



开放科学
(资源服务)
标识码
(OSID)

基于领域语义地图的区块链研究主题发现及演化分析

张爽¹ 刘非凡¹ 罗双玲² 夏昊翔^{1,3}

1. 大连理工大学系统工程研究所 大连 116024;
2. 大连海事大学航运经济与管理学院 大连 116026;
3. 大连理工大学大数据与智能决策研究中心 大连 116024

摘要: [目的/意义] 区块链作为一种新兴的数据共享技术,对推动产业变革、改变社会治理模式具有重要意义。当前对区块链研究现状的分析多侧重应用或技术方面,或以传统的文献计量为主,而缺少对整体研究概貌和发展过程的探究。为此,本研究将基于所构领域语义地图探究区块链技术的研究态势。[方法/过程] 本研究利用区块链相关研究的学术数据,借助文档嵌入表示学习方法 Doc2Vec、流形学习算法 UMAP 从文献语义关联的角度可视化了研究概貌,基于核密度估计识别了主要研究内容,并且结合关键词和领域语义地图的动态变化分析揭示了其发展过程。[结果/结论] 首先,本研究识别出了区块链研究的七个主要研究主题,其领域语义地图显示各个主题发展态势不同,当前侧重于技术应用方面的研究。其次,研究发现区块链的发展经历了从加密货币交易到智能合约开发到区块链应用多元化三个阶段。最后,探究了区块链研究的演变发展特征,即,相关研究主题范围持续扩张,同时领域内部存在主题分化现象。本研究结果对科研人员把握研究现状,推进技术发展,以及对政府和企业制定围绕区块链创新方面的科技管理政策具有一定的参考价值。

关键词: 区块链; Doc2Vec; UMAP; 核密度估计; 研究主题; 发展特征

中图分类号: G35, TP391

基金项目 教育部人文社科规划项目“综合数据解析和动力学建模的科研协作系统演化模式与机理研究”(18YJA630118), 辽宁省社会科学规划基金项目“大数据背景下图书馆科技情报知识服务的支撑方法研究”(L16BTQ003), 国家自然科学基金面上项目“协同创新网络中的集体智能动态机理研究”(71871042)。

作者简介 张爽(1994-), 博士研究生, 研究方向为机器学习、社会网络分析、科技情报分析; 刘非凡(1991-), 博士研究生, 研究方向为复杂网络、机器学习、科技情报分析; 罗双玲(1978-), 博士, 副教授, 研究方向为知识管理、金融系统; 夏昊翔(1972-), 博士, 教授, 研究方向为复杂系统、计算社会科学、知识管理、知识组织, E-mail: hxxia@dlut.edu.cn。

Topic Detection and Evolution Analysis of Blockchain with the Domain Semantic Map

ZHANG Shuang¹ LIU Feifan¹ LUO Shuangling² XIA Haoxiang^{1,3}

1. Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, China;

2. School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China;

3. Center for Big Data and Intelligent Decision-Making, Dalian University of Technology, Dalian 116024, China

Abstract: [Objective/ Significance]. As an emerging data sharing technology, blockchain plays an important role in promoting industrial transformation and changing social governance mode. However, most of existing research only focus on analysis of the application or the technology status of blockchain, with traditional bibliometric methods, and less attentions are paid on the overall landscape and development paths of the blockchain research. To overcome these limitations, we will explore the research status of blockchain technology based on semantic domain map proposed. [Methods/Process]. In this study, we integrate document representation learning and manifold learning to map the landscape of blockchain field from the perspective of semantic correlation of papers and detect research topics with kernel density estimation. The dynamic change of keywords and the constructed domain map are furthermore analyzed. [Results /Conclusions]. Firstly, this study identifies seven major research topics in the blockchain research. The obtained domain map shows that the development trends of each topic are different, and the current studies focus on the research of technology application. Secondly, it is found that the development of blockchain has gone through three stages, i.e., cryptocurrency transaction, intelligent contract development and diversification of blockchain application. Finally, the development characteristics of the blockchain are found, that is, the scope of the blockchain research continues to expand, and meanwhile the topics within the field keep splitting. These results may help researchers grasp the research status and to promote the technical development of the field, and help enterprises and governments make science-and-technology management policies on the block chain innovation.

