区块链和联邦学习赋能下的众包任务个性化安全推荐方案

邬海琴[1]，李梁[1]，李昌峰[2]
（1.华东师范大学软件工程学院，上海 200062；2.中国科学院计算机网络信息中心，北京 100190）

摘要：任务推荐是众包关键服务之一，但却面临单点故障和严峻的安全隐私威胁。现有隐私保护的众包任务推荐算法在数据安全和推荐质量之间难以调和，且未兼顾工作者对任务的个性化偏好需求。针对此，以区块链和联邦学习赋能众包服务，提出了个性化和安全的众包任务推荐框架PBFRec。以联盟链代替传统中心众包平台，工作者基于图神经网络训练本地偏好模型，采用差分隐私实现隐私保护的全局模型聚合更新。此外，提出了基于数据质量的两层模型聚合方法和基于社交关系的模型初始化方法，解决了恶意用户提交低质量模型参数和系统冷启动难题。安全性分析表明，PBFRec满足 $T\varepsilon$-本地差分隐私、任务推荐模型训练完整性和可用性；实验基于Hyperledger Fabric搭建原型系统，结果表明，PBFRec的任务推荐质量和系统性能优于现有方案，召回率最高为现有方案的2.9倍，且社交关系对解决冷启动问题具有积极贡献。

关键词：众包；联盟链；联邦学习；任务推荐；图神经网络
中图分类号：TP391
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260255

Personalized and secure crowdsourcing task recommendation scheme empowered by blockchain and federated learning

Wu Haiqin[1], Li Liang[1], Li Changfeng[2]
1. Software Engineering Institute, East China Normal University, Shanghai 200062, China
2. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China

Abstract: Task recommendation is a key service in crowdsourcing, which suffers from single points of failure and severe security and privacy issues. Existing privacy-preserving crowdsourcing task recommendation algorithms struggle to balance between data security and recommendation accuracy, and fail to accommodate workers' personalized preferences for tasks. To address these, crowdsourcing was empowered with blockchain and federated learning, and a personalized and secure crowdsourcing task recommendation framework PBFRec was proposed. Specifically, the traditional centralized crowdsourcing platform was replaced with a consortium blockchain. Workers train local preference models using graph neural networks, and privacy-preserving global model aggregation and updates were achieved via differential privacy. Furthermore, a two-layer model aggregation method was designed based on data quality and a social relation-based model initialization method, which addressed the situation where malicious users submit low-quality model parameters and the cold-start problem. PBFRec satisfies $T\varepsilon$-local differential privacy and ensures the integrity of task recommendation model training. A prototype experiment on Hyperledger Fabric shows that PBFRec outperforms existing solutions in task recommendation quality and system performance. The recall rate is up to 2.9 times as much as that of existing approaches, and social relationships contribute positively to alleviating the cold-start problem.

Key words: crowdsourcing, consortium blockchain, federated learning, task recommendation, graph neural network

收稿日期：2026–05–08；**修回日期：**2026–08–03
通信作者：李昌峰，cfli@cnic.cn
基金项目：国家自然科学基金资助项目（No.62202167）
Foundation Item：The National Natural Science Foundation of China (No.62202167)

0 引言

众包作为一种基于共享经济的分布式协同模式，通过公开招募大众以低成本、高效率共同完成复杂任务，实现了闲置资源的价值最大化，并成功推动了Amazon Mechanical Turk、美团众包等平台的发展，广泛应用于物流、餐饮服务等领域[1-2]。

任务推荐是众包系统中的关键服务，与工作者提交任务偏好至众包平台执行任务–偏好匹配的方式不同，任务推荐可根据工作者历史参与信息和个人偏好为其推荐感兴趣的任务，从而显著提升用户黏性与平台活跃度[3]。当前，相关研究主要关注众包任务匹配或分配中的激励和隐私保护等问题[4-7]，而针对任务推荐的相关研究总体较少。个人偏好模型训练是实现众包任务推荐的核心，传统推荐模型通常依赖中心平台训练[8-11]。其中，DeepChain[11]采用多智能体模型式强化学习联合优化连续场景下的推荐策略，但其主要面向电商场景，且中心化训练除了存在单点失效、数据篡改等安全隐患外[12]，用户还面临隐私泄露风险。

针对以上问题，一方面，现有研究[13-18]采用联邦学习模式，仅将工作者本地训练的模型梯度上传至中心服务器进行全局聚合以提供推荐，部分方案[18]采用差分隐私对梯度进行扰动以防止攻击者由梯度反推出用户数据。然而，以上方案仅基于序列编码[13]或矩阵分解[18]等传统方法训练用户偏好模型，忽略了众包数据中固有的图拓扑结构，且难以建模众包中的复杂关联关系，导致模型准确性不高；此外，其假设“参与者具备充足历史交互数据”，新用户因数据极度稀疏可能引发推荐失效，因此难以应对新用户的冷启动困境。另一方面，一些研究[19-21]基于区块链技术实现去中心化模型聚合，但链上聚合采用简单的均值聚合，未充分考虑不同训练方提交的模型参数质量差异，且直接将模型参数上传区块链面临梯度推理攻击的风险，采用公有链也面临链上开销过大的问题。

针对上述不足，本文提出了一种区块链与联邦学习赋能的去中心化众包个性化任务安全推荐框架（PBFRec）。与现有去中心化方案相比，PBFRec具有以下显著优势。

首先，打破传统硬性匹配方式，引入图神经网络（graph neural network, GNN）用于个性化偏好模型训练。与矩阵分解等传统方法相比，GNN具有强大的关系建模能力，能深度融合工作者历史参与任务中的时间、任务类型与空间轨迹等多维异构信息，更全面地捕捉工作者的个性化偏好，在此基础上结合用户可达距离，实现了主动且全面的个性化任务推荐。其次，依托联盟链与联邦学习实现底层隐私架构革新，与公有链完全公开透明不同，联盟链准入机制隔离了潜在恶意节点，且对于Hyperledger Fabric联盟链，用户数据仅在通道内部可见，一定程度上保护了用户隐私；此外，联邦学习确保用户数据不出本地，结合Hyperledger Fabric的通道特征，提出基于动态质量评估的通道模型与全局模型两层模型聚合机制，有效应对恶意训练用户提交低质量模型参数，并基于差分隐私实现隐私保护的最终模型安全链上聚合。最后，针对GNN也难以解决的冷启动挑战，创新性地利用社交关系为新用户分配初始模型与动态权重来攻克该难题。本文的主要贡献可归纳为以下4个方面。

（1）提出了PBFRec。PBFRec以联盟链代替传统集中式众包平台，各工作者根据本地历史任务参与数据训练偏好模型，基于双扰动的本地差分隐私[22]保护本地模型梯度，实现隐私保护的本地模型聚合和全局模型更新，同时智能合约确保了模型聚合和更新的计算完整性。

（2）设计了基于GNN的时空感知个性化推荐方法。不同于传统静态关键字匹配，利用GNN深度挖掘工作者历史轨迹中任务时间、类型与空间的复杂非线性关系。通过融合工作者实时演进的时空状态，PBFRec能够周期性地生成全面的个性化任务推荐集合。

（3）构建了基于动态质量评估的双层模型聚合机制与冷启动平滑策略。第一层对本地模型质量进行评估，第二层对通道模型质量进行评估，并相应设计了本地模型和通道模型质量动态评估算法，最终将评估结果作为权重用于模型聚合，解决了恶意用户和通道提交低质量模型的问题。与简单均值聚合相比，该策略有效提高了任务推荐质量。同时，创新性地引入社交关系辅助新用户初始化本地模型，在模型聚合时赋予新用户较低权重，有效攻克了系统冷启动难题。

（4）提供了理论安全证明与系统实验验证。理论分析证明，PBFRec满足$T\varepsilon$-本地差分隐私、模型

聚合完整性和可用性。实验基于Hyperledger Fabric搭建原型系统，并采用真实众包数据集，结果表明，PBFRec在任务推荐质量及整体性能方面较优，召回率最高为现有方案的2.9倍，且社交关系对解决冷启动问题具有积极作用。

1　相关工作

1.1　传统任务推荐模型

早期的推荐方案多侧重于简单的绩效评分或单一维度的匹配。例如，文献[23]使用工作者−任务类别的绩效矩阵得到评分，通过评分降序排序实现任务推荐，获得了较高的推荐效率，但未实现个性化的任务推荐。文献[24]利用同态加密等密码学技术实现了双向隐私保护的任务匹配，但只考虑了基于位置的匹配。宋甫元等[25]提出了基于位置和关键词的匹配，龙浩等[26]提出了基于社区的任务分发方法，通过行为模式划分工作者并在社区层面推荐任务，提升了匹配效率，但仍属于被动匹配，难以实现主动推荐。为此，后续研究[8, 27-28]将工作者历史任务轨迹中的任务类型、距离等加入机器学习模型的训练中，实现了更加个性化和精确的推荐。此外，Zhao等[11]提出DeepChain，通过共享用户行为记忆的多智能体模型强化学习联合优化多个连续场景的推荐策略，但其主要面向电商推荐，未考虑众包任务的时空约束和隐私保护。Tian等[29]构建了面向众包任务推荐的协同知识图谱，并结合图卷积与双行为偏好学习，在提升推荐精度的同时缓解了数据稀疏和冷启动问题，但其冷启动对象主要是缺少历史交互的新任务而非工作者。此外，上述方案仍依赖于集中式平台，未能同时解决平台信任、数据隐私和训练过程可验证性问题。

