# 边缘计算环境下基于区块链和联邦学习的医疗健康数据共享模型*

邢 丹 徐 琦 姚俊明

（济宁医学院医学信息工程学院 日照 276826）

〔摘要〕 分析医疗健康数据应用情况，介绍区块链与联邦学习技术在医疗健康领域研究现状，提出基于区块链和联邦学习技术的健康医疗共享体系，阐述系统架构和应用流程，为实现医疗健康数据的安全可靠共享和智能处理提供新的解决方案。

〔关键词〕 数据共享；联邦学习；边缘计算；区块链；智慧医疗；共享模型

〔中图分类号〕R－056 〔文献标识码〕A 〔DOI〕10.3969/j.issn.1673－6036.2021.02.007

**Medical and Health Data Sharing Model Based on Blockchain and Federated Learning in the Edge Computing Environment**

*XING Dan, XU Qi, YAO Junming, School of Medical Information Engineering of Jining Medical University, Rizhao 276826, China*

〔**Abstract**〕 The paper analyzes the applications of medical and health data, introduces the study status of blockchain and federated learning technology in the field of medical and health, proposes the health care sharing system based on the fusion technology of blockchain and federated learning, and expounds the system architecture and application process, so as to provide a new solution for the realization of safe and reliable sharing and intelligent processing of medical and health data.

〔**Keywords**〕 data sharing; federated learning; edge computing; blockchain; smart medical; sharing model

## 1 引言

目前我国医疗健康服务数据平台规模大小不一、缺乏统一标准和规范，数据呈现多源结构特征和跨时空特性，质量参差不齐，分散分布，共享困难[1]。医疗健康数据（例如电子病历）中包含大量医疗信息，对其进行分析挖掘可应用于疾病预测、辅助医疗诊断、个性化信息推荐、临床决策支持、用药模式挖掘等[2]。传统采用云计算方式存储和处理健康数据，一方面可能造成云端网络通信开销和负载压力，另一方面由于云计算使用多个医疗健康机构个人数据，可能造成部门间利益冲突和患者个

〔收稿日期〕 2020－06－05

〔作者简介〕 邢丹，硕士，讲师，发表论文 24 篇，参编论著 3 部；通讯作者：徐琦，硕士，副教授。

〔基金项目〕 济宁医学院医学人文素质专项“移动互联环境下医学生人文素质培养模式研究”（项目编号：35）；济宁医学院科研计划项目“高校产教融合校企合作机制创新研究”（项目编号：JY2015RW015）；济宁医学院教师科研扶持基金“移动云环境下医疗健康服务研究”（项目编号：JYFC2018KJ064）；济宁医学院医学人文素质专项“基于移动群智感知的医学生人文素质教育评价研究”（项目编号：34）。

人数据泄漏。为了挖掘不同机构的数据价值，实现碎片化数据共享和融合，应在机构/组织共赢基础上实现数据共享，通过共性技术研发与管理机制创新，促进碎片化局部数据融合[3-4]。大量移动设备、智能可穿戴设备、医疗健康传感器持续产生海量数据，数以亿计用户使用互联网服务，使边缘测数据呈爆炸式增长趋势，促使以数据为驱动的人工智能实施成为可能，但尚存在“数据孤岛”等问题。为了解决云计算出现的问题，本文采用边缘计算模式将数据初步处理放置在边缘计算服务器上，但尚需解决数据隐私和安全保护问题。传统医疗信息系统存在医疗健康数据安全存储和共享难的问题，不同身份人员在访问和共享医疗健康数据时受到较严格限制，且验证身份和数据的真实性需要大量资源和时间[5]。区块链具有匿名、不可篡改、分布式等特征，是一种共享分布式数字分类记账技术，可以更好地管理数据、溯源和保障安全性，可应用于智慧医疗领域[6]。