Keywords: Blockchain; Doc2Vec; UMAP; kernel density estimation; research topics; development characteristics

引言

区块链技术起源于化名为“中本聪”的黑客于2008年发表的文章《比特币：一种点对点的电子现金系统》。作为比特币的底层技术，区块链正以其去中心化、不可篡改的特性，重构各行各业的体系架构，具有颠覆性的应用价值。目前，全球都在加快区块链技术的发展和“区块链+”产业的布局。2016年中国国务院将区块链作为战略前沿技术写入《“十三五”国家信息化规划》中，此后从工信部、商业部和网

信办等各部门到北京、上海和浙江等地方也陆续出台一系列规划和意见以推进区块链创新和应用。同年，韩国央行在报告中提出鼓励探索区块链技术，并且多家金融和技术公司共同成立了区块链协会。面对世界各国抢占“发展制高点”的紧迫形势，学术界对区块链展开了如火如荼的研究。近年来，区块链研究呈现爆发增长态势，其核心技术和应用都得到了飞速发展。然而相较于其他技术，区块链的相关研究遍布于计算机、通信、金融、管理等多个领域。且目前也缺乏以区块链这一跨学科复合型前沿

技术研究为主的典型核心期刊。这使得学者难以全面准确地把握区块链的研究现状和发展历程,也为创新战略的制定和区块链的应用普及造成了一定的困难。

基于此,近年来国内外学者纷纷对区块链技术进行了综述性研究和探讨。刘硕等^[1]详尽介绍了区块链的模型、核心技术、应用平台以及典型公司,并且对比了其和传统技术系统的差异,对我国区块链技术的发展提出了优化研发环境、制定行业标准和开展应用试点等政策建议。朱岩等^[2],隆苏研等^[3],袁勇和王飞跃^[4]概览了区块链的核心机制和算法并且讨论了其未来发展方向和当前面临的挑战。除了梳理区块链技术的架构、概念等基础知识,学者还分析了区块链技术的应用现状。Casino等^[5]将区块链的应用分为金融、身份验证、政府管理、物联网、健康管理、工业应用、教育和数据管理八方面,并且阐述了各方面的研究现状和局限。也有学者探讨了区块链技术在其研究领域的应用进展,包括信息安全^[6]、能源交易^[7]、教育^[8]等领域。比如,徐嘉辉和马立新^[9]总结了现有基于区块链的分布式电力交易系统的优势和不足。然而,从上述依赖学者对区块链认识的综述研究可获知其发展的局部状况,即理论技术发展或技术在某个具体领域的应用模式,而无法把握区块链的整体研究图景和发展过程。于是,有学者基于专利或论文数据,运用文献计量学方法揭示区块链的研究现状。Firdaus等^[10]通过分析从Scopus数据库获得的2013-2018年的1000篇区块链文献,发现区块链技术能够解决物联网的安全问题,且逐渐应用于健康管理和数据分析领域;在当前区块链研究中,中国、

美国具有较高的论文产出量,而新加坡和瑞士的论文影响力较高。Ante^[11]利用从商业和经济领域收集到的区块链文献,通过耦合引文分析识别出了市场效率、资产定价、匿名交易等五个发展趋势。雷孝平等^[12]通过统计论文和专利数量发现了区块链领域的主要研发国家、研究机构和研究人員,根据关键词和专利分类代码频次揭示了研发热点和应用方向。王发明和朱美娟^[13]以CNKI数据库检索出的432篇区块链相关文献,使用CiteSpace工具对其进行关键词共现分析和突发性检测,追踪了我国区块链研究的热点和前沿问题。然而这些研究大多以传统的文献计量研究为主,聚焦于文献(或专利)的国家分布、机构分布、学科分布以及影响力分析^[14-16],以共词聚类法探究研究热点和趋势^[17,18]。而且,国内大部分研究局限于使用CiteSpace工具,基于百篇文献的小规模数据集(或仅利用CNKI数据)分析研究领域^[19-21],而缺少基于较大规模的区块链相关学术数据的研究主题发现与领域分析。把握区块链的发展趋势,理解其发展特征,有助于科技管理者制定发展政策、理解其他新兴领域的发展模式。然而,当前鲜有研究探究领域的发展态势。

因此,本研究拟利用较为完整的区块链研究的学术数据,借助新兴的机器学习算法绘制领域地图,实现从文献语义层面探究整体研究地貌。并且基于领域语义地图中文献聚集性差异,探测区块链研究的主题结构和发展特点。本研究所用方法突破了传统的共词分析在捕获文献语义关系的局限。相较于共词网络存在边稠密性的问题,领域语义地图更加便于观察主题结构的动态变化,可直观地获知区块链研究

的发展趋势。

1 数据与方法

1.1 数据集描述

以 Web of Science 核心数据库为数据源, 设置检索式 TS=(“blockchain” or “block chain” or “block-chain”)、检索时间跨度为 2008-2020 年, 一共检索到 6,572 篇文献。在化学领域, 根据以上关键词能检出与区块链无关其他专门文献, 因此, 根据文献的期刊来源, 去除数据集中属于化学领域 (如高分子、聚合物) 的文献, 最终得到 5,951 篇文献及其标题、摘要、关键词和年份等题录信息。尽管所得数据集未包含区块链领域的所有文献, 但是检索方式和数据源保证所得数据集包含了与区块链研究直接相关的、学术质量高的科学文献数据, 具有较好的覆盖率与代表性, 可用于区块链研究的文献计量分

