧蜜源的适应度值, 利用贪婪算法选择优者.

(3) 观察蜂阶段. 观察蜂阶段在跟随蜂阶段结束后开始, 采蜜蜂采蜜回巢后在舞蹈区分享蜜源信息, 即各可行解的适应度值. 观察蜂分析各蜜源的适应度值, 计算各蜜源的相对分数:

$$\text{accuFitness} = \frac{0.9 * Q}{\max(Q)} + 0.1 \quad (32)$$

随机从相对分数大于随机阈值的蜜源中选择跟随. 在开采过程中, 利用式(32)更新蜜源, 择优留下.

在更新蜜源的过程中, 蜜源包含参数 $\text{invalidcount}$ , 记录蜜源没有被更新的次数.

(4) 侦察蜂阶段. 如果一个蜜源经过多次开采没有被更新, 也就是在经过迭代后并没有搜索到更优分配,  $\text{invalidcount}$ 超过了设定的阈值, 则需抛弃这个蜜源, 启动侦察蜂阶段. 这体现了人工蜂群算法的自组织的负反馈和波动属性<sup>[29]</sup>. 在该阶段里, 侦察蜂重新寻找新的蜜源来代替被抛弃蜜源.

## 5 实验

### 5.1 数据集描述

本文实验数据主要来自两个部分的数据集, 第一部分数据2<sup>[32][33]</sup>来源于上海电信数据集的基站分布和用户接入信息, 第二部分数据3<sup>[34]</sup>来源于开源数据集中不同边缘服务的 QoS 数据. 上海电信数据集主要包括基站分布和用户接入基站的情况, 利用上海电信数据集获取边缘服务器的位置信息以及

2 <http://sguangwang.com/TelecomDataset.html>

3 <https://wsdream.github.io>

用户接入信息. 不同边缘服务的 QoS 数据包括响应时间 RT 和吞吐量 TP 两个属性, 利用响应时间属性计算用户获得的 QoS 值来衡量优化效果.

其中第一部分数据来自上海电信系统中 30 天内 9481 名用户移动设备访问 3233 个基站生成的 720 多万份记录, 上海电信数据集的区域划分情况如图 4 所示, 在 1:2km 的地图中对各区域的基站点进行了标注. 不同边缘服务器及其接入用户信息如表 2 所示; 第二部分数据是在 16 小时内 64 个不同时间段上来自 4500 个边缘服务上的 142 个用户的真实 QoS 评估结果, 传统环境下, 时间感知的边缘服务 QoS 数据集信息如表 3 所示. 本实验所用的边

缘 QoS 数据集主要分为服务调用数据和用户接入数据两个部分, 都以矩阵的形式描述. 本实验基于真实的边缘移动设备接入基站的信息模拟了 142 个用户在 274 个基站区域内调用 4500 个边缘服务的 QoS 数据, 通过使用上海电信数据集提取真实的用户接入信息从而反映出边缘网络的移动性, 再将传统的 QoS 数据提取出来模拟分配到所有用户中, 形成边缘 QoS 数据集, 在考虑移动性的同时模拟了更真实的 QoS 数据. 我们从上海电信数据集中随机选取 16 小时内的数据, 选取 142 个不同 ID 的用户移动设备的基站接入记录进行实验.