为了在学习偏好的同时保护隐私，国内外学者引入了差分隐私[30-31]及联邦学习[13-17]。其中，李洪涛等[30]提出了基于差分隐私的连续位置保护机制，有效抵御了轨迹推断攻击。针对移动群智感知场景，杨明川等[31]利用本地差分隐私（local differential privacy，LDP）对工作者位置等敏感信息加入随机噪声，随后由中央服务器训练获得兼顾平台与工作者收益的最优分配策略。同时，针对加噪易引发恶意节点伪造位置的问题，田有亮等[32]在保护用户位置隐私的前提下，通过节点间的相互协作完成了位置真实性认证。然而，对原始数据加噪会严重破坏数据的可用性，导致分配模型精度大幅下降。另一方面，国内学者针对众包场景下的诸多衍生问题给出了解决方案。文献[14]通过边缘参与方的协同推理有效避免了云端数据泄露。徐天承等[15]利用深度时空位置预测网络动态预判工作者与任务的未来分布，实现了空间众包的智能分配。考虑到众包设备异构与本地数据质量差异，张宇等[16]设计了基于联邦学习的参与者动态选择机制。此外，针对未知任务流行度与在线决策挑战，Xu等[17]提出基于多臂老虎机的隐私保护空间众包推送机制。PPTRecS-FL[33]将差异化梯度噪声、个性化隐私预算与分层区域推荐相结合，在联邦学习框架下权衡了隐私与任务推荐效率。然而，其仍采用边云协同架构，未同时考虑模型聚合完整性、基于图结构的时空偏好学习以及社交关系辅助的新用户冷启动问题。

总体来说，现有方案高度依赖中心化服务器，面临单点故障与网络攻击风险。部分联邦架构[13]普遍强假设节点已具备充足的本地数据，这不仅无法解决新用户的冷启动难题，且极易因强行聚合无历史记录节点的低质量梯度，严重破坏全局模型性能。

1.2　基于区块链的任务匹配/推荐模型

近年来，为消除集中式众包平台面临的单点故障和信任危机，诸多研究将区块链技术引入众包系统。文献[6, 34-36]直接使用区块链代替传统众包平台，构建了完全去中心化的众包框架。为了将多个独立的众包服务联合起来，文献[7, 37-38]基于公有链和代理可搜索加密等技术构建了隐私保护的众包任务匹配系统。然而，公有链上大量的计算导致系统性能严重下降，为了缓解高昂的链上开销，文献[4-5]采用了混合存储区块链架构，将复杂计算和大数据存储转移至链下，链上仅保留必要的验证信息。文献[22]提出了一种双扰动的LDP，要求用户在任务信息上链前加入随机噪声以隐匿真实轨迹。然而，以上方案或无法根据工作者的个性化偏好提供智能化主动推荐，或因为LDP直接对原始数据注入噪声而降低了数据可用性。近期研究开始探索在众包场景中结合区块链及人工智能（如行为预测[19]、模型众包训练[39]与数据质量评估[40]等）。为实现主动任务推荐，Kadadha等[19]提出一种针对众包任务推荐的链上行为预测模型，采用“链下训练、链上

推理”的模式，保证了任务分配执行过程的透明性与不可篡改性，但该模型训练依然依赖于平台。Lin等[20]提出了一种区块链与深度强化学习融合的空间众包任务分配方案，该方案在物理层面上仍高度依赖软件定义网络的中心化边缘与云控制器，未彻底摆脱中心化信任假设。针对此，Ma等[21]进一步将联邦学习引入区块链，智能合约取代中心化聚合服务器，对工作者的本地模型梯度进行安全聚合。但是，该方案要求工作者两两之间预先协商一个掩码，用户端开销较大，不适用于资源受限的众包场景。

综上所述，现有传统推荐框架均不可避免地依赖中心化第三方服务器，新兴基于区块链的推荐模型虽在一定程度上打破了中心化信任壁垒，但依然面临严峻的技术瓶颈：复杂的密码学操作给资源受限的终端设备带来沉重开销；在链上聚合过程中，依然面临严重的梯度泄露风险。此外，以上架构普遍建立在“参与节点已具备充足本地交互数据”的理想假设之上，难以应对新用户的冷启动问题。因此，在去中心化众包场景下，亟须一种既能实现轻量级隐私保护聚合，又能有效甄别劣质数据并克服冷启动难题的高质量主动个性化推荐框架。

2 前置知识

2.1 联邦学习

联邦学习（federated learning，FL）是一种面向隐私保护的分布式机器学习新范式，其核心设计理念为“数据不动模型动”与“数据可用不可见”。在标准联邦学习框架中，各参与节点不需要将本地原始数据上传至云端，而是直接在本地设备上利用私有数据对下发的全局模型进行局部迭代训练。

各节点本地训练完成后，仅将本地模型参数（或梯度）上传至聚合服务器。服务器对收集到的全网参数进行聚合，更新全局模型并再次下发，如此循环，直至模型收敛。该机制从物理层面切断了原始数据的泄露途径，保护了用户隐私数据。

2.2 差分隐私

差分隐私（differential privacy，DP）是一个严格的数学框架，旨在给定相关性的情况下将随机噪声注入敏感用户数据来实现隐私保护。

差分隐私技术体系可依据噪声注入位置的不同分为两类核心范式：中心差分隐私（centralized differential privacy，CDP）和本地差分隐私（local differential privacy，LDP）。在CDP中，用户将自己的真实数据上传给聚合服务器，由服务器采用差分隐私算法对数据加噪，向查询用户提供隐私保护查询服务，但CDP假设聚合服务器是可信的。在LDP中，用户可在本地对数据加噪，再上传聚合服务器，由于服务器无法看到原始数据，因此不需要聚合服务器可信。相比CDP，LDP由于不需要信任第三方平台，现已成为主流的差分隐私机制。LDP的形式化定义如下。

定义1 ε-本地差分隐私。给定隐私预算ε和差分隐私算法M，M的定义域为$\mathrm{Dom}(M)$，值域为$\mathrm{Ran}(M)$；给定n个用户，每个用户对应一条记录。如果M在任意两条记录t和$t'(t,t'\in \mathrm{Dom}(M))$上获得相同的输出$t^*(t^*\in \mathrm{Ran}(M))$，则称算法$M$满足$\varepsilon$-本地差分隐私，可形式化为：

$$\Pr\left[M(t)=t^*\right]\leqslant \mathrm{e}^{\varepsilon}\times \Pr\left[M(t')=t^*\right] \tag{1}$$

其中，隐私预算ε用于控制隐私保护的强度，ε越小，表示隐私保护越强。

定义2 双扰动的本地差分隐私。双扰动的本地差分隐私（double disturbance local differential privacy，DDLDP）是针对区块链的透明性，为保护链上隐私而设计的一种隐私保护方法[22]。其基于LDP生成虚假数据l_i'，将l_i'加入δ-扰动因素，再以概率$P(l_i|l_i')$将l_i'或l_i上传到区块链，其中$P(l_i|l_i')$满足：

$$P\left(l_i|l_i'\right)=\begin{cases}\dfrac{\mathrm{e}^{\varepsilon}}{M'-1+\mathrm{e}^{\varepsilon}}, l_i=l_i'\\ \dfrac{1}{M'-1+\mathrm{e}^{\varepsilon}}, l_i\neq l_i'\end{cases} \tag{2}$$

其中，M'为用户数据数量。在DDLDP中，假设隐私参数为ε，用户数量为n，对数据进行干扰的机制为LP，则输出真实数据和虚假数据的概率之比小于或等于e^{ε}，表示为：

$$\frac{\Pr[\mathrm{LP}(l_i,n,\varepsilon=l_i)]}{\Pr[\mathrm{LP}(l_i,n,\varepsilon=l_i')]}\leqslant \mathrm{e}^{\varepsilon} \tag{3}$$

根据定义1，DDLDP满足ε–本地差分隐私。

2.3 联盟链

区块链作为一种由多个区块组成的分布式账本，具有去中心化、不可篡改、不可否认、透明、可追溯的特点[41]。按照开放程度的不同，区块链

可分为公有链、联盟链和私有链。

联盟链是由多个组织或实体联合维护的区块链网络。与公有链完全开放相比，联盟链中只有被授权的组织或节点才能加入网络，其不依赖于工作量证明等高能耗的共识机制，而采用拜占庭容错、Raft等更高效的共识算法，因此联盟链比公有链扩展性更好。相比私有链的完全中心化，联盟链兼顾了去中心化和可扩展性，适用于多组织间的合作和数据共享。Hyperledger Fabric作为可编程联盟链框架，支持通道技术和多种编程语言，成为一种高度灵活且适应性强的企业级联盟区块链解决方案。

3　系统模型

3.1　系统架构

PBFRec框架系统模型如图1所示，其包含4类实体：请求者、工作者、PBFRec客户端和联盟链。各实体具体描述如下。

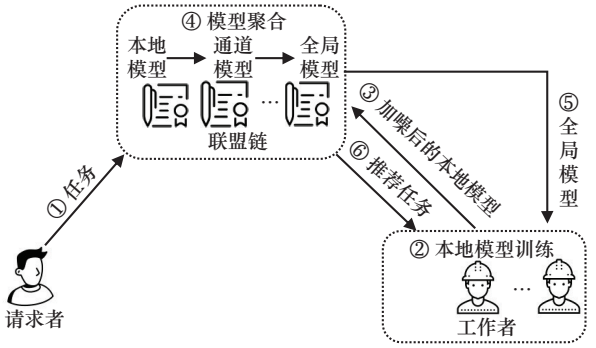


图 1　PBFRec框架系统模型

（1）请求者。请求者为任务需求方，包括个人或者组织，其根据所在地址位置加入相应的联盟链通道，将任务发布在通道内，系统根据训练的个性化任务推荐模型将任务推荐给工作者。

（2）工作者。工作者为任务执行方，其利用存储在本地的历史轨迹数据进行偏好模型训练，得到任务推荐模型参数。工作者根据所在地址位置加入相应的联盟链通道，同一通道内的工作者周期性地将各自训练的本地模型参数和本地数据集特征上传到区块链进行通道模型和全局模型迭代聚合，得到最终模型后，根据任务推荐模型给工作者推荐任务，工作者选择感兴趣的任务参与。