析。从图 1 的历年文献数量增长态势, 可以看出区块链研究目前正处于快速发展的阶段。表 1 列举了 30 个高频论文关键词, 从中可以得知当前研究热点, 如 “Smart Contract” (智能合约), “Bitcoin” (比特币), “Internet Of Things” (物联网) 和 “Distributed Ledger” (分布式账本) 等。需要说明的是, 为了使结果更加清晰, 在后续词云图分析中去除了高频词 “Blockchain”。

1.2 研究方法

为实现从文本语义关联的角度探测区块链研究领域的发展过程, 本研究借助文档嵌入表示学习和流形学习算法, 提出了一种新的领域语义地图构建方式。从而, 通过绘制区块链研究的领域地图可视化其研究地貌, 根据文献分布密度识别研究主题, 结合文献时间标签追溯其发展过程。具体方法流程主要包括文档向量化、领域可视化、词云图映射研究内容和累积地图演化分析等。

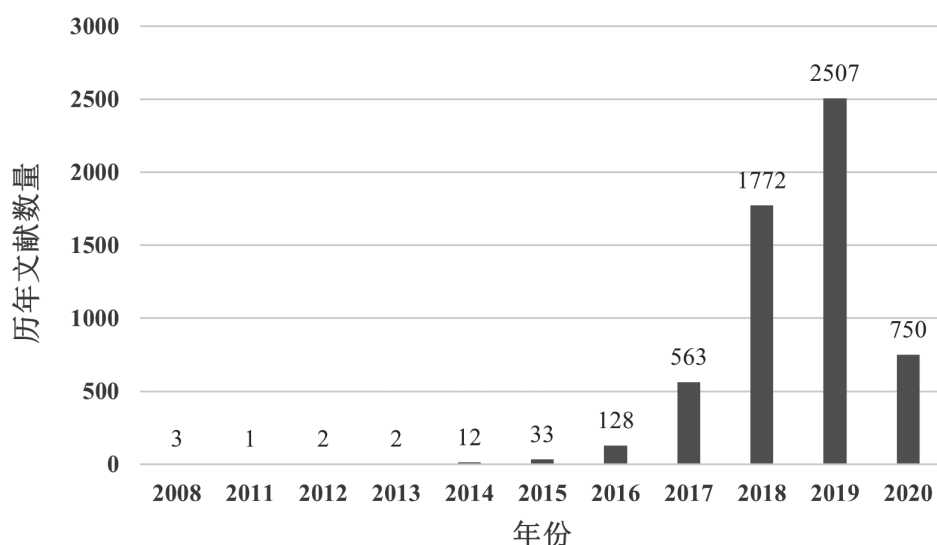


图 1 历年文献数量

表 1 区块链研究的高频关键词 (top30)

关键词	词频	关键词	词频
Blockchain	4041	Decentralization	92
Smart Contract	852	Artificial Intelligence	91
Internet Of Things	676	Authentication	85
Bitcoin	495	Big Data	81
Cryptocurrency	324	Cryptography	76
Security	321	Smart City	74
Ethereum	290	Machine Learning	71
Distributed Ledger	252	Hyperledger Fabric	69
Privacy	227	Healthcare	68
Consensus	208	Smart Grid	66
Supply Chain	159	Scalability	62
Trust	113	Fintech	58
Access Control	108	Data Sharing	51
Cloud Computing	101	Fog Computing	49
Edge Computing	97	Traceability	49

(1) 首先利用 Doc2Vec 捕获区块链科研文献的潜在语义特征。Doc2Vec 作为一种新兴的文档嵌入表示学习算法, 基于深度学习框架能够自动学习文献的词频、语序和语义信息, 将其表示为低维实值向量, 有效地解决了语义鸿沟和计算复杂度问题^[22]。近年来在诸如性别偏见分析^[23]、文化演化^[24]等复杂的文本分析任务中都展现出了强大的潜在特征学习能力。

在获得文献的标题和摘要后, 本研究首先对其进行预处理, 具体步骤包括: 合并标题和摘要、转换大小写、移除符号标点和数字、剔除学术停用词和版权声明。接着采用 PV-DM 框架训练 Doc2Vec 模型。PV-DM 方法的核心思想是根据某个单词的上下文可以预测这个单词^[22]。在 Doc2Vec 中, 每篇文档被标识为一个向量。在训练某个文档向量时, 用滑动窗口的方法产生局部训练集, 选取某个词当预测词, 训练集

中的其他词以及文档向量作为输入词, 然后依据预测结果来调整输入向量。随着训练的进行, 文档向量不断被调整。接着依据自相似性原则判定模型是否收敛以及调整模型参数, 即向模型输入某一文献的语料, 根据模型输出的文档向量计算与该文献最相似的文献是否是它本身, 并统计这种情况的占比。最终, 本研究通过 30 轮训练, 取得 98% 的自相似度, 获得了语料库中每篇文献的 100 维语义向量。

