软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, 2022,33(3):931−949 [doi: 10.13328/j.cnki.jos.006455]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

基于大数据的分布式社会治理智能系统*

吕卫锋[1,2,3,6], 郑志明[1,3,4,5,6,7], 童咏昕[1,2,6], 张瑞升[1,2,6], 魏淑越[1,2,6], 李卫华[1,4,5,7]

[1](软件开发环境国家重点实验室(北京航空航天大学), 北京 100191)

[2](北京航空航天大学 计算机学院, 北京 100191)

[3](北京航空航天大学 人工智能研究院, 北京 100191)

[4](北京航空航天大学 数学科学学院, 北京 100191)

[5](数学、信息与行为教育部重点实验室(北京航空航天大学), 北京 100191)

[6](大数据科学与脑机智能高精尖创新中心(北京航空航天大学), 北京 100191)

[7](鹏城实验室, 广东 深圳 518055)

通信作者: 童咏昕, E-mail: yxtong@buaa.edu.cn

摘 要: 近年来, 推动社会治理的协同化、智能化, 完善共建共治共享的社会治理制度, 是国家的重要发展方向. 数据作为一种生产要素, 在社会治理中起着愈发关键的作用. 如何实现多方海量数据的安全查询、协同管理、智能分析, 是提升社会治理效果的关键问题. 在新冠疫情防控等重大公共事件中, 分布式社会治理面临着安全计算效率低、多方可信协同差、复杂任务决策难的三大挑战. 针对以上挑战, 基于安全多方计算、区块链技术与精准智能理论, 提出了一种基于大数据的分布式社会治理智能系统. 所提出的系统能够支撑社会治理的各类应用, 为新时代社会治理水平的提升提供决策支撑.

关键词: 社会治理; 大数据系统; 区块链; 精准智能; 分布式系统

中图法分类号: TP311

中文引用格式: 吕卫锋, 郑志明, 童咏昕, 张瑞升, 魏淑越, 李卫华. 基于大数据的分布式社会治理智能系统. 软件学报, 2022, 33(3): 931–949. http://www.jos.org.cn/1000-9825/6455.htm

英文引用格式: Lü WF, Zheng ZM, Tong YX, Zhang RS, Wei SY, Li WH. Intelligent System for Distributed Social Governance Based on Big Data. Ruan Jian Xue Bao/Journal of Software, 2022, 33(3): 931–949 (in Chinese). http://www.jos.org.cn/1000-9825/6455.htm

Intelligent System for Distributed Social Governance Based on Big Data

LÜ Wei-Feng[1,2,3,6], ZHENG Zhi-Ming[1,3,4,5,6,7], TONG Yong-Xin[1,2,6], ZHANG Rui-Sheng[1,2,6], WEI Shu-Yue[1,2,6], LI Wei-Hua[1,4,5,7]

[1](State Key Laboratory of Software Development Enviroment (Beihang University), Beijing 100191, China)

[2](School of Computer Science and Engineering, Beihang University, Beijing 100191, China)

[3](Institute of Artficial Intelligence, Beihang University, Beijing 100191, China)

[4](School of Mathematics Sciences, Beihang University, Beijing 100191, China)

[5](Key Laboratory of Mathematics, Informatics and Behavioral Semantics (Beihang University), Ministry of Education, Beijing 100191, China)

[6](Advanced Innovation Center for Big Data and Brain Computing (Beihang University), Beijing 100191, China)

* 基金项目: 国家重点研发计划(2018AAA0102300); 国家自然科学基金(61822201, U1811463, 62076017); CCF-华为数据库创新研究计划(CCF-HuaweiDBIR2020008B); 软件开发环境国家重点实验室(北京航空航天大学)开放课题(SKLSDE-2020ZX-07)
本文由“数据库系统新型技术”专题特约编辑李国良教授、于戈教授、杨俊教授和范举教授推荐.
收稿时间: 2021-06-30; 修改时间: 2021-07-31; 采用时间: 2021-09-13; jos 在线出版时间: 2021-10-21

[7](Peng Cheng Laboratory, Shenzhen 518055, China)

Abstract: In recent years, promoting the synergy and intelligence of social governance, and improving the social governance system of co-construction, co-governance and sharing are important development directions for the country. As a production factor, data plays an increasingly critical role in social governance. How to realize the secure query, collaborative management, and intelligent analysis of multi-party massive data is the key issue to improve the effectiveness of social governance. In major public events such as the prevention and control of the COVID-19, distributed social governance faces low computing efficiency, poor multi-party credible coordination, and difficult decision-making for complex tasks. In response to the above challenges, this study proposes on big data based distributed social governance intelligent system based on secure multi-party computing, blockchain technology, and precise intelligence theory. The proposed system can support various applications of social governance that provide decision-making support for the improvement of social governance in the new era.

Key words: social governance; big data system; blockchain; refined intelligence; distributed system

随着信息技术的迅速发展, 各政府职能部门积累了大规模数据. 基于海量数据实现多部门的协同配合、智能决策, 提高城市信息化服务水平, 对推进社会治理现代化建设具有重要意义. “科技支撑”是社会治理体系的重要组成部分, 而大数据、区块链、云计算、人工智能等新一代信息技术将为社会治理提供关键科技支撑[1]. 目前, 在我国社会治理中, 政府是社会治理的主导力量, 但已不是唯一参与方, 企事业单位、社会组织、城乡社区居民组织等都已经成为社会治理的重要力量[2], 这些多元的参与方共同构成了社会治理中的治理单元. 然而, 不同的治理单元往往存储着多类型的海量数据, 数据的权限管理也情况复杂, 这都对社会治理中高效的数据管理、智能分析带来了挑战.

基于大数据的分布式社会治理是指多方数据在自治管理的场景下, 通过不同层级的治理单元的协同合作实现统一确权管控、数据智能分析, 从而对重大的公共事件做出迅速决策. 以我国抗疫治理实践为例, 在疫情早期的治理中, 部分基层机构存在惯性思维、应急响应慢、决策周期长的问题, 难以及时应对迅速发展的疫情. 北京市基于分布式大数据的治理方式则更具灵活性. 在北京疫情防控中, 以市区、街道等为单位建立了疫情治理子系统, 管控与集成人群的流动数据, 在此基础上, 结合多种人工智能技术追溯传播途径, 及时发现并控制了传播源头, 治理优势明显.

然而, 分布式的社会治理系统在实践中面临着安全计算效率低、多方可信协同差、复杂任务决策难三大挑战.

- 首先, 治理单元所拥有的社会治理数据往往涉及公民隐私, 因此必须保障数据的隐私安全. 例如, 交通部门的地图数据、医疗部门的病患数据、金融部门的财务数据等具有很强的隐私要求, 有些敏感数据禁止离开本地, 需要通过安全计算技术来实现数据查询; 同时, 在社会治理中, 治理单元数目众多、数据规模庞大, 因此, 多方的海量数据安全计算面临着计算效率低的挑战.
- 其次, 不同治理单元的数据具有自治管理的特点, 社会治理通常需要各地、各部门的协同计算, 各职能部门间难以实现可信的确权管控, 难以支撑在重大公共事件中的互信合作.
- 最后, 社会治理中的复杂情景, 如重大公共安全事件、资源竞争矛盾等, 通常具有复杂、动态、随机的特点, 其演化规律往往是非线性的, 基于统计的简单线性化建模方法往往难以适用于非线性的社会演化特征.

本文针对以上需求与挑战, 设计了一种分布式社会治理智能系统. 在大数据的场景下, 该系统基于安全多方计算、区块链技术与精准智能理论, 能够实现对敏感数据的安全计算、数据跨级权限的一网通管和非线性复杂任务的智能决策.

- 本文首先通过安全多方计算为分布式数据库构建安全计算基础算子, 在高效完成安全操作的基础上, 实现多方的查询接口, 从而为后续的智能分析奠定数据基础.
- 然后, 基于区块链技术设计多方数据的共享访问机制, 构建自治多方在互不信任环境中的权限验证机制, 从而实现数据访问的可信权限管理. 例如企事业单位、社会组织、政府等都可以作为区块链系统中的节点参与到社会治理中, 各治理单元可通过多方安全的高效访问接口, 实现跨级数据查询的

一网通管, 为面向社会治理的精准智能算法提供平台支撑.

- 最后, 通过精准智能对复杂任务进行建模, 利用动力学模型精准构建非线性复杂网络, 为社会治理提供决策支持.

本文的主要贡献如下:

- 提出了分布式社会治理智能框架的概念, 并将其应用在卫生安全领域, 为新冠疫情防控中寻找超级传播者、阻断病毒传播链条提供了决策支撑.
- 构建了基于大数据的分布式社会治理智能系统. 该系统基于安全多方计算、区块链和精准智能技术, 能够实现高效安全计算、多方确权管控和智能社会决策.
- 文章在真实数据集上对所提系统进行了实验分析, 从运行时间、通信开销、治理效果等方面验证了所提系统在多方数据查询下的高效性和社会治理应用中的实用性.

本文第 1 节阐述分布式社会治理智能框架. 第 2 节基于所提出的理念设计分布式社会治理智能系统.第 3 节介绍系统的安全计算层. 第 4 节介绍系统的区块链层. 第 5 节介绍系统的社会治理层. 第 6 节对提出的系统进行实验验证. 第 7 节介绍相关工作. 第 8 节对全文进行总结并展望未来工作.

1 分布式社会治理智能框架

为了有效地应对大数据场景下安全计算效率低、多方可信协同差、复杂任务决策难给社会治理所带来的挑战,为新时代社会治理提供平台和决策支撑, 本节首先针对社会治理所面临的复杂场景, 在第 1.1 节提出社会治理框架的核心理念; 然后在第 1.2 节对治理框架进行概述; 最后, 在第 1.3 节以疫情为例介绍治理框架的应用案例.

1.1 社会治理框架的核心理念

为有效应对现有挑战, 基于大数据的分布式社会治理智能框架需要具备五大基本特征. 具体而言, 在结构方面构建动态可信网络, 在协同方面实现跨域确权管控, 在可靠性方面实现强抗毁高安全, 在鲁棒性方面做到强一致高容错, 在智能化方面实现实时精准智能. 下面对上述五大基本特征展开具体论述.