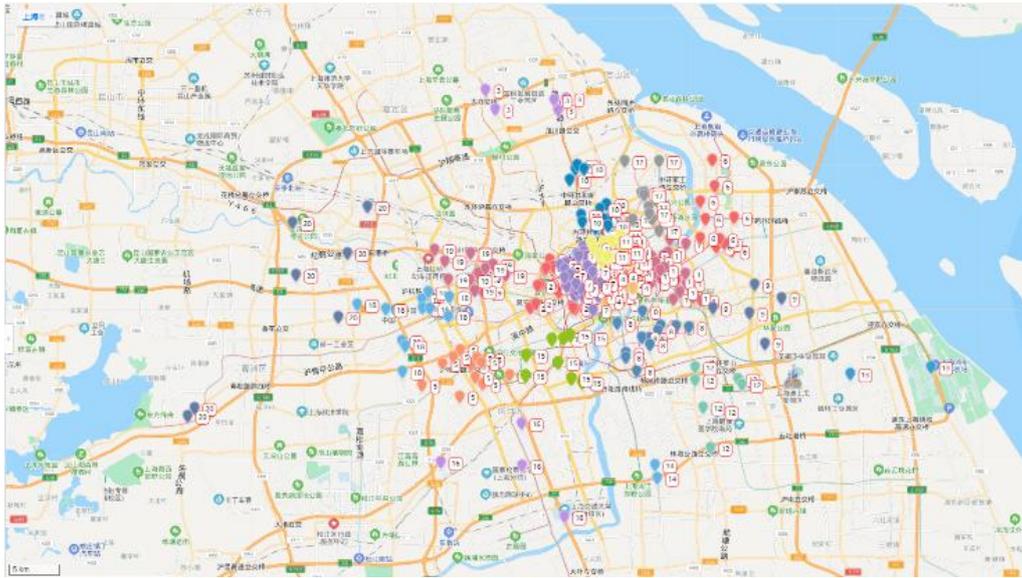


图 4 AP 聚类区域划分情况

表 2 上海电信数据集

区域 ID	基站 ID	接入时间	基站位置	用户 ID
5	68	22:19-22:42	31.240874/121.518086	1
1	1	17:17-17:18	31.237872/121.470259	3

表 3 传统环境时间感知的边缘服务 QoS 数据集

用户 ID	时间 ID	服务 ID	响应时间	吞吐量
1	41	0	0.721	0.5034674
25	54	1012	0.309	4.579288

## 5.2 实验设置

实验在具有以下配置的 PC 上运行: Intel(R) Core(TM) i5-1035G1 CPU @ 1.00GHz 1.19GHz, 16 GB RAM 和 Windows 10 OS. 使用 MySQL8.0 进行数据处理, 在 Python3.6 PyCharm Professional2019.2 中进行开发. 实验各参数如表 4 所示, 研究数据集中用户数量为 142, 边缘服务器数量为 274, 数据集中边缘服务数量为 4500, 构成边缘 QoS 数据集. 在模型

训练中, 用户和边缘区域的特征维度设置为 2, 模型训练次数设置为 10000 次. 在路径预测模型中, 设置插值步长为 500. 适应度值计算参数中距离和资源的比重参数分别设置为 0.5 和 0.2. 改进的人工蜂群算法中解的最大无效次数设置为 1, 设置不同迭代次数(0,1,...,20)和蜂群规模(2,4,...,14). 参与实验的各个聚类区域的特征如表 5 所示, 共选取 4 个规模不同的区域进行实验, 实验基站分布如图 5 所示. 图

5 标记了实验的四个区域, 部署的基站数量分别为 44、16、25 和 15 个, 主要分布在上海市静安区和虹口区, 以 1:2km 的比例尺标记了四个区域的基站部署情况。

表 4 实验参数

参数名称	参数意义	参数设置
$n$	边缘服务器数量	274
$m$	区域内用户数量	142
$k$	边缘服务数量	4500
$p$	用户特征维度	2
userIterCount	用户模型训练次数	10000
step	插值步长	500
$\alpha$ 、 $\beta$	QoS 计算参数	0.5、0.2
invalidcount	人工蜂群最大无效次数	1

表 5 各区域特征

区域编号	用户数量	基站数量	记录数量
1	49	44	2,632,619
4	12	16	1,804,082
8	24	25	2,255,050
10	10	15	562,891

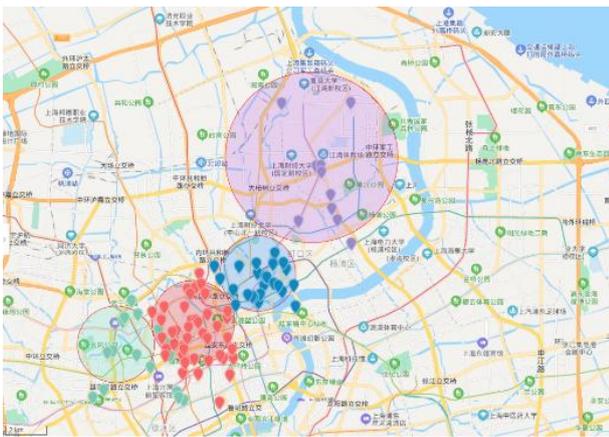


图 5 实验区域基站分布情况

### 5.3 评估参数

#### (1) 适应度值

实验通过适应度值分析实验, 更直观地比较不同方法的优化结果. 适应度值表示这些策略的整体有效性. 计算方式如下:

$$fitness = \sum_{u_i \in U, es_j \in ES, ws_l \in WS} q_{ijl} \quad (33)$$

#### (2) 优化量

为了更直观的反映优化方法产生的最优解相对于所有可行解的适应度值的优势, 引入优化量这一指标, 计算最优方案适应度值和其余方案的平均适

应度值的差值作为优化量. 即最优解适应度值和平局适应度值的差值, 以反映该方法的优化量:

$$\delta = fitness_{opt} - fitness_{avg} \quad (34)$$

### 5.4 实验过程

本文希望通过实验证明移动边缘环境下基于联邦学习的动态 QoS 优化方法通过优化边缘服务器缓存, 考虑用户移动性在动态环境下为用户分配边缘服务器, 通过比较不同种群规模和迭代次数的优化效果, 探索优化方法的最佳参数, 在保护用户隐私的同时实现区域服务质量的优化. 通过对比实验阐述本方法的性能.

通过数据预处理得到上海电信数据集和 QoS 属性数据集后, 提取各个区域数据, 训练用户-边缘服务器接入信息还原用户移动路径、确定备选基站, 训练用户-服务调用信息确定缓存, 比较区域缓存列表对优化结果的影响; 调整人工蜂群优化的种群规模和迭代次数, 与基线方法和其它优化算法比较, 本方法的优化效果、隐私保护方面效果更好. 实验步骤简单描述如下:

(1) 在同一边缘区域和蜂群规模下, 通过实验探究人工蜂群算法的迭代次数对优化效果的影响.

(2) 在同一边缘区域内基于相同的迭代次数, 探索优化效果随蜂群规模大小的变化情况.

(3) 探索边缘区域规模对算法优化效果的影响. 通过实验比较在相同的蜂群规模和迭代次数下, 边缘区域的规模是否和优化效果存在线性关系, 以说明本方法的适用性.

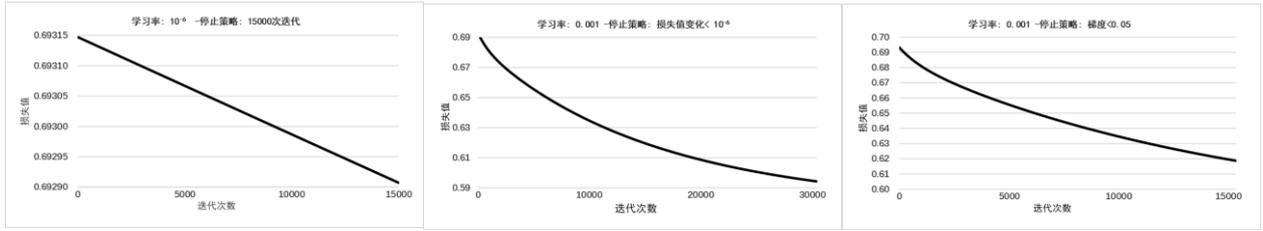
(4) 比较不同方法的优化效果. 通过比较相同种群规模和边缘区域规模下不同优化方法的优化效果, 以及随着迭代次数的增加, 各方法优化效果变化趋势的比较, 确定本方法是否优于现有的优化方法.

### 5.5 结果分析

#### 5.5.1 模型训练

用户偏好模型使用不同停止策略的训练效果如图 6 所示. 以迭代次数为停止策略损失值较高且呈下降趋势; 以损失值为停止策略设置阈值为  $10^{-6}$ , 需训练 30000 余次, 在 20000 次处损失值趋于平缓; 以梯度变换为停止策略设定阈值为 0.05, 需训练 15000 余次, 在 10000 次处损失值趋于平缓. 故选择以梯度变换为用户模型训练的停止策略.

用户轨迹在不同插值步长时的预测效果如图 7, 步长为 500 时既能保证预测准确性, 又能还原用户移动的复杂性, 故选择步长为 500.



