

联邦服务：基于分布式数据共治的智能服务范式

贾晓丰¹, 秦蕊^{2,3}, 王寿文⁴, 齐红威⁵, 王飞跃^{3,4}, 李娟娟^{2,3}, 章敏¹, 梁小龙^{2,3}

(1. 北京市大数据中心, 北京 100101;

2. 中国科学院自动化研究所多模态人工智能系统全国重点实验室, 北京 100190;

3. 中国科学院自动化研究所复杂重点系统管理与控制国家实验室, 北京 100190;

4. 澳门科技大学创新工程学院, 澳门 999078;

5. 数据堂(北京)科技股份有限公司, 北京 100192)

摘要: 随着数字化转型的加速, 数据已成为驱动科技创新与经济增长的核心要素。然而, 广泛存在的数据孤岛现象严重阻碍了跨领域、跨组织间的数据流通与合作, 使得数据供应难以满足多样化的用户需求, 数据的供需匹配面临严峻挑战。作为联邦生态的核心环节和对外赋能的重要窗口, 联邦服务基于分布式数据共治原则, 促进了跨领域、跨组织的数据流通与协同, 为智能的供需匹配提供了一种有效的技术框架。首先, 探讨联邦服务在联邦生态中的重要地位以及与联邦生态各环节的关系。其次, 提出联邦服务的5层基本框架, 并探讨其核心优势。在此基础上, 提出联邦服务的关键支撑技术, 主要包括区块链和去中心化自治组织与运营、大语言模型和场景工程、分布式计算和边缘计算、加密技术和隐私计算。最后, 以数据与场景的供需匹配为应用案例, 介绍联邦服务如何实现智能的供需匹配。提出的联邦服务理论与技术对于构建基于分布式数据共治的智能服务范式具有重要意义。

关键词: 联邦服务; 区块链; 去中心化自治组织与运营; 大语言模型; 去中心化物理基础设施网络; 联邦数据
中图分类号: C931

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202416

Federated services: smart service paradigm based on distributed data co-governance

JIA Xiaofeng¹, QIN Rui^{2,3}, WANG Shouwen⁴, QI Hongwei⁵, WANG Fei-Yue^{3,4},

LI Juanjuan^{2,3}, ZHANG Min¹, LIANG Xiaolong^{2,3}

1. Beijing Big Data Centre, Beijing 100101, China

2. The State Key Laboratory of Multimodal Artificial Intelligence Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

3. The State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

4. Faculty of Innovation Engineering, Macau University of Science and Technology, Macao 999078, China

5. Datang (Beijing) Technology Co., Ltd., Beijing 100192, China

Abstract: The acceleration of digital transformation has made data a core element driving technological innovation and economic growth. However, the widespread phenomenon of data island severely hinders data circulation and collaboration across different fields and organizations, making it difficult for data supply to meet the diverse needs of users, thus creating a dilemma in matching data supply and demand. Serving as the core link in the federated ecology and an impor-

收稿日期: 2024-03-02; 修回日期: 2024-05-24

通信作者: 秦蕊, rui.qin@ia.ac.cn

基金项目: 国家重点研发计划项目 (No.2021YFB2104800); 国家自然科学基金项目 (No.62103411); 澳门特别行政区科学技术发展基金项目 (No.0093/2023/RIA2, No.0050/2020/A1)

Foundation Items: National Key Research and Development Program of China (No.2021YFB2104800), The National Natural Science Foundation of China (No.62103411), The Science and Technology Development Fund of Macau SAR (No.0093/2023/RIA2, No.0050/2020/A1)

tant window for external empowerment, federated services provide an innovative solution for data circulation and utilization under the principle of distributed data co-governance, thereby enabling the intelligent matching of supply and demand. Firstly, this paper delves into the pivotal role of federated services within the federated ecology, and examines the relationship between federated services and federated ecology's key components. Secondly, it presents a five-layer basic framework of federated services and discusses its core advantages. On this basis, the paper proposes the key supporting technologies for federated services, including blockchain and decentralized autonomous organizations and operations, large language models and scenario engineering, distributed computing and edge computing, as well as encryption technology and privacy computing. Finally, taking the supply-demand matching problem of data and scenarios as a case study, the paper illustrates how federated services achieve intelligent matching of demand and supply. The federated services proposed in this paper holds significant implications for constructing a smart service paradigm based on distributed data co-governance.

Key words: federated services, blockchain, decentralized autonomous organization and operation, large language models, decentralized physical infrastructure networks, federated data

0 引言

在当今技术迅猛发展和加速迭代的时代背景下，大数据、通用人工智能以及大语言模型等技术的广泛应用，已使数据成为各行各业的创新发展与战略决策服务的核心要素，其内在价值与日俱增^[1]。国家高度重视数据要素的发展，2022年12月19日发布的《中共中央 国务院关于构建数据基础制度更好发挥数据要素作用的意见》明确指出，要加快构建数据基础制度，充分发挥我国海量数据规模和丰富应用场景优势，激活数据要素潜能，做强做优做大数字经济，增强经济发展新动能，构筑国家竞争新优势。

然而，数据供给与场景需求之间匹配失衡以及数据流通机制不畅通，导致丰富的数据资源难以释放其巨大的实际价值^[2]。一方面，尽管数据总量持续增长，但由于数源的分散性与数据的异质性、场景的交叉性与规则的多变性、隐私的敏感性与数据的安全性等问题，这些数据往往难以精准契合特定场景的具体要求，这就增加了数据利用的复杂度与难度^[3-4]。另一方面，应用场景的多元化与需求的多样化，对数据融合与服务供给提出了更高要求。物理世界强调数据与实际需求的无缝整合，而数字世界的多场景耦合数据引发严峻的语义异构问题，这就导致数据供给与场景需求的“语境错位”问题，拉大了物理世界与数字世界之间的数据交流鸿沟。以交通领域的道路拥堵场景为例，交通拥堵可能由交通事故、恶劣天气、医院和学校等人员聚集路段、道路施工、临时交通管制以及各种突发状况等导致。解决道路拥堵问题不仅需要交通数据，还需要天气