- 动态可信网络. 智能协同治理中需要多方成员的动态协作参与. 例如, 政府部门、医院、院校等机构的本地信息具有很强的保密性, 这就为在缺乏中心服务器的环境中实现协同合作增加了应用难度. 只有超越部门、组织、机构的边界, 构建可信组织形态, 才能实现以特定治理目标为导向的治理指挥形态. 同时, 面临突发应急事件, 社会治理要求特定的节点快速加入治理框架中, 共享信息、协作拟定整体方案, 这对于治理框架的动态更新提出了更高的要求. 因此, 构建可信治理指挥平台、实现去中心化下的动态接入分布式治理单元, 是智能治理框架的结构目标.
- 跨域确权管控. 社会治理中, 重大决策的制定需要跨部门、跨编制的参与, 要求在整体利益最大化的前提下各方能够各司其职、各尽所能. 因此, 实现跨基础单元的身份认证、确权管控, 建立跨域身份认证联盟链, 是协同治理的必要条件. 例如在新冠疫情期间, 完善对卫生部门、信息部门、公安部门、防控部门等政府部门的权限分配, 统筹安排疫情防控措施, 实现真正意义上的“整体智治”, 是提升社会协同治理效果的重要因素.
- 强抗毁高安全. 安全性、抗毁性是一个分布式治理系统的基础要求. 安全性要求在恶意节点的参与下保证其他节点之间的安全通信、交流信息的完整性以及保密性. 抗毁性是指系统在治理节点、通信链路受毁下的可靠性. 在传统的中心化系统中, 中心服务器必须始终处于在线状态, 否则一切需要与中心节点的交互都将无法进行. 在新时代的社会治理中, 各节点往往由社区、居委会等基层治理单元组成, 每个节点都有可能出现暂时离线、通信受阻等问题. 提高系统的抗毁性, 是保证治理框架在复杂多变的环境下有效运行的基础. 而区块链系统分布式、可扩展、跨网络分布、强加密等特点, 可以提升治理网络的安全性、抗毁性, 增强治理体系的弹性韧劲.
- 强一致高容错. 在分布式系统中, 一个重要的问题是如何保障各节点在事务处理时快速达成一致, 实

现高容错. 容错是指当部分节点出现故障或采取恶意行为时, 整个系统仍然可以达成共识, 其功能实现不受到影响. 例如在“拜占庭问题”中, 恶意的将军可以通过发送不利信息或者散布虚假信息来干扰协议的达成, 破坏整体的有利行动. 在分布式社会协同治理框架中, 区块链中的共识机制可以做到有效保证所有治理节点的数据一致性. 基于工作量证明机制、权益证明机制、拜占庭容错协议等共识算法, 治理框架可以在去中心化的环境下建立信任、达成共识, 避免由于节点的恶意行为而导致的系统崩塌.

- 实时精准智能. 现实生活中的社会治理需要在复杂、随机的环境中分析大规模数据. 目前, 基于线性统计方法的模型无法精确地还原随机化的非线性关联关系. 通过基于统计物理、动力系统、应用数学等理论的精准智能技术感知复杂的数据结构, 内嵌先验知识构建复杂系统, 智能分析复杂行为, 是人工智能技术未来的发展方向. 在跨平台、跨治理单元的治理数据实时共享的基础上, 通过基于智能合约的动态治理规则实现实时态势感知、通过精准智能理论实现精准集群智能, 是智能治理框架的根本目标.

1.2 社会治理框架概述

分布式社会治理智能框架涉及空间和时间两个维度, 其中, 空间维度具体指基于区块链技术的集成治理体系, 描述了社会治理数据管理的空间结构; 时间维度具体指基于精准智能的复杂行为演化建模, 描述了社会系统非线性动力学时间演化过程. 通过该两个维度形成对社会治理数据的分布式协同管理和对社会系统演化的智能模拟预测, 构成粒度异构的时空网络组织框架. 下面具体介绍.

- 在结构方面, 如图 1 左半部分所示, 利用区块链技术及其共识机制, 将分布式治理单元协同为集成治理体系. 分布式治理单元是处于社会治理第一线的基层行政机关, 如村委会、居委会、街道办事处等. 它们直接接触和处理大量社会治理相关数据, 是社会治理体系延伸出的“触角”. 利用区块链的共识机制, 可将来自各个不同分布式治理单元的数据协同起来, 实现集成治理, 从而基于协同后的数据有效把握社会系统整体运行状态, 及早发现突发性、随机性事件, 及时做出应对决策;
- 在行为方面, 如图 1 右半部分所示, 利用精准智能进行优化迭代, 实现从初始系统到优化系统的转变, 为社会治理提供决策支撑. 初始系统指基于协同后的社会治理数据构建的社会动力学演化系统. 正如第 1 节所论述的, 社会系统演化具有非线性特征, 并被事件的突发性、随机性所加剧, 导致社会系统演化具有高度的复杂性和不确定性, 基于现有基于统计线性化动态建模的人工智能技术难以对系统演化结果进行可靠预测和对系统演化方向进行干预. 为此引入精准智能技术, 通过复杂系统精准构建反演具有非线性复杂逻辑关系的多层次、多尺度、可解释的人工智能动力学模型, 从而为通过政策干预等社会治理手段, 引导社会系统向特定目标方向演化提供智能决策支撑.

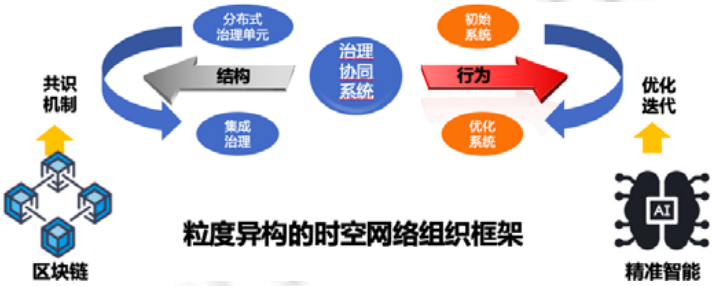

图 1 粒度异构时空网络组织框架

1.3 面向卫生安全的治理框架应用

基于上述分布式社会治理框架, 本节以新冠疫情防控为例, 介绍面向卫生安全的分布式社会治理案例. 首先介绍框架的结构和功能, 然后介绍如何通过系统管控新冠肺炎疫情中的超级传播子以阻断病毒传播链

条, 从而为疫情防控提供决策支撑.

本节以 2020 新冠疫情防控为例, 介绍面向卫生安全的分布式社会治理智能系统. 面对突发的新冠疫情, 新增确诊病例的发现时间、地点具有很强的随机性, 其确诊前 14 天之内的活动轨迹可能会涉及,小区、学校、娱乐场所、批发市场等人员密集场所, 进行病例筛查、统筹调控需要卫健委、公安部门、交通部门等政府部门的协同合作. 本系统结合区块链技术与精准智能理论, 有效解决了以上问题, 为未来的分布式社会智能治理提供了可行的解决方案.

如图 2 所示, 该系统以各医院、学校、居委会等机构作为治理域进行态势感知, 通过区块链技术构建动态可信网络结构, 搭建可信治理指挥平台. 对各地的突发病例, 各基层治理单元迅速完成密切接触者排查、核酸试剂检测, 并将采集到的信息根据区块链管理权限上传到数据权限管理模块中, 实现协同治理、科学溯源, 提高防控效率. 各基层治理单元通过共识协议对区块链中数据摘要与管理权限进行更新、备份, 实现存储数据的一致性、高容错, 确保即使在部分节点因通信因素瘫痪时, 仍能做出最有利的决策. 对不同政府部门的统筹指挥, 利用区块链去中心化的特点实现跨域确权管控, 将各治理单元的 CA 认证中心组网, 构成跨域身份认证联盟链, 实时感知突发性事件. 对疫情防控中各节点中的信息, 通过一些成熟的密码学技术例如哈希算法、同态加密、安全多方计算、秘密共享来保护用户隐私, 实现强抗毁、高安全.

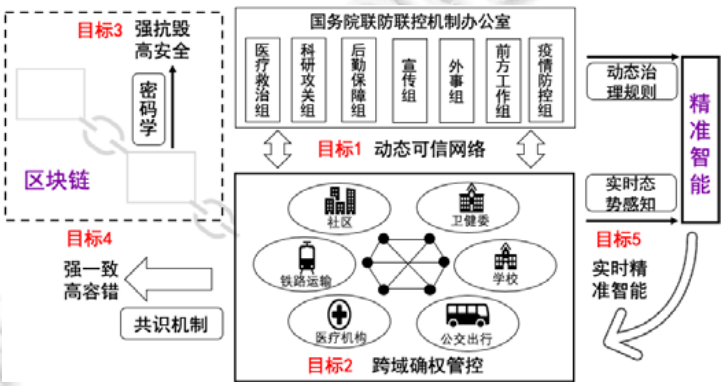

图 2 面向卫生安全的分布式社会协同治理智能系统架构图

在区块链技术作为底层空间结构的基础上, 面对随机性、非线性的突发事件, 还需要通过精准智能理论进行预测、模拟、分析, 做出实时应对. 随着交通技术的不断进步, 在很短的时间内, 确诊患者或无症状感染者可能会携带病毒出现在各类人员密集性场所, 造成大范围的可疑接触者, 这对各地的人员排查、核酸检测带来了巨大的压力. 为防止疫情防控中“超级传播者”带来的大面积感染, 本系统深入依托随机性, 对确诊患者的行动轨迹、可能感染的疑似人员进行精准定位, 以最小的代价阻止新冠病毒的进一步传播.