(2) 在高维文档向量空间中, 研究内容相近的文献, 其语义向量会彼此相近形成聚类, 而研究内容不同的则相距较远。也就是说, 高维文档语义向量间拓扑结构对应区块链相关研究的知识结构。所以本研究将文献向量刻画出的高维知识结构映射为二维语义地图进行可视化。UMAP 是一种基于黎曼几何理论的流形学习算法, 它依据“高维空间相近的点在低维空间中也相近”的假设, 将学习到的高维数据的拓扑结构进行低维映射^[25]。不同于 tSNE 等其他算法, UMAP 的理论架构^[25]以及近来的实证分析^[26]都表明, UMAP 具有同时保留数据集全局和局部拓扑结构的能力。

本研究首先运用 UMAP 算法将 100 维文献语义向量转化为二维坐标。运行 UMAP 算法时, 为了度量出各个文献向量在不同维度上的差异, 采用余弦相似度度量文献向量在高维空间中的距离。UMAP 算法以邻居数量和最小距离两个参数实现对数据中局部和全局结构的学习, 本研究参考 UMAP 开发者的测试结果^[25,27], 将其分别设置为 15 和 0.1。接着本研究通过计算文献散点的高斯核密度, 并以等高线刻画文献在二维语义空间中分布的聚散性, 这样所构建出

的领域语义地图即可表征科研数据集的知识空间。地图上特定坐标区域具有特定的语义信息,文献在地图上的临近性反映了其语义层面上知识的相似或相近性。同时,地图上等高线闭环又刻画出了文献分布的异质性,反映了主题结构,即闭环内区域的文献密度明显不同于周围区域,且密度的高低表明该区域是研究热点或是值得深探的研究“蓝海”。不同于常见基于密度的聚类算法以文献聚集性最高的区域为聚类中心划分主题,本研究依据文献聚集的局部异质性,通过提取等高线闭环实现主题识别,即可获得不同研究热度的主题。基于领域地图的主题划分可根据研究目的灵活处理:若了解领域的主要研究内容,可按照密度值提取低密度、区域覆盖范围大的等高线作为主题划分结果;若探究领域内细粒度的研究主题,可按照等高线闭环数量最大化、同时保证一定的文献覆盖率的原则进行提取。通过提取等高线内文献的关键词绘制词云图,可获知该区域所代表的主要研究内容。

(3) 在获得数据集的文档向量后,根据文献发表年份,使用某一年及以前的文献利用所提方法绘制历年的累积领域地图。领域地图上每块区域都具有特定的语义信息,由于文献向量是同时训练获得的,所以历年地图上相同坐标区域表征的语义是相同的。这意味着通过对比历年地图间的地貌差异,可获悉领域演化过程中的研究范围的扩增以及主题涌现等现象。

2 领域语义地图分析

本研究构建了领域语义地图可视化“区块

链”研究概貌,以期能对当前主要研究内容、研究重点,主题的发展态势和主题间的关联性有更直观的认识。图2是区块链领域语义地图,其中,图2(a)为文献分布热力图。依据前述方法可知,该领域地图是区块链相关研究的知识结构在低维空间中的投影,地图邻近性映射出各研究主题间的语义相关性,而地图上坐标区域颜色的深浅以及等高线的数值表明了文献聚集程度,也反映了该区域被探索的程度。整体来看,该领域地图周围分散五个高文献密度的“山峰”,中央形成密度较低的“平台”,且“山峰”之间稀疏地分布少量的文献。这表明区块链的相关研究已形成了多个研究主题,而跨主题、交叉主题的研究仍然较少。

图2(a)中等高线刻画了闭环内文献分布与周围区域文献分布的异质性,反映了区块链相关研究的主题结构,也刻画出了文献聚类结果。为了获得细粒度的研究主题,本研究依据最大化主题数量同时保证较高文献覆盖率的原则,主要提取了6个文献密度显著高于周围区域的和1个密度明显低于周围的区域代表主要研究主题。并且借助这7个区域内文献的关键词和其TFIDF值绘制词云图表征研究内容(如图2(b))。根据词云图可以看出该领域的研究主要分为两部分,地图左侧的四个主题主要关于区块链的应用研究,地图右侧则是区块链的基础技术研究。除此之外,可以看到地图(图2(a))左侧的文献核密度大于地图右侧的三个主题,这意味着目前对区块链的研究并不平衡。整体上来看,当前学者侧重于区块链技术应用的研究,而对基础技术的发展关注较弱。

区块链基础架构主要由应用执行层、共识层、网络层、数据层和硬件设施层组成。分析发现目前大多研究聚焦于共识层的共识机制（Consensus）、网络层的 P2P 网络（Peer to Peer Network）和应用层的智能合约（Smart Contract）等方面。由图 2 中文献分布可知，相较于其他底层技术，当前对共识机制的研究较多。共识机制是用以决定如何使异步网络中的所有节点达成共识以增添新的节点，它是区块链技术的核心。当前的共识算法主要包括两类，一类是证明类共识，它需要网络中的节点证明自己比其他节点更有“资格”增加节点，常见的算法有工作证明（Proof of Work）、股权证明（Proof of Stake）以及其他基于 PoW 和 PoS 的改进或混合算法。另一类是选举类共识，即每一轮节点通过投票确定记账节点，该类主流

