·研究综述与前沿进展·

国际区块链研究主题挖掘及演化分析

罗 棋 闵 超 颜嘉麒 王嘉杰
（南京大学信息管理学院，江苏 南京 210023）

摘 要：［目的/意义］区块链研究越来越受到社会各界的重视，对国际区块链研究主题及其演化的分析有利于我国学者把握其发展趋势，精准把握面向区块链学科领域的国际前沿研究方向。［方法/过程］本文围绕国际区块链研究主题及演化情况，以 Web of Science 核心合集 SCI-EXPANDED 和 SSCI 中 2008—2020 年相关论文为数据源，使用 LDA 主题模型，从热点主题和主题演化两个方面对国际区块链研究进行了分析。［结果/结论］国际区块链领域中的热点主题包括“医疗健康”“数据隐私保护”“能源交易与共识算法”和“物联网安全”；“商务智能合约”“加密货币”两个主题具有明显演化特征，具体表现为在重要文献发表之后两者的热度开始上升；区块链的研究从其技术本身向包括医疗、能源在内的多个领域拓展渗透；目前，以融合解决其他领域问题为特征的区块链 3.0 的研究占比已经超过了传统的区块链 1.0 和以智能合约为代表的区块链 2.0。

关键词：区块链领域；主题挖掘；主题演化；LDA
DOI:10.3969/j.issn.1008-0821.2021.09.016
〔中图分类号〕TP311 〔文献标识码〕A 〔文章编号〕1008-0821（2021）09-0157-10

Topic Mining and Evolution Analysis of Blockchain Research

Luo Qi Min Chao Yan Jiaqi Wang Jiajie
（School of Information Management，Nanjing University，Nanjing 210023，China）

Abstract: ［Purpose/Significance］The research of blockchain has been paid more and more attention. The analysis of blockchain research topics and their evolution is helpful for Chinese scholars to grasp its development trend and do research facing the international frontier. ［Method/Process］Focusing on the topic and evolution of blockchain research，this paper analyzed from two aspects：hot topics and topic evolution by using the LDA topic model，with the web of science core collection SCI-EXPANDED and SSCI papers from 2008 to 2020 as data sources. ［Result/Conclusion］The hot topics in the field of blockchain included “health care”，“data privacy protection”，“energy transaction and consensus algorithm” and “Internet of things security”. The two topics of “business smart contract” and “cryptocurrency” have obvious evolutionary characteristics，which were shown in the rising popularity of the two topics after the publication of influential works. The research of blockchain extended and penetrated from its technology itself to many fields including medical treatment and energy. At present，the proportion of research on blockchain 3.0，which is characterized by integrating and solving problems in other fields，has exceeded that of traditional blockchain 1.0 and blockchain 2.0 represented by smart contracts.

Key words：blockchain；topic mining；topic evolution；LDA

自 2008 年中本聪（Satoshi Nakamoto）发表《Bitcoin：A Peer-to-Peer Electronic Cash System》[1] 以来，各界研究人员对其底层核心技术——“区块链”的探索和研究热情水涨船高，由于其“可追溯”“防

收稿日期：2021-03-01
基金项目：国家自然科学基金项目“供应链质量管理中基于区块链的智能系统模型研究”（项目编号：71701091）；教育部人文社会科学项目“区块链虚拟组织信息资源的知识表示方法研究”（项目编号：17YJC870020）。
作者简介：罗棋（1997-），男，硕士研究生，研究方向：区块链与科学计量。闵超（1990-），男，助理研究员，博士，研究方向：科学计量、创新扩散。颜嘉麒（1983-），男，副教授，博士，硕士生导师，研究方向：区块链、信息系统、数据分析。王嘉杰（2000-），男，本科生，研究方向：创新网络、组织学习。

篡改”等特性，它在其他领域的重要程度日益凸显。但在这种备受各界学者关注的情况下，王江等[2]指出，虽然我国在区块链研究方面生产力占据世界第一，然而最具有影响力的区块链研究的来源期刊、论文、作者等都来自于国外，因此，把握国际学者对于该领域的研究现状及热点主题的演化，有利于我国的学者发现研究新趋势，学习和借鉴有益成果，为我国的“区块链”研究提供参考。

1 区块链主题的相关科学计量研究及其不足

近些年来，学界有许多关于区块链研究现状、研究热点以及主题演化等方面的研究。Firdaus A 等[3]以 Scopus 数据库收录的 2013—2018 年的区块链相关文章为研究对象，运用文献计量的方法进行分析，发现最活跃的国家是美国，其次是中国和德国。Dabbagh M 等[4]分析了 Web of Science 数据库中 2013—2018 年的相关论文，指出了其主要学科分布包括：计算机科学、工程学、电信学、商学、经济学等。王发明等选取“CNKI 期刊库”2015—2017 年 5 月的论文，使用 Cite Space 可视化工具，从关键词、作者共现等角度，分析了我国区块链的研究热点，该研究认为我国区块链领域尚处于探索期，并且将热点主题概括为基础研究和应用研究两个方面[5]。汪园等也运用 Cite Space 可视化工具，对 2015—2017 年的相关文献进行了分析，从文献类型（科普评论类、探索研究类）、期刊分布、学科分布等方面对区块链相关研究进行描述总结[6]。花敏等通过对 2015—2019 年 CNKI 数据库和 WOS 数据库相关文献的对比分析，从发文量、高产机构等多个角度展开，该研究认为中国和美国是两个开展区块领域研究的主力国家，2015—2019 年，我国在区块链领域发表的外文文章的数量始终高居榜首并迅猛增长[7]。但是正如王江等[2]的发现，最具有影响力的区块链研究的来源期刊、论文、作者等都来自于国外，所以本文以国际区块链研究为研究对象，分析其热点主题演化情况，以期为我国学者提供借鉴参考。