2 分布式社会治理系统架构

本节介绍分布式社会治理智能系统的系统架构. 如图 3 所示, 分布式社会治理智能系统共包括 3 层, 自底端向上依次为安全计算层、区块链层、社会治理层. 首先, 在安全计算层, 针对多类型的海量数据, 系统进行分布式多节点存储, 通过安全多方计算技术实现多方的安全数据查询; 在此基础上, 在区块链层, 通过区块链的共享访问控制, 实现数据访问权限的一网通管, 支撑社会治理层的智能算法; 最后, 在社会治理层, 结合复杂网络模型的构建理论与精准智能理论, 实现社会治理的智能决策. 各层具体功能如下.

- 安全计算层: 本层面向各社会治理参与部门、组织、社区等多元治理节点, 收集并存储包括关系数据、时空数据、文本数据、图像数据等多种类型的多源数据. 其中, 本层各治理节点可能使用不同类型的数据库系统. 在此基础上, 由于各方数据的安全性要求, 在数据存储上实现安全基础算子, 例如多方的安全求和、安全比较、安全集合求并, 从而能够支持系统调用查询接口, 完成多方范围查询、多方

近邻查询等操作，为社会治理提供数据支撑.

- 区块链层：本层在安全多方计算层提供的数据查询接口上构建可信协同的区块链权限管理平台，在提供高安全性的同时，还要保证系统效率. 本文基于控制策略生成机制和智能合约来完成共享访问控制，并通过节点划分方法和分层 Merkle 索引来优化区块链的访问、存储、查询效率，从而为社会治理提供平台支撑.
- 社会治理层：本层在区块链权限管理平台的基础上，通过复杂网络模型构建理论和精准智能技术实现实时精准智能. 在本层中，系统基于 SIR 模型、动力学数学模型等构建复杂网络，通过复杂数据感知、节点关系挖掘、复杂行为分析实现社会治理过程中的精准决策，为社会治理提供决策支撑.

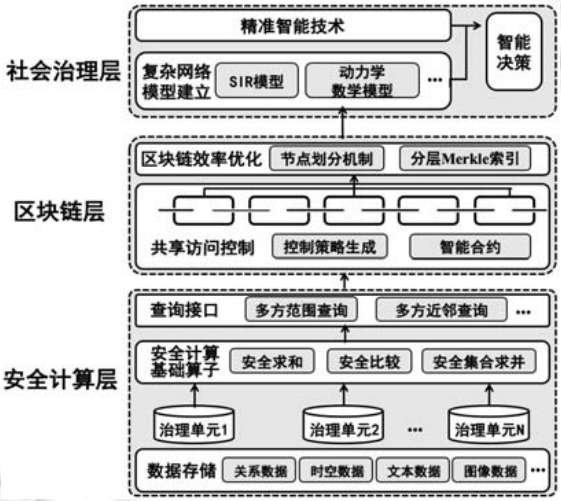

图 3　分布式社会治理智能系统架构

3　安全计算层

系统的安全计算层主要负责在分布式存储的数据基础上设计统一的安全基础算子，并实现高效的多方查询接口，为区块链层的确权管控和社会治理层的智能决策提供数据支撑. 下面首先基于安全多方计算技术介绍本系统实现的安全计算基础算子，然后对本层提供的多方查询接口进行详细阐述.

3.1　安全计算基础算子

不同治理单元的数据分布式存储如何在隐私保护的限制下完成数据查询，是社会治理的首要问题. 安全多方计算[3]是一种能够保证多方数据计算过程安全的密码学技术，本系统基于安全多方计算实现了安全求和、安全比较、安全集合求并这 3 种基础算子. 下面分别进行介绍.

- 安全求和. 安全求和的目标是对多方的数据进行求和，同时保证在计算过程中，各方的数据都不会泄露给其他任意一方. 本系统通过秘密共享[4]实现这一操作：首先，每方随机生成一个 n–1 次的多项式，并将本地存储的数据设为多项式的常数项；接着，各方将 n 个特定值代入多项式中计算出结果；最后，在各方的多项式结果基础上，通过拉格朗日插值公式可以计算出求和结果. 这一过程中，各方原始数据没有离开本地，保证了数据安全；
- 安全比较. 安全比较的目标是对多方的数据求和，并输出求和结果与给定数值的大小关系，同时保证在计算过程中，各方的数据都不会泄露给其他任意一方. 假定多方存储的数据是 a_i，给定的比较数值是 b，我们首先将比较数值 b 平均分为 n 份，然后令 $c_i = a_i - \frac{b}{n}$. 从 c_i 的定义可以发现：只需判断出 $\sum_i c_i$ 是否为正数，就可以得到比较结果. 为保证计算过程的安全性，各方生成随机的正数 d_i，通过秘

密共享协议计算$\left(\sum_i c_i\right)\left(\sum_i d_i\right)$, 可以判断出$\sum_i c_i$的正负性;

- 安全集合求并. 安全集合求并的目标是对多方的数据集合求并集, 并输出并集中的元素, 同时保证在计算过程中, 各方的数据都不会泄露给其他任意一方. 安全集合求并算子主要包含 3 个步骤: 首先, 各方通过差分隐私机制在本地数据集合中添加一些噪声数据, 然后将加噪后数据发送给其他方; 然后, 每方对收集到的数据集合进行求并操作; 最后, 各方通过移除本地添加的噪声数据来得到最终的数据并集.

3.2 多方查询接口

在安全计算基础算子上, 为向社会治理中的数据分析提供数据接口, 本系统实现了一系列数据查询基础操作. 下面以多方范围查询和多方近邻查询为例进行介绍.

- 多方范围查询. 多方范围查询的输入是查询范围, 输出是各方在查询范围内的数据. 为保障在范围查询过程中不泄漏各方的隐私数据, 查询机制通过安全集合求并算子完成: 首先, 对于给定的查询范围, 各方在本地进行范围查询得到自身的查询结果; 然后, 通过安全集合求并算子对多方的查询结果集合计算并集.
- 多方近邻查询. 多方近邻查询的输入是查询数据和正整数 k, 输出是各方数据中与查询数据距离最近的 k 个数据. 为保障在近邻查询过程中保护数据隐私, 查询机制通过安全比较和安全集合求并两种基础算子完成: 首先, 根据给定的查询数据, 生成一个随机半径, 得到初始的查询范围; 接着, 基于初始查询范围进行多方范围查询, 得到查询结果集合. 通过安全比较算子比较查询结果集合大小与 k 的大小关系: 如果查询结果集合中数据数量大于 k, 则将查询范围减少至一半再次进行范围查询; 如果查询结果集合中数据数量小于 k, 则将查询范围扩大至 2 倍再次进行范围查询; 否则, 查询结果集合中数据数量等于 k,直接输出查询结果集合. 通过以上的迭代算法, 可以更高效地完成多方近邻查询.

4 区块链层

系统的区块链层主要负责在安全多方计算技术提供的查询接口上建立区块链系统, 实现确权管控: 首先, 针对各治理单元的访问权限等级不同, 设计“一网通管”的共享访问控制机制; 然后, 针对区块链系统占用资源高、运行效率低的问题, 设计可信协同的区块链系统架构, 能够在保障安全性的同时兼顾系统效率. 以下分别从共享访问控制机制设计、区块链系统效率优化两个方面进行展开.

4.1 共享访问控制机制

如图 4 所示, 共享访问控制机制主要分为访问控制策略制定和访问控制自动授权两个模块: 访问控制策略制定模块基于区块链可信存储的特点实现多方控制策略的透明一致; 访问控制自动授权模块基于智能合约技术实现多方自治数据的可信访问控制. 下面分别进行介绍.

- 访问控制策略制定. 由于社会治理中存在多个自治的数据管理方, 传统基于角色的访问控制产生大量治理单元间的控制策略, 导致访问控制混乱. 因此, 本文提出了基于特征属性的访问控制策略生成算法, 通过统一社会治理系统中的相关属性生成访问控制策略, 同时设计高效算法实现访问控制策略的快速查找. 当某个治理单元 U(如居委会)发起访问某数据 D_i 的请求时, 拟将其归纳为四维属性$A(U)=\langle attr_i,attr_d,attr_a,attr_e\rangle$, 分别表示此次访问的发起方属性、访问数据属性(数据所属单位、数据描述等)、操作属性和环境属性, 数据所在方通过发起访问的属性信息 $A(U)$进行访问控制的判断. 通过在治理中心采用 Apriori 算法等频繁模式项算法挖掘常用访问控制策略, 使社会治理系统中各治理单元就属性归纳与策略归纳达成一致共识. 同时, 基于挖掘得到的高频策略构建属性前缀树, 对基于属性的访问策略进行快速判定.
- 访问控制自动授权. 智能合约能够支持程序的可信执行, 因此访问控制的可信自动授权执行可以通过智能合约实现. 多个区块链平台均设计提供了图灵完备脚本语言用以编写智能合约, 可实现循环、

条件判断等多种代码逻辑. 实现自动执行的智能合约应以访问请求的发起作为触发条件, 在合约执行中需要两步: 一是根据访问请求的数据属性 $A(U)$查找对应访问控制策略集合; 二是计算策略内属性要求与访问请求属性交集判断授权结果. 基于区块链中的属性与策略和智能合约, 可以完成系统的可信访问控制, 实现确权管控.

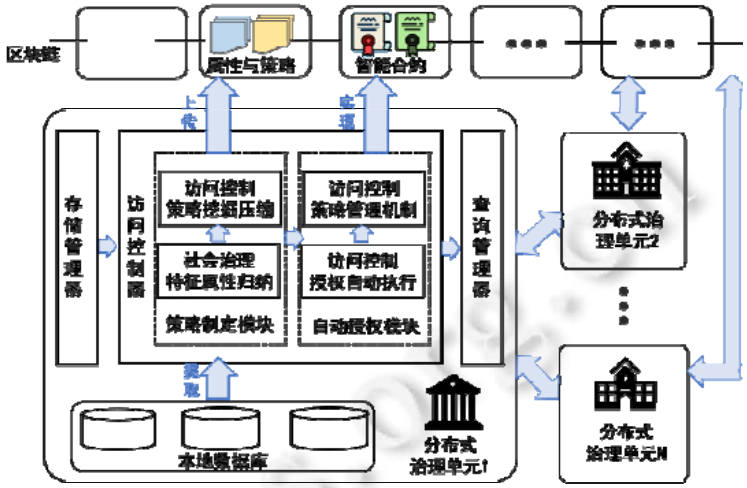

图 4 共享访问控制机制

4.2 区块链系统效率优化

通过前文所述的共享访问控制机制, 可以实现社会治理系统中的确权管控. 但由于区块链系统中各节点需存储完整的权限管理数据, 因此直接部署面临着本地存储资源不足的技术挑战. 为解决这一问题, 本文提出了针对系统效率的优化方案: 首先, 利用具有轻节点、全节点两类节点的结构搭建区块链系统; 然后, 针对社会治理数据类型广泛且各方数据结构不统一的现状, 设计层次化存储的异构数据索引, 从而实现高效的共享访问控制.