的算法包括有实用拜占庭容错算法（Byzantine Fault Tolerance）和 Raft 共识算法。其下方出现了文献聚集密度低于周围区域的特定区域，通过分析其词云图中的高频词，如 Peer to Peer Network、Client-Server，发现该区域是网络层的相关研究。图 2 中该主题下文献核密度较低，这表明当前对该主题的探索较浅。网络层主要负责内部节点间的通信，区块链采用 P2P 网络（Peer to Peer Network）作为网络层的通信协议，通过令所有节点共同负责网络中通信，达到了分布式、去中心化的目的。相邻的应用执行层的研究也相对较少，主要涉及各种脚本、交互接口、分布式程序及智能合约，是区块链可编程的基础研究。以太坊（Ethereum）是受关注和应用较多的开源底层平台，它支持由 Solidity 语言编写的智能合约的执行。

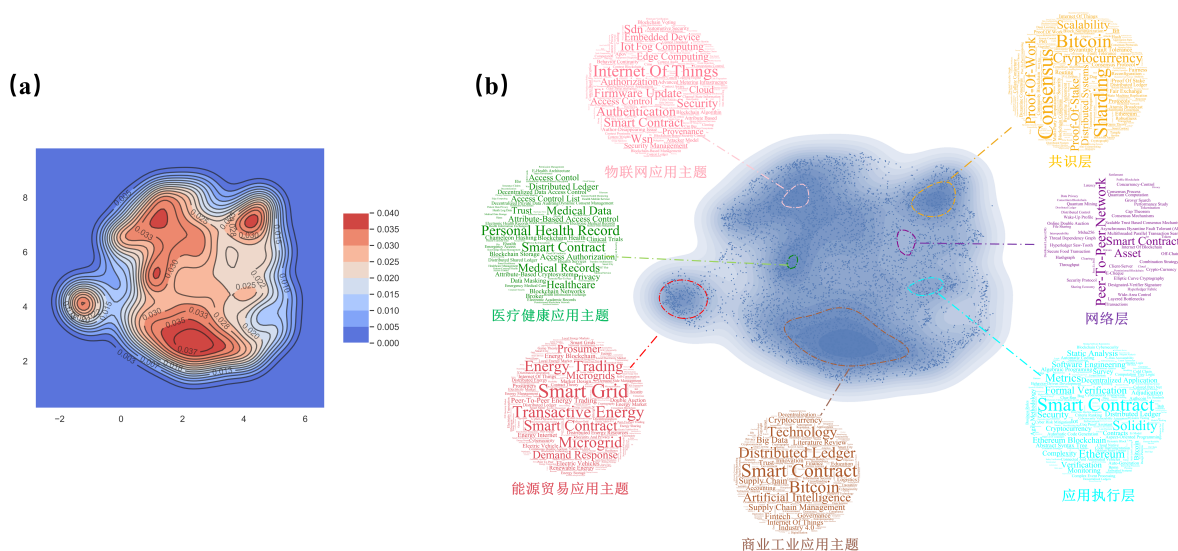


图 2 区块链领域地图

目前区块链技术主要应用于商业工业、能源贸易、物联网和医疗健康等方面。在所构建的领域语义地图中（图 2），商业工业主题的

文献分布最多且“区域面积”较大，这意味着它是区块链应用的热门领域，且其研究范围十分广泛。通过分析词云图发现，商业工业主题

主要包括区块链技术在金融科技 (Fintech)、供应链管理 (Supply Chain) 和治理 (Governance) 等方面的应用研究。相较而言, 能源贸易方面的研究比较集中, 并且和其他应用主题的语义相似性较弱。在能源交易 (Energy Trading) 方面, 区块链技术由其去中心化和不可篡改的特性, 改变了能源 (如电力) 的供给输配、交易方式和消费模式, 提高了能源电力系统的效率和安全性。如引入区块链技术的微电网 (Microgrids)、智能电网系统 (Smart Grid) 可实时供电、实现了自我控制。区块链与物联网的融合应用也是当前关注的焦点。如, 使用区块链技术构建的分布式物联网网络 and 平台具有较高的扩展性, 区块链边缘云 (Edge Computing) 为物联网设备提供了安全且高效的数据交换服务。在医疗健康 (Healthcare) 领域, 区块链技术和智能合约也得到了广泛应用。如, 将医疗数据 (Medical Data) 存储于区块链上便于互操与共享, 基于智能合约的加密访问机制加强了对健康档案 (Personal Health Record) 的隐私保护。在地图中, 物联网、医疗和商业工业应用主题之间的也分布了较高密度的文献, 这意味着当前研究中存在较多的区块链技术的跨行业场景应用研究, 尤其是跨物联网和医疗领域的研究。融合区块链、物联网的智慧医疗也是近来医疗体系发展的主要方向。