## 2 研究现状

### 2.1 区块链技术应用

医疗记录管理是区块链技术的重要应用领域之一，区块链医疗记录侧重于管理跨领域医疗数据共享的各参与方，同时保护数据来源、出处和隐私，可以实现更强大的数据和人口健康分析[4]。医院或医疗健康机构的目标是从存储的电子健康档案（Electronic Health Records，EHR）系统中，通过学习模型预测患者的健康风险或可能患有疾病等结果。仅靠一家医院/机构的电子健康档案无法实现通用模型数据学习，医院/机构可以共享其数据以扩大记录数量，但直接共享患者数据存在隐私信息泄漏风险。为此可使用预测建模方法保护隐私，医疗机构可以仅共享部分受训练的机器学习模型（即一组聚合参数）进行协作建立预测模型。上述方法主要采取客户端－服务器的集中化架构，可能导致服务器单点故障。为解决此类问题可将区块链和隐私保护预测模型结合起来，使医院/机构可以协作并训练通用预测模型[7-8]而无需交换患者数据。该解决方案基于区块链技术，用户参与医院或医疗机构进行跨机构模型学习。用户输入数据是来自EHR的患者级数据，具有相同格式和语义。通过交换模型可解决隐私保护等问题，避免单点故障并生成学习过程的不变日志。在数据共享方面，模型及其元信息（例如模型局部训练错误）在链上共享，而不进行链下数据共享。在治理方面，只有参与医院和机构才包含在区块链网络中。最终基于区块链的学习目标健康记录管理方法包括支持比较有效性研究、生物医学研究以及最终患者护理。

### 2.2 联邦学习

2.2.1 基本概述 谷歌公司于2016年率先提出基于移动设备的联邦学习技术，其借助移动设备进行本地模型训练，避免原始数据移动带来的弊端。联邦学习是新兴机器学习技术，使用本地模型进行分布式模型训练大型节点（例如移动设备）共享数据集，实现仅更新模型而不上传原始训练数据。基于该技术能够提供隐私保护设备同时提高学习性能[3]。现有大部分工作集中在设计高级学习算法阶段以获得更好的学习效果。有研究者提出一种有效的激励机制，在合同理论中享有声誉，激励高声誉且具有高质量数据的移动设备参与模型学习。方案可显著提高联邦学习准确性[3]。

2.2.2 存在问题 联邦学习在通讯参数频繁传输带来链路传输开销、参与用户互信、参与方提供参数的质量验证、参数传输及存储隐私性等方面存在问题。区块链可以增强数据安全性、共享性、互操作性和完整性并可实现实时更新和访问，区块链和智能合约有望提供解决方案，通过共享和访问保护患者电子病历数据[5]。针对上述问题，本文将区块链与联邦学习相结合，建立一种安全可靠、智能隐私的机制。

## 3 基于区块链和联邦学习融合技术的健康医疗共享框架

### 3.1 联邦学习模型

联邦学习是一种分布式隐私保护机器学习技

术，无需将移动设备本地私有数据上传至中心服务器就可共同训练全局模型。每个移动设备从任务发布者处获取全局共享模型，在其本地数据基础上训练模型。移动设备将新的权重或梯度上传给任务发布者，用来更新全局模型。联邦学习的目标，是通过实现每个移动设备本地数据损失函数的平均权重最小最终达到全局损失函数最优。本地训练数据具有高精确性和可靠性，可在少量的训练时间和能量消耗情况下实现高效的学习效果[3]。训练模式下每个节点彼此独立且享有本地数据控制权，服务器端不必直接访问各节点中的本地数据，仅需在参数层面进行模型整合与发布。

## 3.2 系统设计方案

为了保证模型训练中参数安全性，实现较少通信开销和提高计算效率，采用区块链存储模型训练参数，通过区块链为联邦学习各参与方提供一种可信机制。联邦学习模型参数可存储在区块链中，保证其安全性与可靠性；以联邦学习方式在边缘环境构成的医联体内训练模型，对原始医疗健康数据进行处理，仅存储模型训练计算参数，降低区块链存储资源开销，还可通过联邦学习对区块链交易的认证计算、传输通信等进行优化，提升区块链运行效率。

## 3.3 系统架构组成

3.3.1 概述 主要分为用户层和边缘服务层，用户层主要由物联网设备、移动终端组成；服务端主要由配备移动边缘计算服务器并具备一定存储与计算能力的基站构成。按照功能可划分为联邦学习层、区块链层，见图 1。联邦学习本地训练运行在用户侧，依据用户侧数据学习本地模型参数。区块链则运行在边缘服务侧，接收并存储联邦学习模型参数，通过共识协议对参数进行认证。