数据、医疗和教育资源的分布数据、道路施工规划数据等多种数据源。这些领域的交叉使得数据的规模指数级增长，且数据间的关系变得更加错综复杂。

为了消除数据供给与应用场景需求之间的巨大鸿沟，国家数据局会同中央网信办、科技部等17个部门在2023年12月31日联合印发了《“数据要素×”三年行动计划（2024—2026年）》，指出以推进数据要素协同优化、复用增效、融合创新作用发挥为重点，强化场景需求牵引，充分发挥数据要素乘数效应。作为推动数据要素应用创新的关键技术，联邦服务基于分布式数据共治逻辑构建智能服务范式，在保证数据隐私与安全的前提下，为消除数据供给与应用场景需求之间的鸿沟提供一种有效的技术框架^[5]。联邦服务通过区块链^[6-7]、去中心化自治组织与运营（decentralized autonomous organizations and operations, DAO）^[8-10]、加密技术、大语言模型（large language model, LLM）^[11-12]、场景工程等技术，不但能够在确保数据隐私与安全的前提下突破数据孤岛的限制，促进跨域数据的流通与协同，而且能够通过灵活的数据融合与定制化服务，精准对接复杂多变的应用场景需求。联邦服务既能跨越物理与数字世界界限，又能精准适配个性化与多元化需求，从而实现供需双方数据资源的精准匹配。

基于此，本文提出联邦服务的概念和框架。作为联邦生态的核心环节和联邦生态对外赋能的窗口，联邦服务通过先进的技术架构与协作模式，为跨越多个组织边界的多样化场景高效赋能。与传统服务模式相比，联邦服务不再依赖于数据的集中存储与处理，而是在保持数据不出本地的情况下，通

过区块链、加密技术、隐私计算等技术实现数据的智能分析和决策支持。联邦服务突破了传统服务模式的局限,通过数据所有权和使用权的分离,保护数据提供方的数据隐私与安全,以“数据可用不可见”的方式解决数据孤岛问题,实现了数据的安全共享与交互,为解决复杂多变的现实世界数据供需匹配难题提供了一种有效的技术框架^[5]。

本文的结构安排如下:第1节概述联邦服务与联邦生态各环节的关系;第2节提出联邦服务的基本框架,并介绍其核心优势;第3节介绍联邦服务的关键支撑技术;第4节给出联邦服务的应用案例;第5节对本文工作进行总结。

1 联邦服务概述

随着大数据技术的广泛应用,跨组织、跨领域、跨行业的数据流通与协同需求日益迫切^[13-14]。然而,数据流通的便捷性也增加了数据隐私泄露风险与数据安全威胁。如何在保障数据隐私权益和维护数据安全的同时促进数据的有效利用,成为亟待解决的核心问题。

联邦服务作为联邦生态^[15]对外赋能的重要窗口,旨在通过先进的技术架构与协作模式,在保持数据隐私和数据主权的前提下,构建一个安全、合规、高效的数据协作体系,实现分布式数据的协作与利用,为跨越多个组织边界的多样化场景高效赋能,从而使数据价值最大化。联邦服务突破了传统服务模式的局限,是一种全新的生态型服务理念,为解

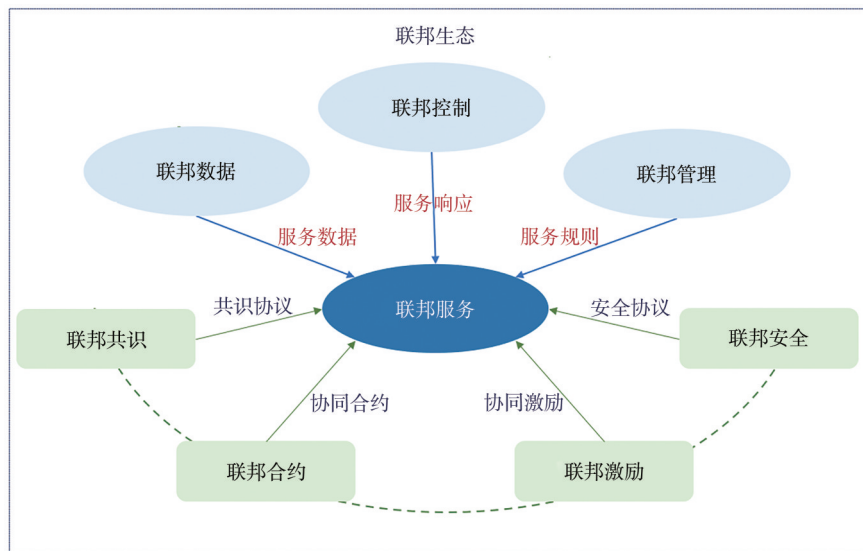
决复杂多变的现实世界问题提供了新的思路 and 工具。

针对特定场景与需求,联邦服务依托于联邦数据^[16],通过联邦控制^[17-18]响应联邦服务需求,采用联邦管理^[19]管理联邦服务过程,利用联邦合约自动化数据访问和非隐私数据的交互过程,依靠联邦共识机制确保分布式决策的一致性,通过联邦激励机制提升节点贡献数据与算力的积极性,并利用联邦安全保障数据在传输与处理过程中的隐私性与完整性。联邦服务与联邦生态各环节的关系如图1所示。

联邦数据作为联邦服务的基石,可以确保联邦服务过程中数据的隐私性与安全性,以及数据所有权不受侵犯。通过加密技术,联邦数据在不出本地的情况下实现数据的安全传输与处理,并且只有计算结果或中间特征被共享,不会暴露敏感信息,防止了数据流通过程中的隐私泄露,为联邦服务提供充足的数据支撑。