(a)以迭代次数为停止策略的训练情况(b)以损失值为停止策略的训练情况(c)为梯度变化停止策略的训练情况  
图 6 不同停止策略训练情况

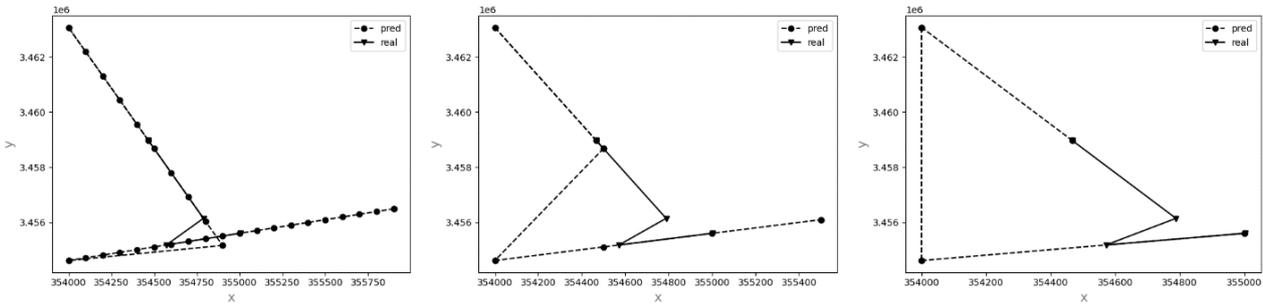


图 7(a)step=100

图 7(b) step=500

图 7(c) step=1000

图 7 不同步长用户轨迹预测情况

### 5.5.2 优化效果

#### 5.5.2.1 迭代次数

不同迭代次数下最优解适应度值和优化量实验结果图 8 所示. 随着迭代次数的增加, 图 8(a)、8(b) 所示实验在区域规模 (即区域内用户数量) 为 14 的区域中进行, 可以明显看出最优解的适应度值随着迭代次数的增加而增加, 且在迭代到一定次数时趋于平缓, 说明在迭代到一定次数时该方法能够找到

最优解; 而优化量在随着迭代次数的增加波动下降, 说明迭代到一定次数后, 产生的解的适应度值也有所提升, 最优解的适应度值的优势减弱. 本方法在第 2 次和第 7 次迭代时找到了更优解, 且第 7 次找到的最优解提高了相对于平均适应度值的优化量, 且在第 7 次迭代后优化量持续下降, 说明在第 7 次迭代时找到了最优解.

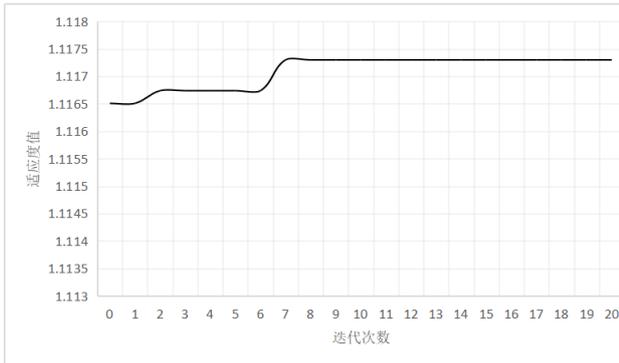


图 8(a)迭代次数对适应度值的影响

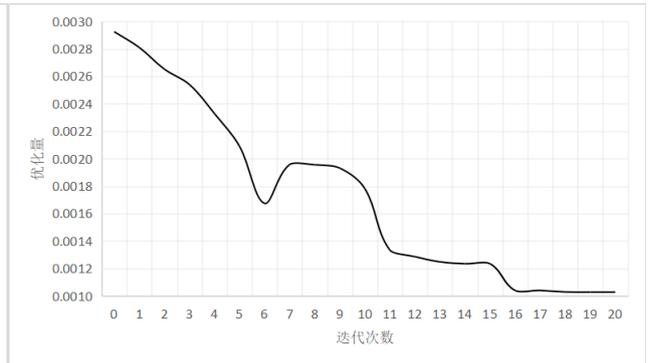


图 8(b)迭代次数对优化量的影响

图 8 迭代次数对优化效果的影响

#### 5.5.2.2 蜂群规模

随着人工蜂群算法蜂群规模的增加, 如图 9(a)、9(b), 为区域规模为 8、迭代次数为 8 次时, 不同蜂群规模下本方法的优化效果. 可以观察得出, 不同蜂群规模下最优解的适应度值在波动上升, 整体略呈上升趋势, 但整体的浮动范围不大, 反映出随着蜂群规模的增大, 本方法获得的最优解的适应度值在

小范围内呈上升趋势, 考虑到扩大蜂群规模会造成优化成本的提高, 因此不建议通过扩大蜂群规模来提高优化效果; 而优化量随蜂群规模的增加而增加, 说明蜂群规模越大, 最优解适应度值和平均适应度值的差值越大, 这可能由于大规模的蜂群产生的解的数量更多, 造成了更多非优解的产生, 从而提高了优化量. 由于本方法优化的目的在于寻找适