3 区块链研究演化分析

为了了解区块链的发展态势以及当前研究图景的形成过程, 本研究从关键词和领域语义地图两个角度分析了区块链研究在 2014-2020

年间演化过程。

3.1 历年高频关键词

考虑到数据集中 2014 年之前的文献较少, 2020 年只有部分数据, 所以在本小节以及后续的演化研究中, 将分别合并 2014 年前的文献, 2019 年和 2020 年的数据进行分析。图 3 为 2014-2020 年每年发表文献的关键词的词频词云图。根据图中高频词的变化可知, 区块链研究的发展大致可以分为三个阶段。2008-2015 为第一阶段, 即区块链 1.0 时代。该阶段的核心概念是 “Bitcoin” 和 “Cryptocurrency”, 研究主要以比特币等加密货币的交易为主。2016-2017 为区块链 2.0 时代。该阶段的核心是以 “Bitcoin” 为代表的 “Smart Contract”, 研究围绕着智能合约的开发和应用。其中最突出的技术开发是以太坊平台 (Ethereum), 技术应用场景主要集中于金融领域。2018-2020 是区块链 3.0 时代。区块链技术应用模式升级, 已经从金融领域扩展到物联网、工业、医疗等全社会领域。特别的, 区块链和 IoT 协同共生, 衍生出许多新的研究分支。综上所述, 区块链研究的发展迄今经历了从加密货币交易到智能合约开发到区块链应用多元化三个阶段。该高频词演变分析结果也印证了现今大众对区块链发展阶段的认知。

3.2 历年累积领域语义地图分析

按前文方法所述, 所构领域语义地图是区块链相关研究形成的知识空间在低维平面上的投影, 且历年累积地图的同一坐标区域具有相同的语义信息。上一章的整体累积领域语义地图分析以及词云图表征, 也进一步明确了地图