联邦控制负责响应联邦服务需求,是确保联邦服务过程中联邦数据流通和协作的关键^[20]。联邦控制根据联邦服务需求,动态调度联邦数据并管理其数据权限,实现了对联邦数据的访问控制,通过隐私数据不出本地、非隐私数据使用权和所有权分离的方式,保障联邦服务过程中隐私联邦数据的安全性以及非隐私联邦数据的原始性和归属权不受侵犯。

联邦管理负责联邦服务规则与协作框架的制定,通过法律法规、经济效益和技术实践等多个维度的融合,将服务需求精准转化为既具有可操作性又遵循法律框架的具体协作框架与执行规则体系,



保证联邦服务在安全稳定的前提下高效合规地运转，并通过联邦激励使节点积极参与，从而促进跨区域数据流通与服务协同，实现收益最大化。

2 联邦服务的基本框架及其优势

为了有效应对海量分布式数据供给与复杂多样化问题需求的匹配挑战，联邦服务采用分布式智能框架，将传统集中式的数据管理模式和业务功能服务逻辑与处理逻辑分散到联邦生态网络中的多个节点上，每个节点都能独立执行特定的任务并与其他节点协同工作，共同完成复杂的系统功能，形成全局的联邦智能。在这种架构下，联邦服务通过区块链网络连接分布式节点，并融合以大模型为代表的基础智能技术以及以场景工程为代表的设施智能技术，通过高效的数据分析、精准的数据理解以及安全的数据交换，实现数据共治前提下的问题智能分析、方案灵活协同与服务敏捷调配，从而赋能多维度共生、跨领域耦合的复杂应用场景，并兼顾高度整合的服务功能与个性化需求。

2.1 基本框架

按照联邦服务的功能模块及其与联邦生态各环节的关系，将联邦服务的分布式智能框架的核心要素划分为基础层、设施层、数据层、管控层以及应用层。联邦服务的基本框架如图2所示。这5层通过标准化接口和协议紧密耦合，下层为上层提供支撑，上层需求又反向驱动下层技术的创新与优化，

形成了一个安全可信、动态灵活和可持续的联邦服务系统。

(1) 基础层

基础层为联邦服务网络提供底层技术、算法与协议，以支撑网络通信、计算资源管理、数据存储与处理、模型训练与信息交换等，主要包括加密算法、多模态大模型、分布式计算框架以及联邦学习算法等。基础层能够保障分布式联邦服务网络通信的高效性、计算资源管理的灵活性、数据交互的安全性以及服务功能的稳定性，从而为完整服务功能的实现与创新提供基础技术支撑。

(2) 设施层

设施层为联邦服务网络提供软硬件基础设施，以支撑联邦生态的稳定运行与服务的高效交付，主要包括区块链、场景工程、去中心化物理基础设施网络（decentralized physical infrastructure network, DePIN）及其连接的物理世界基础设施（如物联网设备、传感器网络等）以及数字世界的数字资源（如数据资源、边缘计算资源等）。设施层为联邦服务网络构建可靠且可扩展的运行环境，从而为应用创新提供必要的物理与数字基础设施。

(3) 数据层

数据层专注于联邦服务网络中的联邦数据治理，利用数据目录链、联邦数据协议、数据共治机制等，实现覆盖数据采集、存储、计算、流转等全生命周期的治理，确保在保护数据隐私和权属前提

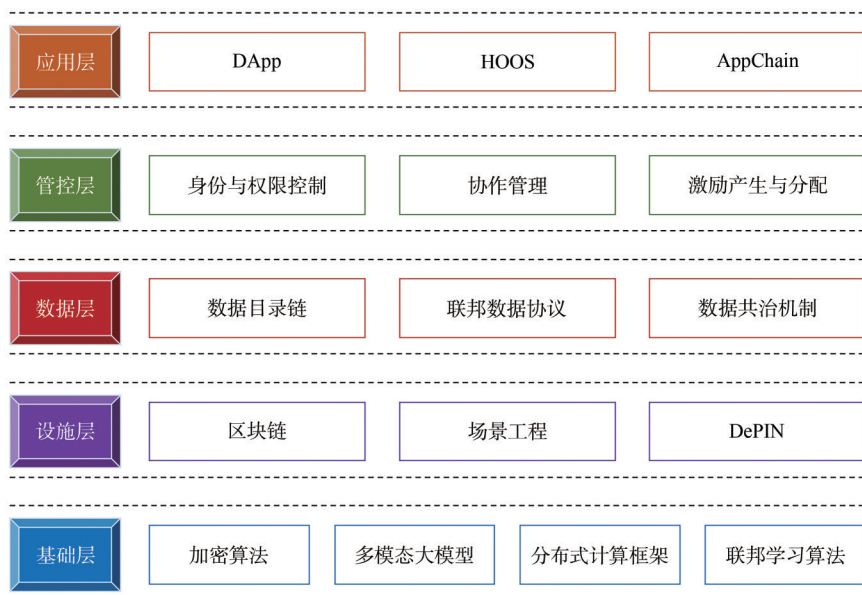


图2 联邦服务的基本框架

下的跨域流转与高效交互。数据层为联邦服务网络提供丰富、高质量的数据,为服务场景提供必要的知识基础和决策依据。

(4) 管控层

管控层专注于联邦服务网络中的协作任务管控,利用基于智能合约与DAO的身份与权限控制、协作管理、激励产生与分配等,实现开放服务网络中面向复杂场景的数据与任务无边界多方协同,从而支撑多对多的去中心化服务供需匹配与优化。管控层旨在实现面向联邦服务的联邦控制与联邦管理,为联邦服务群体提供必要的协作框架与规则。

(5) 应用层

应用层直接面向最终用户,通过联邦服务注册与调用、智能服务编排、场景化应用定制等将下层的各项功能转化为解决实际业务的方案,支持多样化的应用场景,是联邦服务对外赋能的窗口。应用层通过构建去中心化应用(decentralized applications, DApp)、面向人的操作系统(human-oriented operating systems, HOOS)^[21]以及应用链(application-specific blockchain, AppChain)等,打破传统应用的中心化瓶颈,提供用户友好且高效透明的服务接口,在保障安全和公平的前提下使用户接入并获取联邦服务资源。