（3）PBFRec客户端。PBFRec客户端是用户（包括工作者或请求者）与联盟链交互的媒介。

（4）联盟链。联盟链是准入式区块链，任务需求发布在联盟链通道中。工作者本地模型训练后向联盟链发送加噪后的模型参数，由智能合约完成联邦模型聚合和个性化任务推荐。

3.2　威胁模型

在PBFRec系统中，任务请求者是半诚实的，其会诚实地执行既定的协议，但对工作者的历史任务参与信息（如参与时间、地点和任务类型等）好奇。工作者是半诚实甚至恶意的，一方面，工作者间好奇彼此的具体历史任务参与信息，但用户活跃度（如用户历史参与任务次数）等统计信息不对用户隐私构成威胁，其在众包中对激励用户积极参与具有重要作用；另一方面，恶意工作者可能上传伪造数据，企图破坏训练的模型质量。对于联盟链，只有在同一地理区域（如同一城市）的用户才能加入同一通道。部分联盟链节点可能是恶意节点或拜占庭节点，但整条联盟链具有拜占庭容错能力。智能合约被认为是可信的，即其执行结果正确且不可篡改。链上数据仅对同一通道内成员可见，用户和联盟链节点可能想根据上链的模型参数反推出其他用户的历史任务参与数据。

3.3　设计目标

针对以上威胁模型，结合任务推荐功能需求，PBFRec旨在实现以下目标。

（1）工作者数据隐私保护：工作者的历史任务参与时间、地点和任务类型等历史轨迹敏感数据信息不应该泄露给除工作者自身以外的其他实体；此外，工作者的本地偏好模型训练参数也应受到保护。

（2）任务推荐模型训练的完整性：应能保证任务推荐模型整个训练过程的完整性，即训练结果按照既定的联邦学习算法和对应输入正确计算。

（3）个性化和高质量任务推荐：PBFRec应能根据工作者对任务的历史偏好，实现对工作者的个性化和高质量任务推荐，并能应对部分恶意工作者上传虚假数据企图破坏模型质量。

3.4　问题定义

定义 3　众包任务。一个众包任务表示为 $s=(l,c,t)$，其中 $l=(l_{\text{lat}},l_{\text{lng}})$ 为任务位置，l_{lat} 和 l_{lng} 分别表示纬度和经度，c 和 t 分别为任务类型和发布时间。

定义 4　工作者。一个工作者表示为 $w=(\text{ch},r,g,S)$，包括工作者所在的联盟链通道ch、其

任务参与的可达半径 r 和当前位置 $g=\left(g_{\text{lat}},g_{\text{lng}}\right)$，其中，$g_{\text{lat}}$ 和 g_{lng} 分别代表纬度坐标和经度坐标，$S=\left\{s_1,s_2,\cdots,s_m\right\}$ 表示工作者的历史参与任务集合。

定义 5 工作者历史任务轨迹。令工作者的历史参与任务集合为 $S=\left\{s_1,s_2,\cdots,s_m\right\}$，其中任务按照参与先后排序，则工作者历史任务轨迹表示为 $\text{HT}=\left(l_1,l_2,\cdots,l_m\right)$，其中 $l_i=(l_{\text{lat}}^i,l_{\text{lng}}^i),i\in[1,m]$ 对应任务 s_i 的任务执行地点，l_{lat}^i 为纬度，l_{lng}^i 为经度。

定义 6 个性化任务推荐。对于任意工作者 $w=\left(\text{ch},r,g,S\right)$，假设其基于本地历史众包参与任务集合 $S=\left\{s_1,s_2,\cdots,s_m\right\}$，共同训练个性化任务推荐模型 θ，令 $\text{Predict}\left(\theta,S\right)=\text{PR}$ 为基于模型预测的推荐任务集合，其中 $\text{Predict}(\cdot)$ 为预测算法，则系统对工作者 w 的个性化任务推荐表示为 $\text{Match}\left(\text{PR},r\right)=R$，其中 $\text{Match}(\cdot)$ 为条件感知算法，R 为推荐任务集合。对于任意推荐任务 $s\in R$，均有 $s\in\text{PR}$，s 发布在通道 ch 内，且 $\|g-l\|\leqslant r$，其中 $\|g-l\|$ 表示工作者当前位置 g 和任务位置 l 的距离（采用球面余弦定理计算）。

4 个性化安全的众包任务推荐框架 PBFRec

PBFRec 框架总体包含两个阶段：数据预处理和个性化任务推荐。工作者在本地个性偏好模型训练前先进行数据预处理，训练后提交模型参数和相关信息至区块链，再进行任务推荐模型聚合和任务推荐。本文用到的关键符号说明如表 1 所示。

表 1 符号说明

符号	说明
R	任务推荐集合
$\boldsymbol{\theta}_l$	本地偏好模型
$\boldsymbol{\theta}_c$	通道模型
$\boldsymbol{\theta}_g$	全局模型
ε	隐私预算
D	工作者数据集特征
S	工作者历史任务集
DS	数据集特征字典

4.1 数据预处理

链下数据预处理由每个工作者执行，旨在将本地非独立同分布、多模态的数据编码，提取其时空相关性，学习其高维表示。其主要包括 5 个子步骤：数据集特征大小统计、空间轨迹嵌入、时间轨迹嵌入、任务类别嵌入和图特征结构化。

数据集特征大小统计主要是提取工作者在最近时间窗口内的参与任务相关数据的特征大小信息，可反映其工作质量。给定通道 ch 下的工作者集合 W，每个工作者 $w\in W$ 基于其历史任务集合 $S=\left\{s_1,s_2,\cdots,s_m\right\}$，其中 $s_i=(l_i,c_i,t_i)\in S$。构建 3 个集合 H_T、H_L 和 H_C，其中 $H_T=\left\{t_1,t_2,\cdots,t_m\right\}$ 存储历史任务执行时间（均不相同，按时间先后排序），$H_L=\left\{l_1,l_2,\cdots,l_p\right\}\subseteq\text{HT}$ 存储历史任务执行地点，$l_1\neq\cdots\neq l_p$ 且有 $p\leqslant m$，$H_C=\left\{c_1,c_2,\cdots,c_q\right\}$（$c_1\neq\cdots\neq c_q,q\leqslant m$）存储历史任务类别。对于时间周期 τ（本文设置 $\tau=30$，表示近 30 天内的历史任务信息），令 H_T^τ、H_L^τ 和 H_C^τ 分别表示 τ 周期内的任务、地点和类别集合，w 得到数据集特征大小统计 $D=(|H_T^\tau|,|H_L^\tau|,|H_C^\tau|)$，其中 $|\cdot|$ 表示集合中的元素个数。w 通过通道 ch 下的数据发布智能合约，将 (w,D) 上传至区块链，并存储在数据集特征字典 DS 中。

空间轨迹嵌入表示工作者历史轨迹的变化规律，并保留其周期性。工作者将其历史任务轨迹 HT 中的位置经纬度 $(l_{\text{lat}}^i,l_{\text{lng}}^i)$ 按照式(4)和式(5)编码为 4 维向量 $\boldsymbol{l}_i$。

$$\begin{aligned}\boldsymbol{l}_i=(&\text{Ensine}(l_{\text{lat}}^i),\text{Ensine}(l_{\text{lng}}^i),\\&\text{Encose}(l_{\text{lat}}^i),\text{Encose}(l_{\text{lng}}^i))\end{aligned}$$

$$\text{Ensine}(p)=\sin\left(p\times\frac{\pi}{180}\right)\tag{4}$$

$$\text{Encose}(p)=\cos\left(p\times\frac{\pi}{180}\right)\tag{5}$$

时间轨迹嵌入旨在将离散的时间连续化，同时兼容多时区的差异，并提取时间的周期特性。考虑到不同工作者历史轨迹可能涉及多国家区域、多时区，本文以 UTC+8 时区为基准，定义偏移量。对于历史轨迹中离散的时间点集合 $H_T=\left\{t_1,\cdots,t_m\right\}$，令 w 所在时区对应的偏移量为 o，w 计算 $t_i'=t_i+o,i\in[1,m]$，并从中获取时间的周期性（以周为单位）。假设工作者偏好在不同年（如闰年、平年）具有稳定性，将 t_i' 编码为 $174=6+7\times24$ 维向量 $\boldsymbol{t}_i$，代表一年中的小时数，满足 $\boldsymbol{t}_i\in\mathbb{R}^{174}$，其中右端 168 位定义为低位，左端 6 位定义为高位（代表一

年中的周数编号）。例如，时间2025-01-11 21:00:00位于2025年第2周的141时，则低位第141处置1，高6位为000010（表示第2周），其余处置0。