- 节点划分

如图 5 所示, 将区块链节点分为全节点与轻节点两类. 全节点负责收集、存储整个区块链系统中的计算结果数据, 并通过认证数据结构生成数据摘要; 在社会治理系统中, 可以建立专门的社会治理数据节点作为全节点. 轻节点是社会治理系统中的分布式治理单元, 只存储区块链最新区块数据的区块头, 区块头中包含全节点通过认证数据结构生成的摘要值. 认证数据结构是外包数据库[5−7]中的一种常用技术, 可实现用户对数据查询结果的验证, 其核心思想是: 由数据拥有者生成并在数据查询方实时维护一个通过哈希算法生成的数据摘要, 在进行数据更新与查询时, 双方通过摘要进行完整性验证.

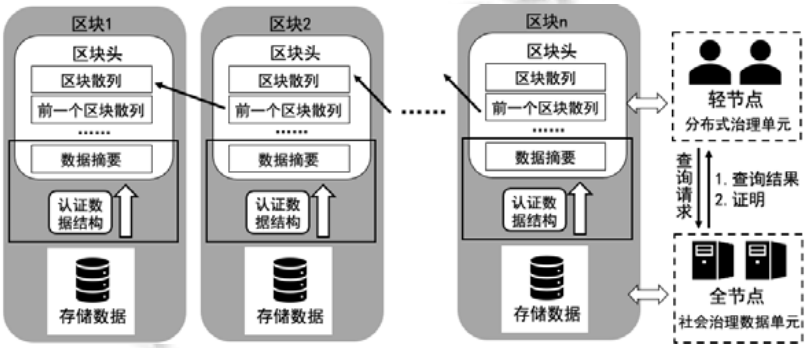

图 5 基于区块链的集成治理结构

基于上述区块链系统结构, 区块链数据的存储更新过程为: 轻节点首先将更新数据发送给全节点; 接着, 全节点广播更新后的区块及数据摘要; 最后, 所有节点通过验证摘要的正确性达成共识, 更新本地的存储数据. 区块链数据的查询处理过程为: 各节点通过拜占庭容错(BFT)[8]共识机制获取最新数据摘要, 并将自身需求发送至全节点, 全节点执行相应查询操作并做出应答, 通过密码学算法生成一个证明结果. 轻节点根据证明结果与查询输出验证查询应答的正确性与完整性. 针对本地数据存储资源、计算资源需求量大的挑战, 基于认证数据结构, 通过两类节点的信息交互完成数据的存储、更新与查询操作, 减少对本地节点资源的依赖, 提高系统的工作效率, 提升了区块链系统在现代社会治理体系中的实用性. 全节点、轻节点两类节点的分类化存储提高了系统效率, 为复杂系统中大量节点的事务处理提供了结构基础.

- 索引结构设计

由于治理数据来源多方、结构异质, 区块链在数据存储时存在支持查询少、访问效率低等问题. 因此, 基于前述匹配模式, 本文构建了层次化存储的高效查询索引: 首先, 根据全局模式和本地模式属性的匹配关系构建层次化索引; 然后, 分别对其生成相关联的 Merkle 值, 实现索引的可信验证. 其中, 全局模式同时包括 Merkle 树树根, 其 Merkle 值由对应的本地模式的 Merkle 值经过 SHA-2 等加密散列函数得出; 本地模式存储 Merkle 树中间节点, 其 Merkle 值由其所有本地模式属性的 Merkle 值经过加密散列函数得出; 本地模式属性的 Merkle 值由该本地模式拥有方生成. 索引拥有者可以通过将子节点的 Merkle 值经过散列函数后与索引中父节点 Merkle 值比较来检查索引是否被篡改, 从而保证索引的可信度. 同时, 为了对后续查询和分析优化, 链下拟基于各治理单元的本地数据形成数据摘要, 包含各属性的总数量、总和、方差等信息, 从而构造全局模式-本地模式-本地数据摘要的层次化异构数据索引, 查询者可以通过该索引快速访问到本地数据, 并基于数据摘要优化查询计划.

5 社会治理层

系统的社会治理层主要负责在区块链层提供的共享访问控制的基础上, 通过精准智能技术实现社会治理的智能决策. 社会治理中, 智能决策的重要问题是对社会系统参与者的复杂行为进行建模和预测. 下面首先针对社会治理中复杂问题的突发性、随机性特征介绍如何建立复杂网络模型, 然后结合疫情防控中超级传播子的寻找算法, 介绍精准智能技术的应用案例.

5.1 复杂网络模型构建

当代社会系统可以看成是一个由独立个体和个体间通过接触、社会联系、互联网等关系建立的连接而形成的复杂信息系统, 这种实际的复杂信息系统通常具有巨大规模和动态规律等特性. 现今社会各种安全问题越来越具有全局突发性和随机特征, 应对这些新的安全挑战, 要求我们对社会中的各种复杂行为的动态传播过程有系统性的掌控. 大规模复杂系统群体行为具有突发性、随机性和动态等演化特征, 是在复杂框架下通过动态迭代演化实现的. 研究复杂系统的智能演化, 需要给出群体在多尺度、多层次下的表示模型, 并揭示个体之间在随机性和非线性关系的作用下的演化过程的动力学规律和特征. 信息扩散是社会治理问题中具有代表性的典型案例, 下面以此为例总结社会治理中复杂问题的突发性和随机性特征, 并建立动力学数学模型.

信息传播是复杂系统中的重要动态过程, 也是分布式社会协同治理体系的关键组成部分. 这种信息传播包括自上而下的传播和自组织传播等方式. 个人用户通过社交媒体讨论某些相关自然灾害的信息, 有助于提前预测自然灾害的演化规律和传播路径, 制定相关预防和应对策略. 搜索引擎的索引数据的动态分析, 有助于了解当下社会中的流行病传播和讨论情况. 在保证信息来源的可靠性的基础上, 通过在线社交媒体等手段向全社会及时通告疫情的发生, 对于疾病预防、稳定社会秩序等诸多社会治理方面具有不可忽视的作用. 对于社交媒体上的真实消息和虚假信息的识别, 也成为近年来社会治理的重点关注对象和研究内容. 虚假的政治新闻, 有关恐怖主义、自然灾害或金融信息的虚假新闻, 对于受众群体更具有煽动性. 复杂系统的行为特征, 本质上是研究系统要素之间的随机非线性关系, 建立动力学数学模型, 揭示系统全局演化规律, 以达到高效、分布式、协同控制动态系统的目的. 对于复杂系统的动力学过程, 一般可以抽象成由代表个体状态集合

的互不相交的若干群体组成, 由于个体成员的状态发生持续的随机变化, 群体中个体数量可以用如下原理进行刻画:

$$\partial_t X^{[m]} = \sum_{h,g} v_{h,g}^m a_{h,g} N^{-1} X^{[h]} X^{[g]} + \sum_h v_h^m a_h X^{[h]}.$$

其中, $X^{[m]}$是状态 m 的个体数量; $a_{h,g}$ 和 a_h 是动态过程群体之间转化率; $v_{h,g}^m$ 和 v_h^m 取值 1, 0 或–1, 用来表示根据群体间的耦合交互左右产生的群体 m 个体的变化数量; N 是系统个体总数. 这一类系统包含了各群体之间的一阶线性关系和二阶非线性关系, 其难点在于如何确立解析的耦合关系表达式, 以及如何拟合动力学方程中的各项参数. 确定各群体间的耦合关系, 可以利用动力系统的数学方法预测复杂系统的周期轨道和稳态, 从系统初始状态得到群体的动态统计规律. 确定系统参数集合后, 能够进一步分析系统的可预测性和混沌效应, 获得精确的系统演化固有规律, 为科学调控和协同治理提供理论模型.

5.2 精准智能技术应用

下面以疫情防控中寻找超级传播者的问题为例, 介绍精准智能在社会治理中的应用. 在复杂网络的传播模型中, 在异质网络上的 SIR 传播过程是一类经典模型. 在疾病传播模型中, 网络中的个体分为 3 类, 即易感者(susceptible)、感染者(infected)和康复者(recovered). 每个单位时间中, 感染者有一定概率β传染一位邻居节点的易感者, 而有概率μ成为康复者. 而在信息传播模型中, 网络中的个体可以被定义为相对应的群体, 即信息未知者、信息传播者和信息厌恶者. 每个单位时间中, 传播者有一定概率β将信息传播给一位邻居节点的未知者, 使其成为信息传播者, 而传播者有概率μ成为信息厌恶者并停止传播信息. 在此基础上, 我们考虑了一类典型的多维度独立传播者模型, 这类问题的动力学模型可表示为

$$\frac{\mathrm{d}i_k(t)}{\mathrm{d}t} = -\mu i_k(t) + \beta k s_k(t)\theta_k(t) + \sigma(t) f(i_k(t), s_k(t)),$$

其中, $i_k(t)$是度为 k 的感染者占总人口的比率; $s_k(t)$是易感者的比率; $\theta_k(t) = \dfrac{\sum_{k'}(k'-1)p_k i_{k'}(t)}{\langle k\rangle}$, 其中, p_k 代表度为 k 的节点占总节点的比率. 这考虑到了由于网络节点度的不均匀分布, 从随机节点沿网络中连边到达指定节点的概率不同, 正比于该节点的度数.