当前研究大多以科学数据库中的文献及引文数据为研究对象，特别是关键词，使用文献计量的方法及工具，特别是关键词共现分析，从作者、期刊及机构等角度分析区块链研究热点。但是题录数据中，摘要包含的信息没有得到有效的利用，仅仅靠关键词只能反映文章的大致方向，难以挖掘其隐含的语义信息。

LDA（Latent Dirichlet Allocation）主题模型能够很好地解决这一问题，通过抽取摘要中隐含的主题信息，为后续研究提供研究主题分布上的参考。Chen H 等[8]运用 LDA 模型对截至 2015 年发表在 MIS Quarterly 等 3 本信息系统领域顶级刊物上的文章进行了主题建模，深入分析了信息系统领域的研究问题，以及各研究问题间的关联。赵紫鹃等运用 LDA 模型对“第十三届全国复杂网络大会”的会议摘要文本进行了文本挖掘，得到了 10 类研究主题[9]。李跃艳等选取 SIGIR 会议论文为研究对象，使用 LDA 模型，分析了近 10 年信息检索领域的研究热点与演化趋势[10]。可见，使用 LDA 主题模型探究某具体领域的热点主题可以从更细的粒度分析推断文章内容，挖掘隐含的语义信息，得到更加细致的结论，因此，本研究采用 LDA 主题模型来挖掘国际区块链领域研究的热点主题，并分析其随时间演化情况，以期为我国学者把握研究前沿和热点提供参考。

2 模型与方法

本研究以 Web of Science 核心合集 SCI-EXPANDED 和 SSCI 中区块链相关的文献数据作为数据来源，根据研究目的对其进行清洗，保留对分析有用的字段，使用 LDA 主题模型对文献的研究内容（标题、摘要、关键词）进行主题挖掘，计算困惑度以确定最优主题数，根据高概率的词对主题进行标注；并计算主题强度，划分出热点主题，并按时间窗口进行离散化处理，分析热点主题随时间的演化情况。本研究整体框架如图 1 所示。

2.1 LDA 主题模型

挖掘科研文献主题的方法有很多，传统的词频分析或者共词分析的方法也可达到揭示科研文献数据集的研究主题的目的，但是关键词之间可能存在“共生现象”，可能有多个高频的关键词同属于一个主题，导致词频较低的关键词所属的主题难以发掘[11]。并且传统的方法以关键词为研究对象，本身损失了很多语义信息（例如摘要中包含的信息），只能大致反映文章的方向，难以挖掘其隐含的语义

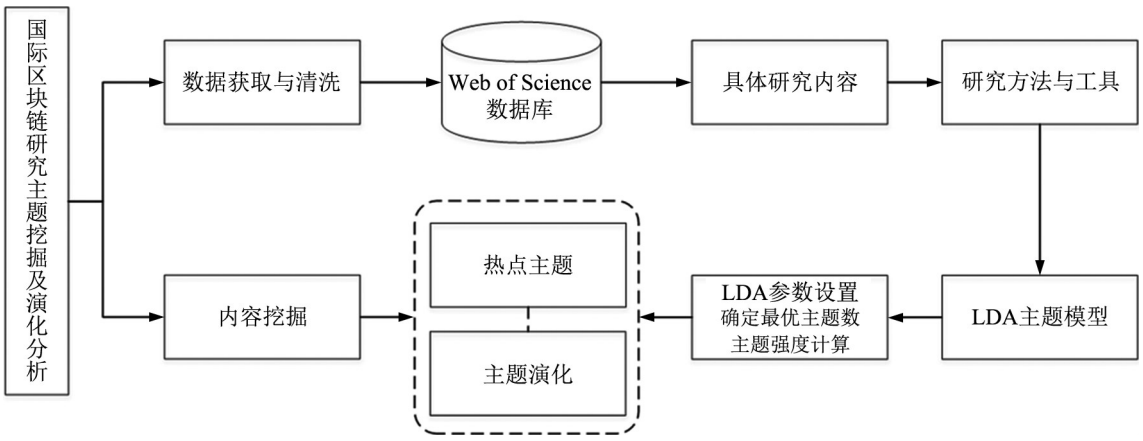

图 1 研究框架

信息，分析文本的规模也有限[12]。而主题模型的出现，较好地解决了这一问题，不仅能够处理大规模的文本数据，还能挖掘出语料中潜在的语义信息，因此，本文采用 LDA 主题模型来挖掘国际区块链研究的热点主题。

（Latent Dirichlet Allocation，LDA）潜在狄利克雷分配模型，是一种常见的主题模型，2003 年由 Blei D M 等共同提出[13]。可以认为 LDA 是 PLSA（Probabilistic Latent Semantic Analysis，概率潜在语义分析）的拓展，LDA 使用了先验分布，克服了学习过程中的过拟合问题。该模型假设：①主题由词的多项分布表示；②文档由主题的多项分布表示；③主题—词分布和文档—主题分布，两者的先验分布都是狄利克雷分布[14]。借由狄利克雷分布是多项分布的共轭先验分布这一特性，可以通过观测的单词序列，推断出文档—主题分布和主题—词分布，挖掘出隐含的主题层，其生成过程如图 2 所示。

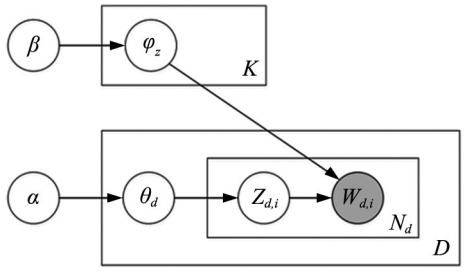

图 2 LDA 的板块表示

LDA 模型将代表文本的词频向量（文档—词频矩阵）作为输入，通过迭代输出推断出的文档—主题分布、主题—词分布，即每个文档由各个主题生成的概率、每个主题包含各个词的概率。图 2 中的节点表示随机变量：实心节点表示观测变量，空心节点表示隐变量；有向边表示概率依存的关系；矩形板块表示重复，板块内数字表示重复次数[14]。图 2 中使用的符号及其含义如表 1 所示。