任务类别嵌入旨在获取任务类别间的关联，同时将离散的类别映射到向量空间，具体步骤如算法1所示。首先，对于集合H_C中的任务类别，将其对应的字符串按照字典序排序，并将排序后的任务类别放到集合H_C'中。进一步地，初始化矩阵$\boldsymbol{C}$和映射表T_c，对于类别$\forall c_{i'} \in H_C'$，若$c_{i'}$不在T_c中，则将$c_{i'}$及其对应的索引i'分别作为key和value添加到表T_c中。对于任意类别$c_i \in H_C$，定义嵌入向量维度d和映射函数Embedding:$\{1,\cdots,q\} \to \mathbb{R}^d$，将$c_i$对应的索引$i \in \{1,\cdots,q\}$映射到$d$维向量$\boldsymbol{e}$，作为矩阵第$i$行$\boldsymbol{C}[i]$，最终得到$q \times d$维的任务类别嵌入矩阵$\boldsymbol{C} \in \mathbb{R}^{q \times d}$。定义可学习嵌入权重矩阵$\boldsymbol{W} \in \mathbb{R}^{q \times d}$，矩阵中的元素均采用均匀分布$U(-\sqrt{\frac{1}{d}}, \sqrt{\frac{1}{d}})$初始化。对于索引$i \in \{1,\cdots,q\}$，映射函数Embedding$(i)$输出为：

$$\boldsymbol{e} = \boldsymbol{W}[i] \tag{6}$$

算法1　任务类别嵌入算法

输入 工作者历史任务类别集合H_C

输出 任务类别嵌入矩阵$\boldsymbol{C}$

1) 对H_C中的任务类别，基于类别字符串的字典序排序（即ASCII码升序排序），赋值给H_C'
2) 初始化$q \times d$维矩阵$\boldsymbol{C} \leftarrow \mathbb{R}^{q \times d}$
3) 初始化映射表$T_c \leftarrow \{\}$
4) for each $c_{i'} \in H_C'$
5)　if$(c_{i'}, i')$ not in T_c then
6)　　$T_c[c_{i'}] \leftarrow i'$
7)　end if
8) end for
9) for each $c_i \in H_C$
10)　$i \leftarrow T_c[c_i]$
11)　Embedding:$\boldsymbol{W}[i] \to e$
12)　$\boldsymbol{C}[i] \leftarrow e$
13) end for

对于任务$s_i = (l_i, c_i, t_i)$，根据以上计算出的3个向量，得到任务特征向量$\boldsymbol{g}_i = (\boldsymbol{l}_i, \boldsymbol{c}_i, \boldsymbol{t}_i)$，进一步得到任务特征矩阵$\boldsymbol{G}$，且有$\boldsymbol{G} \in \mathbb{R}^{m \times u}$，$\boldsymbol{G}[i] = \boldsymbol{g}_i$，其中$u = 4 + 174 + d$。

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_m \end{bmatrix} \tag{7}$$

图特征结构化旨在提取历史任务轨迹集合中的任务连接关系，进一步构造图结构。定义$m \times m$的任务轨迹邻接矩阵$\boldsymbol{F}$，对于历史参与任务集合$S = \{s_1,\cdots,s_m\}$，相邻任务s_i和s_{i+1}被视为图中两个具有连接关系的节点，其中$i \in [1, m-1]$，则$\boldsymbol{F}[i][i+1] = 1$，$S$可表示为$m$个节点构成的有向图。例如，当$S = \{s_1, s_2, s_3, s_4\}$时，有向图和邻接矩阵如图2所示。

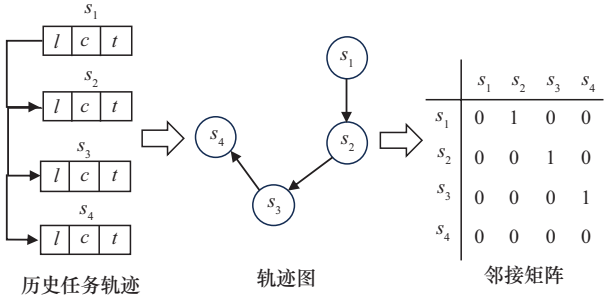


图2　图特征结构化示意

4.2　个性化任务推荐

个性化任务推荐由3个阶段组成，分别为工作者本地个性化偏好学习、联邦模型聚合和任务推荐。

4.2.1　个性化偏好学习

个性化偏好学习分为两个部分：本地GNN建模和模型训练。

（1）本地GNN建模

如图3所示，本地模型网络结构由输入层、输出层、隐藏层和批归一化（batch normalization，BN）层组成。其中，激活函数被用于BN层后进行非线性转换，将输入空间映射到高维特征空间并维持梯度传播。早期激活函数（如Sigmoid、Tanh）会导致梯度消失问题，使本地模型无法学习到有效的特征表示。ReLU激活函数的导数在正区恒为1，可以缓解梯度消失问题，其次简单的阈值判断对计算资源要求低（在同等硬件加速下耗时仅为Sigmoid的1/4），所以本文将ReLU作为激活函数。下面对网络结构进行具体描述。

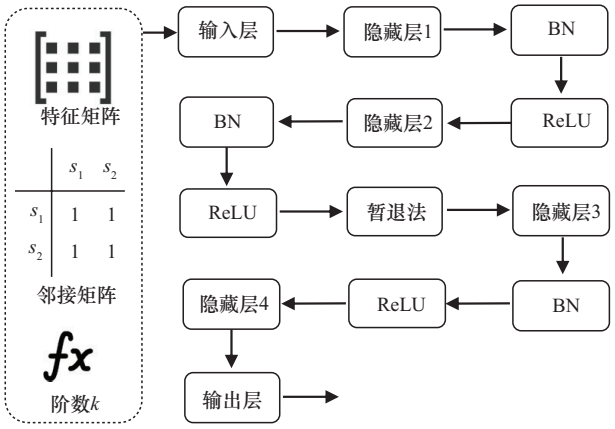


图3　本地模型网络结构

① 输入层。输入层的数据输入包括任务特征矩阵$\boldsymbol{G}$、任务轨迹邻接矩阵$\boldsymbol{F}$和多项式阶数k。

② 隐藏层。模型包含隐藏层1～4，本文采用ChebConv图卷积层实现，ChebConv采用多项式滤波器，解决了传统非参数滤波器在空间上非局部化和学习复杂度高的局限性，可表示为：

$$x' = \sum_{i=1}^{k} z^{(i)} \cdot \kappa^{(i)} \tag{8}$$

其中，k为多项式的阶数，$\boldsymbol{\kappa}$为多项式系数的向量，$\boldsymbol{z}$为可学习参数。

③ BN层。BN层旨在实现特征向量的批归一化，加速模型的训练，缓解梯度消失和爆炸问题，提高模型的泛化能力。给定任务特征矩阵$\boldsymbol{G}$（包含m个特征样本），对于$\boldsymbol{G}$中第i个特征行向量$\boldsymbol{g}_i = (\boldsymbol{l}_i,\boldsymbol{c}_i,\boldsymbol{t}_i)$，令$\boldsymbol{g}_i = (v_i^1,v_i^2,\cdots,v_i^n)$。针对第$j$个特征维度，分别计算该维度下所有样本的均值$\mu_j = \frac{1}{m}\sum_{i=1}^{m} v_i^j$和方差$\sigma_j^2 = \frac{1}{m}\sum_{i=1}^{m}(v_i^j - \mu_j)^2$。接着，根据式(9)对第$i$个样本的第$j$个特征值$v_i^j$进行归一化得到$\hat{v}_i^j$（即特征向量$\boldsymbol{g}_i$对应归一化得到向量$\hat{\boldsymbol{g}}_i$）。

$$\hat{v}_i^j = \frac{v_i^j - \mu_j}{\sqrt{\sigma_j^2 + \varphi}} \tag{9}$$

其中，φ为常数。最后计算$\boldsymbol{y}_i = \gamma \times \hat{\boldsymbol{g}}_i + \beta$，即对$\hat{g}_i$进行缩放和平移，$\gamma$、$\beta$均为一组可学习参数。

④ 输出层。在输出层中，输入的特征矩阵经转换，输出符合定义维度的特征向量。

⑤ ReLU。给定输入值x，经过ReLU激活函数得到$\mathrm{ReLU}(x) = \max(x,0)$。

⑥ 暂退法。给定正则化概率为p，对于一个具有n层神经元的网络结构，在每个迭代周期，神经元均有p的概率被置0，$1-p$的概率被保留。

（2）本地GNN模型训练

本地模型训练流程如图4所示。给定损失函数$\mathrm{loss}(\cdot)$，在第t轮，对历史任务集合S_t进行数据预处理，得到矩阵$\boldsymbol{G}$、$\boldsymbol{F}$作为输入层输入，输出层得到本地模型类别预测向量$\boldsymbol{x}$，其中每个元素$\boldsymbol{x}[j]$表示模型对类别j的预测分数值，本地模型梯度$\boldsymbol{\theta}_l^t$基于第$t-1$周期的梯度$\boldsymbol{\theta}_l^{t-1}$和$\mathrm{loss}(\mathrm{ct},\boldsymbol{x})$更新，在此期间计算$K$值的召回率直至趋于稳定。其中，$\mathrm{loss}(\mathrm{ct},\boldsymbol{x})$可表示为：

$$\mathrm{loss}(\mathrm{ct},\boldsymbol{x}) = -\boldsymbol{x}(\mathrm{ct}) + \log\left(\sum_j \exp(\boldsymbol{x}[j])\right) \tag{10}$$

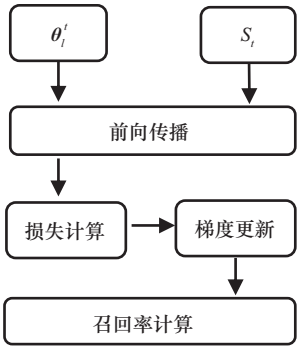


图4　本地模型训练流程

4.2.2　联邦模型聚合

与已有方案对所有本地模型计算均值进行聚合不同，本文将模型分为本地模型$\boldsymbol{\theta}_l$、通道模型$\boldsymbol{\theta}_c$和全局模型$\boldsymbol{\theta}_g$，并提出了基于模型质量的两层个性化模型聚合方法。