2.2 核心优势

联邦服务融合以区块链为代表的可信技术、以大模型为代表的智能技术以及以DePIN为代表的物理世界与数字世界虚实融合技术^[22],构建了多元、共治和可持续的服务生态,不仅推动了跨域数据的安全共享与可信融合,还在此基础上构建了激励驱动的多边服务资源协同与主体协作网络,从而灵活、智能地将分散数据转化为复杂场景中多样化服务需求所需的知识、决策与功能。

(1) 数据安全与可信

联邦服务基于联邦数据模式,允许各个参与节点在不泄露原始数据的情况下共享信息和模型参数,在确保数据隐私的同时实现跨组织的数据协作与模型训练。首先,通过隐私保护技术、加密算法以及分布式计算框架等,实现在隐私与权属保护前提下的数据共享与交互,增强了联邦服务的信任基础,在避免数据集中风险的同时解决了数据孤岛问题。其次,联邦管理与联邦控制能够很好地支撑数据访问和使用的细粒度管理,减少了潜在的信息泄露风险与合规性隐患,从而强化了联邦服务所需的

数据安全防线。此外,基于DAO和智能合约的数据共治机制,使得服务相关方共同参与数据治理过程,并通过智能合约自动执行治理规则与协议,使得联邦服务能够在无须中心化权威干预的情况下,实现数据资源的透明与公平管理,提升了数据治理的效率与可信度。联邦服务从根本上改变了传统数据的治理模式,为在高度分散环境下实现数据价值并支撑多元化服务开辟了可行路径。

(2) 资源协同与主体协作

联邦服务是一个面向资源协同与主体协作的生态型服务网络,鼓励不同主体在保护自身利益的同时,利用自身资源积极参与价值共创的服务生态。一方面,通过区块链、DAO以及DePIN等技术,联邦服务不仅促进了数字世界中数字资源的无缝协同,还高效整合了物理世界中的实体资源,使得物理世界的实体资产、设备与设施能够与数字世界的数据、算法和服务紧密对接。这种深层次的融合不仅优化了虚实资源配置,还促进了实体产业与数字经济的协同发展。另一方面,联邦服务的协同与协作不仅体现在技术层面,而且根植于其治理理念中。综合利用区块链与AI技术的优势,构建多方参与、共识驱动以及虚实融合的综合协同与协作生态,有助于突破复杂、多变的服务决策面临的知识局限与决策瓶颈。

(3) 服务智能化与灵活性

联邦服务通过集成多模态大模型、大数据分析技术等,提供动态灵活的智能解决方案。首先,联邦服务采用分布式智能架构,利用模块化设计和接口标准化的DApp、内嵌大模型且能够精准理解自然语言的HOOS以及面向特定应用场景的AppChain,支持快速集成多样化的服务组件和外部资源,提高了服务的人机交互智能化水平,并加快了复杂服务需求的响应速度。其次,联邦服务可以为不同粒度的需求提供从单一设备级到跨行业协作的解决方案,覆盖视觉、语音、文字等多模态服务,适应不同规模和复杂性的应用场景。此外,联邦服务将供需双方纳入统一的协作框架中,利用标准化的语义映射机制与策略,将数据供给和服务需求转化为高度互操作的语义规范以及合约规则,在不直接交换原始数据的情况下也能确保信息的准确传达与应用,从而解决复杂服务场景面临的供需语义错位问题,提高了服务质量和效率。

3 关键支撑技术

联邦服务依赖于区块链与DAO、LLM与场景工程、分布式计算与边缘计算、加密技术与隐私计算等技术，通过这些技术的协同，实现数据隐私与安全下的智能供需匹配。

(1) 区块链与DAO

区块链与DAO作为联邦服务生态系统的核心技术，实现了联邦节点间的身份验证与权限控制，提升了数据流通的透明性、安全性和隐私保护，并推动了节点间的高效协同与协作^[23-24]。

区块链技术具有难以篡改、透明和可追溯等优势，为联邦服务生态构建了一套去中心化的身份验证体系。为每个联邦节点生成独一无二的去中心化身份标识及数字签名，确保了联邦服务生态中每一个参与成员身份的真实性与可靠性。针对每一次数据交换或服务请求，借助区块链与智能合约迅速验证对方身份，实现用户权限的自动化管理与数据流通协议的自动执行，确保数据交互和操作在经验证的身份框架内合法、安全地进行，并确保数据流转历史完全可追踪，增加了整个联邦服务生态的透明度和可信度。

DAO采用去中心化的民主治理结构，为联邦生态内的多元参与主体提供了一个无须依赖中心化权威的去中心化共治平台，促进了跨组织、跨部门乃至跨领域、跨行业的联邦节点之间资源的高效整合与协同，显著加快了联邦服务的响应速度并提高了执行效率。此外，DAO激励机制通过表彰为联邦服务贡献高质量数据或资源的节点，激励更多节点积极参与，从而促进了数据或资源的积极贡献与协作共享。这一机制构建了一个多方共赢、共识导向的合作与协作网络，使得联邦服务能够动态适应并智能地应对复杂多变的服务场景挑战，最大化服务的灵活性与效能。

(2) LLM与场景工程

LLM与场景工程的融合，不仅极大地提升了联邦服务的个性化、定制化和智能化水平，还显著增强了联邦服务的灵活性、多样性和场景适应性，从而优化用户体验与提高服务效率。

LLM^[25-26]凭借其强大的自然语言处理与理解能力，准确捕捉并深入理解用户的需求，并通过分析用户的历史行为与交互模式，构建用户的精准画像，从而为每位用户提供量身定制的个性化联邦服

务方案，满足用户多样化的服务需求。场景工程^[27-29]可以根据用户的特定需求，设计与之对应的服务需求场景，利用所构建的场景设计并优化联邦服务方案，确保联邦服务方案与用户实际需求的精准匹配。利用LLM的自适应学习机制和场景工程的实时动态自适应调整能力，联邦服务通过持续学习和交互反馈，灵活地调整与优化其服务方案，确保服务内容的时效性和智能性。