表 1 LDA 模型中符号含义

符号	含 义
α	整个文档上主题的 Dirichlet 先验分布
β	所有主题上词的 Dirichlet 先验分布
θ_d	文档 d 上主题的多项式分布
φ_z	主题 z 上词的多项式分布
D	文档个数
K	主题个数
$Z_{d,i}$	文档 d 上第 i 个词的主题
N_d	文档 d 中词的个数
$W_{d,i}$	文档 d 上第 i 个词

LDA 主题模型的参数估计过程其实就是根据观测变量的取值估计隐变量的值，其参数估计的方法主要有 3 种，分别是：吉布斯采样算法（Gibbs Sampling）、变分推断算法（Variational Bayesian Inference）和最大期望算法（Expectation Maximization），张健伟[15]通过实验发现，期望最大算法在某些关键的预测能力指标上（例如：预测混淆度）优于其他两种算法，并且可以在较短的时间内收敛，因此本研究采用期望最大算法来进行 LDA 主题模型的参数估计。

2.2 热点主题挖掘及演化分析

热点主题的挖掘，即判断某主题是否为热点主

题有一个主要依据的指标——主题强度。该指标专用于描述一个主题的热门程度，另一关键指标是主题强度阈值，如果某主题强度高于阈值则认为该主题为热门主题，反之则非热门主题。关于主题强度的计算，孙孟孟在其学位论文中进行了详细的讨论[16]。主要有以下 3 种方法：①基于主题支持文档数量；②基于语料库中主题概率；③基于文本主题显著性。3 种计算方法各有特点，比较常用的是第 2 种基于语料库中主题概率的方法，孙孟孟只给出了伪码，吴查科等将其提炼[12]，具体公式为：

$$\theta_z^t=\frac{\sum_{d=1}^{D^t}\theta_z^d}{D^t} \tag{1}$$

式（1）中，符号 D^t 的含义是时间窗口 t 下的文档数目；θ_z^d 的含义是文档 d 由主题 z 生成的概率；θ_z^t 的含义是时间窗口 t 下主题 z 的强度。

关于主题强度阈值的确定，张斌采用直接指派的方法[17]，该方法在该文中较为适用，但本文从数据驱动出发，采取吴查科等的主题强度阈值的计算方法[12]，结合本文的推导，发现主题强度阈值并不随时间窗口的变化而变化，具体的计算公式为：

$$T=\frac{\sum_{d=1}^{D^t}\sum_{z=1}^{K}\theta_z^d}{D^t\cdot K}=\frac{1}{K}=\frac{\sum_{d=1}^{D}\sum_{z=1}^{K}\theta_z^d}{D\cdot K} \tag{2}$$

式（2）中，符号 K 的含义是主题的个数，D 的含义是数据集中文档的个数；T 的含义是主题强度阈值；其他符号与式（1）重复，不再赘述。

基于 LDA 的主题演化分析，主要有 3 种不同的演化方法：①将时间信息结合到 LDA 模型中；②后离散分析；③先离散方法[18]。本文采用后离散方法，先忽略时间变量，用 LDA 对整个数据集进行建模，然后按不同的时间窗口测算主题强度，比对主题强度阈值进行分析。

3 实证分析

3.1 数据采集与整理

本研究为了减少数据搜集的偏差，使用 Tranfield D 等[19]和 Webster J 等[20]提出的两阶段综述方法来收集国际区块链相关的研究文献，具体的文献收集筛选流程如图 3 所示。

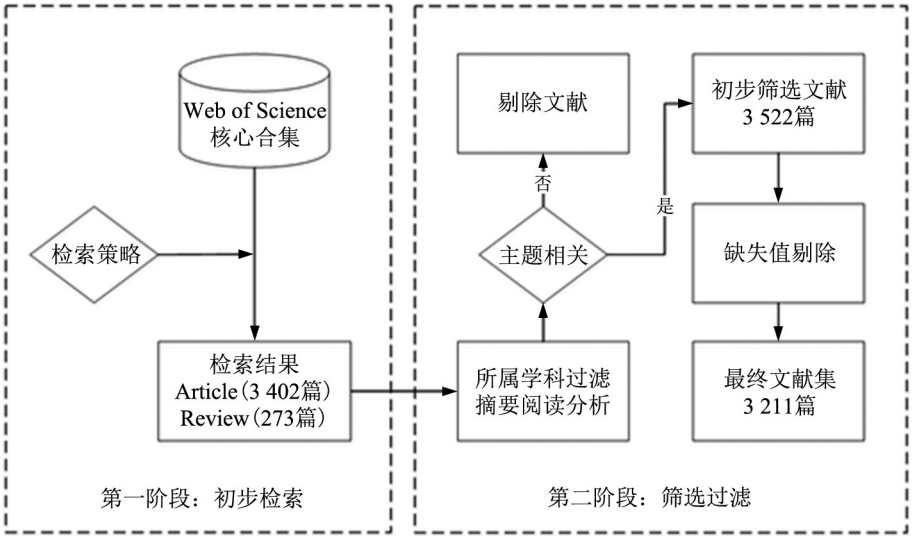

图 3 文献收集筛选过程图

第一阶段，为了得到较为可靠的数据，本研究选择了 Web of Science 核心合集中的 SCI-EXPANDED 和 SSCI 作为数据来源，选择其中的“主题”字段，检索式为 TS = “Blockchain$” or TS = “Block Chain$”，时间段是 2008—2020 年，文献类型为 Article 和 Review。检索策略制定的原因如下：①时间：之所以选择 2008 年作为起始时间，是因为区块链的概念于 2008 年中本聪（Satoshi Nakamoto）的《Bitcoin：A Peer-to-Peer Electronic Cash System》（常译作“比特币白皮书”）一文中首次提出[20]，之前并未有这个词汇；②检索式：中本聪提出“区块链”时称之为“Chain of Blocks”，经中文翻译为“区块链”，后学者多用“Blockchain”作为其英文称谓，但是也有部分学者使用“Block