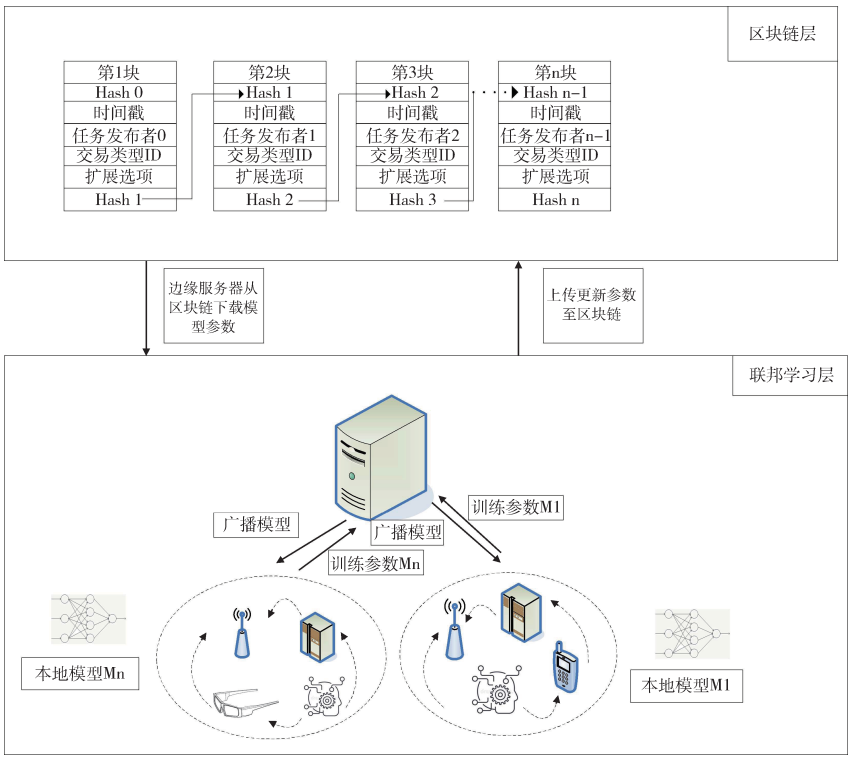


**图 1 区块链和联邦学习融合的医疗健康共享架构**

3.3.2 联邦学习层 该层由无线通信组成的通用移动网络基础设施（例如具有存储与边缘计算功能的基站）和一组移动设备用户构成（例如智能可穿戴传感器设备、手机、平板电脑等）[9]。可穿戴式传感器设备通过对患者进行连续生理监测而提供生理锻炼和饮食习惯调整建议。可在医联体内存储多

名患者健康数据，不仅能够帮助医生判断患者健康状况、提出治疗建议和进行早期诊断，还可以通过对数据的智能处理进行群体疾病预测。广泛部署的通信基础架构可视为边缘节点，配备的计算和通讯设备可从移动应用程序生成各种用户健康数据并收集传感器健康监测数据。模型训练流程为：一是区块链中的任务发布者根据实际应用（对慢性病患者的疾病预测）下发要训练的模型参数。二是用户本地训练，服务器端接收训练任务后，用户在本地根据持有的医疗健康数据，利用智能算法寻找模型参数，每个用户反复训练共享全局模型Φ，通过其局部数据生成局部模型更新Φn。三是服务器端参数收集，所有用户将其本地模型通过无线网络传输至边缘服务层，边缘服务器上传至区块链，任务发布者从区块链获取参数并更新全局模型，基站收集来自各用户参数，以交易的形式存储在各区块链节点。训练重复此过程，直到全局模型准确性达到预定期望值为止。分布广泛边缘节点使用户能够与任务及时进行通信。四是产生交易区块，在区块链层，各参与节点收集来自用户层的模型参数，加密签名打包进区块。节点间通过运行共识机制决定块权所属。获得出块权的节点将区块广播至全网，认证通过后加入区块链。五是模型聚合，任务发布者的聚合节点依据区块链记录聚合模型参数并更新整体模型。进一步将该模型下发至各参与用户开始新一轮训练学习。

3.3.3 区块链层 与传统集中式数据库不同，区块链上数据可以分布在多个数据库或计算机节点上，以便各用户持有相同交易副本。数据“块”通过数字、随机字母和数字签名构成的散列以形成包含完整记录的“数据链”交易，使其具有防篡改功能。区块链数据通过加密技术保护，参与者可以信任“数据块”是经过身份验证和可验证的。以上技术特征保证数据采用分布式处理并且具有较高可信任度，同时允许区块链限制参与和访问或数据共享。数据块通过数据加密哈希链接在一起。区块链结构中主要包括哈希值、时间戳、任务发布者、交易类型、扩展选项。扩展选项主要包括区块链类型、共识机制、权限结构类型、数据存储位置、区块链治理等相关信息。