应度值更高的可行分配,非优解的产生会导致优化成本的增加,而不会对最优解的适应度值产生影响,因此从优化量的角度,通过扩大蜂群规模来提高优化效果也不合适。

从图 8 和图 9 可以看出,在迭代次数更多或种群规模更大,即算法的时间和规模消耗增加时,优

化效果是在增强的.但优化效果的增强是以时空资源的消耗为代价,因为降低时空消耗和提升算法性能是一对矛盾.在时空资源允许的条件下,可以尽可能提升算法性能,在时空资源紧张的情况下,智能在资源允许范围内尽可能提升算法的效果。

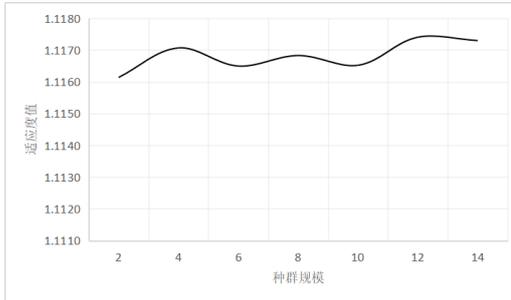


图 9(a)蜂群规模对适应度值的影响

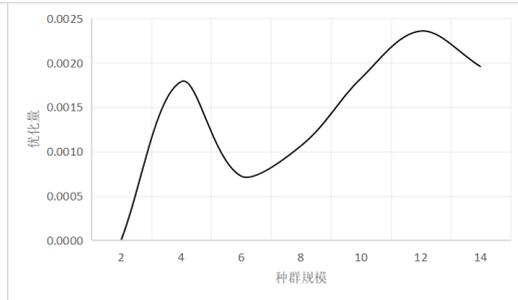


图 9(b)蜂群规模对优化量的影响

图 9 蜂群规模对优化效果的影响

### 5.5.2.3 边缘区域规模

随着边缘区域规模的增加,如图 10(a)、10(b)所示,边缘区域规模的大小对该区域的 QoS 优化没有明显的规律,但适应度值和优化量的变化都在同一个区域规模处取得峰值和谷值,故适应度值和优化量在边缘区域规模变化时具有相同的变化趋势,即

在区域的较大适应度值的最优解较大时也能取得较大的优化量;而边缘区域的规模对优化效果无明显影响,从公式(33)也可看出优化效果与区域规模  $m$  非线性或非线性关系,即本优化方法适用于不同的边缘区域。

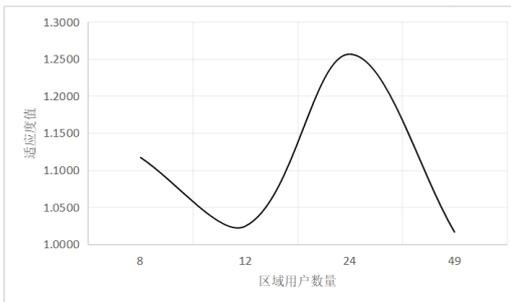


图 10(a)边缘区域规模对适应度值的影响

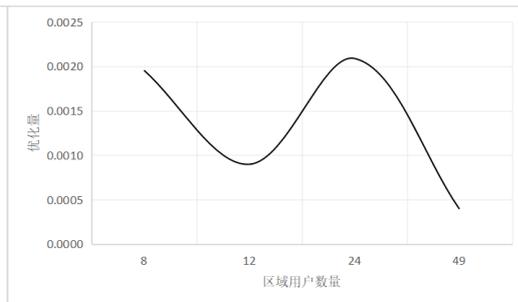


图 10(b)边缘区域规模对优化量的影响

图 10 边缘区域规模对优化效果的影响

### 5.5.3 不同方法的优化效果

在实验模块,我们主要是验证 MECFLD\_QoS 方法的有效性和优化效果.在优化效果方面,我们比较了四种方法:(1)基线方法(Random):随机为用户分配边缘服务器;(2)无缓存的移动感知动态 QoS 优化方法(MECMA\_QoS):不进行缓存优化,只考虑用户的移动性利用人工蜂群算法为用户分配边缘服务器;(3)基于遗传算法的 QoS 优化(GA\_QoS):在考虑区域服务缓存优化的基础上考虑用户的移动性利用遗传算法为用户分配边缘服务器;(4)本文提出的方法基于联邦学习的动态 QoS 优化方法(MECFLD\_QoS):在考虑区域服务缓存优化的基础