此外，个性化与定制化的可插拔服务组件的灵活配置、升级或替换，使联邦服务能够敏锐地识别、感知并理解用户跨行业、跨领域的多样化需求，并为具体需求场景提供高度定制化的接口和交互方式以及一系列灵活多变、具有高度针对性的解决方案，从而提升联邦服务的广泛性和场景适应性。

(3) 分布式计算与边缘计算

分布式计算与边缘计算技术可以显著加快联邦服务响应速度，降低数据传输成本，提升联邦服务的实时性、稳定性和高效性。

分布式计算技术具有强大的并行处理能力^[30]，通过将联邦服务中复杂的计算任务分散到多个节点上并行处理，可以协调联邦服务中多个参与者的计算资源，大幅减少数据在不同地理位置间的传输需求，减少数据在网络中的传输时间和成本，有效控制网络延迟，极大地增强了联邦服务的可扩展性和弹性。此外，分布式计算的动态负载均衡特性确保了资源的高效配置，各节点可根据实时负载自动调整任务量，从而提升联邦服务的稳定性。

边缘计算将模型训练和数据处理任务部署在附近边缘设备或节点，使得智能分析和决策过程能够在边缘设备上实时进行^[31-33]，极大地缩短了数据分析与模型训练的时间，为联邦服务提供实时的数据分析与决策反馈，使联邦服务能够快速适应环境变化。此外，边缘计算还可以促进联邦服务过程中模型的轻量化与本地化训练，使联邦服务能够利用最新数据快速进行自我迭代与优化，以精准匹配个性化的用户需求和应对瞬息万变的市场环境。这能加快联邦服务的响应速度并提升其性能，减少数据传输导致的延迟和数据丢失风险，从而提高了联邦服务的实时性和高效性。

(4) 加密技术与隐私计算

加密技术与隐私计算在保证联邦服务过程中数据隐私与安全的关键技术，可以确保模型训练与数据分析在数据保密性、完整性与可用性的前提下进

行, 能够提升联邦服务的安全性和隐私保护水平。

加密技术可以为联邦服务中数据和模型的参数保护提供重要支撑。采用同态加密技术, 数据在加密状态下仍能直接参与计算过程, 使得联邦服务的各参与方能够安全地上传加密的模型参数, 而服务端则可直接在加密数据上执行聚合操作, 无须解密, 实现了“数据可用不可见”。此外, 端到端加密和密钥管理机制可以进一步增强数据传输和存储的安全性, 数据即便在传输过程中被截获, 也无法被非法访问, 从而极大增强了联邦服务中数据交互的隐私性。

隐私计算^[34]可以为联邦服务中的多方数据协作提供有效解决方案。多方安全计算、差分隐私等手段可以确保各参与方在不泄露自身敏感数据的前提下协同进行数据分析和模型训练。具体而言, 多方安全计算允许多个参与方在保持数据加密的状态下执行联合计算任务, 仅共享汇总的计算结果, 避免了其他参与方直接访问数据内容。差分隐私技术在联邦服务的模型训练和数据分析阶段为输出结果添加随机噪声, 使攻击者无法通过分析模型参数逆推出原始数据的具体信息, 从而保护了联邦服务过程中的数据隐私与安全。

4 联邦服务的应用场景

4.1 根本问题: 数据与场景的供需匹配

数字世界的的数据与物理世界的应用场景之间的供需匹配, 是数据为场景赋能的关键。然而, 供需双方的“语义错位”, 严重制约了二者之间数据的高效流通与协同。数据与场景之间的供需匹配难题的根源主要包括3个方面, 如图3所示。

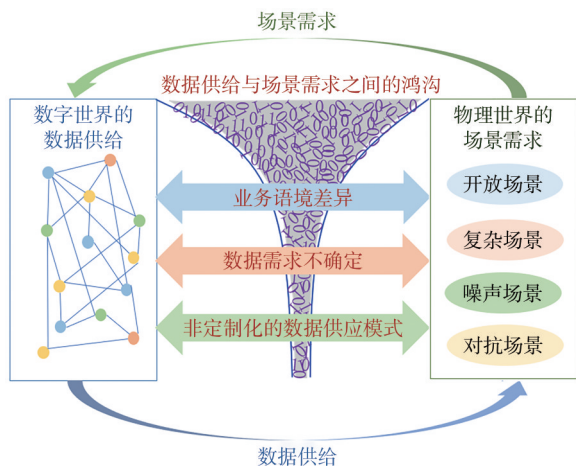


图3 数据与场景之间的供需匹配难题

(1) 业务语境差异

供需双方业务语境的独特性与差异性是导致数据语义错位的首要因素。数据供应方与场景需求方在行业知识、业务逻辑及技术认知上往往存在显著差异, 这导致了供需双方在数据理解方面的语义错位, 即数据产生场景的语义和使用场景的语义存在不同。例如, 在金融风控领域, 数据供应方可能将“交易频率”理解为单一账户每日交易次数, 而场景需求方(如银行)则将“交易频率”刻画为客户群体在特定时间段的活跃度, 包括但不限于交易次数、交易金额、类型等。这种业务语境的差异容易导致数据应用的误解和偏差, 进而影响供需匹配的准确性和效率。