本地模型→通道模型。假设通道ch下的本地模型梯度向量为$\boldsymbol{\theta}_l = (\boldsymbol{\theta}_{l1},\cdots,\boldsymbol{\theta}_{ln})$。对于工作者$w_i$，在上传本地训练出的$\boldsymbol{\theta}_{li}$至区块链之前，通过DDLDP算法保护梯度隐私。具体地，先基于LDP算法，生成干扰梯度$\boldsymbol{\theta}_{li}^* = \mathrm{LDP}(\boldsymbol{\theta}_{li},\varepsilon)$，其中$\varepsilon$为隐私预算；再对干扰梯度$\boldsymbol{\theta}_{li}^*$添加干扰因子$\delta$，即计算$\boldsymbol{\theta}_{li}^* \leftarrow \boldsymbol{\theta}_{li}^* + \delta$；最后，$w_i$以概率$P(\boldsymbol{\theta}_{li}|\boldsymbol{\theta}_{li}^*)$将真实梯度$\boldsymbol{\theta}_{li}$或干扰梯度$\boldsymbol{\theta}_{li}^*$上传至区块链。

接着，由ch下的通道聚合智能合约随机在通道内挑选一个对等节点$\mathrm{Peer}_{\mathrm{ch}}$作为临时通道聚合服务器，并通过本地模型质量动态评估算法（算法2）生成本地模型权重集合WE，相应地为双扰动后的本地模型$\boldsymbol{\theta}_{li}^*$分配权重$\mathrm{weight}_i^l \in \mathrm{WE}$，生成权重向量

$\mathbf{we}_l = (\text{weight}_1^l, \cdots, \text{weight}_n^l)$，根据权重分配结果进行第一层聚合得到通道模型 $\boldsymbol{\theta}_c = \mathbf{we}_l \cdot \boldsymbol{\theta}_l^*$。

算法 2 从数据量稳定性（基于四分位数，low 和 high 进行分割）、数据内在质量（q 值）和信任值（trust 值）等多个维度对用户本地模型进行综合评估，防止恶意用户伪造数据发起投毒污染后续全局模型。

算法 2　本地模型质量动态评估算法

输入　ch 下的工作者 $\tau = 30$ 周期内的数据集特征字典 DS

输出　本地模型权重集合 WE

1) $d \leftarrow []$, trust $\leftarrow []$, $q \leftarrow []$
2) for each $w_i \in W$
3)　DS$[w_i] = (|H_T^\tau|, |H_L^\tau|, |H_C^\tau|)$
4)　$d[i] \leftarrow |H_T^\tau|$
5)　if $d[i] \leqslant 5$ then
6)　　trust$[i] \leftarrow 0.3$
7)　else
8)　　$\text{div}_L \leftarrow \dfrac{|H_L^\tau|}{|H_T^\tau|}$, $\text{div}_C \leftarrow \dfrac{|H_C^\tau|}{|H_T^\tau|}$
9)　　if $\text{div}_L < 0.3$ or $\text{div}_C < 0.3$ then
10)　　　trust$[i] \leftarrow 0$
11)　　else
12)　　　trust$[i] \leftarrow 0.7 + 0.3 \times \min(\text{div}_L, \text{div}_C)$
13)　　end if
14)　end if
15)　　norm_entropy_T $\leftarrow \min\left(\dfrac{|H_T^\tau|}{10}, 1\right)$
16)　　category $\leftarrow 1 - \max\left(\dfrac{1}{|H_C^\tau|}, 0.25\right)$
17)　　$q[i] \leftarrow 0.4 \times$ norm_entropy_T $+ 0.3 \times \text{div}_L + 0.3 \times$ category
18) end for
19) low $\leftarrow d[0.25 \times \text{len}(d)]$
20) high $\leftarrow d[0.75 \times \text{len}(d)]$
21) weight $\leftarrow \{\}$
22) for each index $\in d$
23)　if $d[\text{index}] = 0$ then
24)　　weight $\leftarrow 0$
25)　if $d[\text{index}] < \text{low}$ then
26)　　weight $\leftarrow \dfrac{d[\text{index}]}{\text{low}}$
27)　if $d[\text{index}] > \text{high}$ then
28)　　weight $\leftarrow 1 - \left(\dfrac{d[\text{index}] - \text{high}}{2\text{high} - \text{low}}\right)$
29)　else
30)　　weight $\leftarrow 1$
31)　end if
32)　if $q[\text{index}] < 0.6$ or trust$[\text{index}] < 0.5$ then
33)　　weight $\leftarrow 0$
34)　end if
35)　WE$[\text{index}] \leftarrow \max(0, \min(\text{weight}, 1))$
36) end for
37) return WE

通道模型→全局模型。给定一组通道模型集合 $\boldsymbol{\theta}_c = \{\boldsymbol{\theta}_{c1}, \boldsymbol{\theta}_{c2}, \cdots, \boldsymbol{\theta}_{cN}\}$ 和一组全局对等节点集合 $\text{Peer}_a = \{\text{Peer}_{a1}, \cdots, \text{Peer}_{aN}\}$，在全局聚合前，由全局聚合智能合约随机选择 $\text{Peer}_{ai} \in \text{Peer}_a$ 作为全局聚合服务器，并通过通道模型质量动态评估算法（算法 3）生成通道模型权重集合 CG，为各通道模型梯度 $\boldsymbol{\theta}_{ci}$ 分配权重 $\text{weight}_i^c \in \text{CG}$，根据权重分配结果进行第二层聚合得到全局模型，表示为：

$$\boldsymbol{\theta}_g = \text{weight}_1^c \boldsymbol{\theta}_{c1} + \cdots + \text{weight}_N^c \boldsymbol{\theta}_{cN} \tag{11}$$

算法 3　通道模型质量动态评估算法

输入　所有通道下的数据集特征字典集合 $\text{DS}' = (\text{DS}_1, \cdots)$

输出　通道模型权重集合 CG

1) counts $\leftarrow \{\}$, CG $\leftarrow \{\}$
2) for each $\text{DS}_{\text{ch}} \in \text{DS}'$
3)　flag $\leftarrow$ ch
4)　counts[flag] $\leftarrow |\text{DS}_{\text{ch}}|$
5) end for
6) for each flag $\in$ count
7)　CG[flag] $= \dfrac{\text{counts}[\text{flag}]}{|\text{counts}|}$
8) end for
9) return CG

工作者的加入和退出，会导致通道工作数据集特征总量变化，算法 3 旨在每次聚合全局模型前，动态地为通道模型梯度分配权重。

此外，针对新用户存在的冷启动问题，本文采用基于社交关系的模型初始化方法（在实际众包场景中，可利用区块链预言机获取第三方社交平台中

用户间社交关系）。假设工作者社交关系（用有向图表示，节点为工作者）已知，将其采用邻接矩阵$\boldsymbol{H}$存储。给定一组ch下的工作者集合W，工作者$w_i \in W$对$w_{i+1} \in W$有社交关系（即存在从w_i至w_{i+1}的有向边）可表示为$\boldsymbol{H}[i][j]=1$；否则，$\boldsymbol{H}[i][j]=0$。定义$\alpha_{i,j}=\boldsymbol{H}[i][j]+\boldsymbol{H}[j][i]$，$w_i$的度之和表示为$d'=\boldsymbol{H}[i,:]+\boldsymbol{H}[:,i]$。假设$w_i$为新人，其存在朋友工作者集合$\{w_{i-1},w_{i+1}\}$，则$w_i$的本地模型梯度可初始化为：

$$\boldsymbol{\theta}_{li}=\frac{\alpha_{i,i-1}\boldsymbol{\theta}_{l(i-1)}+\alpha_{i,i+1}\boldsymbol{\theta}_{l(i+1)}}{d'} \tag{12}$$

4.2.3 任务推荐

任务推荐包括模型预测和条件感知两个阶段，给定通道ch下的在线工作者$W=\{w_1,\cdots,w_n\}$，可达半径分别为（$r_1,\cdots,r_n$）和一组在线众包任务$S=\{s_1,\cdots,s_n\}$，对工作者w_i的任务推荐步骤如算法4所示。

算法4 通道任务推荐算法

输入 工作者w_i可达半径r_i，在线任务集合S

输出 工作者w_i任务推荐集合R_i

1) for each iteration $t \in [1,T]$
2) for each $w_i \in \{w_1,w_2,\cdots,w_n\}$
3) $\mathrm{PR}_i \leftarrow \mathrm{Predict}(\boldsymbol{\theta}_{li},S)$
4) $R_i \leftarrow \mathrm{Match}(\mathrm{PR}_i,r_i)$
5) end for
6) end for
7) return R_i

模型预测阶段核心任务为获取ch内符合工作者偏好的任务集合。对于工作者$w_i \in W$，其更新本地模型为当前全局模型，即有$\boldsymbol{\theta}_{li}=\boldsymbol{\theta}_g$。预测算法$\mathrm{Predict}(\cdot)$的输入为$\boldsymbol{\theta}_g$和$S$，输出为模型预测结果任务集合$\mathrm{PR}_i$，集合元素按照偏好程度从高到低排序。

条件感知阶段在模型预测的基础上，根据工作者当前状态（位置、时间）引入限制条件（可达距离），进一步过滤得到最终推荐任务集合。具体地，条件感知算法$\mathrm{Match}(\cdot)$输入w_i的模型预测结果集合PR_i和可达半径r_i，对于$\forall s_j \in \mathrm{PR}_i$，其根据工作者当前位置$g_i$，计算距离$\mathrm{len}_{ij}=\|g_i-l_j\|$。若$\mathrm{len}_{ij} \leqslant r_i$，则表示任务$s_j$在符合$w_i$偏好的同时是可达的，将$s_j$加入任务推荐集合$R_i$中。由于工作过程中$w_i$和任务状态是动态变化的，因此，算法4中的任务推荐被设置为一个周期性过程。

5 安全性分析

定理1 PBFRec保护偏好模型和工作者原始数据隐私（即数据和模型机密性），若每轮满足ε-LDP，设总训练轮数为T，则PBFRec总体满足$T\varepsilon$-LDP。