不同坐标区域所代表的语义信息或者研究内容。于是，本研究绘制了各年累计领域地图，以期直观地展现主题的发展过程。图 4 展现了区块链这

一新兴领域的生长发展过程，可以看出区块链相关研究的发展是一个领域范围由“研究核”不断向外扩张，同时领域内部不断分化的过程。

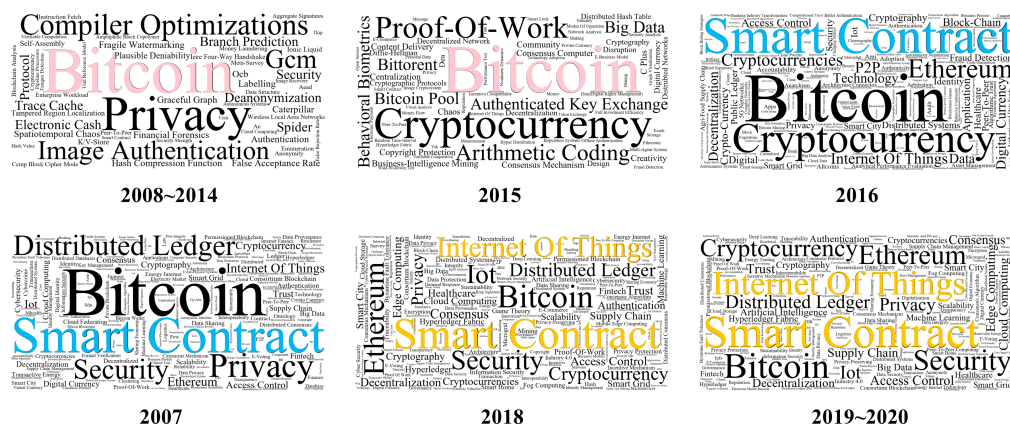


图 3 历年文献的关键词词云图

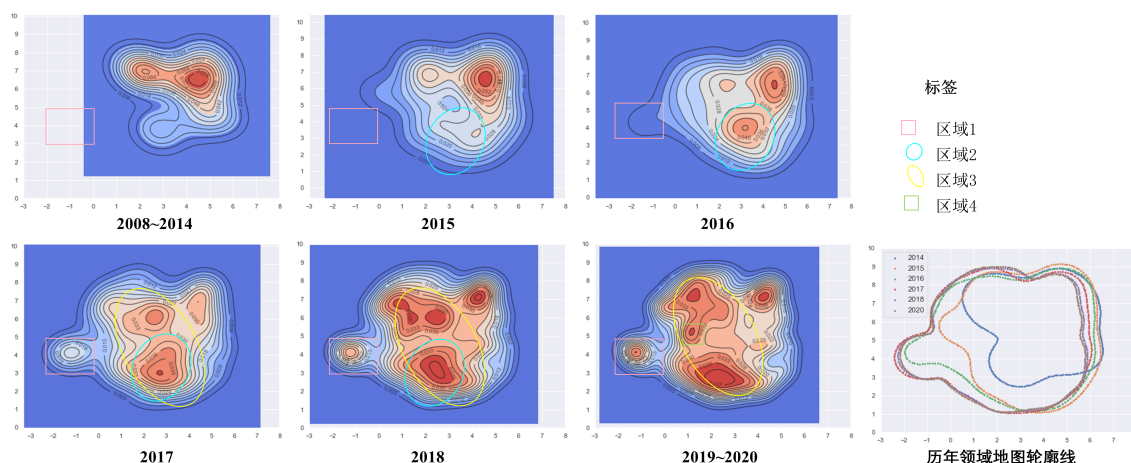


图 4 区块链领域的历年累积地图

首先，由图 3 可以看出，从 2014-2020 年，区块链研究的领域语义地图的整体轮廓线范围在不断扩大，即研究范围不断拓增。2014-2017 年，研究范围扩张较为明显，涌现出新主题。2014 年，即在区块链发展的初期，研究范围主要集中于基础技术研究（地图右上方）。到 2015、2016 年，区块链研究从基础技术研究慢慢向商业工业的应用研究扩展（地

图右下方）。尤其至 2017 年，领域扩张出新的研究区域（区域 1），即能源贸易主题涌现，并且该主题自出现以来持续发展。结合前述关键词分析可知，智能合约的开发和应用推动了该阶段区块链技术应用的拓展。2017 年后，领域轮廓线相对稳定，没有拓展出新方向。然而，随着研究重点的迁移，领域范围仅在小幅度调整。



图5 商业工业主题的演变

其次,伴随着区块链“研究面”的扩张,其领域语义地图内部的地貌也在不断调整变化。从图4可以看出在区块链技术的发展过程中存在较多的分化涌现现象,而主题间的融合较少。如2015-2017年商业工业应用主题的涌现(区域2)。2017年,区块链在物联网和商业工业领域的应用研究还为完全分化为两个研究方向(区域3),被包围在同一条密度较高的等高线闭环内。随着研究的深入,到2018年这两个领域的研究分裂成两个主题。2018-2019年,分裂出医疗健康主题(区域4)。除此之外,各主题的“区域形态”,“区域重心”也在不断变化。结合论文关键词可以探知各个主题研究内容的变迁,如图5显示商业工业主题(区域2)在发展过程中逐渐应用了人工智能算法。

图4所揭示的区块链研究领域扩张的模式对进一步发展区块链学科领域有所借鉴:后续发展一方面将致力于区块链关键技术的进一步突破(表现为领域地图的右侧部分的提升),另一方面致力于新的应用领域的拓展(表现为整个领域地图的扩张和新热点的涌现)。这一领域地图的快速扩张特征也反映了区块链领域近年来的快速发展。

4 结论与展望

随着区块链技术日益受到世界各国的重视

和关注,了解区块链技术的研究现状、认识其发展过程对推动技术发展、制定科技政策具有重要意义。针对当前数据范围以及研究手段方面的局限,本研究基于WOS数据库中收集到的5,951篇文献,借助新近的文档嵌入表示学习和流形学习算法识别了区块链的主要研究主题,可视化了研究领域图景,展现并揭示了其演化过程。主要发现:

(1) 近来区块链研究呈现快速增长态势,研究主题主要分为基础技术研究和应用型研究。其中,基础技术研究集中于共识层、网络层以及应用执行层;区块链的应用研究已渗透到商业工业、医疗健康、物联网以及能源交易等领域。

(2) 区块链领域地图清晰地展现了研究“地貌”。领域内各个主题的研究相对独立,跨主题研究较少,只有物联网和医疗领域之间的跨行业应用研究相对较多。各个主题被探索的程度不同,当前研究多侧重于技术应用方面。

(3) 区块链的研究呈现出应用导向的发展趋势。其发展经历了三个阶段,即以比特币交易为主的区块链1.0时代,以智能合约为代表的区块链2.0时代,技术应用多元化的区块链3.0时代。

(4) 区块链领域的发展过程呈现领域扩张的特征,且该过程伴随着领域内部主题的分化。发展初期,区块链领域的研究范围集中于基础技术研究,后来不断向外拓展涌现新主题,然

后领域范围相对稳定。于此同时,随着研究的深入,各主题内部和主题之间的“地貌”不断变化,比如分裂出医疗主题。

本研究获悉了区块链领域的整体研究概貌、主要研究主题以及发展趋势,有望对我国区块链的发展提供参考,也有助于进一步理解类似区块链领域等新兴领域的发展规律。除此之外,本文在方法上拓展了当前的领域主题及演化分析的研究手段。然而本研究未能对区块链研究演化形态进行更为精细的量化分析,后续将尝试借助地理时空分析指标,以期更加客观准确地识别主题演化类型。且本研究仅利用了学术论文数据,未来可结合专利数据,进一步全面地分析区块链技术在产业界和学术界两方面的研究现状以及发展过程。