Chain”，结合各自的单复数形式，所以采用此检索式，共收集到3 675篇文献（检索时间为 2021 年 1 月 3 日）。

第二阶段，通过所属学科的过滤，和对文献的摘要阅读，分析其是否与“区块链”主题相关，比如：学科类别为“PHYSICS PARTICLES FIELDS”（物理粒子场）的文章《A Multipoint Conformal Block Chain in d Dimensions》经过对其摘要的阅读，与“区块链”并无关联，便将此篇文献剔除。用相同方法过滤了与“区块链”主题无关的文献，经过初步筛选得到3 522篇文献；接着将年份和国别等关键字段缺失的文献剔除，得到最终文献集3 211篇，下面将使用此文献数据集进行进一步的主题挖掘与分析。

3.2 研究主题分析

欲深入探究国际区块链领域的研究内容，挖掘其潜在的语义信息，需要借助 LDA 模型从摘要数据集中抽取主题，发现热门主题，参考马永红等[21]的研究框架，本节研究具体的分析处理框架如图 4 所示。

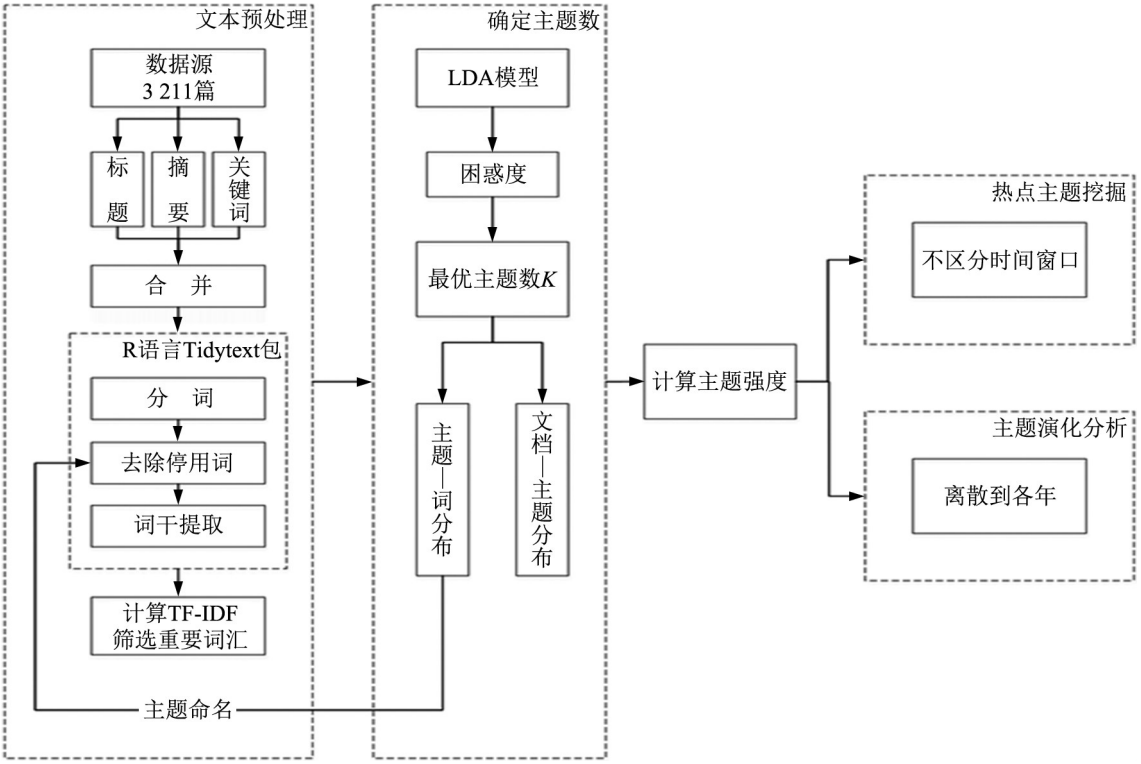

图 4 研究主题分析框架图

3.2.1 文本预处理

对科学文献数据集进行 LDA 主题建模的预处理一般步骤包括：①提取文献的“摘要”字段；②分词；③去除停用词；④构建“文档—词”矩阵。本研究基于以上步骤，且为了提升 LDA 主题模型的聚类效果，进行了以下 4 步处理：

1）将文献的“标题”“摘要”“关键词”合并作为待分析文本，由于 3.1 数据采集与整理得到的3 211篇文献中有 56 篇文献缺失了“摘要”数据，为了在更大程度上保留原有的信息，本研究不剔除缺失的记录，而是参考 Chen H 等[8]的做法，将“标题”“摘要”“关键词”合并后作为一个整体而后进行分词，分词后获得的词的集合用于下一步处理。

2）在“去除停用词”步骤时，除了使用 R 语言中 Tidytext 包默认的停用词，也根据主题建模的结果反馈，将“主题—词分布”中无意义的高频词汇（例如：“Paper”）加入“自定义停用词表”，排除其对结果的干扰。