(2) 数据需求不确定

数据需求的不确定性是导致语义错位的另一个关键因素。即使场景需求方能够清晰把握其业务的核心场景需求, 但他们往往难以确定实现这些业务目标所需的数据集。数据转化为业务成效路径的不确定性, 使得需求方难以将其业务逻辑中的隐性知识显性转化为具体的数据需求。与此同时, 数据供应方由于对需求方业务模型缺乏深刻认识, 难以跨越知识与理解的鸿沟将高层次的业务需求转化为精确的数据需求。以电商平台为例, 需求方的目标是提升商品个性化推荐的水平以提高商品转化率, 但其不确定实现这一目标所需的数据。而数据供应方因缺乏对个性化推荐逻辑的深入理解, 可能仅提供常规的用户行为和商品浏览数据, 却忽略了用户的实时兴趣波动、特定时段的购买偏好等关键数据, 从而难以满足需求方的场景需求。供需双方在数据需求表达上的不确定性, 在市场环境快速演变、业务场景需求持续变化的情况下显得尤为突出。数据需求的不确定性进一步加剧了数据应用的语义错位问题, 导致即使有大量数据, 也难以精准匹配并为目标业务场景高效赋能。

(3) 非定制化的数据供应模式

非定制化的数据供应模式也是导致供需双方语义错位的重要因素。这主要是因为数据供应方遵循的通用标准与需求方特定业务场景要求不协调。数据供应方通常基于其内部业务逻辑生成、分类和组织数据, 并非针对需求方的特定业务场景进行定制, 这使得供应方提供的数据在结构设计、细粒度处理及信息关联性上, 往往与需求方的实际场景需求不匹配, 导致需求方无法直接应用。以智慧医疗

场景为例，数据供应方可能仅提供了患者基本生命体征的汇总报告，然而，医疗机构作为数据的需求方，更希望这些数据能与先进的疾病诊断算法无缝对接，从而支持精准医疗决策。这种非定制化的数据供应模式导致数据无法被直接应用，使得需求方认为数据缺乏实用价值，从而进一步加深了供需双方的语义错位。

4.2 方法应用：基于联邦服务的供需匹配

联邦服务融合大语言模型、区块链、DAO、场景工程等技术，可以有效解决物理世界的场景需求与数字世界的的数据供给之间的供需匹配难题。基于联邦服务的供需匹配框架如图4所示。

首先，针对业务语境差异导致的数据语义错位问题，联邦服务通过集成多模态大语言模型、自然语言理解与情景感知等技术，解析并统一不同业务语境下的数据定义。基于物理世界和数字世界的跨行业、跨领域知识，构建一个综合的、统一的、具有高度可解释性的跨域数据语义知识库，该知识库能够根据供需语境实现供需双方的智能语义映射。此外，区块链、DAO等技术提升了语义知识库的透明性、可审计性和可追溯性。参与联邦服务的节点共同维护和协同更新语义知识库，确保了数据定义的实时更新与跨领域共识，使得数据在多领域间流通时能够保持语义的高度一致性，从而有效解决

由业务语境差异导致的供需不匹配问题。

其次，针对数据需求不确定导致的语义错位问题，联邦服务采用机器学习、领域知识工程等技术构建了一个智能的需求解析与转化平台，自动将需求方的隐性知识和业务场景需求转化为具体、明确的数据需求规范，加速了从场景需求到数据需求的转化，提升了数据需求定义的精准度。此外，通过强化学习、时间序列分析等技术建立的动态数据需求预测模型，能够精准预测需求变化趋势，并利用情境感知、边缘计算与物联网等技术，实时感知并适应需求场景变化，再结合区块链智能合约，自动优化数据供应，确保数据供应与不断演化的场景需求实时对齐并精准对接，从而实现服务的敏捷性和场景的适应性。

最后，针对非定制化的数据供应模式导致的数据语义错位问题，联邦服务融合联邦学习、多模态大模型、场景工程等技术，在保障数据隐私的前提下，通过数据的跨域协同训练，深度挖掘跨领域数据之间的关联，为需求方量身定制高度个性化、与场景相适应的数据服务，使得数据服务与需求方业务场景紧密贴合。同时，利用安全多方计算与差分隐私技术实现数据在加密状态下的安全处理与分析，能够避免信息泄露，增加数据交互的信任度。此外，通过设计场景感知的智能组件，联邦服务能够实现