参 考 文 献

- [1] 刘硕,胡泽文,李玉平. 区块链模型、技术与典型应用探究[J]. 情报工程, 2019, 5(6):15-36.
- [2] 朱岩,王巧石,秦博涵,等. 区块链技术及其研究进展[J]. 工程科学学报, 2019, 41(11):1361-1373.
- [3] 隆苏妍,李辉,靳晓宏,等. 区块链时代与机遇伴生的安全威胁[C]. 2018年北京科学技术情报学会学术年会—智慧科技发展情报服务先行论坛, 2018, 中国北京.
- [4] 袁勇,王飞跃. 区块链技术发展现状与展望[J]. 自动化学报, 2016, 42(4):481-494.
- [5] Casino F, Dasaklis T K, Patsakis C. A systematic literature review of blockchain-based applications: Current status, classification and open issues[J]. Telematics and informatics, 2019(36):55-81.
- [6] 刘敖迪,杜学绘,王娜,等. 区块链技术及其在信息安全领域的研究进展[J]. 软件学报, 2018, 29(7):2092-2115.
- [7] 谢敬东,陆池鑫,孙欣,等. 区块链技术在能源与电力系统领域的应用和展望[J]. 电测与仪表, 2020, 42(8):1-12.
- [8] 杨现民,李新,吴焕庆,等. 区块链技术在教育领域的应用模式与现实挑战[J]. 现代远程教育研究, 2017(2):34-45.
- [9] 徐嘉辉,马立新. 区块链技术在分布式能源交易中的应用[J]. 电力自动化设备, 2020(8):1-7.
- [10] Firdaus A, Razak M F A, Feizollah A, et al. The rise of “blockchain”: Bibliometric analysis of blockchain study[J]. Scientometrics, 2019, 120(3):1289-1331.
- [11] Ante L. A place next to Satoshi: Foundations of blockchain and cryptocurrency research in business and economics[J]. Scientometrics, 2020, 124(2):1305-1333.
- [12] 雷孝平,张海超,桂婕,等. 基于论文和专利的区块链技术研发状况分析[J]. 情报工程, 2017, 3(2):20-32.
- [13] 王发明,朱美娟. 国内区块链研究热点的文献计量分析[J]. 情报杂志, 2017, 36(12):69-74.
- [14] Zeng S, Ni X, Yuan Y, et al. A bibliometric analysis of blockchain research, 2018 [C]. 2018 IEEE Intelligent Vehicles Symposium (IV), June 26-30, Changshu, Suzhou, China, 2018: 102-107.
- [15] Dabbagh M, Sookhak M, Safa N S. The evolution of blockchain: A bibliometric study[J]. IEEE Access, 2019(7):19212-19221.
- [16] Miao S, Yang J. Bibliometrics-based evaluation of the Blockchain research trend [J]. Technology analysis & strategic management, 2018, 30(9):1029-1045.
- [17] 乔鹏程. 基于知识图谱的区块链与云会计比较及融合研究[J]. 科技管理研究, 2019, 39(5):166-173.
- [18] 顾佳依. 基于文献计量的区块链热点及趋势分析[J]. 技术与创新管理, 2020, 41(1):46-50.
- [19] 程晨,张毅,宁晓静,等. 国外区块链研究主题及展望[J]. 电子政务, 2018(6):11-21.
- [20] 程豪,张峥. 基于 CiteSpace 分析的区块链技术可视化研究[J]. 物流科技, 2019, 42(2):7-11, 22.
- [21] 汪园,王学东,李金鑫. 基于文献计量的我国区

- 区块链研究的知识网络与结构分析 [J]. 现代情报, 2018, 38(1):147-153.
- [22] Mikolov T. Distributed Representations of Words and Phrases and their Compositionality[J]. Advances in Neural Information Processing Systems, 2013(26):3111-3119.
- [23] Garg N, Schiebinger L, Jurafsky D, et al. Word embeddings quantify 100 years of gender and ethnic stereotypes [J]. Proceedings of the National Academy of Sciences, 2018, 115(16):3635-3644.
- [24] Kozłowski A C, Taddy M, Evans J A. The Geometry of Culture: Analyzing the Meanings of Class through Word Embeddings [J]. American Sociological Review, 2019, 84(5):905-949.
- [25] McInnes L, Healy J, Melville J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction [J]. Journal of Open Source Software, 2018, 3(29):861.
- [26] Becht E, McInnes L, Healy J, et al. Dimensionality reduction for visualizing single-cell data using UMAP [J]. Nature Biotechnology, 2019, 37(1):38.
- [27] Coenen A, Pearce A. Understanding UMAP [EB/OL]. [2020-10-23] <https://pair-code.github.io/understanding-umap/>