3）加入了“词干提取”步骤：由于许多词汇含义相同，却拥有不同的形式（单复数、词性等），造成词频过于分散，影响聚类效果，所以提取真正代表其含义的词干，降低稀疏性。

4）在构建“文档—词”矩阵之前，先使用

TF-IDF（Term Frequency-Inverse Document Frequency）得分对词汇进行筛选，将不重要的词汇剔除，减少词项（特征数目），降低“文档—词”矩阵的维度，提高聚类效果。实际操作时，通过多次实验，发现每个文档取 TF-IDF 得分前 20 的词汇，维度损失较少，聚类结果较好。

3.2.2 确定主题数

主题模型中主题数目是一个关键参数，但是关于如何确定主题数目，学者们众说纷纭，主要有两大类方式：①Blei D M 等[13]提出的困惑度（Perplexity）的方法，该指标反映了模型的拟合程度，困惑度越小，模型的拟合程度越好，可以通过多次实验找到其极小值的方法来确定主题数目；②计算“主题相似度”的方式，常见的有计算 Jensen-Shannon 散度（JS 散度）的方法，关鹏等[22]对其做了详细的研究，当主题数接近最优值时，JS 散度较小，反之则较大。综合前人的研究，本研究采用学者使用较多的困惑度的方式来确定最优主题数。

使用 R 语言中的 Topicmodels 包进行主题建模，主题数目的变化区间为 [2,30]，计算的结果如图 5 所示。从图像可以看出，当主题数目小于 8 时，随着主题数目的增加，困惑度不断减小，拟合效果越来越好；当大于 8 时，困惑度逐渐稳定在高点；所以，本研究确定的最优主题数为 8。

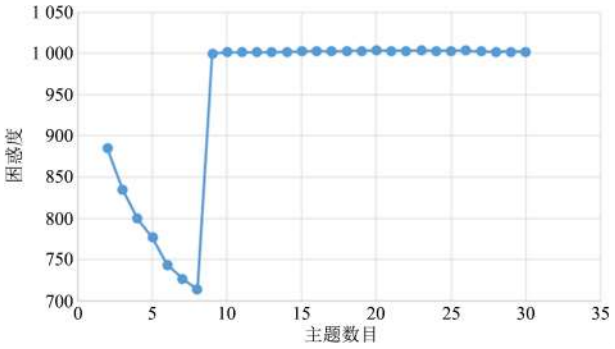

图 5 不同主题数目下困惑度图像

确定最优主题数为 8 之后，代入 LDA 模型，使用期望最大算法估计参数取值，最终得到“文档—主题分布”和“主题—词分布”。各个主题中概率较高的特征词如表 2 所示。

表 2 各主题高概率特征词汇表

Topic1	Topic2	Topic3	Topic4	Topic5	Topic6	Topic7	Topic8
Blockchain	Blockchain	Smart	Blockchain	Data	Energi	Iot	Chain
Vehicl	Technologi	Contract	Bitcoin	Blockchain	Blockchain	Blockchain	Blockchain
Secur	System	Blockchain	Cryptocurr	Secur	Trade	Secur	Suppli
Network	Research	Technologi	Digit	Base	Consensu	Network	Technologi
System	Health	Busi	Transact	Privaci	Transact	Comput	Industri

根据上表中展示的高概率特征词，对每个主题进行“命名”（标注），然后结合“文档—主题分布”对命名结果进行验证。例如：Topic1 中概率较高的词是“Blockchain” “Vehicl” “Secur” “Network”，根据词干的提示，可以将其命名为“区块链”+“车联网安全”，然后将文档按照由 Topic1 生成的概率进行排序，概率较高的 3 篇代表性文献分别是《A Blockchain Based Certificate Revocation Scheme for Vehicular Communication Systems》[23]《Physical Layer Security of Autonomous Driving: Secure Vehicle-to-Vehicle Communication in A Security Cluster》[24]《Blockchain-Based Dynamic Key Management for Heterogeneous Intelligent Transportation Systems》[24]，经过对其阅读研判，确为研究“区块链”在“车联网安全”方面应用的文章，印证了标注的准确性。按照此模式分别对 8 个主题进行标注，结果如表 3 所示。因为本研究对象为国际区块链研究，为了简便起见，后文在提到主题标注时，将省去“区块链+”。

表 3 主题标注表

主 题	主题标注
Topic1	“区块链”+“车联网安全”
Topic2	“区块链”+“医疗健康”
Topic3	“区块链”+“商业智能合约”
Topic4	“区块链”+“加密货币”

表 3（续）

主 题	主题标注
Topic5	“区块链”+“数据隐私保护”
Topic6	“区块链”+“能源交易与共识算法”
Topic7	“区块链”+“物联网安全”
Topic8	“区块链”+“工业供应链”

3.2.3 热点主题挖掘

仅仅对主题进行标注是不够的，还需要根据主题强度对热点主题进行挖掘，为我国学者研究选题提供参考。根据式(2)得出主题强度阈值为 0.125，根据公式 1 计算出各个主题的主题强度，具体的结果如图 6 所示。

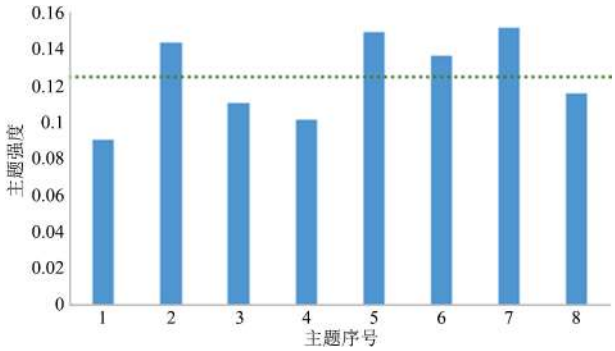

图 6 国际区块链领域各主题强度图

从图中可以看出，Topic2、Topic5、Topic6、Topic7 的主题强度值高于主题强度阈值，属于“热门主题”，下面结合有代表性的论文对热点主题逐个进行解析。