上考虑用户的移动性利用人工蜂群算法为用户分配边缘服务器。

图 11 为区域规模为 8 时本方法和其它对比方法的适应度值变化.由于优化工作的目的时找到更优的用户分配,所以对四种方法的最优解适应度值进行比较.实验结果表明,随机分配(Random)方法获得的分配结果在较低范围浮动;基于遗传算法的 QoS 优化(GA\_QoS)方法在第二次迭代取得一个较优解后没有获得更优解,反映了遗传算法在搜索时容易陷入局部最优解;无缓存的移动感知动态 QoS 优化(MECMA\_QoS)方法在第二次迭代时取得了较高的用户分配,但与本文提出的

MECFLD\_QoS 方法相比, 区域的整体 QoS 较低, 由此可以分析得出 MECFLD\_QoS 方法能够取得更好的优化效果, 且不易陷入局部最优解。

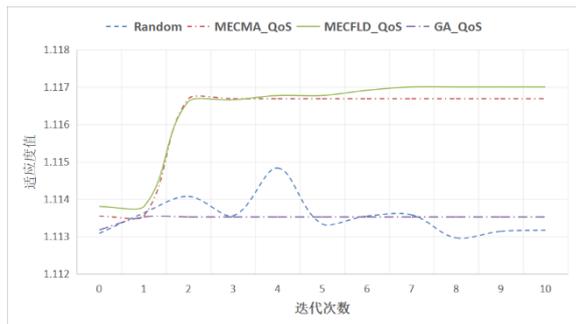


图 11 不同方法的优化效果比较

## 6 结论与未来工作

现有的 QoS 优化方法没有从移动边缘网络的角度出发考虑边缘用户的分配, 没有同时考虑到用户特征隐私和位置隐私的保护, 并且研究场景较为单一. 针对上述问题, 本文提出了一种移动边缘环境下基于联邦学习的动态服务质量优化方法. 通过优化边缘服务缓存和用户移动场景, 有效地保护了用户的特征隐私和位置隐私; 提出改进的人工蜂群算法, 基于二维解对边缘网络中的用户分配问题进行优化. 实验证明边缘缓存对区域 QoS 起到优化效果, 改进的人工蜂群算法针对其多变量多峰值的特点有效地优化了用户分配. 综上所述, 本文提出的 MECFLD\_QoS 方法在保证优化效果的同时, 在隐私保护方面也取得很好的效果.

在未来工作中, 将重点对以下几个问题进行研究: (1) 当前只考虑响应时间的因素对区域 QoS 进行优化, 未来考虑多 QoS 属性值进行 QoS 优化; (2) 相比于本方法只在聚类区域内进行局部优化, 未来还需考虑研究区域内的 QoS 和负载均衡, 进行全局优化; (3) 考虑时间因素, 在真实的 5G 网络移动边缘场景中动态更新用户模型, 实现动态优化.

### 参考文献

[1] ZhangPeng-Cheng, Jin Hui-Ying. Privacy protection QoS forecasting in mobile edge environment. *Chinese Journal of Computers*, 2020, 43(8):1555-1571(in Chinese)  
(张鹏程, 金惠颖. 一种移动边缘环境下向隐私保护 QoS 预测方法. *计算机学报*, 2020, 43(8):1555-1571)

[2] Shi Wei-Song, Sun Hui, Cao Jie, Zhang Quan, Liu Wei. Edge

computing:an emerging computing model for the internet of everything era. *Journal of Computer Research and Development*, 2017,54(5):907-924 (in Chinese)

(施巍松, 孙辉, 曹杰, 张权, 刘伟. 边缘计算:万物互联时代新型计算模型. *计算机研究与发展*, 2017, 54(5): 907-924)