在图 6 中, 图 6(A)是初始传播时刻, 有一个传播者出现; 图 6(B)中, 该传播者将疾病传染给其邻居节点, 使其成为一个普通传播者; 图 6(C)中, 有一个非传统传播渠道感染、与现有感染者没有关联的传播者, 即独立传播者.

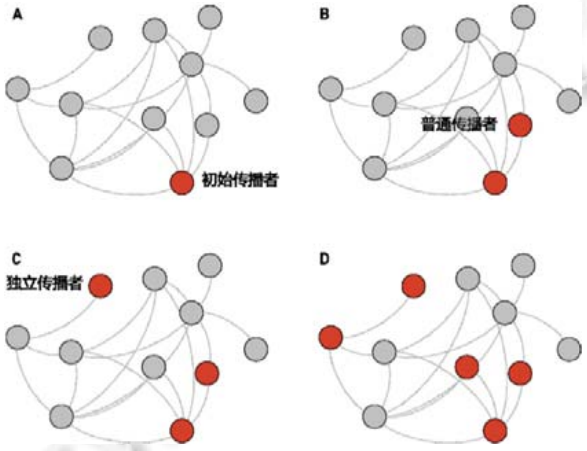

图 6 有独立传播者参与的复杂网络上的传播过程

有别于传统传播模型, 我们考虑现实中一类具有突发性的未知传染源的节点, 称为独立传播者. 独立传播者的引入, 有助于精准刻画这一类突发性和随机性感染事件. 我们定义$\sigma(t)$是独立传播者产生的概率, 正比

于非线性项 $f(i_k(t),s_k(t))$, 其具体表达式取决于疾病传播的动力学特征、可能传输途径、防控策略等诸多因素. 一个简单的例子是定义 $f=i_k(t),s_k(t)$. 这一类模型很好地刻画了网络异质性带来的与均质网络的传播动力学差异. 我们将这类建模推广到 SI, SIS 等疾病传播和信息传播模型中, 发展了多维复杂系统传播过程的非线性和时空耦合的动力学理论[9−11].

在现实社会治理中, 网络的拓扑结构呈现动态演化规律, 我们需要充分考虑结构的动态特征, 对传统模型加以修正和改进, 达到智能、精准预测和社会治理的目标.

例如, 我们考虑复杂网络随时间变化的生成函数:

$$G(x,t)=\sum_{k=0}^{\infty}p_k(t)x^k\ ,$$

其中, $p_k(t)$为 t 时间度为 k 的节点占总节点的比率. 网络结构的动态变化可以通过 $p_k(t)=\frac{1}{k!}\frac{\partial_k G(x,t)}{\partial x^k}\Big|_{x=0}$ 进行动态统计刻画. 例如, 在疾病大流行期间进行动态居家隔离, 减少复杂系统中节点度的平均值以及核心节点的度数, 可以有效控制疾病的传播速度和范围. 而上述表达范式可以刻画这种动态的网络度分布, 进而提高系统控制和预测的准确性和鲁棒性.

在精准构建复杂信息系统拓扑结构的基础上, 寻找复杂网络的中心节点, 即超级传播子, 是针对网络拓扑结构控制系统成员的动态行为的关键方法. 其中, 网络节点的中心性的传统定义有度、中介中心性、特征向量中心性等等, 也有诸如 PageRank[12−14]、Collective Influence (CI)[15−17]、High Degree Adaptive (HDA)[18,19] 等较新的方法. PageRank 可以给出每个节点的中心性的数值; CI 和 HDA 方法虽然不能给出节点中心性的数值, 但可以给出一个反映节点中心性的排名. 其中, HDA 方法是基于度分布的一类有效寻找超级传播子的方法, 比较 CI 而言更为简洁和高效. HDA 方法的伪代码如算法 1 所示.

算法 1. HDA 方法.

Input: G: 有向图.

Output: L: 有序的节点列表.

1: $N\leftarrow G$ 的节点数量
2: **for** $1\leqslant i\leqslant N$ **do**
3: $v_{\max}\leftarrow\varnothing$
4: **for** v' in G **do**
5: **if** $v'.degree>v_{\max}.degree$ **then**
6: $v_{\max}\leftarrow v'$
7: 将节点 $v_{\max}$ 以及其所有连边移除, 并更新图 G
8: 节点 v 的中心性排名设为 i
9: 按中心性排名依次添加节点到有序列表 L 中
10: **return** L

与度中心性不同, HDA 方法保证每一次寻找到的节点都是排除中心性排名更高节点之后的子图中度最大的节点, 这使得该方法在寻找超级传播子中要绝对优于度中心性. 而在实际复杂网络中, HDA 方法也同样接近现有最优方法的结果. 在新冠疫情大流行期间, 各国政府在制定防控政策时需要考虑两个互相矛盾的层面: 让更多的人居家隔离以控制疫情, 同时还要最小化防控措施对经济发展产生的负面影响. 居家隔离等防控措施不可避免地让很多经济活动无法正常进行, 造成人员失业和经济损失. 在精准控制个人接触网络的基础上, 利用区块链建立的分布式大数据库, 我们可以依据个人工作类型对群体进行细分, 并决定大流行病期间如何安排某些特定群体居家. 我们可以选择让可居家办公的群体, 例如某些专业的教师和学生、政府机关文案人员、信息产业从业人员等等, 保持社交距离的同时继续进行工作. 在此基础上, 优先隔离一定比例剩余个体中的超级传播子. 超级传播子的识别算法可参照前述各类模型, 并在实证网络中验证其有效性. 隔离比例需要

根据疫情发展规模、局部地区的网络联通性等性质进行大规模数值模拟, 以确定最佳比例. 在此基础上, 我们还可以参考工作种类对其他国民经济部门的贡献和相互依赖关系, 以修正超级传播子的选取策略.

6 系统实验验证

本节针对安全多方计算的基础操作与精准智能治理效果进行实验验证. 本系统的安全计算层可以实现高效的数据查询, 如寻找特定区域特定时间范围内出现过的密接人员, 助力于疫情防控中的人员排查. 下文首先以范围查询、最近邻查询为例, 验证安全计算层的查询效率、通信开销; 然后, 通过疫情仿真实验, 验证社会治理层 HDA 超级传播子控制下的防疫效果.

6.1 安全计算效率实验

6.1.1 实验设定

- 实验环境

本文通过 5 台服务器搭建安全多方计算场景. 服务器操作系为 Ubuntu 18.04.5 LTS, 内存 64 GB, 配有 32 核 Intel(R) Xeon(R) Gold 5118 2.30 GH 的 CPU. 4 台服务器通过运行多个进程模拟不同层级的多方治理单元, 1 台服务器作为治理的应用方, 通过集成多方数据执行查询操作. 实验在带宽 10 GB/s 的网络环境中进行.

- 实验数据

本实验采用北京市 10 家出租车公司的轨迹数据, 数据集中包含 1 029 081 辆出租车 2 个月的轨迹数据. 数据的空间范围是北纬 39.5°–北纬 42.0°, 东经 115.5°–东经 117.2°, 数据集包含多方数据, 因此不需要重新进行数据划分.

- 比较算法

本系统基于 Hu-Fu 中的安全查询算子[20]实现多方查询操作. 在实验中比较的基线查询算法包括明文查询算法、近期的安全查询系统 Conclave[21]和 SMCQL[22].

- 明文查询. 明文查询不使用任何安全加密技术, 直接聚合多方的明文查询结果. 虽然明文查询算法难以满足数据安全要求, 但是可将其作为查询算法计算效率与通信开销的最优数值.
- Conclave 系统. Conclave 是一种基于秘密共享[4]的安全多方计算技术, 其基于 Sharemind 实现基本的安全查询操作. 本实验将其应用于多方范围查询与多方近邻查询并与之比较.
- SMCQL 系统. SMCQL 是一种能够将 SQL 查询原语转换为 ObliVM[23]的安全计算原语, 在此基础上实现多方安全查询, 进而实现安全的聚合操作. 由于 ObliVM 仅支持两方操作, 因此 SMCQL 系统仅支持两方查询. 本实验将其应用于多方范围查询与多方近邻查询并与之比较.

- 实验参数

在多方范围查询与多方近邻查询中, 默认参与方数目为 6. 变化参数见表 1, 在多方范围查询中分别改变查询范围与数据规模, 其中, 加粗的参数为默认参数; 在多方近邻查询中, 分别改变查询的近邻数目与数据规模. 由于 SMCQL 仅支持两方实验, 故在治理单元为两方的场景下单独进行比较.

表 1 实验参数

操作名称	参数名称	参数范围
多方范围查询	查询范围	10^{-5}%, 10^{-4}%, **10^{-3}%**, 10^{-2}%, 10^{-1}%
	数据规模	10^4, 10^5, 10^6, **10^7**, 10^8
多方近邻查询	近邻数目	4, 8, **16**, 32, 64
	数据规模	10^4, 10^5, 10^6, **10^7**, 10^8

6.1.2 多方范围查询

- 改变查询范围

首先分析改变查询范围计算多方范围查询的运行时间. 从图 7 中可以看出, 相比于 Conclave 系统, 本系统实现的安全查询算子运行时间明显更短, 本系统的多方范围查询计算时间约是 Conclave 系统的 43%. 本系

统在多方范围查询上的计算时间是明文计算的 3.4 倍, 而 Conclave 系统则是明文计算的 8.8 倍. 其次分析不同查询范围下的通信开销. 本系统多方范围查询的通信开销是明文计算的 4.2 倍, 而 Conclave 的通信开销可达明文计算的 69 倍. 因此, 无论是查询效率方面还是在通信开销方面, 本系统相比 Conclave 具有明显优势.