1）Topic2（医疗健康领域）：

医疗健康领域随着老龄化的发展，越来越受到各国的关注，但是医疗健康是一个复杂的系统，至少需要三方的参与：医疗服务的核心提供方（医生、护士等）、关联服务提供方（医学研究、保险等）、医疗服务的用户（病人、公众等）。这样一个多方参与的系统，其中数据的管理共享、隐私保护的问题亟待解决，催生了大量相关研究：Dhagarra D 等[26]试图通过区块链技术构建一个综合的医疗保险框架来整合碎片化的健康记录，改善医疗服务的均衡性；Zhang P 等[27]构建了一个基于区块链的去中心化应用程序来进行安全和可扩展的数据共享，协助临床诊断。

2）Topic5（数据隐私保护）：

随着云存储等技术的不断发展和云服务提供商的涌现，极大地降低了用户存储数据的成本，但是云服务提供商能否对数据的安全和隐私保护负责，始终是一个困扰用户的难题。“棱镜门”事件、“夜莺计划”等隐私泄露事件层出不穷[28]。这一关键问题吸引了大量学者研究：Huang P 等[29]提出了一种协作审核的区块链框架，引入了共识节点代替单个的第三方，试图解决数据所有者和云服务提供商之间的信任问题；Yang X 等[30]则利用区块的不可预测性构造挑战信息，来防止恶意的审核第三方和云服务器串通。

3）Topic6（能源交易与共识算法）：

能源问题特别是电能的分布式整合问题长久以来困扰着工业界和学界，随着区块链技术特别是其实用共识算法的出现，使得分布式的整合和配电成为可能，越来越多的框架被提出并进行了小范围的试点：Hayes B P 等[31]提出了一种配电网络和本地对等能源交易平台结合的仿真方法，采用基于区块链的双拍卖机制，使用欧洲郊区的配电网案例演示了该方法；Cai W 等[32]将传统的拜占庭容错算法改进，大大提高了交易速度，使其适用于能源领域实时处理交易的需求。

4）Topic7（物联网安全）：

包括射频识别技术（RFID）、传感器技术在内的物联网技术的飞速发展，也产生了许多网络常见问题，易受攻击、劫持，安全性和网络性能都面临考验，学者们运用区块链技术提出了多种方法来提高物联网的安全性并保障其网络性能：Rathore S 等[33]利用区块链提供分散式的攻击检测，来缓解现有架构中的“单点故障”问题；Sahay R 等[34]运用区块链上的智能合约来生成实时警报，能够有效地识别被篡改的节点。

花敏等[7]的研究表明，国外区块链领域的三大研究热点为“智能合约”“物联网”“隐私问题”，也印证了本文的研究发现，但是囿于其采用的关键词聚类方法，对语义信息损失较多，无法对热点领域进行更深入的分析，本研究由于采用 LDA 主题模型，可以挖掘篇名、摘要和关键词中的语义信息，能够从更细的粒度上挖掘发现热点主题。

3.2.4 主题演化分析

根据后离散方法，离散到各个年份后，计算了各个主题对应的主题强度，结果如图 7 所示，由于数据源中 2010 年和 2011 年没有“区块链”相关的文章，所以图示中跳过了该年份。图中横坐标表示年份，纵坐标表示主题强度值，柱体的高度反映主题强度的大小。

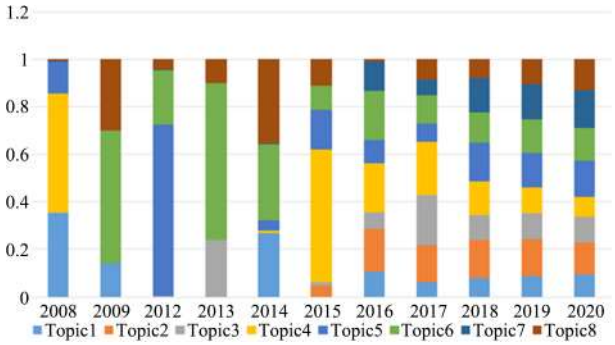

图 7 国际区块链研究主题演化图

通过图 7 可以发现，随着时间的推移，堆积图从原有的单调的几个颜色，开始变得色彩丰富，即区块链研究从原有的仅涉及其技术本身的研究，如：Topic4（加密货币）、Topic5（数据隐私保护），开始向其他领域如 Topic2（医疗健康）等进行渗透和拓展。这也向我国研究学者提出了更高的要求：除了在区块链底层技术（如：共识算法等方面）发力，也要重点关注区块链技术在其他领域（如：医疗健康等）的融合拓展研究。除了整体的趋势变化，其中 Topic3、Topic4 随时间演化特点较为明显，下面详细进行分析

1）Topic3（商务智能合约）：

2013 年，以太坊白皮书的问世，使人们看到了区块链的应用潜力，不只是可以分布式记账，还可以部署合约，把区块链带入了 2.0 时代，“智能合约”开始受到广泛关注，从图中也可以看出从 2013 年开始，代表“智能合约”的 Topic3（灰色）相关文章开始出现。Chang S E 等[35] 从信用支付的角度，研究了区块链技术在国际贸易过程中的适用性；Eenmaa-Dimitrieva H 等[36] 从合同法学者的角度出发，倡议使用智能合约来提供比传统交易更便宜快捷的交易服务。