- [3] Shen H, Bai G, Hu Y, et al. P2TA: Privacy-preserving task allocation for edge computing enhanced mobile crowdsensing. *Journal of Systems Architecture*, 2019, 97: 130-141
- [4] Wang Z, Zhao Z, Min G, et al. User mobility aware task assignment for mobile edge computing. *Future Generation Computer Systems*, 2018, 85(8):1-8
- [5] Wang S, Tuor T, Salonidis T, et al. Adaptive federated learning in resource constrained edge computing systems. *IEEE Journal on Selected Areas in Communications*, 2018, 37(6): 1205-1221
- [6] Miao Y, Wu G, Li M, et al. Intelligent task prediction and computation offloading based on mobile-edge cloud computing. *Future Generation Computer Systems*, 2020, 102:925-931
- [7] Xu J, Li X, Liu X, et al. Mobility-aware workflow offloading and scheduling strategy for mobile edge computing//*Proceedings of the International Conference on Algorithms and Architectures for Parallel Processing*. Melbourne,Australia, 2019: 184-199
- [8] Deng S, Xiang Z, Taheri J, et al. Optimal application deployment in resource constrained distributed edges. *IEEE Transactions on Mobile Computing*, 2020, PP(99): 1-1
- [9] Qian Y, Hu L, Chen J, et al. Privacy-aware service placement for mobile edge computing via federated learning. *Information Sciences*, 2019, 505:562-570
- [10] Wu C, Peng Q, Xia Y, et al. Mobility-aware tasks offloading in mobile edge computing environment//*Proceedings of the 2019 Seventh International Symposium on Computing and Networking*. Nagasaki, Japan, 2019: 204-210
- [11] Zhang Kai-Yuan, Gui Xiao-Lin, Ren De-Wang, Li Jing, Wu Jie, Ren Dong-Sheng. Survey on computation offloading and content caching in mobile edge networks. *Journal of Software*, 2019, 30(8): 2491-2516 (in Chinese)  
(张开元, 桂小林, 任德旺, 李敬, 吴杰, 任东胜. 移动边缘网络中计算迁移与内容缓存研究综述. *软件学报*, 2019, 30(8):2491-2516)
- [12] Lim W Y B, Luong N C, Hoang D T, et al. Federated learning in mobile edge networks: a comprehensive survey. *IEEE Communications Surveys and Tutorials*, 2020, 22(3):2031-2063
- [13] Bayardo RJ, Agrawal R. Data privacy through optimal k-anonymization//*Proceedings of the 21st International Conference on Data Engineering*. Tokyo, Japan, 2005: 217-228
- [14] Lai P, He Q, Cui G, et al. Edge user allocation with dynamic quality

- of service//Proceedings of theInternational Conference on Service-Oriented Computing. Toulouse, France, 2019: 86-101
- [15] Lai P, He Q, Abdelrazek M, et al. Optimal edge user allocation in edge computing with variable sized vector bin packing//Proceedings of the16th International Conference on Service Oriented Computing. Hangzhou, China, 2018: 230-245
- [16] He Q, Cui G, Zhang X, et al. A game-theoretical approach for user allocation in edge computing environment. *IEEE Transactions on Parallel and Distributed Systems*, 2020, 31(3): 515-529
- [17] Peng Q, Xia Y, Feng Z, et al. Mobility-aware and migration-enabled online edge user allocation in mobile edge computing//Proceedings of the2019 IEEE International Conference on Web Services. Milan, Italy, 2019: 91-98
- [18] Liang Y, Ge J, Zhang S, et al. Interaction-oriented service entity placement in edge computing. *IEEE Transactions on Mobile Computing*, 2019, PP(99):1
- [19] You K, Tang B, Qian Z, et al. QoS-aware placement of stream processing service. *Journal of Supercomputing*, 2013, 64(3): 1-23
- [20] Shi W, Cao J, Zhang Q, et al. Edge computing: vision and challenges. *IEEE Internet of Things Journal*, 2016, 3(5):637-646
- [21] McMahan H B, Moore E, Ramage D, et al. Federated learning of deep networks using model averaging.arXiv preprint arXiv:1602.05629, 2016
- [22] Li B, Zhang H, Lu H. User mobility prediction based on Lagrange's interpolation in ultra-dense networks//Proceedings of the2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications. Valencia, Spain, 2016: 1-6
- [23] Thomas, Sauer, Yuan, et al. On multivariate Lagrange interpolation. *Mathematics of Computation*, 1995, 64(211): 1147-1170
- [24] Wang X, Wang Z, Xu X. An improved artificial bee colony approach to QoS-aware service selection//Proceedings of the2013 IEEE 20th International Conference on Web Services. Santa Clara, USA, 2013: 395-402
- [25] Armbrust M, Fox A, Griffith R, et al. A view of cloud computing. *Communications of the ACM*, 2010, 53(4): 50-58
- [26] Grafarend E. Geodetic Theory Today: The optimal universal transverse mercator projection. Berlin , Germany: International Association of Geodesy Symposia, 1995
- [27] Xiao Y, Yu J. Semi-supervised clustering based on affinity propagation algorithm. *Journal of Software*, 2008, 19(11): 2803-2813
- [28] Kooperberg C, Ruczinski I, Leblanc M L, et al. Sequence analysis using logic regression. *Genet Epidemiol.* 2001, 21 ( Suppl 1) :S626-31
- [29] Smith A E. Swarm intelligence: from natural to artificial systems. *IEEE Transactions on Evolutionary Computation*, 2000, 4(2):192-193
- [30] Bejinariu S I, Costin H, Rotaru F, et al. Performance analysis of artificial bee colony optimization algorithm//Proceedings of the2017 International Symposium on Signals, Circuits and Systems.Iasi, Romania, 2017: 1-4
- [31] Karaboga D, Akay B. A comparative study of artificial bee colony algorithm. *Applied Mathematics and Computation*, 2009, 214(1):108-132
- [32] Wang S, Zhao Y, Huang L, et al. QoS prediction for service recommendations in mobile edge computing. *Journal of Parallel and Distributed Computing*, 2017, 127(5):134-144
- [33] Wang S, Zhao Y , Xu J, et al. Edge server placement in mobile edge computing. *Journal of Parallel and Distributed Computing*, 2018, 127(5):160-168
- [34] Zheng Z, Zhang Y, Lyu M R. Investigating QoS of Real-World Web Services. *IEEE Transactions on Services Computing*, 2014, 7(1):32-39