证明 根据式(3)，PBFRec提供的差分隐私保护处理体现在以下阶段：工作者w_i在每轮上传本地模型至链上参与模型聚合更新前，使用DDLDP对本地模型$\boldsymbol{\theta}_{li}$加入差分扰动，防止众包用户或联盟链节点根据上链的本地模型参数反推出其他用户本地存储的历史任务参与数据（即模型反演攻击），即保护工作者原始数据和本地模型机密性。即使存在部分恶意工作者，由于各工作者独立上传扰动后的本地模型参数，因此不影响以上机密性。

假设每轮算法隐私预算为ε，随机算法为M_t: $D \to R$，$D_1,D_2 \in D$为第t轮任意两个相邻数据集合（即有$|D_1 \oplus D_2|=1$），根据定义1，对于$\forall r \in R$，满足：

$$\Pr\left[M_t(D_1)=r\right] \leqslant \mathrm{e}^{\varepsilon \times |D_1 \oplus D_2|} \times \Pr\left[M_t(D_2)=r\right] \tag{13}$$

本文每轮模型参数上传使用的DDLDP方法满足ε-LDP，因此PBFRec每轮也满足ε-LDP。对于T轮训练，随机算法分别对应$M_1,M_2,\cdots,M_T$，则由差分隐私的组合定理可知，PBFRec总体满足$T\varepsilon$-LDP。证毕。

定理2 在联盟链具有拜占庭容错能力下，PBFRec通道模型和全局模型聚合结果具有完整性，且整个模型具有可用性，恶意工作者的比例不会影响聚合完整性和模型可用性。

证明 在通道模型聚合机制中，PBFRec采用通道聚合合约动态评估通道模型质量并进行通道模型聚合。由于联盟链具有拜占庭容错能力，即使有部分恶意节点，智能合约的执行过程也是自动且不可篡改的，执行结果经由共识机制被多个对等节点验证存储，因此聚合后的通道模型$\boldsymbol{\theta}_c$具有完整性。

同理，采用全局模型聚合合约实现全局模型质量态评估，进一步聚合得到全局模型$\boldsymbol{\theta}_g$，聚合结果具有完整性，且整个模型也具有可用性。即使存在恶意工作者，算法2也能有效识别其恶意行为（如发送伪造数据），将恶意工作者在本地模型聚合过程中的权重置为0，即其不会影响后续的通道模型和全局模型，因此恶意工作者的比例不会影响聚合完整性和模型可用性。

6　实验

6.1　实验设计

本文使用 Foursquare 最新的签到数据集模拟工作者的众包任务轨迹，数据集拥有 114 324 个用户群、22 809 624 条签到记录，同时包含用户之间的社交关系。签到记录包含签到时间、签到位置和签到地点类型。本文将所有用户视为工作者，将每位工作者最近的一次签到位置作为工作者的当前位置。区块链采用 Hyperledger Fabric，本文设置 3 个通道，用户均匀分布到每个通道中，每个用户只属于 1 个通道。所有实验均在 AMD Ryzen 7 5800H with Radeon Graphics@3.2 GHz 和 NVIDIA GeForce RTX 3060 Laptop GPU 上实现。

本文将签到时间作为任务完成时间，统计了每个用户签到过的地点个数，其中用户签到地点个数的最大值为 1 567，最小值为 1，平均值为 92.27。图 5 体现了数据集在“工作者–地点类别–签到数量”分布上的稀疏性。其中，地点类别的数字（0~500）代表 Foursquare 数据集中原始地点类别的索引编号，可以看出，用户在大部分地点的签到次数极少，少数用户在特定地点类别上签到次数较多。为探究不同活跃度用户对隐私预算 ε 和推荐列表大小 K 的敏感性差异，评估模型在不同数据稀疏度条件下的适应性，本文基于签到地点数量平均值（92.27），将用户划分为高稀疏度群体和低稀疏度群体，设置多组隐私预算 ε（如 $\varepsilon\in\{0.1,0.2,0.8\}$）和 K 值（如 $K\in\{1,5,10\}$）的实验组合。此外，为了验证社交关系在不同稀疏群体下的推荐效果，评估模型在冷启动场景下的有效性，本文分层提取测试集进行对照实验，具体为：从高稀疏度群体和低稀疏度群体中分别随机抽取等规模的用户样本，从每个用户的时间序列中提取最后 10（$K\leqslant 10$）条签到记录作为测试样本。针对每个分层后的测试用户，构建两种对比实验场景：完全删除用户的历史签到记录，仅依靠社交关系网络进行推荐；完全删除用户的历史签到记录和社交记录。接下来分别对两个场景进行实验。

本文将任务偏好模型推荐的质量、社交关系对解决冷启动问题带来的平均贡献度、链下预处理时间和链上开销作为系统的评估指标。

模型对工作者任务偏好预测质量使用召回率（Recall@K）作为具体衡量指标，衡量的是模型推荐的前 K 个结果中，工作者实际感兴趣的任务个数在测试集里感兴趣任务总数中所占的比例。该指标越高，表明模型在 Top-K 推荐列表中的预测质量越高（即推荐更全面）。计算式为：

$$\text{Recall@}K \leqslant \frac{\left|R_i \cap R_i'\right|}{\left|R_i'\right|} \tag{14}$$

其中，R_i 是对工作者 w_i 推荐的 K 个任务，R_i' 是 w_i 实际在整个测试集中感兴趣的任务。本文将数据集的 70% 作为训练集，30% 作为测试集，重点分析了不同隐私预算 ε、K 值和不同稀疏度对方案推荐召回率的影响，并选择以下代表性方法进行比较。

FedAvg[42]。FedAvg 将本地模型进行平均汇总，用于全局模型的更新。

CTP[43]。CTP 假设所有工作者属于同一个平台中心，同时中央服务器采用集中式架构。

PLP[44]。PLP 基于工作者的历史时空轨迹预测其位置偏好，工作者分布在多个平台中心，将任务匹配转换为最大权值匹配，中央服务器采用集中式架构，不需要和所有平台同时保持通信。

工作者间社交关系的平均贡献度定义为 contribution@K = Recall1@K − Recall2@K，其中 Recall1@K 和 Recall2@K 分别为实验场景 1 和实验场景 2 下的召回率。此外，链上开销包括链码部署时间、交易吞吐量（即每秒钟的交易确认数量）和时延。

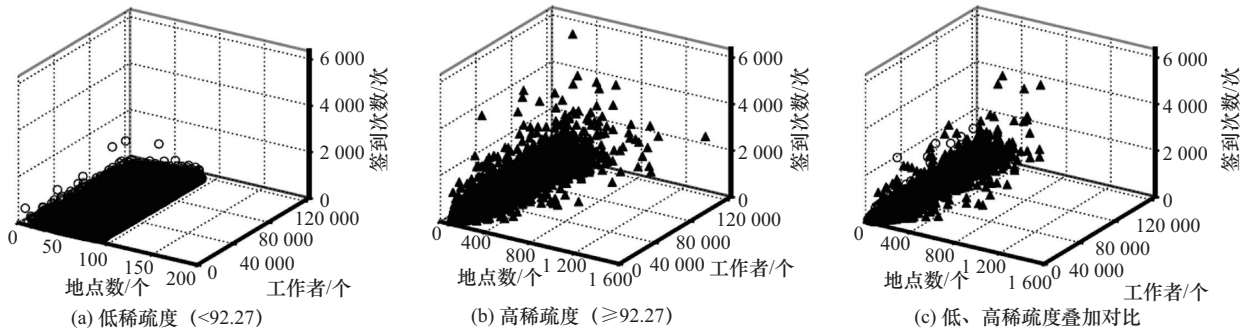


图 5　用户签到数量分布

6.2　实验结果

6.2.1　推荐质量

表2和表3分别对比了PBFRec和其他方法在低稀疏度和高稀疏度下的召回率Recall@K。可以看出，不管何种稀疏度，4种方法的召回率均随K值的增加而提高。FedAvg、CTP和PLP在相同K值下的召回率近似，远低于PBFRec，其中FedAvg的召回率最低。表2和表3表明，基于GNN的PBFRec框架更能全面地表示工作者的偏好，且随着ε的增大，同一K值下的召回率增高，这是因为ε越大，隐私保护程度越低，因此召回率越高。

表2　低稀疏度下的召回率Recall@K

方法	K=1	K=5	K=10
FedAvg	0.098 6	0.206 8	0.273 1
CTP	0.113 8	0.218 3	0.276 3
PLP	0.111 9	0.228 7	0.289 2
PBFRec(ε = 0.8)	0.547 3	0.751 2	0.794 3
PBFRec(ε = 0.2)	0.410 6	0.681 3	0.702 5
PBFRec(ε = 0.1)	0.385 7	0.519 3	0.531 7

表3　高稀疏度下的召回率Recall@K

方法	K=1	K=5	K=10
FedAvg	0.136 6	0.286 9	0.378 4
CTP	0.157 4	0.302 8	0.383 1
PLP	0.155 1	0.316 9	0.400 9
PBFRec(ε = 0.8)	0.748 5	0.892 0	0.935 0
PBFRec(ε = 0.2)	0.635 0	0.781 4	0.827 0
PBFRec(ε = 0.1)	0.592 7	0.713 5	0.748 1

对比表2和表3可以看出，在相同设置下，PBFRec在高稀疏度下的召回率均高于低稀疏度下的召回率。当K=10、ε = 0.8时，高稀疏度下的召回率比低稀疏度下的增加约15%。这是由于高稀疏度下的用户签到地点数量更集中，对用户的推荐偏好也更加集中，模型能够更全面地捕捉其兴趣模式。

6.2.2　社交关系对解决冷启动的贡献度

图6对比了ε =0.8时不同稀疏度和K值下工作者间社交关系的平均贡献度。可以看出，高稀疏度下的平均贡献度总体高于低稀疏度，且均为正值，这表明实验场景1下的召回率更高，即社交关系对解决冷启动问题、提高召回率具有重要作用，该作用在高稀疏度下更为明显。当K值增大时，高稀疏度下的社交关系平均贡献度相应增加，而低稀疏度下的平均贡献度呈不稳定变化趋势。这是由于低稀疏度群体社交关系比较简单，当K值大于某个值时，反而会影响平均贡献度。