2）Topic4（加密货币）：

Topic4 比较有代表性，“加密货币”是区块链的传统主题，区块链正是由中本聪于 2008 年在比特币白皮书中首次提出，所以在 2008 年的时候主题强度比较高，但是随着区块链在其他领域应用研究的蓬勃发展，渐渐势微，但是 2015 年以来，随着以太坊（ETH）、门罗币（XMR）、达世币（DASH）等多种加密货币的涌现，使得公众对加密货币的关注度空前提高，学界也从其安全性、经济性等多角度开始了如火如荼的研究：Wu Y 等[37] 提出了一种识别可疑比特币地址的框架，可以发现犯罪网络并提供可视化功能；Bousfield D[38] 从经济学和网络演化的角度对加密货币，特别是比特币及其替代货币的持久性和可行性进行了分析。

本研究还参考王发明等[5] 对于区块链应用研究的划分，将区块链应用研究划分为 3 大类型，也将挖掘出的主题与之对应：①区块链 1.0，诸如虚拟数字货币等对于区块链的传统应用（对应 Topic4 加密货币）；②区块链 2.0，主要涉及智能合约的使用，例如在证券登记、期货、票据等金融市场的应用（对应 Topic3 商务智能合约）；③区块链 3.0，区块链在其他更广阔的领域的应用，特别是用于解决各领域的信任、共享等问题（对应其余的 6 个主题）。不难发现，上述两个演化特点明显的主题，正是代表了区块链 1.0 和区块链 2.0 的演化特点。为了解析当今学界对于各个类别研究的占比情况，将 2020 年各主题强度求和，代表各个类别的应用研究的热度，结果如图 8 所示。

图 8 2020 年各类别区块链研究强度占比图

从图中可以看出，如今对于区块链的研究已经不只局限于诸如“加密货币”“智能合约”等传统领域，而是拓展到其他领域，正如对图 7 分析得到的结论一样，区块链的研究呈现多样化的态势。如今对于区块链 3.0 的研究如火如荼，但区块链 1.0 和区块链 2.0 的研究并未消亡，究其原因，正是对于其传统领域应用研究的逐步深入，带动和启发了

更多应用场景的实施，我国的学者在拓展更多应用场景的同时，也要关注其技术发展带来的新特性，有针对性地寻找其新的应用场景。

4 结 语

4.1 结 论

本研究收集 Web of Science 核心合集 SCI-EXPANDED 和 SSCI 中 2008—2020 年区块链领域的文献，运用 LDA 主题建模，从热点主题和主题演化两个方面对国外区块链研究进行了分析，得出以下结论。

1）国际区块链研究自 2008 年开始，经过 10 余年的发展，如今已经形成非常丰富的概念内涵。国际学者比较关心的区块链研究领域包括商业智能合约、数字货币、数据隐私保护、能源交易与共识算法、物联网安全、工业供应链、车联网安全、医疗健康等。这些研究极大扩展了区块链的内涵，也奠定了该领域的理论与实践研究基础。

2）在全部国际区块链研究主题中，医疗健康、数据隐私保护、能源交易与共识算法和物联网安全 4 个主题的主题强度高于阈值，即 4 个主题作为当下区块链研究的热点主题，代表着国际区块链领域学者最关心的热点话题。在未来一段时间内仍然是区块链研究中的热点。

3）商务智能合约和加密货币两个主题都是在区块链技术发展早期出现，其共同演化特征都是在相关重要文献发表之后开始受到更多关注，从此研究热度开始上升。另外，也发现了区块链领域研究早期的话题大多与区块链本身技术相关，如加密货币和数据隐私保护；而到了发展后期，其研究热点开始向应用研究转移，如医疗健康、车联网等。

4）从主题分布上看，国际区块链领域主题热度分布近年来逐渐趋于均衡，说明领域研究的结构相较于早期已开始变得稳定。

4.2 建 议

基于本研究的发现，结合上述分析结论与我国区块链领域研究现状和行业发展需求，提出以下建议。

1）重视国外研究成果，从中获取国外区块链研究前沿，以此指导我国学者、企业界相关从事者抓住区块链领域的发展现状，追踪最前沿的研究热点。本研究对国际区块链文献进行主题分析，结果正是国外当前的研究热点，了解、分析这些热点出现的背景以及对社会、经济的影响，可以快速了解国外区块链研究的现有布局，以提升我国研究的战略视野和竞争力。

2）加快研究成果的转化与落地。本研究展示国际区块链研究从早期的纯技术理论研究逐渐转向了应用研究。由于区块链的产业价值更多体现在市场应用方面，解决具体社会、经济问题，因此国际研究兴趣的转变说明国际学者开始更多地关注区块链技术的市场化和产业化，而在这方面，我国能力较弱。应当加强高校与企业之间的合作创新，加速科研成果的技术、应用转化，促进区块链研究价值最大化。

4.3 不 足

本研究的不足之处是数据源较为单一，只选取了期刊数据库，如今技术迭代加快，高质量的会议论文也具有很高的研究价值。未来考虑结合会议论文、专利和替代计量学指标，对主题进行深度的挖掘，并结合深度学习算法，进行技术发展的预测研究。

参 考 文 献

[1] Nakamoto S. Bitcoin: A Peer-to-Peer Electronic Cash System [J]. Cryptography Mailing List at https://metzdowd.com, 2009.

[2] 王江，束正琦. 区块链何去何从?——基于网络分析 [J]. 科技管理研究，2020，40（22）：32-38.

[3] Firdaus A，Razak M F，Feizollah A，et al. The Rise of “Blockchain”: Bibliometric Analysis of Blockchain Study [J]. Scientometrics，2019，120（3）：1289-1331.

[4] Dabbagh M，Sookhak M，Safa N S. The Evolution of Blockchain: A Bibliometric Study [J]. IEEE Access，2019，（7）：19212-19221.

[5] 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.

[6] 汪园，王学东，李金鑫. 基于文献计量的我国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.

[7] 花敏，卢恒. 基于科学知识图谱的国内外区块链研究热点分析 [J]. 情报科学，2020，38（11）：70-79.