**ZHANG Peng-Cheng**, Ph.D., professor.

His research interests include software engineering, service computing and data mining.

**WEI Xin-Miao**, masterstudent. Her research interests include service computing and data mining.

**JIN Hui-Ying**, Ph.D. candidate. Her research interests include service computing and data mining.

## Background

As the government report accelerating the deployment of new infrastructure such as 5G networks and the Internet of Things. 5G technology has also developed into the country's future strategic development direction. Low-latency, localized services sink to the edge of the network, which solves the problem of large overall network traffic and time extension, and provides better support for bandwidth-intensive and real-time services in 5G networks.

Existing optimization methods carry out research on service quality optimization in terms of computing offloading, user allocation and task scheduling. By considering edge node resource consumption constraints, service business logic constraints, and user mobility constraints, the task scheduling strategy is formed, mainly optimizing task acceptance rates, optimization strategy costs, optimization of computational offloading, and optimization of resource allocation. However, the existing work has not considered the protection of user characteristics and location privacy at the same time.

This paper proposes a dynamic QoS optimization method in mobile edge environment based on federated learning (MECFLD\_QoS). Based on the essence of federated learning, uses gradient descent to solve the logistic regression model, trains the regional public model to optimize the Web service cache of the edge area, and allocates edge servers according to the user location in the dynamic mobile scenario, effectively protecting user privacy. The regional service quality is optimized, and it has better adaptability to dynamic user

mobility scenarios. MECFLD\_QoS mainly optimized the following aspects: (1) Optimize the traditional QoS data set, mapped the data set to the edge network environment, and fully considered the mobility, distributed, real-time, and complex scenarios features of edge computing, forming an edge QoS feature data set; (2) Optimize the edge server cache, train the user preference model on the user terminals, only transmit model parameters when interacting with the regional public model, and encapsulate the user's private data in the user terminal to avoid data transmission, which can effectively protect user privacy; (3) Optimize user mobile scenarios, collect user mobile information in dynamic mobile scenarios, and use the geographic location of the user to access the base station to fit the user's movement trajectory for prediction, which effectively blurs the user's reality location, while predicting the trajectory, effectively protects the users' location privacy; (4) Optimize the user allocation method, and proposed an improved two-dimensional solution-based artificial bee colony algorithm(ABC) to optimize the user allocation problem in the edge network, Facts have proved that the improved artificial bee colony algorithm effectively optimizes user allocation according to its multi-variable and multi-peak characteristics, and achieves a better allocation effect.

This research is supported by the National Natural Science Foundation of China (61572171,61702159), the Natural Science Foundation of Jiangsu Province (BK20191297) and the Fundamental Research Funds for the Central Universities of China (B210202075).