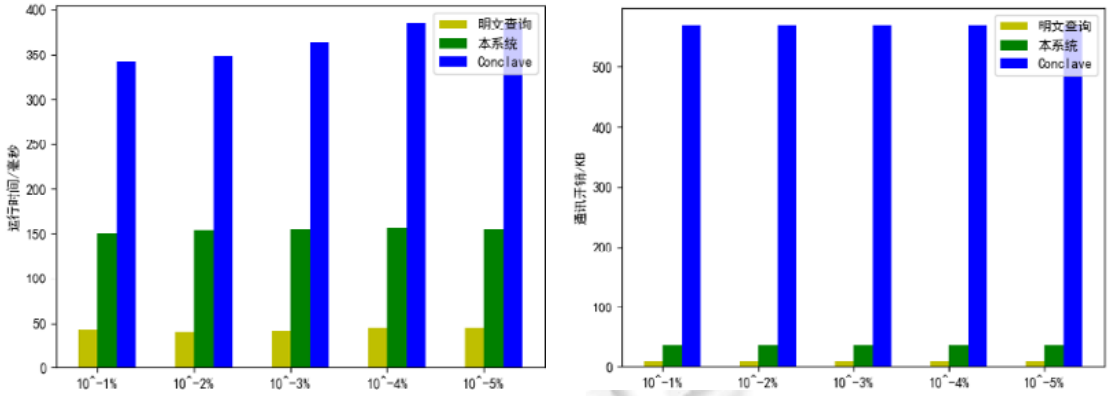

图 7 不同查询范围下的多方范围查询实验结果

- 改变数据规模

从图 8 中可以看出, 在不同数据规模下的实验结果与不同查询范围实验结果相似, 本系统在查询效率与通信开销上均优于 Conclave 系统. 在查询效率方面, 本系统的运行时间是 Conclave 系统的 38.8%–40.9%; 在通信开销方面, 本系统的优势更加明显, 通信开销仅为 Conclave 系统的 4.4%. 相比于明文计算, 通信开销平均增长了 3.3 倍, 处于可接受范围.

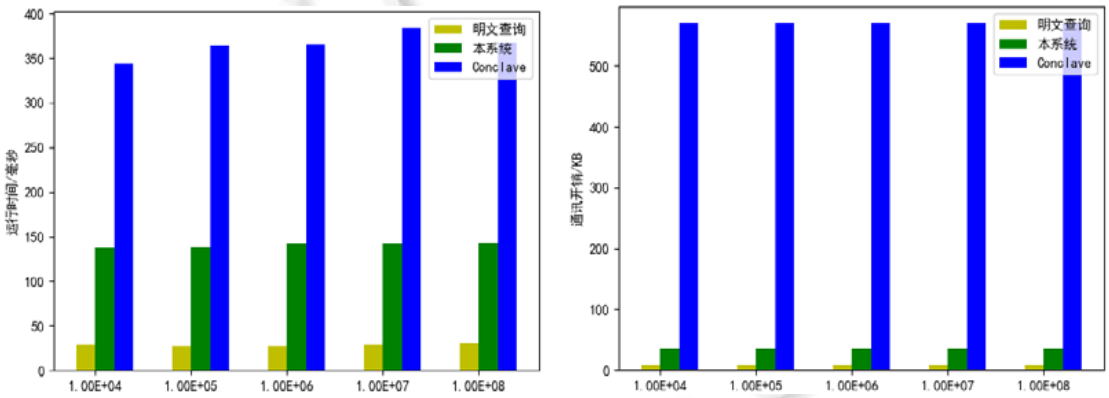

图 8 不同数据规模下的多方范围查询实验结果

- 与 SMCQL 比较

本系统与 SMCQL 在参与方数目为两方时进行比较, 其他参数选择默认参数. 从表 2 中可以看出, 在多方范围查询上, 本系统相比 SMCQL 系统在运行时间与通信开销方面优势明显. 本系统范围查询时间是 SMCQL 的 31.3%, 而通信开销仅为 SMCQL 的 63%.

表 2 SMCQL 比较结果

名称	参数名称	运行时间(ms)	通信开销(KB)
明文计算	范围查询	25.05	2.88
	近邻查询	21.82	9.43
SMCQL	范围查询	425.07	56.61
	近邻查询	1 604.21	2 072
本系统	范围查询	133.17	35.29
	近邻查询	26.07	39.41

6.1.3 多方近邻查询

- 改变近邻数目

首先分析不同近邻数目下各系统的安全查询时间. 从图 9 中可以看出, 在近邻较小时, 如近邻数为 2 时, Conclave 系统的查询时间少于本系统; 但是随着近邻数目增多, Conclave 系统查询时间增长迅速, 本系统查询效率优势明显. 当查询近邻数为 8 时, 本系统的查询时间为 Conclave 的 68.3%; 当查询近邻数目为 64 时, Conclave 系统的多方近邻查询的运行时间是本系统的 23.7 倍. 在通信开销方面, Conclave 系统多方近邻查询所需的通信开销巨大, 高达 55 656 KB, 超过可接受范围; 而明文计算与本系统所需的通信开销均小于 100 KB,属于可接受范围.

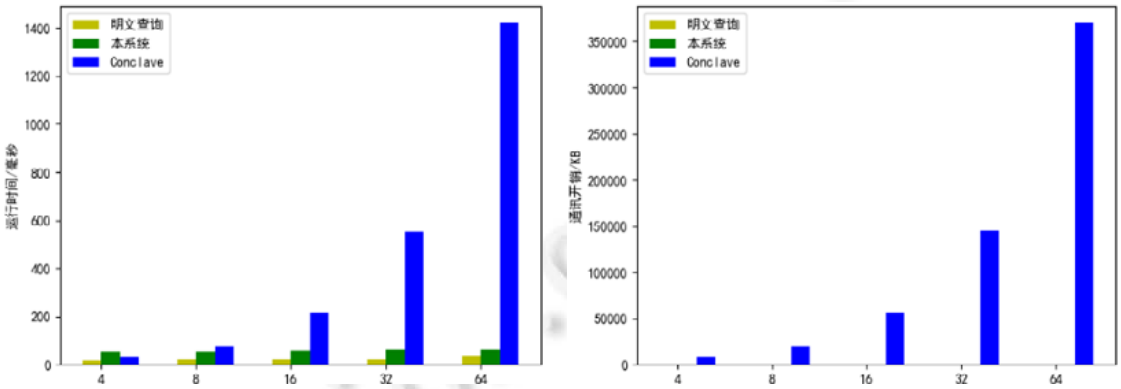

图 9 不同近邻数目下的多方近邻查询

- 改变数据规模

如图 10 所示, 改变数据规模对系统运行时间与通信开销影响有限. 本系统在运行时间接近明文计算, 仅为 Conclave 系统的 23.9%. 在通信开销方面, 与图 9 中结果类似, Conclave 系统执行多方近邻查询开销巨大, 分别是本系统和明文计算开销的 1 600 倍和 6 500 倍. 本系统的通信开销小于 100 KB, 可以接受.

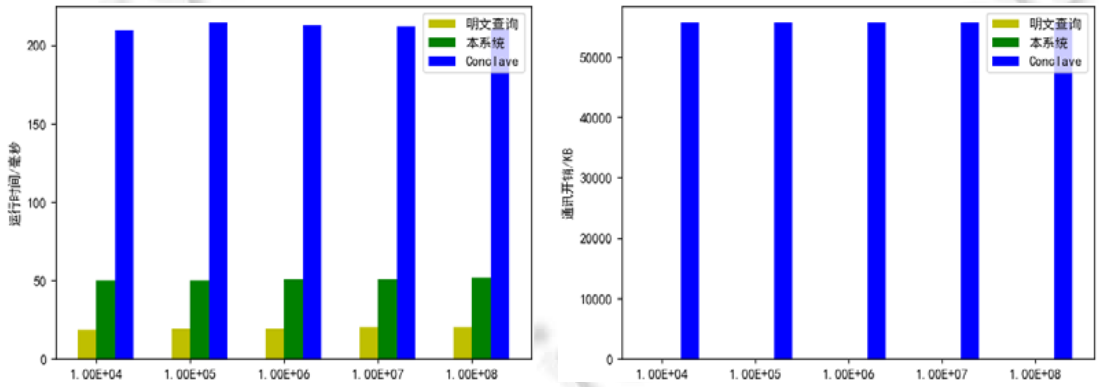

图 10 不同数据规模下的多方近邻查询实验结果

- 与 SMCQL 比较

从表 2 中可得: 在两方场景下, 本系统在查询效率与通信开销方面均优于 SMCQL 系统. SMCQL 的运行时间是本系统的 61.7 倍, 通信开销是本系统的 53 倍, 而本系统的各项指标均与明文计算处于同一量级, 优势明显.

6.2 社会治理仿真实验

基于 HDA 方法进行仿真实验. 模拟一个具有 20 万从业人员的城市, 不同工作种类在工作中与人接触次

数不同, 以此作为度分布建立一个接触网络. 接触的类型包括在家接触、交通接触和工作接触. 最初, 随机选取网络单一节点感染, 并根据 SEIR 模型进行传播模拟. 图 11 中显示了 4 种防疫策略下的病毒传播曲线, 其中, 粗实线代表该策略下 1 000 次模拟的平均值, 每一条细线代表了单次模拟. 无干预策略是对病毒传播不采取任何人为措施; 居家办公策略是指让大约 40%可以居家办公的从业人员在家办公; 居家办公+随机隔离 10%节点策略是指在有条件人员居家办公的基础上, 随机隔离 10%的员工; 居家办公+隔离 10%中心节点策略是指在有条件人员居家办公的基础上, 将剩余员工中的网络中心指数最高的 10%员工隔离.