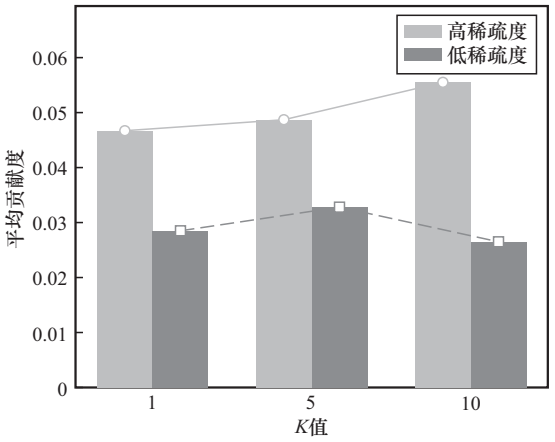


图6　ε=0.8时不同稀疏度和K值下工作者间社交关系的平均贡献度

6.2.3　链下开销

本次测试选取了7个工作者样本进行实验，图7评估了不同数据样本量（即不同任务集合大小$|S|$）下的数据预处理时间。可以看出，$|S|$与数据预处理时间之间存在正相关关系，但非线性。这主要归因于样本间数据复杂度的显著差异及其非独立同分布特性。在不同区间内，样本时空关联性与特征维度不同，使轨迹嵌入与图特征结构化等步骤产生的计算开销有显著差异，从而引起该非线性增长。

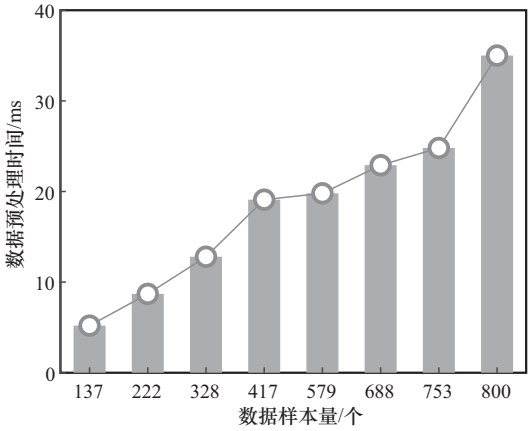


图7　数据预处理时间

6.2.4　链上开销

对于链码部署开销，本文以任务池链码task和准入鉴权模块链码user为例，对其在Hyperledger Fabric网络中打包、安装、审批和提交进行测试，并记录

各个阶段的平均用时。具体地，给定排序节点order、对等节点PEER = { $peer_1$,$peer_2$}、组织集合ORG = { org_1,org_2}，其中PEER、ORG位于ch内。

在测试过程中，链代码首先会安装依赖并被打包，然后由 org_1 和 org_2 分别安装在 $peer_1$ 和 $peer_2$，得到链码安装ID，安装成功后，由 org_1 和 org_2 批准并提交到ch内。链码部署的各阶段开销如图8所示。在链码安装阶段，task和user链码的开销均达峰值，随后审批阶段用时大幅下降（分别为221 ms和219 ms），最后提交阶段与审批阶段平均用时大致接近。

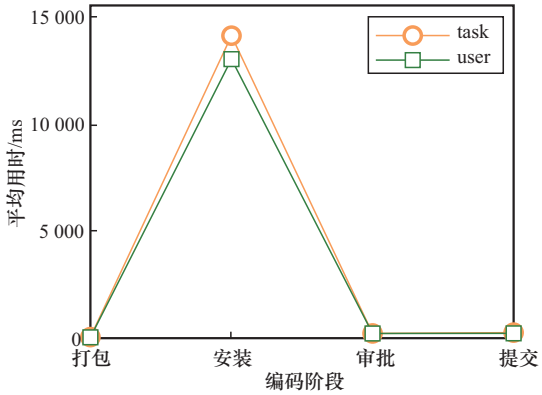


图8　链码部署的各阶段开销

图9展示了在通道数量为10的情况下，单个通道内工人数量对模型聚合算法的吞吐量和平均时延的影响。可以看出，随着通道内工人数量的增加，吞吐量下降，时延上升。这一性能损耗主要是由于工人数量增加，聚合时需要处理的模型数量也随之线性增加，进而增大单次交易执行开销，影响了系统整体效率。然而，本文方案依然保持了较高的可用性与稳定可接受的时延。

图10展示了在单通道工人数量固定为50时，模型聚合交易吞吐量与平均时延随通道数量的变化趋势。结果表明，当通道数量从10增至50时，系统吞吐量呈衰减趋势（由25.3 TPS降至10.8 TPS），但平均时延逐渐上升（由0.27 s增至0.66 s）。这一开销源于Fabric的运行机制：维护大量独立通道引发了剧烈的磁盘I/O竞争（多账本寻址）与Raft共识机制的通信过载。尽管多通道架构带来了不可忽视的算力与网络开销，但是本文方案在50个通道的高并发场景下，依然将平均时延严格控制在0.7 s内。

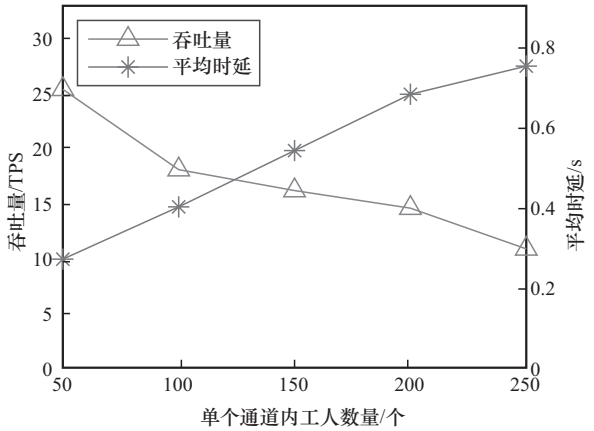


图9　单个通道内工人数量对模型聚合算法的吞吐量和平均时延的影响

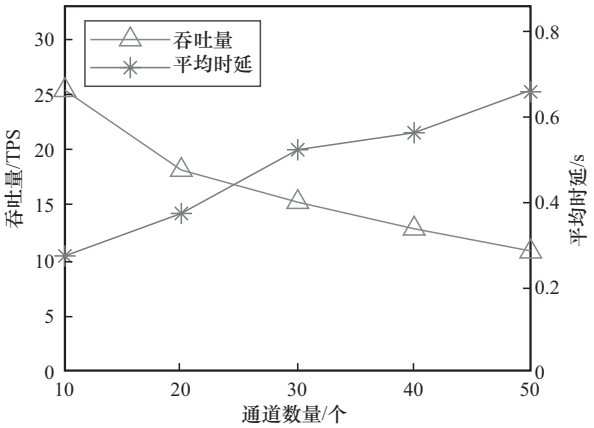


图10　模型聚合交易吞吐量与平均时延随通道数量的变化趋势

7　结束语

本文针对众包服务中的个性化任务推荐和安全问题，提出了一种去中心化和隐私保护的个性化众包任务推荐框架PBFRec。PBFRec采用区块链代替传统众包平台，基于GNN和联邦学习训练用户个性化推荐模型。在模型聚合阶段，设计了基于动态质量评估的双层模型聚合机制，并采用双扰动差分隐私保护本地模型参数。理论分析表明，PBFRec满足 $T\varepsilon$-本地差分隐私和模型聚合完整性；实验验证了PBFRec的推荐质量和性能。

参考文献：

[1] 赵江华, 穆舒婷, 王学志, 等. 科学数据众包处理研究[J]. 计算机研究与发展, 2017, 54(2): 284-294.
Zhao J H, Mu S T, Wang X Z, et al. Crowdsourcing-based scientific data processing[J]. Journal of Computer Research and Development, 2017, 54(2): 284-294.

[2] Zhang L C, Zhang Y Y, Li T. Dynamic food delivery problem based on spatial crowdsourcing[J]. IEEE Transactions on Services Computing,

2025, 18(5): 2768-2780.

[3] Bhatti S S, Gao X F, Chen G H. General framework, opportunities and challenges for crowdsourcing techniques: a comprehensive survey[J]. Journal of Systems and Software, 2020, 167: 110611.

[4] Wu H Q, Düdder B, Jiang S R, et al. VP--match: verifiable privacy-aware and personalized crowdsourcing task matching via blockchain[J]. IEEE Transactions on Mobile Computing, 2024, 23(10): 9913-9930.

[5] Li L, Wu H Q, Shen J C, et al. PGVMatch: privacy-preserving and fine-grained crowdsourcing task matching with lightweight on-chain public verifiability[J]. IEEE Transactions on Mobile Computing, 2025, 24(9): 8642-8655.

[6] Li L, Wu H Q, He L G, et al. Matching as you want: a decentralized, flexible, and efficient realization for crowdsourcing with dual-side privacy[J]. IEEE Transactions on Network Science and Engineering, 2025, 12(2): 1026-1040.

[7] Guo Y, Xie H C, Miao Y B, et al. FedCrowd: a federated and privacy-preserving crowdsourcing platform on blockchain[J]. IEEE Transactions on Services Computing, 2022, 15(4): 2060-2073.

[8] Deng L W, Zhao Y, Cui Y, et al. Task recommendation in spatial crowdsourcing: a trade-off between diversity and coverage[C]//Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE). Piscataway: IEEE Press, 2024: 276-288.

[9] Xu C H, Song W. Intelligent task allocation for mobile crowdsensing with graph attention network and deep reinforcement learning[J]. IEEE Transactions on Network Science and Engineering, 2023, 10(2): 1032-1048.