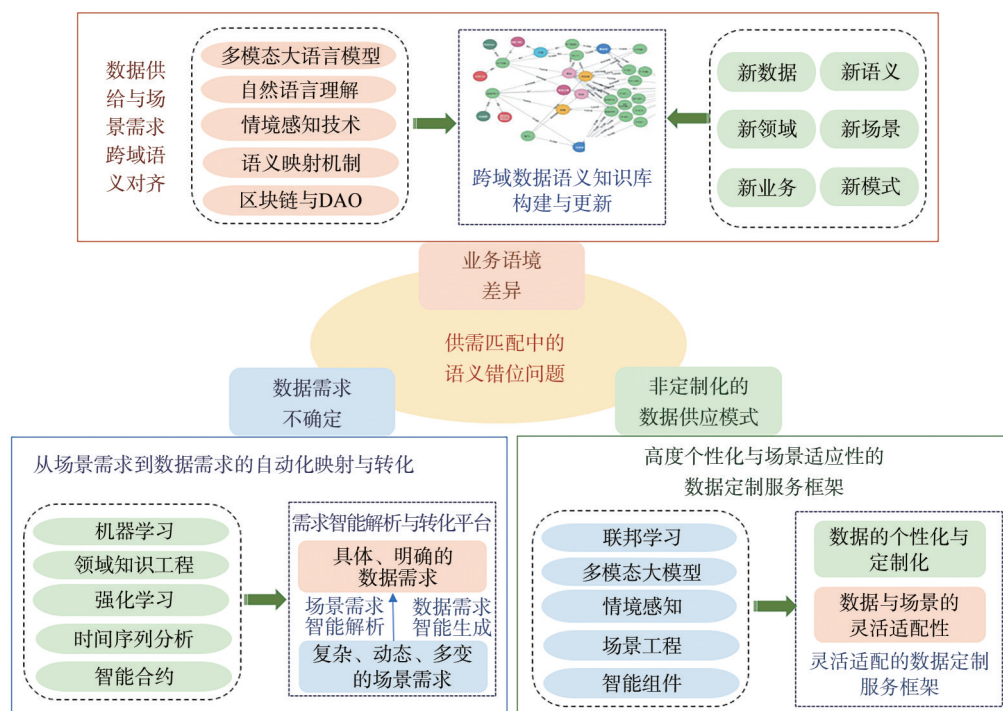


图4 基于联邦服务的供需匹配框架

服务的灵活配置和快速迭代,从而解决了非定制化数据供应模式带来的语义错位问题,确保数据服务能够与场景需求精准匹配。

5 结束语

随着大数据与人工智能等技术的发展,数据已成为至关重要的生产要素,如何在保护数据隐私与安全的前提下实现数据的流通与协同,充分挖掘数据的潜在价值,成为当前的重要挑战。联邦服务作为联邦生态对外赋能的重要窗口,可以在保护数据隐私和安全的情况下,实现数据供应与场景需求的智能匹配,从而推动分布式数据向智能化服务转化。本文探讨了联邦服务在联邦生态中的地位与作用,以及与联邦生态各关键环节的关系,提出了包含基础层、设施层、数据层、管控层、应用层的联邦服务5层框架,并分析了其在数据安全与可信、资源协同与主体协作、服务智能化与灵活性等方面的核心优势。在此基础上,本文探讨了联邦服务的关键支撑技术,包括区块链与DAO、LLM与场景工程、分布式计算与边缘计算、加密算法与隐私计算等。最后,围绕数据与场景的供需匹配,介绍了联邦服务的方法应用。本文提出的联邦服务为解决大数据时代的数据孤岛问题,促进在隐私保护和安全前提下的数据流通,提供了一种可行的技术框架,对于创建安全、协同、高效的未来智能服务生态具有重要意义。

参考文献:

- [1] 刘昕, 王晓, 张卫山, 等. 平行数据: 从大数据到数据智能[J]. 模式识别与人工智能, 2017, 30(8): 673-681.
LIU X, WANG X, ZHANG W S, et al. Parallel data: from big data to data intelligence[J]. Pattern Recognition and Artificial Intelligence, 2017, 30(8): 673-681.
- [2] 贾晓丰, 高嵩, 周琰, 等. 一种面向超大城市治理的数据高效跨域流通技术框架[J]. 数据与计算发展前沿, 2023, 5(5): 35-45.
JIA X F, GAO S, ZHOU Y, et al. A technical framework of efficient cross-domain data circulation for mega-city governance[J]. Frontiers of Data & Computing, 2023, 5(5): 35-45.
- [3] 贾晓丰, 梁郑丽, 任锦鸾. 多源信息协同[M]. 北京: 清华大学出版社, 2016.
JIA X F, LIANG Z L, REN J L. Multi-source information collaboration[M]. Beijing: Tsinghua University Press, 2016.
- [4] TU J, FAN J, TANG N, et al. Unicorn: a unified multi-tasking model for supporting matching tasks in data integration[C]//Proceedings of the ACM on Management of Data. New York: ACM, 2023: 1-26.
- [5] 王飞跃, 王艳芬, 陈意竹, 等. 联邦生态: 从联邦数据到联邦智能[J]. 智能科学与技术学报, 2020, 2(4): 305-311.
WANG F Y, WANG Y F, CHEN Y Z, et al. Federated ecology: from federated data to federated intelligence[J]. Chinese Journal of Intelligent Science and Technology, 2020, 2(4): 305-311.
- [6] LI J J, QIN R, GUAN S T, et al. Blockchain intelligence: intelligent blockchains for web 3.0 and beyond[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024: 1-10.
- [7] QIN R, YUAN Y, WANG F Y. Blockchain-based knowledge automation for CPSS-oriented parallel management[J]. IEEE Transactions on Computational Social Systems, 2020, 7(5): 1180-1188.
- [8] 李娟娟, 秦蕊, 丁文文, 等. 基于Web3的去中心化自治组织与运营新框架[J]. 自动化学报, 2023, 49(5): 985-998.
LI J J, QIN R, DING W W, et al. A new framework for Web3-powered decentralized autonomous organizations and operations[J]. Acta Automatica Sinica, 2023, 49(5): 985-998.
- [9] WANG F Y, DING W W, QIN R, et al. Parallel philosophy for MetaOrganizations with MetaOperations: from Leibniz's Monad to HanoiDAO[J]. IEEE Transactions on Computational Social Systems, 2022, 9(3): 658-666.
- [10] QIN R, DING W W, LI J J, et al. Web3-based decentralized autonomous organizations and operations: architectures, models, and mechanisms[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2023, 53(4): 2073-2082.
- [11] WANG F Y, LI J J, QIN R, et al. ChatGPT for computational social systems: from conversational applications to human-oriented operating systems[J]. IEEE Transactions on Computational Social Systems, 2023, 10(2): 414-425.
- [12] QIN R, WANG F Y, ZHENG X L, et al. Sora for computational social systems: from counterfactual experiments to artificiofactual experiments with parallel intelligence[J]. IEEE Transactions on Computational Social Systems, 2024, 11(2): 1531-1550.
- [13] 石庆松, 秦蕊, 肖鹏, 等. 基于联盟链的电力物资采购数据共享激励机制[J]. 中国电力, 2022, 55(3): 87-96.
SHI Q S, QIN R, XIAO P, et al. Incentive mechanism for data sharing of power material procurement based on consortium blockchain[J]. Electric Power, 2022, 55(3): 87-96.
- [14] LI J J, LI J Q, WANG X, et al. Multi-blockchain based data trading markets with novel pricing mechanisms[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(12): 2222-2232.
- [15] WANG F Y, QIN R, CHEN Y Z, et al. Federated ecology: steps toward confederated intelligence[J]. IEEE Transactions on Computational Social Systems, 2021, 8(2): 271-278.
- [16] WANG F Y, ZHANG W S, TIAN Y L, et al. Federated data: toward new generation of credible and trustable artificial intelligence[J]. IEEE Transactions on Computational Social Systems, 2021, 8(3): 538-545.
- [17] 朱静, 王飞跃, 王戈, 等. 联邦控制: 面向信息安全和权益保护的分布式控制方法[J]. 自动化学报, 2021, 47(8): 1912-1920.
ZHU J, WANG F Y, WANG G, et al. Federated control: a distributed control approach towards information security and rights protection[J]. Acta Automatica Sinica, 2021, 47(8): 1912-1920.
- [18] WANG F Y, ZHU J, QIN R, et al. Federated control: toward information security and rights protection[J]. IEEE Transactions on Computational Social Systems, 2021, 8(4): 793-798.
- [19] WANG F Y, QIN R, LI J J, et al. Federated management: toward federated services and federated security in federated ecology[J]. IEEE Transactions on Computational Social Systems, 2021, 8(6): 1283-1290.
- [20] 贾晓丰, 高嵩, 江茜, 等. 面向城市复杂系统的社会计算关键技术研究[J]. 智能科学与技术学报, 2021, 3(2): 228-233.
JIA X F, GAO S, JIANG X, et al. Research on key technologies of social computing for urban complex system[J]. Chinese Journal of Intelligent Science and Technology, 2021, 3(2): 228-233.
- [21] QIN R, LI J J, LIANG X L, et al. Management-oriented operating sys-