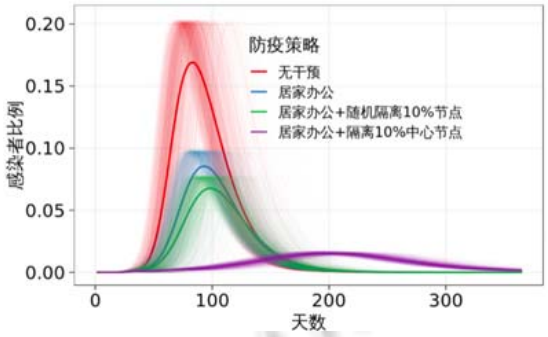

图 11 模拟新冠病毒传播示意图

可以看出, 不同策略的病毒曲线差异巨大, 而同种策略下的病毒曲线也存在较大的随机性差异. 从绿色实线中, 对比蓝色实线, 我们可以看出复杂网络拓扑结构对于传播动力学的显著影响. 紫色实线表示我们安排可居家办公人员居家办公, 并在此基础上隔离剩余复杂网络中的前 10%的中心节点. 这一策略要明显优于隔离同样人数但随机选取隔离人员的绿色实线. 该实验结果充分表明: 结合大数据构建精准的复杂系统底层结构, 并对其采取有效的结构分析, 能够智能分析复杂系统中的非线性、随机性的演化过程, 控制随机突发事件在社会系统中的爆发式传播, 显著提高社会治理效果.

7 相关工作

如前文所述, 安全多方计算、区块链与精准智能是分布式社会治理智能系统中的关键技术. 下面, 本节分别从安全多方计算、区块链和精准智能这 3 个方面介绍相关工作.

7.1 安全多方计算

安全多方计算的目标是在无可信第三方的条件下, 以各方的数据为输入联合完成目标函数的计算, 完成计算后, 各数据拥有方只能获得函数最终计算结果, 而不能获得其他信息[3]. 在分布式社会治理的场景中, 各方数据的隐私安全受到严格要求, 安全多方计算技术可以为社会治理的数据计算、查询提供解决方案. 混淆电路[24]和秘密共享[4]是安全多方计算的两类核心技术: 混淆电路将复杂目标函数分解为逻辑电路门, 每个基础门电路中, 通过真值表混淆与不经意传输(oblivious transfer, OT)保护输入数据的安全; 秘密共享技术将秘密即需要保护的数据划分为多个份额, 分别分发给多个参与方, 只有多个参与方共同拿出份额才能重构还原出秘密. 一般来说, 混淆电路技术在计算通用性上更强, 而秘密共享技术在参与方数量的可扩展性上更强. 目前已发展出多个安全多方计算的工具库, 例如 SPDZ[25]、Sharemind 等. 近年来出现了安全多方计算的开发工具库, 如 ObliVM[23]与 Obliv-C[26], 其主要面向缺乏专业密码学背景知识的开发人员. 使用时只要按开发工具库所定义的编程语言模式, 编写计算流程模板即可. 开发工具库将模板翻译为混淆电路. 这些开发工具库大大提升了使用安全多方计算技术的开发效率.

数据联邦是基于安全多方计算技术实现对多数据拥有方联合计算的安全保护. 2017 年提出的 SMCQL[22]系统是基于 ObliVM 工具库搭建的. SMCQL 由可信第三方与多个数据拥有方组成. 可信第三方接收用户的 SQL 查询, 并对该查询进行解析生成可执行查询计划, 然后将该计划分发给各数据拥有方执行, 各方将最终

执行结果发送给可信第三方, 得到最终的查询结果. 该过程由各数据拥有方在本地执行查询, 从而避免了数据离开本地. 而随后出现的 Conclave 系统进一步增强了数据联邦系统在数据规模上的可扩展性. Conclave 系统的各数据拥有方支持 Spark 大数据计算引擎, 并通过 Obliv-C 与 Sharemind 工具库实现安全保护功能. 然而 SMCQL 系统与 Conclave 系统仅能支持 2–3 个数据拥有方参与, 且系统计算效率仍有较大提升空间.

7.2 区块链技术

区块链是一种基于密码学的分布式存储技术. 2009 年, 《软件可信性动力学特征及其演化复杂性》[27]一文揭示了分布式可信软件系统的动态复杂性, 探讨了构建分布式可信系统的基本科学问题和建立系统可信性的度量理论等. 在此之后, 区块链技术成为学术界与工业界的研究热点方向, 其发展从以比特币为代表的电子货币逐渐发展为以智能合约为核心的 2.0 阶段. 包括 IBM[28,29]、Oracle[30]、华为[31]等企业在内, 均建立了自身的区块链系统, 并应用到实际的生产生活中. 未来的区块链 3.0 阶段将以可编程社会为核心特征, 实现去中心化的社会治理, 成为实现社会协同治理的重要工具.

区块链技术是利用块链式数据结构来验证与存储数据、利用分布式节点共识算法来生成和更新数据、利用密码学的方式保证数据传输和访问的安全、利用由自动化脚本代码组成的智能合约来编程和操作数据的一种全新的分布式基础架构与计算范式[32]. 区块链技术的核心理论包括共识机制算法[33,34]、安全与隐私保护算法[35,36]、高效扩展算法[37,38]这 3 个方面. 共识协议是在群体中动态达成一致的算法. 与多数表决相比, 共识协议更加强调整个群体可以通过达成一致来实现最大收益, 避免由于群体中的恶意行为而导致的系统崩塌. 本系统通过区块链技术中的共识算法、智能合约等理论, 搭建一个去中心化的安全可信环境, 实现多方数据访问权限的一网通管, 为分布式的社会治理智能算法提供数据平台基础.

7.3 精准智能理论

大数据时代, 人工智能理论得到了飞速的发展, 一些成熟的深度学习技术已经可以应用于图像识别[39]、自然语言处理[40]、无人驾驶[41]等各个领域. 然而, 现有的基于统计方法的动态线性学习技术在复杂场景中仍有不足之处. 在现实场景中, 数据之间的关系往往是随机、非线性的, 仅通过动态的线性学习技术去逼近非线性的系统具有不足之处. 现有的人工智能技术在复杂环境下的可解释性、泛化性与可复现性仍有不足. 因此, 设计基于复杂系统环境的动态非线性智能技术, 研究智能算法在极端、恶意情况下的稳定性与可靠性, 探索基于数理规律和先验知识的非线性关系表征, 是人工智能领域未来的潜在突破口.

对此, 文献[42]中提出了精准智能的概念, 创建了基于复杂性与多尺度分析的新一代人工智能理论. 精准智能理论主要包括复杂数据科学感知、复杂系统精准构建、复杂行为智能分析这 3 个层次. 下面进行简要介绍. 复杂数据科学感知主要解决无法真实反映内部数据结构的问题. 复杂系统中的数据具有规模大、噪音多、分布异构的特点, 因此, 直接通过现有的人工智能算法对数据进行训练会阻碍模型的收敛, 大大降低算法效率. 精准智能理论则聚焦于研究复杂大数据的内部结构, 基于数据的内部数理规律设计精准的智能算法. 复杂系统精准构建主要研究复杂系统内部各组成部分之间非线性的关联关系. 在复杂系统中, 各节点之间的关系错综复杂, 每个状态的产生不仅受历史数据的影响, 同时还具有随机性, 基于线性的智能技术难以进行有效分析. 因此需要利用面向全局的非线性系统建模方法, 在复杂数据科学感知的基础上实现对复杂系统的精准构建. 复杂行为智能分析的主要目标是在现有的人工智能理论的基础上, 研究基于复杂系统的行为学习方法. 在复杂系统中, 各对象的行为具有多样性、不可控性、复杂性、随机性的特点, 精准智能理论融合了动力系统和统计物理理论, 实现对系统演化模式的精准模拟、预测.

8 总结与展望

本文提出了一种基于大数据的分布式社会智能治理系统, 以应对社会治理复杂问题带来的挑战. 该系统通过安全多方计算技术实现安全计算基础算子, 为社会治理提供多方安全查询接口; 基于区块链技术搭建去中心化空间结构, 为突发事件的高效处理提供数据平台支撑; 通过精准智能构建非线性模型, 建立面向系统

行为演进和全局动态分析的可解释、可调控人工智能, 为社会治理中的非线性因素提供科学决策支撑. 本文提出的分布式社会治理智能系统将进一步支撑包括公共卫生安全、智慧交通等社会治理的各类应用, 为新时代社会治理水平提升提供平台和决策助力.

References:

[1] Chen S. What new opportunities and challenges are facing social governance in the context of the new round of scientific and technological revolution. Governance, 2019, 43: 34−38 (in Chinese with English abstract).

[2] Zhang Y. Take the road of socialist social governance with Chinese characteristics. Qiushi, 2018, (6): 57−58 (in Chinese with English abstract).

[3] Yao AC. Protocols for secure computations. In: Proc. of the 23rd Annual Symp. on Foundations of Computer Science. IEEE, 1982. 160−164.

[4] Beimel A. Secret-sharing schemes: A survey. In: Proc. of the 2011 Int'l Conf. on Coding and Cryptology. Berlin, Heidelberg: Springer, 2011. 11−46.

[5] Li FF, Hadjieleftheriou M, Kollios G, Reyzin L. Dynamic authenticated index structures for outsourced databases. In: Proc. of the 2006 ACM SIGMOD Int'l Conf. on Management of Data. Chicago: ACM, 2006. 121−132.

[6] Miller A, Hicks M, Katz J, Shi E. Authenticated data structures, generically. In: Proc. of the Symp. on Principles of Programming Languages. San Diego: ACM, 2014. 411−424.

[7] Yang Y, Papadias D, Papadopoulos S, Kalnis P. Authenticated join processing in outsourced databases. In: Proc. of the 2009 ACM SIGMOD Int'l Conf. on Management of Data. Providence: ACM, 2009. 5−18.

[8] Yin M, Malkhi D, Reiter MK, Gueta GG, Abraham I. Hotstuff: BFT consensus with linearity and responsiveness. In: Proc. of the 2019 ACM Symp. on Principles of Distributed Computing. Toronto: ACM, 2019. 347−356.

[9] Li W, Tang S, Pei S, Yan S, Jiang S, Teng X, Zheng Z. The rumor diffusion process with emerging independent spreaders in complex networks. Physica A: Statistical Mechanics and Its Applications, 2014, 397: 121−128.

[10] Ma K, Li W, Guo Q, Zheng X, Zheng Z, Gao C, Tang S. Information spreading in complex networks with participation of independent spreaders. Physica A: Statistical Mechanics and Its Applications, 2018, 492: 21−27.