[8] Chen H，Zhao J. ISTopic: Understanding Information Systems Research Through Topic Models [J]. Social Science Electronic Publishing，2015.

[9] 赵紫娟，李小珂，郭强，等. 基于 LDA 的复杂网络整体研究

态势主题分析 [J]. 电子科技大学学报，2019，48（6）：931-938.

[10] 李跃艳，王昊，邓三鸿，等. 近十年信息检索领域的研究热点与演化趋势研究—— 基于 SIGIR 会议论文的分析 [J]. 数据分析与知识发现，2021，5（4）：1-14.

[11] 邓淑卿，徐健. 我国情报学研究主题内容分析 [J]. 情报科学，2017，35（11）：83-88.

[12] 吴查科，王树义. 基于 LDA 的国内图书馆学研究主题发现及演化研究 [J]. 新世纪图书馆，2019，（7）：90-96.

[13] Blei D M，Ng A Y，Jordan M I. Latent Dirichlet Allocation [J]. J. Mach. Learn. Res.，2003，3：993-1022.

[14] 李航. 统计学习方法（第 2 版） [M]. 北京：清华大学出版社，2019.

[15] 张健伟. 主题模型 LDA 推理算法对比与改进研究 [D]. 苏州：苏州大学，2017.

[16] 孙孟孟. 基于名词短语提取与词条权重分析的话题提取算法研究 [D]. 杭州：浙江大学，2014.

[17] 张斌. 国内共享科研数据热点主题及演化分析：从主题模型视角 [J]. 图书馆学研究，2020，（14）：11-18.

[18] 单斌，李芳. 基于 LDA 话题演化研究方法综述 [J]. 中文信息学报，2010，24（6）：43-49.

[19] Tranfield D，Denyer D，Smart P. Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review [J]. British Journal of Management，2003，14（3）：207-222.

[20] Webster J，Watson R T. Analyzing the Past to Prepare for the Future：Writing a Literature Review [J]. MIS Quarterly，2002，26（2）：xiii-xxiii.

[21] 马永红，孔令凯，林超然，等. 基于专利挖掘的关键共性技术识别研究 [J]. 情报学报，2020，39（10）：1093-1103.

[22] 关鹏，王曰芬. 科技情报分析中 LDA 主题模型最优主题数确定方法研究 [J]. 现代图书情报技术，2016，（9）：42-50.

[23] Lei A，Cao Y，Bao S，et al. A Blockchain Based Certificate Revocation Scheme for Vehicular Communication Systems [J]. Future Generation Computer Systems-The International Journal of Escience，2020，110：892-903.

[24] Ahn N，Lee D H. Physical Layer Security of Autonomous Driving：Secure Vehicle-to-Vehicle Communication in A Security Cluster [J]. Ad Hoc & Sensor Wireless Networks，2019，45（3-4）：293-336.

[25] Lei A，Cruickshank H，Cao Y，et al. Blockchain-Based Dynamic Key Management for Heterogeneous Intelligent Transportation Systems [J]. IEEE Internet of Things Journal，2017，4（6）：1832-1843.

[26] Dhagarra D，Goswami M，Sarma P R S，et al. Big Data and Blockchain Supported Conceptual Model for Enhanced Healthcare Coverage the Indian Context [J]. Business Process Management Journal，2019，25（7）：1612-1632.

[27] Zhang P，White J，Schmidt D C，et al. FHIRChain：Applying Blockchain to Securely and Scalably Share Clinical Data [J]. Computational and Structural Biotechnology Journal，2018，16：267-278.

[28] 朱光，刘蕾，李凤景. 基于 LDA 和 LSTM 模型的研究主题关联与预测研究——以隐私研究为例 [J]. 现代情报，2020，40（8）：38-50.

[29] Huang P，Fan K，Yang H，et al. A Collaborative Auditing Blockchain for Trustworthy Data Integrity in Cloud Storage System [J]. IEEE Access，2020，（8）：94780-94794.

[30] Yang X，Pei X，Wang M，et al. Multi-Replica and Multi-Cloud Data Public Audit Scheme Based on Blockchain [J]. IEEE Access，2020，（8）：144809-144822.

[31] Hayes B P，Thakur S，Breslin J G. Co-simulation of Electricity Distribution Networks and Peer to Peer Energy Trading Platforms [J]. International Journal of Electrical Power & Energy Systems，2020，115：105419.

[32] Cai W，Jiang W，Xie K，et al. Dynamic Reputation-based Consensus Mechanism：Real-time Transactions for Energy Blockchain [J]. International Journal of Distributed Sensor Networks，2020，16（3）：1-13.

[33] Rathore S，Kwon B W，Park J H. BlockSecIoTNet：Blockchain-based Decentralized Security Architecture for IoT Network [J]. Journal of Network and Computer Applications，2019，143：167-177.

[34] Sahay R，Geethakumari G，Mitra B. A Novel Blockchain Based Framework to Secure IoT-LLNs Against Routing Attacks [J]. Computing，2020，102（11）：2445-2470.

[35] Chang S E，Chen Y，Wu T. Exploring Blockchain Technology in International Trade Business Process re-engineering for Letter of Credit [J]. Industrial Management & Data Systems，2019，119（8）：1712-1733.

[36] Eenmaa-Dimitrieva H，Schmidt-Kessen M J. Creating Markets in No-trust Environments：The Law and Economics of Smart Contracts [J]. Computer Law & Security Review，2019，35（1）：69-88.

[37] Wu Y，Luo A，Xu D. Identifying Suspicious Addresses in Bitcoin Thefts [J]. Digital Investigation，2019，31：200895.

[38] Bousfield D. Crypto-coin Hierarchies：Social Contestation in Blockchain Networks [J]. Global Networks-A Journal of Transnational Affairs，2019，19（3）：291-307.

（责任编辑：孙国雷）