tems: harnessing the power of DAOs and foundation models[C]//Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC). Piscataway: IEEE Press, 2023: 3180-3185.

- [22] 李娟娟, 管桑田, 秦蕊, 等. 智能区块链与区块链智能: 构筑DePIN的基础设施智能[J]. 智能科学与技术学报, 2024, 6(1): 5-16.
LI J J, GUAN S T, QIN R, et al. Intelligent blockchains and blockchain intelligence: the infrastructure intelligence for DePIN[J]. Chinese Journal of Intelligent Science and Technology, 2024, 6(1): 5-16.
- [23] WANG G, QIN R, LI J J, et al. A novel DAO-based parallel enterprise management framework in Web3 era[J]. IEEE Transactions on Computational Social Systems, 2024, 11(1): 839-848.
- [24] QIN R, LI J J, WANG F Y. MetaEconomics and MetaManagement for MetaCities and MetaSocieties in Metaverse[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024: 1-10.
- [25] CHANG Y P, WANG X, WANG J D, et al. A survey on evaluation of large language models[J]. ACM Transactions on Intelligent Systems and Technology, 15(3): 1-45.
- [26] WANG L, MA C, FENG X Y, et al. A survey on large language model based autonomous agents[J]. Frontiers of Computer Science, 2024, 18(6): 186345.
- [27] LI X, SONG R, FAN J F, et al. Development and testing of advanced driver assistance systems through scenario-based systems engineering[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(8): 3968-3973.
- [28] CHANG C, CAO D P, CHEN L, et al. MetaScenario: a framework for driving scenario data description, storage and indexing[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(2): 1156-1175.
- [29] DE GELDER E, PAARDEKOOPER J P, SABERI A K, et al. Towards an ontology for scenario definition for the assessment of automated vehicles: an object-oriented framework[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(2): 300-314.
- [30] JIANG Y N, KANG J W, NIYATO D, et al. Reliable distributed computing for Metaverse: a hierarchical game-theoretic approach[J]. IEEE Transactions on Vehicular Technology, 2023, 72(1): 1084-1100.
- [31] 施巍松, 张星洲, 王一帆, 等. 边缘计算: 现状与展望[J]. 计算机研究与发展, 2019, 56(1): 69-89.
SHI W S, ZHANG X Z, WANG Y F, et al. Edge computing: state-of-the-art and future directions[J]. Journal of Computer Research and Development, 2019, 56(1): 69-89.
- [32] GONG C, LIN F H, GONG X W, et al. Intelligent cooperative edge computing in Internet of things[J]. IEEE Internet of Things Journal, 2020, 7(10): 9372-9382.
- [33] HAMDAN S, AYYASH M, ALMAJALI S. Edge-computing architectures for Internet of things applications: a survey[J]. Sensors, 2020, 20(22): 6441.
- [34] 张鹏, 李美鹏, 郭兆中, 等. 基于隐私计算技术的数据安全共享交互框架[J]. 网络安全与数据治理, 2023, 42(12): 54-59.
ZHANG P, LI M P, GUO Z Z, et al. A data security sharing and interaction framework based on privacy computing technology[J]. Cyber Security and Data Governance, 2023, 42(12): 54-59.

[作者简介]



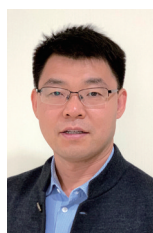
贾晓丰 (1985-), 男, 博士, 北京市大数据中心数据管理部负责人、教授级高级工程师, 主要研究方向为复杂系统数据治理、联邦智能。



秦蕊 (1983-), 女, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室副研究员, 主要研究方向为区块链智能、DAO与平行管理。



王寿文 (1986-), 男, 澳门科技大学创新工程学院博士生, 主要研究方向为复杂系统与平行智能、区块链智能、智能合约及基于DAO的管理与治理。



齐红威 (1975-), 男, 博士, 数据堂 (北京) 科技股份有限公司创始人兼CEO, 正高级工程师, 中国自动化学会联邦数据与联邦智能专委会主任, 主要研究方向为人工智能数据采集、处理以及联邦数据。



王飞跃 (1961-), 男, 中国科学院自动化研究所复杂系统管理与控制国家重点实验室主任, 主要研究方向为平行系统的方法与应用、社会计算、平行智能、知识自动化。



李娟娟 (1986-), 女, 博士, 中国科学院自动化研究所多模态人工智能系统全国重点实验室副研究员, 主要研究方向为区块链智能、DAO与平行管理。



章敏 (1992-), 女, 北京市大数据中心数据管理部高级工程师, 主要研究方向为数据治理、知识工程、联邦智能。



梁小龙 (1987-), 男, 博士, 多模态人工智能系统全国重点实验室助理研究员, 主要研究方向为平行智能、区块链与DAO。