[11] Ding Q, Li WH, Hu XM, Zheng ZZ, Tang S. The SIS diffusion process in complex networks with independent spreaders. Physica A: Statistical Mechanics and Its Applications, 2020, 546(14): Article No.122921.

[12] Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the Web. Stanford InfoLab, 1999. http://ilpubs.stanford.edu:8090/422/

[13] Langville AN, Meyer CD. Deeper inside pagerank. Internet Mathematics, 2004, 1(3): 335−380.

[14] Xing W, Ghorbani A. Weighted pagerank algorithm. In: Proc. of the 2nd Annual Conf. on Communication Networks and Services Research. IEEE, 2004. 305−314.

[15] Morone F, Min B, Bo L, Mari R, Makse HA. Collective influence algorithm to find influencers via optimal percolation in massively large social media. Scientific Reports, 2016, 6(1): 1−11.

[16] Pei S, Teng X, Shaman J, Morone F, Makse HA. Efficient collective influence maximization in cascading processes with first-order transitions. Scientific Reports, 2017, 7(1): 1−13.

[17] Zhang H, Zhang H, Wu C. Identification of essential proteins based on centrality methods using improved collective influence algorithm. In: Proc. of the 2019 IEEE Global Conf. on Signal and Information Processing. IEEE, 2019. 1−5.

[18] Cohen R, Erez K, Ben-Avraham D, Havlin S. Breakdown of the internet under intentional attack. Physical Review Letters, 2001, 86(16): Article No.3682.

[19] Xu H, Yuan H, Duan K, Xie W, Wang Y. Adaptive high-degree cubature Kalman filter in the presence of unknown measurement noise covariance matrix. The Journal of Engineering, 2019, 19: 5697−5701.

[20] Tong YX, Pan XC, Zeng YX. Hu-Fu: Efficient and secure spatial queries over data federation. 2021. https://github.com/BUAA-BDA/Hu-Fu

[21] Volgushev N, Schwarzkopf M, Getchell B, Varia M, Lapets A, Bestavros A. Conclave: Secure multi-party computation on big data. In: Proc. of the 14th EuroSys Conf. ACM, 2019. 1−18.

[22] Bater J, Elliott G, Eggen C, Goel S, Kho AN, Rogers J. SMCQL: Secure query processing for private data networks. Proc. of the VLDB Endowment, 2017, 10(6): 673−684.

[23] Liu C, Wang XS, Nayak K, Huang Y, Shi E. Oblivm: A programming framework for secure computation. In: Proc. of the 2015 IEEE Symp. on Security and Privacy. IEEE, 2015. 359−376.

[24] Bellare M, Hoang VT, Rogaway P. Foundations of garbled circuits. In: Proc. of the 2012 ACM Conf. on Computer and Communications Security. ACM, 2012. 784−796.

[25] Keller M. MP-SPDZ: A versatile framework for multi-party computation. In: Proc. of the 2020 ACM SIGSAC Conf. on Computer and Communications Security. ACM, 2020. 1575−1590.

[26] Zahur S, Evans D. Obliv-C: A language for extensible data-oblivious computation. Cryptology ePrint Archive, Report 2015/1153, 2015. https://ia.cr/2015/1153

[27] Zheng ZM, Ma SL, Li W, Wei W, Jiang X, Zhang ZL, Guo BH. Dynamical characteristics of software trustworthiness and their evolutionary complexity. Science in China (Series F, Information sciences), 2009, 39(9): 946−950 (in Chinese with English abstract).

[28] IBM Blockchain. Enterprise blockchain solutions and services. 2018. https://www.ibm.com/blockchain

[29] Ahram T, Sargolzaei A, Sargolzaei S, Daniels J, Amaba B. Blockchain technology innovations. In: Proc. of the 2017 IEEE Technology & Engineering Management Conf. ACM, 2017. 137−141.

[30] Oracle. Transforming the enterprise with oracle blockchain platform. 2018. https://www.oracle.com/cloud/blockchain

[31] Huawei. Huawei blockchain whitepaper, toward a trusted digital world. 2018. https://static.huaweicloud.com

[32] China Blockchain Technology and Industrial Development Forum. The white paper of China blockchain technology and application development. 2016 (in Chinese with English abstract).

[33] Sun ZX, Zhang X, Xiang F, Chen L. Survey of storage scalability on blockchain. Ruan Jian Xue Bao/Journal of Software, 2021, 32(1): 1−20 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6111.htm [doi: 10.13328/j.cnki.jos.006111]

[34] Zhang ZW, Wang GR, Xu JL, Du XY. Survey on data management in blockchain systems. Ruan Jian Xue Bao/Journal of Software, 2020, 31(9): 2903−2925 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6091.htm [doi: 10.13328/j.cnki.jos.006091]

[35] Xia Q, Dou WS, Guo KW, Liang G, Zuo C, Zhang FJ. A survey on blockchain consensus protocol. Ruan Jian Xue Bao/Journal of Software, 2021, 32(2): 277−299 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/6150.htm [doi: 10.13328/j.cnki.jos.006150]

[36] Li F, Li ZR, Zhao H. Research on the progress in cross-chain technology of blockchains. Ruan Jian Xue Bao/Journal of Software, 2019, 30(6): 1649−1660 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5741.htm [doi: 10.13328/j.cnki.jos.005741]

[37] Qian WN, Shao QF, Zhu YC, Jin CQ, Zhou AY. Research problems and methods in blockchain and trusted data management. Ruan Jian Xue Bao/Journal of Software, 2018, 29(1): 150−159 (in Chinese with English abstract). http:// www.jos.org.cn/1000-9825/5434.htm [doi: 10.13328/j.cnki.jos.005434]

[38] Liu AD, Du XH, Wang N, Li SZ. Blockchain-based access control mechanism for big data. Ruan Jian Xue Bao/Journal of Software, 2019, 30(9): 2636−2654 (in Chinese with English abstract). http://www.jos.org.cn/1000-9825/5771.htm [doi: 10.13328/j.cnki.jos.005771]

[39] LeCun Y, Bengio Y. Convolutional networks for images, speech, and time series. In: The Handbook of Brain Theory and Neural Networks. 1995.

[40] Pannu A. Artificial intelligence and its application in different areas. Artificial Intelligence, 2015, 4(10): 79−84.

[41] Fridman L, Brown DE, Glazer M, Angell W, Dodd S, Jenik B, Terwilliger J, Kindelsberger J, Ding L, Seaman S, Abraham H. Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation. arXiv preprint arXiv:1711.06976, 2019.

[42] Zheng ZM, Lü JH, Wei W, Tang ST. Refined intelligence theory: Artificial intelligence regarding complex dynamic objects. Scientia Sinica Informationis, 2021, 51(4): 678−690 (in Chinese with English abstract).

附中文参考文献:

[1] 陈升. 新一轮科技革命背景下社会治理面临哪些新机遇和新挑战. 国家治理, 2019, 43: 34–38.

[2] 张翼. 走中国特色社会主义社会治理之路. 求是, 2018, (6): 57–58.

[27] 郑志明, 马世龙, 李未, 韦卫, 姜鑫, 张占利, 郭炳晖. 软件可信性动力学特征及其演化复杂性. 中国科学(F 辑: 信息科学), 2009, 39(9): 946–950.

[32] 中国区块链技术和应用发展白皮书. 2016.

[33] 孙知信, 张鑫, 相峰, 陈露. 区块链存储可扩展性研究进展. 软件学报, 2021, 32(1): 1–20. http://www.jos.org.cn/1000-9825/6111.htm [doi: 10.13328/j.cnki.jos.006111]

[34] 张志威, 王国仁, 徐建良, 杜小勇. 区块链的数据管理技术综述. 软件学报, 2020, 31(9): 2903–2925. http://www.jos.org.cn/1000-9825/6091.htm [doi: 10.13328/j.cnki.jos.006091]

[35] 夏清, 窦文生, 郭凯文, 梁赓, 左春, 张凤军. 区块链共识协议综述. 软件学报, 2021, 32(2): 277–299. http://www.jos.org.cn/1000-9825/6150.htm [doi: 10.13328/j.cnki.jos.006150]

[36] 李芳, 李卓然, 赵赫. 区块链跨链技术进展研究. 软件学报, 2019, 30(6): 1649–1660. http://www.jos.org.cn/1000-9825/5741.htm [doi: 10.13328/j.cnki.jos.005741]

[37] 钱卫宁, 邵奇峰, 朱燕超, 金澈清, 周傲英. 区块链与可信数据管理: 问题与方法. 软件学报, 2018, 29(1): 150–159. http://www.jos.org.cn/1000-9825/5434.htm [doi: 10.13328/j.cnki.jos.005434]

[38] 刘敖迪, 杜学绘, 王娜, 李少卓. 基于区块链的大数据访问控制机制. 软件学报, 2019, 30(9): 2636–2654. http://www.jos.org.cn/1000-9825/5771.htm [doi: 10.13328/j.cnki.jos.005771]

[42] 郑志明, 吕金虎, 韦卫, 唐绍婷. 精准智能理论: 面向复杂动态对象的人工智能. 中国科学: 信息科学, 2021, 51(4): 678–690.

吕卫锋(1972－), 男, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为时空大数据分析处理, 智慧城市, 众包计算, 群体智能.

郑志明(1953－), 男, 博士, 教授, 博士生导师, 中国科学院院士, CCF 会士, 主要研究领域为动力系统, 群体智能, 区块链, 网络信息安全.

童咏昕(1982－), 男, 博士, 教授, 博士生导师, CCF 高级会员, 主要研究领域为联邦学习, 时空大数据分析处理, 智慧城市, 众包计算, 群体智能, 隐私保护.

张瑞升(1998－), 男, 硕士生, 主要研究领域为联邦学习, 时空大数据分析处理, 群体智能, 隐私保护.

魏淑越(1998－), 男, 硕士生, CCF 学生会员, 主要研究领域为时空大数据分析处理, 群体智能,激励机制, 隐私保护.

李卫华(1989－), 男, 博士, 讲师, 主要研究领域为复杂系统, 网络科学, 群体智能, 计算社会学, 大数据科学及建模.