### 3.4 系统优点

3.4.1 数据认知功能强大 基于移动边缘计算环境的移动医疗场景下，医疗健康物联网设备可以将复杂任务交给边缘服务器节点，平衡通信和计算性能。大量医疗健康智能设备和边缘节点可以充分感知和获取丰富和个性化医疗健康数据以用于模型训练。基于原始数据使用联邦学习技术而不是采用集中式处理方式，表现出移动边缘计算的强大认知功能。边缘节点还可获取包括计算负载、存储空间、无线通信数量、任务队列状态等数据。

3.4.2 稳健性良好 与数据并行化训练方式相比联邦学习无需具有独立同分布的数据样本，且在各节点的数据量不平衡的情况下依然可以进行边缘模型训练。还可在超大规模无线环境下处理处理非标准的数据。

3.4.3 灵活性高 在联邦学习模块中可使用其他计算以减少交流轮训的次数模型。增加计算的有效方法之一是添加每轮本地随机梯度下降算法（Stochastic Gradient Descent，SGD）训练次数。为了进一步减少通讯费，具有强大计算和能量的用户设备可以决定执行更多批次训练。

3.4.4 安全性强 区块链具有安全、可信等特点，联邦学习具有分布式智能、保护数据隐私等特点，两者进行优势互补提升系统整体安全性。

## 4 结语

分析现有医疗健康数据共享难、存在数据孤岛等问题，采取联邦学习保证数据不出本地即可实现数据共享和使用；利用区块链对联邦学习参数进行存储及认证，提高联邦学习安全性与可靠性。提出系统的层次性架构，包括区块链和联邦学习层，明确模型训练流程。利用区块链的分布式、防篡改特点实现医疗健康数据安全存储和共享并降低管理成本，同时可加强对碎片化数据的利用，为数据挖掘和临床决策提供支持和辅助诊断。但由于联邦学习处于起步阶段，尚存在通信带宽受限、需提高模型

收敛速度、移动设备海量数据存储受限等问题[10]，在区块链中选择联邦学习边缘服务器算法方面仍需研究，可尝试采取具有激励机制、基于声誉的服务器选择算法[11-13]。

## 参考文献


1 冯涛，焦滢，方君丽，等．基于联盟区块链的医疗健康数据安全模型［J］．计算机科学，2020，47（4）：305-311.

2 胡满满，陈旭，孙毓忠，等．基于动态采样和迁移学习的疾病预测模型［J］．计算机学报，2019，42（10）：2339-2354.

3 Kang J，Xiong Z，Niyato D，et al. Incentive Mechanism for Reliable Federated Learning：a joint optimization approach to combining reputation and contract theory［J］. IEEE Internet of Things Journal，2019，6（6）：10700-10714.

4 Tim K Mackey，Tsung-Ting Kuo，Basker Gummadi，et al. 'Fit-for-purpose?'-challenges and opportunities for applications of blockchain technology in the future of healthcare［J］. BMC Medicine，2019，17（1）：1-17.

5 Hussien H M，Yasin S M，Udzir S NI，et al. A Systematic Review for Enabling of Develop a Blockchain Technology in Healthcare Application：taxonomy，substantially analysis，motivations，challenges，recommendations and future direction［EB/OL］.［2019-09-14］. https：//doi.org/10.1007/s10916-019-1445-8.

6 华中生，刘作仪，孟庆峰，等．智慧养老服务的国家战略需求和关键科学问题［J］．中国科学基金，2016，30（6）：535-545.

7 Wang S，Jiang X，Wu Y，et al. Expectation Propagation Logistic Regression（explorer）：distributed privacy-preserving online model learning［J］. J Biomed Inform，2013，46（3）：480-496.

8 Wu Y，Jiang X，Kim J，et al. Grid Binary Logistic Regression（GLORE）：building shared models without sharing data［J］. JAMIA，2012，19（5）：758-764.

9 张彦，卢云龙，黄小红．区块链与联邦学习：融合与互补［J］．中国计算机学会通讯，2020，16（2）：17-22.

10 刘俊旭，孟小峰．机器学习的隐私保护研究综述［J］．计算机研究与发展，2020，57（2）：346-362.

11 Patel，Vishal. A Framework for Secure and Decentralized Sharing of Medical Imaging Data via Blockchain Consensus［EB/OL］.［2018-04-30］. https：//doi.org/10.1177/1460458218769699.

12 Dagher G G，Mohler J，Milojkovic M，et al. Ancile：privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology［EB/OL］.［2018-05-31］. https：//doi.org/10.1016/j.scs.2018.02.014.

13 Chen L，Lee W K，Chang C C，et al. Blockchain Based Searchable Encryption for Electronic Health Record Sharing［J］. Future Generation Computer Systems，2019，95（6）：420-429.