[10] Tan Y H, Peng T, Wang G J, et al. LO-GDRL: privacy-preserving online task allocation based on Lyapunov optimization and graph-based deep reinforcement learning in mobile crowdsensing[J]. Computer Networks, 2026, 282: 112215.

[11] Zhao X Y, Xia L, Zou L X, et al. Whole-chain recommendations[C]//Proceedings of the 29th ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2020: 1883-1891.

[12] 熊金波, 毕仁万, 田有亮, 等. 移动群智感知安全与隐私: 模型、进展与趋势[J]. 计算机学报, 2021, 44(9): 1949-1966.
Xiong J B, Bi R W, Tian Y L, et al. Security and privacy in mobile crowdsensing: models, progresses, and trends[J]. Chinese Journal of Computers, 2021, 44(9): 1949-1966.

[13] Miao H, Zhong X L, Liu J X, et al. Task assignment with efficient federated preference learning in spatial crowdsourcing[J]. IEEE Transactions on Knowledge and Data Engineering, 2024, 36(4): 1800-1814.

[14] 金歌, 魏晓超, 魏森茂, 等. FPCBC: 基于众包聚合的联邦学习隐私保护分类系统[J]. 计算机研究与发展, 2022, 59(11): 2377-2394.
Jin G, Wei X C, Wei S M, et al. FPCBC: federated learning privacy preserving classification system based on crowdsourcing aggregation[J]. Journal of Computer Research and Development, 2022, 59(11): 2377-2394.

[15] 徐天承, 乔少杰, 武俊, 等. 一种基于时空位置预测的空间众包任务分配方法[J]. 计算机研究与发展, 2022, 59(2): 310-328.
Xu T C, Qiao S J, Wu J, et al. A spatial crowdsourcing task assignment approach based on spatio-temporal location prediction[J]. Journal of Computer Research and Development, 2022, 59(2): 310-328.

[16] 张宇, 江海峰, 杨浩文, 等. 移动群智感知中基于联邦学习的参与者选择机制[J]. 计算机应用研究, 2023, 40(4): 1172-1177, 1183.
Zhang Y, Jiang H F, Yang H W, et al. Participant selection mechanism based on federated learning in mobile crowd sensing[J]. Application Research of Computers, 2023, 40(4): 1172-1177, 1183.

[17] Xu Y, Xiao M J, Wu J, et al. Privacy preserving task push in spatial crowdsourcing with unknown popularity[J]. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(11): 2039-2053.

[18] Wang E, Li B J, Liu W B, et al. Federated learning-based distributed data completion in sparse mobile CrowdSensing[J]. IEEE Transactions on Mobile Computing, 2026, 25(5): 6784-6800.

[19] Kadadha M, Otrok H, Mizouni R, et al. On-chain behavior prediction machine learning model for blockchain-based crowdsourcing[J]. Future Generation Computer Systems, 2022, 136: 170-181.

[20] Lin H, Garg S, Hu J, et al. Blockchain and deep reinforcement learning empowered spatial crowdsourcing in software-defined Internet of vehicles[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(6): 3755-3764.

[21] Ma H Y, Huang S L, Guo J L, et al. Blockchain-based privacy-preserving federated learning for mobile crowdsourcing[J]. IEEE Internet of Things Journal, 2024, 11(8): 13884-13899.

[22] Sun Z C, Wang Y J, Cai Z P, et al. A two-stage privacy protection mechanism based on blockchain in mobile crowdsourcing[J]. International Journal of Intelligent Systems, 2021, 36(5): 2058-2080.

[23] Mills J, Hu J, Min G Y. Multi-task federated learning for personalised deep neural networks in edge computing[J]. IEEE Transactions on Parallel and Distributed Systems, 2022, 33(3): 630-641.

[24] Zhao B W, Tang S H, Liu X M, et al. iTAM: bilateral privacy-preserving task assignment for mobile crowdsensing[J]. IEEE Transactions on Mobile Computing, 2021, 20(12): 3351-3366.

[25] 宋甫元, 丁思洋, 王威, 等. 面向众包的隐私保护任务匹配方案[J]. 通信学报, 2025, 46(5): 115-132.
Song F Y, Ding S Y, Wang W, et al. Privacy-preserving task matching scheme for crowdsourcing[J]. Journal on Communications, 2025, 46 (5): 115-132.

[26] 龙浩, 张书奎, 张洋, 等. 移动群智感知中基于社区的任务分发算法[J]. 通信学报, 2019, 40(10): 42-54.
Long H, Zhang S K, Zhang Y, et al. Task distribution algorithm based on community in mobile crowd sensing[J]. Journal on Communications, 2019, 40(10): 42-54.

[27] Zhao Y, Zheng K, Yin H Z, et al. Preference-aware task assignment in spatial crowdsourcing: from individuals to groups[J]. IEEE Transactions on Knowledge and Data Engineering, 2022, 34(7): 3461-3477.

[28] Zhao Y, Deng L W, Zheng K. AdaTaskRec: an adaptive task recommendation framework in spatial crowdsourcing[J]. ACM Transactions on Information Systems, 2023, 41(4): 1-32.

[29] Tian Y H, Huang S Q, Zhang Y, et al. Crowdsourcing task recommendation model based on collaborative knowledge graph[J]. Computers & Industrial Engineering, 2025, 210: 111516.

[30] 李洪涛, 任晓宇, 王洁, 等. 基于差分隐私的连续位置隐私保护机制[J]. 通信学报, 2021, 42(8): 164-175.

Li H T, Ren X Y, Wang J, et al. Continuous location privacy protection mechanism based on differential privacy[J]. Journal on Communications, 2021, 42(8): 164-175.

[31] 杨明川, 朱敬华, 李元婧, 等. 一种考虑隐私保护的深度强化学习任务分配模型[J]. 计算机研究与发展, 2023, 60(11): 2650-2659.
Yang M C, Zhu J H, Li Y J, et al. Task allocation model based on deep reinforcement learning considering privacy protection[J]. Journal of Computer Research and Development, 2023, 60(11): 2650-2659.

[32] 田有亮, 田茂清, 高鸿峰, 等. 面向群智感知应用的基于协作的位置认证方案[J]. 通信学报, 2022, 43(9): 121-133.
Tian Y L, Tian M Q, Gao H F, et al. Cooperation-based location authentication scheme for crowdsensing applications[J]. Journal on Communications, 2022, 43(9): 121-133.

[33] Zhang J, Zhong X X, Huang Z H, et al. PPTRecS-FL: privacy-preserving task recommendation strategy based on federated learning in mobile crowdsensing[J]. Computer Networks, 2026, 287: 112561.

[34] Xu H T, He Z, Lan D P. Revolutionizing machine learning: blockchain-based crowdsourcing for transparent and fair labeled datasets supply[J]. Future Generation Computer Systems, 2024, 161: 106-118.

[35] 张明瑞, 王涛, 杨波. 一种基于区块链的群智网络框架[J]. 密码学报, 2021, 8(2): 248-259.
Zhang M R, Wang T, Yang B. A blockchain-based framework for crowd intelligence[J]. Journal of Cryptologic Research, 2021, 8(2): 248-259.

[36] 朱向荣, 吴鸿祜, 胡伟. FactChain: 一个基于区块链的众包知识融合系统[J]. 软件学报, 2022, 33(10): 3546-3564.
Zhu X R, Wu H H, Hu W. FactChain: a blockchain-based crowdsourcing knowledge fusion system[J]. Journal of Software, 2022, 33(10): 3546-3564.

[37] Zhang C, Guo Y, Du H W, et al. PFcrowd: privacy-preserving and federated crowdsourcing framework by using blockchain[C]//Proceedings of the 2020 IEEE/ACM 28th International Symposium on Quality of Service (IWQoS). Piscataway: IEEE Press, 2020: 1-10.

[38] Zhang C, Guo Y, Jia X H, et al. Enabling proxy-free privacy-preserving and federated crowdsourcing by using blockchain[J]. IEEE Internet of Things Journal, 2021, 8(8): 6624-6636.

[39] Alagha A, Otrok H, Singh S, et al. Blockchain-based crowdsourced deep reinforcement learning as a service[J]. Information Sciences, 2024, 679: 121107.

[40] He L G, Wu H Q, Li L, et al. UNITE^ +: privacy-aware data quality assessment with public verifiability via federated learning in blockchain-empowered crowdsourcing[J]. IEEE Transactions on Network Science and Engineering, 2026, 13: 522-538.

[41] 邵奇峰, 金澈清, 张召, 等. 区块链技术: 架构及进展[J]. 计算机学报, 2018, 41(5): 969-988.
Shao Q F, Jin C Q, Zhang Z, et al. Blockchain: architecture and research progress[J]. Chinese Journal of Computers, 2018, 41(5): 969-988.

[42] McMahan B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data[C]//Proceedings of the 20th International Conference on Artificial Intelligence and Statistics. New York: PMLR, 2017: 1273-1282.

[43] Liu J X, Deng L W, Miao H, et al. Task assignment with federated preference learning in spatial crowdsourcing[C]//Proceedings of the 31st ACM International Conference on Information & Knowledge Management. New York: ACM Press, 2022: 1279-1288.

[44] Zhong X L, Miao H, Qiu D Z, et al. Personalized location-preference learning for federated task assignment in spatial crowdsourcing[C]//Proceedings of the 32nd ACM International Conference on Information and Knowledge Management. New York: ACM Press, 2023: 3534-3543.

[作者简介]



邬海琴（1992– ），女，博士，华东师范大学副教授，主要研究方向为数据安全、隐私保护和区块链应用安全。



李梁（2001– ），男，华东师范大学博士生，主要研究方向为公钥密码学和区块链应用安全。



李昌峰（1998– ），男，中国科学院计算机网络信息中心助理工程师，主要研究方向为众包安全和区块链应用。